TOPICAL REVIEW

Stochastic switching in biology: from genotype to phenotype

Paul C. Bressloff[1]

[1]Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City UT 84112

Abstract. There has been a resurgence of interest in non-equilibrium stochastic processes in recent years, driven in part by the observation that the number of molecules (genes, mRNA, proteins) involved in gene expression are often of order 1-1000. This means that deterministic mass-action kinetics tends to break down, and one needs to take into account the discrete, stochastic nature of biochemical reactions. One of the major consequences of molecular noise is the occurrence of stochastic biological switching at both the genotypic and phenotypic levels. For example, individual gene regulatory networks can switch between graded and binary responses, exhibit translational/transcriptional bursting, and support metastability (noise-induced switching between states that are stable in the deterministic limit). If random switching persists at the phenotypic level then this can confer certain advantages to cell populations growing in a changing environment, as exemplified by bacterial persistence in response to antibiotics. Gene expression at the single-cell level can also be regulated by changes in cell density at the population level, a process known as quorum sensing. In contrast to noise-driven phenotypic switching, the switching mechanism in quorum sensing is stimulus-driven and thus noise tends to have a detrimental effect. A common approach to modeling stochastic gene expression is to assume a large but finite system and to approximate the discrete processes by continuous processes using a system-size expansion. However, there is a growing need to have some familiarity with the theory of stochastic processes that goes beyond the standard topics of chemical master equations, the system-size expansion, Langevin equations and the Fokker-Planck equation. Examples include stochastic hybrid systems (piecewise deterministic Markov processes), large deviations and the Wentzel-Kramers–Brillouin (WKB) method, adiabatic reductions, and queuing/renewal theory. The major aim of this review is to provide a self-contained survey of these mathematical methods, mainly within the context of biological switching processes at both the genotypic and phenotypic levels. However, applications to other examples of biological switching are also discussed, including stochastic ion channels, diffusion in randomly switching environments, bacterial chemotaxis, and stochastic neural networks.

Submitted to: *J. Phys. A: Math. Gen.*

Contents

1. Introduction

Intrinsic variability in gene expression due to small numbers of molecules (low copy numbers) is prevalent in many cellular processes. Hence, genetically identical cells exposed to the same environmental conditions can show significant variation in molecular content and significant differences in phenotypic characteristics. Although intrinsic variability was originally viewed as being detrimental to cellular function, and a potential cause of certain diseases, it has now been realized that intrinsic noise may provide the flexibility needed by cells to adapt to fluctuating environments or respond to sudden stresses, and could also be a mechanism whereby population heterogeneity is established during cell differentiation and development. Following the discovery of a functional role for stochastic gene expression in λ-phage [14], there has been an explosion of studies on the molecular mechanisms of noise generation at the single gene level, and its effects on larger regulatory networks, see the reviews [145,173,202,211,220,241]. *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.

One of the major features of stochasticity in gene expression at the molecular level are various forms of noise-induced switching, including metastability in nonlinear feedback networks, translational and transcriptional bursting, and binary responses to external stimuli. Switching at the genotypic level is also thought to play a role at the phenotypic level, as exemplified by bacterial persistence in randomly switching environments, and quorum sensing. In order to model these phenomena, there is a growing requirement to have some familiarity with the theory of stochastic processes that goes beyond the standard topics of chemical master equations, the system-size expansion, stochastic differential equations (SDEs) and the Fokker-Planck (FP) equation. Examples include stochastic hybrid systems, large deviation theory, the Wentzel-Kramers–Brillouin (WKB) method, adiabatic reductions, singular perturbation theory, and queuing theory. The aim of this review is to provide a self-contained survey of these methods within the context of genetically-based biological switching.† In the remainder of this introduction, we provide a brief biological

† There are some obvious parallels with neuroscience [161, 232], since individual neurons can act as binary switches, exhibit various forms of bursting [64], and contribute to stochastic dynamics at the population level [41, 42, 57, 58]. There is also a distinction between intrinsic noise (driven mainly by the stochastic opening and closing of ion channels) and extrinsic noise (synaptic inputs from other neurons). However, since including such topics would double the length of the review, we have decided to focus on genetic-based switching. Nevertheless, we touch on some of the connections with other areas of biology in section 8.

background and summarize the contents of the review.

1.1. Biological background

We begin by summarizing the *central dogma* of molecular biology, which describes the fundamental process whereby proteins are synthesized from DNA [69]. There are two main stages involved in the expression of a single gene, see Fig. 1(a).

(i) *Transcription*: synthesis of a *messenger RNA* molecule (mRNA) 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*. 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. (Prokaryotes such a bacteria do not have a cell nucleus.) The key steps in transcription are binding of RNA polymerase (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 RNA polymerase reads one of the exposed strands. These steps can be represented by the reaction scheme

$$P + D \underset{k_-}{\overset{k_+}{\rightleftharpoons}} PD_c \overset{k_{\rm open}}{\longrightarrow} PD_0 \overset{k_{\rm escape}}{\longrightarrow} \text{transcription}.$$

The binding and unbinding of polymerase is very fast, $k_\pm \gg k_{\rm open}$ so that the first step happens many times before formation of an open complex. Hence, one can treat the RNA polymerase as in quasi-equilibrium with the promoter characterized by an

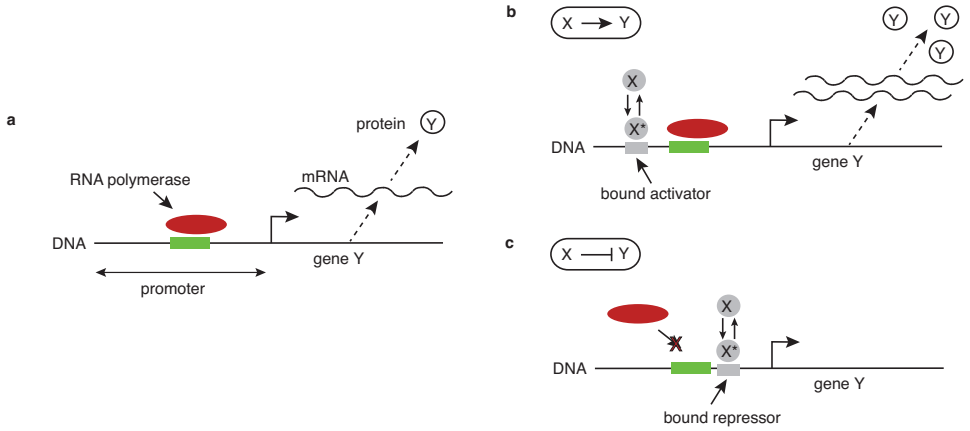


Figure 1. Transcriptional regulation due to the binding of a repressor or activator protein to a promoter region along the DNA. (a) Unregulated transcription of a gene Y following binding of RNA polymerase to the promoter region. The resulting mRNA exits the nucleus and is then translated by ribosomes 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 stopped by a repressor protein X binding to the promoter and blocking the binding of RNA polymerase.

equilibrium constant $K_P = k_+/k_-$. The rate of transcription will thus be proportional to the fraction of bound RNA polymerase, $k_+/(k_+ + k_-)$.

(ii) *Translation*: 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, 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. (Note that there are also non-protein coding genes such as small nuclear RNA (snRNA) whose products are functional RNA.)

A major requirement of cellular organisms is the ability to increase or decrease the level of particular gene products (protein or RNA) at different times and spatial locations, a process known as *gene regulation*. Almost every step of gene expression can be regulated, including the initiation and termination of transcription, RNA processing, and the post-translational modification of a protein. The control of transcription (switching on or off a gene) is mediated by DNA-binding proteins known as transcription factors, see Fig. 1(b,c). Positive control (activation) is mediated by activators that increase the probability of RNA polymerase binding to the promoter, whereas negative control (repression) is mediated by repressors that bind to a promoter region along the DNA where RNA polymerase has to bind in order to initiate transcription - it thus inhibits transcription. The presence of transcription factors means that cellular processes can be controlled by extremely complex gene networks with multiple negative and positive feedback loops. Identifying functional modules and motifs within such networks is a major component of systems biology [2]. Gene regulation plays an essential role in viruses and bacteria, since it increases the versatility and adaptability of an organism within an environment. In multicellular organisms, gene regulation drives cellular differentiation and morphogenesis in the embryo, producing different cell types with different gene expression profiles from the same genome sequence.

Whenever the number of molecules (genes, mRNA, proteins) is large, one can represent the dynamics of a gene regulatory network in terms of a system of coupled ordinary differential equations (ODEs). The resulting dynamical system, which describes the evolution of the associated molecular concentrations based on mass-action reaction kinetics, can exhibit multistability (genetic switches) and limit cycle oscillations (genetic oscillators) in the presence of nonlinear feedback, both of which are thought to play an important role in cell function. However, one typically finds low copy numbers of at least some of the molecular players, so that the discrete and stochastic nature of the dynamics needs to be taken into account. For example, 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 means that on average there are 10 mRNA molecules per cell. Generation of the mRNA molecule occurs at a rate of 50 nucleotides per second. Hence, a typical gene of around 1000 nucleotides will be transcribed in about 20 seconds.

Thus there are around 3 RNA polymerase per gene at any one time, suggesting the number fluctuations will be significant. A major motivation for the explosion of interest in stochastic gene expression has been the development of experimental methods to observe the real-time production of single protein molecules in living cells. For example Yu *et al.* [256] used fluorescent imaging to observe the expression of individual protein molecules in single *Escerichia* cells under the control of a repressed *lac* promotor. They found that proteins were produced in bursts from stochastically transcribed single mRNA molecules, with the distribution of the protein copy number per burst following a geometric distribution. This so-called translational bursting is a consequence of the fact that the lifetime of mRNA is usually much shorter than that of a protein, and each mRNA produces a random integer number of proteins before degrading [59, 101, 178]. This results in a broadening of the distribution of proteins compared to a simple Poisson process. Bursts of transcriptional activity have also been observed, which can be caused by extrinsic factors such as fluctuations in the number

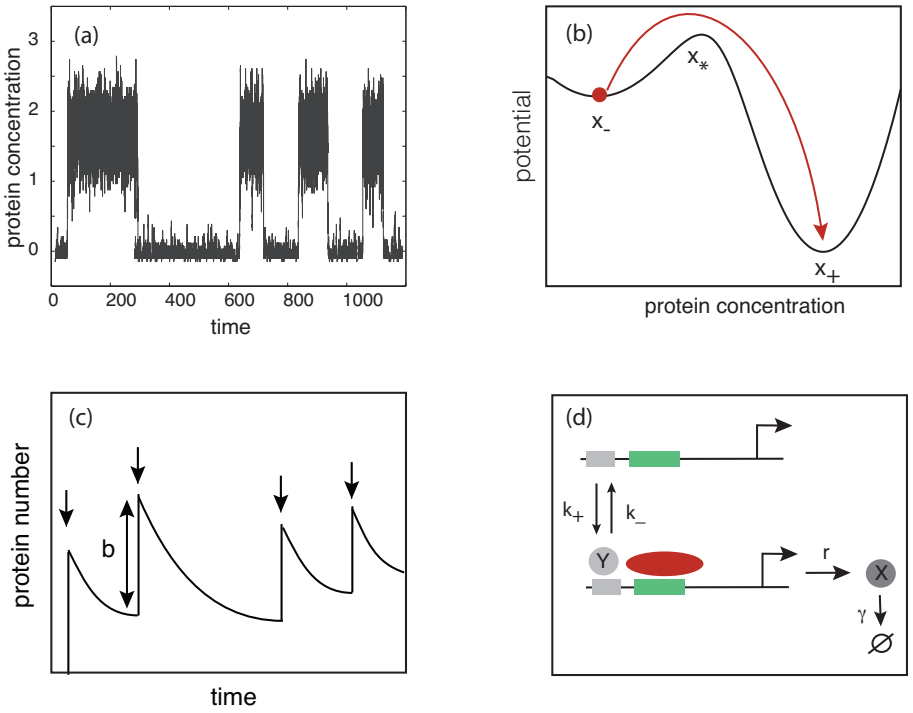


Figure 2. Effects of noise in gene expression. (a) Schematic illustration of the time course of the level of protein in a single cell, which randomly switches between low- and high-expression states. Time is in units of the inverse protein degradation rate. (b) Bistable potential for a simple genetic switch. In the absence of noise, the system evolves to one of the minima $x_\pm$ of the potential, depending on initial conditions. Intrinsic noise can induce a transition between the minima (metastability) by crossing over the potential barrier at x_*. (c) 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. (d) One source of transcriptional bursting. A gene randomly switches between an on-state and an off-state at rates $k_\pm$. In the on-state proteins are produced at a rate r and degrade at a rate γ.

of RNA polymerase or intrinsic fluctuations in regulatory proteins. Finally, noise can also induce transitions between metastable states of a genetic switch [106]; in the absence of noise, the particular state of the switch is determined by initial conditions. A simple conceptual framework for understanding metastability is in terms of the stochastic dynamics of a particle in a double potential well. Various examples of noise-induced switching at the molecular level are shown in Fig. 2.

Fluctuations arising from low copy numbers are usually identified as a form of *intrinsic noise*, whereas fluctuations in cellular environmental factors such as the activity of ribosomes and polymerases, metabolite concentrations, cell size, cell age and stage of the cell cycle are typically treated as a form of *extrinsic noise* [14,86]. Elowitz et al. [86] developed a *two-reporter assay* that can discriminate between extrinsic and intrinsic noise 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. 3a,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 are 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. There are, however, some potential limitations

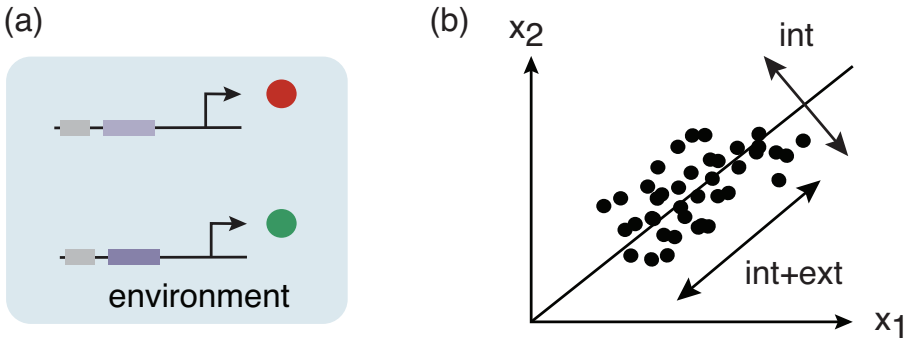


Figure 3. Measuring intrinsic and extrinsic noise in gene expression. (a) Two almost identical genes, which encode red and green fluorescent proteins, are expressed from identical promoters, and are influenced identically by cell-specific factors, such as gene-regulatory signals. Cells with equal amounts of the two proteins appear yellow, indicating that the level of intrinsic noise is low. 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. (b) Scatter plot of fluorescence in a population. 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.

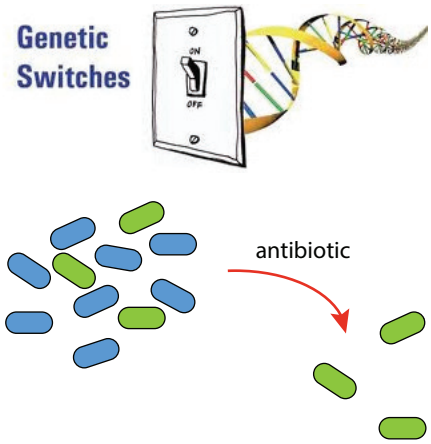


Figure 4. Stochastic genetic switching at the single cell level can lead to phenotypic switching at the population level. This is one mechanism for bacterial persistence, where a slow-growing drug resistant phenotype (green) coexists with a fast-growing phenotype (blue) that is susceptible to the antibiotic.

of the experimental protocol. 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.

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. For example, switching between phenotypic states with different growth rates might be an important factor in the phenomenon of persistent bacterial infections after treatment with antibiotics, see Fig. 4. Although most of the population is rapidly killed by the treatment, a small genetically identical subset of dormant persister cells can survive an extended period of exposure. When the drug treatment is removed, the surviving persisters randomly transition out of the dormant state, causing the infection to reemerge.

Another example of switching at the population level is *quorum sensing*, which is a form of system stimulus and response that is correlated to population density. Many species of bacteria use quorum sensing to coordinate various types of behavior including bioluminescence, biofilm formation, virulence, and antibiotic resistance, based on the local density of the bacterial population [71,80,102,163,181,233,250,252]. In an analogous fashion, some social insects use quorum sensing to determine where to nest [66]. Roughly speaking, quorum sensing can function as a decision-making process in any decentralized system, provided that individual components have (i) some mechanism for determining the number or density of other components they interact with and (ii) a stereotypical response once some threshold has been reached. In the case of bacteria, quorum sensing involves the production and extracellular secretion of certain signaling molecules called *autoinducers*. Each cell also has receptors that can

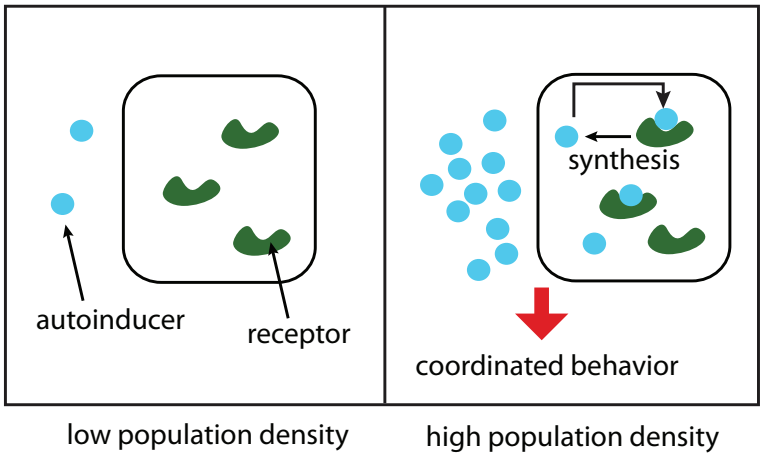


Figure 5. A schematic illustration of quorum sensing at the single-cell level.

specifically detect the signaling molecule (inducer) via ligand-receptor binding, which then activates transcription of certain genes, including those for inducer synthesis. However, since the concentration of autoinducer produced by a single cell is likely to be below the threshold for detection, the cell must encounter signaling molecules secreted by other cells in its environment in order for gene transcription to be activated. When only a few other bacteria of the same kind are in the vicinity (low bacterial population density), diffusion reduces the concentration of the inducer in the surrounding medium to almost zero, resulting in small amounts of inducer being produced. On the other hand, as the population grows, the concentration of the inducer passes a threshold, causing more inducer to be synthesized. This generates a positive feedback loop that fully activates the receptor, and induces the up-regulation of other specific genes. Hence, all of the cells initiate transcription at approximately the same time, resulting in some form of coordinated behavior. The basic process at the single-cell level is shown in Fig. 5.

1.2. Organization of the review

In the following sections, we survey various topics in stochastic processes that are currently being used to investigate various aspects of genetically-based biological switching. We begin by reviewing discrete Markov processes and stochastic hybrid systems, which provide the mathematical underpinning of stochastic models of gene expression (section 2). We consider two important examples of discrete Markov processes, namely Poisson processes and chemical reaction networks. In the latter case we describe how to reduce the corresponding discrete Markov process to a continuous one using some form of perturbation expansion (diffusion approximation), which exploits either a system-size expansion or a separation of time scales. We also describe the stochastic simulation algorithm (SSA) introduced by Gillespie to simulate sample trajectories of a gene or biochemical network, and discuss various extensions, including a numerical scheme for stochastic hybrid systems. Some simple examples of switching in gene regulatory networks are presented in section 3. We first consider in some detail the behavior of a feedforward regulatory network operating in a switching environment, and show how it can exhibit both graded and binary

responses. We show how the correlated statistics of a population of such networks operating in the same environment can be determined in terms of moments of a piecewise deterministic partial differential equation (PDE). We then consider a few simple examples of bistability in nonlinear feedback regulatory networks, including an autoregulatory network and a mutual repressor network. We also discuss some spatial aspects of gene regulation. In section 4 we consider various mathematical models of translational and transcriptional bursting, and provide a brief introduction to queuing theory. In section 5 we provide an introduction to WKB methods and large deviations for analyzing escape problems in SDEs. We then illustrate how to extend these methods to metastability in a simple autoregulatory gene network. The latter is either modeled in terms of a discrete Markov process or a stochastic hybrid system. We turn to stochastic switching at the phenotypic level in section 6, where we explore mathematical models of how bacterial populations adapt to randomly fluctuating environments by stochastic phenotype-switching mechanisms. In section 7 we consider models of bacterial quorum sensing, and introduce another switching mechanism based on *ultrasensitivity* in phosphorylation-dephosphorylation cycles. We describe how ultrasensitivity can lead to noise amplification, which is then mitigated by population averaging within a quorum sensing system. Finally, in section 8, we present some other examples of stochastic switching in biology that share the same basic mathematical structures as highlighted in this review. Examples include stochastic ion channels and spontaneous neuronal firing, diffusion in randomly switching environments, bacterial chemotaxis, and stochastic neural networks.

We have tried to make the review as self-contained as possible. However, we do assume some working knowledge of continuous Markov processes [105], in particular the interplay between Fokker-Planck equations and stochastic differential equations. We also only consider very simple gene networks, in order to develop the basic mathematical and conceptual tools. One of the outstanding challenges is finding ways to apply these methods to more complicated gene regulatory networks with multiple feedback loops. This will require combining various model reduction schemes with efficient numerical algorithms. One important aspect of biological switching that we do not consider is its potential computational role [56,126]. For once one has a genetic switch, one can start constructing synthetic gene circuits out of logic gates, by analogy with binary neural switches and silicon-chip computers. Finally, for a very readable introduction to biological aspects of genetic switches, as exemplified by the λ-phage, see the book by Ptashne [206].

2. Discrete Markov processes and stochastic hybrid systems.

One of the driving forces for developing stochastic models of gene expression has been the observation that the number of molecules (genes, mRNAs, proteins) can be small so that the discrete nature of the underlying processes must be taken into account. This naturally leads to the theory of discrete Markov processes and chemical master equations. Sometimes, at least one component can be treated as a continuous variable (eg. the concentration of proteins), which is coupled to the remaining set of discrete processes, resulting in a so-called piecewise deterministic Markov process (PDMP). The continuous variable evolves deterministically according to mass-action kinetics except at discrete times where there is a jump in one of the discrete variables resulting in a modified kinetic equation. However, the continuous process could also be a stochastic differential equation (SDE) or a partial differential equation (PDE), so

we will refer to the general class of models as stochastic hybrid systems (SSHs), a term also found in the control theory literature. In this section we provide a mathematical introduction to discrete Markov processes and stochastic hybrid systems, which will serve as a useful background in the remainder of the review. For a much more extensive discussion of discrete Markov processes see the book by Grimmett and Stirzaker [121]. A rigorous introduction to PDMPs can be found in Refs. [70, 153], whereas a wide range of applications of hybrid systems to cell biology can be found in Ref. [45].

2.1. Markov chains

We start with discrete-time processes. Let $\{X_0, X_1, \ldots\}$ be a sequence of discrete random variables that take one of N_0 values in some countable set Γ with $N_0 = |\Gamma|$. (Note that it is possible for $N_0 = \infty$.) Without loss of generality, we take $\Gamma = \{0, 1, \ldots, N_0 - 1\}$. The stochastic process X is said to be a *Markov chain* if it satisfies the Markov property

$$\mathbb{P}[X_l = n|X_0, \ldots, X_{l-1}] = \mathbb{P}[X_l = n|X_{l-1}]$$

for all discrete times $l \geq 1$ and $n \in \Gamma$. The evolution of the chain is described by its transition probabilities

$$K_{nm} = \mathbb{P}[X_{l+1} = n|X_l = m]$$

for all n, m, l.† For simplicity, we assume that the Markov chain X is homogeneous, that is, K_{nm} is independent of discrete time l. $\mathbf{K}$ is a stochastic matrix, since it satisfies the following two properties:

(i) $\mathbf{K}$ has non-negative entries, $K_{nm} \geq 0$

(ii) $\sum_n K_{nm} = 1$.

Introduce the l-step transition matrix $\mathbf{K}(l)$ with components

$$K_{nm}(l) = \mathbb{P}[X_{l_0+l} = n|X_{l_0} = m].$$

Clearly $\mathbf{K}(1) = \mathbf{K}$. We can now derive a discrete version of the Chapman-Kolomogorov equations:

$$\begin{aligned} K_{nm}(l + l') &= \mathbb{P}[X_{l+l'} = n|X_0 = m] = \sum_k \mathbb{P}[X_{l+l'} = n, X_l = k|X_0 = m] \\ &= \sum_k \mathbb{P}[X_{l+l'} = n|X_l = k, X_0 = m]\mathbb{P}[X_l = k|X_0 = m] \\ &= \sum_k \mathbb{P}[X_{l+l'} = n|X_l = k]\mathbb{P}[X_l = k|X_0 = m]. \end{aligned}$$

from the Markov property, and hence

$$K_{nm}(l + l') = \sum_k K_{nk}(l')K_{km}(l). \tag{2.1}$$

In matrix form, the Chapman-Kolomogorov equations (2.1) imply that $\mathbf{K}(l + l') = \mathbf{K}(l)\mathbf{K}(l')$ and hence $\mathbf{K}(l) = \mathbf{K}^l$. Finally, introducing the l-dependent probability $P_l(n) = \mathbb{P}[X_l = n]$, we have

$$\begin{aligned} P_l(n) = \mathbb{P}[X_l = n] &= \sum_m \mathbb{P}[X_l = n|X_0 = m]P_0(m) \\ &= \sum_m K_{nm}(l)P_0(m) = \sum_m [\mathbf{K}^l]_{nm}P_0(m). \end{aligned}$$

† Probabilists usually write the transition matrix in the adjoint form, that is, $K_{mn} = \mathbb{P}[X_{l+1} = n|X_l = m]$.

This establishes that the probability vector $\mathbf{P}_l = (P_l(1), \ldots P_l(N))$ can be determined from the initial probability vector $\mathbf{P}_0$ according to $\mathbf{P}_l = \mathbf{K}^l \mathbf{P}_0$.

Recurrent and transient states. A state n of a Markov chain is called *recurrent* or *persistent* if

$$\mathbb{P}[X_l = n \text{ for some } l \geq 1 | X_0 = n] = 1.$$

That is, the probability of eventual return to n, having started from n is unity. If the probability of return is strictly less than one, then n is called *transient*. Let

$$f_{nm}(l) = \mathbb{P}[X_1 \neq n, X_2 \neq n, \ldots, X_{l-1} \neq n, X_l = n | X_0 = m]$$

be the probability that the first visit to state n, starting from m, occurs at the lth step. This is analogous to the first passage time distribution. It follows that

$$f_{nm} = \sum_{l=1}^{\infty} f_{nm}(l)$$

is the probability that the chain ever visits n starting from m. We see that the state n is persistent if and only if $f_{nn} = 1$. In order to derive a criterion for persistence, it is useful to introduce the generating functions

$$G_{nm}(z) = \sum_{l=0}^{\infty} z^l K_{nm}(l), \quad F_{nm} = \sum_{l=1}^{\infty} z^l f_{nm}(l).$$

Here $G_{nm}(0) = \delta_{nm}$ and $F_{nm}(0) = 0$ for all n, m. Clearly $f_{nm} = F_{nm}(1)$. We now derive an important relationship between the two generating functions. Let A be the event that $X_l = n$ given $X_0 = m$, and let $B_{nm}(s)$ be the event that the first arrival at n occurs at the sth step. Then

$$\mathbb{P}[A] = \sum_{s=1}^{l} \mathbb{P}[A | B_{nm}(s)] \mathbb{P}[B_{nm}(s)].$$

Since $\mathbb{P}[B_{nm}(s)] = f_{nm}(s)$ and $\mathbb{P}[A | B_{nm}(s)] = K_{nn}(l - s)$, we have

$$K_{nm}(l) = \sum_{s} f_{nm}(s) K_{nn}(l - s).$$

Multiplying both sides by z^l and summing over l shows that

$$\begin{aligned} \sum_{l=1}^{\infty} z^l K_{nm}(l) &= \sum_{l=1}^{\infty} \sum_{s=1}^{l} (z^s f_{nm}(s))(z^{l-s} K_{nn}(l - s)) \\ &= \sum_{s=1}^{\infty} \sum_{l=s}^{\infty} (z^s f_{nm}(s))(z^{l-s} K_{nn}(l - s)), \end{aligned}$$

that is,

$$G_{nm}(z) = \delta_{nm} + F_{nm}(z) G_{nn}(z). \tag{2.2}$$

Equation (2.2) allows us to derive a condition for persistence of a state n. If $|z| < 1$, then $G_{nn}(z) = (1 - F_{nn}(z))^{-1}$. Hence, if $f_{nn} = F_{nn}(1) = 1$ (persistence), then as $z \to 1^-$, we have $G_{nn}(z) \to \infty$. We now use a well known theorem from power series known as Abel's Theorem.

Theorem 1 *Let $\{a_k\}$ be a sequence of real or complex numbers and let*

$$G(z) = \sum_{k=0}^{\infty} a_k z^k,$$

with z real (or in a restricted region of the complex plane). If the series $\sum_k a_k$ is convergent, then

$$\lim_{z \to 1^-} G(z) = \sum_{k=0}^{\infty} a_k.$$

An application of Abels' Theorem establishes that n is persistent if $\sum_l K_{nn}(l) = \infty$. It then follows from equation (2.2) for $m \neq n$ that $\sum_l K_{nm}(l) = \infty$ for all m such that $f_{nm} > 0$. An immediate consequence of this result is that n is transient if $\sum_l K_{nn}(l) < \infty$ and if this holds then $\sum_l K_{nm}(l) < \infty$ for all m. Moreover, if n is transient then $K_{nm}(l) \to 0$ as $l \to \infty$.

Mean recurrence time. Given that $X_0 = m$, let

$$T_{nm} = \min\{l \geq 1 : X_l = n\}$$

be the time of the first visit to n, with the convention that $T_{nm} = \infty$ if the visit never occurs. Define the *mean recurrence time* μ_n of the state n as

$$\mu_n = \mathbb{E}[T_{nn}] = \begin{cases} \sum_l l f_{nn}(l) & \text{if } n \text{ is persistent} \\ \infty & \text{if } n \text{ is transient} \end{cases}.$$

Note that μ_n may be infinite even when i is persistent, so a persistent state n is called *null* if $\mu_n = \infty$ and *non-null* or *positive* if $\mu_n < \infty$. It can be shown that a persistent state is null if and only if $K_{nn}(l) \to 0$ as $l \to \infty$ [121]. One final definition is useful: a state n is said to be periodic if there exists an integer $d(n) > 1$ such that $K_{nn}(l) = 0$ unless l is a multiple of $d(n)$. If $d(n) = 1$ then the state n is said to be aperiodic.

Irreducible Markov chains. Consider a set C of states. The set C is said to be *closed* if $K_{nm} = 0$ for all $m \in C$ and $n \notin C$. A closed set containing exactly one state is said to be *absorbing*, that is, if the Markov chain reaches the absorbing state then it stays there. The set C is said to be *irreducible* if any two states $m, n \in C$ can be connected together in a finite number of steps. That is, there exist $l, l \geq 0$ for which $K_{nm}(l)K_{mn}(l) > 0$. If the whole state space Γ is irreducible, then the Markov chain is said to be irreducible. Any state space Γ can be partitioned uniquely as $\Gamma = T \cup C_1 \cup C_2 \cup \ldots$, where T is a set of transient states and the C_r are irreducible closed sets of persistent states. In the case of a finite state space, not all states can be transient. For if they were, then $\sum_n K_{nm}(l) = 1$ but $K_{nm}(l) \to 0$ as $l \to \infty$ leading to the contradiction

$$1 = \lim_{l \to \infty} \sum_n K_{nm}(l) = 0!.$$

A similar argument shows that all persistent states are non-null.

Stationary distributions and the limit theorem. An important issue is whether or not a Markov chain converges to a unique stationary distribution, assuming the latter exists. A stationary distribution of a Markov chain is a vector $\mathbf{p}^*$ whose components satisfy

(i) $p^*(n) \geq 0$ for all n, $\quad \sum_n p^*(n) = 1$
(ii) $p^*(n) = \sum_m K_{nm} p^*(m)$ for all n.

It is stationary since $\mathbf{p}^* = \mathbf{K}^n \mathbf{p}^*$. Note that in the case of an irreducible chain, $p^*(n) > 0$ for all n. We then have the following theorem [121]:

Theorem 2 *An irreducible Markov chain has a stationary distribution $\mathbf{p}^*$ if and only if all the states are non-null persistent: $\mathbf{p}^*$ is the unique stationary distribution and $p^*(n) = \mu_n^{-1}$ for each in $\in \Gamma$, where μ_n is the mean recurrence time of n.*

In the case of a finite state space irreducibility is sufficient to establish existence of a unique stationary state. The next theorem establishes the link between the existence of a stationary distribution and the limiting behavior of $K_{nm}(l)$.

Theorem 3 *For an irreducible, aperiodic Markov chain (all states n have period $d(n) = 1$), we have that*

$$K_{nm}(l) \to \frac{1}{\mu_n} \text{ as } l \to \infty \text{ for all } m, n.$$

In the case of a transient or null persistent chain, the theorem is satisfied, since $K_{nm}(l) \to 0$ for all n, m and $\mu_n = \infty$. In the case of a non-null persistent Markov chain, we have

$$\lim_{l \to \infty} K_{nm}(l) \to p^*(n) = \frac{1}{\mu_n},$$

where $\mathbf{p}^*$ is the unique stationary distribution. It immediately follows that the limiting transition probability does not depend on its starting point $X_0 = m$ and

$$P_l(n) = \mathbb{P}[X_l = n] = \sum_m K_{nm}(l) P_0(m) \to \frac{1}{\mu_n} \text{ as } l \to \infty.$$

Theorem 3 is an example of an ergodic theorem for Markov chains.

Perron-Frobenius Theorem. The theory of Markov chains is considerably simplified if Γ is finite. For if Γ is irreducible then it is necessarily non-null persistent, and the probability distribution converges to a unique stationary distribution when it is aperiodic. This can also be established using the well-known Perron-Frobenius Theorem for finite square matrices.

Theorem 4 *If $\mathbf{K}$ is the transition matrix of a finite, irreducible chain with period d then*

(i) $\lambda_1 = 1$ *is an eigenvalue of* $\mathbf{K}$
(ii) the d complex roots of unity, $\lambda_k = \mathrm{e}^{2\pi i k/d}$ for $k = 0, 1, \ldots, d-1$, are eigenvalues of $\mathbf{K}$
(iii) the remaining eigenvalues $\lambda_{d+1}, \ldots, \lambda_N$ satisfy $|\lambda_j| < 1$.

The left eigenvector corresponding to λ_1 is $\psi = (1, 1, \ldots, 1)$, whereas the right-eigenvector is a stationary state $\mathbf{p}^*$. In the aperiodic case, there exists a simple real eigenvalue $\lambda_1 = 1$ and all other eigenvalues satisfy $|\lambda_j| < 1$. One can thus establish that the probability distribution converges to $\mathbf{p}^*$ in the large-n limit.

Continuous-time Markov chains. Let us now turn to continuous-time Markov chains whose probability distributions evolve according to a master equation. Let $X = \{X(t); t \geq 0\}$ be a family of random variables taking values in some subset Γ of the integers. X is called a *continuous-time Markov chain* if it satisfies the Markov property

$$\mathbb{P}[X(t_l) = n | X(t_1), \ldots, X(t_{l-1})] = \mathbb{P}[X(t_l) = n | X(t_{l-1})]$$

for any $n \in \Gamma$ and any sequence $t_1 < t_2 < \ldots < t_l$ of times. Introduce the transition probability (assuming a homogeneous chain)

$$K_{nm}(t) = \mathbb{P}[X(t+s) = n | X(s) = m],$$

which is independent of the initial time s. Let $\mathbf{K}_t$ denote denote the $N \times N$ matrix with entries $K_{nm}(t)$. The family of matrices $\{\mathbf{K}_t : t \geq 0\}$ satisfies the following:

(i) $\mathbf{K}_0 = \mathbf{I}$, the identity matrix;

(ii) $\mathbf{K}_t$ is stochastic, that is, it has non-negative entries and each column sums to unity, that is, $\sum_n K_{nm}(t) = 1$ for all $t \geq 0$ and m;

(iii) the Chapman-Kolmogorov equation

$$\mathbf{K}_{s+t} = \mathbf{K}_s \mathbf{K}_t \text{ if } s, t \geq 0.$$

The family $\{\mathbf{K}_t : t \geq 0\}$ is called a *stochastic semigroup*. We will restrict our discussion to standard semigroups, for which $\lim_{t \to 0^+} \mathbf{K}_t = \mathbf{I}$, which holds if and only if the elements $K_{nm}(t)$ are continuous functions of t.

Suppose that the chain is in state $X(t) = m$ at time t. One of two distinct events can then occur in the small time interval $(t, t + \Delta t)$:

(i) The chain is in the same state at time $t + \Delta t$ with probability $K_{mm}(\Delta t) + o(\Delta t)$;

(ii) The chain jumps to a new state n with probability $K_{nm}(\Delta t)$;

It can be shown that the probability of two or more transitions in the small time interval is $o(\Delta t)$.

Introduce the transition rates A_{nm} according to

$$K_{nm}(\Delta t) = A_{nm}\Delta t + o(\Delta t) \text{ for } m \neq n, \quad K_{mm}(\Delta t) = 1 + A_{mm}\Delta t + o(\Delta t). \tag{2.3}$$

We are assuming that the transition probabilities are differentiable at $t = 0$. From properties of the transition matrix $\mathbf{K}$ we see that $A_{nm} \geq 0$ for $m \neq n$, $A_{mm} \leq 0$ for all m, and $\sum_n A_{nm} = 0$ for all m. The matrix $-\mathbf{A}$ is known as the *generator* of the Markov process†. Introducing the probability distribution

$$P(n, t) = \mathbb{P}[X(t) = n],$$

we then have

$$P(n, t + \Delta t) = \sum_{m \neq n} A_{nm}\Delta t P(m, t) + (1 + A_{nn}\Delta t)P(n, t).$$

Rearranging this equation, dividing through by Δt and taking the limit $\Delta t \to 0$ yields the master equation for the continuous-time Markov chain:

$$\frac{dP}{dt} = \sum_m A_{nm} P(m, t). \tag{2.4}$$

† Note that probabilists define the generator in terms of the adjoint operator $-\mathbf{A}^\dagger$.

Using the fact that $A_{mm} = -\sum_{n\neq m} A_{nm}$ and setting $W_{nm} = A_{nm}$ for $n \neq m$, we can rewrite the master equation as

$$\frac{dP(n,t)}{dt} = \sum_{m\neq n} W_{nm}P(m,t) - \left(\sum_{k\neq n} W_{kn}\right) P(n,t). \tag{2.5}$$

This version of the master equation can be interpreted as follows: the first sum on the right-hand side involves all transitions $m \to n$, whereas the second sum involves all transitions $n \to k$ for $k \neq n$.

A stationary solution $\mathbf{p}^*$ of the master equation (2.4) is a right null-vector of $\mathbf{A}$, that is, $\sum_i A_{nm}p^*(m) = 0$. We then have the following ergodic theorem for continuous-time Markov chains:

Theorem 5 *Let X be irreducible with a standard semigroup $\{\mathbf{K}_t\}$ of transition probabilities.*

(i) If there exists a stationary distribution $\mathbf{p}^*$ then it is unique and

$$P(n,t) \to p^*(n) \text{ as } t \to \infty$$

for all n.

(ii) If there is no stationary distribution then $P(n,t) \to 0$ as $t \to \infty$ for all n.

The basic point is that in the case of a continuous-time Markov chain, periodicity is no longer an issue. For a finite chain, irreducibility is sufficient to guarantee existence of a unique stationary state.

2.2. Poisson processes

One important example of a continuous-time discrete Markov process is the Poisson process, which is one of the best-known examples of a so-called counting process. A random process $\{N(t),\, t \in [0,\infty)\}$ is said to be a *counting process* if $N(t)$ is the number of events that have occurred in the interval $[0,t]$ such that

(i) $N(0) = 0$;

(ii) $N(t) \in \{0,1,2,\ldots\}$ for all $t \geq 0$;

(iii) for $0 \leq s < t$, $N(t) - N(s)$ is the number of events in the interval $(s,t]$.

One example of an event is the random arrival of a customer at some service station resulting in the formation of a queue. (In section 4.3 queuing theory will be used to study transcriptional bursting in gene networks. Another example is the firing of an action potential spike when the membrane voltage of a neuron reaches threshold or the arrival of that spike at the synapse of another neuron, see section 8.) A counting process can be characterized in terms of the occurrence or arrival times T_i of the ith event. This leads to two further definitions useful in characterizing counting processes. First, let $\{X(t), t \in [0,\infty)\}$ be a continuous-time stochastic process. We say that $X(t)$ has *independent increments* if, for all $0 \leq t_1 < t_2 \cdots < t_n$, the random variables $X(t_j) - X(t_{j-1})$, $j = 2,\ldots,n$, are independent. In the case of a counting process this means that the numbers of arrivals in non-overlapping time intervals are independent. Following on from this, $X(t)$ is said to have *stationary increments* if, for all $t_2 > t_1 \geq 0$ and all $r > 0$, the random variables $X(t_2) - X(t_1)$ and $X(t_2 + r) - X(t_1 + r)$ have the same distributions. In particular, a counting process has stationary increments if, for

all $t_2 > t_1 \geq 0$, $N(t_2) - N(t_1)$ has the same distribution as $N(t_2 - t_1)$. In other words, the distribution of the number of events in an interval depends only on the length of the interval.

We can now give one of the standard definitions of a (homogeneous) Poisson process. Let $\lambda > 0$ be fixed. The counting process $\{N(t), t \in [0, \infty)\}$ is called a Poisson process with rate λ if the following conditions hold:

(i) $N(0) = 0$;

(ii) $N(t)$ has independent, stationary increments;

(iii) the number n of arrivals in any interval of length τ has the Poisson distribution

$$P_n(\tau) = \frac{(\lambda\tau)^n}{n!}\mathrm{e}^{-\lambda\tau}. \tag{2.6}$$

Another classical way to obtain a Poisson process is to consider the occurrence of arrivals or events in infinitesimal time intervals. That is, suppose we divide a given time interval τ into M bins of size $\Delta\tau = \tau/M$ and assume that $\Delta\tau$ is small enough so that the probability of finding two events within any one bin can be neglected. Given a rate λ take the probability of finding one event in a given bin to be $\lambda\Delta\tau$. The probability $P_n(\tau)$ of finding n independent events in the interval τ is then given by the Binomial distribution

$$P_n(\tau) = \lim_{\Delta\tau\to 0} \frac{M!}{(M-n)!n!}(\lambda\Delta\tau)^n(1-\lambda\Delta\tau)^{M-n}.$$

This consists of the probability $(\lambda\Delta\tau)^n$ of finding n events in n specific bins multiplied by the probability $(1-\lambda\Delta\tau)^{M-n}$ of not finding events in the remaining bins. The binomial factor is the number of ways of choosing n out of M bins that contain an event. Using the approximation $M - n \approx M = \tau/\Delta\tau$ and defining $\varepsilon = -\lambda\Delta\tau$, we have that

$$\lim_{\Delta\tau\to 0}(1-\lambda\Delta\tau)^{M-n} = \lim_{\varepsilon\to 0}\left((1+\varepsilon)^{1/\varepsilon}\right)^{-\lambda\tau} = \mathrm{e}^{-\lambda\tau}.$$

For large M, $M!/(M-n)! \approx M^n = (\tau/\Delta\tau)^n$, so that we recover the Poisson distribution (2.6).

A simple method for calculating the moments of the Poisson distribution is to introduce the moment generating function

$$G(s;\tau) = \sum_{n=0}^{\infty} P_n(\tau)\mathrm{e}^{sn}. \tag{2.7}$$

Differentiating with respect to s shows that

$$\left.\frac{d^k G(s;\tau)}{ds^k}\right|_{s=0} = \sum_{n=0}^{\infty} P_n(\tau)n^k \equiv \langle n^k(\tau)\rangle. \tag{2.8}$$

The generating function for the Poisson process can be evaluated explicitly as

$$G(s;\tau) = \exp(-\lambda\tau)\exp(\lambda\tau\mathrm{e}^s), \tag{2.9}$$

from which we deduce that the mean and variance are

$$\langle n(\tau)\rangle = \lambda\tau, \quad \sigma_n^2(\tau) = \lambda\tau. \tag{2.10}$$

The ratio of the standard deviation to the mean is called the *coefficient of variation* $C_V = \sigma_\tau/\langle n(\tau)\rangle$.

Another useful quantity is the distribution of inter-arrival times $\tau_n = T_n - T_{n-1}$, where T_n is the nth arrival time, which can be defined iteratively according to

$$T_n = \inf\{t \geq 0 | N(t + T_{n-1}) = n\}, \quad T_0 = 0.$$

Suppose that an event last occurred at time T_n. The probability that the next event occurs in the interval $T_n + \tau \leq T_{n+1} \leq T_n + \tau + \Delta\tau$ is equal to the probability that no event occurs for a time τ, which is $p = \mathrm{e}^{-\lambda\tau}$ multiplied by the probability $\lambda\Delta\tau$ of an event within the following interval $\Delta\tau$:

$$\mathbb{P}[\tau \leq T_{n+1} - T_n \leq \tau + \Delta\tau] = \lambda\Delta\tau\mathrm{e}^{-\lambda\tau}.$$

The inter-arrival time probability density is thus an exponential, $\rho(\tau) = \lambda\mathrm{e}^{-\lambda\tau}$. It follows that the mean is

$$\langle\tau\rangle = \int_0^\infty \lambda\mathrm{e}^{-\lambda\tau}\tau d\tau = \frac{1}{\lambda}$$

and the variance is

$$\sigma_\tau^2 = \int_0^\infty \lambda\mathrm{e}^{-\lambda\tau}\tau^2 d\tau - \langle\tau\rangle^2 = \frac{1}{\lambda^2}.$$

Finally, note that the fastest way to generate a Poisson sequence of events for constant λ is to iterate the arrival times according to $T_{n+1} = T_n - \log(x_{rand})/\lambda$ with x_{rand} uniformly distributed over $[0, 1]$.

It is possible to generalize the above Poisson model to the case of a time-dependent rate $\lambda(t)$. The simplest way to analyze this inhomogeneous Poisson process is to consider the joint probability density of n arrival times, $\rho(T_1, \ldots T_n)$. In particular, $\rho(T^1, \ldots T^n)(\Delta t)^n$ is given by the product of the probabilities $\lambda(T_j)\Delta t$ that the jth event occurs in the time interval $T_j \leq t \leq T_j + \Delta t$ and the probabilities of not firing during the inter-arrival intervals. The latter is given by

$$\mathbb{P}[\text{no events in } (T_j, T_{j+1})] = \prod_{m=1}^{M} (1 - \lambda(T_j + m\Delta t)\Delta t),$$

where we have partitioned the interval (T_j, T_{j+1}) into M bins of size Δt. Taking the logarithm,

$$\log\mathbb{P}[\text{no events in } (T_j, T_{j+1})] = \sum_{m=1}^{M} \log(1 - \lambda(T_j + m\Delta t)\Delta t) \approx -\sum_{m=1}^{M} \lambda(T_j + m\Delta t)\Delta t.$$

Taking the limit $\Delta t \to 0$ and exponentiating again shows that

$$\mathbb{P}[\text{no events in } (T_j, T_{j+1})] = \exp\left(-\int_{T_j}^{T_{j+1}} \lambda(t)dt\right).$$

Hence, the probability density that during the time interval $[0, T]$ there are exactly n events at times $T_1, \ldots, T_n$ is

$$\rho(T_1, \ldots T_n; T) = \left(\prod_{i=1}^{n} \lambda(T_i)\right) \exp\left(-\int_0^T \lambda(t)dt\right). \tag{2.11}$$

We now note that the probability of n events occurring in $[0, T]$ in particular time bins can also be written as

$$\rho(T_1, \ldots T_n; T)(\Delta t)^n = n! \frac{\prod_{i=1}^{n} \lambda(T_i)\Delta t}{\int_0^T \lambda(t)dt} P_n[T],$$

which establishes that the distribution for an inhomogeneous Poisson process is given by

$$P_n(T) = \frac{\Lambda(T)^n}{n!}\mathrm{e}^{-\Lambda(T)}, \quad \Lambda(T) = \int_0^T \lambda(t)dt. \tag{2.12}$$

2.3. Chemical reaction networks.

Consider a well-mixed system with K chemical species, $\mathcal{X} = \{X_1, \ldots, X_K\}$, undergoing R single-step chemical reactions labeled $a = 1, \ldots, R$. Let $\mathbf{s}_a$ be the vector specifying the number of molecules of each species forming the a-th reactant complex, and let $\mathbf{r}_a$ be the corresponding vector of the a-th product complex. The a-th chemical reaction thus takes the form

$$s_{1a}X_1 + s_{2a}X_2 + \ldots \to r_{1a}X_1 + r_{2a}X_2 + \ldots,$$

where s_{ja}, r_{ja} are known as *stochiometric coefficients*. Let n_j be the number of molecules of X_j and set $\mathbf{n} = (n_1, \ldots, n_K)$. When one such reaction occurs the state $\mathbf{n}$ is changed according to

$$\mathbf{n} \to \mathbf{n} + \mathbf{r}_a - \mathbf{s}_a.$$

The reverse reaction (if it exists)

$$\sum_j r_{ja}X_j \to \sum_j s_{ja}X_j,$$

would then have $\mathbf{n} \to \mathbf{n} - \mathbf{r}_a + \mathbf{s}_a$. If X_k is a catalyst or enzyme, then it plays the role of facilitating reactions between other chemical species such that $s_{ka} = r_{ka} \neq 0$. More complicated multi-step reactions can always be decomposed into these fundamental single-step reactions with appropriate stochiometric coefficients. In practice, most reactions involve the collisions of a pair of molecules so that $\sum_j s_{ja} = 1$ or 2.

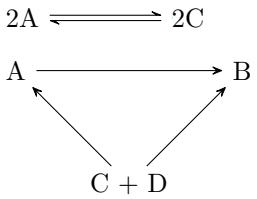
In the case of a large number of molecules, one can describe the dynamics of a single step chemical reaction in terms of a kinetic or rate equation involving the concentrations $x_j = n_j/\Omega$ – the *law of mass action*. Here Ω is a dimensionless quantity representing the system size, which in gene networks is typically taken to be the characteristic number of proteins. (Alternatively, it could represent some volume scale factor.) From basic physical principles, one usually finds that

$$\frac{dx_i}{dt} = \sum_{a=1}^{R} \kappa_a (r_{ia} - s_{ia}) \prod_{j=1}^{K} x_j^{s_{ja}} \equiv \sum_{a=1}^{R} S_{ia} f_a(\mathbf{x}), \quad i = 1, \ldots, K). \tag{2.13}$$

Here κ_a is a rate constant that depends on the probability that a collision of the relevant molecules actually leads to a reaction. The product of concentrations is motivated by the idea that in a well-mixed container there is a spatially uniform distribution of each type of molecule, and the probability of a collision depends on the probability that each of the reactants is in the same local region of space. Ignoring any statistical correlations, the latter is given by the product of the individual concentrations. Also note that $\mathbf{S}$ is the so-called $K \times R$ stochiometric matrix for K 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 as transition intensities or *propensities*.

Analyzing a system of nonlinear ODEs is generally a daunting task, even with regards finding a general set of conditions for the existence and stability of fixed points. However, in the case of the mass-action kinetic equations (2.13), there is an underlying network structure that makes the task much easier. This network structure is associated with a directed graph whose nodes are complexes (sets of chemical species that occur on the left-hand (reactants) or right-hand (products) side of single-step reactions, and whose links represent these reactions [93, 131, 132]. One of the most important theorems regarding deterministic mass action kinetics is the *deficiency zero theorem* of Feinberg [94, 95]. Roughly speaking, this theorem states that if certain easily verifiable properties of the network hold, then within each invariant manifold of the dynamics (compatibility class) there is exactly one fixed point with strictly positive components, and the fixed point is locally asymptotically stable. For a review of deterministic chemical reaction network theory see [122]. Here we briefly summarize some of the main ideas needed to state the zero deficiency theorem.

Let $\mathcal{C} = \{\mathbf{r}_a, \mathbf{s}_a;\, a = 1, \dots R\}$ and $\mathcal{R} = \{\mathbf{s}_a \to \mathbf{r}_a;\, a = 1, \dots R\}$ denote the sets of complexes and reactions, respectively. (Note that the same complex can act as a product or reactant in more than one reaction.) The triple $\{\mathcal{X}, \mathcal{C}, \mathcal{R}\}$ is called a chemical reaction network. Corresponding to each reaction network is a unique, directed graph constructed as follows. The nodes of the graph are given by the set of distinct complexes $\mathbf{z} \in \mathcal{C}$. A directed edge is then placed from a complex $\mathbf{z}$ to a complex $\mathbf{z}'$ if and only if $\mathbf{z} \to \mathbf{z}' \in \mathcal{R}$. Each connected component of the graph is called a linkage class of the graph, with the number of linkage classes denoted by l. A network is said to be *reversible* if for every forward reaction $\mathbf{z} \to \mathbf{z}' \in \mathcal{R}$ there is a corresponding backward reaction $\mathbf{z}' \to \mathbf{z} \in \mathcal{R}$. A network is said to be *weakly reversible* if for any reaction $\mathbf{z} \to \mathbf{z}' \in \mathcal{R}$, there is a sequence of directed reactions starting with $\mathbf{z}'$ as a reactant complex and ending with $\mathbf{z}$ as a product complex. As an illustration, consider the following set of chemical reactions:



There are $K = 4$ chemical species (A, B, C, D), $m = 5$ complexes $(A, B, C{+}D, 2A, 2C)$, and $R = 5$ reactions. There are two disconnected graphs so the number of linkages $l = 2$. The top graph is reversible, whereas the second is weakly reversible.

Another important notion is the span of the stochiometric vectors $\mathrm{v}_a = \mathbf{r}_a - \mathbf{s}_a$, that is,

$$\mathcal{S} = \mathrm{span}_{a=1,\dots,R}\{\mathrm{v}_a\} \subset \mathbb{R}^K.$$

In general, $\mathcal{S}$ will be a proper subset of $\mathbb{R}^K$ so that $s \equiv \dim[\mathcal{S}] < K$. For the above example,

$$\mathcal{S} = \mathrm{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -2 \\ 0 \end{bmatrix} \right\}$$

$$= \text{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 2 \\ 0 \end{bmatrix} \right\},$$

so that $s = 3$. Since the vector space $\mathcal{S}$ specifies the possible combinations of changes in the number of molecules of each species over the set of reactions, it follows that if the initial concentrations are given by the vector $\mathbf{x}(0) = x_0$, then the solution $\mathbf{x}(t)$ of the mass-action kinetic equations (2.13) will lie in the space $\mathbf{x}_0 + \mathcal{S}$. This invariant manifold of the dynamics is called a stochiometric compatability class, and since the concentration (number) of each chemical species is positive, any solution lies in the positive stochiometric compatability class. We need one final definition to state the zero deficiency theorem, namely, the *deficiency* of a chemical reaction network is

$$\delta = m - l - s,$$

where m is the number of complexes, l is the number of linkage classes, and s is the dimension of the stochiometric subspace of the network. For example, the above reaction network has a zero deficiency $\delta = 0$, since $m = 5, l = 2, s = 3$.

Theorem 6 (The deficiency zero theorem [95]) *Consider a weakly reversible, deficiency zero chemical reaction network $(\mathcal{X}, \mathcal{C}, \mathcal{R})$ with dynamics given by the mass-action kinetics (2.13). Then for any choice of rate constants, within each positive stochiometric compatability class there is precisely one fixed point, which is locally asymptotically stable with respect to that compatability class.*

Note that a crucial step in the proof of the deficiency zero theorem and its extensions is to reformulate the mass-action kinetics in terms of the space of complexes $\mathbb{R}^m$ rather than the space of chemical species $\mathbb{R}^n$. As a first step, we rewrite equation (2.13) as

$$\frac{dx_i}{dt} = \sum_{\mathbf{z} \to \mathbf{z}'} \kappa_{\mathbf{z} \to \mathbf{z}'} \left(\prod_{j=1}^{K} x_j^{s_j} \right) (z_i' - z_i).$$

That is, we specify the reactions in terms of the reactant and product complexes rather than the reaction label a. We construct a new vector space $\mathbb{R}^m$ of complexes by introducing the set of basis vectors $\{\omega_{\mathbf{z}}, \mathbf{z} \in \mathcal{C}\}$ and decomposing any $v \in \mathbb{R}^m$ as $v = \sum_{\mathbf{z} \in \mathcal{C}} v_{\mathbf{z}} \omega_{\mathbf{z}}$. Let $Z : \mathbb{R}^m \to \mathbb{R}^K$ be the linear map defined by $Z(\omega_{\mathbf{z}}) = \mathbf{z}$. Let $\Psi : \mathbb{R}^K \to \mathbb{R}^m$ be the nonlinear map given by

$$\Psi(\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{C}} \left(\prod_{j=1}^{K} x_j^{z_j} \right) \omega_{\mathbf{z}}. \tag{2.14}$$

Finally, introduce the map $A_\kappa : \mathbb{R}^m \to \mathbb{R}^m$,

$$A_\kappa(v) = \sum_{\mathbf{z} \to \mathbf{z}'} \kappa_{\mathbf{z} \to \mathbf{z}'} v_{\mathbf{z}} (\omega_{\mathbf{z}'} - \omega_{\mathbf{z}}). \tag{2.15}$$

The mass-kinetics can now be written in the form

$$\frac{d\mathbf{x}}{dt} = Y(A_\kappa(\Psi(\mathbf{x}))). \tag{2.16}$$

From equation (2.13), a fixed point $\mathbf{x}_*$ exists if and only if $\sum_a S_{ia} f_a(\mathbf{x}_*) = 0$ for all $i = 1, \ldots, K$. In terms of the decomposition (2.16), it is sufficient to show that

$A_\kappa(\Psi(\mathbf{x}_*)) = 0$. This is equivalent to the so-called *complex balance condition*: for each $\mathbf{z} \in \mathcal{C}$

$$\sum_{a:\mathbf{r}_a=\mathbf{z}} \kappa_a \left(\prod_{j=1}^{K} x_j^{s_{ja}} \right) = \sum_{b:\mathbf{s}_b=\mathbf{z}} \kappa_b \left(\prod_{j=1}^{K} x_j^{s_{jb}} \right). \tag{2.17}$$

Complex balance corresponds to the condition that for each $\mathbf{z} \in \mathcal{C}$, the rates of all reactions into the complex $\mathbf{z}$ are balanced by the rates of all reactions out of $\mathbf{z}$. This is distinct from the standard condition of detailed balance in reversible networks. The requirement of zero deficiency and weak reversibility establishes that there exists at least one fixed point for which $A_\kappa(\Psi(\mathbf{x}^*)) = 0$. One can the establish that this fixed point is unique and asymptotically stable.

There have been a number of more recent results in chemical reaction network theory. For example, in the case of deterministic models, mathematical analysis has been used to investigate conditions for the existence of multiple equilibria [67,68], as well as dynamical properties such as persistence and global stability [6,12]; persistence is the condition that if all chemical species are initially present then none of them completely disappear during the time-evolution of the network.

2.4. Chemical master equation and the system-size expansion

In the case of gene networks at least some of the reacting species occur with low copy numbers (see section 3). This means that one can no longer model the dynamics in terms of a deterministic system of ODEs. Instead, one keeps track of changes in the number of molecules of each species according to a continuous time Markov chain. The possible transitions and reaction rates are determined by the stochiometric matrix and propensities introduced above. The corresponding chemical master equation takes the form

$$\frac{dP(\mathbf{n},t)}{dt} = \Omega \sum_{a=1}^{R} \left(\prod_{i=1}^{K} \mathbb{E}^{-S_{ia}} - 1 \right) f_a(\mathbf{n}/\Omega) P(\mathbf{n},t), \tag{2.18}$$

Here $\mathbb{E}^{-S_{ia}}$ is a step or ladder operator such that for any function $g(\mathbf{n})$,

$$\mathbb{E}^{-S_{ia}} g(n_1,\ldots,n_i,\ldots,n_N) = g(n_1,\ldots,n_i - S_{ia},\ldots,n_N). \tag{2.19}$$

One point to note is that when the number of molecules is sufficiently small, the characteristic form of a propensity function $f(\mathbf{x})$ in equation (2.13) has to be modified:

$$\left(\frac{n_j}{\Omega}\right)^{s_j} \to \frac{1}{\Omega^{s_j}} \frac{n_j!}{(n_j - s_j)!}.$$

In general it is not possible to obtain exact solutions of the master equation (2.18) even in the case of a stationary solution. (Note, however, that recent progress has been made by generalizing the theory of deterministic chemical reaction networks briefly described in section 2.3 to stochastic models, see [5].) Therefore, one often resorts to some form of approximation scheme. The most common is the diffusion approximation obtained by carrying out an expansion in the system size Ω [83,105,242]. 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 a stochastic differential equation (SDE). A rigorous analysis of the diffusion approximation has

been carried out by Kurtz [158]. The basic idea is to set $f_a(\mathbf{n}/\Omega)P(\mathbf{n},t) \to f_a(\mathbf{x})p(\mathbf{x},t)$ with $\mathbf{x} = \mathbf{n}/\Omega$ treated as a continuous vector so that

$$\prod_{i=1}^{K} \mathbb{E}^{-S_{ia}} h(\mathbf{x}) = h(\mathbf{x} - \mathbf{S}_a/\Omega)$$
$$= h(\mathbf{x}) - \Omega^{-1} \sum_{i=1}^{K} S_{ia} \frac{\partial h}{\partial x_i} + \frac{1}{2\Omega^2} \sum_{i,j=1}^{K} S_{ia} S_{ja} \frac{\partial^2 h(\mathbf{x})}{\partial x_i \partial x_j} + O(\Omega^{-3})$$

Carrying out a Taylor expansion of the master equation to second order thus yields the multivariate FP equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{K} \frac{\partial V_i(\mathbf{x}) p(\mathbf{x},t)}{\partial x_i} + \frac{1}{2\Omega} \sum_{i,j=1}^{K} \frac{\partial^2 D_{ij}(\mathbf{x}) p(\mathbf{x},t)}{\partial x_i \partial x_j}, \tag{2.20}$$

where

$$V_i(\mathbf{x}) = \sum_{a=1}^{R} S_{ia} f_a(\mathbf{x}), \quad D_{ij}(\mathbf{x}) = \sum_{a=1}^{R} S_{ia} S_{ja} f_a(\mathbf{x}). \tag{2.21}$$

The FP equation (2.20) corresponds to the multivariate Langevin equation

$$dX_i = V_i(\mathbf{X})dt + \frac{1}{\sqrt{\Omega}} \sum_{a=1}^{R} B_{ia}(\mathbf{X}) dW_a(t), \tag{2.22}$$

where $W_a(t)$ are independent Wiener processes [105],

$$\langle dW_a(t) \rangle = 0, \quad \langle dW_a(t) dW_b(t') \rangle = \delta_{a,b} \delta(t-t') dt\, dt', \tag{2.23}$$

and $\mathbf{D} = \mathbf{B}\mathbf{B}^T$, that is,

$$B_{ia} = S_{ia} \sqrt{f_a(\mathbf{x})}. \tag{2.24}$$

The system-size expansion has been used extensively to analyze the effects of molecular noise on protein fluctuations in gene expression [83, 202, 237, 240] as well as other cellular processes such as the opening/closing of ion channels [63, 97], see section 8.1. It works particularly well when the underlying deterministic system (2.13) has a unique stable fixed point since, after a transient phase, the dynamics consists of Gaussian-like fluctuations about the fixed point (see the linear noise approximation). However, in the case of bistable molecular switches with weak noise (see section 3), the transitions between different metastable states typically involve rare transitions that lie in the tails of the associated probability distributions where the Gaussian approximation no longer necessarily holds. Therefore, in order to ensure accurate estimates of transition rates, one has to use alternative approximation schemes based on WKB and large deviation theory (see section 5).

Linear noise approximation. Now suppose that the deterministic system (2.13), 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

$$A_{ij} = \left. \frac{\partial V_i}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x}_*} \tag{2.25}$$

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 (2.22) and keeping only lowest order terms in $\Omega^{-1/2}$ yields the Ornstein-Uhlenbeck (OU) process [105]

$$dY_i = \sum_{j=1}^{K} A_{ij} Y_j dt + \sum_{a=1}^{R} B_{ia}(\mathbf{x}_*) dW_a(t). \tag{2.26}$$

Writing this equation in matrix form

$$d\mathbf{Y} = \mathbf{A}\mathbf{Y}dt + \mathbf{B}\, d\mathbf{W}(t),$$

and performing the change of variables $\mathbf{Z}(t) = \mathrm{e}^{-\mathbf{A}t}\mathbf{Y}(t)$ gives $d\mathbf{Z} = \mathrm{e}^{-\mathbf{A}t}\mathbf{B}\, d\mathbf{W}(t)$. Formal integration then shows that

$$\mathbf{Y}(t) = \mathrm{e}^{-\mathbf{A}t}\mathbf{Y}(0) + \int_0^t \mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B}\, d\mathbf{W}(t').$$

Taking averages then establishes that

$$\mathbf{Y}(t) - \langle \mathbf{Y}(t) \rangle = \int_0^t \mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B} d\mathbf{W}(t').$$

Introduce the correlation function $\mathbf{C}(t,s) = \langle \mathbf{Y}(t), \mathbf{Y}^T(s) \rangle$ with components

$$C_{ij}(t,s) = \langle Y_i(t), Y_j(s) \rangle = \langle [Y_i(t) - \langle Y_i(t) \rangle][Y_j(s) - \langle Y_j(s) \rangle] \rangle.$$

It follows that

$$\begin{aligned}
\mathbf{C}(t,s) &= \left\langle \int_0^t \mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B} d\mathbf{W}(t') \int_0^s d\mathbf{W}^T(s') \mathbf{B}^T \mathrm{e}^{\mathbf{A}^T(s-s')} \right\rangle \\
&= \int_0^t \int_0^s \mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B} \left\langle d\mathbf{W}(t') d\mathbf{W}^T(s') \right\rangle \mathbf{B}^T \mathrm{e}^{\mathbf{A}^T(s-s')}
\end{aligned}$$

Now

$$\left\langle d\mathbf{W}(t') d\mathbf{W}^T(s') \right\rangle_{ab} = \langle dW_a(t') dW_b(s') \rangle = \delta_{ab}\delta(t'-s')dt'\, ds'$$

Therefore,

$$\mathbf{C}(t,s) = \int_0^{\min(t,s)} \mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B}\mathbf{B}^T \mathrm{e}^{\mathbf{A}^T(s-t')} dt'$$

Finally, defining the covariance matrix according to $\Sigma(t) = \mathbf{C}(t,t)$ we have

$$\Sigma(t) = \int_0^t \mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B}\mathbf{B}^T \mathrm{e}^{\mathbf{A}^T(t-t')} dt'.$$

Differentiating both sides with respect to t shows that

$$\begin{aligned}
\frac{d\Sigma}{dt} &= \left[\mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B}\mathbf{B}^T \mathrm{e}^{\mathbf{A}^T(t-t')} \right]_{t'=t} + \int_0^t \mathbf{A}\mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B}\mathbf{B}^T \mathrm{e}^{-\mathbf{A}^T(t-t')} dt' \\
&\quad + \int_0^t \mathrm{e}^{\mathbf{A}(t-t')}\mathbf{B}\mathbf{B}^T \mathrm{e}^{\mathbf{A}^T(t-t')}\mathbf{A}^T dt' \\
&= \mathbf{A}\Sigma(t) + \Sigma(t)\mathbf{A}^T + \mathbf{B}\mathbf{B}^T
\end{aligned}$$

Stability of the fixed point x_* means that $\mathbf{A}$ has K negative definite eigenvalues λ_j. For the sake of illustration, suppose that the eigenvalues are distinct so that $\mathbf{A}$ is diagonalizable, that is, there exists an orthogonal matrix $\mathbf{U}$ such that $\mathbf{U}^T\mathbf{A}\mathbf{U} =$

$\mathbf{A}_d \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$. Left-multiplying both sides of the above equation by $\mathbf{U}^T$ and right-multiplying by $\mathbf{U}$ gives

$$\frac{d\widehat{\Sigma}}{dt} = \mathbf{A}_d\widehat{\Sigma}(t) + \widehat{\Sigma}(t)\mathbf{A}_d^T + \widehat{\mathbf{B}\mathbf{B}^T},$$

where $\widehat{\Sigma} = \mathbf{U}^T\Sigma\mathbf{U}$ and $\widehat{\mathbf{B}\mathbf{B}^T} = \mathbf{U}^T\mathbf{B}\mathbf{B}^T\mathbf{U}$. In component form,

$$\frac{d\widehat{\Sigma}_{jk}}{dt} = (\lambda_j + \lambda_k)\widehat{\Sigma}_{jk}(t) + \widehat{\mathbf{B}\mathbf{B}^T}_{jk},$$

with $\lambda_j + \lambda_k < 0$. Hence, in the limit $t \to \infty$ we have $\widehat{\Sigma}(t) \to \widehat{\Sigma}_0$ where

$$(\lambda_j + \lambda_k)\widehat{\Sigma}_{0,jk} = -\widehat{\mathbf{B}\mathbf{B}^T}_{jk}.$$

In the original basis, we thus find that the stationary covariance matrix Σ_0 satisfies the so-called Ricatti equation

$$\mathbf{A}\Sigma_0 + \Sigma_0\mathbf{A}^T = -\mathbf{B}\mathbf{B}^T. \tag{2.27}$$

Equation (2.27), which is a form of fluctuation-dissipation theorem, is often used to estimate the size of protein fluctuations due to intrinsic noise. We will give an example in section 3.4.

2.5. The stochastic simulation algorithm (SSA)

The SSA, which was originally developed by Gillespie [110–112], is an efficient numerical scheme for generating exact sample paths of a continuous-time Markov process whose probability distribution evolves according to the chemical master equation (2.18). In the following we eliminate the global factor of Ω by rescaling time $t \to \Omega t$.

The basic idea underlying the SSA is to introduce 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. It follows that 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 $[t, t+\tau)$. From the definitions of P_0 and the propensities f_a, we have

$$P_0(\tau + d\tau|\mathbf{x}, t) = P_0(\tau|\mathbf{x}, t)\left[1 - \sum_{a=1}^{R} f_a(\mathbf{x})d\tau\right],$$

where the right-hand side involves 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

$$\frac{dP_0(\tau|\mathbf{x}, t)}{d\tau} = -F(\mathbf{x})P_0(\tau|\mathbf{x}, t), \quad F(\mathbf{x}) = \sum_{a=1}^{R} f_a(\mathbf{x}).$$

Given the initial condition $P(0|x, t) = 1$, this equation has the solution $P_0(\tau|\mathbf{x}, t) = \exp(-F(\mathbf{x})\tau)$ which, when combined with the result $p(\tau, a|\mathbf{x}, t)d\tau = P_0(\tau|\mathbf{x}, t)f_a(\mathbf{x})d\tau$, implies

$$p(\tau, a|\mathbf{x}, t) = F(\mathbf{x})\exp(-F(\mathbf{x})\tau)\frac{f_a(\mathbf{x})}{F(\mathbf{x})}. \tag{2.28}$$

This establishes that 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 $[0,1]$ with

$$\tau = -\frac{1}{F(\mathbf{x})} \ln r_1 \tag{2.29a}$$

$$a = \text{ the smallest integer for which } \sum_{s=1}^{a} f_a(\mathbf{x}) > r_2 F(\mathbf{x}). \tag{2.29b}$$

The direct method of implementing the SSA is as follows:

(i) Initialize the time $t = t_0$ and the chemical state $\mathbf{x} = \mathbf{x}_0$

(ii) Given the state $\mathbf{x}$ at time t, determine the $f_a(\mathbf{x})$ for $a = 1, \ldots, R$ and their sums $F(\mathbf{x})$

(iii) Generate values for τ and a using equations (2.29a) and (2.29b)

(iv) Implement the next reaction by setting $t \to t' = t + \tau$ and $x_j \to x_j' = x_j + S_{ja}/\Omega$.

(v) Return to step 2 with $(\mathbf{x}, t)$ replaced by $(\mathbf{x}', t')$, or else stop.

There have been several modifications of the basic SSA that differ in the implementation of step 2. These include the next reaction method [109] and the modified next reaction method [4]. The latter is based on the random time change representation of Kurtz [7, 159]. Finally, note that SSAs have also been developed for stochastic hybrid systems, which are defined in the next section, see for example Ref. [258]

tau-leaping. One of the limitations of the standard SSA is that it is very time expensive. For example, in many applications the mean time between reactions, $1/F(\mathbf{x})$, is very small so that simulating every reaction becomes computationally infeasible. One way to improve efficiency is to sacrifice accuracy for a gain in computational speed, which is the basis of the *tau-leaping* method introduced originally by Gillespie [111]. 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 (see section 2.2). 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 increases x_j by S_{ja}/Ω, the state at time $t+\tau$ will be

$$X_j(t+\tau) = \mathbf{x} + \sum_{a=1}^{R} \mathscr{N}_a(f_a(\mathbf{x})\tau) S_{ja}, \tag{2.30}$$

where the $\mathscr{N}_a$ are independent Poisson processes. This equation is known as the *tau-leaping formula*. One major issue with the original formulation of tau-leaping is that one has to be careful in choosing an appropriate value of τ at each iteration of the algorithm. If τ is too large then a sufficiently large change in propensities could cause one or more components x_j to become negative. On the other hand, if τ is too small

then the algorithm becomes prohibitively expensive. This issue has been addressed in various modifications of the tau-leaping procedure, see for example Cao et al. [60]. More recently, multi-level methods that couple tau-leaping approximations at different resolutions have been used to reduce variances in Monte Carlo estimators [8, 9].

Trajectory sampling methods. Another situation where the standard SSA is computationally expensive is when dealing with rare events, such as transitions between metastable states in the weak-noise limit (see section 5). That is, the switching of a bistable switch from one state to another may happen so infrequently that running a stochastic simulation long enough to see transitions would be very computationally prohibitive. A number of computational methods have been developed to address this problem, based on some form of efficient sampling of trajectories in state space transitioning from a state A to a state B, say [40, 127, 253]. For example, the *weighted sampling* method uses a multiple-trajectory strategy, whereby state space is partitioned into local regions or bins, and individual trajectories generate multiple daughter trajectories upon reaching a new bin region; the daughters are suitably weighted to ensure statistical rigor [38, 79, 134, 259]. One finds that weighted sampling Monte Carlo simulations can yield rigorous rate estimates for processes that occur on much longer timescales than the simulations themselves. Another method is *umbrella sampling* [74, 249]. This also divides state space into a set of bins but now treats simulations within each bin independently. Any trajectory that hits the boundary is re-injected into the domain in such a way that it preserves the distribution of fluxes from neighboring bins.

2.6. Stochastic hybrid systems

Let us now turn to the definition of a stochastic hybrid system and, in particular, a piecewise deterministic Markov process (PDMP) [70, 96, 153]. For the sake of illustration, consider a system whose states are described by a pair $(x, n) \in \Sigma \times \{0, \cdots, N_0 - 1\}$, where x is a continuous variable in a bounded interval $\Sigma \subset \mathbb{R}$ and n a discrete stochastic variable taking values in the finite set $\Gamma \equiv \{0, \cdots, N_0 - 1\}$. (Note that one could easily extend the analysis to higher-dimensions, $x \in \mathbb{R}^d$. In this case Σ is taken to be a connected, bounded domain with a regular boundary $\partial\Omega$. However, for notational simplicity, we restrict ourselves to the case $d = 1$. 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.) When the internal state is n, the system evolves according to the ordinary differential equation (ODE)

$$\dot{x} = F_n(x), \tag{2.31}$$

where the vector field $F_n : \mathbb{R} \to \mathbb{R}$ is a continuous function, locally Lipschitz. That is, given a compact subset $\mathscr{K}$ of Σ, there exists a positive constant K_n such that

$$|F_n(x) - F_n(y)| \leq K_n|x - y|, \quad \forall x, y \in \Sigma \tag{2.32}$$

for some constant K_n. We assume that the dynamics of x is confined to the domain Σ so that we have existence and uniqueness of a trajectory for each n. For fixed x, the discrete stochastic variable evolves according to a homogeneous, continuous-time Markov chain with generator $-\mathbf{A}(x)$. We make the further assumption that the chain is irreducible for all $x \in \Sigma$, that is, for fixed x there is a non-zero probability

of transitioning, possibly in more than one step, from any state to any other state of the Markov chain (section 2.1). This implies the existence of a unique invariant probability distribution on Γ for fixed $x \in \Sigma$, denoted by $\rho^*(x)$, such that

$$\sum_{m \in \Gamma} A_{nm}(x)\rho^*_m(x) = 0, \quad \forall n \in \Gamma. \tag{2.33}$$

The above stochastic model defines a one-dimensional PDMP. It is also possible to consider generalizations of the continuous process, in which the ODE (2.31) is replaced by a stochastic differential equation (SDE) or even a partial differential equation (PDE). In order to allow for such possibilities we will refer to all of these processes as examples of a stochastic hybrid system.

Let us decompose the transition matrix of the Markov chain as

$$W_{nm}(x) = P_{nm}(x)\lambda_m(x),$$

with $\sum_{n \neq m} P_{nm}(x) = 1$ for all x. Hence $\lambda_m(x)$ determines the jump times from the state m whereas $P_{nm}(x)$ determines the probability distribution that when it jumps the new state is n for $n \neq m$. The hybrid evolution of the system with respect to $x(t)$ and $n(t)$ can then be described as follows. Suppose the system starts at time zero in the state (x_0, n_0). Call $x_0(t)$ the solution of (2.31) with $n = n_0$ such that $x_0(0) = x_0$. Let θ_1 be the random variable such that

$$\mathbb{P}(\theta_1 < t) = 1 - \exp\left(-\int_0^t \lambda_{n_0}(x_0(t'))dt'\right).$$

Then in the random time interval $s \in [0, \theta_1)$ the state of the system is $(x_0(s), n_0)$. We draw a value of θ_1 from $\mathbb{P}(\theta_1 < t)$, choose an internal state $n_1 \in \Gamma$ with probability $P_{n_1 n_0}(x_0(\theta_1))$, and call $x_1(t)$ the solution of the following Cauchy problem on $[\theta_1, \infty)$:

$$\begin{cases} \dot{x}_1(t) &= F_{n_1}(x_1(t)), \quad t \geq \theta_1 \\ x_1(\theta_1) &= x_0(\theta_1) \end{cases}$$

Iterating this procedure, we construct a sequence of increasing jumping times $(\theta_k)_{k \geq 0}$ (setting $\theta_0 = 0$) and a corresponding sequence of internal states $(n_k)_{k \geq 0}$. The evolution $(x(t), n(t))$ is then defined as

$$(x(t), n(t)) = (x_k(t), n_k) \quad \text{if } \theta_k \leq t < \theta_{k+1}. \tag{2.34}$$

Note that the path $x(t)$ is continuous and piecewise C^1. In order to have a well-defined dynamics on $[0, T]$, it is necessary that almost surely the system makes a finite number of jumps in the time interval $[0, T]$. This is guaranteed in our case.

The above formulation is the basis of a simulation algorithm for PDMPs [258]. If $\lambda_m(x) = \kappa_m$ is independent of x for all m, then the stochastic jump time from state m is given by

$$\tau = \frac{1}{\kappa_m} \ln(1/u),$$

where $u \in [0, 1]$ is a realization of a uniform random variable. The simulation of the PDMP is then very similar to the SSA for a chemical master equation. On the other hand, in the x-dependent case one has to numerically solve the integral equation

$$\int_0^\tau \lambda_{n_k}(x_k(t))dt = \ln(1/u). \tag{2.35}$$

at the k-th iteration. This can be achieved by solving the ODE system

$$\begin{aligned}\frac{dx_k}{dt} &= F_{n_k}(x_k(t)) \\ \frac{dR_k}{dt} &= \lambda_{n_k}(x_k(t))[1 - R_k(t)],\end{aligned}$$

and setting $R_k(t) = u$. In contrast to discrete Markov processes, the time spent in a discrete state (the sojourn time τ) of a PDMP may become infinite. That is,

$$\mathbb{P}(\tau = \infty) = c > 0, \quad \lim_{t\to\infty} \mathbb{P}(\tau < \tau) \to 1 - c.$$

Chapman-Kolmogorov equation. Given the above iterative definition of a PDMP, let $X(t)$ and $N(t)$ denote the stochastic continuous and discrete variables, respectively, at time t, $t > 0$, given the initial conditions $X(0) = x_0, N(0) = n_0$. Introduce the probability density $\rho_n(x, t|x_0, n_0, 0)$ with

$$\mathbb{P}\{X(t) \in (x, x + dx),\, N(t) = n|x_0, n_0) = \rho_n(x, t|x_0, n_0, 0)dx.$$

It follows that p evolves according to the forward differential Chapman-Kolmogorov (CK) equation [45, 105]

$$\frac{\partial \rho_n}{\partial t} = \mathbb{L}\rho_n, \tag{2.36}$$

with the generator $-\mathbb{L}$ (dropping the explicit dependence on initial conditions) defined according to

$$\mathbb{L}\rho_n(x, t) = -\frac{\partial F_n(x)\rho_n(x, t)}{\partial x} + \frac{1}{\varepsilon}\sum_{m\in\Gamma} A_{nm}(x)\rho_m(x, t). \tag{2.37}$$

The first term on the right-hand side represents the probability flow associated by the piecewise deterministic dynamics for a given n, whereas the second term represents jumps in the discrete state n. Note that we have rescaled the matrix $\mathbf{A}$ by introducing the dimensionless parameter ε, $\varepsilon > 0$. This is motivated by the observation that in gene networks one often finds a separation of time-scales between the relaxation time for the dynamics of the continuous variable x and the rate of switching between the different discrete states n. The fast switching limit then corresponds to the case $\varepsilon \to 0$. Let us now define the averaged vector field $\overline{F} : \mathbb{R} \to \mathbb{R}$ by

$$\overline{F}(x) = \sum_{n\in\Gamma} \rho_n^*(x)F_n(x)$$

It can be shown [92] that, given the assumptions on the matrix $\mathbf{A}$, the functions $\rho_n^*(x)$ on Σ belong to $C^1(\mathbb{R})$ for all $n \in \Gamma$ and that this implies that $\overline{F}(x)$ is locally Lipschitz. Hence, for all $t \in [0, T]$, the Cauchy problem

$$\left\{\begin{array}{lcl} \dot{x}(t) &=& \overline{F}(x(t)) \\ x(0) &=& x_0 \end{array}\right. \tag{2.38}$$

has a unique solution for all $n \in \Gamma$. Intuitively speaking, one would expect the stochastic hybrid system (2.31) to reduce to the deterministic dynamical system (2.38) in the fast switching limit $\varepsilon \to 0$. That is, for sufficiently small ε, the Markov chain undergoes many jumps over a small time interval Δt during which $\Delta x \approx 0$, and thus the relative frequency of each discrete state n is approximately $\rho_n^*(x)$. This can be made precise in terms of a Law of Large Numbers for stochastic hybrid systems proven in [92, 153].

Boundary conditions. From a PDE perspective the CK equation is an N_0^{th}-order quasilinear equation on Σ, which for the sake of concreteness we take to be $\Sigma = [0, L]$. In general, well-posed boundary conditions for a quasilinear PDE have to be determined using the theory of characteristics. However, for the applications considered in this review, boundary conditions are relatively straightforward due to the fact that the drift terms $F_n(x)$ do not change sign in the interval Σ. In particular, there exists an integer m, $1 \leq m \leq N_0 - 1$, such that for all $0 < x < L$ we have $F_n(x) < 0$ for $0 \leq n \leq m-1$ and $F_n(x) > 0$ for $m \leq n \leq N_0 - 1$. Given the fact that Σ is an invariant manifold, it follows that $F_n(0) = 0$ for $0 \leq n \leq m-1$ and $F_n(L) = 0$ for $m \leq n \leq N_0 - 1$. No-flux boundary conditions at the ends $x = 0, L$ take the form $J(0,t) = J(L,t) = 0$ with

$$J(x,t) = \sum_{n=0}^{N_0-1} F_n(x)\rho_n(x,t) = 0. \tag{2.39}$$

It follows that $\rho_n(0,t) = 0$ for $m \leq n \leq N_0 - 1$ and $\rho_n(L,t) = 0$ for $0 \leq n \leq m-1$. In our analysis of metastability (section 5.2), it will be necessary to impose an absorbing boundary condition at some interior point x_* of the domain Σ, that is,

$$\rho_n(x_*,t) = 0, \quad m \leq n \leq N_0 - 1, \quad \rho_n(L,t) = 0, \quad 0 \leq n \leq m-1.$$

In contrast to the no-flux conditions, there are non-zero fluxes through x_*. The nature of boundary conditions in the case of x-dependent drift terms $F_n(x)$ that switch sign is more complicated because the partitioning of left and right moving states may differ at the boundaries $x = 0$ and $x = L$. One has to look more closely at the characteristics of the underlying quasilinear PDE.

Stationary solution of two-state PDMPs. In general it is difficult to obtain an analytical steady-state solution (assuming it exists) unless $N_0 = 2$. The steady-state version of Eqs. (2.36) reduces to

$$0 = -\frac{\partial}{\partial x}(F_0(x)\rho_0(x)) + \beta(x)\rho_1(x) - \alpha(x)\rho_0(x) \tag{2.40}$$

$$0 = -\frac{\partial}{\partial x}(F_1(x)\rho_1(x,t)) - \beta(x)\rho_0(x,t) + \alpha(x)\rho_1(x), \tag{2.41}$$

Adding the pair of Eqs. yields

$$\frac{\partial}{\partial x}(F_0(x)\rho_0(x)) + \frac{\partial}{\partial x}(F_1(x)\rho_1(x)) = 0, \tag{2.42}$$

that is,

$$F_0(x)\rho_0(x) + F_1(x)\rho_1(x) = c,$$

for some constant c. The reflecting boundary conditions imply that $c = 0$. Since $F_n(x)$ is non-zero for all $x \in \Sigma$, we can express $\rho_1(x)$ in terms of $\rho_0(x)$:

$$\rho_1(x) = -\frac{F_0(x)\rho_0(x)}{F_1(x)}. \tag{2.43}$$

Substituting into Eq. (2.40) gives

$$0 = \frac{\partial}{\partial x}(F_0(x)\rho_0(x)) + \left(\frac{\alpha(x)}{F_1(x)} - \frac{\beta(x)}{F_0(x)}\right) F_0(x)\rho_0(x). \tag{2.44}$$

This yields the solutions

$$\rho_n(x) = \frac{1}{Z|F_n(x)|} \exp\left(- \int_{x_*}^{x} \left(\frac{\alpha(y)}{F_1(y)} + \frac{\beta(y)}{F_0(y)} \right) dy \right). \tag{2.45}$$

where $x_* \in \Sigma$ is arbitrary and assuming that the normalization factor Z exists.

Dichotomous noise. A special case of a two-state PDMP is a so-called velocity jump process for which $F_0(x) = -v_- < 0$ and $F_1(x) = v_+ > 0$ with $\pm v_\pm$ constant "velocities." The piecewise deterministic dynamics takes the simple form

$$\frac{dx}{dt} = \xi(t) \equiv [v_+ + v_-]n(t) - v_-.$$

The discrete state $n(t)$ evolves according to a two-state Markov chain with matrix generator

$$\mathbf{A} = \begin{pmatrix} -k_+ & k_- \\ k_+ & -k_- \end{pmatrix}.$$

In the physics literature $\xi(t)$ is called a dichotomous Markov noise process (DMNP), see the review by Bena [27]. If $P_{nn_0}(t) = \mathbb{P}[N(t) = n | N(0) = n_0]$ then the master equation for $n(t)$ takes the form

$$\frac{dP_{nn_0}}{dt} = \sum_{m=0,1} A_{nm} P_{mn_0}$$

Using the fact that $P_{0n_0}(t) + P_{1n_0}(t) = 1$ we can solve this pair of equations to give

$$P_{0n_0}(t) = \delta_{0,n_0} \mathrm{e}^{-t/\tau_c} + \frac{k_-}{\tau_c}(1 - \mathrm{e}^{-t/\tau_c}), \quad \tau_c = \frac{1}{k_- + k_+}.$$

A number of results follow from this. First τ_c is the relaxation time of the DMNP with $P_{mn_0}(t) \to \rho_m$ in the limit $t \to \infty$ and

$$\rho_0 = \frac{k_-}{k_+ + k_-}, \quad \rho_1 = \frac{k_+}{k_+ + k_-}. \tag{2.46}$$

In the stationary state, the dichotomous noise term has the mean

$$\langle \xi(t) \rangle = (\nu_+ + \nu_-)\langle n(t) \rangle - \Gamma_- = \rho_1 \nu_+ - \rho_0 \nu_-. \tag{2.47}$$

Suppose, in particular, that the DMNP is unbiased so that $\langle \xi(t) \rangle = 0$. The stationary autocorrelation function is then given by

$$\begin{aligned} \langle \xi(t)\xi(t') \rangle &= \nu_-^2 - 2\nu_-(\nu_+ + \nu_-)\rho_1 \\ &\quad + (\nu_+ + \nu_-)^2 \langle n(t) n(t') \rangle \\ &= \frac{D}{\tau_c} \mathrm{e}^{-|t-t'|/\tau_c}, \end{aligned} \tag{2.48}$$

with noise amplitude $D = k_+ k_- \tau_c^3 (\nu_+ + \nu_-)^2$. This shows that the DMNP provides an alternative form of colored noise to an Ornstein-Uhlenbeck process [105].

Given the initial conditions $x(0) = x_0, n(0) = n_0$, introduce the probability density $p_n(x, t | x_0, n_0, 0)$ the forward CK equation reduces to

$$\frac{\partial p_0}{\partial t} = v_- \frac{\partial p_0}{\partial x} + k_+ p_1 - k_- p_0 \tag{2.49a}$$

$$\frac{\partial p_1}{\partial t} = -v_+ \frac{\partial p_1}{\partial x} + k_- p_0 - k_+ p_1. \tag{2.49b}$$

In applications one is typically interested in the marginal density $p(x,t) = p_0(x,t) + p_1(x,t)$, which can be used to calculate moments of p such as the mean and variance,

$$\langle X(t)\rangle = \int xp(x,t)dx,\ \mathrm{Var}[X(t)] = \int x^2 p(x,t)dx - \langle x(t)\rangle^2.$$

In the unbiased case $v_\pm = v, k_\pm = k$, the marginal probability density $p(x,t)$ satisfies the telegrapher's equation [25, 114]

$$\left[\frac{\partial^2}{\partial t^2} + 2k\frac{\partial}{\partial t} - v^2\frac{\partial^2}{\partial x^2}\right] p(x,t) = 0. \tag{2.50}$$

(The individual densities $p_{0,1}$ satisfy the same equations.) The telegrapher's equation can be solved explicitly for a variety of initial conditions. More generally, the short-time behavior (for $t \ll \tau_c = 1/2k$) is characterized by wave-like propagation with $\langle x(t)\rangle^2 \sim (Vt)^2$, whereas the long-time behavior ($t \gg \tau_c$) is diffusive with $\langle x^2(t)\rangle \sim 2Dt$, $D = v^2/2k$. As an explicit example, the solution for the initial conditions $p(x,0) = \delta(x)$ and $\partial_t p(x,0) = 0$ is given by

$$\begin{aligned} p(x,t) = {} & \frac{e^{-kt}}{2}[\delta(x-vt) + \delta(x+vt)] \\ & + \frac{k\mathrm{e}^{-kt}}{2v}\left[I_0(k\sqrt{t^2 - x^2/v^2}) + \frac{t}{\sqrt{t^2 - x^2/v^2}} I_0(k\sqrt{t^2 - x^2/v^2})\right] \\ & \times [\Theta(x+vt) - \Theta(x-vt)], \end{aligned}$$

where I_n is the modified Bessel function of n-th order and Θ is the Heaviside function. The first two terms clearly represent the ballistic propagation of the initial data along characteristics $x = \pm vt$, whereas the Bessel function terms asymptotically approach Gaussians in the large time limit. The steady-state equation for $p(x)$ is simply $p''(x) = 0$, which from integrability means that $p(x) = 0$ point-wise. This is consistent with the observation that the above explicit solution satisfies $p(x,t) \to 0$ as $t \to \infty$. Examples of dichotomous noise processes in biological switching will be given in section 8.3.

Quasi-steady-state (QSS) diffusion approximation. For small but non-zero ε, one can use perturbation theory to derive lowest order corrections to the deterministic mean field equation, which leads to a Langevin equation with noise amplitude $O(\sqrt{\varepsilon})$ [188]. More specifically, perturbations of the mean-field equation (2.38) can be analyzed using a quasi-steady-state (QSS) diffusion or adiabatic approximation, in which the CK equation (2.36) is approximated by a Fokker-Planck (FP) equation for the total density $C(x,t) = \sum_n \rho_n(x,t)$. The QSS approximation was first developed from a probabilistic perspective by Papanicolaou [201], see also [105]. It has subsequently been applied to a wide range of problems in biology, including bacterial chemotaxis [87, 128, 200], wave-like behavior in models of slow axonal transport [99, 100, 212], molecular motor-based models of random intermittent search [188, 189], and gene networks [151, 191]. The QSS reduction proceeds as follows:

(i) Decompose the probability density as

$$\rho_n(x,t) = C(x,t)\rho_n^*(x) + \varepsilon w_n(x,t),$$

where $\sum_n \rho_n(x,t) = C(x,t)$ and $\sum_n w_n(x,t) = 0$. Substituting into (2.36) yields

$$\rho_n^*(x)\frac{\partial C}{\partial t} + \varepsilon\frac{\partial w_n}{\partial t} = -\frac{\partial F_n(x)[C\rho_n^*(x) + \varepsilon w_n]}{\partial x} + \frac{1}{\varepsilon}\sum_{m\in\Gamma} A_{nm}(x)[C\rho_m^*(x) + \varepsilon w_m]$$

Summing both sides with respect to n then gives

$$\frac{\partial C}{\partial t} = -\frac{\partial \overline{F}(x)C}{\partial x} - \varepsilon \sum_{n\in\Gamma} \frac{\partial F_n(x) w_n}{\partial x}. \tag{2.51}$$

(ii) Using the equation for C and the fact that $\mathbf{A}\rho^* = 0$, we have

$$\varepsilon \frac{\partial w_n}{\partial t} = \sum_{m\in\Gamma} A_{nm}(x) w_m - \frac{\partial [F_n(x)\rho_n^*(x)C]}{\partial x} + \rho_n^*(x)\frac{\partial \overline{F}(x)C}{\partial x}$$
$$- \varepsilon \sum_{m\in\Gamma} [\delta_{m,n} - \rho_n^*(x)] \frac{\partial F_m(x) w_m}{\partial x}$$

(iii) Introduce the asymptotic expansion

$$\mathbf{w} \sim \mathbf{w}^{(0)} + \varepsilon \mathbf{w}^{(1)} + \varepsilon^2 \mathbf{w}^{(2)} + \ldots$$

and collect $O(1)$ terms:

$$\sum_{m\in\Gamma} A_{nm}(x) w_m^{(0)} = \frac{\partial [F_n(x)\rho_n^*(x)C(x,t)]}{\partial x} - \rho_n^*(x)\frac{\partial \overline{F}(x)C}{\partial x}. \tag{2.52}$$

The Fredholm alternative theorem show that this has a solution, which is unique on imposing the condition $\sum_n w_n^{(0)}(x,t) = 0$.

(iv) Combining equations (2.52) and (2.51) shows that C evolves according to the FP equation

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial x}(\overline{F}(x)C) + \varepsilon \frac{\partial}{\partial x}\left(D(x)\frac{\partial C}{\partial x}\right) \tag{2.53}$$

with the diffusion coefficient $D(x)$ given by

$$D(x) = \sum_{n\in\Gamma} Z_n(x) F_n(x), \tag{2.54}$$

where $Z_n(x)$ is the unique solution to

$$\sum_{m\in\Gamma} A_{nm}(x) Z_m(x) = [\overline{F}(x) - F_n(x)]\rho_n^*(x). \tag{2.55}$$

with $\sum_m Z_m(x) = 0$. We have dropped $O(\varepsilon)$ corrections to the drift term. For $N_0 > 2$ one typically has to solve equation (2.55) numerically in order to find the pseudo-inverse of $\mathbf{A}$. However, in the special case of a two-state discrete process ($n = 0, 1$), with

$$\mathbf{A}(x) = \begin{pmatrix} -\alpha(x) & \beta(x) \\ \alpha(x) & -\beta(x) \end{pmatrix},$$

one has the explicit solution

$$D(x) = \frac{\beta(x)[F_0(x) - \overline{F}(x)]F_0(x) + \alpha(x)[F_1(x) - \overline{F}(x)]F_1(x)}{2[\alpha(x) + \beta(x)]^2}. \tag{2.56}$$

At a fixed point x_* of the deterministic equation $\dot{x} = \overline{F}(x)$, we have $\overline{F}(x) = 0$ and $\beta(x_*)F_0(x_*) = -\alpha(x_*)F_1(x_*)$. Hence, we have the reduced expression

$$D(x_*) = \frac{|F_0(x_*)F_1(x_*)|}{\alpha(x_*) + \beta(x_*)}. \tag{2.57}$$

This result will be used in our study of metastability in simple gene networks (see section 5).

One subtle point is the nature of boundary conditions under the QSS reduction, since the FP equation is a second-order parabolic PDE and one typically has to specify one boundary condition at either end. It follows that for $N_0 > 2$, there is a mismatch in the number of boundary conditions between the CK equation and the FP equation. This implies that the QSS reduction may break down in a small neighborhood of either boundary, as reflected by the existence of boundary layers. One way to eliminate the existence of boundary layers is to ensure that the boundary conditions of the CK equation are compatible with the QSS reduction. For example at $x = L$,

$$\rho_n(L,t) = \rho_n^*(L)C(L,t) + O(\varepsilon)$$

This holds for the particular problems considered in this review.

3. Simple examples of gene regulatory networks

We now turn to some simple stochastic models of gene expression, based on the two major components of the central dogma, namely, transcription and translation, see Fig. 1. We begin by considering unregulated gene expression, and then turn to some simple regulatory networks with or without feedback. In the former case, the resulting network can act as a bistable switch. We also briefly discuss one of the spatial aspects of gene regulation, namely, facilitated diffusion underlying DNA-protein interactions. For a more detailed discussion of modeling gene regulatory networks see the book by Sneppen [230]

3.1. Unregulated transcription and translation

Suppose that we ignore any regulation of the promoter as in Fig. 1(a), 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

$$m \xrightarrow{k\Omega} m+1, \quad m \xrightarrow{\gamma} m-1$$

with corresponding kinetic equation for the concentration $x = m/\Omega$, where Ω is proportional to 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

$$\frac{dP(m,t)}{dt} = -k\Omega P(m,t) + k\Omega P(m-1,t) - \gamma m P(m,t) + \gamma(m+1)P(m+1,t) \quad (3.1)$$

for $m \geq 0$ and $P(-1,t) \equiv 0$. Equation (3.1) 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 (see also section 2.2). The simplest way to see this is to introduce the generating function†

$$G(z,t) = \sum_{m \geq 0} z^m P(m,t),$$

† The application of generating functions to determine the probability distribution of more complicated gene networks has been developed by Swain and collaborators [205, 225].

and use equation (3.1) to show that

$$\frac{\partial G}{\partial t} + \gamma(z-1)\frac{\partial G}{\partial z} = k\Omega(z-1)G.$$

This is a linear first-order PDE with non-constant coefficients. A standard method for solving such equations is the *method of characteristics* [219]. 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

$$\frac{dz}{dt} = \gamma(z-1), \quad \frac{dG}{dt} = k\Omega(z-1)G.$$

Solving for $z(t)$, we have $z(t) = 1 + s\mathrm{e}^{\gamma t}$, where s parameterizes the initial data. Then

$$\frac{dG}{dt} = k\Omega s\mathrm{e}^{\gamma t}G, \quad G(t) = F(s)\exp\left(k\Omega s\mathrm{e}^{\gamma t}/\gamma\right)$$

for some function F determined by the initial data. In order to obtain the solution $G(z,t)$ we eliminate s in terms of z, which gives

$$G(z,t) = F([z-1]\mathrm{e}^{-\gamma t})\exp\left(k\Omega(z-1)/\gamma\right). \tag{3.2}$$

Since $G(1,t) = 1$, we require $F(0) = 1$. Moreover, given the initial condition $\rho_m(0) = \delta_{m,0}$, we have $G(z,0) = 1$ and $F(z) = \mathrm{e}^{-k\Omega z/\gamma}$. It follows that

$$G(z,t) = \mathrm{e}^{k\Omega\left(1-\mathrm{e}^{-\gamma t}\right)(z-1)/\gamma}, \tag{3.3}$$

If we now Taylor expand $G(z,t)$ in powers of z we find that

$$P(m,t) = \mathrm{e}^{-\lambda(t)}\frac{\lambda(t)^m}{m!}, \quad \lambda(t) = \frac{k\Omega}{\gamma}(1-\mathrm{e}^{-\gamma t}), \tag{3.4}$$

which is a time-dependent Poisson distribution of rate $\lambda(t)$. Hence, in the limit $t \to \infty$ we obtain a stationary Poisson process (see section 2.2) with

$$P(m) = \mathrm{e}^{-\lambda_0}\frac{\lambda_0^m}{m!}, \quad \lambda_0 = k\Omega/\gamma. \tag{3.5}$$

It follows that $\langle m\rangle = \lambda_0$ and $\mathrm{var}[m] = \lambda_0$. 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 networks is the so-called Fano factor:

$$\text{Fano factor} = \frac{\langle m^2\rangle - \langle m\rangle^2}{\langle m\rangle}. \tag{3.6}$$

For the unregulated process, the Fano factor is one. This is a baseline value for quantifying the effects of gene regulation on the level of noise.

3.2. Two-state gene regulatory network

We now consider the simple gene regulatory network shown in Fig. 2(d). This consists of a gene that can be in either an active or inactive state. 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 when the gene is inactive. For simplicity, we do not explicitly keep track of the amount of mRNA, that is, we lump together the stages of transcription and translation. The resulting reaction scheme is thus

$$I \underset{k_-}{\overset{k_+}{\rightleftharpoons}} A \xrightarrow{r} X \xrightarrow{\gamma} \emptyset,$$

where A and I denote the active and inactive states of the gene. Suppose that the number of X proteins is sufficiently large so the dynamics can be expressed in terms of a continuous-valued protein concentration x [147, 258]. The latter then evolves according to the (piecewise) deterministic equation

$$\frac{dx}{dt} = F_n(x) \equiv rn(t) - \gamma x, \tag{3.7}$$

where the discrete random variable $n(t)$ represents the current state of the gene with $n(t) = 1$ (active) or $n(t) = 0$ (inactive). Note that at the single gene level one can equally well identify $n(t)$ with the state of a randomly switching environment - this equivalence breaks down at the population level, see section 3.3. It is also possible to have a nonzero protein production rate in both states with $0 < r_0 < r_1$ [228]. Equation (3.7) is a simple example of a PDMP (see section 2.6), whose corresponding CK equation is

$$\frac{\partial \rho_0}{\partial t} = -\frac{\partial}{\partial x}(-\gamma x \rho_0(x,t)) + k_- \rho_1(x,t) - k_+ \rho_0(x,t) \tag{3.8a}$$

$$\frac{\partial \rho_1}{\partial t} = -\frac{\partial}{\partial x}([r - \gamma x]\rho_1(x,t)) + k_+ \rho_0(x,t) - k_- \rho_1(x,t), \tag{3.8b}$$

The CK equation is supplemented by the no-flux boundary conditions $J(x,t) = 0$ at $x = 0, r/\gamma$, where $J(x,t) = F_0(x)\rho_0(x,t) + F_1(x)\rho_1(x,t)$. That is, $\rho_1(0,t) = 0$ and $\rho_0(r/\gamma, t) = 0$. 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

$$\frac{dx}{dt} = r\langle n \rangle - \gamma x = \frac{rk_+}{k_+ + k_-} - \gamma x. \tag{3.9}$$

Following Ref. [147], we will characterize the long-time behavior of the system in terms of the steady-state solution, which satisfies

$$\frac{d}{dx}(-\gamma x \rho_0(x)) \quad = k_- \rho_1(x) - k_+ \rho_0(x) \tag{3.10a}$$

$$\frac{d}{dx}([r - \gamma x]\rho_1(x)) = k_+ \rho_0(x) - k_- \rho_1(x). \tag{3.10b}$$

The no-flux boundary conditions imply that $\rho_0(r/\gamma) = 0$ and $\rho_1(0) = 0$. First, note that we can take $x \in [0, r/\gamma]$ and impose the normalization condition

$$\int_0^{r/\gamma} [\rho_0(x) + \rho_1(x)]dx = 1.$$

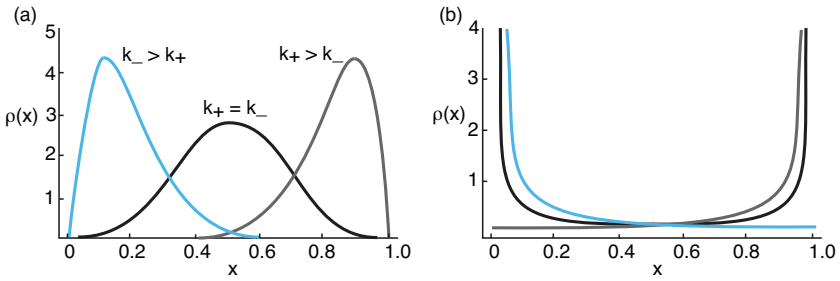


Figure 6. Sketch of steady-state protein density $\rho(x)$ for a simple regulated network in which the promoter transitions between an active and inactive state at rates $k_\pm$. (a) Case $k_\pm/\gamma > 1$: there is a graded density that is biased towards $x = 0, 1$ depending on the ratio k_+/k_-. (b) Case $k_\pm/\gamma < 1$: there is a binary density that is concentrated around $x = 0, 1$ depending on the ratio k_+/k_-

Integrating equations (3.10a) and (3.10b) with respect to x then leads to the constraints

$$\int_0^{r/\gamma} \rho_0(x)dx = \frac{k_-}{k_- + k_+}, \quad \int_0^{r/\gamma} \rho_1(x)dx = \frac{k_+}{k_- + k_+}.$$

Adding equations (3.10a) and (3.10b) we can solve for $\rho_0(x)$ in terms of $\rho_1(x)$ and then generate a closed differential equation for $\rho_1(x)$, see the derivation of equation (2.45). We thus obtain a solution of the form

$$\rho_0(x) = C\,(\gamma x)^{-1+k_+/\gamma}(r-\gamma x)^{k_-/\gamma}, \quad \rho_1(x) = C\,(\gamma x)^{k_+/\gamma}(r-\gamma x)^{-1+k_-/\gamma} \tag{3.11}$$

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 $\rho(x) = \rho_0(x) + \rho_1(x)$ is given by [147]

$$\rho(x) = \frac{x^{k_+/\gamma - 1}(1-x)^{k_-/\gamma - 1}}{B(k_+/\gamma, k_-/\gamma)}. \tag{3.12}$$

The steady-state density $\rho(x)$ for various values of $K_\pm = k_\pm/\gamma$ is sketched in Fig. 6. When the switching rates $k_\pm$between the active and inactive gene states are faster than the rate of degradation γ then the steady-state density is unimodal (graded). On the other hand, 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. This suggests that if switching between promoter states is much slower than other processes then one can have a transcriptional contribution to protein bursting [147], see also section 4. Slow switching is more likely to occur in eukaryotic rather than prokaryotic gene expression. That is, the presence of nucleosomes and the packing of DNA-nucleosome complexes into chromatin tends to render promotors inaccessible to the transcriptional machinery in eukaryotic cells, thus slowing down transitions between active and repressed promotor states.

Hufton et al. [135] have recently generalized the two-state environmental switching model to include the effects of molecular noise and nonlinearities. These can be included by carrying out a system-size expansion of the master equation for protein synthesis when the environment is in state n. This leads to modified CK equations that describe the time evolution of the pdfs for the piecewise stochastic differential equation

$$dX(t) = F_n(X)dt + \sqrt{\frac{\sigma_n(X)}{\Omega}}dW(t) \tag{3.13}$$

for $n(t) = n \in \{0, 1\}$, Ω the system size, and

$$F_n(x) = r_n - \gamma x, \quad \sigma_n(x) = r_n + \gamma x. \tag{3.14}$$

We now have a combination of discrete environmental noise and continuous intrinsic molecular noise. Hufton *et al.* show how to approximate the steady-state densities by carrying out a linear noise approximation, which can be applied even when F_n, σ_n and $k_\pm$ are nonlinear functions of x (see also [237]).

It is important to note that both the PDMP and its extension to include demographic noise are approximations of the full chemical master equation describing the stochastic variations in the state of the gene and the number of proteins. Rigorous error bounds on the accuracy of a PDMP approximation to a chemical master equation based on a partial thermodynamic limit have been obtained elsewhere [141]. In particular one finds that under suitable conditions, the PDMP approximates the marginal distribution of the discrete species (in our case the discrete state n) and moments of the continuous species (in our case protein concentration) up to an error of $O(\Omega^{-1})$.

3.3. Population-level correlations in gene expression.

As we have already indicated in section 3.2, there are two distinct mechanisms for switching (see Fig. 7):

(i) The first is intrinsic switching within a cell that is driven by fluctuations in the binding and unbinding of a transcription factor Y [147, 151, 258]. The dynamics of Y is assumed to be independent of protein X, that is, there is no feedback. Each copy of the gene within a cell or across a population of cells switches independently. 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.

(ii) The second is extrinsic switching driven by a randomly switching environment [228]. Whether or not a promoter site is occupied now depends on the state $n(t)$ of the environment, and will be common to all cells evolving in the same environment. The discrete environmental states could represent the presence of some extracellular metabolite or signaling molecule, perhaps arising from changes in the physiological or hormonal state that a cell experiences in a multicellular organism. For concreteness, we will assume that each cell carries a single copy of the gene of interest. However, one could equally well have multiple copies of the gene in each cell, and a combination of intrinsic and extrinsic switching.

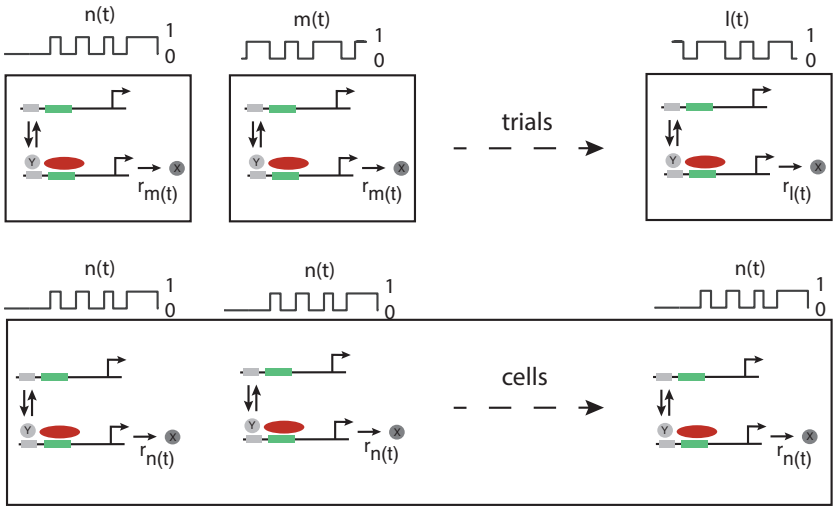


Figure 7. Diagram illustrating the non-equivalence of population averaging over many cells evolving in the same switching environment and averaging over trails of a single cell evolving in multiple realizations of a switching environment. Here $r_{n(t)}$ denotes the switching protein production rate in a given realization $n(t)$ of the environment.

These two mechanisms are equivalent at the single cell level, but are significantly different at the population level, since the common random environment introduces statistical correlations between the cells.

Numerically speaking, in the case of a single gene one could determine the stationary density by carrying out many independent realizations of the doubly stochastic process given by equation (3.13) (after allowing for transients) and constructing a histogram of the fraction of trials at some given time $T \gg 0$ for which $n(T) = n$ and $X(t) \in (x_l, x_l + \Delta x)$, after partitioning the continuous variable x into bins labeled by l. In many stochastic models, an equivalent way of determining the statistics is to consider a large population of identical, independent cells and to carry out an ensemble average over a single trial. However, the situation is more complicated in the case of cells evolving in the same randomly switching environment. That is, for a single realization of the stochastic process for the population, the cells are not independent. In other words, trial averaging and ensemble averaging are not equivalent unless one averages the latter over multiple realizations of the environment. Thus there are additional statistics contained within a single realization of a population of cells. This issue has recently been explored in a different context, namely, diffusion processes [50, 53, 54] and random walks [164] in switching environments (see section 8.2). Here we briefly sketch how the mathematical framework for studying diffusion in random environments can be extended to gene networks.

Consider the piecewise deterministic PDE

$$\frac{\partial P(x,t)}{\partial t} = \mathbb{L}_n P(x,t) \equiv -\frac{\partial}{\partial x}(F_n(x)P(x,t)) \tag{3.15}$$

for $n(t) = n \in \{0, 1\}$. Equation (3.15) is supplemented by no-flux boundary conditions at $x = 0, 1$. We are imagining that there is an infinite population of identical cells all evolving in the same switching environment, such that in between switches, the

probability density that any one cell has concentration x at time t is $P(x,t)$. (A finite population of cells would add yet another level of stochasticity.) The important point to note is that $P(x,t)$ is itself a random variable. In order to analyze equation (3.15), we follow the approach of [50] by discretizing x using a finite-difference scheme so that the system is converted to a PDMP. Introduce the lattice spacing a such that $(\mathscr{N}+1)a = r/\gamma$ for integer $\mathscr{N}$, and let $P_j(t) = P(aj,t)$ etc., $j = 0,\ldots,\mathscr{N}+1$. Also set $F_j^{(n)} = F_n(ja)$. Then for $n(t) = n$

$$\frac{dP_i}{dt} = -\sum_{j=1}^{\mathscr{N}} K_{ij}^{(n)} P_j, \quad i = 1,\ldots,\mathscr{N} \tag{3.16}$$

Away from the boundaries ($i \neq 1, \mathscr{N}$),

$$K_{ij}^{(n)} = \frac{1}{a}[\delta_{i,j-1} - \delta_{i,j}]F_j^{(n)}. \tag{3.17}$$

At the boundaries we require $P_0(t) = 0$ when $n = 1$ and $P_{N+1}(t) = 0$ when $n = 0$. These conditions can be implemented by taking

$$K_{1j}^{(1)} = \frac{1}{2a}\delta_{j,2}F_j^{(1)}, \quad K_{\mathscr{N}j}^{(0)} = \frac{1}{2a}\delta_{j,\mathscr{N}-1}F_j^{(0)}.$$

Let $\mathbf{P}(t) = (P_1(t),\ldots,P_{\mathscr{N}}(t))$ and introduce the probability density

$$\text{Prob}\{\mathbf{P}(t) \in (\mathbf{P}, \mathbf{P}+d\mathbf{P}), n(t) = n\} = \varrho_n(\mathbf{P},t)d\mathbf{P}, \tag{3.18}$$

where we have dropped the explicit dependence on initial conditions. Following the analysis of PDMPs in section 2.6, the CK for the discretized piecewise deterministic PDE is

$$\frac{\partial \varrho_n}{\partial t} = \sum_{i=1}^{\mathscr{N}} \frac{\partial}{\partial P_i}\left[\left(\sum_{j=1}^{\mathscr{N}} K_{ij}^{(n)} P_j\right)\varrho_n(\mathbf{P},t)\right] + \sum_{m=0,1} A_{nm}\varrho_m(\mathbf{P},t), \tag{3.19}$$

with $A_{00} = -A_{10} = -k_+$ and $A_{01} = -A_{11} = k_-$. Since the Liouville term in the CK equation is linear in $\mathbf{P}$, we can derive a closed set of equations for the moments of $\mathbf{P}$. For the sake of illustration, we will calculate the first and second moments. Let

$$p_{n,j}(t) = \mathbb{E}[P_j(t)1_{n(t)=n}] = \int \varrho_n(\mathbf{P},t)P_j(t)d\mathbf{P}, \tag{3.20}$$

where $1_{n(t)=n}$ is the indicator function that is equal to one if $n(t) = n$ and is zero otherwise. Multiplying both sides of the CK equation (3.19) by $P_k(t)$ and integrating with respect to $\mathbf{P}$ gives (after integrating by parts and using $\varrho_n(\mathbf{P},t) \to 0$ as $\mathbf{P} \to \infty$)

$$\frac{dp_{n,k}}{dt} = -\sum_{j=1}^{\mathscr{N}} K_{kj}^{(n)} p_{n,j} + \sum_{m=0,1} A_{nm}p_{m,k}.$$

We have assumed that the initial discrete state is distributed according to the stationary distribution of the matrix $\mathbf{A}$. If we now retake the continuum limit $a \to 0$, we recover the original CK equation (2.36) after making the identification

$$\rho_n(x,t) = \mathbb{E}[P(x,t))1_{n(t)=n}]. \tag{3.21}$$

Hence, we can equate the probability density $\rho_n(x,t)$, obtained by trial averaging over multiple realizations of the stochastic process (3.7), with the expectation of the stochastic density $P(x,t)$ of a population averaged over multiple realizations of the environment such that $n(t) = n$. However, from the population perspective, there are

additional higher-order statistics arising from a large population evolving in the same random environment. In particular, consider the second-order moments

$$p_{n,kl}(t) = \mathbb{E}[P_k(t)P_l(t)1_{n(t)=n}] = \int \varrho_n(\mathbf{P},t)P_k(t)P_l(t)d\mathbf{P}.$$

Multiplying both sides of the CK equation (3.19) by $P_k(t)P_l(t)$ and integrating with respect to $\mathbf{P}$ gives (after integration by parts)

$$\frac{dp_{n,kl}}{dt} = -\sum_{j=1}^{\mathscr{N}} K_{kj}^{(n)} p_{n,jl} - \sum_{j=1}^{\mathscr{N}} K_{lj}^{(n)} p_{n,jk} + \sum_{m=0,1} A_{nm} p_{m,kl}.$$

If we now retake the continuum limit $a \to 0$, we obtain a system of equations for the equal-time two-point correlations

$$C_n(x,y,t) = \mathbb{E}[P(x,t)P(y,t)1_{n(t)=n}], \tag{3.22}$$

given by

$$\frac{\partial C_n}{\partial t} = -\frac{\partial}{\partial x}(F_n(x)C_n) - \frac{\partial}{\partial y}(F_n(y)C_n) + \sum_{m=0,1} A_{nm}C_m. \tag{3.23}$$

The boundary conditions are

$$C_1(0,y,t) = C_1(x,0,t) = 0, \quad C_0(L,y,t) = C_0(x,L,t) = 0,$$

and we have the marginal densities

$$\int_0^{r/\gamma} C_n(x,y,t)dx = \int_0^{r/\gamma} C_n(y,x,t)dx = \rho_n(y.t).$$

One can interpret $C_n(x,y,t)$ as the joint probability (with respect to trials) that two cells evolving in the same environment have concentrations x and y at time t when the environment is in state n [50]. Similarly the r-th moments determine the joint probability density for r cells, which evolves according to an r-th order quasilinear system. The latter can be analyzed using the method of characteristics. In the case of the second-order moments, one finds that the variances $\sigma_n^2(x) = C_n(x,x)$ have the asymptotic behavior

$$\sigma_0^2(x) \sim k_- x^{-2+k_+/\gamma}, \quad x \sim 0,$$
$$\sigma_1^2(x) \sim k_+[r-\gamma x]^{-2+k_-/\gamma}, \quad x \sim r/\gamma,$$

Piecewise deterministic PDEs will be explored further in section 8.2, within the context of diffusion in randomly switching environments.

Note that the above analysis of population-level statistical correlations can be extended to more complicated systems. First, there could be multiple discrete states of the environment, $n \in \{0,1,\dots N_0\}$, with associated functions $F_n(x)$ such that $\dot{x} = F_n(x)$ for $n(t) = n$. Second, the individual gene networks could involve several interacting genes and protein products resulting in a nonlinear feedback system on $\mathbb{R}^m$, $m > 1$ (see section 3.5). Now the piecewise-deterministic dynamics is nonlinear and could exhibit multistability and limit cycle oscillations. Third, the switching rates of the environment could depend on the total concentration of proteins produced in a population of cells, as occurs in quorum sensing (see section 7). In this case, the CK Eq. (3.19) for the higher-level PDMP would be a nonlinear function of the probability distribution, resulting in a moment closure problem. Finally, one could include demographic noise in the protein concentration along the lines of Ref. [135]. The piecewise-deterministic first-order PDE (3.15) is replaced by a piecewise-deterministic Fokker-Planck equation. The derivation of the moment equations via finite differences then involves the discrete Laplacian, see also section 8.2.

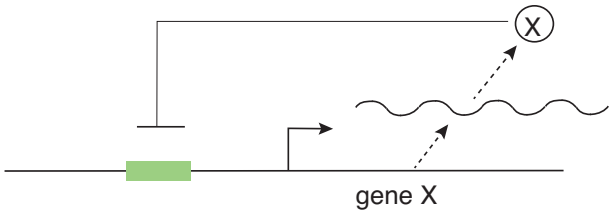


Figure 8. An autoregulatory network. A gene X is repressed by its own protein product.

3.4. Autoregulatory network

One of the simplest examples of a gene network with regulatory feedback is autoregulation, in which a gene is directly regulated by its own gene product [235], see Fig. 8. A simple kinetic model of autoregulatory feedback is

$$\frac{dx_1}{dt} = -\gamma x_1 + F(x_2), \quad \frac{dx_2}{dt} = rx_1 - \gamma_p x_2, \tag{3.24}$$

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 $F(x)$ represents the nonlinear feedback effect of the protein on the transcription of mRNA. A typical choice for F in the case of an activator (a) or repressor (r) is to express it in terms of a Hill function. That is, $F = F_a$ or $F = F_r$ with

$$F_a(x) = \sigma_0 + \frac{\sigma_1 \kappa x^2}{1 + \kappa x^2}, \quad F_r(x) = \frac{\sigma_r}{1 + \kappa x^2} \tag{3.25}$$

One can view the quadratic terms as arising from dimerization. A fixed point (x_1^*, x_2^*) of the kinetic equations satisfies $x_1^* = F(x_2^*)/\gamma$ with x_2^* a root of the cubic

$$f(x_2^*) \equiv -\gamma_p x_2^* + \frac{r}{\gamma} F(x_2^*) = 0.$$

One can thus determine the number of fixed points and their stability using a graphical construction, see Fig. 9. In the case of negative feedback there is a single stable fixed point, whereas with positive feedback the network can be monostable or bistable. For the moment, suppose that there is only one stable fixed point.

Linear noise approximation. In order to take into account the effects of molecular noise due to finite copy numbers, we need to write down the corresponding master equation. Let $P = P(m, n, t)$ denote the probability that there are m mRNA and n proteins at time t. Then (see section 2.4)

$$\begin{aligned} \frac{dP}{dt} &= \Omega F(n) P(m-1, n, t) + \gamma(m+1) P(m+1, n, t) \\ &\quad + rmP(m, n-1, t) + \gamma_p(n+1) P(m, n+1, t) \\ &\quad - \left[\Omega F(n) + \gamma m) + rm + \gamma_p n\right] P(m, n, t). \end{aligned} \tag{3.26}$$

We will estimate the resulting Fano factor by carrying out a linear noise approximation as outlined in section 2.4. First, rewrite the kinetic equations in the general form (2.13) with two chemical species ($K = 2$) and four single-step reactions ($R = 4$). For example, taking $a = 1, 2$ to be mRNA production and degradation, respectively, we

have $S_{i,1} = \delta_{i,1}, S_{i,2} = -\delta_{i,1}, f_1(\mathbf{x}) = F(x_2)$, and $f_2(\mathbf{x}) = \gamma x_1$. Expressing the master equation as (2.18) and carrying out a diffusion approximation then leads to the FP equation (2.20) with drift terms

$$V_1(\mathbf{x}) = F(x_2) - \gamma x_1, \quad V_2(\mathbf{x}) = r x_1 - \gamma_p x_2. \tag{3.27}$$

and a diagonal diffusion matrix $\mathbf{D}$ with non-zero components

$$D_{11} = F(x_2) + \gamma x_1, \quad D_{22} = r x_1 + \gamma_p x_2 \tag{3.28}$$

Linearizing the corresponding Langevin equation about the unique fixed point by setting $X_i(t) = x_i^* + \Omega^{-1/2} Y_i(t)$ then yields the OU process (2.26) 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

$$\mathbf{A}\Sigma + \Sigma \mathbf{A}^T = -\mathbf{D} \tag{3.29}$$

with

$$\mathbf{A} = \begin{pmatrix} -\gamma & \mu \\ r & -\gamma_p \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} F(x_2^*) + \gamma x_1^* & 0 \\ 0 & r x_1^* + \gamma_p x_2^* \end{pmatrix}, \quad \mu = F'(x_2^*).$$

Note that $\mu > 0$ for an activator and $\mu < 0$ for a repressor. Solving the matrix equation (3.29) yields

$$\Sigma_{12} = \Sigma_{21} = \frac{\eta}{1+\eta}\left(1 - \frac{\phi}{1+b\phi}\right) x_2^*, \quad \Sigma_{22} = x_2^* + \frac{r}{\gamma_p}\Sigma_{12},$$

where

$$b = \frac{r}{\gamma}, \quad \eta = \frac{\gamma_p}{\gamma}, \quad \phi = -\frac{\mu}{\gamma_p}.$$

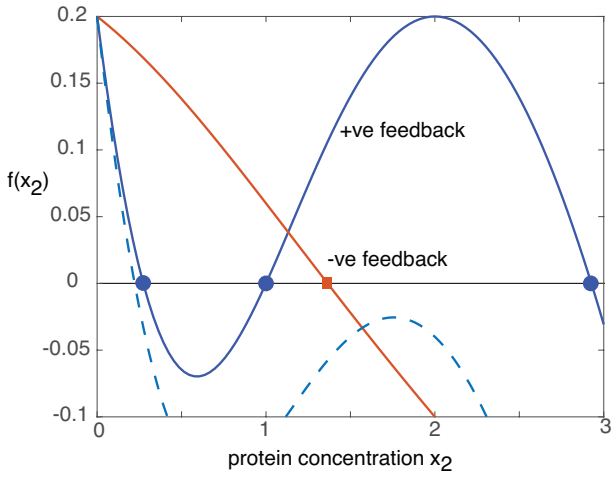


Figure 9. Fixed points of deterministic autoregulatory network. Network is monostable in the case of negative feedback (red curve) and weak positive feedback (dashed blue curve), but can exhibit bistability in the case of strong positive feedback (solid blue curve).

Here b is the so-called burst size (see section 4.1), η is the ratio of degradation rates, and ϕ describes the strength and sign of the feedback. It follows that the Fano factor for proteins is

$$\frac{\text{var}[n]}{\langle n\rangle} = 1 + \frac{b}{1+\eta}\left(1 - \frac{\phi}{1+b\phi}\right). \tag{3.30}$$

In the absence of feedback ($\phi = 0$), the Fano factor is $1 + b/(1+\eta)$, which is larger than the unregulated model of section 3.1, because we have included the effects of protein fluctuations. When negative feedback is included, the Fano factor decreases since $\phi > 0$. This establishes that negative feedback can reduce fluctuations in protein number.

Promoter noise. Let us assume that the associated gene can exist in two states, an off state ($s = 0$) where there is a baseline level σ_0 of mRNA transcription and an on state ($s = 1$) that produces mRNA at a higher rate σ_1. Transitions between the on and off states of the gene are given by a two-state Markov process

$$(\text{off}) \underset{\beta(x)}{\overset{\alpha(x)}{\rightleftharpoons}} (\text{on}) \quad \alpha(x) = \alpha_0 x^2, \quad \beta(x) = \beta_0, \tag{3.31}$$

with $\alpha_0/\beta_0 = \kappa$ and $x = n/N$ the (number) concentration of proteins. Here N is a characteristic number of proteins, say, and acts as a dimensionless system-size parameter. For simplicity, we will ignore mRNA dynamics by taking the concentration of mRNA at quasi-equilibrium and absorbing the factor r/γ into the production rates σ_0, σ_1. Let $P_s(n,t)$ be the probability that there are n proteins and the gene is in state s at time t. We then have the pair of equations

$$\frac{dP_0(n,t)}{dt} = \mathbb{L}_0 P_0(n,t) - \alpha(n/N)P_0(n,t) + \beta(n/N)P_1(n,t) \tag{3.32a}$$

$$\frac{dP_1(n,t)}{dt} = \mathbb{L}_1 P_1(n,t) + \alpha(n/N)P_0(n,t) - \beta(n/N)P_1(n,t) \tag{3.32b}$$

where

$$\mathbb{L}_s[f(n)] = N(s\sigma_1 + \sigma_0)[(f(n-1) - f(n)] + \gamma_p[(n+1)f(n+1) - nf(n)].$$

Equations (3.32a) and (3.32b) can either be viewed as a master equation for the discrete variables (s,n) or as a Chapman-Kolmogorov equation for the discrete variables n that evolves in a switching environment labeled by $s = 0, 1$. The latter interpretation links the model to a stochastic hybrid system that is obtained by carrying out a system-size expansion with respect to N, as detailed below equation (3.26). That is, for large N

$$\frac{\partial \rho_0(x,t)}{dt} = -\frac{\partial V_0(x)\rho_0(x,t)}{\partial x} + \frac{1}{2N}\frac{\partial^2 D_0(x)\rho_0(x,t)}{\partial x^2} - \frac{\alpha(x)}{\varepsilon}\rho_0(x,t) + \frac{\beta(x)}{\varepsilon}\rho_1(x,t) \tag{3.33a}$$

$$\frac{\partial \rho_1(x,t)}{dt} = -\frac{\partial V_1(x)\rho_1(x,t)}{\partial x} + \frac{1}{2N}\frac{\partial^2 D_1(x)\rho_1(x,t)}{\partial x^2} + \frac{\alpha(x)}{\varepsilon}\rho_0(x,t) - \frac{\beta(x)}{\varepsilon}\rho_1(x,t), \tag{3.33b}$$

where

$$V_s(x) = s\sigma_1 + \sigma_0 - \gamma_p x, \quad D_s(x) = s\sigma_1 + \sigma_0 + \gamma_p x.$$

We have also performed the rescalings $\alpha \to \alpha/\varepsilon$ and $\beta \to \beta/\varepsilon$ such that the transitions rates are $O(1)$ with respect to γ_p. Equations (3.33a) and (3.33b) describe the CK equation for the piecewise Langevin equation

$$dX = V_s(X)dt + \frac{1}{\sqrt{N}}\sqrt{D_s(X)}dW(t), \tag{3.34}$$

which reduces to a piecewise deterministic ODE in the limit $\Omega \to 0$.

Another important limit is when the promoter switching rates are much faster than the other transition rates so that $0 < \varepsilon \ll 1$. Applying the adiabatic limit $\varepsilon \to 0$ to (3.33a) and (3.33b) as described in section 2.6 shows that we recover the effective SDE

$$dX = V(X)dt + \frac{1}{\sqrt{N}}\sqrt{D(X)}dW(t), \tag{3.35}$$

with

$$V(x) = \sum_{s=0,1} V_s(x)\rho_s^*(x), \quad D = \sum_{s=0,1} D_s(x)\rho_s^*(x), \tag{3.36}$$

where ρ_s^* is the steady-state distribution of the two-state Markov process (3.31):

$$\rho_0^*(x) = \frac{\beta(x)}{\alpha(x)+\beta(x)}, \quad \rho_1^*(x) = 1 - \rho_0^*(x) = \frac{\alpha(x)}{\alpha(x)+\beta(x)}. \tag{3.37}$$

We thus recover equations (3.27) and (3.28) for $x_1 = F(x)/\gamma$ and $x_2 = x$.

3.5. Mutual repressor model of a bistable genetic switch

The idea of a genetic switch was first proposed over 40 years ago by Jacob and Monod [140], 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. Significant insight into genetic switches such as the *lac operon* 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 [106]. The flipping of the switch can thus be observed by measuring the fluorescent level of the cells. The synthetic gene circuit consists of two repressor proteins whose transcription is mutually regulated, see Fig. 10. That is, the protein product of one gene binds to the promoter of the other gene and represses its output - a so-called mutual repressor model. For simplicity, suppose that the dynamics of transcription and translation are ignored so that only the mutual effects of the proteins on protein production are taken

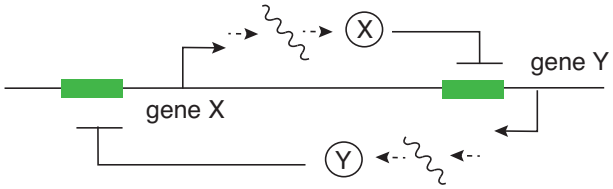


Figure 10. Mutual repressor model of a genetic switch. A gene X expresses a protein X that represses the transcription of gene Y and the protein Y represses the transcription of gene X.

into account. Denoting the concentrations of the proteins by $x(t), y(t)$, the resulting kinetic equations are

$$\frac{dx}{dt} = -\gamma x + \frac{r}{1+Ky^n}, \quad \frac{dy}{dt} = -\gamma y + \frac{r}{1+Kx^n}. \tag{3.38}$$

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 auto regulation, negative feedback is modeled in terms of a Hill function with Hill coefficient n. It is convenient to rewrite the equations in non-dimensional form by measuring x and y in units of $K^{-1/n}$ and time in units of γ^{-1}:

$$\frac{du}{dt} = -u + \frac{\alpha}{1+v^n}, \quad \frac{dv}{dt} = -v + \frac{\alpha}{1+u^n}, \tag{3.39}$$

with $\alpha = rK^{1/n}/\gamma$. In order to determine whether or not the mutual repressor model can act as a bistable switch, it is necessary to investigate the existence and stability of fixed points. 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 $u = \alpha/(1+u^2) = v$. 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.

So far we have assumed that the number of proteins is sufficiently large so that deterministic mass action kinetics can be used. In the case of relatively low copy numbers it is necessary to construct a chemical master equation along similar lines to the autoregulatory model. For the sake of illustration, we will consider 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 [151, 191]. 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 non-dimensional 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

$$\frac{d}{dt}P_j(n,m,t) = \sum_{k=0,1,2}\sum_{n',m'}\left[\delta_{n,n'}\delta_{m,m'}A_{jk} + \delta_{j,k}W^j_{nm,n'm'}\right]P_k(n',m',t), \tag{3.40}$$

where

$$\mathbf{A} = \kappa\begin{pmatrix} -n(n-1)-m(m-1) & \beta & \beta \\ n(n-1) & -\beta & 0 \\ m(m-1) & 0 & -\beta \end{pmatrix}, \tag{3.41}$$

and

$$\begin{aligned}&\sum_{n',m'} W^0_{nm,n'm'}P_0(n',m',t)\\ &= \gamma[(n+1)P_0(n+1,m,t) + (m+1)P_0(n,m+1,t) - (n+m)P_0(n,m,t)]\\ &\quad + \alpha(P_0(n-1,m,t) + P_0(n,m-1,t) - 2P_0(n,m,t))\end{aligned} \tag{3.42a}$$

$$\begin{aligned}&\sum_{n',m'} W^1_{nm,n'm'}P_1(n',m',t)\\ &= \gamma[(n+1)P_1(n+1,m,t) + (m+1)P_1(n,m+1,t) - (n+m)P_1(n,m,t)]\\ &\quad + \alpha(P_1(n-1,m,t) - P_1(n,m,t))\end{aligned} \tag{3.42b}$$

$$\begin{aligned}&\sum_{n',m'} W^2_{nm,n'm'}P_1(n',m',t)\\ &= \gamma[(n+1)P_2(n+1,m,t) + (m+1)P_2(n,m+1,t) - (n+m)P_2(n,m,t)]\\ &\quad + \alpha(P_2(n,m-1,t) - P_2(n,m,t))\end{aligned} \tag{3.42c}$$

Kepler and Elston [151] 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 quasi-steady-state (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 $\rho_j(x,y,t)$ takes the form

$$\frac{\partial}{\partial t}\rho_j(x,y,t) = \sum_{k=0,1,2}\left[\frac{1}{\varepsilon}A_{jk} + N\delta_{j,k}\mathbb{W}^j\right]\rho_k(x,y,t), \tag{3.43}$$

where $\varepsilon = \gamma^3/\kappa\alpha^2$ and $b = \beta\gamma^2/\alpha^2$ are dimensionless parameters,

$$\mathbf{A} = \begin{pmatrix} -x(x-1/N) - y(y-1/N) & b & b \\ x(x-1/N) & -b & 0 \\ y(y-1/N) & 0 & -b \end{pmatrix}, \tag{3.44}$$

and $\mathbb{W}^j$ are differential shift operators

$$\mathbb{W}^0 = \left(\mathrm{e}^{\partial_x/N} - 1\right)x + \left(\mathrm{e}^{\partial_y/N} - 1\right)y + \left(\mathrm{e}^{-\partial_x/N} + \mathrm{e}^{-\partial_y/N} - 2\right) \tag{3.45a}$$

$$\mathbb{W}^1 = \left(\mathrm{e}^{\partial_x/N} - 1\right)x + \left(\mathrm{e}^{\partial_y/N} - 1\right)y + \left(\mathrm{e}^{-\partial_x/N} - 1\right) \tag{3.45b}$$

$$\mathbb{W}^2 = \left(\mathrm{e}^{\partial_x/N} - 1\right)x + \left(\mathrm{e}^{\partial_y/N} - 1\right)y + \left(\mathrm{e}^{-\partial_y/N} - 1\right). \tag{3.45c}$$

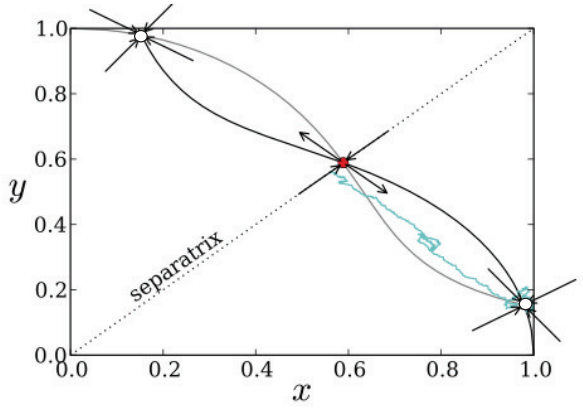


Figure 11. Phase-plane dynamics of mutual repressor model analyzed by Kepler and Elston [151] and Newby [191] with $b = 0.15$. The black curve shows the y-nullcline and the grey curve shows the x-nullcline. The open circles show the stable fixed points, the filled circle shows the unstable saddle. The irregular curve shows a stochastic trajectory leaving the lower basin of attraction to reach the separatrix.

The 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

$$\frac{dx}{dt} = f(x,y), \quad \frac{dy}{dt} = f(y,x), \quad \text{with } f(x,y) = \frac{1}{1+\frac{y^2}{b+x^2}} - x. \tag{3.46}$$

One finds that the deterministic system is bistable for $0 < b < b_c = 4/9$, see Fig. 11. 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 [151, 191]

$$\frac{\partial \rho_j}{\partial t} = -\frac{\partial F_j(x)\rho_j}{\partial x} - \frac{\partial G_j(y)\rho_j}{\partial y} + \frac{1}{\varepsilon}\sum_{k=0,1,2} A_{jk}(x,y)\rho_k \tag{3.47}$$

with

$$\begin{aligned} F_0(x) &= 1-x, \quad F_1(x) = 1-x, \quad F_2(x) = -x \\ G_0(y) &= 1-y, \quad G_1(y) = -y, \quad G_2(y) = 1-y, \end{aligned}$$

and

$$\mathbf{A} = \begin{pmatrix} -x^2 - y^2 & b & b \\ x^2 & -b & 0 \\ y^2 & 0 & -b \end{pmatrix}. \tag{3.48}$$

In a similar fashion to equations (3.33a) and (3.33b), the CK equation (3.47) 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. The analysis of metastability in the mutual repressor model can be found in [151, 191].

3.6. Spatial aspects of gene regulation

So far we have assumed that the various molecular players are well mixed within the cytoplasm of a bacterium or the nucleus of a eukaryotic cell, so that spatial effects can be ignored. However, there is at least one major step in gene regulation where spatial affects cannot be ignored. This concerns the mechanism whereby a protein transcription factor searches for a specific promoter binding site (specific target sequence of base pairs) on a long DNA molecule.† One of the striking features of these protein-DNA interactions is that experimentally observed binding rates ($k = 10^{10} M^{-1} s^{-1}$ [215]) are several orders of magnitude higher than predicted from the classical Smoluchowski theory of diffusion–limited reactions [149]. A possible explanation for this discrepancy was originally proposed by Berg, Winter and von Hippel (BHW) in their model of facilitated diffusion [33, 246]. The main 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. 12. 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.

A rough estimate of the effective reaction rate of facilitated diffusion in the BHW model can be obtained as follows [180]. 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 of switching such that during 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 with $i = 1, \ldots, R$. It follows that the total search time is $T = \sum_{i=1}^{R} (T_{3,i} + T_{1,i})$, and the mean search time is $\tau = r(\tau_3 + \tau_1)$, where 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

† In the case of eukaryotic cells one also has to consider the time taken for a newly synthesized transcription factor to be transported back into the nucleus. This process is thought to be an important factor in the clock gene underlying circadian rhythms [116].

is the $1D$ diffusion coefficient, and we have [180]

$$\tau = \frac{N}{n}(\tau_1 + \tau_3). \tag{3.49}$$

Since τ_3 depends primarily on the cellular environment, it is unlikely to vary significantly between different protein molecules. This suggests minimizing the mean search time with respect to τ_1 while keeping τ_3 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 $n/2$. Additional insights into facilitated diffusion may be obtained by using the Smoluchowski formula [149] for the rate at which a diffusing protein can find any one of N binding sites of size b, namely, $\tau_3^{-1} = 4\pi D_3 N b[DNA]$, where $[DNA]$ is the concentration of DNA and the 3D geometry of DNA is ignored. Using this to eliminate N shows that the effective reaction rate of facilitated diffusion is [180]

$$k \equiv \frac{1}{\tau[DNA]} = 4\pi D_3 \left(\frac{\tau_3}{\tau_1 + \tau_3}\right) nb \tag{3.50}$$

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* [133]. 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 non-specific sites that are far from the target. Note that typical experimental values are $D_3 = 10\mu m^2 s^{-1}$, $b = 0.34nm$, $n = 200$, and one has to convert k into units of inverse molar per sec. Improvements in these estimates have been obtained by Coppey et al. [65], who take into account the fact that diffusion occurs in a bounded domain.

There is some experimental support for the BHW model of facilitated diffusion. In particular, recent advances in single–molecule spectroscopy have allowed one to observe the sliding of fluorescently labeled proteins along DNA chains with high precision, although most of these studies have been performed *in vitro*. However,

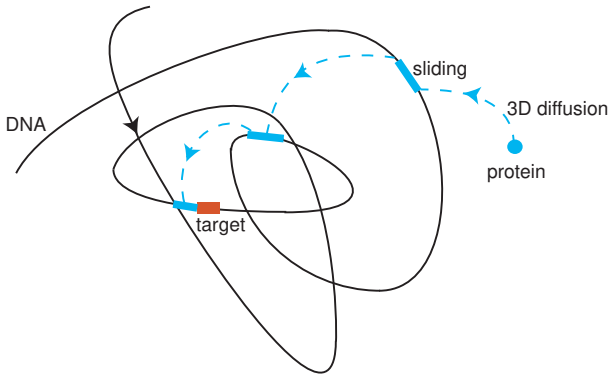


Figure 12. Basic mechanism of facilitated diffusion involving alternating phases of 3D diffusion and 1 D diffusion (sliding along the DNA). (b) 1D representation of facilitated diffusion.

a quantitative comparison of the BHW model with experimental data leads to a number of discrepancies. 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 $D_1 \ll D_3$ [255]. 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 [133]. 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 [155]. One proposed mechanism in bacteria is based on the observation that 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 [180, 207]. 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.

Another interesting issue is the so-called speed-stability paradox [180, 226], which arises from the observation that sequence-dependent protein-DNA interactions generate a rugged energy landscape during sliding motion of the protein [180]. 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 $\sigma \sim 5 k_B T$. One possible 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 [180]. 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 [226]). Other effects include changes in the conformational state of DNA and the possibility of correlated association/dissociation of the protein [133], and molecular crowding along DNA [166] or within the cytoplasm [139].

4. Transcriptional and translational bursting

We now turn to one important example of stochastic switching at the genetic level, namely, transcriptional and translational bursting. The latter is a consequence of the fact that lifetime of mRNA is usually much shorter than the expressed proteins [32, 59, 178], whereas the former tends to be associated with the switching between active and inactive states of a gene.

4.1. Translational bursting

Consider a single mRNA molecule with a degradation rate γ, which starts synthesizing a protein at time $t = 0$. Let $P_0(n, t)$ denote the probability that there are n proteins at time t and the mRNA has not decayed, and let $P_c(n, t)$ be the corresponding probability when the mRNA has decayed. If protein degradation is ignored, then the master equation is

$$\frac{dP_0(n,t)}{dt} = -\gamma P_0(n,t) + r[P_0(n-1,t) - P_0(n,t)] \tag{4.1a}$$

$$\frac{dP_c(n,t)}{dt} = \gamma P_0(n,t), \tag{4.1b}$$

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 (4.1b) 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 $\rho_0(n,t)$, integrate equation (4.1a) with respect to time using $P_0(n,0) = \delta_{n,0}$:

$$-\delta_{n,0} = -P(n) + \frac{r}{\gamma}[P(n-1) - P(n)].$$

Setting $n = 0$ gives $P(0) = \gamma/(r+\gamma)$, and for $n \geq 1$, we have the recurrence relation

$$P(n) = \frac{r}{r+\gamma}P(n-1) \to P(n) = \left(\frac{r}{r+\gamma}\right)^n \frac{\gamma}{r+\gamma}.$$

One important quantity is the mean number of proteins produced per mRNA, which is known as the burst size b. The latter can be calculated by introducing the generating function

$$G(z) \equiv \sum_{n\geq 0} z^n P(n) = \frac{\gamma}{r+\gamma}\frac{1}{1-zr/(r+\gamma)} = \frac{\gamma}{r+\gamma-zr},$$

such that

$$b = \sum_{n\geq 0} nP(n) = G'(1) = \left.\frac{r\gamma}{[r+\gamma-zr]^2}\right|_{z=1} = \frac{r}{\gamma}.$$

The notion of a translational burst refers to the observation that a single mRNA generates a burst of protein production before it decays, see Fig. 2(c).

Multiple mRNAs. 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 [204]. For example, if $m = 2, 3$ then

$$P_2(N) = \sum_{n=0}^{N} P(n)P(N-n), \quad P_3(N) = \sum_{n=0}^{N} P(n) \sum_{n'=0}^{N-n} P(n')P(N-n-n').$$

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 integral formulation allows us to use Laplace transforms and the convolution theorem. In particular,

$$\widehat{P}_m(s) \equiv \int_0^\infty P_m(n)\mathrm{e}^{-sn}dn = \left[\widehat{P}(s)\right]^m,$$

where

$$\widehat{P}(s) = \int_0^\infty \frac{b^n}{(1+b)^{n+1}}\mathrm{e}^{-sn}dn = \frac{1}{1+b}\int_0^\infty \left(\frac{b}{1+b}\mathrm{e}^{-s}\right)^n dn$$
$$= \frac{1}{1+b}\int_0^\infty \exp\left(n\ln\left(\frac{b}{1+b}\mathrm{e}^{-s}\right)\right)dn = -\frac{1}{1+b}\left[\ln\left(\frac{b}{1+b}\right) - s\right]^{-1}.$$

Hence,

$$\widehat{P}_m(s) = \left(\frac{-1}{1+b}\right)^m \left[\ln\left(\frac{b}{1+b}\right) - s\right]^{-m} = \left(\frac{1}{1+b}\right)^m [s-\beta]^{-m},$$

where $\beta = \ln(b/(1+b))$. Using the Laplace identities

$$\mathscr{L}(\mathrm{e}^{\beta n}) = \frac{1}{s-\beta},\ \left(-\frac{d}{d\beta}\right)^k \mathscr{L}(\mathrm{e}^{\beta n}) = (-1)^k \mathscr{L}(n^k \mathrm{e}^{\beta n}) = \frac{k!}{[s-\beta]^{k+1}},$$

establishes that

$$\widehat{P}_m(s) = \left(\frac{1}{1+b}\right)^m \frac{1}{(m-1)!}\mathscr{L}(n^{m-1}\mathrm{e}^{\beta n})$$

It follows that

$$P_m(n) = \left(\frac{1}{1+b}\right)^m \frac{1}{(m-1)!} n^{m-1}\mathrm{e}^{\beta n} = \left(\frac{b}{1+b}\right)^n \left(\frac{1}{1+b}\right)^m \frac{n^{m-1}}{\Gamma(m)}.$$

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:

$$P_m(n) \equiv F(n; m, b^{-1}) = \frac{n^{m-1}\mathrm{e}^{-n/b}}{b^m \Gamma(m)}. \tag{4.2}$$

It immediately follows from properties of the gamma distribution that, for a given number m of mRNA molecules, we have $\langle n\rangle = mb$ and $\mathrm{var}(n) = mb^2$. Thus, 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).

Chapman-Kolomorov equation for population bursting. An alternative approach to analyzing protein bursting is to start from the Chapman-Kolmogorov (CK) equation [101]

$$\frac{\partial \rho(x,t)}{\partial t} = \frac{\partial}{\partial x}[\gamma_0 x \rho(x)] + k\int_0^x w(x-x')\rho(x',t)dx', \tag{4.3}$$

where $\rho(x,t)$ is the probability density for x protein molecules (treating x as a continuous variable) at time t, and

$$w(x) = \frac{1}{b}\mathrm{e}^{-x/b} - \delta(x). \tag{4.4}$$

The first term on the right-hand side of the CK equation represents protein degradation, wheres the second term represents the production of proteins from exponentially distributed bursts. The gamma distribution (4.2) with $n \to x$ is obtained as the stationary solution of the CK equation, which can be established using Laplace transforms (see below). 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 [101, 171, 172, 257]:

$$\frac{\partial \rho(x,t)}{\partial t} = \frac{\partial}{\partial x}[\gamma_0 x \rho(x,t)] + k\int_0^x w(x-x')c(x')\rho(x',t)dx'. \tag{4.5}$$

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

$$\rho(x) = Ax^{m(1+\varepsilon)-1}\mathrm{e}^{-x/b}[1 + (x/k)^s]^{-m/s}.$$

Plots of the stationary density show that negative autoregulation sharpens the density (noise reduction), whereas positive feedback broadens the density and can lead to bistability, see Fig. 13.

Suppose that $c(x) = 1$ (no autoregulatory feedback). Laplace transforming with respect to the protein number x, $\mathscr{L}(\rho) = \int_0^\infty \mathrm{e}^{-zx}\rho(x,t)dx = \widehat{\rho}(z,t)$, and using the convolution theorem gives

$$\frac{\partial\widehat{\rho}(z,t)}{\partial t} = \gamma_0 z\mathscr{L}[x\rho](z,t) + k\widehat{w}(z)\widehat{\rho}(z,t).$$

Now

$$\int_0^\infty \mathrm{e}^{-zx}x\rho(x,t)dx = -\frac{\partial}{\partial z}\int_0^\infty \mathrm{e}^{-zx}\rho(x,t)dx = -\frac{\partial}{\partial z}\widehat{\rho}(z,t),$$

and

$$\widehat{w}(z) = \int_0^\infty \mathrm{e}^{-zx}\left[\frac{1}{b}\mathrm{e}^{-x/b} - \delta(x)\right]dx = \frac{1}{1+bz} - 1 = -\frac{bz}{1+bz}.$$

Therefore,

$$\frac{\partial\widehat{\rho}(z,t)}{\partial t} + \gamma_0 z\frac{\partial}{\partial z}\widehat{\rho}(z,t) = k\widehat{w}(z)\widehat{\rho}(z,t).$$

This is a quasilinear equation, which can be solved using the method of characteristics [219]. That is, the corresponding characteristic equations are

$$\frac{dz}{dt} = \gamma_0 z, \quad \frac{d\widehat{\rho}}{dt} = k\widehat{w}(z)\widehat{\rho}.$$

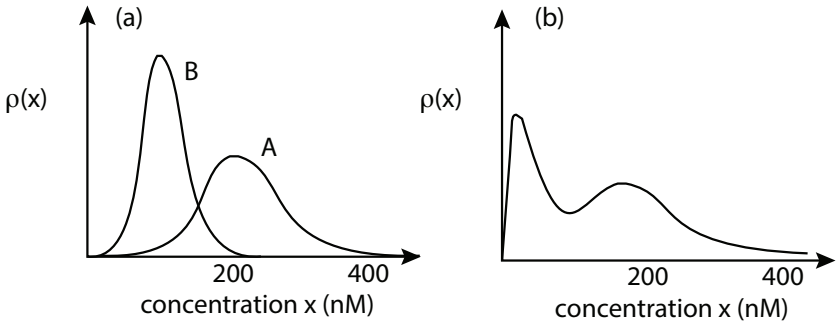


Figure 13. Sketch of typical steady-state probability densities obtained in a model of protein autoregulation [100]. Parameter values are $m = 10, b = 20, k = 70nM$. (a) Negative feedback. Curves A and B correspond to the two cases of no regulation ($c \equiv 1$) and regulation $c = k^s/(k^s + x^s) + \varepsilon$ with $s = 1$ and $\varepsilon = 0.05$. (b) Positive feedback with $\varepsilon = 0.2$ and $s = -4$.

Solving for z, we have $z(t) = z_0 \mathrm{e}^{\gamma_0 t}$ where z_0 parameterizes the initial data. Hence

$$\frac{d\widehat{\rho}}{dt} = -k \frac{b z_0 \mathrm{e}^{\gamma_0 t}}{1 + b z_0 \mathrm{e}^{\gamma_0 t}} \widehat{\rho},$$

which can be integrated to give

$$\widehat{\rho}(t) = F(z_0) \exp\left(-k \int_0^t \frac{b \mathrm{e}^{\gamma_0 t'}}{1 + b \mathrm{e}^{\gamma_0 t'}} dt' \right) = F(z_0) \exp\left(-\frac{k}{\gamma_0} \ln \left[\frac{1 + b z_0 \mathrm{e}^{\gamma_0 t}}{1 + b z_0} \right] \right)$$

$$= F(z_0) \left[\frac{1 + b z_0 \mathrm{e}^{\gamma_0 t}}{1 + b z_0} \right]^{-k/\gamma_0}$$

Setting $z_0 = z \mathrm{e}^{-\gamma_0 t}$ then yields

$$\widehat{\rho}(z,t) = F(z \mathrm{e}^{-\gamma_0 t}) \left[\frac{1 + bz}{1 + bz \mathrm{e}^{-\gamma_0 t}} \right]^{-k/\gamma_0}.$$

Since $\int_0^\infty \rho(x,t) dx = 1$ for all $t > 0$, we see that $\widehat{\rho}(0,t) = 1$ for all $t > 0$. In the limit $t \to \infty$,

$$\widehat{\rho}(z,t) \to \widehat{\rho}(z) = F(0) \left[1 + bz\right]^{-k/\gamma_0}.$$

Since $\int_0^\infty \rho(x) dx = 1$, we see that $\widehat{\rho}(0) = 1$, which implies that $F(0) = 1$. Using the inverse Laplace transform

$$\mathscr{L}^{-1}(b^{-1} + z)^{-m} = \mathrm{e}^{-x/b} \frac{x^{m-1}}{\Gamma(m)},$$

we obtain the result that the stationary density is

$$\rho(x) = \frac{1}{b^m \Gamma(m)} x^{m-1} \mathrm{e}^{-x/b}, \quad m = \frac{k}{\gamma_0}.$$

Now suppose that $c(x)$ is given by the Hill function

$$c(x) = \frac{k^s}{k^s + x^s} + \varepsilon.$$

Laplace transforming the steady-state equation gives

$$\frac{\partial \widehat{\rho}}{\partial z} = -\frac{k}{\gamma_0} \frac{1}{b^{-1} + z} \int_0^s \widehat{c}(z - s) \widehat{\rho}(s) ds,$$

where we have applied the convolution theorem. Multiplying both sides by $z + b^{-1}$ and taking the inverse Laplace transform gives

$$\frac{\partial x\rho}{\partial x} + \frac{x\rho}{b} = \frac{k}{\gamma_0} c(x) \rho(x).$$

This can be solved for $p(x)$:

$$\rho(x) = A x^{-1} \mathrm{e}^{-x/b} \exp\left(\frac{k}{\gamma_0} \int \frac{c(y)}{y} dy \right),$$

where A is a normalization factor. Performing the integral with respect to y for the given form of $c(y)$ finally yields the result

$$\rho(x) = A x^{m(1+\varepsilon)-1} \mathrm{e}^{-x/b} [1 + (x/k)^s]^{-m/s}.$$

Finally, note that a rigorous analysis of the existence of a stationary density and convergence to the stationary density has been carried out for a more general class of integral operator equations than (4.5) [171, 172, 257].

4.2. Transcriptional bursting and queuing theory

In section 3.2 we considered a simple two-state model of gene regulation. In this model, a gene switches between an on- and off-state and mRNA is only produced during the on-phase. However, in eukaryotic cells transcription appears to follow an ordered, multistep and cyclic process, which involves transitions between distinct chromatin states [36, 251]. Schwabe et al. [223] refer to this complex transcription mechanism as a *molecular ratchet model*, in which the time a single gene spends in the on- and off-phases, and the time between consecutive transcription initiation events are random variables with corresponding waiting time densities $f(t), g(t)$ and $h(t)$, see Fig. 14. The complexity of the underlying molecular mechanisms means that these waiting time densities are not necessarily exponential.

Suppose that an on-state persists for some random time $T_{\text{on}} = t$ and let $B(t)$ be the number of transcription initiation events (number of mRNA) that occur over this time interval. (For simplicity, we assume that each mRNA generates the same number of proteins so that burst size is measured in terms of the number of mRNA produced during an on-phase.) Let $F(b,t) = \mathbb{P}[B(t) \geq b | T_{\text{on}} = t]$. If the waiting time density between events is $h(\tau)$ and τ_j, $j = 1, \ldots, b$ denotes the time of the j-th event, then

$$F(b,t) = \int_0^t \ldots \int_0^{\tau_2} h(t-\tau_b) H(\tau_b - \tau_{b-1}) \cdots h(\tau_2 - \tau_1) d\tau_1 \cdots d\tau_b \equiv \int_0^t h(\tau)^{(b)} d\tau,$$

where $h(t)^{(b)}$ denotes the b-th convolution of $h(t)$. The probability that exactly b initiation events occur in the time interval t is

$$P(b,t) = F(b,t) - F(b+1,t) = \int_0^t \left[h(\tau)^{(b)} - h(\tau)^{(b+1)} \right] d\tau.$$

The mean burst-size given t is then

$$\langle b(t) \rangle = \sum_{b=0}^{\infty} b \int_0^t \left[h(\tau)^{(b)} - h(\tau)^{(b+1)} \right] d\tau.$$

Taking Laplace transforms of both sides and using the convolution theorem shows that

$$\langle \widehat{b}(s) \rangle = s^{-1}(1 - \widehat{h}(s)) \sum_{b=0}^{\infty} b \widehat{h}(s)^b = s^{-1} \widehat{h}(s)(1 - \widehat{h}(s)) \frac{d}{d\widehat{h}(s)} \sum_{b=0}^{\infty} \widehat{h}(s)^b$$

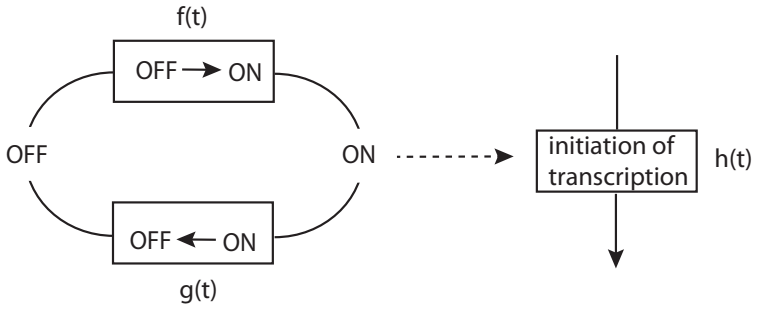


Figure 14. Schematic illustration of molecular ratchet model of transcription at the single gene level, where the waiting time distribution of the on-state, off-state, and transcription initiation are denoted by $f(t), g(t), h(t)$.

$$= s^{-1}\widehat{h}(s)(1-\widehat{h}(s))\frac{d}{d\widehat{h}(s)}\frac{1}{1-\widehat{h}(s)} = s^{-1}\frac{\widehat{h}(s)}{1-\widehat{h}(s)}.$$

Hence

$$\langle b(t)\rangle = \mathscr{L}^{-1}\left[\frac{1}{s}\frac{\widehat{h}(s)}{1-\widehat{h}(s)}\right](t). \tag{4.6}$$

A similar analysis establishes that

$$\langle b(t)^2\rangle = \mathscr{L}^{-1}\left[\frac{1}{s}\frac{\widehat{h}(s)(1+\widehat{h}(s))}{(1-\widehat{h}(s))^2}\right](t). \tag{4.7}$$

One now obtains the moments of the burst-size distribution by integrating with respect to the on times $T_{\rm on} = t$:

$$\langle b\rangle = \int_0^\infty \langle b(t)\rangle g(t)dt, \quad \langle b^2\rangle = \int_0^\infty b(t)^2 g(t)dt. \tag{4.8}$$

In the case of exponential waiting times $g(t) = k_g \mathrm{e}^{-k_g t}$ and $h(t) = k_h \mathrm{e}^{-k_h t}$, we have

$$\widehat{h}(s) = \frac{k_h}{k_h + s},$$

so that

$$\langle b(t)\rangle = \mathscr{L}^{-1}\left[\frac{k_h}{s^2}\right](t) = k_h t, \quad \langle b(t)^2\rangle = \mathscr{L}^{-1}\left[\frac{k_h}{s^2}\frac{2k_h+s}{s}\right](t) = k_h^2 t^2 + k_h t.$$

It follows that

$$\langle b\rangle = \frac{k_h}{k_g}, \quad \langle b^2\rangle = \langle b\rangle + 2\langle b\rangle^2, \tag{4.9}$$

and thus the Fano factor is

$$Q_b \equiv \frac{\mathrm{Var}[b]}{\langle b\rangle} = 1 + \langle b\rangle. \tag{4.10}$$

The corresponding distribution of burst sizes is a geometric distribution,

$$\eta_b \equiv \int_0^\infty P(b,t)g(t)dt = \eta(1-\eta)^b, \quad \eta = \frac{k_g}{k_h + k_g}.$$

Schwabe et al. [223] explore the effects of non-exponential on-state time distributions, and show that they can lead to a burst size distribution that is peaked (rather than exponentially decreasing) and this reduces transcriptional noise.

Now suppose that one combines the stochastic process that generates a distribution of burst sizes with the waiting time density $f(\tau)$ of being in the off-phase. The resulting system can be mapped into one of the classical models of queuing theory [156], see Fig. 15. Queuing theory concerns the mathematical analysis of waiting lines formed by customers randomly arriving at some service station, and staying in the system until they receive service from a group of servers. Different types of queuing process are defined in terms of (i) the stochastic process underlying the arrival of customers, (ii) the distribution of the number of customers (batches) in each arrival, (iii) the stochastic process underlying the departure of customers (service-time distribution), and (iv) the number of servers. The above model of transcriptional bursting can be mapped to a queuing process as follows: individual mRNAs are analogous to customers, transcriptional bursts correspond to customers

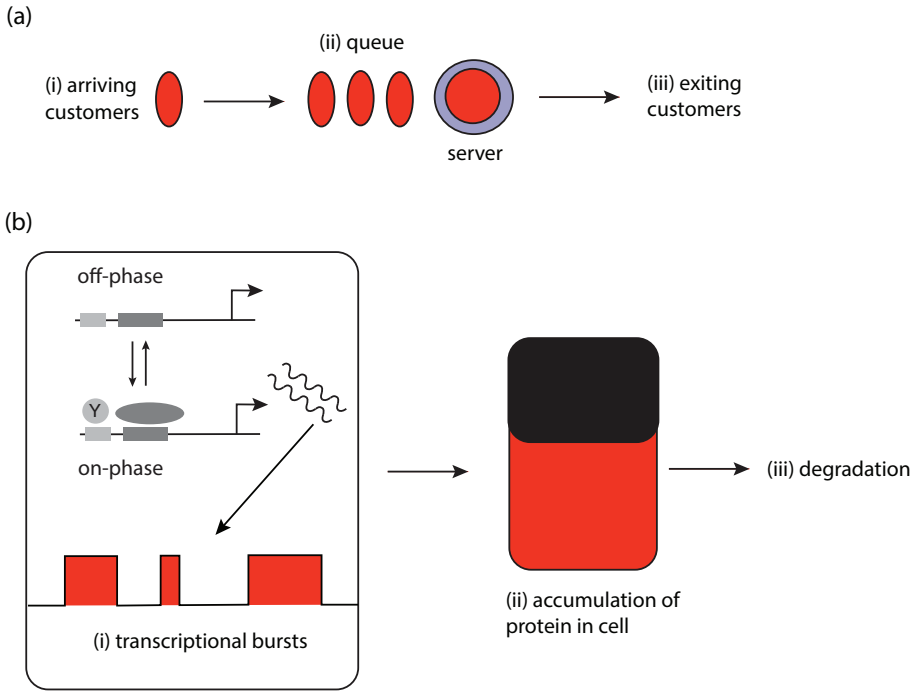


Figure 15. Diagram illustrating the mapping between queuing theory and transcriptional bursting. (a) Example of a single-server queue. (b) Transcriptional bursting. The stochastic switching of a gene between an on-phase and an off-phase generates a a sequence of transcriptional bursts that is analogous to the arrival of customers in the queuing model. This results in the accumulation of proteins within the cell, which is the analog of a queue. Degradation corresponds to exiting of customers after being serviced by an infinite number of servers.

arriving in batches, and the degradation of mRNAs is the analog of customers exiting the system after being serviced. Thus, the waiting-time density for mRNA degradation is the analog of the service-time distribution. Finally, since the mRNAs are degraded independently of each other, the effective number of servers in the corresponding queuing model is infinite, that is, the presence of other customers does not affect the service time of an individual customer.

The particular queuing model that maps to the model of transcriptional bursting is the $GI^X/M/\infty$ system, Here the symbol G denotes a general waiting time density for the arrival process, that is, the waiting time density $f(\tau)$ for the gene to switch from the off-phase to the on-phase, and I^X refers to the fact that customers (mRNAs) arrive in batches of independently distributed random sizes B (mRNA burst sizes). The symbol M stands for a Markovian or exponential service-time density (mRNA degradation-time density), and '∞' denotes infinite servers. Exploiting this mapping, exact results for the moments of the mRNA steady-state distribution can be obtained from the corresponding known expressions for the moments of the steady-state distribution of the number of current customers in the $GI^X/M/\infty$ model [156]. The latter were originally derived by Liu et al. [167], and we will follow their analysis in section 4.3. Here we simply quote the results for the mean and variance of the steady-state mRNA

copy number N:

$$\langle m\rangle = \frac{\lambda}{\mu}\langle b\rangle, \tag{4.11}$$

and

$$\mathrm{Var}[m] = \langle m\rangle - \langle m\rangle^2 + \frac{\mu}{\lambda}\langle m\rangle^2\left(\frac{\widehat{f}(\mu)}{1-\widehat{f}(\mu)} + \frac{\langle b^2\rangle - \langle b\rangle}{2\langle b\rangle^2}\right) \tag{4.12}$$

In the special case that the arrival times are also exponentially distributed, $\widehat{f}(s) = \lambda/(s+\lambda)$, and the variance simplifies to

$$\mathrm{Var}[m] = \frac{\lambda}{\mu}\langle b^2\rangle. \tag{4.13}$$

4.3. Moments of $GI^X/M/\infty$ queuing model

Let $F(t)$ and $H(t)$ be the probability distributions for the inter-arrival times and the service times, respectively. It is assumed that the mean arrival rate λ and the mean service rate μ are finite. Since the distribution of service times is taken to be exponential, we have $S(t) = 1 - \mathrm{e}^{-\mu t}$. The arrivals are taken to occur in batches of variable size X, where

$$\mathbb{P}[X = b] = \eta_b, \quad b = 1, 2, \ldots,$$

with finite mean and variance, and generating function

$$A(z) = \sum_{b=1}^{\infty} \eta_b z^b.$$

It follows that for any positive integer k, the k-th factorial moment of X is given by

$$A_k = \left.\frac{d^k A(z)}{dz^k}\right|_{z=1} = \sum_{b=k}^{\infty} b(b-1)\cdots(b-k+1)\eta_b. \tag{4.14}$$

Let T_n be the time of the n-th group arrival, with X_n denoting the size of the batch and S_{ni} the service time of the ith member of the batch. The number of busy servers at time is then

$$N(t) = \sum_{0\le T_n\le t} \chi(t - T_n, X_n), \tag{4.15}$$

where

$$\chi(t-T_n, X_n) = \sum_{i=1}^{X_n} I(t-T_n, S_{ni}), \quad I(t-T_n, S_{ni}) = \begin{cases} 1 & \text{if } t - T_n \le S_{ni} \\ 0 & \text{if } t - T_n > S_{ni} \end{cases} \tag{4.16}$$

In other words, we are keeping track of all batches that arrived prior to time t and the number of customers in each batch that haven't yet exited the system. Introduce the generating function

$$G(z,t) = \sum_{k=0}^{\infty} z^k \mathbb{P}[N(t) = k], \tag{4.17}$$

and the binomial moments

$$B_r(t) = \sum_{k=r}^{\infty} \frac{k!}{(k-r)!r!}\mathbb{P}[N(t) = k], \quad r = 1, 2, \ldots. \tag{4.18}$$

Suppose that the system is empty at time $t = 0$. We shall derive an integral equation for the generating function $G(z,t)$. Conditioning on the first arrival time by setting $T_1 = y$, it follows that $N(t) = \chi(t-y, X_1) + N^*(t-y)$ if $y \le t$ and zero otherwise, where $\chi(t-y, X_1)$ and $N^*(t-y)$ are independent of each other, and $N^*(t)$ has the same distribution as $N(t)$. Moreover,

$$\begin{aligned}\mathbb{P}[I(t-y, S_{1i}) = k] &= \mathbb{P}[t-y \le S_{1i}]\delta_{k,1} + \mathbb{P}[t-y > S_{1i}]\delta_{k,0} \\ &= [1 - H(t-y)]\delta_{k,1} + H(t-y)\delta_{k,0},\end{aligned}$$

so that

$$\sum_{k=0}^{\infty} z^k P[I(t-y, S_{1i}) = k] = z + (1-z)H(t-y).$$

Since $I(t-y, S_{1i})$ for $i = 1, 2, \ldots$ are independent and identically distributed, the theorem of total expectation implies that

$$\begin{aligned}\mathbb{E}[z^{\chi(t-y,X_1)}] &= \mathbb{E}[\mathbb{E}[z^{\chi(t-y,X_1)}|X_1]] = \sum_{b=1}^{\infty} \eta_b \mathbb{E}[z^{\sum_{i=1}^{b} I(t-y,S_{1i})}] \\ &= \sum_{b=1}^{\infty} \eta_b \prod_{i=1}^{b} \mathbb{E}[z^{I(t-y,S_{1i})}] = \sum_{b=1}^{\infty} \eta_b \prod_{i=1}^{b} \left(\sum_{k=0}^{\infty} z^k P[I(t-y, S_{1i}) = k] \right) \\ &= \sum_{b=1}^{\infty} \eta_b [z + (1-z)H(t-y)]^b = A[z + (1-z)H(t-y)].\end{aligned}$$

Another application of the theorem of total expectation gives

$$\begin{aligned}G(z,t) &= \mathbb{E}[z^{N(t)}] = \mathbb{E}[\mathbb{E}[z^{N(t)}|y]] \\ &= \int_t^{\infty} \mathbb{E}[z^0] dF(y) + \int_0^t \mathbb{E}[z^{\chi(t-y,X_1)}]\mathbb{E}[z^{N^*(t-y)}] dF(y) \\ &= 1 - F(t) + \int_0^t G(z, t-y) A[z + (1-z)H(t-y)] dF(y). \qquad (4.19)\end{aligned}$$

One can now obtain an iterative equation for the Binomial moments by differentiating equation (4.19) multiple times with respect to z and using

$$B_r(t) = \frac{1}{r!} \left. \frac{d^r G(z,t)}{dz^r} \right|_{z=1}.$$

Noting that

$$\left. \frac{d^r}{dz^r} A[z + (1-z)H(t-y)] \right|_{z=1} = [1 - H(t-y)]^r A_r, \quad r = 1, 2, \ldots$$

we obtain the integral equation

$$B_r(t) = \int_0^t B_r(t-y) dF(y) + \sum_{k=1}^{r} \frac{A_k}{k!} \int_0^t B_{r-k}(t-y)[1 - H(t-y)]^k dF(y). \qquad (4.20)$$

This can be written in the more compact form

$$B_r(t) = \int_0^t B_r(t-y) dF(y) + \int_0^t \mathscr{H}_r(t-y) dF(y), \qquad (4.21)$$

where

$$\mathscr{H}_r(t) = \sum_{k=1}^{r} \frac{A_k}{k!} B_{r-k}(t)[1 - H(t)]^k. \qquad (4.22)$$

Equation (4.21) is an example of a renewal-type equation, see below, which has the unique solution

$$B_r(t) = \mathscr{H}_r(t) + \int_0^t \mathscr{H}_r(t-y)dm(y), \tag{4.23}$$

where

$$m(y) = \sum_{n=1}^{\infty} F_n(y), \quad F_n(y) = \mathbb{P}[T_n \leq y] \tag{4.24}$$

is the so-called renewal function.

The steady-state Binomial moments can now be obtained by taking the limit $t \to \infty$ in equation (4.23) and using the Key Renewal Theorem (see section 4.4)

$$B_r = \lambda \sum_{k=1}^{r} \frac{A_k}{k!} \int_0^{\infty} B_{r-k}(t)[1-H(t)]^k dt, \tag{4.25}$$

where $\lambda^{-1} = \mathbb{E}[T_1]$. In the case of an exponential service-time distribution, $1 - H(t) = \mathrm{e}^{-\mu t}$ so that

$$B_r = \lambda \sum_{k=1}^{r} \frac{A_k}{k!} \widehat{B}_{r-k}(k\mu), \tag{4.26}$$

where $\widehat{B}_r(s)$ is the Laplace transform of $B_r(t)$. An iterative equation for the Laplace transforms can be obtained by Laplace transforming equations (4.21) and (4.22), and using the convolution theorem:

$$\widehat{B}_r(s) = \widehat{B}_r(s)\widehat{f}(s) + \widehat{\mathscr{H}_r}(s)\widehat{f}(s),$$

which, on rearranging, yields

$$\widehat{B}_r(s) = \frac{\widehat{f}(s)}{1-\widehat{f}(s)}\widehat{\mathscr{H}_r}(s), \quad \widehat{\mathscr{H}_r}(s) = \sum_{k=1}^{r} \frac{A_k}{k!}\widehat{B}_{r-k}(s+k\mu), \tag{4.27}$$

Note that $B_0(t) = 1$ so $\widehat{B}_0(s) = 1/s$ and, hence,

$$\widehat{B}_1(s) = \frac{\widehat{f}(s)}{1-\widehat{f}(s)} \frac{A_1}{s+\mu}.$$

Equations (4.26) and (4.27) determine completely the steady-state Binomial moments. In particular,

$$B_1 \equiv \langle N \rangle = \frac{\lambda}{\mu} A_1, \tag{4.28}$$

and

$$\begin{aligned} B_2 &\equiv \frac{1}{2}\left(\langle N^2 \rangle - \langle N \rangle\right) = \lambda \left(A_1 \widehat{B}_1(\mu) + \frac{A_2}{4\mu} \right) \\ &= \frac{\lambda A_1}{2\mu} \left(\frac{\widehat{f}(\mu)}{1-\widehat{f}(\mu)} A_1 + \frac{A_2}{2A_1} \right). \end{aligned} \tag{4.29}$$

Combining these results with equation (4.14) for the specific bursting model yields the expressions (4.11) and (4.12) for the mean and variance of the mRNA copy number. For higher order moments see [156].

4.4. The renewal equation.

A *renewal process* $N(t)$ is defined according to

$$N(t) = \max\{n : T_n \leq t\}, T_0 = 0, \quad T_n = X_1 + \ldots + X_n \tag{4.30}$$

for $n \geq 1$, where $\{X_j\}$ is a sequence of independent identically distributed non-negative random variables. It immediately follows that

$$N(t) \geq n \text{ if and only if } T_n \leq t.$$

Typically, $N(t)$ represents the number of occurrences of some event in the time interval $[0, t]$. In the case of queuing theory, this could be the number of customer batches that have arrived for service. Thus T_n is the n-th arrival time and X_n is the n-th inter-arrival time. In neurobiology, $N(t)$ could represent the number of times a neuron has fired an action potential in a time interval $[0, t]$, T_n is the n-th firing time, and X_n is the n-th inter-spike interval. In the following we will assume that X_j is strictly positive, which ensures that $\mathbb{P}[N(t) < \infty] = 1$ for all t.

Introduce the inter-arrival distribution $F(t) = \mathbb{P}[X_1 \leq t]$ and let $F_k(t) = \mathbb{P}[T_k \leq t]$ be the distribution function of the k-th arrival time T_k. Clearly $F_1 = F$. From the identity $T_{k+1} = T_k + X_{k+1}$, one has the iterative equation

$$F_{k+1}(t) = \int_0^t F_k(t - y)dF(y), \quad k \geq 1. \tag{4.31}$$

Moreover,

$$\begin{aligned}\mathbb{P}[N(t) = k] &= \mathbb{P}[N(t) \geq k] - \mathbb{P}[N(t) \geq k + 1] = \mathbb{P}[T_k \leq t] - \mathbb{P}[T_{k+1} \leq t] \\ &= F_k(t) - F_{k+1}(t).\end{aligned}$$

An important object in renewal theory is the *renewal function* m

$$m(t) \equiv \mathbb{E}[N(t)] = \sum_{k=1}^{\infty} F_k(t). \tag{4.32}$$

The last expression results by expressing $N(t)$ as a sum of Heaviside functions

$$N(t) = \sum_{k=1}^{\infty} H(t - T_k),$$

so that

$$m(t) = \mathbb{E}\left[\sum_{k=1}^{\infty} H(t - T_k)\right] = \sum_{k=1}^{\infty} \mathbb{E}[H(t - T_k)] = \sum_{k=1}^{\infty} F_k(t).$$

A fundamental result of renewal theory is that m satisfies the renewal equation

$$m(t) = F(t) + \int_0^t m(t - x)dF(x). \tag{4.33}$$

This can be established using the theorem of total expectation:

$$\begin{aligned}m(t) &= \mathbb{E}[N(t)] = \mathbb{E}[\mathbb{E}[N(t) \mid X_1]] \\ &= \int_0^{\infty} \mathbb{E}[N(t) \mid X_1 = x]dF(x) \\ &= \int_0^t \mathbb{E}[N(t) \mid X_1 = x]dF(x) + \int_t^{\infty} \mathbb{E}[N(t) \mid X_1 = x]dF(x) \\ &= \int_0^t (1 + \mathbb{E}[N(t - x)])dF(x) = \int_0^t (1 + m(t - x))dF(x).\end{aligned}$$

We have used the fact that $\mathbb{E}[N(t) \mid X_1 = x] = 0$ if $t < x$. It can also be shown that $m(t) = \sum_{k=1}^{\infty} F_k(t)$ is the unique solution to the renewal equation (4.33) that is bounded on finite intervals. An important generalization of (4.33) is the renewal-type equation

$$C(t) = \mathscr{H}(t) + \int_0^t C(t-x)dF(x), \tag{4.34}$$

where $\mathscr{H}$ is a uniformly bounded function. The solution of the latter equation is specified by the following theorem (see [121]):

Theorem 1 *The function*

$$C(t) = \mathscr{H}(t) + \int_0^t \mathscr{H}(t-y)dm(y) \tag{4.35}$$

is a solution of the renewal-type equation (4.34). Moreover, if $\mathscr{H}$ is bounded on finite intervals then so is C, and the solution is unique.

Proof If $h : [0,\infty) \to \mathbb{R}$, then define the functions $h * m$ and $h * F$ by

$$(h*m)(t) = \int_0^t h(t-x)dm(x), \quad (h*F)(t) = \int_0^t h(t-x)dF(x),$$

assuming the integrals exist. From the definitions, we have

$$m * F = F * m, \quad (h*m)*F = h*(m*F).$$

Moreover, the iterative equation (4.31), the renewal equation (4.33) and the solution (4.35) can be written in the compact forms

$$F_{k+1} = F_k * F = F * F_k, \quad m = F + F*m, \quad C = \mathscr{H} + \mathscr{H}*m.$$

Convolving C with respect to F gives

$$C*F = \mathscr{H}*F + \mathscr{H}*m*F = \mathscr{H}*F + \mathscr{H}*(m-F) = \mathscr{H}*m = C - \mathscr{H},$$

which establishes that the function C is a solution of the renewal-type equation (4.34). It remains to prove that if $\mathscr{H}$ is bounded on finite intervals then the solution is unique.

If $\mathscr{H}$ is bounded on finite intervals $[0,T]$, then from equation (4.35)

$$\begin{aligned}\sup_{0\le t\le T}|C(t)| &\le \sup_{0\le t\le T}|H(t)| + \sup_{0\le t\le T}\left|\int_0^t \mathscr{H}(t-y)dm(y)\right| \\ &\le (1+m(T)) \sup_{0\le t\le T}|H(t)| < \infty.\end{aligned}$$

Finiteness of the renewal function means that the solution C is also bounded on finite intervals. Now suppose that $\widehat{C}$ is another bounded solution of (4.34) and set $\Delta(t) = C(t) - \widehat{C}(t)$. Equations (4.31) and (4.34) imply that

$$\Delta = \Delta * F = \Delta * F * F = \Delta * F_2 = \Delta * F * F_2 = \Delta * F_3 \cdots,$$

that is, $\Delta = \Delta * F_k, \quad k \ge 1$. Therefore,

$$|\Delta(t)| \le F_k(t) \sup_{0\le u\le t}|\Delta(u)|, \quad k \ge 0.$$

Finally, taking the limit $k \to \infty$ and noting that

$$F_k(t) = \mathbb{P}[N(t) \ge k] \to 0 \text{ as } k \to \infty,$$

we conclude that $|\Delta(t)| = 0$ for all t.

We end our brief discussion of renewal theory by stating several limit theorems. Set $\mu = \mathbb{E}[X_1] < \infty$.

Theorem 2 (Elementary Renewal Theorem) *This states that*

$$\frac{N(t)}{t} \overset{\text{a.s.}}{\to} \frac{1}{\mu} \text{ as } t \to \infty.$$

Theorem 3 (Key Renewal Theorem) *Define a random variable Y and its distribution F_Y to be arithmetic with span $\lambda, \lambda > 0$, if Y takes values in the set $\{m\lambda, m \in \mathbf{Z}\}$ with probability one, and λ is the maximal value with this property. Suppose that the first inter-arrival time X_1 is not arithmetic. If $g : [0, \infty) \to [0, \infty)$ is such that g is monotone decreasing and $\int_0^\infty g(t)dt < \infty$, then*

$$\int_0^t g(t-x)dm(x) \to \frac{1}{\mu}\int_0^\infty g(x)dx \text{ as } t \to \infty.$$

5. Metastability in a genetic switch and the WKB approximation

In section 3 we described how a genetic regulatory network with nonlinear feedback can form a genetic switch, consisting of two stable states representing high and low levels of protein production, respectively, see Fig. 9. In the absence of noise, the particular state of the switch is determined by initial conditions. On the other hand, when weak molecular noise is included, fluctuations can induce transitions between the metastable states (the states that are stable fixed points in the deterministic limit). Since the noise tends to be weak, transitions are rare events involving large fluctuations that are in the tails of the underlying probability density function. This means that estimates of mean transition times and other statistical quantities can be sensitive to any approximations, including the Gaussian approximation based on a system-size expansion or adiabatic approximation (see section 2), and can sometimes lead to exponentially large errors. Moreover, one finds that ignoring the stochastic aspects of the promoter-based activation/inactivation of the gene is a major source of error [15, 151, 191, 192, 195].

5.1. Metastability and large deviations in SDEs

The analysis of metastability has a long history [125], particularly within the context of SDEs with weak noise. The underlying idea is that the mean rate to transition from a metastable state in the weak noise limit can be identified with the principal eigenvalue of the generator of the underlying stochastic process, which is a second-order differential operator in the case of a Fokker-Planck equation. Calculating the eigenvalue typically involves obtaining a Wentzel-Kramers–Brillouin (WKB) approximation of a quasistationary solution and then using singular perturbation theory to match the solution to an absorbing boundary condition [124, 174, 176, 187, 222]. The latter is defined on the boundary that marks the region beyond which the system rapidly relaxes to another metastable state, becomes extinct, or escapes to infinity. In one-dimensional systems ($d = 1$), this boundary is simply an unstable fixed point, whereas in higher-dimensions ($d > 1$) it is generically a $(d-1)$-submanifold. For example, if $d = 2$ then the boundary is a closed curve, see Fig. 16. In the weak noise limit, the most likely paths of escape through an absorbing boundary are rare events, occurring in the tails of the associated functional probability distribution. From a mathematical perspective, the rigorous analysis of the tails of a distribution is known as large deviation theory [72, 96, 98, 238, 239], which provides a rigorous probabilistic framework for interpreting the WKB solution in terms of optimal fluctuational paths.

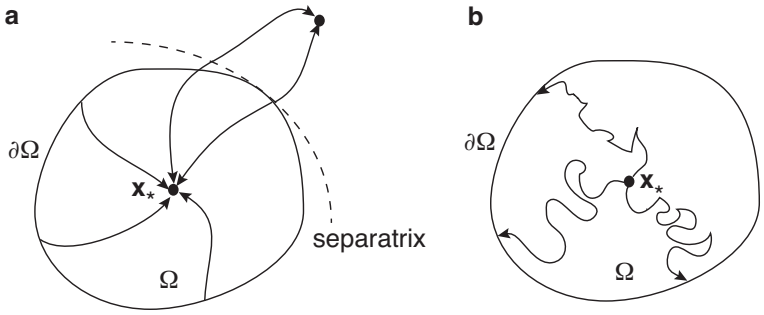


Figure 16. (a) Deterministic trajectories of a multistable dynamical system. The subset Ω is contained within the basin of attraction of a fixed point $\mathbf{x}_s$. The boundary of the basin of attraction consists of separatrices, which are also solution trajectories. Paths that start in a neighborhood 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 fluctuational path hits the boundary $\partial\Omega$.

5.1.1. Large deviation principles. Let us begin with a simple motivating example of large deviations [238, 239]. Consider the sum (sample mean)

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

where $X_1, X_2, \ldots$ is an i.i.d. sequence of random variables generating from the probability density $P(X)$ with $X \in \mathbb{R}$. The joint probability density function (pdf) is

$$P(X_1, \ldots, X_n) = \prod_{j=1}^{n} P(X_j).$$

The corresponding pdf of the sum S_n is obtained as follows:

$$P_n(s) = \mathbb{P}[S_n = s] = \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n \delta(\sum_{i=1}^{n} x_i - ns) P(x_1, \ldots, x_n). \tag{5.1}$$

Introducing the Fourier representation of the Dirac delta function, $\delta(x) = \int_{-\infty}^{\infty} \mathrm{e}^{ikx} dk/2\pi$, and using the product decomposition of the joint pdf,

$$\begin{aligned} P_n(s) &= \int_{-\infty}^{\infty} \frac{dk}{2\pi} \mathrm{e}^{-ikns} \prod_{j=1}^{n} \left(\int_{-\infty}^{\infty} \mathrm{e}^{ikx_j} P(x_j) dx_j \right) \\ &= \int_{-\infty}^{\infty} \widetilde{P}(k)^n \mathrm{e}^{-ikns} \frac{dk}{2\pi} \end{aligned}$$

In the case of a Gaussian pdf with mean μ and variance σ^2,

$$\widetilde{P}(k) = \mathrm{e}^{ik\mu} \mathrm{e}^{-\sigma^2 k^2/2},$$

and in the case of an exponential pdf on $[0, \infty)$ with mean μ,

$$\widetilde{P}(k) = \frac{\mu}{1 - i\mu k}.$$

Substituting into the integral expression for $P_n(s)$ then gives

$$P_n(s) = \int_{-\infty}^{\infty} \mathrm{e}^{ikn(\mu-s)} \mathrm{e}^{-n\sigma^2 k^2/2} \frac{dk}{2\pi} = \frac{1}{\sqrt{2\pi n\sigma^2}} \mathrm{e}^{-n(\mu-s)^2/2\sigma^2}$$

for the Gaussian, and

$$\begin{aligned} P_n(s) &= \int_{-\infty}^{\infty} \mathrm{e}^{-ikns} \left(\frac{\mu}{1-i\mu k}\right)^n \frac{dk}{2\pi} \\ &= \frac{\mu^n}{\mu} \oint \frac{\mathrm{e}^{-izns/\mu}}{[1-iz]^n} \frac{dz}{2\pi} = \frac{\mu^{n-1}}{(n-1)!} \left(\frac{-ns}{\mu}\right)^{n-1} \mathrm{e}^{-ns/\mu} \\ &= \frac{n^{n-1}}{(n-1)!} \left(\frac{s}{\mu}\right)^n \mathrm{e}^{-ns/\mu}. \end{aligned}$$

We have closed the contour in the lower-half complex z-plane and used the following residue theorem for an analytic function $f(z)$:

$$f^{(n-1)}(z_0) = \frac{1}{(n-1)!} \oint \frac{f(z)}{(z-z_0)^n} \frac{dz}{2\pi i}.$$

We now note that in both cases the leading order behavior in n for large n can be expressed as

$$P_n(s) \approx \mathrm{e}^{-nI(s)}, \tag{5.2}$$

where

$$I(s) = \frac{(s-\mu)^2}{2\sigma^2}, \quad s \in \mathbb{R} \tag{5.3}$$

for the Gaussian pdf and

$$\begin{aligned} I(s) &= -\frac{1}{n} \{(n-1)\ln n - \ln[(n-1)!] - n\ln(s/\mu) - ns/\mu\} \\ &= \frac{s}{\mu} - 1 - \ln(s/\mu), \quad s \geq 0 \end{aligned} \tag{5.4}$$

for the exponential pdf (after applying Strirling's formula). In both cases $I(s) \geq 0$ and $I(s) = 0$ only when $s = \mu = \mathbb{E}[X]$. Since the pdf of S_n is normalized, it becomes more and more concentrated around $s = \mu$ as $n \to \infty$, that is, $P_n(s) \to \delta(s-\mu)$ in the large-n limit. The leading order exponential form $\mathrm{e}^{-nI(s)}$ found for the Gaussian and exponential pdfs is the fundamental property of large deviation theory, which is known as the large deviation principle or LDP. (An LDP should be distinguished from the Central Limit Theorem, which is concerned with the asymptotic properties of $\sqrt{n}(S_n - \mu)$.) It arises in a much wider range of stochastic process then sums of i.i.d. random variables. Following Touchette [239], we will avoid the technical aspects of the rigorous formulation of the large deviation principle. For our purposes, a random variable S_n or its pdf P_n satisfies a *large deviation principle* (LDP) if the following limit exists:

$$\lim_{n\to\infty} -\frac{1}{n} \ln[P_n(s)] = I(s), \tag{5.5}$$

with $I(s)$ the so-called rate function. A more rigorous definition involves probability measures on sets rather than in terms of pdfs, and gives lower and upper bounds on these probabilities rather than a simple limit [238,243]. One of the main goals of large deviation theory is to identify stochastic processes that satisfy an LDP, and to develop analytical (and numerical) methods for determining the associated rate function.

Large deviation principle for an SDE. Consider as an example the one-dimensional SDE with additive noise

$$dX(t) = F(X(t) + \sqrt{\varepsilon}dW(t), \quad X(0) = 0, \tag{5.6}$$

with $W(t)$ a Wiener process. We are interested in the pdf of random paths $\{X(t), t \in [-\tau, 0]\}$ of duration τ in the limit where the noise strength ε vanishes. Denote the pdf by the functional $P[x]$. The occurrence of an LDP in the low noise limit is a reflection of the fact that random paths of the SDE should converge in probability to the deterministic path given by the solution to the ODE

$$\dot{x}(t) = F(x(t)), \quad x(-\tau) = 0.$$

The path fluctuations away from the deterministic path are characterized by a functional LDP of the Freidlin-Wentzel form [98]

$$P[x] \approx \mathrm{e}^{-S[x]/\varepsilon}, \quad S[x] = \int_{-\tau}^{0} [\dot{x}(t) - F(x(t))]^2 dt, \tag{5.7}$$

where the rate function $S[x]$ is known as an action functional. (Here the approximation sign means that we are only considering the leading-order behavior and ignoring prefactors.) Note that the above LDP can also be derived formally using Onsager-Machlup path-integrals [119, 120].

If we are only interested in the pdf $P(x, \tau)$ of the state $X(0)$ given $x(-\tau) = x_s$, then a coarser-grained LDP can be derived of the from

$$P(x, \tau) \approx \mathrm{e}^{-\Phi(x,\tau)/\varepsilon}, \quad \Phi(x, \tau) = \inf_{\{x(t):x(-\tau)=x_s, x(0)=x\}} S[x], \tag{5.8}$$

with Φ identified as the so–called *quasipotential.* That is, in the limit $\varepsilon \to 0$ the pdf for $X(0)$ given the initial condition $X(-\tau) = x_s$ is obtained by minimizing the action functional with respect to the set of all trajectories from x_s to x. In particular, if x_s is a metastable state of the system, then taking the limit $\tau \to -\infty$ generates the steady-state density (assuming it exists).

In order to show this, we introduce a Lagrangian L according to

$$S[x] = \int_{-\tau}^{0} L(x, \dot{x})dt, \quad L(x, \dot{x}) = \frac{1}{2}(\dot{x} - F(x))^2. \tag{5.9}$$

The most probable path is the given by the solution to the Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = \frac{\partial L}{\partial x}. \tag{5.10}$$

Substituting for L, we see that the most probable path satisfies

$$\ddot{x} = F(x)F'(x). \tag{5.11}$$

Suppose that in the zero noise limit there is a globally attracting fixed point x_s such that $F(x_s) = 0$. 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 $x(0) = x$. Multiplying both sides of equation (5.11) by $\dot{x}$ and integrating with respect to t shows that $\dot{x}(t)^2 = F(x(t))^2 + \text{constant}$. The initial condition implies that the constant is zero and the end condition implies that the most probable path satisfies $\dot{x} = -F(x)$. It follows that the quasipotential is

$$\Phi(x) = -2\int_{-\infty}^{0} F(x)\dot{x}dt = 2\int_{-\infty}^{0} U'(x)\dot{x}dt = 2\int_{x_s}^{x} U'(x)dx = 2U(x).$$

where we have set $F(x) = -U'(x)$ with $U(x)$ the potential of the deterministic system. Hence, we obtain the expected result that the stationary density is

$$\rho^*(x) \approx \mathrm{e}^{-2U(x)/\epsilon}.$$

In the case of a scalar SDE one can of course solve the steady-state FP equation directly. The power of functional LDPs is that one can extend the definition of the quasipotential to multivariate SDEs and to nonlinear systems having multiple attractors. For example, consider the multivariate SDE

$$dX_i(t) = F_i(\mathbf{X})dt + \sqrt{\epsilon}\sum_j b_{ij}(\mathbf{X})dW_i(t), \tag{5.12}$$

for $i = 1, \ldots, d$ with $W_i(t)$ a set of independent Wiener processes. The associated LDP is given by [98] $P[\mathbf{x}] \approx \mathrm{e}^{-S[\mathbf{x}]}$ with action functional

$$S[\mathbf{x}] = \frac{1}{2}\int_{-\tau}^{0}\sum_{i,j=1}^{d}(\dot{x}_i(t) - F_i(\mathbf{x}(t)))D_{ij}^{-1}(\dot{x}_j(t) - F_j(\mathbf{x}(t)))dt, \tag{5.13}$$

where $\mathbf{D} = \mathbf{b}\mathbf{b}^{\mathrm{tr}}$ is the diffusion matrix. Suppose that the underlying deterministic system has multiple attracting fixed points $\mathbf{x}_r, r = 1, 2....$ The quasipotential $\Phi(\mathbf{x})$ 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 quasipotential obtained by initiating paths at $\mathbf{x}_r$:

$$\Phi_r(\mathbf{x}) = \inf_{\mathbf{x}(-\infty)=\mathbf{x}_r, \mathbf{x}(0)=\mathbf{x}} S[\mathbf{x}].$$

Note that the minimum over r switches abruptly on a separatrix separating the basins of attraction of the fixed points.

5.1.2. Mean first passage times and the WKB method. We now establish a connection between LDPs and the WKB method. Consider the FP equation corresponding to the scalar SDE (5.6):

$$\frac{\partial \rho}{\partial t} = -\frac{\partial[F(x)\rho(x,t)]}{\partial x} + \frac{\epsilon}{2}\frac{\partial^2 \rho(x,t)}{\partial x^2} \equiv -\frac{\partial J(x,t)}{\partial x}, \tag{5.14}$$

where

$$J(x,t) = -\frac{\epsilon}{2}\frac{\partial \rho(x,t)}{\partial x} + F(x)\rho(x,t).$$

Suppose that the deterministic equation $\dot{x} = F(x)$ has a stable fixed point x_-, $F(x_-) = 0$, and a basin of attraction given by the interval $\Omega = (0, x_*)$; the point x_* 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_*. Assume that the stochastic system is initially at x_- so that $\rho(x,0) = \delta(x - x_-)$. In order to solve the first passage time problem for escape from the basin of attraction of x_-, we impose an absorbing boundary condition at x_*, $\rho(x_*, t) = 0$, and a reflecting boundary condition at $x = 0$. Let T denote the (stochastic) first passage time for which the system first reaches x_*, given that it started at x_-. The distribution of first passage times is related to the survival probability that the system hasn't yet reached x_*:

$$S(t) \equiv \int_\Omega \rho(x,t)dx. \tag{5.15}$$

That is, Prob$\{t > T\} = S(t)$ and the first passage time density is

$$f(t) = -\frac{dS}{dt} = -\int_\Omega \frac{\partial \rho}{\partial t}(x,t)dx. \tag{5.16}$$

Substituting for $\partial p/\partial t$ using the FP equation (5.14) shows that

$$f(t) = \int_\Omega \frac{\partial J(x,t)}{\partial x}dx = J(x_*,t) = -\frac{\epsilon}{2}\frac{\partial \rho(x_*,t)}{\partial x}. \tag{5.17}$$

We have used $J(0,t) = 0$ and $\rho(x_*,t) = 0$. The first passage time density can thus be interpreted as the probability flux $J(x_*,t)$ at the absorbing boundary.

The first passage time problem in the weak noise limit ($\epsilon \ll 1$) has been well studied in the case of FP equations, see for example [174, 176, 187, 222]. One of the characteristic features of the weak noise limit is that the flux through the absorbing boundary is exponentially small. Let $\langle T\rangle = \int_0^\infty f(t)tdt$ denote the mean first passage time (MFPT) to reach the absorbing boundary. Then $\lambda = 1/\langle T\rangle \sim \mathrm{e}^{-C/\epsilon}$ for some constant C, which reflects the existence of an underlying LDP. In order to make this connection more explicit, we consider the eigenfunction expansion

$$\rho(x,t) = \sum_r C_r \mathrm{e}^{-\lambda_r t}\phi_r(x), \tag{5.18}$$

where $(-\lambda_r, \phi_r(x))$ are an eigenpair of the linear operator

$$\mathbb{L} = -\frac{\partial}{\partial x}F(x) + \frac{\epsilon}{2}\frac{\partial^2}{\partial x^2}$$

appearing on the right-hand side of (5.14). That is,

$$\mathbb{L}\phi_r(x) = -\lambda_r\phi_r(x), \tag{5.19}$$

together with the absorbing boundary conditions $\phi_r(x_*) = 0$. We also assume that the eigenvalues $\lambda_\varepsilon^{(r)}$ all have positive definite real parts and the smallest eigenvalue $\lambda_\varepsilon^{(0)}$ is real and simple, so that we can introduce the ordering $0 < \lambda_\varepsilon^{(0)} < \mathrm{Re}[\lambda_\varepsilon^{(1)}] \leq \mathrm{Re}[\lambda_\varepsilon^{(2)}] \leq \ldots$. The exponentially slow rate of escape through x_* in the weak-noise limit means that $\lambda_\varepsilon^{(0)}$ is exponentially small, $\lambda_\varepsilon^{(0)} \sim \mathrm{e}^{-C/\epsilon}$, whereas $\mathrm{Re}[\lambda_\varepsilon^{(r)}]$ is only weakly dependent on ε for $r \geq 1$. Under the above assumptions, we have the *quasistationary* approximation for large t

$$\rho(x,t) \sim C_0 \mathrm{e}^{-\lambda_\varepsilon^{(0)}t}\phi_\varepsilon^{(0)}(x), \tag{5.20}$$

and the FPT density takes the form $f(t) \sim \lambda_\varepsilon^{(0)}\mathrm{e}^{-\lambda_\varepsilon^{(0)}t}$ with $1/\lambda_\varepsilon^{(0)}$ identified as the MFPT and

$$\lambda_\varepsilon^{(0)} = \frac{J_0(x_*)}{\int_\Omega \phi_0(x)dx}, \quad J_0(x) = -\frac{\varepsilon}{2}\frac{\partial \phi_\varepsilon^{(0)}}{\partial x}. \tag{5.21}$$

The calculation of the principle eigenvalue λ_0 consists of two major components [174, 176, 187, 222]: (i) a WKB approximation of the quasistationary state, which also provides an alternative method for deriving the quasipotential 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. The first step involves seeking a quasistationary solution of the WKB form

$$\phi_\varepsilon^{(0)}(x) \sim K(x;\epsilon)\mathrm{e}^{-\Phi(x)/\epsilon}, \tag{5.22}$$

with $K(x;\epsilon) \sim \sum_{m=0}^{\infty} \epsilon^m K_m(x)$. Substitute equation (5.22) into the eigenvalue equation $\mathbb{L}\phi_\varepsilon^{(0)}(x) = -\lambda_\varepsilon^{(0)}\phi_\varepsilon^{(0)}(x)$ and Taylor expand with respect to ϵ using the fact that $\lambda_\varepsilon^{(0)}$ is exponentially small. Collecting the $O(1)$ terms gives

$$\frac{1}{2}\left(\frac{\partial\Phi(x)}{\partial x}\right)^2 + F(x)\frac{\partial\Phi(x)}{\partial x} = 0. \tag{5.23}$$

Similarly, collecting $O(\epsilon)$ terms yields the following equation for the leading contribution K_0 to the pre factor:

$$\left[\frac{\partial\Phi}{\partial x} + F(x)\right]\frac{\partial K_0}{\partial x} = -\left[F'(x) + \frac{1}{2}\frac{\partial^2\Phi(x)}{\partial x^2}\right]K_0(x). \tag{5.24}$$

The latter either has the solution

$$K_0(x) = \frac{C_0}{F(x) + \Phi'(x)}, \tag{5.25}$$

along non-optimal paths and $K_0 =$ constant along optimal paths. Equation (5.23) has the form of a Hamilton-Jacobi (HJ) equation for a classical Newtonian particle. That is, introducing the time-independent Hamiltonian

$$H(x,p) = \frac{p^2}{2} + F(x)p, \tag{5.26}$$

we see that equation (5.23) can be rewritten as the "zero-energy" HJ equation

$$H(x,\Phi'(x)) = 0. \tag{5.27}$$

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

$$\dot{x} = \frac{\partial H}{\partial p} = p + F(x) \tag{5.28a}$$

$$\dot{p} = -\frac{\partial H}{\partial x} = -pF'(x). \tag{5.28b}$$

The Hamiltonian is related to a classical Lagrangian $L(x,\dot{x})$ according to the Legendre transformation

$$H(x,p) = p\dot{x} - L(x,\dot{x}), \quad p = \frac{\partial L}{\partial \dot{x}}. \tag{5.29}$$

It follows that we recover the Lagrangian of equation (5.9). Moreover, from the least action principle of classical mechanics, we can identify $\Phi(x)$ as the quasipotential of large deviation theory, see equation (5.8). Note that $\Phi(x)$ is determined along a zero energy trajectory, since we are interested in the quasistationary density, and thus should take $\tau \to -\infty$ and x_s to be the fixed point x_-. Since $p = \Phi'(x)$ along this trajectory, it follows that the quasipotential can be computed from

$$\Phi(x) = \int_{x_-}^{x} p(y)dy. \tag{5.30}$$

We now need to match the outer quasistationary solution with an appropriate inner solution $\Pi(x)$ in a neighborhood of the point $x = x_*$. This is necessary since the quasistationary solution does not satisfy the absorbing boundary condition at x_*. There are a number of different ways of carrying out the matched asymptotics [125,174,176,187,222]. Here we will proceed by fixing the probability flux J_0 through

x_* and carrying out a diffusion approximation of the inner solution. Introducing stretched coordinates $y = (x - x_*)/\sqrt{\varepsilon}$, the inner solution $\Pi(y)$ satisfies the stationary FP equation for constant flux, that is,

$$J_0 = F(x_* + \sqrt{\varepsilon}y)\Pi(y) - \frac{\sqrt{\varepsilon}}{2}\frac{\partial \Pi(y)}{\partial y}, \tag{5.31}$$

which yields the solution

$$\Pi(y) = \frac{2J_0}{\sqrt{\varepsilon}} \mathrm{e}^{-\Phi(x_*+\sqrt{\varepsilon}y)/\varepsilon} \int_y^\infty \mathrm{e}^{\Phi(x_*+\sqrt{\varepsilon}y')/\varepsilon} dy'.$$

with $2F(x) = -\Phi'(x)$. Taylor expanding the potential to second order in $\varepsilon^{1/2}y$ and reintroducing unstreteched coordinates gives the inner solution $\phi_i(x) = \varepsilon^{-1/2}\Pi(x/\sqrt{\varepsilon})$,

$$\phi_i(x) = \frac{2J_0}{\varepsilon} \mathrm{e}^{(x-x_*)^2/\sigma^2} \int_x^\infty \mathrm{e}^{-(x'-x_*)^2/\sigma^2} dx', \tag{5.32}$$

where $\sigma = \sqrt{2\varepsilon/|\Phi''(x_*)|}$ determines the size of the boundary layer around x_*. In order to match with the outer solution we note that for $x_* - x \gg \sigma$ we can take the lower limit in the integral to be $-\infty$ and evaluate the resulting Gaussian integral:

$$\phi_i(x) = \frac{2J_0\sigma\sqrt{\pi}}{\varepsilon} \mathrm{e}^{(x-x_*)^2/\sigma^2} \tag{5.33}$$

Similarly, Taylor expanding the outer solution about x_* and matching shows that

$$J_0 = \frac{C_0}{2}\sqrt{\frac{\varepsilon|\Phi''(x_*)|}{2\pi}} \mathrm{e}^{-\Phi(x_*)/\varepsilon}. \tag{5.34}$$

Finally, using steepest descents we evaluate the normalization of the outer solution:

$$\int_\Omega \phi_\varepsilon^{(0)}(x)dx \approx C_0 \int_{-\infty}^\infty \mathrm{e}^{-\varepsilon^{-1}[\Phi(x_-)+\Phi''(x_-)(x-x_-)^2/2]} = C_0\sqrt{\frac{2\pi\varepsilon}{\Phi''(x_0)}}. \tag{5.35}$$

From equation (5.21), the mean transition rate is then

$$\lambda_\varepsilon^{(0)} = \frac{1}{4\pi}\sqrt{|\Phi''(x_*)|\Phi''(x_-)} \mathrm{e}^{-(\Phi(x_*)-\Phi(x_-))/\varepsilon}. \tag{5.36}$$

Higher-dimensional SDEs. Equation (5.36) will be recognized as the classical result for the inverse MFPT obtained by solving a backwards FP equation [105]. However, the point of using the more complicated analysis is that it can be generalized to higher dimensions with multiplicative nose [35, 174, 222]. The simplest case is when the drift term is given by the gradient of a potential so that

$$dX_i(t) = F_i(\mathbf{X})dt + \sqrt{\varepsilon}\sum_{j=1}^n D_{ij}(\mathbf{x})W_j(t),\ F_i(\mathbf{X}) = -\nabla_i V(\mathbf{X}) \tag{5.37}$$

for $i = 1, \ldots, d$ with $W_i(t)$ a set of independent Wiener processes. Suppose that $\mathbf{x}_s$ is a stable fixed point of the deterministic system, (a minimum of the potential $V(\mathbf{x})$), and that the basin of attraction of the fixed point consists of separatrices linking one or more saddle points and unstable fixed points, see Fig. 16. Also let $D_{ij} = \delta_{ij}$ (additive noise). Suppose that there exists an optimal path that passes through a distinguished saddle $\mathbf{x}_H$. One can then derive the classical Eyring formula

$$\lambda_\varepsilon^{(0)} = \frac{|\mu_1(x_H)|}{2\pi}\sqrt{\frac{\det(\mathbf{Z}(\mathbf{x}_S))}{|\det(\mathbf{Z}(\mathbf{x}_H)|}} \mathrm{e}^{-2(V(\mathbf{x}_H)-V(\mathbf{x}_S))/\varepsilon}, \tag{5.38}$$

where $\mu_1(x_H)$ is the single negative eigenvalue of $\mathbf{Z}(\mathbf{x}_H)$, where $\mathbf{Z}(\mathbf{x})$ is the so-called Hessian matrix with components $Z_{ij} = \partial_i \partial_j V$.

One of the major mathematical challenges is deriving a version of the Eyring formula under more general conditions, including boundaries with multiple saddles and unstable fixed points, non-smooth gradient systems, and non-gradient systems. Maier and Stein [174] have used WKB methods and matched asymptotics to investigate two-dimensional non-gradient systems with multiplicative noise. Now the physical potential V is replaced by a quasipotential $\Phi/2$ satisfying an HJ equation $H(\mathbf{x}, \nabla\Phi) = 0$ with Hamiltonian

$$H(\mathbf{x}, \mathbf{p}) = \frac{1}{2}\sum_{i,j} D_{ij}(\mathbf{x}) p_i p_j + \sum_{i=1}^{n} F_i(\mathbf{x}) p_i, \quad p_i = \frac{\partial \Phi}{\partial x_i}. \tag{5.39}$$

They show that the Eyring formula breaks down unless the drift is locally given by some gradient potential around $\mathbf{x}_H$ and $|\mu_s(\mathbf{x}_H)|/\mu_u(\mathbf{x}_H)| > 1$. (If these conditions do not hold then the Hessian matrix of the quasipotential develops a discontinuity at $\mathbf{x}_H$ so standard Gaussian approximations of the inner solution break down). Here μ_s and μ_u are the negative and positive eigenvalues of the 2D Hessian matrix at $\mathbf{x}_H$.; the only modification is that one has to determine the prefactor $K_0(\mathbf{x}_H)$, since it is no longer a constant. The equation for the prefactor K_0 now takes the form

$$\dot{K_0} \equiv \sum_{i=1}^{n} \frac{\partial H}{\partial p_i}\frac{\partial K_0}{\partial x_i} = -\left[\sum_i \frac{\partial^2 H}{\partial p_i \partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2 \Phi}{\partial x_i \partial x_j}\frac{\partial^2 H}{\partial p_i \partial p_j}\right] K_0. \tag{5.40}$$

We have used the fact that along a trajectory $\mathbf{x}(t)$, $\dot{K}_0(\mathbf{x}(t)) = \dot{\mathbf{x}} \cdot \nabla K_0$. Hence, K_0 can be determined numerically by integrating along trajectories originating from a neighborhood of the fixed point $\mathbf{x}_S$, provided that the Hessian $\mathbf{Z}$ of Φ is known. It turns out that the Hessian also satisfies an evolution equation. That is, differentiating the HJ equation twice using

$$\left(\frac{\partial}{\partial x_i} + \sum_k \frac{\partial p_k}{\partial x_i}\frac{\partial}{\partial p_k}\right)\left(\frac{\partial}{\partial x_j} + \sum_k \frac{\partial p_k}{\partial x_j}\frac{\partial}{\partial p_k}\right) H = 0,$$

we find that [174]

$$\dot{Z}_{ij} = -\sum_{k,l} \frac{\partial^2 H}{\partial p_k \partial p_l} Z_{ik} Z_{jl} - \sum_k \frac{\partial^2 H}{\partial x_j \partial p_k} Z_{ik} - \sum_k \frac{\partial^2 H}{\partial x_i \partial p_k} Z_{jk} - \frac{\partial^2 H}{\partial x_i \partial x_j}. \tag{5.41}$$

Thus one can proceed numerically by simultaneously integrating Hamilton's equations together with equations (5.40) and (5.41).

One of the features of higher-dimensional escape problems is that there tends to be a distribution of fluctuational paths from a metastable state to the boundary of its basin of attraction, see Fig. 16. Following Chan et al. [61], one can define a switching path-distribution that gives the probability density of passing a given point in state space during switching. In certain cases it can be shown both theoretically and experimentally that the switching-path distribution consists of a narrow ridge in state space whose cross-section is Gaussian except in a neighborhood of the metastable state. Moreover, the maximum of the ridge lies along the most probable switching path that is obtained from large deviation theory.

5.2. Metastability in an autoregulatory network

The analysis of metastability in chemical master equations has been developed along analogous lines to SDEs, combining WKB methods [81,85,90,124,129,177,217], large deviation principles [96], and path-integral or operator methods [77,78,203,221,247]. The study of metastability in stochastic hybrid systems (SHS) is more recent, and much of the theory has been developed in a series of papers on stochastic ion channels [47, 190, 193, 194], gene networks [191, 192, 195] and stochastic neural networks [43]. Again there is a strong connection between WKB methods, large deviation principles [48, 91, 92, 153] and formal path-integral methods [46, 49], although the connection is now more subtle. Given that one can often approximate a chemical master equation or a PDMP using an FP equation (see section 2), one might be inclined to estimate the MFPT to escape from a metastable state using the results outlined above. However, it is clear from the Arrhenius-like formula (5.36) that the escape rate is sensitive to the precise form of the quasipotential Φ. That is, since a diffusion approximation of a master equation or PDMP leads to an approximation of the corresponding quasipotential, this can lead to exponentially large errors in estimates of the escape rate. Hence, it is often necessary to apply WKB methods and large deviation principles directly to the underlying master equation or CK equation.

We will illustrate the WKB analysis of metastability in discrete Markov processes and stochastic hybrid systems by considering the simple autoregulatory network of section 3.4, in which the dynamics of mRNA is ignored, see equations (3.32*a*) and (3.32*b*) with the matrix $\mathbf{A} \to \mathbf{A}/\varepsilon$. We will consider separately the two limiting cases $\varepsilon \to 0$ (adiabatic limit) and $N \to \infty$ (thermodynamic limit). Note, however, that one could analyze the more general model by combining the two cases [192]. In the limit $\varepsilon \to 0$ for fixed N, we have $P_s(n,t) = \rho_s^*(n/N)P(n,t)$, where ρ_s^* is given by equation (3.37), and $P(n,t)$ evolves according to the birth-death master equation

$$\frac{dP(n,t)}{dt} = \omega_+(n-1)P(n-1,t) + \omega_-(n+1)P(n+1,t) - [\omega_+(n) + \omega_-(n)]P(n,t). \tag{5.42}$$

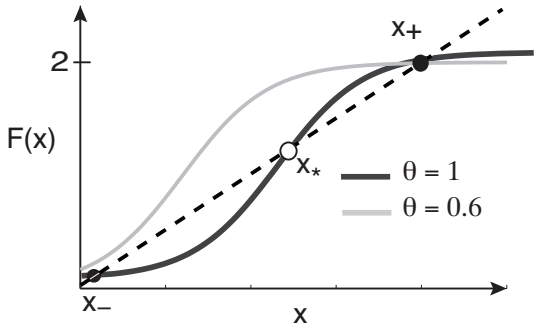


Figure 17. Fixed points of 1D equation $\dot{x} = -x + F(x)$ with $F(x) = 2[1 + \mathrm{e}^{-\kappa(x-\theta)}]^{-1}$. If $\kappa > 1$ then the network can exhibit bistability for a range of thresholds θ. The sigmoid function arises if the corresponding transition rates are taken to be $\alpha(x) = \mathrm{e}^{\kappa(x-\theta)}$ and $\beta(x) = 1$, while $\sigma_0 = 0$ and $\sigma_1 = 2$.

The transition rates are

$$\omega_+(n) = NF(n/N), \quad \omega_-(n) = n, \tag{5.43}$$

with

$$F(x) = \sum_{n=0,1} \rho_n^*(x)(s\sigma_1 + \sigma_0). \tag{5.44}$$

The units of time are fixed by setting the protein degradation rate to unity $\gamma_p = 1$. On the other hand, in the limit $N \to \infty$ for fixed ε, we obtain a stochastic hybrid system evolving according to the CK equation (see (3.33a) and (3.33b))

$$\frac{\partial \rho_n(x,t)}{\partial t} \equiv \mathbb{L}\rho_n(x,t) = -\frac{\partial F_n(x)\rho_n(x,t)}{\partial x} + \frac{1}{\varepsilon}\sum_{m=0,1} A_{nm}(x)\rho_m(x,t), \tag{5.45}$$

where n now represents the state of the gene rather than the number of proteins, and

$$\mathbf{A}(x) = \begin{pmatrix} -\alpha(x) & \beta(x) \\ \alpha(x) & -\beta(x) \end{pmatrix}, \quad F_n(x) = n\sigma_1 + \sigma_0 - x$$

Note that if $x(0) \in [\sigma_0, \sigma_0 + \sigma_1]$ then $x(t) \in \Sigma = [\sigma_0, \sigma_0 + \sigma_1]$ for all $t > 0$ such that $F_0(x) < 0$ and $F_1(x) > 0$ (see discussion of boundary conditions in section 2.6). In the double limit $N \to \infty$ and $\varepsilon \to 0$, irrespective of the order, we obtain the deterministic equation

$$\dot{x} = \sum_{n=0,1} \rho_n^*(x)F_n(x) \equiv -x + F(x). \tag{5.46}$$

Rather than identifying $F(x)$ with the second-order Hill function $F_a(x)$ of equation (3.25), we will consider a more general sigmoid function of the form

$$F(x) = \frac{2}{1 + \mathrm{e}^{-\kappa(x-\theta)}}.$$

We will assume that the deterministic system (5.46) is bistable with stable fixed points $x_\pm$ separated by an unstable fixed point x_*, see Fig. 17.

5.2.1. Metastability in the adiabatic limit. We first analyze metastability for the birth-death process given by the master equation (5.42) with n representing the number of proteins. For large but finite N, we set $n_\pm = Nx_\pm$ and $n_* = Nx_*$ and impose an absorbing boundary condition at $n = n_*$. Let T denote the (stochastic) first passage time for which the birth-death process first reaches n_*, given that it started at n_-. The distribution of first passage times is related to the survival probability that the system hasn't yet reached n_*:

$$S(t) = \sum_{n<n_*} P(n,t). \tag{5.47}$$

That is, $\mathrm{Prob}\{t > T\} = S(t)$ and the first passage time density is

$$f(t) = -\frac{dS}{dt} = -\sum_{n<n_*} \frac{dP(n,t)}{dt}. \tag{5.48}$$

We now note that equation (5.42) can be written in the form

$$\frac{dP(n,t)}{dt} = J(n,t) - J(n+1,t) \tag{5.49}$$

with

$$J(n,t) = \omega_-(n)P(n,t) - \omega_+(n-1)P(n-1,t).$$

Using the reflecting boundary condition $J(0,t) = 0$, it follows that $f(t) = J(n_*,t)$. The first passage time density can thus be interpreted as the probability flux $J(n_*,t)$ at the absorbing boundary.

Suppose that we rewrite the birth-death master equation (5.42) for $0 \leq n \leq n_*$ in the more compact form

$$\frac{dP(n,t)}{dt} = \sum_{m \leq n_*} A_{nm} P(m,t) \tag{5.50}$$

with $-\mathbf{A}$ the generator of the continuous Markov chain (see section 2). For the given model, the matrix $\mathbf{A}$ is irreducible. If the absorbing boundary condition at $n = n_*$ 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 (see section 2.1) then ensures that all other eigenvalues of $-\mathbf{A}$ have positive real part and that equation (5.50) has a globally attracting steady-state p_n^* such that $\sum_m A_{nm} p_m^* = 0$ and $P(n,t) \to p_n^*$ 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 non-zero flux at $n = n_*$ (for large N). The eigenvalues of $-\mathbf{A}$ can now be ordered according to $0 < \lambda_\varepsilon^{(0)} < \mathrm{Re}[\lambda_\varepsilon^{(1)}] \leq \mathrm{Re}[\lambda_\varepsilon^{(2)}] \leq \ldots$ with $\lambda_\varepsilon^{(0)} \sim \mathrm{e}^{-\eta N}$ for $\eta = \mathscr{O}(1)$, whereas $\lambda_\varepsilon^{(r)}$ for $r > 0$ are only weakly dependent on N (polynomial rather than exponential functions of N). The exponentially small principal eigenvalue reflects the fact that the flux through the absorbing boundary is exponentially small, and in the limit $N \to 0$ it reduces to the zero Perron eigenvalue.

Now consider the eigenfunction expansion

$$P(n,t) = \sum_r C_r \mathrm{e}^{-\lambda_\varepsilon^{(r)} t} \phi_\varepsilon^{(r)}(n), \tag{5.51}$$

where $(\lambda_\varepsilon^{(r)}, \phi_\varepsilon^{(r)}(n))$ are eigenpairs of the matrix operator $-\mathbf{A}$, supplemented by the absorbing boundary conditions $\phi_\varepsilon^{(r)}(n_*) = 0$. Since $\lambda_\varepsilon^{(0)} \ll \mathrm{Re}[\lambda_\varepsilon^{(r)}]$, we have the *quasistationary* approximation for large t

$$P(n,t) \sim C_0 \mathrm{e}^{-\lambda_\varepsilon^{(0)} t} \phi_\varepsilon^{(0)}(n), \tag{5.52}$$

Substituting into equation (5.50) and summing over n, $0 \leq n \leq n_*$ shows that the FPT density takes the form $f(t) \sim \lambda_\varepsilon^{(0)} \mathrm{e}^{-\lambda_\varepsilon^{(0)} t}$ with $\lambda_\varepsilon^{(0)}$ identified as the mean transition rate (inverse MFPT) and

$$\lambda_\varepsilon^{(0)} = \frac{J^{(0)}(n_*)}{\sum_{n \leq n_*} \phi_\varepsilon^{(0)}(n)}, \quad J^{(0)}(n) = \omega_-(n)\phi_\varepsilon^{(0)}(n) - \omega_+(n-1)\phi_\varepsilon^{(0)}(n-1). \tag{5.53}$$

As in the case of an SDE (section 5.1), the calculation of the principle eigenvalue $\lambda_\varepsilon^{(0)}$ consists of two major components, namely, a WKB approximation of the principal eigenfunction in the bulk of the domain (outer solution), and the asymptotic matching of the outer solution with an inner solution around x_*, which satisfies the absorbing boundary condition.

WKB approximation of quasistationary state. Write $\phi_\varepsilon^{(0)}(n) = \phi_\varepsilon(n/N)$ and $\omega_\pm(n) = N\Omega_\pm(n/N)$ with $x = n/N$ treated as a continuous variable. Since $\lambda_\varepsilon^{(0)}$ is exponentially small we set it to zero and treat $\phi_\varepsilon(x)$ as a quasistationary solution:

$$0 = \Omega_+(x-1/N)\phi_\varepsilon(x-1/N) + \Omega_-(x+1/N)\phi_\varepsilon(x+1/N) - (\Omega_+(x)+\Omega_-(x))\phi_\varepsilon(x). \tag{5.54}$$

We then seek make the WKB ansatz

$$\phi_\varepsilon(x) \sim K(x;\varepsilon)\mathrm{e}^{-\Phi(x)/\varepsilon} \quad \varepsilon = N^{-1}, \tag{5.55}$$

with $K(x;\varepsilon) \sim \sum_{m=0}^{\infty} \varepsilon^m K_m(x)$. This solution is taken to hold in the bulk of the domain $[0, x_*]$ excluding $O(\varepsilon^{1/2})$ neighborhoods of the stable fixed point x_- and the unstable fixed point x_*. The latter is necessary since the WKB approximation does not satisfy the absorbing boundary condition at x_*. Substituting equation (5.55) into equation (5.54), Taylor expanding with respect to ε, and collecting the $O(1)$ terms gives

$$\Omega_+(x)(\mathrm{e}^{\Phi'(x)} - 1) + \Omega_-(x)(\mathrm{e}^{-\Phi'(x)} - 1) = 0 \tag{5.56}$$

where $\Phi' = d\Phi/dx$. Solving this quadratic equation in $\mathrm{e}^{\Phi'}$ shows that

$$\Phi = \int^x \ln \frac{\Omega_-(y)}{\Omega_+(y)} dy \tag{5.57}$$

or $\Phi =$ constant. 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 (5.56) and solving for K_0 yields the following leading order forms for ϕ_ε:

$$\phi_\varepsilon(x) = \frac{\mathscr{A}}{\sqrt{\Omega_+(x)\Omega_-(x)}}\mathrm{e}^{-N\Phi(x)}. \tag{5.58}$$

with Φ given by (5.57), which is sometimes called the activation solution, and

$$\phi_\varepsilon(x) = \frac{\mathscr{B}}{\Omega_+(x) - \Omega_-(x)}, \tag{5.59}$$

which is sometimes called the relaxation solution. The constants $\mathscr{A}, \mathscr{B}$ are determined by matching solutions around x_0. Clearly, (5.59) 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.

We now observe that equation (5.56) has the form of a stationary Hamilton–Jacobi (HJ) equation [143]:

$$H(x, \Phi'(x)) = 0, \tag{5.60}$$

with Hamiltonian given by

$$H(x,p) = (\mathrm{e}^p - 1)\Omega_+(x) + (\mathrm{e}^{-p} - 1)\Omega_-(x). \tag{5.61}$$

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

$$\dot{x} = \frac{\partial H}{\partial p} = \sum_{r=\pm 1} r\Omega_r(x)\mathrm{e}^{rp} \tag{5.62}$$

$$\dot{p} = -\frac{\partial H}{\partial x} = \sum_{r=\pm 1} \frac{\partial \Omega_r}{\partial x}(x)\left[1 - \mathrm{e}^{rp}\right] \tag{5.63}$$

Defining the Lagrangian $L(x, \dot{x})$ according to the Legendre transformation (5.29) we can express the quasipotential as

$$\Phi(x) = \inf_{x(-\infty)=x_-, x(\tau)=x} \int_0^\tau L(x, \dot{x}) dt. \tag{5.64}$$

Analogous to SDEs, the fact that the quasipotential can be expressed as the infimum over trajectories from x_- to x, is a reflection of an underlying large deviation principle [96]. From a dynamical systems perspective, the most probable fluctuational path of large deviation theory is the unstable manifold of the fixed point $(x_-, 0)$ in the Hamiltonian phase space $[0, x_*] \times \mathbb{R}$. Since $p = \Phi'(x)$ along this trajectory, it follows that the quasipotential can be computed from

$$\Phi(x) = \int_{x_-}^{x} p dx. \tag{5.65}$$

The corresponding stable manifold is obtained by setting $p = 0$ in Hamilton's equations, which recovers the deterministic solution of equation (5.46) with $\lim_{t\to\infty} x(t) = x_-$. In Fig. 18 we illustrate the Hamiltonian phase space for the particular transition rates (5.43), showing the constant energy solutions; the zero energy activation and relaxation trajectories through the fixed points of the deterministic system are highlighted as thicker curves. Note that $N\Phi(x_*)$, where $\Phi(x_*)$ is the area enclosed by the heteroclinic connection from x_- to x_*, gives the leading order contribution to $\log \tau_-$, where τ_- is the mean escape time from x_-. Similarly, the area under the heteroclinic connection from x_+ to x_* gives the corresponding leading contribution to $\log \tau_+$, where τ_+ is the mean escape time from x_+ . We immediately deduce that for the chosen parameter values $\tau_+ < \tau_-$.

Matched asymptotics at x_*. Given the WKB approximation, the rate of escape $\lambda_\varepsilon^{(0)}$ from the metastable state centered about $x = x_-$ can be calculated by matching

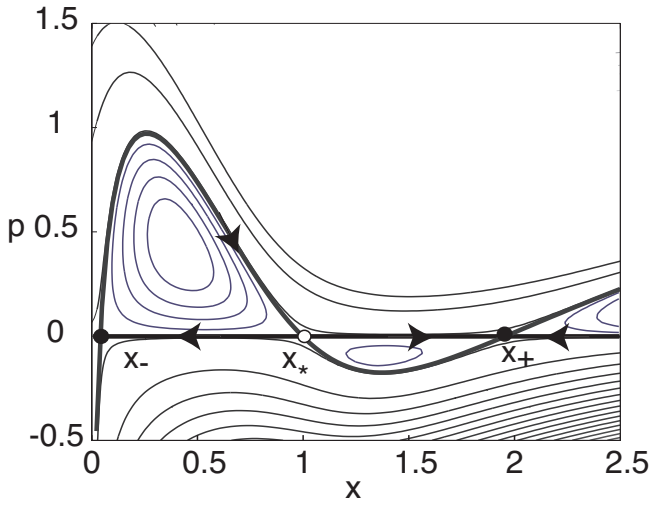


Figure 18. Phase portrait of Hamiltonian equations of motion for $\Omega_+(x) = \sigma_1/(1+\mathrm{e}^{-\kappa(x-\theta)})$ and $\Omega_-(x) = x$ with $\kappa = 4, \theta = 1.0$ and $\sigma_1 = 2$. The zero energy solutions are shown as thicker curves.

the quasistationary solution with an appropriate inner solution in a neighborhood of the fixed point $x = x_*$. This is necessary since the quasistationary solution does not satisfy appropriate boundary conditions at the saddle point separating the two metastable states. There are a number of different ways of carrying out the matched asymptotics [81,85,90,124,129,217]. Here we will proceed by fixing the probability flux J_0 through x_* and carrying out a diffusion approximation of the inner solution [90,129]. Introducing stretched coordinates $y = (x - x_*)/\sqrt{\varepsilon}$, the inner solution $\Pi(y)$ satisfies the stationary FP equation for constant flux, that is,

$$J_0 = V(x_* + \sqrt{\varepsilon}y)\Pi(y) - \frac{\sqrt{\varepsilon}D(x_*)}{2}\frac{\partial \Pi(y)}{\partial y}, \tag{5.66}$$

where

$$V(x) = \Omega_+(x) - \Omega_-(x), \quad D(x) = \Omega_+(x) + \Omega_-(x).$$

This yields the solution

$$\Pi(y) = \frac{2J_0}{\sqrt{\varepsilon}} \mathrm{e}^{-2U(x_* + \sqrt{\varepsilon}y)/D(x_*)\varepsilon} \int_y^\infty \mathrm{e}^{2U(x_* + \sqrt{\varepsilon}y')/D(x_*)\varepsilon} dy'.$$

with $V(x) = -U'(x)$ and $D(x_*) = 2\Omega_+(x_*)$. Taylor expanding the potential to second order in $\varepsilon^{1/2}y$ and reintroducing unstretched coordinates gives the inner solution $\phi_i(x) = \varepsilon^{-1/2}\Pi(x/\sqrt{\varepsilon})$,

$$\phi_i(x) = \frac{2J_0}{\varepsilon D(x_*)} \mathrm{e}^{(x-x_*)^2/\Gamma^2} \int_x^\infty \mathrm{e}^{-(x'-x_*)^2/\Gamma^2} dx', \tag{5.67}$$

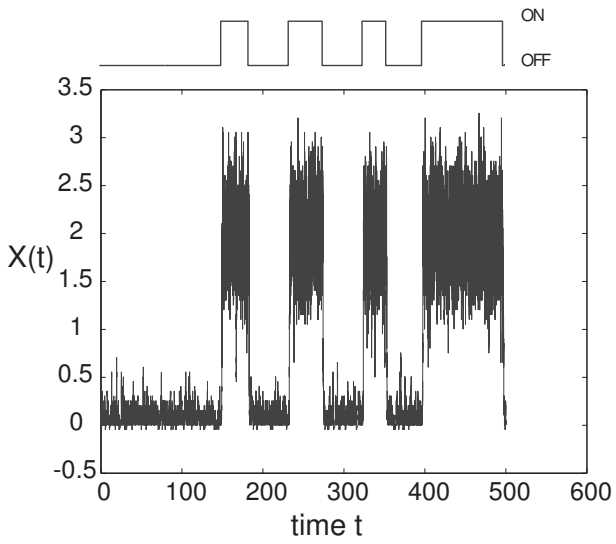


Figure 19. Time series showing a single realization of the stochastic autoregulatory gene network evolving according to the birth-death master equation (5.42) with transition rates (5.43). Parameters are $\theta = 0.86$, $\kappa = 4.0$, $\sigma_1 = 2$ and $N = 20$. The bistable nature of the process is clearly seen.

where

$$\Gamma = \sqrt{\frac{\varepsilon D(x_*)}{|U''(x_*)|}} = \sqrt{\frac{2\Omega_+(x_*)}{N[\Omega'_+(x_*) - \Omega'_-(x_*)]}} \tag{5.68}$$

determines the size of the boundary layer around x_*. In order to match with the outer solution we note that for $x_* - x \gg \Gamma$ we can take the lower limit in the integral to be $-\infty$ and evaluate the resulting Gaussian integral:

$$\phi_i(x) = \frac{J_0 \Gamma \sqrt{\pi}}{\varepsilon \Omega_+(x_*)} \mathrm{e}^{(x-x_*)^2/\Gamma^2} \tag{5.69}$$

Similarly, Taylor expanding the outer solution about x_* using $\Phi^{'}(x_*) = 0$ and $\Phi^{''}(x_*) = -2\varepsilon/\Gamma^2$ and matching shows that (for $\varepsilon^{-1} = N$)

$$J_0 = \frac{A\Omega_+(x_*)}{\sqrt{\Omega_+(x_*)\Omega_-(x_*)}} \sqrt{\frac{|\Phi''(x_*)|}{2\pi N}} \mathrm{e}^{-N\Phi(x_*)}. \tag{5.70}$$

Finally, using steepest descents we evaluate the normalization of the outer solution:

$$\begin{aligned} \int_\Sigma \phi_\varepsilon^{(0)}(x)dx \quad &\approx \frac{A}{\sqrt{\Omega_+(x_*)\Omega_-(x_*)}} \int_{-\infty}^{\infty} \mathrm{e}^{-N[\Phi(x_-)+\Phi''(x_-)(x-x_-)^2/2]} \\ &= \frac{A}{\sqrt{\Omega_+(x_*)\Omega_-(x_*)}} \sqrt{\frac{2\pi}{N\Phi''(x_-)}}. \end{aligned} \tag{5.71}$$

From equation (5.21), the mean transition rate is then

$$r_- \equiv \lambda_\varepsilon^{(0)} = \frac{\Omega_+(x_-)}{2\pi} \sqrt{|\Phi''(x_*)|\Phi''(x_-)} \mathrm{e}^{-N[\Phi(x_*)-\Phi(x_-)]}. \tag{5.72}$$

Similarly, the escape rate from the metastable state x_+ is

$$r_+ = \frac{\Omega_+(x_+)}{2\pi} \sqrt{|\Phi''(x_*)|\Phi''(x_+)} \mathrm{e}^{-N[\Phi(x_*)-\Phi(x_+)]}. \tag{5.73}$$

Finally, note that the long–term behavior of the stochastic bistable gene network can be approximated by a two–state Markov process that only keeps track of which metastable state the system is close to [244], see also Fig. 19:

$$\frac{d}{dt} \begin{pmatrix} P_- \\ P_+ \end{pmatrix} = \widehat{\mathbf{Q}} \begin{pmatrix} P_- \\ P_+ \end{pmatrix}, \quad \widehat{\mathbf{Q}} = \begin{pmatrix} -r_- & r_+ \\ r_- & -r_+ \end{pmatrix} \tag{5.74}$$

where $P_\pm$ are the probabilities of being in a neighborhood of $x_\pm$. The matrix $\widehat{\mathbf{Q}}$ has eigenvectors $\widehat{\lambda}_0 = 0$ and $\widehat{\lambda}_1 = -(r_+ + r_-)$ and corresponding eigenvectors

$$\widehat{\mathbf{v}}_0 = \begin{pmatrix} r_+/(r_+ + r_-) \\ r_-/(r_+ + r_-) \end{pmatrix}, \quad \widehat{\mathbf{v}}_1 = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}. \tag{5.75}$$

Note that the two–state model generates an exponential density for the residence times within a given metastable state. This captures the behavior of the full master equation at large residence times but fails to capture the short–term dynamics associated with relaxation trajectories within a neighborhood of the metastable state. Finally, we note that certain care has to be taken if the fixed point x_- is too close to the boundary at $x = 0$, as highlighted in Ref. [129].

5.2.2. Metastability in the thermodynamic limit. In the thermodynamic limit for fixed ε, the autoregulatory network evolves according to a stochastic hybrid system with corresponding CK equation (5.45). Now $n \in \Gamma = \{0, 1\}$ represents the discrete state of the gene. In order to calculate the mean transition time for escape from the metastable state x_-, we supplement the CK equation (5.45) by the absorbing boundary conditions

$$\rho_0(x_*, t) = 0, \tag{5.76}$$

since $F_0(x_*) < 0$. The initial condition is taken to be $\rho_n(x, 0) = \delta(x - x_-)\delta_{n,n_0}$. Let T denote the (stochastic) first passage time for which the system first reaches x_*, given that it started at x_-. The distribution of first passage times $f(t)$ is related to the survival probability that the system hasn't yet reached x_* according to equation (5.48) with

$$S(t) = \int_\Sigma \sum_{n\in\Gamma} \rho_n(x, t)dx. \tag{5.77}$$

Substituting for $\partial\rho_n/\partial t$ using the CK equation (5.45) shows that

$$f(t) = \int_\Sigma \left[\sum_{n\in\Gamma} \frac{\partial[F_n(x)\rho_n(x, t)]}{\partial x}\right] dx = \sum_{n\in\Gamma} \rho_n(x_*, t)F_n(x_*), \tag{5.78}$$

with $\Gamma = \{0, 1\}$ for the two-state model. We have used $\sum_{n\in\Gamma} A_{nm}(x) = 0$ and $\lim_{x\to-\infty} F_n(x)\rho_n(x, t) = 0$. Note that for the sake of generality we will develop the following analysis in the case of the more general discrete set $\Gamma = \{0, 1, \ldots, N_0 - 1\}$, under the assumption that we can partition the functions $F_n(x)$, $n \in \Gamma$, into positive and negative subsets, see section 2.6. We will then return to the two-state model ($N_0 = 2$).

As in the case of the discrete Markov process, we can identify the mean transition rate with the principal eigenvalue $\lambda_\varepsilon^{(0)}$ of the CK operator $-\mathbb{L}$ in equation (5.45), assuming $\lambda_\varepsilon^{(0)}$ exists and is exponentially small. We can then make the quasistationary approximation

$$\rho_n(x, t) \sim C_0 \mathrm{e}^{-\lambda_\varepsilon^{(0)} t}\phi_\varepsilon^{(0)}(x, n). \tag{5.79}$$

Substituting such an approximation into equation (5.78) gives

$$f(t) \sim C_0 \mathrm{e}^{-\lambda_\varepsilon^{(0)} t} \sum_{n\in\Gamma} F_n(x_*)\phi_\varepsilon^{(0)}(x, n). \tag{5.80}$$

and thus

$$\lambda_\varepsilon^{(0)} = \frac{\sum_{n\in\Gamma} F_n(x_*)\phi_\varepsilon^{(0)}(x_*, n)}{\sum_n \int_\Sigma \phi_\varepsilon^{(0)}(x, n)dx}. \tag{5.81}$$

Proceeding along analogous lines to the birth-death master equation, we seek a WKB approximation of the quasistationary solution of the form

$$\phi_\varepsilon^{(0)}(x, n) \sim Z_n(x) \exp\left(-\frac{\Phi(x)}{\varepsilon}\right), \tag{5.82}$$

where $\Phi(x)$ is the WKB quasipotential. Substituting into the time-independent version of equation (5.45) yields

$$\sum_m \left(A_{nm}(x) + \Phi'(x)\delta_{n,m}F_m(x)\right) Z_m(x) = \varepsilon \frac{dF_n(x)Z_n(x)}{dx}, \tag{5.83}$$

where $\Phi' = d\Phi/dx$. Introducing the asymptotic expansions $\Phi \sim \Phi_0 + \varepsilon\Phi_1$ and $Z \sim Z^{(0)} + \varepsilon Z^{(1)}$, the leading order equation is

$$\sum_{m\in\Gamma} A_{nm}(x)Z_m^{(0)}(x) + \Phi_0'(x)F_n(x)Z_n^{(0)}(x) = 0. \tag{5.84}$$

(Since the prefactor is a component of a vector, we separately expand Φ and $\mathbf{Z}^{(0)}$.) Positivity of the quasistationary density $\phi_\varepsilon^{(0)}$ requires positivity of the corresponding solution $\mathbf{Z}^{(0)}$. One positive solution is the trivial solution $\mathbf{Z}^{(0)}(x) = \rho^*(x)$ for all $x \in \Sigma$, where ρ^* is the unique right eigenvector of $\mathbf{A}$, for which $\Phi_0' = 0$. Establishing the existence of a non-trivial positive solution requires more work, and is related to the fact that the connection of the WKB solution to optimal fluctuational paths and large deviation principles is less direct.

It turns out that one has to consider the eigenvalue problem [46, 49, 91, 92, 153]

$$\sum_{m\in\Gamma} \left[A_{nm}(x) + p\delta_{n,m}F_m(x)\right] R_m(x,p) = \Lambda(x,p)R_n(x,p). \tag{5.85}$$

Assuming that $\mathbf{A}(x)$ is irreducible for all x, one can use the Perron-Frobenius theorem to show that for fixed (x,p) there exists a unique eigenvalue $\Lambda_0(x,p)$ with a positive eigenvector $R_n^{(0)}(x,p)$. The optimal fluctuational paths are obtained by identifying the Perron eigenvalue $\Lambda_0(x,p)$ as a Hamiltonian and finding zero energy solutions to Hamilton's equations

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial x}, \quad H(x,p) = \Lambda_0(x,p). \tag{5.86}$$

Comparison of equation (5.84) with equation (5.85) then shows that there exists a nontrivial positive solution of equation (5.84) given by $Z_n^{(0)}(x) = R_n^{(0))}(x,p)$ with $p = \Phi_0'(x)$ and Φ_0 satisfies the corresponding HJ equation

$$\Lambda_0(x, \Phi_0'(x)) = 0. \tag{5.87}$$

Note that since $\Phi_0'(x)$ vanishes at $x = x_*$ it follows that $\mathbf{Z}^{(0)}(x_*) = \rho^*(x_*)$, and similarly for the other fixed points. The interesting feature of PDMPs is that the WKB method does not generate the correct Hamiltonian of large deviation theory except along fluctuational paths where the Hamiltonian is zero.

Calculation of principal eigenvalue. In order to calculate the principal eigenvalue, we have to determine the first order correction Φ_1 to the quasipotential of the WKB solution (5.82). Proceeding to the next order in the asymptotic expansion of equation (5.83), we have

$$\sum_m \left(A_{nm}(x) + \Phi_0'(x)\delta_{n,m}F_n(x)\right) Z_m^{(1)}(x) = \frac{dF_n(x)Z_n^{(0)}(x)}{dx} - \Phi_1'(x)F_n(x))Z_n^{(0)}(x). \tag{5.88}$$

For fixed x and WKB potential Φ_0, the matrix operator $\bar{A}_{nm}(x) = A_{nm}(x) + \Phi_0'(x)\delta_{n,m}F_m(x)$ on the left-hand side of this equation has a one-dimensional null space spanned by the positive WKB solution $\mathbf{Z}^{(0)}(x)$. The Fredholm Alternative

Theorem† then implies that the right-hand side of (5.88) is orthogonal to the left null vector S of $\bar{A}$. That is, we have the solvability condition

$$\sum_{n\in\Gamma} S_n(x)\left[\frac{dF_n(x)Z_n^{(0)}(x)}{dx} - \Phi_1'(x)F_n(x)Z_n^{(0)}(x)\right] = 0,$$

with S satisfying

$$\sum_{n\in\Gamma} S_n(x)\left(A_{nm}(x) + \Phi_0'(x)\delta_{n,m}F_m(x)\right) = 0. \tag{5.89}$$

Given $\mathbf{Z}^{(0)}, \mathbf{S}$ and Φ_0, the solvability condition yields the following equation for Φ_1:

$$\Phi_1'(x) = \frac{\sum_{n\in\Gamma} S_n(x)[F_n(x)Z_n^{(0)}(x)]'}{\sum_{n\in\Gamma} S_n(x)F_n(x)Z_n^{(0)}(x)}. \tag{5.90}$$

Combining the various results, and defining

$$k(x) = \exp\left(-\Phi_1(x)\right), \tag{5.91}$$

gives to leading order in ε,

$$\phi_\varepsilon^{(0)}(x,n) \sim \mathcal{N}k(x)\exp\left(-\frac{\Phi_0(x)}{\varepsilon}\right)Z_n^{(0)}(x), \tag{5.92}$$

where we choose $\sum_n Z_n^{(0)}(x) = 1$ for all x and $\mathcal{N}$ is the normalization factor,

$$\mathcal{N} = \left[\int_\Sigma k(x)\exp\left(-\frac{\Phi_0(x)}{\varepsilon}\right)\right]^{-1}.$$

The latter can be approximated using Laplace's method to give

$$\mathcal{N} \sim \frac{1}{k(x_-)}\sqrt{\frac{|\Phi_0''(x_-)|}{2\pi\varepsilon}}\exp\left(\frac{\Phi_0(x_-)}{\varepsilon}\right). \tag{5.93}$$

The final step is to use singular perturbation theory to match the outer quasistationary solution to the absorbing boundary condition at x_*. The analysis is quite involved, see [148, 191, 192], so here we simply quote the result for the 1D model:

$$\lambda_0 \sim \frac{1}{\pi}\frac{k(x_*)D(x_*)}{k(x_-)}\sqrt{\Phi_0''(x_-)|\Phi_0''(x_*)|}\exp\left(-\frac{\Phi_0(x_*)-\Phi_0(x_-)}{\varepsilon}\right). \tag{5.94}$$

with $D(x)$ the effective diffusion coefficient (2.54) obtained using a QSS reduction.

Two-state model. We now illustrate the above theory for the simple two-state autoregulatory network of equation (5.45). An example of a larger discrete set ($N_0 > 2$) will be considered in section 8.1 for a model of stochastic ion channels. The specific version of the linear equation (5.85) can be written as the two-dimensional system

$$\begin{pmatrix} -\alpha(x) + pF_0(x) & \beta(x) \\ \alpha(x) & -\beta(x) + pF_1(x) \end{pmatrix}\begin{pmatrix} R_0 \\ R_1 \end{pmatrix} = \Lambda\begin{pmatrix} R_0 \\ R_1 \end{pmatrix}. \tag{5.95}$$

† Consider an M-dimensional linear inhomogeneous system $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{x}, \mathbf{b} \in \mathbb{R}^M$. Suppose that the $M \times M$ matrix $\mathbf{A}$ has a nontrivial null-space and let $\mathbf{v}$ be a null vector of the adjoint matrix $\mathbf{A}^\dagger$, that is, $\mathbf{A}^\dagger\mathbf{v} = 0$. The Fredholm Alternative Theorem states that the inhomogeneous equation has a (non-unique) solution if and only if $\mathbf{v}\cdot\mathbf{b} = 0$ for all null vectors $\mathbf{v}$.

The corresponding characteristic equation is

$$0 = \Lambda^2 + \Lambda[\alpha(x) + \beta(x) - p(F_0(x) + F_1(x))] + (pF_1(x) - \beta(x))(pF_0(x) - \alpha(x)) - \beta(x)\alpha(x).$$

It follows that the Perron eigenvalue is given by

$$\Lambda_0(x,p) = \frac{1}{2}\left[\Sigma(x,p) - \sqrt{\Sigma(x,p)^2 - 4h(x,p)}\right] \tag{5.96}$$

where

$$\Sigma(x,p) = p(F_0(x) + F_1(x)) - [\alpha(x) + \beta(x)],$$

and

$$h(x,p) = p^2 F_1(x)F_0(x) - p[\beta(x)F_0(x) + \alpha(x)F_1(x)].$$

A little algebra shows that

$$\mathscr{D}(x,p) \equiv \Sigma(x,p)^2 - 4h(x,p) = [p(F_0 - F_1) - (\alpha(x) - \beta(x))]^2 + \alpha(x)\beta(x) > 0$$

so that as expected Λ_0 is real. The quasipotential $\Phi_0(x)$ satisfies the HJ equation $\Lambda_0(x,p) = 0$ with $p = \Phi_0'(x)$, which reduces to the condition

$$h(x, \Phi_0'(x)) = 0. \tag{5.97}$$

This has two solutions: the classical deterministic solution $p = 0$ with $\Phi_0'(x) = 0$ and a non-trivial solution whose quasipotential satisfies

$$\Phi_0'(x) = \frac{\beta(x)}{F_1(x)} + \frac{\alpha(x)}{F_0(x)}. \tag{5.98}$$

(Note that $F_n(x)$ does not vanish anywhere and $F_0(x)F_1(x) < 0$.) The quasipotential can be determined by numerically integrating with respect to x. As shown in [192], the resulting quasipotential differs significantly from the one obtained by carrying out a QSS diffusion approximation of the stochastic hybrid system along the lines outlined in section 2.6.

For this simple model, it is also straightforward to determine the various prefactors in equation (5.94). For example, the normalized positive eigenvector $\mathbf{Z}^{(0)}$ has components

$$Z_0^{(0)} = \frac{F_1(x)}{F_1(x) - F_0(x)}, \quad Z_1^{(0)} = \frac{-F_0(x)}{F_1(x) - F_0(x)}.$$

Since $F_0(x) < 0$ and $F_1(x) > 0$ for $0 < x < \sigma$, it follows from equation (5.98) that $Z_0^{(0)}$ is positive. The components of the adjoint eigenvector $\mathbf{S}$ satisfy

$$\frac{S_1}{S_0} = \frac{-\alpha + \Phi_0'(x)F_0(x)}{\alpha} = \frac{-\beta + \Phi_0'(x)F_1(x)}{\beta}.$$

It then follows from equation (5.90) that the first correction to the quasipotential satisfies

$$\Phi_1'(x) = \frac{1}{F_0(x)F_1(x)}\frac{d}{dx}(F_0(x)F_1(x)). \tag{5.99}$$

Hence

$$k(x) \equiv \mathrm{e}^{-\Phi_1(x)} = \frac{1}{|F_0(x)|F_1(x)}. \tag{5.100}$$

Finally, $D(x_*)$ is given by equation (2.57). For extensions to the two-dimensional models considered in section 3, namely, the mutual repressor network and the autoregulatory network with both mRNA and protein dynamics included, see [151, 191, 195].

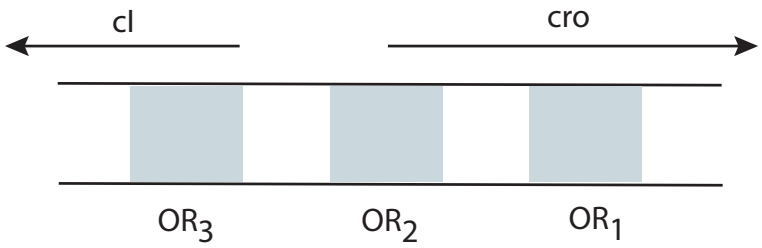


Figure 20. Schematic diagram of operator complex OR with three binding sites OR_j, $j = 1, 2, 3$. The gene cI is transcribed when OR_3 is free and OR_2 is occupied by the protein CI, whereas the gene cro is transcribed when both OR_2 and OR_1 are free. Dimers of CI bind cooperatively to OR_1 and OR_2.

5.3. Metastability in epigenetics

Aurell *et al.* [21, 22] have developed a theoretical framework for studying the metastability of epigenetic states (see also [230]). *Epigenetics* concerns phenotypic states that are not encoded as genes, but as inherited patterns of gene expression originating from environmental factors. Examples of environmental influences range from changes in the supply of nutrients in bacteria to stress in humans. The simplest and best studied example of inherited gene expression involves the infection of *E. coli* by the λ phage DNA virus [206]. Following infection, the λ phage either multiplies and ultimately kills the host (lysis), or it integrates its DNA into that of the host (lysogeny) and is passively replicated over many generations. These two distinct scenarios involve the expression of different genes. In wild-type, one finds that the lyosgenic state is extremely stable with spontaneous loss occurring around 10^{-5} per cell and generation, which corresponds to a lifetime of five years. Hence, the spontaneous escape from the metastable lysogenic state is an excellent example of a rare event. Here we briefly review the model of metastability by Aurell and Sneppen [21].

Lysogeny is maintained by a regulatory network involving λ phage DNA, a pair of regulatory proteins CI and Cro, and an operator complex OR consisting of three binding sites OR_j, $j = 1, 2, 3$, overlapping with two promoter sites P_{RM} and P_R, see Fig. 20. Either CI or Cro can bind to the operators OR_j. The protein CI has the highest affinity for OR_1, and when it is bound it blocks RNA polymerase from binding to the promoter P_R and initiating transcription of the gene cro. On the other hand, the protein Cro mainly binds to OR_3 consequently blocking the promoter P_{RM} and the synthesis of CI. The rate of initiation of cI transcription also depends on whether or not CI is bound to OR_2. We see that the regulatory network is a more complicated version of the mutual repressor model of a bistable switch, see section 3.5. Lysogeny is maintained during bacterial growth provided that the number of CI molecules per bacterial cell is sufficiently high (200-350 per cell [206]). However, if the CI concentration becomes too low, then increased activation of cro increases the concentration of Cro protein and decreases cI activation. Thus lysogeny ends and lysis begins. This can be modeled in terms of a bistable switch, in which one of the metastable fixed points is the lysogenic state, and termination of lysogenesis occurs when there is a noise-induced path that exits the basin of attraction of this state.

First suppose that the numbers of CI and Cro within a cell are large so that the dynamics can be represented in terms of mass action kinetic equations for the

concentrations $x_{\rm CI}$ and x_{Cro} [213]:

$$\frac{dx_{\rm CI}}{dt} = \phi_{\rm CI}(x_{\rm CI}, x_{\rm Cro}), \quad \frac{dx_{\rm Cro}}{dt} = \phi_{\rm Cro}(x_{\rm CI}, x_{\rm Cro}), \tag{5.101}$$

where the net production rates are taken to be of the form

$$\phi_J = S_J f_J(x_{\rm CI}, x_{\rm Cro}) - \frac{x_J}{\tau_J}, \quad J = \text{CI,Cro}. \tag{5.102}$$

Here S_J is the number (per unit volume) of proteins of type J produced from one mRNA of the corresponding gene, whereas the transcription rates f_J are taken to have the following form:

$$\begin{aligned} f_{\rm CI} &= \widehat{R}_{RM}(P_{010} + P_{011} + P_{012}) + R_{RM}(P_{000} + P_{001} + P_{002} + P_{020} + P_{021} + P_{022}) \\ f_{\rm Cro} &= \widehat{R}_R(P_{000} + P_{100} + P_{200}), \end{aligned} \tag{5.103}$$

where P_s is the probability of a state $s = (i_3, i_2, i_1)$ with the three numbers referring respectively to OR_3, OR_2 and OR_1, and are based on the coding $i_n = 0$ if the corresponding site is free, $i_n = 1$ if the site is occupied by a CI dimer, and $i_n = 2$ if the site is occupied by a Cro dimer. Moreover, R_{RM} is the base rate of transcription of cI, $\widehat{R}_{RM}$ is the stimulated rate when CI is bound to OR_2, and R_R is the transcription rate of cro. Finally, the grand canonical ensemble of equilibrium statistical mechanics is used to determine P_s:

$$P_s = \mathscr{N}^{-1} x_{\rm CI}^{m_s} x_{\rm Cro}^{n_s} \mathrm{e}^{-G(s)/RT}, \tag{5.104}$$

where m_s and n_s are the numbers of CI and Cro dimers bound to operator sites in state s, and $G(s)$ is the associated free energy.

In reality, the numbers of CI and Cro proteins involved in lysogenesis are only in the range of hundreds. This has two implications for the analysis of the model. First, one has to take into account conservation of molecules in order to determine the number of free dimers that can bind to sites on the operator OR. Second, intrinsic fluctuations in the number of proteins occur. In Ref. [21], these fluctuations are modeled in terms of an effective SDE of the form

$$dX_{\rm CI} = \phi_{\rm CI} dt + \frac{1}{\sqrt{\Omega}} g_{\rm CI} dW_{\rm CI}, \quad dX_{\rm Cro} = \phi_{\rm Cro} dt + \frac{1}{\sqrt{\Omega}} g_{\rm CI} dW_{\rm Cro}, \tag{5.105}$$

where $W_{\rm CI}(t)$ and $W_{\rm Cro}(t)$ are independent Wiener processes, Ω is the system size, and the noise amplitudes are

$$g_J = \sqrt{S_J^2 f_J + X_J/\tau_J}, \quad J = \text{CI,Cro}. \tag{5.106}$$

This SDE can be derived by carrying out a system-size expansion of the associated chemical master equation along the lines of section 2.4. Aurell and Sneppen then analyze metastability of the lysogenic state by applying the large deviation theory of SDEs described in section 5.1. In particular, the Hamiltonian (5.39) is given by [21]

$$H = \frac{1}{2}(g_{\rm CI}^2 p_{\rm CI}^2 + g_{\rm Cro}^2 p_{\rm Cro}^2) + p_{\rm CI}\phi_{\rm CI} + p_{\rm Cro}\phi_{\rm Cro}, \tag{5.107}$$

where p_J is the momentum conjugate to the generalized coordinate x_J. The optimal path of escape from the metastable lysogenic state is determined from the zero-energy solutions of Hamilton's equations

$$\dot{x}_J = \frac{\partial H}{\partial p_J}, \quad \dot{p}_J = -\frac{\partial H}{\partial x_J}, \quad J = \text{CI,Cro}. \tag{5.108}$$

One of the interesting aspects of the analysis in Ref. [21] is the quantification of the robustness of the lysogenic state in terms of the value of the minimized action $S_{\rm min}$ along the optimal path of escape. That is, since the rate of escape $\sim \mathrm{e}^{-\Omega S_{\rm min}}$, higher stability corresponds to a larger action.

6. Bacterial growth and switching in changing environments

A challenging problem for microbial organisms such as bacteria is how to adapt to a randomly fluctuating environment involving, for example, sudden changes in chemical composition, local temperature, or light illumination. One strategy is to sense these environmental changes and respond appropriately by switching phenotype or behavior. However, there is a cost: each individual must maintain an active sensory system. An alternate way to adapt to randomly fluctuating environments is to maintain population diversity, whereby different sub-populations are well-adapted to different types of environments (see the review [197]). There are variety of microbial organisms that achieve phenotypic diversity by stochastic phenotype-switching mechanisms. One example is the persistence mechanism in *E. coli*, where cells switch spontaneously and reversibly to a phenotype exhibiting slower growth and reduced killing by antibiotics [24]; this allows cells to survive prolonged exposure to antibiotics. Another example is the lactose metabolism of *E. coli*: if an inducer that is not metabolized binds to the repressor of the lac operon, then only a part of the bacteria expresses lac [229]. This allows a more robust response to changes in lactose densities in the environment. A third example occurs in the quorum sensing system of *V. harveyi*, see section 7.2, where it has been found that only a fraction of the cells respond to an autodinducer (a signaling molecule that can be exchanged with the environment and whose concentration depends on cell density) [11].

The potential advantages of stochastically switching phenotypes in fluctuating environments has been known for a long time in ecology and population genetics, where it is known as *bet-hedging* [224]. In this scenario, each phenotype is better suited to a particular environmental condition so that there is a fitness tradeoff. More recently a number of mathematical models of bet-hedging have been developed within the context of microbial populations [1,103,160,186,214,236]. Roughly speaking, when changes in the environment are slow and cells in the wrong state are subject only to moderate stress (weak selection), then stochastic switching can be preferable to active environmental sensing. On the other hand, if environmental changes are unpredictable and more severe, active sensing is generally preferred, provided that cells are capable of responding quickly to new conditions. If the cells cannot respond in a timely fashion then stochastic switching again becomes preferable. Stochastic switching can also be beneficial in cases where the environment is favorable most of the time but occasionally causes a catastrophe that wipes out most of the population [168, 245].

In section 3 we described various mechanisms for genes to exhibit switching between low and high levels of activity, in particular, noise-induced switching in bistable gene regulatory networks. This suggests that stochastic gene expression provides a direct mechanism for populations of cells to undergo different fates. In this section we will simply assume that individual cells can switch between different phenotypes, and describe various stochastic models of cell populations evolving in random environments. We will assume that the switching times of the phenotypes and environmental states are exponentially distributed. Note that by observing the evolution of successive generations one can obtain statistics regarding the distribution of switching or residence times, the relative abundance of different phenotypes and the fitness of the population, see Fig. 21.

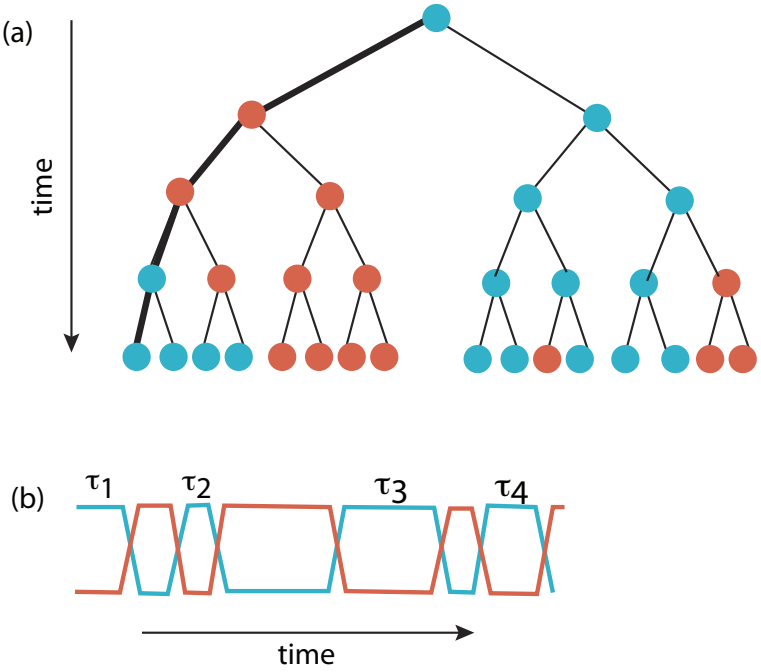


Figure 21. Population-level phenotypic switching. (a) An individual mother cell grows and divides into two daughter cells, producing a family tree over multiple generations. Switching between two phenotypes is shown using two different colors. At a given generation, the relative abundance of each of the two phenotypes can be determined by measuring across the corresponding row. Alternatively, one can keep track of phenotypic switching along one lineage of the expanding tree (indicated by the thicker branches). If this single lineage is observed for a sufficiently long time then one should recover the steady-state population incidence of the two phenotypes, assuming the system is ergodic. (b) Keeping track of the durations τ_l of each phenotype generates the residence time distribution.

6.1. Stochastic population model of bacterial persistence

We begin by considering a stochastic population model of bet-hedging in *E. coli* due to Kussell and Leibler [160], see also [186]. Consider a bacterial population growing in an environment $K(t)$ that switches between a finite number n of different states according to a jump Markov process. Let $P_k(t) = \mathbb{P}[K(t) = k]$ denote the probability that the environment is in state k at time t, which evolves according to the master equation (section 2.4)

$$\frac{dP_k}{dt} = \sum_{l \neq k} W_{kl} P_l(t) - \left(\sum_{l \neq k} W_{lk} \right) P_k(t), \tag{6.1}$$

where $\mathbf{W}$ is the transition rate matrix of the continuous-time Markov chain. Define τ_k and b_{kl} by

$$\tau_k^{-1} = \sum_{j \neq k} W_{jk}, \quad W_{kl} = b_{kl}/\tau_l.$$

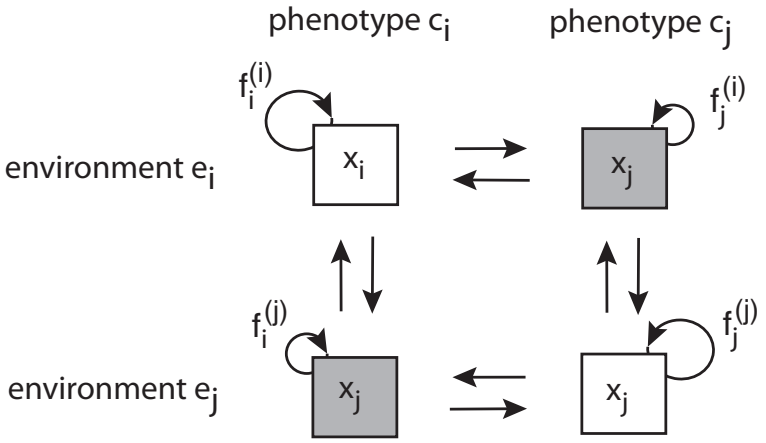


Figure 22. Bet-hedging model of switching between bacterial phenotypes and environmental states. For the sake of illustration two environmental and two phenotypic states are shown. The phenotype c_i (c_j) grows faster in the environmental state e_i (e_j). Irrespective of the labels i, j, the fit state is denoted an open box and the unfit state by a shaded box.

such that $\sum_{l\neq k} b_{lk} = 1$. We can interpret τ_l as the mean of the exponentially distributed waiting times for transitions from environmental state l and b_{kl} as the probability of jumping to state $k \neq l$ when a transition occurs. Setting $p_k = P_k/\tau_k$ we obtain the rescaled master equation

$$\frac{dp_k}{dt} = \sum_{l\neq k} b_{kl} p_l(t) - p_k(t), \tag{6.2}$$

At equilibrium, we have

$$p_k^* = \sum_{l\neq k} b_{kl} p_l^*.$$

in which p_k^* can be interpreted as the steady-state probability that a transition to state k occurs.

Each individual organism can exist in one of n different phenotypes. Suppose that phenotype i grows with rate $f_i^{(k)}$ in environment k (with the possibility of either positive or negative growth), see Fig. 22. Assume that the fastest growing phenotype in environment k is phenotype k, that is $f_k^{(k)} > f_j^{(k)}$ for all $j \neq k$. Let $x_j(t)$ denote the number of bacteria in a large population that has phenotype j at time t. The vector $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$ evolves according to the system of piecewise linear ODEs

$$\frac{dx_i}{dt} = f_i^{(k)} x_i(t) + \sum_j H_{ij}^{(k)} x_j(t) \equiv \sum_j A_{ij}(k) x_j(t), \tag{6.3}$$

for $K(t) = k$. We are assuming that the phenotypes switch according to a jump Markov process with generator $-\mathbf{H}^{(k)}$, which may also depend on the current state of the environment. Equations (6.2) and (6.3) thus provide another example of a stochastic hybrid system, although in this case transitions of the discrete environmental variable do not depend on the continuous variables $\mathbf{x}(t)$. For a given $\mathbf{x}(t)$ the total population size is $N(t) = \sum_{j=1}^n x_j(t)$.

Kussell and Leibler [160] distinguish between two types of phenotypic switching: purely stochastic switching, where the rates are independent of k,

$$H_{ij}^{(k)} = H_{ij} \text{ for all } k,$$

and responsive switching, where sensing mechanisms allow switching to depend strongly on the state of the environment. An extreme version of the latter is where all phenotypes switch at the same rate H_m to phenotype k in environment k:

$$H_{kj}^{(k)} = H_m \text{ for all } j \neq k, \quad H_{ij}^{(k)} = 0 \text{ for all } i \neq k, j \neq i.$$

The effectiveness of these two types of switching can be compared by calculating the so-called Lyapunov exponent Λ, which characterizes the asymptotic growth rate of the total population size. In general, deriving an exact expression for the Lyapunov exponent (assuming it exists) is non-trivial. Therefore, we will follow Ref. [160] and obtain an approximate expression for the Lyapunov exponent.

Suppose that time is divided into consecutive intervals during which the environmental variable $K(t)$ does not change. Let T_l denote the duration of the l-th interval and set $t_L = \sum_{l=1}^{L} T_l$ with $t_0 = 0$. Let n_l be the state of the environment during interval l. Introduce the generalized eigenvectors $\mathbf{v}_r(k)$ of the matrix $\mathbf{A}(k)$ and denote the corresponding eigenvalues by $\lambda_r(k)$ with $\lambda_1(k) > \lambda_2^k \geq \ldots \geq \lambda_n$. The Perron-Frobenius Theorem (see section 2.1) ensures that the principle eigenvalue is non-degenerate and that it's eigenvector is positive. We will take the principal eigenvector to have the normalization $\sum_j (\mathrm{v}_1(k))_j = 1$. To simplify the subsequent analysis, we will assume that all eigenvalues are non-degenerate. However, this is not necessary to derive the final result. Suppose that at the beginning of the l-th interval, we expand the solution $\mathbf{x}(t_{l-1})$ in terms of the eigenvectors $\mathrm{v}_r^{n_l}$:

$$\mathbf{x}(t_{l-1}) = \sum_r C_{l,l-1}^r \mathrm{v}_r(n_l).$$

It follows that for $t_{l-1} < t < t_l$

$$\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}(n_l)(t-t_{l-1})} \sum_r C_{l,l-1}^r \mathrm{v}_r(n_l) = \sum_r \mathrm{e}^{\lambda_r(n_l)(t-t_{l-1})} C_{l,l-1}^r \mathrm{v}_r(n_l).$$

For sufficiently long durations T_l, we can keep only the term involving the principle eigenvalue, that is,

$$\mathbf{x}(t_l) \approx \mathrm{e}^{\lambda_1(n_l)(t_l - t_{l-1})} C_{l,l-1}^1 \mathrm{v}_1(n_l) = N(t_l)\mathrm{v}_1(n_l).$$

Since, we can apply the same argument to all the intervals, we see that

$$\mathbf{x}(t_{l-1}) \approx N(t_{l-1})\mathrm{v}_1(n_{l-1}),$$

where he have used the normalization $\sum_j [\mathrm{v}_1(n_{l-1})]_j = 1$ and the fact that $\sum_j x_j(t) = N(t)$ with $N(t)$ the population size at time t. It follows that

$$C_{l,l-1}^1 = N(t_{l-1})[\mathbf{M}(n_l)^{-1}\mathrm{v}_1(n_{l-1})]_1 \equiv N(t_{l-1})q_{n_l,n_{l-1}},$$

where $\mathbf{M}(k)$ is the matrix whose columns are the vectors $\mathrm{v}_r(k)$.

Iterating the above results, we see that to leading order

$$N(t_L) = \mathrm{e}^{\lambda_1(n_L)(T_L)} q_{n_L,n_{L-1}} N(t_{L-1}) = \prod_{l=1}^{L} \mathrm{e}^{\lambda_1(n_l)(T_l)} q_{n_l,n_{l-1}} N_0.$$

The Lyapunov exponent is then given by the following limit:

$$\Lambda = \lim_{L\to\infty} \frac{\log N(t_L)}{t_L} = \lim_{L\to\infty} \frac{1}{t_L} \left[\sum_{l=1}^{L} \log \left(q_{n_l,n_{l-1}} \mathrm{e}^{\lambda_1(n_l)T_l} \right) + \log N_0 \right]$$

$$= \lim_{L\to\infty} \frac{1}{L\tau} \sum_{l=1}^{L} T_l \lambda_1(n_l) + \lim_{L\to\infty} \frac{1}{L\tau} \sum_{l=1}^{L} \log q_{n_l,n_{l-1}}, \tag{6.4}$$

where $\tau = \lim_{L\to\infty} t_L/L$. is the average duration time. Now recall that transitions between the different environmental states are modeled in terms of a continuous-time Markov chain, see equation (6.2). Let $T_k^{(i)}$ be the duration of the k-th occurrence of environment i, which are exponentially distributed with mean τ_i. For a large number L of elapsed intervals, the number of occurrences of environment i approaches Lp_i^* and the number of transitions $i \to j$ approaches $b_{ij}p_j^*L$. Hence, we can decompose the sums in equation (6.4) for the Lyapunov exponent as follows:

$$\tau\Lambda = \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{n} \sum_{k=1}^{p_i^* L} T_k^{(i)} \lambda_1(i) + \lim_{L\to\infty} \frac{1}{L} \sum_{i,j=1}^{n} p_j^* b_{ij} L \log q_{ij},$$

We thus obtain the final result [160]

$$\tau\Lambda = \sum_{i=1}^{n} p_i^* \tau_i \lambda_1(i) + \sum_{i,j=1}^{n} p_j^* b_{ij} \log q_{ij}, \quad \tau = \sum_{k=1}^{n} p_k^* \tau_k. \tag{6.5}$$

It remains to calculate the principal eigenvalues $\lambda_1(k)$ and the elements q_{ij}. One finds that for responsive switching [160]

$$\tau\Lambda_R = \sum_{i=1}^{n} p_i^* \tau_i f_i^{(i)} - c\tau - \sum_{i,j=1}^{n} p_j^* b_{ij} \log(1 + \Delta_{ji}^R/H_m), \tag{6.6}$$

where

$$\Delta_{ij}^R = \Delta_{ij} \equiv f_j^{(i)} - f_i^{(j)}.$$

Note that the additional term $-c\tau$ has been introduced to take into account the energy costs of requiring environmental sensing machinery, which is assumed to slow the growth rate. The case of stochastic switching is more difficult to analyze. However, if we assume that the switching rates H_{ij} are sufficiently slow, then we can carry out a perturbation expansion in the switching rates to obtain the approximation [160]

$$\tau\Lambda_S = \sum_{i=1}^{n} p_i^* \tau_i f_i^{(i)} - \sum_{i=1}^{n} p_i^* \tau_i \left(\sum_{j\neq i} H_{ij} \right) - \sum_{i,j=1}^{n} p_j^* b_{ij} \log(1 + \Delta_{ji}^S/H_{ij}), \tag{6.7}$$

with

$$\frac{1}{\Delta_{ij}^S} = \frac{1}{\Delta_{ij}} + \frac{1}{\Delta_{ji}}.$$

The first term on the right-hand side of equations (6.6) and (6.7) represents the expected long term growth of the population, whereas the final term takes into account the time taken for the structure of the population to change in response to changes in environment (delay-time cost). The second term in (6.6) is the cost due to sensing the environment, whereas in (6.7) it is the cost associated with switching to slower phenotypes (diversity cost).

Using equation (6.7), we can now determine the optimal phenotype switching rates H_{ij} by maximizing Λ_S. Setting $\partial\Lambda_S/\partial H_{ij} = 0$ and using the approximation $\Delta_{ij}^S/H_{ij} \gg 1$, we find that

$$H_{ij}(\text{optimal}) = \frac{b_{ij}}{\tau_j}. \tag{6.8}$$

Hence, the optimal switching rate is proportional to the probability that the environment undergoes the transition $j \to i$ and is inversely proportional to the average duration of environment j. In other words, optimal rates are precisely tuned to the environmental statistics. This recovers the same basic principle as found in the two-state model.

6.2. Stochastic switching with reset in catastrophic environments

In section 6.1 it was assumed that the environment randomly switches between two or more different states, each favoring a particular bacterial phenotype. Here we consider a distinct scenario, in which there is a single environment that undergoes rare, sudden and instantaneous catastrophes. One example is the sudden flushing out of a microbe from an animal host due to urination or diarrhoea; a more resistant phenotype would be one that is able to attach to the wall of the hist's intestinal or urinary tract. A second example would be the sudden exposure of a population to antibiotics, where cells in a non-growing persister state are more likely to survive. In this section, we describe a particular model due to Visco et al. [245].

Consider two microbial sub-populations A and B corresponding to distinct phenotypes. Between catastrophes, they grow exponentially at at rates γ_A and γ_B with $\gamma_A > \gamma_B$. Thus under normal conditions, sub-population A is fitter. Let n_A and n_B be the number of microbes in the two sub-populations, which evolve between catastrophes according to the pair of equations

$$\frac{dn_A}{dt} = \gamma_A n_A + k_B n_B - k_A n_A, \tag{6.9a}$$

$$\frac{dn_B}{dt} = \gamma_B n_B - k_B n_B + k_A n_A, \tag{6.9b}$$

with k_A, k_B the rates of phenotypic switching. Whenever a catastrophe occurs, the population size n_A decreases instantaneously to some new value $n'_A < n_A$, with a reset probability $\nu(n'_A|n_A)$. The final component of the model is specifying how the rate of catastrophes depends on the population size according to some environmental response function $\beta(n_A, n_B)$. Before specifying the precise form of ν and β, it is useful to introduce the notion of population fitness.

Following [236], define the population fitness f as the fraction of the total population that is in state A:

$$f(t) = \frac{n_A(t)}{n(t)}, \quad n(t) = n_A(t) + n_B(t).$$

A nonlinear dynamical equation for f can be derived by first noting that

$$\frac{dn}{dt} = \gamma_A n_A + n_B \gamma_B = (\gamma_B + \Delta\gamma f)n,$$

where $\Delta\gamma = \gamma_A - \gamma_B > 0$. Differentiating the expression for f with respect to time then shows that

$$\begin{aligned}\frac{df}{dt} &= \frac{1}{n}\frac{dn_A}{dt} - \frac{n_A}{n^2}\frac{dn}{dt} = (\gamma_A - k_A)f + k_B(1-n) - f(\gamma_B + \Delta\gamma f) \\ &= -\Delta\gamma(f - f_+)(f - f_-) \equiv V(f),\end{aligned}$$

where $f_\pm$ are the roots of the quadratic equation

$$f^2 - \left(1 - \frac{k_A + k_B}{\Delta\gamma}\right) f - \frac{k_B}{\Delta\gamma} = 0.$$

It is straightforward to show that $f_- < 0 < f_+ \leq 1$. Hence, the system evolves deterministically, converging towards the fitness value f_+, except at specific times (catastrophes) where it undergoes random jumps. We thus have yet another example of a stochastic hybrid system or piecewise deterministic Markov process. However, in contrast to previous examples, it is not one of the parameters of the deterministic system that jumps, rather it is the deterministic variable itself.

Visco et al. [245] take the rate of catastrophes to depend on the fitness f according to the parameterized sigmoid function

$$\beta_\lambda(f) = \frac{\xi}{2}\left(1 + \frac{f - f^*}{\sqrt{\lambda^2 + (f - f^*)^2}}\right). \tag{6.10}$$

Mathematically speaking, this function is defined for all $f \in \mathbb{R}$, but the physically range is $0 < f < 1$. Note that for finite λ, $\beta_\lambda(f) \to 0$ for $f - f^* \ll \lambda$ and $\beta_\lambda(f) \to \xi$ for $f - f^* \gg \lambda$. The sharpness of the sigmoid increases as λ decreases with $\beta_0(f) = \xi\Theta(f - f^*)$ a Heaviside function. Finally, f^* is the threshold value at which $\beta_\lambda = \xi/2$. The other important element of the model is the probability distribution $\nu(n'_A|n_A)$ for the population reset $n_A \to n'_A$ following a catastrophe. In order to express everything in terms of the fitness f, Visco et al. take

$$\nu(n'_A|n_A) = \frac{1}{n_A}\mathscr{F}(n'_A/n_A)$$

with $\int_0^x \mathscr{F}(x)dx = 1$. That is, when a catastrophe occurs, the fit population A is reduced by a random factor generated from the distribution $\mathscr{F}$. In terms of the fitness f, we have $f \to f'$ with $f' = n'_A/(n'_A + n_B)$. The corresponding distribution of jumps in fitness, $\mu(f'|f)$, then satisfies $\mu(f'|f)df' = \nu(n'_A|n_A)dn'_A$. Since

$$\mathscr{F}(n'_A/n_A) = \mathscr{F}([n'_A/n_B][n_B/n_A]) = \mathscr{F}\left(\frac{f'(1-f)}{f(1-f')}\right),$$

and

$$n_A\frac{df'}{dn'_A} = \frac{n_A}{n'_A + n_B} - \frac{n_A n'_A}{(n'_A + n_B)^2} = \frac{n_A}{n'_A}f'(1-f') = \frac{(1-f')^2 f}{1-f},$$

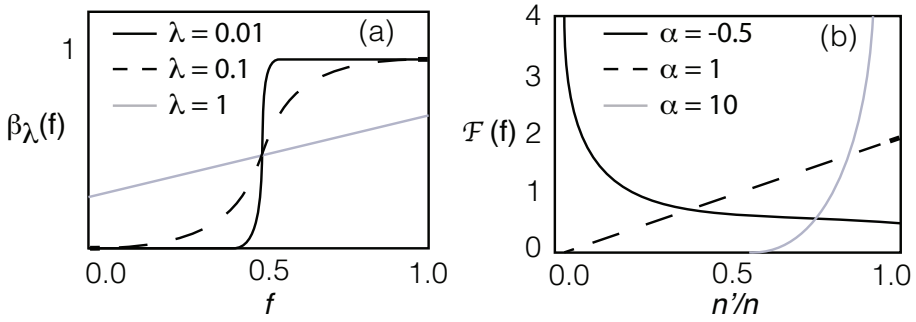


Figure 23. Rate and strength of catastrophes. (a) Sigmoidal rate function $\beta_\lambda(f)$ with $\xi = 1$ and $f^* = 1/2$. (b) Power law jump distribution $\nu(n'_A|n_A)$.

it follows that

$$\mu(f|f') = \mathscr{F}\left(\frac{f'(1-f)}{f(1-f')}\right)\frac{1-f}{(1-f')^2 f}, \quad f > f'. \tag{6.11}$$

In Ref. [245], $\mathscr{F}$ is chosen for simplicity so that $\mu(f|f')$ factorizes: $\mathscr{F}(x) = (\alpha+1)x^\alpha$ with $\alpha > -1$. For this choice,

$$\nu(n'|n) = \frac{\alpha+1}{n}\left(\frac{n'}{n}\right)^\alpha, \tag{6.12}$$

and

$$\mu(f'|f) = \frac{d}{df'}\frac{m(f')}{m(f)}, \quad m(f) = \left(\frac{f}{1-f}\right)^{1+\alpha}. \tag{6.13}$$

The model specific choices for β and ν are shown in Fig. 23.

Steady-state distribution. The steady-state probability distribution for the population fitness, $p(f)$, is determined by a balance condition for the probability flux [245]:

$$V(f)p(f) = \int_f^{f^+}\int_0^f \beta(f')p(f')\mu(f''|f')df''\,df'. \tag{6.14}$$

The left-hand side represents the flux through the state f to higher population fitnesses $f' > f$ due to growth, which must be balanced by the total flux from the states f' to states $F'' < f$ due to catastrophes. Substituting the explicit expression for ν gives

$$V(f)p(f) = \int_f^{f^+}\beta(f')p(f')\frac{m(f')}{m(f)}df'. \tag{6.15}$$

Dividing through by $m(f)$, setting $G(f) = \beta(f)G(f)/V(f)$, and differentiating both sides with respect to f yields

$$\frac{dG}{df} = -\frac{\beta G}{V},$$

which can be solved to give

$$G(f) = C\exp\left(-\frac{\beta(f)}{V(f)}\right),$$

for some constant C. Hence

$$p(f) = \frac{C}{V(f)}\left(\frac{f}{1-f}\right)^{1+\alpha}\exp\left(-\frac{\beta(f)}{V(f)}\right). \tag{6.16}$$

The constant C can then be determined from the normalization of $p(f)$.

The existence of an analytic solution means that we can now investigate under what circumstances random switching is advantageous to the microbial population. One way to characterize this is in terms of the average population fitness

$$\langle f\rangle = \int_0^1 fp(f)df$$

as a a function of the switching rate k_A. In Fig. 24 we sketch results for several values of the gain parameter λ. It can be seen that for large λ, where the catastrophe rate is approximately constant, $\beta = \xi/2$, the average fitness is a monotonically decreasing function of k_A, suggesting that not switching is the best strategy. On the other hand, for sufficiently small λ, there is a local maximum at a non-zero switching rate k_A, implying that switching into the slow-growing state represents an optimal strategy.

6.3. Stochastic model of population extinction

As our final example, we consider a recent WKB analysis of the effect of a catastrophe on extinction in a population model consisting of two distinct phenotypes [168]. As in the previous models, the populations are assumed to be well-mixed so that spatial effects can be ignored. One population consists of so-called "normals" that grow and die, whilst the second population consists of "persisters" that do neither. However, individuals can switch between the two types. The Let n and m denote the number of normals and persisters. Assuming a maximum carrying capacity of N, the normals grow at a rate $B(1-n/N)$ and die at a rate that is set to unity. If the phenotypic switching rates are taken to be constants α, β, then the mean-field kinetic equations (for large populations) are

$$\frac{dn}{dt} = B\left(1-\frac{n}{N}\right)n - n - \alpha n + \beta m, \tag{6.17a}$$

$$\frac{dm}{dt} = \alpha n - \beta m \tag{6.17b}$$

The rate equations have a trivial fixed point at $n = m = 0$, which represents population extinction, and a nontrivial fixed point Q at $n_* = N(1-1/B)$, $m_* = \alpha n/\beta$. The latter fixed point is stable provided that $B > 1$, whereas the zero state is a saddle point. Hence, in the absence of noise and $n(0) > 0$, the system converges to the non-trivial fixed point Q, resulting in a viable steady-state population of phenotypes.

As with biochemical processes, one can write down a stochastic version of the population model based on the discrete nature of individuals. The resulting master equation for the probability distribution $P_{nm}(t) = \mathbb{P}[n(t) = n, m(t) = m]$ is

$$\begin{aligned}\frac{dP_{n,m}}{dt} &= \widehat{A}P_{n,m} \qquad (6.18)\\ &= B(n-1)\left(1-\frac{n-1}{N}\right)P_{n-1,m} + (n+1)P_{n+1,m} - Bn\left(1-\frac{n}{N}\right)P_{n,m} - nP_{n,m}\\ &\quad + \alpha(n+1)P_{n+1,m-1} + \beta(m+1)P_{n-1,m+1} - \alpha n P_{n,m} - \beta(1-\delta_{n,N})P_{n,m}.\end{aligned}$$

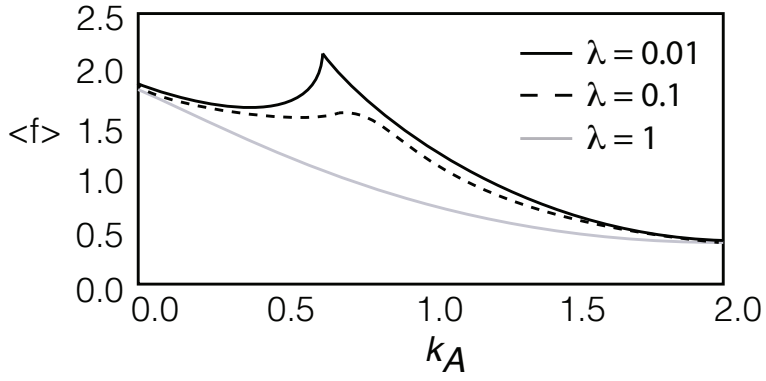


Figure 24. Schematic illustration of results from a model of catastrophes [245]. Sketch of average population fitness $\langle f \rangle$ as a function of the switching rate k_A and various values of λ. Other parameter values are $\lambda = 0$, $k_B = 0.1$, $\Delta\gamma = 1$ $\xi = 1$, $\alpha = -0.99$ and $f^* = 0.5$.

Here $(n,m) \in [0,N] \times [0,\infty)$ with $P_{n<0,m} = P_{n,m<0} = P_{n>N,m} = 0$. In the stochastic model the state $n = m = 0$ is now an absorbing extinction state, with the extinction probability satisfying

$$\frac{dP_{0,0}}{dt} = P_{1,0}. \tag{6.19}$$

The extinction state reflects the existence of a simple zero eigenvalue of the generator $-\widehat{A}$ of the Markov chain with corresponding eigensolution $P_{n,m} = \delta_{n,0}\delta_{m,0}$. Since all other eigenvalues are positive definite, it follows that all other eigensolutions decay to zero, and hence the population goes extinct in the limit $t \to \infty$. We thus have a situation analogous to the stochastic genetic switch considered in section 5. Now the metastable state is the non-trivial fixed point Q, and there is an absorbing boundary at the zero fixed point. We can decompose the probability density for all non-zero states as

$$P_{n,m}(t) = \sum_r \phi^{(r)}_{n,m} \mathrm{e}^{-\lambda^{(r)}t},$$

where the sum is over all non-zero eigenvalues, and

$$P_{0,0}(t) = 1 - \sum_{n,m}{}' P_{n,m}(t).$$

Here $'$ indicates that the sum excludes $(0,0)$. If the system is initially close to A and N is sufficiently large, then we expect extinction to be a rare event. Following from the analysis of section 5, we assume that the smallest non-zero eigenvalue is exponentially small with respect to the system size N, in contrast to all other non-zero eigenvalues. We thus have the quasistationary approximation for $(n,m) \neq (0,0)$,

$$P_{n,m}(t) \approx \Pi_{n,m} \mathrm{e}^{-\lambda^{(0)}t},$$

where $\Pi_{n,m} = \phi^{(0)}_{n,m}$. Choosing $\Pi_{n,m}$ to have unit normalization (with $\Pi_{0,0} = 0$), it follows that

$$P_{0,0} \approx 1 - \mathrm{e}^{-\lambda^{(0)}t}$$

and we can identify $1/\lambda^{(0)}$ as the mean time to extinction [15]. Moreover, equation (6.19) implies that

$$\lambda^{(0)} = \Pi_{1,0}. \tag{6.20}$$

The quasistationary density can be calculated using a WKB approximation along similar lines to section 5.1. Consider the WKB ansatz

$$\Pi_{n,m} = \mathrm{e}^{-N\Phi(x,y)} \tag{6.21}$$

with $x = n/N, y = m/N$ treated as continuous variables. It follows that

$$\lambda^{(0,0)} \approx \mathrm{e}^{-N\Phi(0,0)}. \tag{6.22}$$

Substituting the WKB ansatz into the quasistationary equation $\widehat{A}\Pi_{n,m} = 0$ gives, to leading order in $1/N$, the zero-energy Hamilton-Jacobi equation [168]

$$H(x,y,\partial_x\Phi,\partial_y\Phi) = 0, \tag{6.23}$$

with effective Hamiltonian

$$\begin{aligned} H(x,y,p_x,p_y) \quad &= Bx(1-x)(\mathrm{e}^{p_x}-1) + x(\mathrm{e}^{-p_x}-1) + \alpha x(\mathrm{e}^{-p_x+p_y}-1) \\ &\quad + \beta y(\mathrm{e}^{p_x-p_y}-1) \end{aligned} \tag{6.24}$$

Hamilton's equation are given by

$$\dot{x} = Bx(1-x)\mathrm{e}^{p_x} - x\mathrm{e}^{-p_x} - \alpha x\mathrm{e}^{-p_x+p_y} + \beta y\mathrm{e}^{p_x-p_y} \tag{6.25a}$$

$$\dot{y} = \alpha x\mathrm{e}^{-p_x+p_y} - \beta y\mathrm{e}^{p_x-p_y} \tag{6.25b}$$

$$\dot{p}_x = -B(1-2x)(\mathrm{e}^{p_x}-1) - (\mathrm{e}^{-p_x}-1) - \alpha(\mathrm{e}^{-p_x+p_y}-1) \tag{6.25c}$$

$$\dot{p}_y = -\beta(\mathrm{e}^{p_x-p_y}-1). \tag{6.25d}$$

As in the case of the birth-death process in section 5.1, the zero-energy dynamics in the invariant plane $p_x = p_y = 0$ recovers the rate equations (6.17a) and (6.17b), after rescaling by N. Thus, the two saddles $\mathscr{O} = (0,0,0,0)$ and $\mathscr{Q} = (n_*/N, m_*/N, 0, 0)$ in the four-dimensional phase-plane originate from the fixed points of the two-dimensional deterministic system. There is an additional saddle at $\mathscr{P} = (0, 0, -\ln B, -\ln B)$, which corresponds to the fluctuational extinction point [15, 85, 168].

We now have to determine the appropriate zero-energy path along which to evaluate the action of the Hamiltonian system, which we then identify with the quasipotential (see section 5.1). Given an established population at $t \to -\infty$, the trajectory starts at the saddle $\mathscr{Q}$. The issue is then whether it terminates at $\mathscr{O}$ or $\mathscr{P}$, both of which represent extinction of the two populations. It can be shown that the correct choice is the fluctuational extinction point $\mathscr{P}$ [82, 152], so that the optimal fluctuational path (also known as an instanton) is the heteroclinic connection $\mathscr{C}$ from $\mathscr{Q}$ to $\mathscr{P}$. Hence†,

$$\Phi(0,0) = \int_{\mathscr{C}} [p_x dx + p_y dy]. \tag{6.26}$$

In contrast to the one-degree-of-freedom model of a genetic switch (single protein concentration) with a corresponding two-dimensional Hamiltonian phase space (section 5.1), the extinction model has two-degrees-of-freedom (population densities of two phenotypes) and a four-dimensional phase-space. Since there is only one independent integral of motion (the energy), it follows that in general the instanton solution has to be calculated numerically. It is also possible to carry out a multi-scale perturbation analysis when $B \approx 1$, where all of the fixed points approach each other (due to the fact that for $B < 1$ the deterministic system also becomes extinct). Analytical results can be obtained in both the fast and slow switching limits. Moreover, it is possible to analyze the effects of a catastrophe (temporary reduction in the growth rate B) on the population extinction risk. This issue was previously explored in Ref. [17] for a single population of normals, which is obtained in the fast switching limit $\alpha, \beta \to \infty$ (see below). The main finding of the two-population model is that the presence of a persister subpopulation dramatically reduces the increase in extinction probability due to the same type of catastrophe [168]. Moreover, the reduction is significantly greater in the slow phenotypic switching regime.

One-population model. In spite of the limitations of the one-population model with regards the response to a catastrophe, it is useful to review the analysis of the simpler model [17], since the same basic ideas extend to the more complex model. In the fast

† It is important to note that the assumption of large system size ($n, m \gg 1$) implicit in the WKB analysis breaks down around $x = y = 0$. This means that one has to introduce a boundary layer of width $\sim 1/N$ around the origin. However, since this yields a subleading contribution in N^{-1} to the mean extinction time [18], it can be ignored (see also [129]).

switching limit $\alpha, \beta \to \infty$ we can eliminate the variables y, p_y by taking $y \to \alpha x/\beta$ and $p_y \to p_x$. Hamilton's equations for x, p_x become

$$\dot{x} = Bx(1-x)\mathrm{e}^{p_x} - x\mathrm{e}^{-p_x} \tag{6.27a}$$

$$\dot{p}_x = -B(1-2x)(\mathrm{e}^{p_x}-1) - (\mathrm{e}^{-p_x}-1). \tag{6.27b}$$

and the associated Hamiltonian is

$$H(x, p_x) = Bx(1-x)(\mathrm{e}^{p_x}-1) + x(\mathrm{e}^{-p_x}-1) \tag{6.28}$$

The corresponding phase plane is shown in Fig. 25. We now have an explicit equation for the heteroclinic connection from $\mathscr{P} = (1 - B^{-1}, 0)$ to $\mathscr{P} = (0, -\ln B)$:

$$x = x(p) = 1 - \frac{\mathrm{e}^{-p_x}}{B}.$$

After integration by parts, we have

$$\Phi(0,0) = -\int_0^{-\ln B} x(p)dp = \frac{1-B}{B} + \ln B. \tag{6.29}$$

In the regime $0 < \delta \ll 1$ with $\delta = B - 1$, we have $\Phi(0,0) \approx \delta^2/2$ and the mean extinction time $\tau_e = 1/\lambda^{(0)}$ is

$$\tau_e \sim \mathrm{e}^{N\delta^2/2}. \tag{6.30}$$

For comparison, note that when persisters are also taken into account and phenotypic switching is slow, the extinction time is increased according to [168]

$$\tau_e \sim \mathrm{e}^{N\delta^2(1/2+\alpha/\beta)}. \tag{6.31}$$

That is, spending more time in the persister state delays extinction.

Effect of a catastrophe. Now suppose that an environmental catastrophe occurs, in which either the reproduction rate or the carrying capacity temporarily reduces [17]. The latter can be modeled by taking $Bx(1-x) \to Bx(f(t)-x)$ with a time-dependent factor such that $f(\pm\infty) = 1$. Suppose for the sake of illustration that

$$f(t) = \begin{cases} 1 & \text{if } t < 0 \text{ or } t > T \\ 0 & \text{if } 0 < t < T \end{cases} \tag{6.32}$$

That is, the carrying capacity drops instantaneously to zero at $t = 0$ and subsequently jumps back to the original value at $t = T$. The effective Hamiltonian system now switches between the unperturbed Hamiltonian (6.28) and the perturbed Hamiltonian

$$\widehat{H}(x, p_x) = -Bx^2(\mathrm{e}^{p_x}-1) + x(\mathrm{e}^{-p_x}-1). \tag{6.33}$$

Each of the Hamiltonians is an integral of the motion during the associated time interval. Hence the maximum likelihood fluctuational path to extinction can be determined by matching three separate trajectories corresponding to the pre-catastrophe, catastrophe and post-catastrophe stages. The trajectories for $t \in (-\infty, 0)$ and $t \in (T, \infty)$ are zero energy and thus belong to the original heteroclinic connection from $\mathscr{Q}$ to $\mathscr{P}$. On the other hand, the catastrophe path for $t \in [0, T]$ has some non-zero energy E_c, which is unknown *a priori*. In particular, the points ρ_1, ρ_2 where the non-zero energy path intersects the heteroclinic connection depend on E_c. The full solution is obtained by imposing continuity of $x(t)$ and $p(t)$ at the intersection points

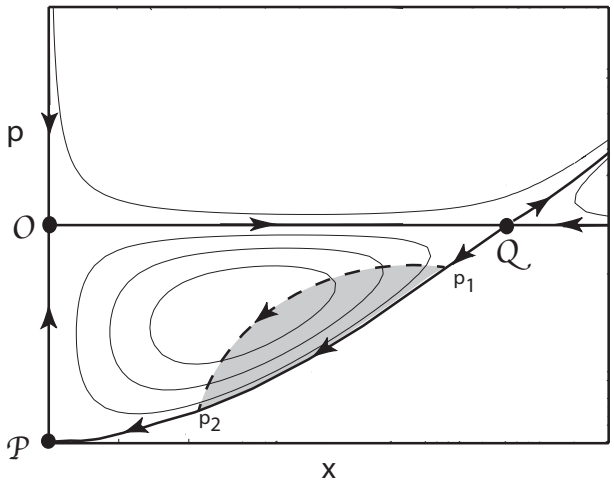


Figure 25. Schematic diagram indicating the phase-plane (x, p) for the one-population model of extinction. The fixed points are the trivial state $\mathscr{O}$, the metastable state $\mathscr{Q}$ and the fluctuational fixed point $\mathscr{P}$. The zero-energy trajectories are indicated by thick solid lines, including the optimal fluctuational path given by the heteroclinic connection $\mathscr{Q}\mathscr{P}$. The dashed thick line indicates a fluctuational path segment with non-zero energy E_c. This arises in response to a catastrophe consisting of the instantaneous reduction of the carrying capacity to zero at $t = 0$ and full recovery at $t = T$ (see text for details).

and imposing the condition that the duration of the catastrophe is T. One thus finds that the modified quasipotential or action is

$$\widehat{\Phi}(0,0) = \Phi(0,0) - \Delta A - E_c(T)T, \tag{6.34}$$

where ΔA is the shaded region in Fig. 25. The reduction in the action leads to an exponential decrease in the mean extinction time. However, this is mitigated by the presence of a persister subpopulation, as shown in Ref [168].

In the above example, the effect of a catastrophe on population extinction was explored by introducing a time-dependent reaction rate into the WKB Hamiltonian and seeing how it effects the optimal fluctuational path to the absorbing state. This is a special case of a more general theory of population extinction in the presence of time-dependent reaction rates that has been developed in a series of papers [16, 19, 20, 23, 89, 146, 216]. Such rates represent the effects of colored environmental or extrinsic noise or some periodic modulation of the environment.)

7. Quorum sensing

In this section, we consider another example of switching at the multicellular level, namely, quorum sensing. This refers to the switching on or off of gene regulatory networks within each cell depending on the cell density within the environment. Most models of bacterial quorum sensing are based on deterministic ordinary differential equations (ODEs), in which both the individual cells and the extracellular medium are treated as well-mixed compartments (fast diffusion limit) [13, 62, 75, 104, 117, 142, 179, 182, 248]. (For a discussion of spatial models that take into account bulk diffusion of the autoinducer in the extracellular domain see Refs. [75,118,154,184,185].) Suppose

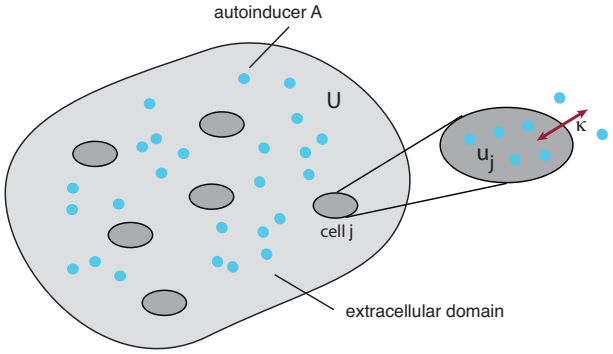


Figure 26. Schematic diagram showing a population of cells labeled $j = 1, \ldots, N$ that can exchange an autodinducer A with the extracellular environment. The extracellular concentration of A is denoted by U and the intracellular concentration of A in the j-th cell is denoted by u_j. The diffusive coupling parameter is κ. All compartments are taken to be well-mixed.

that there are N cells labeled $i = 1, \ldots, N$. Let $U(t)$ denote the concentration of signaling molecules in the extracellular space and let u_i be the corresponding intracellular concentration within the i-th cell, see Fig. 26. Suppose that there are K other chemical species within each cell, which together with the signaling molecule comprise a regulatory network. Introduce the vector $\mathbf{v}_i = (v_{i,1}, \ldots, v_{i,K})$ with $v_{i,k}$ the concentration of species k within the i-th cell. A deterministic model of quorum sensing may be expressed in the general form [182]

$$\frac{du_i}{dt} = F(u_i, \mathbf{v}_i) - \kappa(u_i - U), \tag{7.1a}$$

$$\frac{dv_{i,k}}{dt} = G_k(u_i, \mathbf{v}_i), \tag{7.1b}$$

$$\frac{dU}{dt} = \frac{\alpha}{N} \sum_{j=1}^{N} \kappa(u_j - U) - \gamma U, \tag{7.1c}$$

The nonlinear terms $F(u, \mathbf{v})$ and $G_k(u, \mathbf{v})$ are the reaction rates of the regulatory network based on mass action kinetics, the term $\kappa(u_j - U)$ represents the diffusive exchange of signaling molecules across the membrane of the i-th cell, and γ is the rate of degradation of extracellular signaling molecules. Also $\alpha = V_{\rm cyt}/V_{\rm ext}$ is a cell density parameter equal to the ratio of the total cytosolic and extracellular volume. Note that $V_{\rm cyt} = v_{\rm cyt} N$, where $v_{\rm cyt}$ is the single-cell volume.

The global convergence or synchronization properties of quorum sensing networks, where coupling between nodes in the network is mediated by a common environmental variable, has been analyzed within the context of nonlinear dynamical systems in Ref. [218]. In particular, the authors derive conditions on the nonlinear function F and G_k for which solutions converge exponentially toward each other, that is $u_i(t) \to u(t)$ and $\mathrm{v}_i(t) \to \mathrm{v}(t)$ with

$$\frac{du}{dt} = F(u, \mathbf{v}) - \kappa(u - U), \tag{7.2a}$$

$$\frac{dv_k}{dt} = G_k(u, \mathbf{v}), \tag{7.2b}$$

$$\frac{dU}{dt} = \alpha\kappa(u - U) - \gamma U. \tag{7.2c}$$

From a dynamical systems perspective, two basic forms of collective behavior are typically exhibited by equations (7.2*a*-7.2*c*): either the population acts as a biochemical switch [75, 136, 142, 248] or as a synchronized biochemical oscillator [62,179,182]. Two distinct mechanisms for a biochemical switch have been identified. The first mechanism is bistability in a gene regulatory network (see section 3). One well-known example is a mathematical model of quorum sensing in the bacterium *Pseudomonas aeruginosa* [75]. *P. aeruginosa* is a human pathogen that only releases toxins once the bacterial colony has reached a critical size so that it can resist being neutralized by an efficient host response [71,80]. Multistability also occurs in a related model of quorum sensing in the bioluminescent bacteria *V. fisheri* [142]. In this system, quorum sensing limits the production of bioluminescent luciferase to situations where cell populations are large; this saves energy since the signal from a small number of cells would be invisible and thus useless. Recent experimental studies of quorum sensing in the bacterial species *V. harveyi* and *V. cholerae* [163, 233, 252] provide evidence for an alternative switching mechanism, which can provide robust switch-like behavior without bistability. In these quorum sensing systems two or more parallel signaling pathways control a gene regulatory network via a cascade of phosphorylation-dephosphorylation cycles (PdPCs). Within the context of quorum sensing, the binding of an autodinducer to its cognate receptor switches the receptor from acting like a kinase to one acting like a phosphotase. Thus the PdPCs are driven by the level of autoinducer, which itself depends on the cell density. One source of the switch-like behavior is thus *ultrasenstivity* of the PdPCs [34,107,113,208–210]. In contrast to the noise-driven phenotypic switching considered in section 6, the switching mechanism in quorum sensing is now stimulus-driven (based on cell density), and thus noise tends to have a detrimental effect.

7.1. Bistability in a model of Pseudomonas aeruginosa *quorum sensing*

In *P. aeruginosa* there are two quorum-sensing systems working in series, known as the *las* and *rhl* system, respectively. Following Ref. [75], we consider the upstream *las* system, which is composed of *lasI*, the autoinducer synthase gene responsible for synthesis of the autoinducer 3-oxo-C12-HSL via the enzymatic action of the protein LasI, and the *lasR* gene that codes for transcriptional activator protein LasR (R), see Fig. 27. Positive feedback occurs due to the fact that LasR and 3-oxo-C12-HSL can form a dimer, which promotes both *lasR* and *lasI* activity. (We ignore an additional

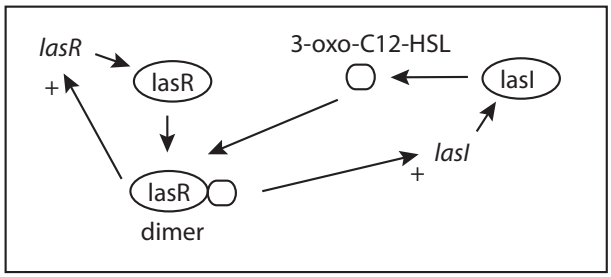


Figure 27. Simplified regulatory network for the *las* system in *P. aerginosa.*

negative feedback loop that is thought to play a relatively minor role in quorum sensing.) Exploiting the fact that the lifetime of each type of mRNA is much shorter than its corresponding protein, we can eliminate the mRNA dynamics, and write down a system of ODES for the concentrations of LasR, the dimer LasR/3-oxo-C12-HSL and the autoinducer 3-oxo-C12-HSL, which we denote by R, P and A, respectively. The resulting system of ODES for mass action kinetics at the single-cell level thus take the form [75]

$$\frac{dP}{dt} = k_{RA}RA - k_P P, \tag{7.3a}$$

$$\frac{dR}{dt} = -k_{RA}RA + k_P P - k_R R + V_R \frac{P}{K_R + P} + R_0, \tag{7.3b}$$

$$\frac{dA}{dt} = -k_{RA}RA + k_P P + V_A \frac{P}{K_A + P} + A_0 - k_A A. \tag{7.3c}$$

Here k_A, k_R are the rates of degradation of A, R, k_{RA}, k_P are the rates of production and degradation of the dimer P, and A_0, R_0 are baseline rates of production of A, R. Finally, the positive feedback arising from the role of the dimer P as an activator protein that enhances the production of LasR and arising from the activation of R and A (with the latter mediated by lasI) is taken to have Michaelis-Menten form. Dockery and Keener [75] carry out a further reduction by noting that formation and degradation of dimer is much faster than transcription and translation so that we can assume P is in quasi steady-state so that

$$P = \frac{k_{RA}}{k_P} RA = kRA.$$

Then we have

$$\frac{dR}{dt} = -k_R R + V_R \frac{kRA}{K_R + kRA} + R_0, \tag{7.4a}$$

$$\frac{dA}{dt} = V_A \frac{kRA}{K_A + kRA} + A_0 - k_A A. \tag{7.4b}$$

Comparison of equations (7.4a) and (7.4b) with the general kinetic system in equations (7.2a) and 7.2b), shows that we have one auxiliary species and we can make the following identifications (after dropping the $k = 1$ index): $u = A, v = R$ with

$$F = V_A \frac{uv}{K_A + kuv} + A_0 - k_A u.$$

$$G = -k_R v + V_R \frac{kuv}{K_R + kuv} + R_0.$$

Now suppose equation (7.4b) is diffusively coupled to an extracellular concentration along the lines of (7.2a-7.2c), and as a further simplification take U to be in quasi-equilibrium. The latter conditions allows us to carry out a phase-plane analysis without changing the essential behavior of the system. Set $\alpha = \rho/(1-\rho)$ and $\kappa = \delta/\rho$, where ρ is the volume fraction of cells and the conductance δ is independent of ρ. The only modification is that the last term on the right-hand side of (7.4b) which is transformed according to the scheme

$$k_A A \to d(\rho) \equiv \left[k_A + \frac{\delta}{\rho} \frac{\gamma}{\gamma + \delta/(1-\rho)} \right] A.$$

Quorum sensing thus arises due to the dependence of the effective degradation rate $d(\rho)$ on ρ. More specifically, when ρ is small the rate $d(\rho)$ is large, whereas $d(\rho)$ is

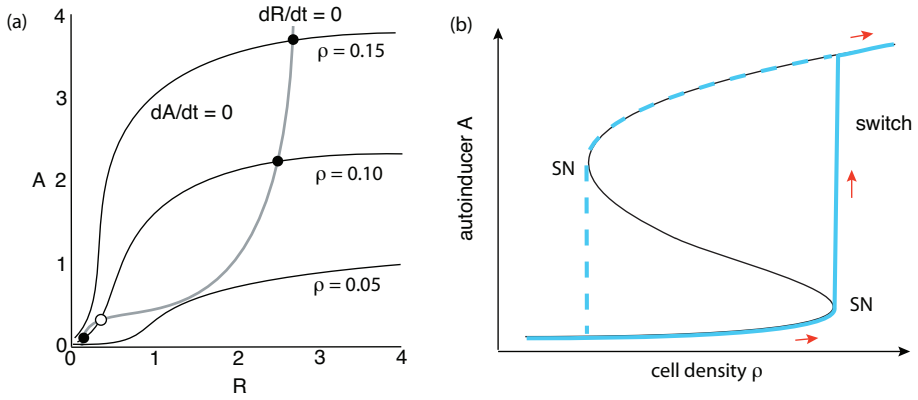


Figure 28. Bistability in planar model of *las* system in *P. aerginosa.*. (a) Nullcline $\dot{R} = 0$ is shown by the gray curve and the ρ-dependent nullclines $\dot{A} = 0$ are shown by the black curves for three different values of ρ. It can be seen that for intermediate ρ values there are two stable fixed points separated by an unstable fixed point. (b) The appearance and subsequent disappearance of a stable/unstable pair of fixed points via saddle-node bifurcations can switch the system from a low activity state to a high activity state. Decreasing the density again leads to hysteresis. Parameter values are $V_R = 2.0$, $V_A = 2.0$, $K_R = 1.0$, $K_A = 1.0$, $R_0 = 0.05$, $A_0 = 0.05$, $\delta = 0.2$, $\gamma = 0.1$, $k_R = 0.7$, and $k_A = 0.02$.

small when $\rho \to 1$. In Fig. 28(a), we plot nullclines of the system at various values of ρ, illustrating that bistability occurs at intermediate values of ρ. In this parameter regime the system acts like a bistable switch, where one stable state has low levels of autoinducer and the other high levels of autoinducer [75]. Switching occurs due to a saddle-node bifurcation as illustrated in Fig. 28(b). When noise is included, the resulting steady-state distribution exhibits peaks around the stable fixed points of the deterministic system, and the transition between a unimodal and a bimodal distribution is used as one indictator of a stochastic bifurcation [108, 231].

7.2. Ultrasensitivity in a model of V. harveyi *quorum sensing*

The bioluminescent bacterium *V. harveyi* has three parallel quorum sensing systems, each consisting of a distinct autoinducer (HAI-1, AI-2, CAI-1), cognate receptor (LuxN, LuxPQ, CqsS), and associated enzyme that helps produce the autoinducer, see Fig. 29. (The human pathogen *V. cholerae* has a similar quorum sensing network, except that there could be up to four parallel pathways [144].) Each autoinducer is freely exchanged between the intracellular and extracellular domains. Extracellular diffusion at low (high) cell densities leads to small (large) intracellular autoinducer concentrations, resulting in a low (high) probability that the autoinducer can bind to its cognate receptor. The receptors target downstream DNA-binding regulatory proteins LuxU and LuxO. At low cell densities the receptors act as kinases and the resulting phosphorylation of LuxU and LuxO activates transcription of the genes encoding five regulatory *small non-coding RNAs* (sRNAs) termed Qrr1-Qrr5, which destabilize the transcriptional activator protein LuxR. This prevents the activation of target genes responsible for the production of various proteins, including bioluminescent luciferase. Hence, at low cell densities the bacteria do not bioluminesce. On the other hand, at high cell densities the receptors switch from kinases to

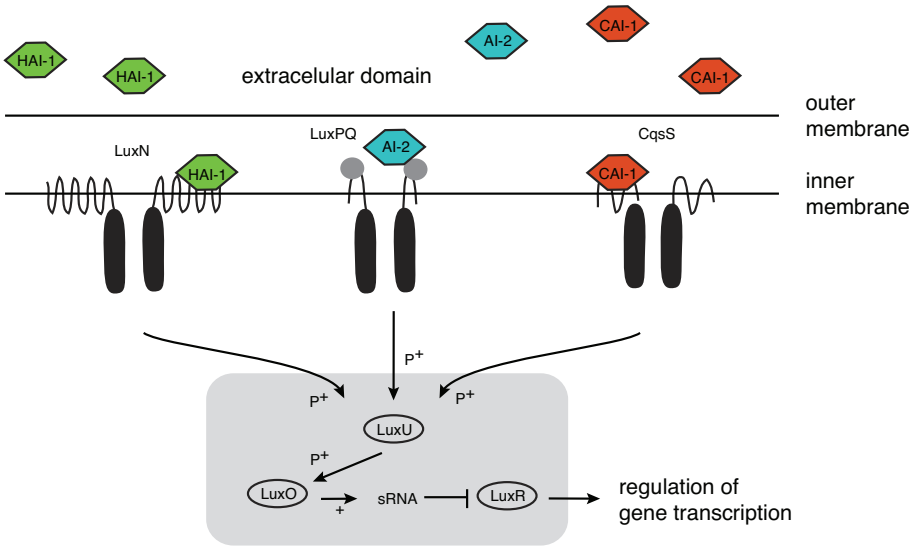


Figure 29. Summary diagram of the *V. harveyi* quorum sensing circuit. Three phosphorylation cascades work in parallel to control the ratio of LuxO to LuxO-P based on local cell-population density. Five sRNA, qrr1-5, then regulate expression of quorum sensing target genes including the master transcriptional regulator LuxR, which upregulates downstream factors.

phosphotases, significantly reducing the levels of LuxU-P and LuxO-P. The sRNAs are thus no longer expressed, allowing the synthesis of LuxR and the expression of bioluminescence, for example. Both the phosphorylation-dephosphorylation cascades and the sRNA regulatory network provide a basis for a sharp, sigmoidal response of the concentration of LuxR to smooth changes in cell density.

Following Ref. [55], we analyze the occurrence of ultrasensitivity at the single cell level by focusing on a single phosphorylation pathway and adapting the Goldbeter-Koshland of phosphorylation-dephosphorylation cycles (PdPCs) [113]. (For a corresponding model of switching due to the action of sRNAs see Hunter et al. [136].) In particular, we consider the phosphorylation-dephosphorylation of LuxU by the enzymatic action of a particular quorum sensing receptor, which is denoted by R when acting as a kinase and by $\widehat{R}$ when it is it is bound by an autoinducer (A) and acts like a phosphotase. Denoting the protein LuxU by W, we define the following reaction schemes:

$$W + R \underset{d_1}{\overset{a_1}{\rightleftharpoons}} WR \overset{k_1}{\to} W^* + R \tag{7.5a}$$

$$W^* + \widehat{R} \underset{d_2}{\overset{a_2}{\rightleftharpoons}} W^*\widehat{R} \overset{k_2}{\to} W + \widehat{R}, \tag{7.5b}$$

$$R + A \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \widehat{R}. \tag{7.5c}$$

For simplicity, we assume that both the phosphorylation and the dephosphorylation steps are irreversible. (See the work of Qian and collaborators for an analysis of more detailed, reversible models of PdPCs [107, 208–210].) Introducing the concentrations $u = [A]$, $w = [W]$, $w^* = [W^*]$, $r = [R]$, $\widehat{r} = [\widehat{R}]$, $v = [WR]$ and $v^* = [W^*\widehat{R}]$, the corresponding kinetic equations for a single cell with a fixed intracellular concentration

of autoinducer, are

$$\frac{dw}{dt} = -a_1 w(r-v) + d_1 v + k_2 v^* \tag{7.6a}$$

$$\frac{dv}{dt} = a_1 w(r-v) - (d_1 + k_1)v \tag{7.6b}$$

$$\frac{dw^*}{dt} = -a_2 w^*(\widehat{r} - v^*) + d_2 v^* + k_1 v \tag{7.6c}$$

$$\frac{dv^*}{dt} = a_2 w^*(\widehat{r} - v^*) - (d_2 + k_2)v^* \tag{7.6d}$$

$$\frac{dr}{dt} = k_- \widehat{r} - k_+ u r. \tag{7.6e}$$

These are supplemented by the conservation equations

$$W_T = w + w^* + v + v^*, \quad R_T = r + \widehat{r}, \tag{7.7}$$

where R_T is the total concentration of receptors and W_T is the total concentration of LuxU.

For simplicity, we will assume that the conversion of the receptors from kinase to phosphotase activity is independent of the PdPC. That is, we ignore any positive feedback pathways, in which the regulation of gene expression by the phosphorylation/dephsophorylation of LuxU alters the production of the autoinducer [196]. This allows us to treat the receptor-ligand dynamics given by Eq. (7.6*e*), or subsequent extensions, independently of the PdPC dynamics given by Eqs. (7.6*a*-7.6*d*). In particular, for fixed concentration u, we can take the concentration of kinases and phosphotases to be at equilibrium:

$$r_{\rm eq} = \frac{k_-}{k_+ u + k_-} R_T \equiv R(u), \quad \widehat{r}_{\rm eq} = R_T - R(u). \tag{7.8}$$

The system of Eqs. (7.6*a*-7.6*e*) then reduces to the classical Goldbeter-Koshland model of PdPCs [113, 157], and we can apply their analysis based on a generalization of Michaelis-Menten kinetics. The first step is to assume that the concentration of W and W^* is much larger than that of the receptor, that is, $W_T \gg R_T$ or equivalently $W_T = w + w^*$. This implies that the time scale for the dynamics of the complexes WR and $W^*\widehat{R}$ is much faster than that for the dynamics of W and W^*. Performing a separation of time-scales, we can treat the concentrations w and w^* as constants when analyzing equations (7.6*b*) and (7.6*b*) , while we can take the steady-state values of the concentrations v, v^* when solving equations. (7.6*a*) and (7.6*c*) . Hence, setting $dv/dt = 0$ and $r = R(u)$ in (7.6*b*) we can solve for v in terms of w. Similarly, setting $dv^*/dt = 0$ and $\widehat{r} = R_T - R(u)$ in (7.6*d*) we can solve for v^* in terms of w^*. We thus obtain the reduced kinetic scheme

$$W \underset{f_2(w^*)}{\overset{f_1(w)}{\rightleftharpoons}} W^*, \tag{7.9}$$

with

$$f_1(w) = \frac{k_1 R(u) w}{K_1 + w}, \quad f_2(w^*) = \frac{k_2 [R_T - R(u)] w^*}{K_2 + w^*}, \tag{7.10}$$

and

$$K_1 = \frac{d_1 + k_1}{a_1}, \; K_2 = \frac{d_2 + k_2}{a_2}.$$

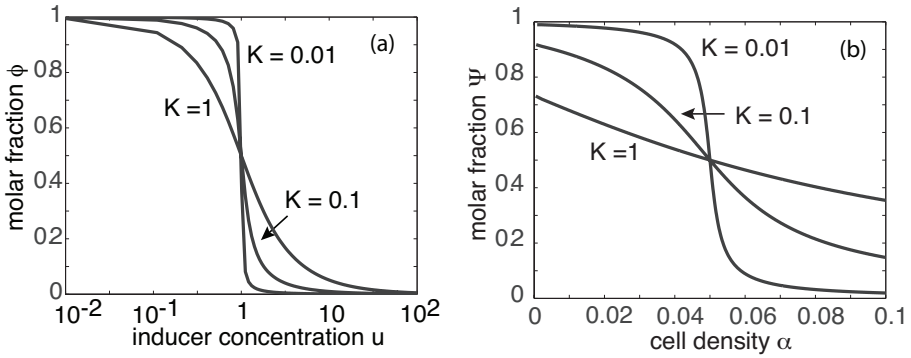


Figure 30. Molar fraction of modified protein W^* at steady-state as a function of (a) the autoinducer concentration u and (b) cell density α for different values of K, with $K = K_1 = K_2$, $k_- k_1 = k_+ k_2$, $\Gamma = 1$, $\kappa = 10$, and $\alpha_c = 0.05$.

Imposing the conservation condition $W_T = w + w^*$ thus yields the single independent kinetic equation

$$\frac{dw^*}{dt} = f_1(W_T - w^*) - f_2(w^*). \tag{7.11}$$

The steady-state concentration $w^*_{\rm eq}$ of LuxU-P is thus obtained by solving $f_1(W_T - w^*) = f_2(w^*)$, which yields a quadratic equation for w^*. Taking the positive root, and expressing it as a function of the fixed autoinducer concentration u, we find that

$$\frac{[\text{LuxU-P}]}{[\text{LuxU-P}]+[\text{LuxU}]} = \phi(V(u)), \tag{7.12}$$

with

$$V(u) = \frac{k_1 R(u)}{k_2[R_T - R(u)]} = \frac{k_-}{k_+ u}\frac{k_1}{k_2} \tag{7.13}$$

and ϕ given by

$$\phi = \frac{-B + \sqrt{B^2 + 4AC}}{2AW_T}, \tag{7.14}$$

for $V \neq 1$, where

$$A = V(u) - 1, \quad B = \frac{1}{W_T}\left(K_1 + K_2 V(u)\right) - \left(V(u) - 1\right), \quad C = \frac{K_2}{W_T} V(u).$$

A plot of ϕ, as a function of the autoinducer concentration u is shown in Fig. 30(a) for $K_1 = K_2 = K$. At low values of K, there is a sharp change from high to low levels of modified protein over a very small change in u (ultrasensitivity); this corresponds to a regime in which the two enzymes are saturated. On the other hand, for large values of K, the curve is relatively shallow, and one obtains a response similar to first-order kinetics.

Now consider a population of N identical cells that are coupled via a common extracellular domain due to the transfer of the autoinducer A across the cell membrane.

Eq. (7.6e) for a single cell is then replaced by a system of equations of the form

$$\frac{du_i}{dt} = \Gamma + k_-(R_T - r_i) - k_+ u_i r_i - \kappa(u_i - U), \tag{7.15a}$$

$$\frac{dr_i}{dt} = k_-(R_T - r_i) - k_+ u_i r_i \quad i = 1, \ldots, N \tag{7.15b}$$

$$\frac{dU}{dt} = \alpha\kappa(u_{\rm av} - U) - \gamma U, \tag{7.15c}$$

where

$$u_{\rm av} = \frac{1}{N}\sum_{j=1}^{N} u_j, \tag{7.16}$$

is the population-averaged intracellular concentration of A, U is the extracellular concentration of A, and Γ is the rate of production of A due to the action of enzymes.

Suppose that the above system is globally convergent with $u_i(t) \to u(t)$ and $r_i(t) \to r(t)$ in the limit $t \to \infty$, with (u, r) evolving according to the effective single-cell model

$$\frac{du}{dt} = \Gamma + k_-[R_T - r] - k_+ u r - \kappa(u - U) \tag{7.17a}$$

$$\frac{dr}{dt} = k_-[R_T - r] - k_+ u r \tag{7.17b}$$

$$\frac{dU}{dt} = \alpha\kappa(u - U) - \gamma U. \tag{7.17c}$$

(Conditions for global convergence have been derived elsewhere [55].) Equations (7.17a)-(7.17c) have a unique stable fixed point $u_{\rm eq}$ with

$$u_{\rm eq} = \Gamma\frac{(\alpha\kappa + \gamma)}{\kappa\gamma} \equiv \psi(\alpha). \tag{7.18}$$

Putting $u = u_{\rm eq}$ in equation (7.12) finally shows that

$$\frac{[\text{LuxU-P}]}{[\text{LuxU-P}]+[\text{LuxU}]} = \phi\left(\frac{k_-}{k_+\psi(\alpha)}\frac{k_1}{k_2}\right) \equiv \Psi(\alpha), \tag{7.19}$$

In order that the system exhibit switch-like behavior as a function of cell density α, we require a critical value α_c, $0 < \alpha_c < \infty$ such that (for $\Gamma = 1$)

$$\chi \equiv \frac{k_-}{k_+}\frac{k_1}{k_2} = \frac{\alpha_c}{\gamma} + \frac{1}{\kappa},$$

that is $\alpha_c = \gamma(\chi - \kappa^{-1})$. Assuming that $\chi = 1$ and taking $\alpha_c = 0.05$ [182], we require $\kappa > 1$ and $\gamma = \gamma_c = 0.05(1 - \kappa^{-1})$. For low cell densities ($\alpha \ll \alpha_c$) we have $\Psi(\alpha) \approx 1$, which follows from the functional form of ϕ, see Fig. 30(a). Hence the fraction of phosphorylated LuxU-P is high, which ultimately means that the expression of the gene regulator protein LuxR is suppressed. On the other hand, for large cell densities ($\alpha \gg \alpha_c$) we find that $\Psi(\alpha) \approx 0$. Now the fraction of phosphorylated LuxU-P is small, allowing the expression of LuxR and downstream gene regulatory networks. The α-dependence is illustrated in Fig. 30(b) with α and u linearly related.

7.3. Noise amplification of intrinsic fluctuations in an ultrasensitive switch

One of the characteristic features of ultrasensitive biochemical signaling networks is that they tend to amplify noise [34, 107, 165, 227]. On the other hand, collective behavior at the population level, as exhibited by quorum sensing networks, can mitigate the effects of noise [218, 234]. The amplification of intrinsic fluctuations in the *V. harveyi* quorum sensing model has recently been investigated within the framework of linear response theory [55].

Linear response of single-cell model. First, consider the single-cell model with fixed concentration u of autoinducer. In the deterministic case, with r given by the equilibrium solution (7.8), the dynamics of LuxU is given by equation (7.11), which we write explicitly in the form

$$\frac{dw^*}{dt} = \mathscr{F}(w^*, r) \equiv \frac{k_1 r(W_T - w^*)}{K_1 + [W_T - w^*]} - \frac{k_2[R_T - r]w^*}{K_2 + w^*}. \tag{7.20}$$

Several studies have analyzed noise signal amplification in ultrasensitive signal transduction networks by considering stochastic versions of equation (7.20) [34, 107, 165, 227]. (A more detailed analysis would need to start from a stochastic version of the full system of equations (7.6*a*-7.6*e*), since $r(t)$ is now time-dependent.) Suppose that $r(t)$ undergoes small fluctuations about the equilibrium state $r_{\rm eq}$ of Eq. (7.8). This can be incorporated by replacing equation (7.6*e*) for fixed u by the Langevin equation

$$\frac{dr}{dt} = k_-(R_T - r) - k_+ ur + \sigma_0 \xi(t), \tag{7.21}$$

with $\xi(t)$ a white noise process,

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

and σ_0 a fixed noise intensity. One possible source of noise is fluctuations in the autoinducer concentration u (see below). Under the linear noise approximation, we set $r(t) = r_{\rm eq} + R(t)$ such that

$$\frac{dR}{dt} = -(k_- + k_+ u)R + \sigma_0 \xi(t). \tag{7.22}$$

Similarly, linearizing equation (7.20) about the equilibrium solution by setting $w^* = w^*_{\rm eq} + W$, $w^*_{\rm eq} = \phi(V(u))$, gives

$$\frac{dW}{dt} = -\beta_1 W + \beta_2 R, \tag{7.23}$$

with

$$\beta_1 = -\left.\frac{\partial \mathscr{F}}{\partial w^*}\right|_{\rm eq} = \frac{k_1 K_1 r_{\rm eq}}{[K_1 + [W_T - w^*_{\rm eq}]]^2} + \frac{k_2 K_2 [R_T - r_{\rm eq}]}{[K_2 + w^*_{\rm eq}]^2} \tag{7.24}$$

and

$$\beta_2 = \left.\frac{\partial \mathscr{F}}{\partial r}\right|_{\rm eq} = \frac{k_1(W_T - w^*_{\rm eq})}{K_1 + [W_T - w^*_{\rm eq}]} + \frac{k_2 w^*_{\rm eq}}{K_2 + w^*_{\rm eq}} \tag{7.25}$$

Note that the gain of the underlying deterministic system (7.20) at equilibrium is given by

$$g = \frac{\Delta W / w^*_{\rm eq}}{\Delta R / r_{\rm eq}} = \frac{\beta_2}{\beta_1} \frac{r_{\rm eq}}{w^*_{\rm eq}}. \tag{7.26}$$

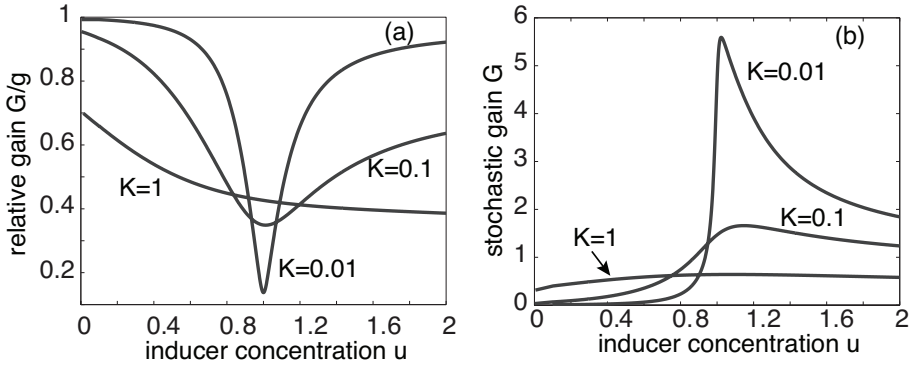


Figure 31. Plot of (a) relative gain G/g and (b) stochastic gain G as a function of autoinducer concentration u for different values of K, $K = K_1 = K_2$. Other parameter values are $k_- = k_+ = k_1 = k_2 = 1$, and $R_T = W_T = 1$.

Incorporating the u-dependence of $r_{\rm eq}$ and $\beta_{1,2}$ one can determine $g = g(u)$ and show that there is a sharp gain around the critical concentration $u = 1$. (Note that the units of autoinducer concentration are fixed by taking the PdPC switch to occur at $u = 1$. Typical intracellular molar concentrations are of order nM.)

In order to characterize noise amplification in the system, one can Fourier transform the linear equations (7.22) and (7.23) with

$$R(\omega) = \int_{-\infty}^{\infty} {\rm e}^{i\omega t} R(t) dt$$

etc.. This gives

$$W(\omega) = \frac{\beta_2}{\beta_1 + i\omega} R(\omega), \quad R(\omega) = \frac{\sigma_0}{k_- + k_+ u + i\omega} \xi(\omega),$$

where $\xi(\omega)$ is the Fourier transform of a white noise process with

$$\langle \xi(\omega) \rangle = 0, \quad \langle \xi(\omega) \overline{\xi(\omega')} \rangle = 2\pi \delta(\omega - \omega').$$

Using the Wiener-Khinchine theorem, the variance of the receptor concentration is given by the integral of the power spectrum defined by

$$2\pi S_R(\omega) \delta(\omega - \omega') = \langle R(\omega) \overline{R(\omega')} \rangle$$

That is,

$$\sigma_R^2 = \int_{-\infty}^{\infty} S_R(\omega) \frac{d\omega}{2\pi} = \int_{-\infty}^{\infty} \frac{\sigma_0^2}{[k_- + k_+ u]^2 + \omega^2} \frac{d\omega}{2\pi} = \frac{\sigma_0^2}{2[k_- + k_+ u]}. \tag{7.27}$$

Similarly, the variance of the Lux-P concentration is

$$\begin{aligned} \sigma_W^2 &= \int_{-\infty}^{\infty} S_W(\omega) \frac{d\omega}{2\pi} = \int_{-\infty}^{\infty} \frac{\beta_2^2}{\beta_1^2 + \omega^2} \frac{\sigma_0^2}{[k_- + k_+ u]^2 + \omega^2} \frac{d\omega}{2\pi} \\ &= \frac{\beta_2^2 \sigma_0^2}{2\beta_1 [k_- + k_+ u]} \frac{1}{k_- + k_+ u + \beta_1}. \end{aligned} \tag{7.28}$$

If we interpret $\sigma_W / w_{\rm eq}^*$ as the relative noise intensity of the output and $\sigma_R / r_{\rm eq}$ as the relative noise intensity of the output, then the noise amplification of the PdPC in

response to receptor fluctuations is defined by the stochastic gain [227]

$$G = \frac{\sigma_W/w^*_{\rm eq}}{\sigma_R/r_{\rm eq}} = \frac{r_{\rm eq}}{w^*_{\rm eq}}\sqrt{\frac{\beta_2^2}{\beta_1}\frac{1}{k_- + k_+ u + \beta_1}} = g\sqrt{\frac{\beta_1}{k_- + k_+ u + \beta_1}}, \tag{7.29}$$

where g is the deterministic gain (7.26). In Fig. 31 we plot the relative gain G/g and the stochastic gain G as a function of u. It can be seen that the sharp amplification around the critical density $u = 1$ in the ultrasensitive regime ($K = 0.01$) is suppressed relative to the deterministic gain.

Linear response of population model. In order to extend the analysis of noise amplification to the population model, it is necessary to explicitly model the stochastic dynamics of the autodinducer concentration. In a stochastic model of quorum sensing, it is no longer possible to identify the state of all the cells, even if the deterministic system (7.1*a*-7.1*c*) is globally convergent. Therefore, in the case of *V. harveyi*, we have to consider a stochastic version of equations (7.15*a*) and (7.15*b*)

$$\frac{du_i}{dt} = \Gamma + k_-(R_T - r_i) - k_+ u_i r_i - \kappa(u_i - U) + \theta_0 \xi_i(t), \tag{7.30a}$$

$$\frac{dr_i}{dt} = k_-(R_T - r_i) - k_+ u_i r_i, \tag{7.30b}$$

$$\frac{dU}{dt} = \alpha\kappa(u_{\rm av} - U) - \gamma U \tag{7.30c}$$

for $i = 1, \ldots, N$, with

$$\langle \xi_i(t) \rangle = 0, \quad \langle \xi_i(t)\xi_j(t') \rangle = \delta(t - t')\delta_{ij}.$$

We assume that each cell is driven by an independent white noise term with constant noise intensity θ_0; one source of the noise could be fluctuations in the production of the autoinducer. Linearizing about the global steady-state by setting

$$u_i(t) = u_{\rm eq} + \mathscr{U}_i(t),\ r_i(t) = r_{\rm eq} + R_i(t),\ U(t) = U_{\rm eq} + \mathscr{V}(t),$$

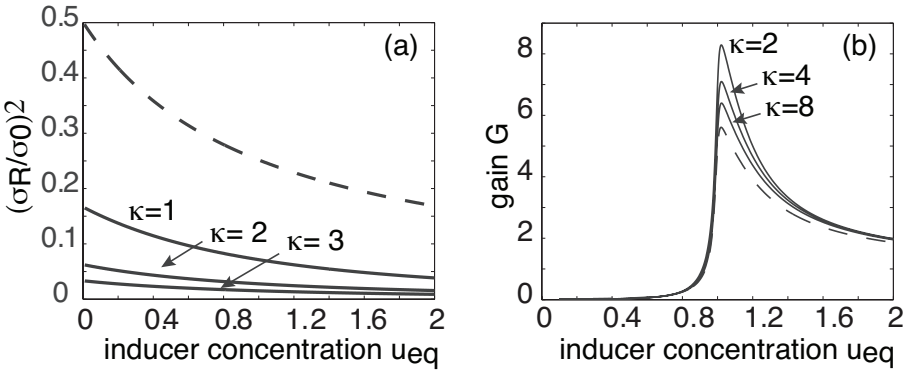


Figure 32. (a) Plot of variance σ_R^2 in receptor fluctuations as a function of autoinducer concentration $u_{\rm eq}$ for different values of κ. Also shown is the corresponding quantity in the single-cell model (dashed curve). (b) Plot of stochastic gain G of population model as a function of autoinducer concentration $u_{\rm eq}$ for different values of κ and $K = 0.01$. Other parameter values are $k_- = k_+ = k_1 = k_2 = 1$, and $R_T = W_T = 1$. Dashed curve is gain of single-cell model from Fig. 31(b).

yields

$$\frac{d\mathscr{U}_i}{dt} = -k_- R_i - k_+(\mathscr{U}_i r_{\rm eq} + u_{\rm eq} R_i) - \kappa(\mathscr{U}_i - \mathscr{V}) + \theta_0 \xi_i(t), \quad (7.31a)$$

$$\frac{dR_i}{dt} = -k_- R_i - k_+(\mathscr{U}_i r_{\rm eq} + u_{\rm eq} R_i), \quad (7.31b)$$

$$\frac{d\mathscr{V}}{dt} = \alpha\kappa(\mathscr{U}_{\rm av} - \mathscr{V}) - \gamma\mathscr{V}, \quad \mathscr{U}_{\rm av} = \frac{1}{N}\sum_{i=1}^{N} \mathscr{U}_i. \quad (7.31c)$$

These linear equation can be analyzed along similar lines to the single-cell case using Fourier transforms and contour integration [55]. Here, we simply state the final results (for large N). First, the variance in the receptor concentration r_i at the single cell level is

$$\sigma_R^2 = \frac{\sigma_0^2}{2z_-(z_+^2 - z_-^2)} - \frac{\sigma_0^2}{2z_+(z_+^2 - z_-^2)} = \frac{\sigma_0^2}{2z_+ z_-(z_+ + z_-)}, \quad (7.32)$$

where $\sigma_0 = k_+ r_{\rm eq}\theta_0/\Gamma$ with $\Gamma = 1$, and

$$z_\pm = \frac{1}{2}\left[(k_- + k_+(u_{\rm eq} + r_{\rm eq}) + \kappa) \pm \sqrt{(k_- + k_+(u_{\rm eq} + r_{\rm eq}) + \kappa)^2 - 4\kappa(k_- + k_+ u_{\rm eq})}\right].$$

The corresponding stochastic gain of individual cell PdPCs driven by the receptor fluctuations of the population model is

$$G = g\sqrt{\frac{\beta_1}{(\beta_1^2 - z_+^2)(\beta_1^2 - z_-^2)}}\sqrt{\beta_1[\beta_1^2 - (z_-^2 + z_- z_+ + z_+^2)] + (z_+ + z_-)z_+ z_-}, \quad (7.33)$$

One finds that $\sigma_R^2 \to 0$ as $\kappa \to \infty$, see Fig.32(a). On the other hand, expressing $r_{\rm eq}$ and $\beta_{1,2}$ in terms of $u_{\rm eq}$, one finds that the stochastic gain G is only weakly dependent on κ, and approaches the gain of the single-cell model as κ increases, see Fig. 32(b). Given that the receptor fluctuations σ_R^2 are suppressed in the quorum sensing model for large κ, while the gain is hardly changed, we conclude that the fluctuations in the LuxU-P concentration w^* are greatly suppressed compared to an isolated cell.

8. Other examples of biological switching processes

In this final section we describe some other common examples of stochastic switching models in biology that have similar mathematical structures to the genotypic and phenotypic examples of previous sections.

8.1. Stochastic ion channels and spontaneous action potentials

The generation and propagation of a neuronal action potential arises from nonlinearities associated with active membrane conductances [45]. 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 voltage x 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

$$I_{\text{con}} = \sum_s g_s(x - V_s), \tag{8.1}$$

where g_s is the conductance due to channels of type s and V_s is the corresponding reversal potential.

Recordings of the current flowing through single channels indicate that channels fluctuate rapidly between open and closed states in a stochastic fashion, thus acting as stochastic switches with respect to ion current flow, see Fig. 33. Consider a simple two-state model of an ion channel that can exist either in a closed state (C) or an open state (O). Transitions between the two states are governed by the continuous-time discrete Markov process

$$C(\text{closed}) \underset{\beta(x)}{\overset{\alpha(x)}{\rightleftharpoons}} O(\text{open}). \tag{8.2}$$

with transition rates $\alpha(x), \beta(x)$ depending on the membrane voltage x. For the moment, we assume that the voltage is fixed. Let $Z(t)$ be a discrete random variable taking values $Z \in \{C, O\}$ and set $P(t) = \text{Prob}\,[Z(t) = O] = 1 - \text{Prob}\,[Z(t) = C]$. The master equation for P is

$$\frac{dP}{dt} = \alpha(1 - P) - \beta P,$$

which has the unique stable steady-state (for fixed x) $P^*(x) = a(x) \equiv \alpha(x)/(\alpha(x) + \beta(x))$. In spite of stochasticity at the level of single channels, 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

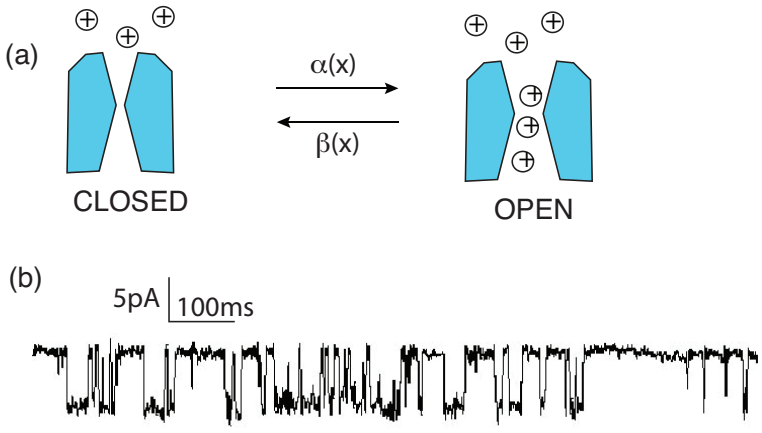


Figure 33. Ion channel as a stochastic switch. (a) Internal open and closed states of an ion channel. (b) Illustration of a typical time-course of a glycine receptor showing stochastic variations in current due to the opening and closing of the ion channel.

$g_s = \bar{g}_s P_s$ 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 with

$$\frac{dP_s}{dt} = -\alpha_s(1 - P_s) + \beta_s P_s. \tag{8.3}$$

A simple model of neural excitability is the Morris-Lecar model [183], which takes the form of a planar dynamical system

$$\frac{dx}{dt} = a(x) f_{\rm Na}(x) + w f_{\rm K}(x) - g(x) \equiv f(x, w) \tag{8.4a}$$

$$\frac{dw}{dt} = \frac{w_\infty(x) - w}{\tau_w(x)} \equiv \epsilon g(x, w), \tag{8.4b}$$

where w represents the fraction of open K^+ channels, $f_i(x) = g_i(v_i - x)$, $i = {\rm Na}, {\rm K}$, $g(x) = g_0(x - v_0)$ and

$$a(x) = \frac{\alpha_{\rm Na}(x)}{\alpha_{\rm Na}(x) + \beta_{\rm Na}(v)}, \quad w_\infty(x) = \frac{\alpha_{\rm K}(x)}{\alpha_{\rm K}(x) + \beta_{\rm K}(x)}, \quad \tau_w(x) = \frac{1}{\alpha_{\rm K}(x) + \beta_{\rm K}(x)}.$$

The fraction of Na^+ channels (or Ca^{2+} channels in the original formulation of the model) is assumed to be in quasi steady-state. The deterministic dynamics can be analyzed using a slow/fast analysis of the planar system, under the assumption that the dynamics of w is slow relative to that of x ($\epsilon \ll 1$). The fast variable x has a cubic-like nullcline (along which $\dot{x} = 0$) and the slow variable has a monotonically increasing nullcline (along which $\dot{w} = 0$), see Fig. 34. Suppose that the nullclines have a single intersection point at (x^*, w^*) on the left-hand branch of the x-nullcline. This corresponds to a stable fixed point of the system, which is identified with the resting state of the neuron. The neuron is said to be 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 - this is

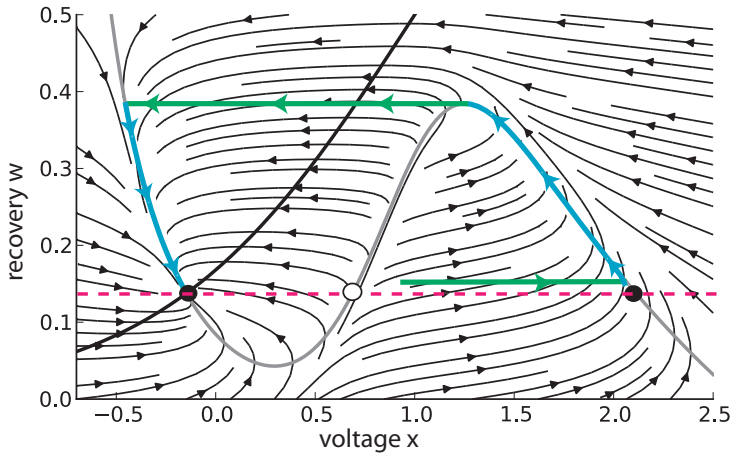


Figure 34. Deterministic phase plane dynamics. Thick curves show the nullclines: $\dot{x} = 0$ as grey and $\dot{w} = 0$ as black. Black stream lines represent deterministic trajectories. Green/blue curves represent an action potential trajectory in the limit of slow w. If the slow variable w is fixed then the voltage dynamics along the dashed horizontal line exhibits bistability.

an action potential - see Fig. 34. On the other hand, 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 sub-threshold perturbations result in a simple return to the resting state, whereas super-threshold perturbations generate an action potential. Note, in particular, that decreasing the external input $I_{\rm ext}$ increases the firing threshold.

Almost all studies of ion channel fluctuations and their effect on action potential generation are based on a system size expansion with respect to the number N of ion channels, in which one reduces the dynamics to an effective Langevin equation [63, 97, 115], see section 2.4. However, such a reduction breaks down if the number of ion channels is small. Moreover, it can lead to exponentially large errors when estimating the rate of spontaneous action potential generation. Recently, Keener and Newby [148, 190] have used the WKB method for stochastic hybrid systems (see section 5.2) in order to analyze spontaneous action potential generation, which holds for small N and fast channel switching. They consider the simplified problem of how ion channel fluctuations affect the initiation of an action potential due to the opening of a finite number of Na^+ channels. The slow K^+ channels are assumed to be frozen, so that they effectively act as a leak current, and each sodium channel is treated as a simple two-state system. The Morris-Lecar equations (8.4a) and (8.4b) reduce to a piecewise deterministic equation for the membrane voltage

$$\frac{dx}{dt} = F_n(x) \equiv \frac{n}{N} f(x) - g(x), \tag{8.5}$$

where $f(x) = g_{\rm Na}(V_{\rm Na} - x)$ represents the gated sodium currents, $g(x) = -g_{\rm eff}[V_{\rm eff} - x] - I_{\rm ext}$ represents the sum of effective leakage currents and external inputs, and n is the number of open sodium channels. Note that the right-hand side of (8.5) is negative for large x and positive for small x. This implies that the stochastic voltage x is confined to some bounded domain $[x_1, x_2]$. The opening and closing of the ion channels is described by a birth-death process of the form (5.42) with x-dependent transition rates

$$\omega_+(x,n) = \alpha(x)(N-n), \quad \omega_-(n) = \beta n$$

with

$$\alpha(x) = \beta \exp\left(\frac{2(x-v_1)}{v_2}\right)$$

for constants β, v_1, v_2.

The associated probability density $\rho_n(x,t)$ satisfies the differential Chapman-Kolmogorov (CK) equation (see also (5.45))

$$\frac{\partial \rho_n}{\partial t} = -\frac{\partial [F_n(x)\rho_n(x,t)]}{\partial v} + \frac{1}{\epsilon}[\omega_+(x,n-1)\rho_{n-1}(x,t) + \omega_-(n+1)\rho_{n+1}(x,t) - (\omega_+(x,n) + \omega_-(n))\rho_n(x,t)],$$

with boundary condition $\rho_{-1}(x,t) \equiv 0$ and $\rho_{N+1}(x,t) = 0$. The equation for the steady–state distribution $\rho^*(x)$ of the discrete Markov process for fixed x can be obtained from the zero current condition

$$J(x,n) \equiv \omega_-(n)\rho_n^*(x) - \omega_+(x,n-1)\rho_{n-1}^*(x) = 0, \quad n \geq 0.$$

Solving this equation iteratively using the unit normalization condition on $\mathbf{p}^*$ gives the binomial distribution

$$\rho_n^*(x) = \frac{N!}{(N-n)!n!} a(x)^n b(x)^{N-n}, \quad a(x) = \frac{\alpha(x)}{\alpha(x)+\beta}, \; b(x) = \frac{\beta}{\alpha(x)+\beta}. \tag{8.6}$$

The mean number of open channels is then

$$\langle n\rangle = \sum_{n=1}^{N} n\rho_n^*(x) = Na(x). \tag{8.7}$$

It follows that the mean-field equation obtained in the $\epsilon \to 0$ limit is

$$\frac{dx}{dt} = \sum_{n=0}^{N} \rho_n^*(x)F_n(x) = \overline{F}(x) = a(x)f(x) - g(x). \tag{8.8}$$

It is straightforward to find physiologically reasonable parameter values under which the deterministic equation exhibits bistability. (For example, fixing w such that the dynamics of x takes place along the horizontal dashed line in Fig. 34.) Defining the effective potential $\Psi(x)$ according to $\Psi'(x) = -\overline{F}(x)$, we thus obtain a double well potential, see Fig. 2(b), with x_- representing a stable resting state and x_+ a stable active state.

The spontaneous initiation of an action potential now reduces to a first-passage time problem of noise-induced escape from the resting metastable state x_- to the active state x_+ by crossing the barrier at x_*. (Once this transition has occurred, the opening of the K^+ channels has to be taken into account in order that the system eventually returns to the rest state.) The calculation proceeds along identical lines to the genetic switch in section 5.3 [148]. In particular the mean rate of escape $\lambda_\varepsilon^{(0)}$ is given by equation (5.94). Hence, in order to determine $\lambda_\varepsilon^{(0)}$ for the ion channel model, one has to calculate the quasipotential $\Phi_0(x)$, the prefactor $k(x)$ and the effective diffusivity $D(x)$. First note that the generator of the discrete Markov process is now a tridiagonal matrix with

$$A_{n,n-1}(x) = \omega_+(x, n-1),\ A_{n,n}(x) = -\omega_+(x,n) - \omega_-(n),\ A_{n,n+1}(x) = \omega_-(n+1)$$

for $n = 0, 1, \ldots, N$. In the case of the stochastic ion channel model, the eigenvalue equation (5.85) with $R_n^{(0)}(x,p) = \psi_n(x,p)$ takes the explicit form

$$\begin{aligned}(N-n+1)\alpha\psi_{n-1} - [\lambda_0 + n\beta + (N-n)\alpha]\psi_n \\ + (n+1)\beta\psi_{n+1} = -p\left(\frac{n}{N}f - g\right)\psi_n\end{aligned} \tag{8.9}$$

Consider the trial solution

$$\psi_n(x,p) = \frac{\Gamma(x,p)^n}{(N-n)!n!}, \tag{8.10}$$

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

$$p = -\frac{N}{f(x)}\left(\frac{1}{\Gamma(x,p)} + 1\right)(\alpha(x) - \beta(x)\Gamma(x,p)), \tag{8.11}$$

and

$$\Lambda_0(x,p) = -N(\alpha(x) - \Gamma(x,p)\beta(x)) - pg(x). \tag{8.12}$$

Eliminating Γ from these equation gives

$$p = \frac{1}{f(x)}\left(\frac{N\beta(x)}{\Lambda_0(x,p) + N\alpha(x) + pg(x)} + 1\right)(\Lambda_0(x,p) + pg(x))$$

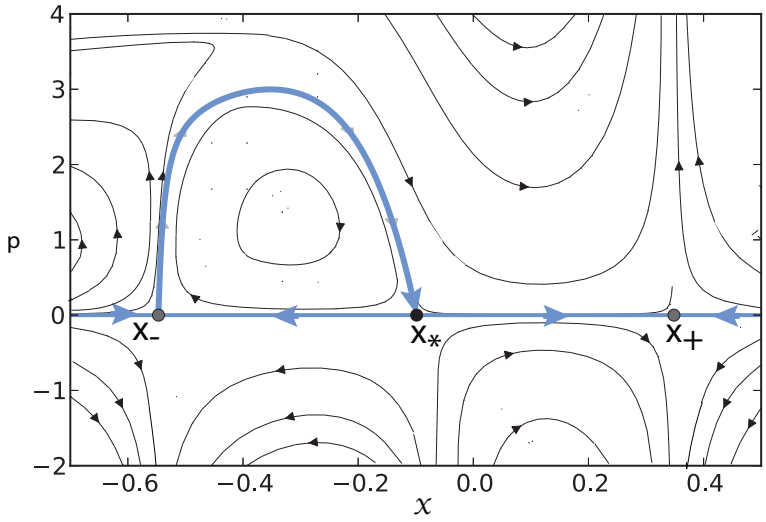


Figure 35. Phase portrait of Hamilton's equations of motion for the ion channel model with Hamiltonian given by the Perron eigenvalue (8.12). (x and p are taken to be dimensionless.) The zero energy solution representing the maximum likelihood path of escape from x_- is shown as the gray curve. (The corresponding path from x_+ is not shown.) Parameter values are $V_{\rm Na} = 120$ mV, $V_L = -62.3$ mV, $g_{\rm Na} = 4.4$ mS/cm^2, $g_L = 2.2$ mS/cm^2, and $\alpha(x) = \beta \exp[(x - v_1)/v_2]$ with $\beta = 0.8$ s^{-1}, $v_1 = -1.2mV$, $v_2 = 18mv$.

This yields a quadratic equation for Λ_0 of the form

$$\Lambda_0^2 + \sigma(x)\Lambda_0 - h(x,p) = 0. \tag{8.13}$$

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

$$p = 0 \text{ and } p = p(x) \equiv N\frac{\alpha(x)f(x) - (\alpha(x) + \beta)g(x)}{g(x)(f(x) - g(x))}. \tag{8.14}$$

It follows that the non-trivial WKB quasipotential (or action) is given by

$$\Phi_0(x) = \int_{x_-}^{x} p(y)dy. \tag{8.15}$$

In Fig. 35 we show solutions to Hamilton's equations in the (x,p)-plane, highlighting the zero energy maximum likelihood curve linking x_- and x_*. Note that $N\Phi(x_*)$, where $\Phi(x_*)$ is the area enclosed by the heteroclinic connection from x_- to x_*, gives the leading order contribution to $\log \tau$, where τ is the mean escape time.

Keener and Newby [148] also calculate the subleading-order contributions to the mean escape time. First, the null eigenfunction $\eta_n(x) = S(x,n)$ of equation (5.89),

which becomes

$$(N-m)\alpha\eta_{m+1} - [(N-m)\alpha + m\beta]\eta_m + m\beta\eta_{m-1} \\ = \mu\left(\frac{m}{N}f(x) - g(x)\right)\eta_m. \tag{8.16}$$

Trying a solution of the form $\eta_m(x) = \Gamma(x)^m$ yields

$$(N-m)\alpha\Gamma - ((N-m)\alpha + m\beta) + m\beta\Gamma^{-1} = \mu_1\left(\frac{m}{N}f(x) - g(x)\right). \tag{8.17}$$

Γ is then determined by canceling terms independent of m, which gives

$$\eta_n(x) = \left(\frac{b(x)g(x)}{a(x)(f(x)-g(x))}\right)^n. \tag{8.18}$$

The prefactor $k(x)$ may now be determined using equations (5.90) and (5.91). Finally, the effective diffusion coefficient $D(x)$ is obtained from equations (2.54) and (2.55) to give [148]

$$D(x_*) = \frac{f(x_*)^2\alpha(x_*)\beta}{N(\alpha(x_*)+\beta)^3}. \tag{8.19}$$

Keener and Newby show that the resulting expression for the mean escape time agrees very well with Monte Carlo simulations of the stochastic model [148]. On the other hand, estimating the mean escape time using a Gaussian approximation of the PDMP (see section 2.6) leads to exponentially large errors when the resting state of the neuron is well below the firing threshold (small external input).

As in the case of metastability in genetic switches (section 5), there have been various extensions of the above analysis to more complicated ion channel models, including a stochastic version of the full Morris-Lecar model [193, 194] and a model of dendritic action potentials [47]. Once again it is necessary to deal with higher-dimensional models, where numerical methods become crucial.

8.2. Diffusion in randomly switching environments

In the analysis of the two-state regulatory gene network in sections 3.2 and 3.3, we considered a piecewise deterministic PDE based on the Fokker-Planck equation, which described a population of independent cells operating in the same randomly switching environment. A very similar type of mathematical model occurs within the context of molecular diffusion within bounded environments such as the cell cytoplasm, the cell nucleus [254], and the branched network of tracheal tubes forming the passive respiration system in insects [54, 162]. Now the switching environment is generated by the random opening and closing of pores or ion channels within the boundary of the domain, and so one has the problem of analyzing the diffusion (or advection-diffusion) equation in a domain with a (partially) randomly switching boundary [50–52, 162].

As a simple example, consider a single Brownian particle drift-diffusing in the finite interval $[0, L]$ with a fixed absorbing boundary at $x = 0$ and a randomly switching gate at $x = L$; the right-hand boundary randomly switches between absorbing and reflecting. In order to keep track of the boundary state, we introduce the discrete random variable $n(t) \in \{0, 1\}$ such that

$$dX(t) = -V'(X)dt + \sqrt{2D}dW(t), \tag{8.20}$$

where $V(x)$ is a potential energy function, with an inhomogeneous Dirichlet boundary condition at $x = L$ if $n(t) = 0$ and a reflecting boundary at $x = L$ if $n(t) = 1$.

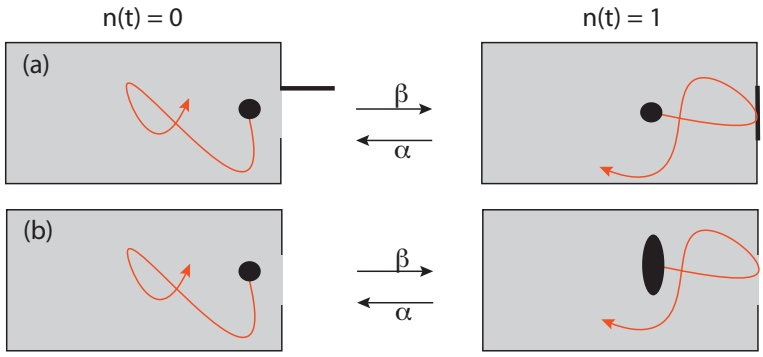


Figure 36. (a) Switching gate and (b) switching conformational state.

Assume that transitions between the two states are given by the two-state Markov process $0 \underset{\alpha}{\overset{\beta}{\rightleftharpoons}} 1$ with fixed transition rates α, β. One way to analyze the given system, is to view it as a piecewise SDE and to consider the differential Chapman-Kolmogorov (CK) equation for the joint probability density $\rho_n(x,t) = \mathbb{E}[p(x,t)1_{n(t)=n}]$:

$$\frac{\partial \rho_n(x,t)}{\partial t} = \frac{\partial}{\partial x}[V'(x)\rho_n(x,t)] + D\frac{\partial^2 \rho_n(x,t)}{\partial x^2} + \sum_{m=0,1} A_{nm}\rho_m(x,t), \tag{8.21}$$

with $\mathbf{A}$ the matrix

$$\mathbf{A} = \begin{pmatrix} -\beta & \alpha \\ \beta & -\alpha \end{pmatrix}. \tag{8.22}$$

Equation (8.21) is supplemented by the joint boundary conditions

$$\rho_0(0,t) = \rho_1(0,t) = 0, \quad \rho_0(L,t) = 0, \quad -\Phi'(L)\rho_1(L,t) - D\left.\frac{\partial \rho_1(x,t)}{\partial x}\right|_{x=L} = 0,$$

and the initial condition

$$\rho_n(x,0) = \delta(x-y)\rho_n^*, \quad \rho_0^* = 1 - \rho_1^* = \frac{\alpha}{\alpha+\beta},$$

where ρ_n^* is the stationary distribution of the two-state Markov process.

Equation (8.21), and its higher-dimensional analog, is the starting point for analyzing the escape of a single particle from a bounded domain with switching gates in the boundary [51, 52]. From this single-particle perspective, one could equally well interpret the source of the switching to be changes in the conformational state of the particle, rather than the gate(s), such that it can only pass through a gate when in one of the two states, see Fig. 36. (In this case, the diffusivity of the particle could also depend on n.) However, this equivalence breaks down when there are multiple particles and switching gates. Even when the particles are non-interacting, statistical correlations arise due to the fact that they all move in the same randomly switching environment; such correlations would not occur if each particle independently switched between different conformational states with fixed gates. In order to explore the multi-particle case, it is necessary to consider an alternative formulation of the problem, in which equation (8.21) is replaced by a piecewise deterministic diffusion equation:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \quad x \in [0,L], t > 0 \tag{8.23a}$$

with u satisfying the boundary conditions

$$u(0,t) = 0, \quad u(L,t) = 0 \text{ for } n = 0, \quad \partial_x u(L,t) = 0 \text{ for } n = 1. \tag{8.23b}$$

For simplicity we have dropped the drift term $V'(x)$. One can view a numerical solution of equation (8.23a) up to some time t as determining the probability density $u(x,t)$ conditioned on a single realization $\{n(s), 0 \leq s < t\}$ of the stochastic gate. Thus the conditional probability density $u(x,t)$ can be interpreted as determining the density of multiple particles moving in the same random environment. Each realization of the gate will typically generate a different solution $u(x,t)$ so that $u(x,t)$ is a random field variable. The relationship between the single particle and multi-particle perspectives is then established by averaging over realizations of the stochastic gate:

$$\rho_n(x,t) = \mathbb{E}[u(x,t)1_{n(t)=n}]. \tag{8.24}$$

The CK equation (8.21) is the analog of equations (3.8a) and (3.8b) for a single gene network, whereas the piecewise deterministic PDE (8.23a) is the analog of equation (3.15) for a population of gene networks evolving in the same random environment. Following [50] and section 3.3, we can analyze (8.23a) by discretizing space and constructing the differential CK equation for the resulting finite-dimensional stochastic hybrid system. The first step is to spatially discretize the piecewise deterministic PDE (8.23a) using a finite-difference scheme. One of the useful features of this discretization for switching boundaries is that we can incorporate the boundary conditions into the resulting discrete Laplacian. Introduce the lattice spacing a such that $(N+1)a = L$ for integer N, and let $u_j = u(aj)$, $j = 0, \ldots, N+1$. This yields the piecewise deterministic ODE

$$\frac{du_i}{dt} = \sum_{j=1}^{N} \Delta_{ij}^n u_j + \eta_a \delta_{i,N}\delta_{n,0}, \quad i = 1, \ldots, N, \quad \eta_a = \frac{\eta D_0}{a^2} \tag{8.25}$$

for $n = 0, 1$. Away from the boundaries ($i \neq 1, N$), Δ_{ij}^n is given by the discrete Laplacian

$$\Delta_{ij}^n = \frac{D}{a^2}[\delta_{i,j+1} + \delta_{i,j-1} - 2\delta_{i,j}]. \tag{8.26a}$$

On the left-hand absorbing boundary we have $u_0 = 0$, whereas on the right-hand boundary we have

$$u_{N+1} = \eta \text{ for } n = 0, \quad u_{N+1} - u_{N-1} = 0 \text{ for } n = 1.$$

These can be implemented by taking

$$\Delta_{1j}^0 = \frac{D}{a^2}[\delta_{j,2} - 2\delta_{j,1}], \quad \Delta_{Nj}^0 = \frac{D}{a^2}[\delta_{N-1,j} - 2\delta_{N,j}], \quad \Delta_{1j}^1 = \frac{D}{a^2}[\delta_{j,2} - 2\delta_{j,1}] \tag{8.26b}$$

and

$$\Delta_{Nj}^1 = \frac{2D}{a^2}[\delta_{N-1,j} - \delta_{N,j}]. \tag{8.26c}$$

Let $\mathrm{u}(t) = (u_1(t), \ldots, u_N(t))$ and introduce the probability density

$$\text{Prob}\{\mathrm{u}(t) \in (\mathrm{u}, \mathrm{u} + d\mathrm{u}), n(t) = n\} = \varrho_n(\mathrm{u}, t)d\mathrm{u}, \tag{8.27}$$

where we have dropped the explicit dependence on initial conditions. The probability density evolves according to the following differential CK equation for the stochastic hybrid system (8.25),

$$\frac{\partial \varrho_n}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial u_i}\left[\left(\sum_{j=1}^{N} \Delta_{ij}^n u_j + \eta_a \delta_{i,N}\delta_{n,0}\right)\varrho_n(\mathrm{u},t)\right] + \sum_{m=0,1} A_{nm}\varrho_m(\mathrm{u},t), \quad (8.28)$$

One can now proceed equation along identical lines to the analysis of equation (3.19) by deriving moment equations from (8.28) and then taking the continuum limit [50]. For example, defining the first moment

$$V_n(x,t) = \mathbb{E}[u(x,t)1_{n(t)=n}], \quad (8.29)$$

we obtain the system of equations

$$\frac{\partial V_0}{\partial t} = D\frac{\partial^2 V_0}{\partial x^2} - \beta V_0 + \alpha V_1 \quad (8.30a)$$

$$\frac{\partial V_1}{\partial t} = D\frac{\partial^2 V_1}{\partial x^2} + \beta V_0 - \alpha V_1 \quad (8.30b)$$

with

$$V_0(0,t) = V_1(0,t) = 0, \quad V_0(L,t) = \rho_0^*\eta > 0, \quad \partial_x V_1(L,t) = 0. \quad (8.31)$$

The resulting steady-state solution for $V = V_0 + V_1$ is [50, 162]

$$V(x) = \frac{x}{L}\frac{\eta}{1 + (\rho_1^*/\rho_0^*)(\xi L)^{-1}\tanh(\xi L)}, \quad \xi = \sqrt{\alpha+\beta} \quad (8.32)$$

Although one has the expected linear gradient in concentration, the dependence of the slope on model parameters is non-trivial. However, one recovers the classical result in the fast switching limit $\xi \to \infty$:

$$V(x) = \frac{x}{L}\eta,$$

This reflects a more general result that a switching gate is equivalent to an open gate in the fast switching limit, provided the fraction of time it is open is non-zero.

Similarly, we find that the higher-order moments

$$C_n^{(r)}(x,y) = \mathbb{E}[u(x_1,t)u(x_2,t)\dots u(x_r,t)1_{n(t)=n}], \quad (8.33)$$

satisfy the system of equations

$$\frac{\partial C_0^{(r)}}{\partial t} = D_0\sum_{l=1}^{r}\frac{\partial^2 C_0^{(r)}}{\partial x_l^2} - \beta C_0^{(r)} + \alpha C_1^{(r)} \quad (8.34a)$$

$$\frac{\partial C_1^{(r)}}{\partial t} = D_1\sum_{l=1}^{r}\frac{\partial^2 C_0^{(r)}}{\partial x_l^2} + \beta C_0^{(r)} - \alpha C_1^{(r)} \quad (8.34b)$$

Note that the r-point correlations couple to the $(r-1)$-order moments via the boundary conditions:

$$C_0^{(r)}(x_1,\dots,x_r,t)\Big|_{x_l=0} = C_1^{(r)}(x_1,\dots,x_r,t)\Big|_{x_l=0} = \partial_{x_l}C_1^{(r)}(x_1,\dots,x_r,t)\Big|_{x_l=L} = 0,$$

and

$$C_0^{(r)}(x_1,\dots,x_r,t)\Big|_{x_l=L} = \eta C_0^{(r-1)}(x_1,\dots,x_{l-1},x_{l+1}\dots,x_r,t),$$

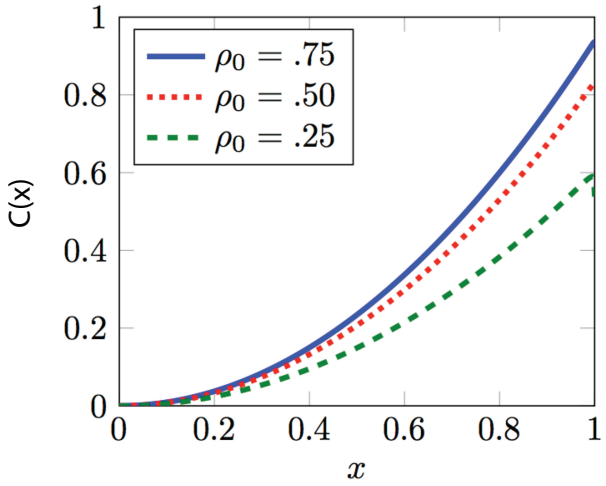


Figure 37. Plot of variance $C(x)$ for $L = \eta = 1$, $\xi = 10$, and either $\rho_0^* = 0.75$, 0.5, or 0.25.

for $l = 1, \ldots, r$. The solution of the second-order moment equations can be found in Ref. [50]. In Fig. 37 we show example plots of the steady-state variance $C(x) = \mathbb{E}[u(x)u(x)] = C_0^{(2)}(x,x) + C_1^{(2)}(x,x)$.

The rth moments of the piecewise deterministic PDE can be related to the statistics of r Brownian particles diffusing in the same randomly switching environment. The first observation is that the first-moment equations (8.30a) and (8.30b) are identical in form to the CK equation (8.21) describing a single Brownian particle diffusing under the same switching boundary conditions with $\Phi(x) = 0$. That is, $\rho_n(x,t) = V_n(x,t)$. One remaining issue is the interpretation of the inhomogeneous term η from the single particle perspective. This can be addressed by noting that if we set

$$\pi_n(x) = \frac{1}{\rho_n^* \eta} V_n(x),$$

where $V_n(x)$ is the steady-state solution of equations (8.30a) and (8.30b), then

$$0 = D\frac{\partial^2 \pi_0}{\partial y^2} - \beta[\pi_0 - \pi_1], \tag{8.35a}$$

$$0 = D\frac{\partial^2 \pi_1}{\partial y^2} + \alpha[\pi_0 - \pi_1] \tag{8.35b}$$

with boundary conditions

$$\pi_0(0) = \pi_1(0) = 0, \quad \pi_0(L) = 1, \quad \partial_y \pi_1(L) = 0.$$

We recognize equations (8.35a) and (8.35b) as steady-state backward equations for the CK equation (8.21) when $\Phi(x) = 0$ and $\eta = 0$, with $\pi_n(x)$ corresponding to the hitting probability that the particle exits at the end $x = L$ and $n(0) = n$. Furthermore, let $\pi_n^r(x_1, \ldots, x_r)$ be the probability that r Brownian particles all exit at $x = L$ given that the initial positions of the Brownian particles are $x_1, \ldots, x_r$ and $n(0) = n$. Then

$$\pi_n^r(x_1, \ldots, x_r) = \frac{1}{\rho_n^* \eta^r} \lim_{t \to \infty} C_n^{(r)}(x_1, \ldots, x_r, t), \tag{8.36}$$

where $C_n^{(r)}$ is the rth moment defined in equation (8.33). Though the particles are non-interacting, the probability that all r particles exit at $x = L$ is not the product of the probabilities of each particle exiting at $x = L$ because the particles are all diffusing in the same randomly switching environment. Equation (8.36) follows from writing down the backward equation for the joint probability density for r particles. The crucial step is determining the appropriate inhomogeneous boundary condition for the resulting r-dimensional time-independent PDE that determines the splitting probability. The boundary condition takes the form

$$\pi_0^{(r)}(x_1,\ldots,x_r)\Big|_{x_l=L} = \pi_0^{(r-1)}(x_1,\ldots,x_{l-1},x_{l+1}\ldots,x_r), \tag{8.37}$$

for $l = 1,\ldots,r$. This ensures that if one of the particles starts on the right-hand boundary when the latter is in the state $n = 0$, then the particle is immediately removed and thus one just has to determine the splitting probability that the $r-1$ remaining particles also exit at the right-hand boundary. Finally, performing a similar scaling to the first-moments yields the desired result.

8.3. Bacterial chemotaxis and velocity-jump processes

Another important class of biological switching processes involves switching between different velocity states. A classical example is the chemotaxis of the bacteria *E. coli* [31,39]. 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.) When all of the flagellar motors are rotating counter-clockwise, 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. 38(a). 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. 38(b). Tuning of the swimming behavior by environmental signaling molecules allows the bacterium to swim either toward 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 counter-clockwise rotation, see Fig. 38(c). This is another example of an ultrasensitive switch, which is based on cooperative receptor binding rather than a phophorylation-dephosophorylation cycle, as in the case of *V. harveyi* quorum sensing.

Early models of chemotaxis tended to be phenomenological in nature, representing the dynamics of cells in terms of an advection-diffusion equation for the cell density $u(\mathbf{x},t)$, in which the velocity is taken to depend on the concentration gradient of some chemotactic substance [150]. For example,

$$\frac{\partial u}{\partial t} = \nabla\cdot(D\nabla u - u\chi(c)\nabla), \tag{8.38}$$

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 u if cells secrete their own chemoattractant. An alternative, stochastic formulation of bacterial motion has

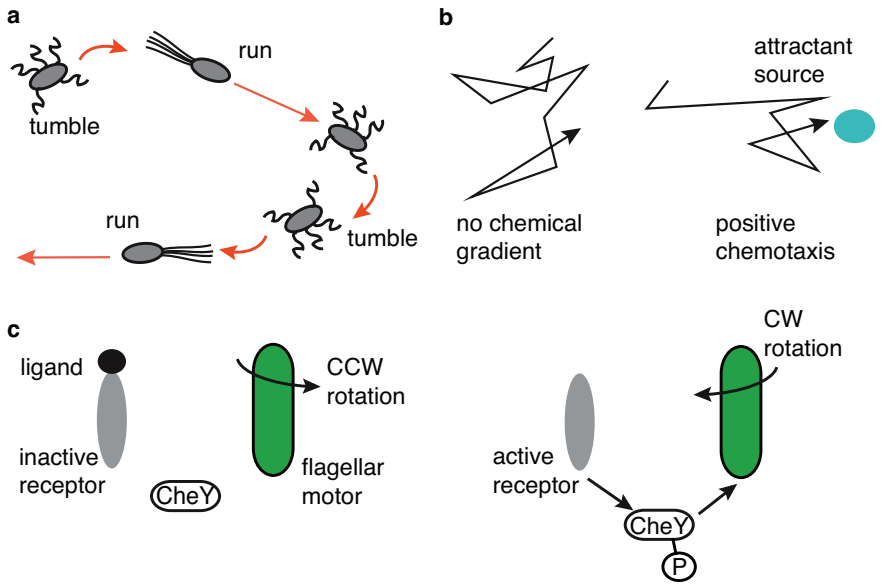


Figure 38. Bacterial chemotaxis. (a) A schematic showing the motion of a bacterium that consists of a series of runs and tumbles. (b) The sequence of runs and tumbles can be altered by an external chemical gradient so that the motion is biased towards (away from) an attractant (a repellant). (c) The 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.

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 [3, 73, 128, 199, 200]. For example, let $\rho(\mathbf{x}, \mathrm{v}, t)$ denote the probability density of cells at position $\mathbf{x} \in \mathbb{R}^d$ and velocity $\mathrm{v} \in \mathbb{R}^d$ at time t. Then ρ evolves according to an equation of the form

$$\frac{\partial}{\partial t}\rho(\mathbf{x}, \mathrm{v}, t) + \mathrm{v} \cdot \nabla \rho(\mathbf{x}, \mathrm{v}, t) = -\lambda \rho(\mathbf{x}, \mathrm{v}, t) + \lambda \int T(\mathrm{v}, \mathrm{v}')\rho(\mathbf{x}, \mathrm{v}', t)d\mathrm{v}'. \tag{8.39}$$

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 $T(\mathrm{v}, \mathrm{v}')$ is the conditional probability of a velocity jump from v' to v given that a jump occurs. If motion is restricted to 1D, then there are just two velocity states $\pm v$ and equation (8.39) reduces to a version of the velocity jump (or dichotomous noise) process considered in section 2.6:

$$\frac{\partial \rho_+}{\partial t} + v\frac{\partial \rho_+}{\partial x} = -\lambda \rho_+ + \lambda \rho_-, \tag{8.40a}$$

$$\frac{\partial \rho_-}{\partial t} - v\frac{\partial \rho_-}{\partial x} = \lambda \rho_+ - \lambda \rho_-, \tag{8.40b}$$

where $\rho_\pm(x, t)$ 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 also similar in form to the CK equations (3.8a) and (3.8a) of a two-state gene regulatory

network in the absence of degradation, with x now representing the spatial position of a bacterium rather than protein concentration. Several other biological systems can be modeled according to the pair of equations (8.40a) and (8.40b):

(i) The Dogterom-Leibler model of microtubule catastrophes [76]. Here x represents the length of a microtubule and the +-state (−-state) represents a recovery phase with growth rate v (a catastrophe phase with shrinkage rate v).

(ii) A model of bidirectional intracellular motor transport along a microtubule [44, 188, 189]. Here x represents the position of a motor along the filament, and the motor can exist in two motile states with velocities $\pm v$, corresponding to anterograde and retrograde transport respectively. (There could also be a third, non-motile or pausing state.) As with bacterial chemotaxis, there are higher-dimensional versions of bidirectional transport in which motors walk along along 2D and 3D cytoskeletal networks.

(iii) The previous example is a molecular version of a more general type of switching process known as a random intermittent search [28–30, 198], in which a particle randomly switches between a slow search phase and a faster non–search phase. There is evidence that such a search strategy is 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 [26], transcription factor proteins searching for particular sites on DNA [33, 123] (see section 3.6), and biochemical reaction kinetics [170]. Theoretical calculations of the mean first passage time to find a hidden target establishes the relative efficiency of the random search process. At the molecular level, the random switching between different motile states can increase the efficiency of biochemical reactions.

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 $\dot{c} = \pm v dc/dx$ [39]. This yields the pair of equations

$$\frac{\partial \rho_+}{\partial t} + v\frac{\partial \rho_+}{\partial x} = -\frac{1}{2}r(vc'(x))\rho_+(x,t) + \frac{1}{2}r(-vc'(x))\rho_-(x,t) \quad (8.41a)$$

$$\frac{\partial \rho_-}{\partial t} - v\frac{\partial \rho_-}{\partial x} = \frac{1}{2}r(vc'(x))\rho_+(x,t) - \frac{1}{2}r(-vc'(x))\rho_-(x,t). \quad (8.41b)$$

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

$$v\frac{\partial \rho_+}{\partial x} = \frac{1}{2}r(-vc'(x))\rho_-(x) - \frac{1}{2}r(vc'(x))\rho_+(x)$$

and

$$-v\frac{\partial \rho_-}{\partial x} = \frac{1}{2}r(vc'(x))\rho_+(x) - \frac{1}{2}r(-vc'(x))\rho_-(x).$$

Adding these two equations gives

$$v\frac{\partial \rho_+}{\partial x} - v\frac{\partial \rho_-}{\partial x} = 0,$$

which implies that the difference $\rho_+(x) - \rho_-(x) =$ constant. Assuming that $-\infty < x < \infty$, normalizability of the probability densities requires this constant to be zero. Hence, $\rho_\pm(x) = \rho(x)/2$ with $\rho(x)$ satisfying the single equation

$$v\frac{\partial \rho}{\partial x} = \frac{1}{2}\left[r(-vc'(x)) - r(vc'(x))\right]\rho(x).$$

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

$$v\frac{\partial \rho}{\partial x} = -r'(0)c'(x)\rho(x).$$

This has the straightforward solution

$$\rho(x) = \frac{1}{Z}\mathrm{e}^{-r'(0)c(x)}, \tag{8.42}$$

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 (8.42) coincide with maxima of the concentration $c(x)$. Conversely, $r'(0) > 0$ for a chemorepellant and maxima of $\rho(x)$ coincide with minima of the concentration. Note that Erban and Othmer [87] have developed a more detailed 1D model of chemotaxis that incorporates aspects of the biochemical signal transduction pathways.

8.4. Stochastic neural networks

As our final example, we briefly consider one approach to incorporating noise into local populations of neurons, which explicitly makes a formal connection with gene networks [43], see also [41, 42, 57, 58]. The basic idea is to partition a network of synaptically-coupled spiking neurons into M homogeneous subpopulations labeled $\alpha = 1, \ldots, M$ and to represent the output activity of the α-th population in terms of the number $N_\alpha(t)$ of neurons that are currently firing. The discrete stochastic variables $N_\alpha(t)$ are taken to evolve according to a birth-death Markov process:

$$N_\alpha(t) \to N_\alpha(t) \pm 1: \quad \text{transition rate } \omega_\pm, \tag{8.43}$$

The corresponding transition rates are

$$\omega_+ = \frac{1}{\tau_a}F(U_\alpha), \quad \omega_- = \frac{n_\alpha}{\tau_a}, \tag{8.44}$$

where F is a sigmoidal firing-rate function, the time constant τ_a determines the relaxation rate of a local population to the instantaneous firing rate, and $U_\alpha(t)$ is the net synaptic current into population α. The latter evolves according to the piecewise deterministic ODE

$$\tau\frac{dU_\alpha(t)}{dt} = -U_\alpha(t) + \sum_{\beta=1}^{M} w_{\alpha\beta}N_\beta(t), \tag{8.45}$$

where $w_{\alpha\beta}$ represents the effective strength or weight of synaptic connections from population β to population α, and τ is a synaptic time constant. We see that the

resulting stochastic process defined by equations (8.43)–(8.45) is a PDMP at the population neuron level, just as stochastic ion channel gating provides an example at the cellular level, see section 8.1. The formal connection between the stochastic neural network and a stochastic gene network is illustrated in Fig. 39. Given this connection, all of the mathematical methods highlighted in this review carry over to the neural population model. We briefly sketch a few of the details.

Denote the random state of the model at time t by $\{(U_\alpha(t), N_\alpha(t)); \alpha = 1, \ldots, M\}$. Introduce the corresponding probability density

$$\text{Prob}\{U_\alpha(t) \in (u_\alpha, u_\alpha + du), N_\alpha(t) = n_\alpha; \alpha = 1, \ldots, M\} = \rho(\mathrm{u}, \mathbf{n}, t)d\mathrm{u}, \tag{8.46}$$

with $\mathbf{n} = (n_1, \ldots, n_M)$ and $\mathrm{u} = (u_1, \ldots, u_M)$. It follows from equations (8.43)–(8.45) that the probability density evolves according to the Chapman-Kolmogorov (CK) equation [43]

$$\frac{\partial \rho}{\partial t} + \frac{1}{\tau} \sum_\alpha \frac{\partial [v_\alpha(\mathrm{u}, \mathbf{n})\rho(\mathbf{u}, \mathbf{n}, \mathbf{t})]}{\partial u_\alpha} \tag{8.47}$$

$$= \frac{1}{\tau_a} \sum_\alpha \left[(\mathbb{T}_\alpha - 1)\left(\omega_-(n_\alpha)\rho(\mathrm{u}, \mathbf{n}, t)\right) + (\mathbb{T}_\alpha^{-1} - 1)\left(\omega_+(u_\alpha)\rho(\mathrm{u}, \mathbf{n}, t)\right) \right],$$

$$\equiv \frac{1}{\tau_a} \sum_{\mathbf{m}} A(\mathbf{n}, \mathbf{m}; \mathrm{u})\rho(\mathrm{u}, \mathbf{m}, t) \tag{8.48}$$

with

$$\omega_+(u_\alpha) = F(u_\alpha), \quad \omega_-(n_\alpha) = n_\alpha, \quad v_\alpha(\mathrm{u}, \mathbf{n}) = -u_\alpha + \sum_\beta w_{\alpha\beta} n_\beta. \tag{8.49}$$

The continuous-time Markov process for fixed u,

$$\frac{d\rho(\mathrm{u}, \mathbf{n}, t)}{dt} = \frac{1}{\tau_a} \sum_{\mathbf{m} \in I} A(\mathbf{n}, \mathbf{m}; \mathrm{u})\rho(\mathrm{u}, \mathbf{m}, t),$$

has a globally attracting steady-state $\rho^*(\mathrm{u}, \mathbf{n})$ such that $\rho(\mathrm{u}, \mathbf{n}, t) \to \rho^*(\mathrm{u}, \mathbf{n})$ as $t \to \infty$. The steady-state equation is

$$0 = \sum_{\mathbf{m}} A(\mathbf{n}, \mathbf{m}; \mathrm{u})\rho^*(\mathrm{u}, \mathbf{m})$$

$$= \sum_{\alpha=1}^{M} \left[(n_\alpha + 1)\rho^*(\mathrm{u}, \mathbf{n} + \mathbf{e}_\alpha) - n_\alpha \rho(\mathrm{u}, \mathbf{n}) + F(u_\alpha)(\rho(\mathrm{u}, \mathbf{n} - \mathbf{e}_\alpha) - \rho^*(\mathrm{u}, \mathbf{n})) \right],$$

where $[\mathbf{e}_\alpha]_\beta = \delta_{\alpha,\beta}$. The solution can be factorized as $\rho^*(\mathrm{u}, \mathbf{n}) = \prod_{\beta=1}^M \rho_1(u_\beta, n_\beta)$ with

$$0 = \sum_{\alpha=1}^{M} \left[\prod_{\beta \neq \alpha} \rho_1(u_\beta, n_\beta) \right] \left[J(u_\alpha, n_\alpha + 1) - J(u_\alpha, n_\alpha) \right],$$

where

$$J(u, n) = n\rho_1(u, n) - F(u)\rho_1(u, n - 1).$$

Since $\rho_1(u, -1) \equiv 0$, it follows that $J(u, n) = 0$ for all n. Hence,

$$\rho_1(u, n) = \rho_1(u, 0) \prod_{m=1}^{n} \frac{F(u)}{m} = \rho_1(u, 0) \frac{F(u)^n}{n!}, \tag{8.50}$$

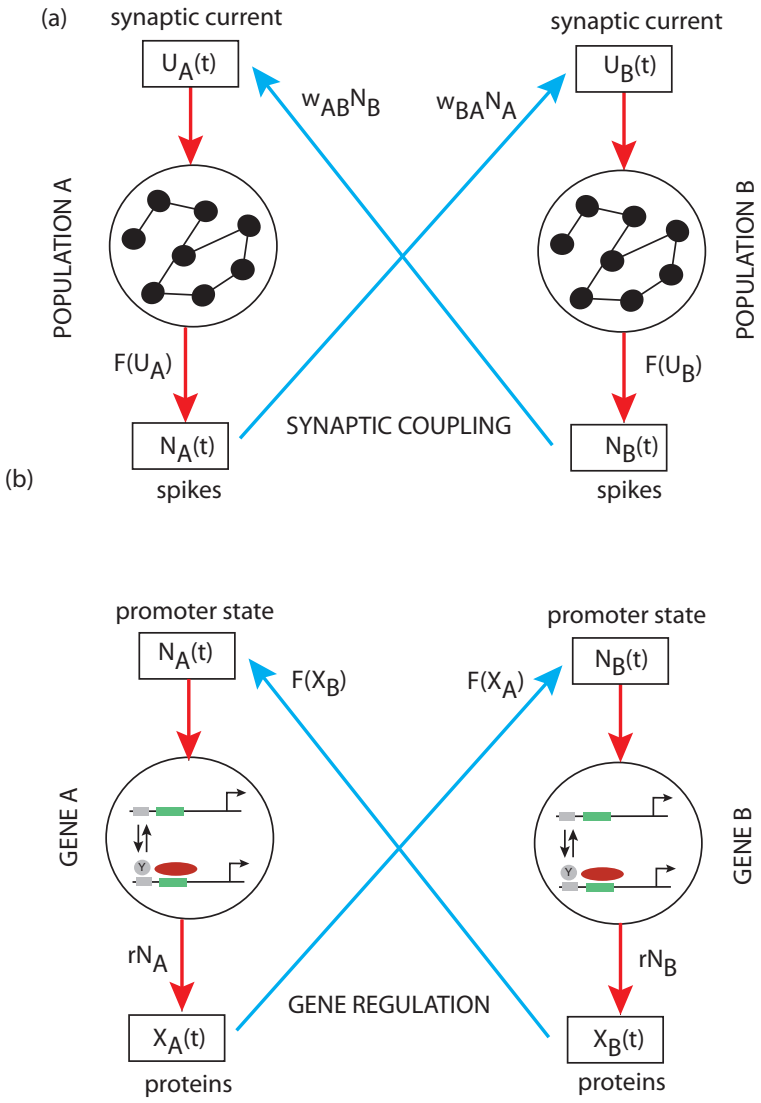


Figure 39. Comparison of mathematical structures for a pair of synaptically coupled neural populations and a pair of mutually regulated genes. (a) In the neural case the input synaptic current U is the piecewise deterministic continuous variable, whereas the number of output spikes N is a discrete Markov process. The nonlinear sigmoid-like function F appears in the transition rates of the discrete process. Population A can either excite ($w_{BA} > 0$) or inhibit ($w_{BA} < 0$) population B, and vice versa. It is also possible to have autofeedback with weight w_{AA}. (b) In the genetic case, the output concentration X of proteins is the piecewise deterministic continuous variable, whereas the state N of the promoter (active or inactive) is a discrete Markov process. The nonlinear sigmoid-like function F again appears in the transition rates of the discrete process. Proteins from gene A can either activate or repress gene B and vice versa, depending on the precise form of F. Proteins can also regulate their own gene (autoregulation).

and the corresponding normalized density is a Poisson process with rate $F(u)$

$$\rho_1(u,n) = \mathrm{e}^{-F(u)}\frac{F(u)^n}{n!}. \tag{8.51}$$

There are two time-scales in the CK equation (8.47), the synaptic time constant τ and the time constant τ_a, which characterizes the relaxation rate of population activity. In the limit $\tau_a \to 0$ for fixed τ, we obtain deterministic mean-field equations

$$\begin{aligned}\tau\frac{du_\alpha}{dt} &= \langle v_\alpha\rangle(\mathrm{u}(t)) \equiv \sum_{\mathbf{n}} v_\alpha(\mathrm{u}(t),\mathbf{n})\rho^*(\mathrm{u}(t),\mathbf{n}) \\ &= -u_\alpha(t) + \sum_{\beta=1}^{M} w_{\alpha\beta}\sum_{\mathbf{n}} n_\beta \rho^*(\mathrm{u}(t),\mathbf{n}). \end{aligned} \tag{8.52}$$

Since $\rho^*(\mathrm{u},\mathbf{n})$ is given by product of independent Poisson processes with rates $F(u_\alpha)$, it follows that

$$\langle n_\beta\rangle = F(u_\beta), \tag{8.53}$$

and (8.52) reduces to the standard Hopfield neural network equation [130].

By analogy with gene regulatory networks, a single neural population with positive synaptic feedback, or a pair of mutually inhibitory neural populations can exhibit bistability in the deterministic limit. When intrinsic noise is included one has noise-induced transitions that can be analyzed along identical lines to section 5.2 [42, 43].

9. Discussion

In this review we have presented current state-of-the-art analytical methods and mathematical tools to deal with the stochastic processes giving rise to biological switches. One of our main goals was to go beyond the standard and often-used techniques of the linear noise approximation and system size expansion, by covering more advanced topics in the theory of stochastic processes, including continuous time Markov chains, chemical reaction network theory, stochastic hybrid systems, queuing theory, large deviations and the WKB method, adiabatic reductions, and diffusion in switching environments. In order to develop the main ideas, we focused on applications to simple gene regulatory networks. However, in section 8 we illustrated how the same techniques can be extended to other examples of biological switches.

One of the major challenges in furthering our understanding of biological switches, as well as other processes in systems biology, is dealing with the complexity of most gene regulatory and biochemical signaling networks. For example, typical reaction networks involve multiple nodes (complexes) and links (reactions) that represent a hierarchy of nonlinear feedback loops. Identifying characteristic modules and motifs within these networks (including switches) is a subject of intense research. The theory of chemical reaction networks touched upon in section 2.3 could be important in identifying non-equilibrium steady-states in complex networks, as could multi-scale analyses that exploit any separation of time-scales. Another source of complexity arises when spatial effects become important. We touched on this within the context of gene regulatory networks, where some form of facilitated diffusion is thought to play a role in speeding up the search of a transcription factor for a specific binding site on DNA (see section 3.6). In eukaryotic cells, one also has to take into account the fact that gene regulation involves the exit of mRNA from the nucleus through the nuclear pore complex, and the subsequent reentry of newly synthesized transcription factors into the nucleus [254]. More generally, one needs to consider hybrid systems involving the

coupling of stochastic chemical reactions within well mixed compartments with bulk diffusion in complex, crowded environments. This in turn will require the development of efficient numerical schemes for simulating reaction-diffusion systems with non-trivial boundary conditions. Much of the current work in this area is based on constructing spatial extensions of the SSA introduced in section 2.5 by considering reaction-diffusion master equations [10, 37, 84, 88, 137, 138, 175].

Finally, we would like to highlight the increasingly important role that stochastic hybrid systems such as PDMPs are playing in biological modeling, see also [45]. Throughout the review we have encountered examples of systems that involve a coupling between one or more continuous variables, evolving either deterministically or stochastically, and a discrete Markov process. In the case off gene networks, this occurs if there is a mixture of molecules with low copy numbers (eg. genes, mRNA) and molecules with high copy numbers (eg. certain proteins). Other examples were presented in section 8, including membrane voltage fluctuations induced by the stochastic opening and closing of ion channels, bacterial chemotaxis, stochastic neural networks, and diffusion in randomly switching environments. The theory of stochastic hybrid systems is underdeveloped compared to that of SDEs and discrete Markov processes, although similar techniques can be applied, including large deviations and WKB methods, and diffusion approximations. It is a rich area for further exploration.

Acknowledgments:

PCB would like to thank the National Science Foundation for its continuing support.

References

[1] Acar M, Mettetal J T and van Oudenaarden A 2008 Stochastic switching as a survival strategy in fluctuating environments *Nat. Gen.* **40** 471-475

[2] Alon U 2007 *An introduction to systems biology: Design principles of biological circuits* Chapman and Hall/CRC

[3] Alt W 1980 Biased random walk model for chemotaxis and related diffusion approximation *J. Math. Biol.* **9** 147-177

[4] Anderson D F 2007 A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J. Chem. Phys.* **127** 214107

[5] Anderson D F Craciun G and Kurtz T G 2010 Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.* **72** 1947-1970 (2010)

[6] Anderson D F and Shiu A (2010) The dynamics of weakly reversible population processes near facets. *SIAM J. Appl. Math.* **70** 1840-1858

[7] Anderson D F and Kurtz T G 2011 Continuous time Markov chain models for chemical reaction networks In: *Design and Analysis of Biomolecular Circuits* pp. 3-42

[8] Anderson D F and Higham D J 2012 Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics *Multiscale Modeling and Simulation* **10** 146-179

[9] Anderson D F, Higham D J and Sun Y 2014 Complexity of multilevel Monte Carlo tau-leaping *SIAM Journal on Numerical Analysis* **52** 3106-3127

[10] Andrews S S and Bray D 2004 Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys Biol* **1** 137-151

[11] Anetzberger C, Pirch T and Jung K 2009 Heterogeneity in quorum sensing- regulated bioluminescence in *Vibrio harveyi*. *Molecular Microbiology* **73** 267-277

[12] Angeli D, De Leenheer P and Sontag E D 2007 A petri net approach to the study of persistence in chemical reaction networks. *Math. Biosci.* **210** 598-618

[13] Anguige K, King J R and Ward J P 2006 A multi-phase mathematical model of quorum-sensing in a maturing Pseudomonas aeruginosa biofilm. *Math.Biosci.* **203**,240-276

[14] Arkin A, Ross J and McAdams H H 1998 Stochastic kinetic analysis of developmental pathway bifurcation in phage infected *escherichia coli* cells. *Genetics* **149** 1633-1648
[15] Assaf M and Meerson B 2006 Spectral theory of metastability and extinction in birth-death systems *Phys. Rev. Lett.* **97** 200602
[16] Assaf M, Kamenev A and Meerson B 2008 Population extinction in a time-modulated environment *Phys. Rev. E* **78** 041123
[17] Assaf M, Kamenev A and Meerson B 2009 Population extinction risk in the aftermath of a catastrophic event *Phys. Rev. E* **79** 011127
[18] Assaf M and Meerson B 2010 Extinction of metastable stochastic populations *Phys. Rev. E* **81** 021116
[19] Assaf M, Roberts E, Shulten Z L and Goldenfeld N 2013 Extrinsic noise driven phenotype switching in a self-regulating gene*Phys. Rev. E* **111** 058102
[20] Assaf M, Mobilia M and Roberts E 2013 Cooperation dilemm in finite populations under fluctuating environments. *Phys. Rev. E* **111** 238101
[21] Aurell E and Sneppen K 2002 Epigenetics as a first exit problem *Phys Rev Lett* **88** 048101
[22] Aurell E and Sneppen K 2002 Stability puzzles in phage λ *Phys Rev E* **65** 051914
[23] Bacaer N 2015 On the stochastic SIS epidemic model in a periodic environment *J. Math. Biol.* **71** 491-511
[24] Balaban N Q, Merrin J, Chait R, Kowalik L and Leibler S 2004 Bacterial persistence as a phenotypic switch. *Science* **205**1578-1579
[25] Balakrishnan V and Chaturvedi S 1988 Persistent diffusion on a line *Physica A* **148** 581
[26] Bell J W 1991 *Searching behavior, the behavioral ecology of finding resources.* London: Chapman and Hall
[27] Bena I 2006 Dichotomous Markov noise: exact results for out-of-equilibrium systems. *Int. J. Mod. Phys. B* **20** 2825
[28] Benichou O, Coppey M, Moreau M, Suet P and Voituriez R 2005 A stochastic model for intermittent search strategies. *J. Phys. Cond. Matter* **17**, S4275-4286
[29] Benichou O, Loverdo C, Moreau M and Voituriez R 2007 A minimal model of intermittent search in dimension two *J. Phys. A* **19** 065141
[30] Benichou O, Loverdo C, Moreau M and Voituriez R 2011 Intermittent search strategies. *Rev. Mod. Phys.* **83** 81-
[31] Berg H C and Purcell E M 1977 Physics of chemoreception. *Biophys. J.* **20** 93-219
[32] Berg O G 1978 A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* **4** 587-603
[33] Berg O G, Winter R B and von Hippel P H 1981 Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory *Biochem.* **20**, 6929–6948
[34] Berg O G, Paulsson J and Ehrenberg M 2000 Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. *Biophys. J.* **79** 1228-1236
[35] Berglund N 2011 Kramers' law: Validity, derivations and generalizations *Markov Processes Related Fields* **19** 459-490
[36] Berger S L 2007 The complex language of chromatin regulation during transcription. *Nature* **447** 407-412
[37] Bhalla U S 2004 Signaling in small subcellular volumes. I. stochastic and diffusion effects on individual pathways. *Biophys. J.* **87**, 733-744
[38] Bhatt D, Zhang B W,and Zuckerman D M 2010 Steady state via weighted ensemble path sampling. *J. Chem Phys.* **133** 014110
[39] Bialek W 2012 *Biophysics* Princeton University Press, Princeton
[40] Bolhius P G, Chandler D, Dellago C and Geissler P L 2002 Transition path sampling: throwing ropes over rough mountain passes, in the dark *Ann. Rev. Phys. Chem.* **53** 291-318
[41] Bressloff P C 2009 Stochastic neural field theory and the system-size expansion. *SIAM J. Appl. Math* **70** 1488-1521
[42] Bressloff P C 2010 Metastable states and quasicycles in a stochastic Wilson-Cowan model of neuronal population dynamics. *Phys. Rev. E* **85** 051903
[43] Bressloff P C and Newby J M 2013 Metastability in a stochastic neural network modeled as a velocity jump Markov process. *SIAM Appl. Dyn. Syst.* **12** 1394-1435
[44] Bressloff P C and Newby J M 2013 Stochastic models of intracellular transport. *Rev. Mod. Phys.* **85** 135-196
[45] Bressloff P C 2014 *Stochastic Processes in Cell Biology* Springer
[46] Bressloff P C and Newby J M 2014 Path-integrals and large-deviations in stochastic hybrid systems. *Phys. Rev. E* **89** 042701
[47] Bressloff P C and Newby J M 2014 Stochastic hybrid model of spontaneous dendritic NMDA

spikes. Phys. Biol. *Phys. Biol.* **11** 016006
[48] Bressloff P C and Faugeras O 2014 On the Hamiltonian structure of large deviations in stochastic hybrid systems. arXiv preprint arXiv:1410.2152
[49] Bressloff P C 2015 Path-integral methods for analyzing the effects of fluctuations in stochastic hybrid neural networks. *J. Math. Neurosci* **5** 33pp
[50] Bressloff P C and Lawley S D 2015 Moment equations for a piecewise deterministic PDE *J. Phys. A* **48** 105001
[51] Bressloff P C and Lawley S D 2015 Escape from a potential well with a randomly switching boundary *J. Phys. A* **48** 225001
[52] Bressloff P C and Lawley S D 2015 Escape from subcellular domains with randomly switching boundaries. *Multiscale Model. Simul.* **13** 1420-1445
[53] Bressloff P C and Lawley S D 2015 Stochastically-gated diffusion-limited reactions for a small target in a bounded domain. *Phys. Rev. E* **92** 062117
[54] Bressloff P C and Lawley S D 2016 Diffusion on a tree with stochastically-gated nodes. *J. Phys. A* **49** 245601
[55] Bressloff P C 2016 Ultrasensitivity and noise amplification in a model of *V. harveyi* quorum sensing *Phys Rev E* **93** 062418
[56] Brophy J A N and Voight C A 2014 Principles of genetic circuit design *Nature Methods* **11** 508-520
[57] Buice M and Cowan J D 2007 Field-theoretic approach to fluctuation effects in neural networks *Phys. Rev. E* **75** 051919
[58] Buice M, Cowan J D and Chow C C 2010 Systematic fluctuation expansion for neural network activity equations *Neural Comp.* **22** 377-426
[59] Cai L, Friedman N and Xies X S 2006 Stochastic protein expression in individual cells at the single molecule level. *Nature* **440** 358-362
[60] Cao Y, Gillespie D T, Petzold L R 2006 Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.* **124** 044109
[61] Chan H B, Dykman M I and Stambaugh C 2008 Switching-path distribution in multidimensional systems *Phys. Rev. E* **78** 051109
[62] Chiang W Y, Li Y X and Lai P Y 2011 Simple models for quorum sensing: Nonlinear dynamical analysis. *Phys. Rev. E* **84** 041921
[63] Chow C C and White J A 1996 Spontaneous action potentials due to channel fluctuations. *Biophys. J.* **71** 3013–3021
[64] Coombes S and Bressloff P C (eds.) 2005 *Bursting: The Genesis of Rhythm in the Nervous System* World Scientific Press
[65] Coppey M, Benichou O, Voituriez R and Moreau M 2004 Kinetics of target site localization of a protein DNA: a stochastic approach. *Biophys J.* **87** 1640-1649
[66] Couzin I D 2009 Collective cognition in animal groups. *Trends. Cogn. Sci.* **13**, 36-43
[67] Craciun G and Feinberg M 2005 Multiple equilibria in complex chemical reaction networks: I. The injectivity property. *SIAM J. Appl. Math.* **65** 1526-1546
[68] Craciun G and Feinberg M 2006 Multiple equilibria in complex chemical reaction networks: II. The species-reactions graph. *SIAM J. Appl. Math.* **66** 1321-1338
[69] Crick F 1970 The central dogma of molecular biology. *Nature* **227** 561-563
[70] Davis M H A 1984 Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Society, Series B (Methodological)* **46** 353-388
[71] Davies D G, Parsek M R, Pearson J P, Iglewski B H, Costerton J W and Greenberg E P 1998 The involvement of cell-to-cell signals in the development of bacterial biofilm. *Science* **280** 295-298
[72] Dembo A and Zeitouni O 1998 *Large Deviations Techniques and Applications*. Applications of Mathematics, vol. 38. Springer Berlin
[73] Dickinson R B and Tranquillo R T 1995 Transport equations and indices for random and biased cell migration based on single cell properties *SIAM J. Appl. Math.* **55** 1419-1454
[74] Dickson, Warmflash A and Dinner A R 2009 Separating forward and backward pathways in nonequilibrium umbrella sampling *J. Chem. Phys.* **130** 074104
[75] Dockery J D and Keener J P 2001 A mathematical model for quorum-sensing in Pseudomonas aeruginosa. *Bull. Math. Biol.* **63**, 95-116
[76] Dogterom M and Leibler S 1993 Physical aspects of the growth and regulation of microtubule structures. *Phys. Rev. Lett.* **70** 1347-1350
[77] Doi M 1976 Second quantization representation for classical many-particle systems. *J. Phys. A.* **9** 1465-1477
[78] Doi M 1976 Stochastic theory of diffusion controlled reactions. *J. Phys. A.* **9** 1479-1495

[79] Donovan R M, Sedgewick, A J, Faeder J R, Zuckerman D M 2013 Efficient stochastic simulation of chemical kinetic networks using a weighted ensemble of trajectories. *J. Chem. Phys.* **129** 115105
[80] Dunlap P V 1999 Quorum regulation of luminescence in Vibrio fischeri. *J. Mol. Microbiol. Biotechnol.* **1** 5-12
[81] Dykman M I, Mori E, Ross J and Hunt P M 1994 Large fluctuations and optimal paths in chemical kinetics. *J. Chem. Phys. A* **100** 5735-5750
[82] Dykman M I, Schwartz I B and Landsman A S 2008 Disease extinction in the presence of random vaccination. *Phys. Rev. Lett.* **101** 078101
[83] Elf J and Ehrenberg M 2003 Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.* **13** 2475-2484
[84] Elf J, and Ehrenberg M 2004 Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst Biol* **1** 230-236
[85] Elgart V and Kamenev A 2004 Rare event statistics in reaction–diffusion systems. *Phys. Rev. E* **70** 041106
[86] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 Stochastic gene expression in a single cell. *Science* **297** 1183-1186
[87] Erban R and Othmer H 2005 From individual to collective behavior in bacterial chemotaxis. *SIAM J. Appl. Math.* **65** 361-391
[88] Erban R and Chapman J 2009 Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. *Phys. Biol.* **6**, 046001
[89] Escudero C and Rodriguez J A 2008 Persistence of instanton connections in chemical reactions with time-dependent rates. *Phys. Rev. E* **77** 011130
[90] Escudero C and Kamanev A 2009 Switching rates of multistep reactions. *Phys. Rev. E* **79** 041149
[91] Faggionato A, Gabrielli D and Crivellari M R 2009 Non-equilibrium thermodynamics of piecewise deterministic Markov Processes. *J Stat Phys* **137** 259-304
[92] Faggionato A, Gabrielli D and Crivellari M R 2010 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
[93] Feinberg M 1972 Complex balancing in general kinetic systems. *Arch. Rational Mech. Anal.* **49** 187-194 (1972).
[94] Feinberg M 1979 Lectures on chemical reaction networks, Delivered at the Mathematics Research Center, Univ. Wisc. Madison. Available for download at http://www.che.eng.ohio-state.edu/?feinberg/LecturesOnReactionNetworks
[95] Feinberg M 1987 Chemical reaction network structure and the stability of complex isothermal reactors - I. the deficiency zero and deficiency one theorems. *Chem. Eng. Sci.* **42** 2229-2268 (1987).
[96] Feng J and Kurtz T G 2006 *Large Deviations for Stochastic Processes*. American Mathematical Society
[97] Fox R F and Lu Y N 1994 Emergent collective behavior in large numbers of globally coupled independent stochastic ion channels. *Phys. Rev. E* **49** 3421-3431
[98] Freidlin M I and Wentzell A D 1998 *Random perturbations of dynamical systems*. Springer, New York
[99] Friedman A and Craciun G 2005 A model of intracellular transport of particles in an axon *J. Math. Biol.* **51** 217-246
[100] Friedman A and Craciun G 2006 Approximate traveling waves in linear reaction-hyperbolic equations. *SIAM J. Math. Anal.* **38** 741-758
[101] Friedman N, Cai L and Xie X S 2006 Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.* **97** 168302
[102] Fuqua C, Winans S C and Greenberg E P 1996 Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annu. Rev. Microbiol.* **50** 727-751 (1996).
[103] Gander M J, Mazza C and Rummler H 2007 Stochastic gene expression in switching environments. *J. Math. Biol.* **55** 259-294
[104] Garde C, Bjarnsholt T, Givskov M, Jakobsen T H, Hentze M, Claussen A, Sneppen K, Ferkinghoff-Borg J and Sams T 2010 Quorum sensing regulation in Aeromonashydrophila. *J. Mol. Biol.* **396** 849-857
[105] Gardiner C W 2009 *Handbook of stochastic methods, 4th edition* Springer Berlin
[106] Gardner T S, Cantor C R and Collins J J 2000 Construction of a genetic toggle switch in *E. coli*. *Nature* **403** 339-342

[107] Ge H and Qian M 2008 Sensitivity amplification in the phosphorylation-dephosphorylation cycle: nonequilibrium steady states, chemical master equation and temporal cooperativity. *J. Chem. Phys.* **129** 015104
[108] Ghosh S, Pal A K and Bose I 2013 Noise-induced regime shifts: A quantitative characterization *Eur. Phys. J. E* **36** 123
[109] Gibson M and Bruck J 2000 Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.* **104** 1876-1889
[110] Gillespie, D. T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340-2361 (1977).
[111] Gillespie, D. T.: Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. **115**, 1716-1733 (2001).
[112] Gillespie, D. T., Hellander, A., Petzold, L. R.: Perspective: stochastic algorithms for chemical kinetics. J. Chem. Phys. **138**, 170901 (2013).
[113] Goldbeter A and Koshland D E 1981. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. USA* **78** 6840-6844
[114] Goldstein S 1951 On diffusion by discontinuous movements and on the telegraph equation. *Quart. J. Mech. Appl. Math.* **4** 129-156
[115] Goldwyn J H and Shea-Brown E 2011 The what and where of adding channel noise to the Hodgkin-Huxley equations. *PLoS Comp. Biol.* **7)** e1002247
[116] Gonze D, Halloy J and Gaspard P 2002 Biochemical clocks and molecular noise: theoretical study of robustness factors. *J. Chem. Phys.* **116** 10997-11010
[117] Goryachev A B 2009 Design principles of the bacterial quorum sensing gene newtorks. *WIRE Syst. Biol. Med.* **1** 45-60
[118] Gou J and Ward M J 2016 An asymptotic analysis of a 2-D model of dynamically active compartments coupled by bulk diffusion. *J. Nonlin. Sci.* (2016).
[119] Graham R and Tel T 1984 On the weak-noise limit of Fokker-Planck models. *J. Stat. Phys* **35** 729
[120] Graham R and Tel T 1985 Weak-noise limit of Fokker-Planck models and non-differentiable potentials for dissipative dynamical systems. *Phys. Rev.* **31** 1109
[121] Grimmett G R and Stirzaker D R 2001 *Probability and Random Processes 3rd ed.* Oxford University Press, Oxford
[122] Gunawardena J 2003 Chemical reaction network theory for in-silico biologists. Notes available for download at http://vcp.med.harvard.edu/papers/crnt.pdf
[123] Halford S E and Marko J F 2004 How do site-specific DNA-binding proteins find their targets? *Nucl. Acid Res.* **32** 3040-3052
[124] Hanggi P, Grabert H, Talkner P and Thomas H 1984 Bistable systems: master equation versus Fokker–Planck modeling. *Phys. Rev. A* **29** 371-378
[125] Hanggi P, Talkner P and Borkovec M 1990 Reaction rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62** 251-341
[126] Hasty J, McMillen D and Collins J J 2002 Engineered gene circuits *Nature* **420** 224-230
[127] Heymann M and Vanden-Eijden E 2008 The geometric minimum action method: a least action principle on the space of curves. *Comm. Pure Appl. Math.* **61** 1052-1117
[128] Hillen T and Othmer H 2000 The diffusion limit of transport equations derived from velocity-jump processes. *SIAM J. Appl. Math.* **61** 751-775
[129] Hinch R and Chapman S J 2005 Exponentially slow transitions on a Markov chain: the frequency of calcium sparks. *Eur. J. Appl. Math.* **16** 427-446
[130] Hopfield J J 1984 Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA* **81** 3088-3092
[131] Horn F J M 1972 Necessary and sufficient conditions for complex balancing in chemical kinetics *Arch. Rat. Mech. Anal.* **49** 172-186
[132] Horn F J M and Jackson R 1972 General mass action kinetics. *Arch. Rat. Mech. Anal.* **47** 81-116
[133] Hu T, Grossberg AY, Shklovskii B II 2006 How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys J.* **90**, 2731-2744
[134] Huber G A and Kim S 1996 Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **70** 97-110
[135] Hufton P G, Lin Y T, Galla T and McKane A J 2016 Intrinsic noise in systems with switching environments *Phys. Rev. E* **93** 052119
[136] Hunter G A M, Guevara Vasquez F and Keener J P 2013 A mathematical model and quantitative comparison of the small RNA circuit in the *Vibrio harveyi* and *Vibrio cholerae* quorum sensing systems. *Phys. Biol.* **10** 046007

[137] Isaacson S and Peskin C 2006 Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations. *SIAM J. Sci. Comp.* **28**, 47-74
[138] Isaacson S A 2009 The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target. *SIAM J. Appl. Math.* **7**, 77-111
[139] Isaacson S A, McQueen D M and Peskin C S 2011 The influence of volume exclusion by chromatin on the time required to find specific DNA binding sites by diffusion. *Proc. Natl. Acad. Sci. U. S. A.* **108** 3815-3820
[140] Jacob F and Monod J 1961 Genetic regulatory mechanisms in the synthesis of proteins *J. Mol. Biol.* **3** 318-356
[141] Jahnke T and Kreim M 2012 Error bounds for piecewise deterministic processes modeling stochastic reaction systems. *Multiscale Model. Simul.* **10** 1119-1147
[142] James S, Nilsson P, James G, Kjelleberg S and Fagerstrom T 2001 Luminescence control in the marine bacterium Vibrio fischeri: an analysis of the dynamics of lux regulation. *J. Mol. Biol.* **296** 1127-1137
[143] Jose J V and Saletan E J 2013 *Classical dynamics: a contemporary approach* Cambridge University Press, Cambridge
[144] Jung S A, Hawver L A and Ng W-L 2016 Parallel quorum sensing signaling pathways in *Vibrio cholerae Curr. Genet.* **62** 255.
[145] Kaern M, Elston T C, Blake W J and Collins J J 2005 Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genetics* **6** 451-464
[146] Kamenev A, Meerson B and Shklovskii B 2008 How colored environmental noise affects population extinction *Phys. Rev. E* **101** 268103
[147] Karmakar R and Bose I 2004 Graded and binary responses in stochastic gene expression *Phys. Biol.* **1**197-204
[148] Keener J P and Newby J M 2011 Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. *Phy. Rev. E* **84** 011918
[149] Keizer J 1982 Nonequilibrium statistical thermodynamics and the effect of diffusion on chemical reaction rates. *J. Phys. Chem.* **86** 5052-5067
[150] Keller E and Segel L 1970 Initiation of slime mold aggregation viewed as an instability. *J. Theoret. Biol.* **26**, 399-415
[151] Kepler T B and Elston T C 2001 Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.* **81** 3116-3136
[152] Khasin M, Dykman M I and Meerson B 2010 Speeding up disease extinction with a limited amount of vaccine*Phys. Rev. E* **81** 051925
[153] Kifer Y 2009 *Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging* Memoirs of the AMS **201** issue 944
[154] Klapper I and Dockery J 2010 Mathematical description of microbial biofilms. *SIAM Rev.* **52**
[155] Kolomeisky A B 2011 Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* **13** 2088-2095
[156] Kumar N, Singh A and Kulkarni R V 2015 Transcriptional bursting in gene expression: analytical results for general stochastic models. *PLoS Comp. Biol.* **11** e1004292
[157] Koshland D E, Goldbeter A, and Stock J B 1982 Amplification and adaptation in regulatory and sensory systems. *Science* **217** 220-225
[158] Kurtz T G 1976 Limit theorems and diffusion approximations for density dependent Markov chains. *Math. Prog. Stud.* **5** 67-78
[159] Kurtz T G 1980 Representations of Markov processes as multiparameter changes. *Ann. Prob.* **8** 682-715
[160] Kussell E and Leibler S 2005 Growth and information in fluctuating environments. Science **309** 2075-2078
[161] Laing C and Lord G 2009 *Stochastic methods in neuroscience* Oxford University Press, Oxford
[162] Lawley S D, Mattingly J C and Reed M C 2015 Stochastic switching in infinite dimensions with applications to random parabolic PDE. *SIAM J. Math. Anal.* **47** 3035-3063
[163] Lenz D D H, Mok K C, Lilley B N, Kulkarni R V, Wingreen N S and Bassler B L 2004 The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae Cell* **118** 69-82
[164] Levien E and Bressloff P C 2016 A stochastic hybrid framework for obtaining statistics of many random walkers in a switching environment *Multiscale Model. Simul.* **14** 1417-1433
[165] Levine J, Kueh H Y and Mirny L 2007 Intrinsic fluctuations, robustness, and tunability in signaling cycles. *Biophys. J.* **92**, 4473-4481
[166] Li G W, Berg O G and Elf J 2009 Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **4**, 294-297

[167] Liu L, Kashyap B R K and Templeton J G C 1990 On the GI X /G/∞ system. *J. Appl. Prob.* **27** 671-683
[168] Lohmar I and Meerson B 2011 Switching between phenotypes and population extinction. *Phys. Rev. E* **84** 051901
[169] Lohmiller W and Slotine J J E 1998 On contraction analysis of nonlinear systems. *Automatica* **34** 683-696
[170] Loverdo C, Benichou O, Moreau M and Voituriez R 2008 Enhanced reaction kinetics in biological cells. *Nat. Phys.* **4** 134-137
[171] Mackey M C and Tyran-Kaminska M 2008 Dynamics and density evolution in piecewise deterministic growth processes. *Ann. Polon. Math.* **94** 111-129
[172] Mackey M C, Tyran-Kaminska M and Yvinec R 2013 Dynamic behavior of stochastic gene expression in the presence of bursting. *SIAM J. Appl. Math.* **73** 1830-1852
[173] Maheshri N and O'Shea E K 2007 Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36** 413-434
[174] Maier R S and Stein D L 1997 Limiting exit location distribution in the stochastic exit problem *SIAM J. Appl. Math.* **57** 752-790
[175] Marquez-Lago T T and Burrage K 2007 Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics. *J Chem Phys* **127** 104101
[176] Matkowsky B J and Schuss Z 1977 The exit problem for randomly perturbed dynamical systems *SIAM J. Appl. Math.* **33** 365-382
[177] Knessl C, Matkowsky B J, Schuss Z and Tier C 1985 An asymptotic theory of large deviations for Markov jump processes. *SIAM J. Appl. Math.* **46** 1006-1028
[178] McAdams H H and Arkin A 1997 Stochastic mechanisms in gene expression *Proc. Natl. Acad. Sci. USA* **94** 814-819
[179] Mina P, di Bernardo M, Savery N J and Tsaneva-Atanasova K 2013 Modelling emergence of oscillations in communicating bacteria: a structured approach from one to many cells. *J. Roy. Soc. Interface* **10** 20120612
[180] Mirny L, Slutsky M, Wunderlich Z, Tafvizi A, Leith J and Kosmrlj A 2009 How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A* **42** 434013
[181] Miyashiro T and Ruby E G 2012 Shedding light on bioluminescence regulation in Vibrio fischeri. *Mol.Microbiol.* **84** 795-806
[182] De Monte S, Ovidio F, Dano S and Sorensen P G 2007 Dynamical quorum sensing: Population density encoded in cellular dynamics. *Proc. Natl. Acad. Sci. USA* **104** 18377-18381
[183] Morris C and Lecar H 1981 Voltage oscillations in the barnacle giant muscle fiber. *J. Biophys.* **35** 193-213
[184] Muller J, Kuttler C, Hense B A, Rothballer M and Hartmann A 2006 Cell-cell communication by quorum sensing and dimension- reduction *J. Math. Biol.* **53** 672-702
[185] Muller J and Uecker H 2013 Approximating the dynamics of communicating cells in a diffusive medium by ODEs - homogenization with localization *J. Math. Biol.* **67** 1023-1065
[186] Muller J, Hense B A, Fuchs T M, Utz M and Potzsche Ch 2013 Bet-hedging in stochastically switching environments. *J. Theor. Biol.* **336** 144-157
[187] Naeh T, Klosek M M, Matkowsky B J and Schuss Z 1990 A direct approach to the exit problem. *SIAM J. Appl. Math.* **50** 595-627
[188] Newby J M and Bressloff P C 2010 Quasi-steady state reduction of molecular-based models of directed intermittent search. *Bull Math Biol* **72** 1840-1866
[189] Newby J M and Bressloff P C 2010 Local synaptic signalling enhances the stochastic transport of motor-driven cargo in neurons. *Phys. Biol.* **7** 036004
[190] Newby J M and Keener J P 2011 An asymptotic analysis of the spatially inhomogeneous velocity-jump process. *SIAM Multiscale Model. Simul.* **9** 735-765
[191] Newby J M 2012 Isolating intrinsic noise sources in a stochastic genetic switch. *Phys. Biol.* **9** 026002
[192] Newby J M and Chapman S J. 2014 Metastable behavior in Markov processes with internal states. *J. Math Biol.* **69** 941-976
[193] Newby J M, Bressloff P C and Keeener J P 2013 Breakdown of fast-slow analysis in an excitable system with channel noise. *Phys. Rev. Lett.* **111** 128101
[194] Newby J M 2014 Spontaneous excitability in the Morris–Lecar model with ion channel noise. *SIAM J. Appl. Dyn. Syst.* **13** 1756-1791
[195] Newby J M 2015 Bistable switching asymptotics for the self regulating gene. *J. Phys. A* **48** 185001
[196] Ng W L and Bassler B L 2009 Bacterial quorum-sensing architectures. *Annu. Rev. Genet.* **43** 197-222

[197] Norman T M, Lord N D, Paulsson J and Losick R 2015 Stochastic switching of cell fate in microbes. *Annu. Rev. Microbiol.* **69**, 381-403
[198] Oshanin O, Lindenberg K, Wio H S and Burlatsky S 2009 Efficient search by optimized intermittent random walks. *J. Phys. A* **42**, 434008 (2009).
[199] Othmer H, Dunbar S and Alt W 1988 Models of dispersal in biological systems. *J. Math. Biol.* **26** 263-298
[200] Othmer H G and Hillen T 2002 The diffusion limit of transport equations II: Chemotaxis equations *SIAM J. Appl. Math.* **62** 1222-1250
[201] Papanicolaou G C 1975 Asymptotic analysis of transport processes. *Bull. Amer. Math. Soc.* **81** 330-392
[202] Paulsson J 2005 Models of stochastic gene expression. *Phys. Life Rev.* **2** 157-175
[203] Peliti L 1985 Path integral approach to birth-death processes on a lattice. *Journal de Physique* **46** 1469-1483
[204] Phillips R, Kondev J, Theriot J and Garcia H 2012 *Physical biology of the cell 2nd. ed.* Garland Science
[205] Popovic N, Marr C and Swain P S 2016 A geometric analysis of fast-slow models for stochastic gene expression. *J. Math. Biol.* **72** 87-122
[206] Ptashne M 2004 *A genetic switch* Cold Spring Harbor Laboratory Press; 3rd edition
[207] Pulkkinen O and Metzler R 2013 Distance matters: the impact of gene proximity in bacterial gene regulation *Phys. Rev. Lett.* **110** 198101
[208] Qian H 2003 Thermodynamic and kinetic analysis of sensitivity amplification in biological signal transduction. *Biophys. Chem.* **105** 585-593
[209] Qian H 2007 Phosphorylation energy hypothesis: open chemical systems and their biological functions. *Annu. Rev. Phys. Chem.* **58** 113-142
[210] Qian H 2012 Cooperativity in cellular biochemical processes: Noise-enhanced sensitivity, fluctuating enzyme, bistability with nonlinear feedback, and other mechanisms for sigmoidal responses *Annu. Rev. Biophys.* **41** 179-204
[211] Raj A and van Oudenaarden A 2008 Nature, nurture, or chance: Stochastic gene expression and its consequences *Cell* **135** 216-226
[212] Reed M C, Venakides S and Blum J J 1990 Approximate traveling waves in linear reaction-hyperbolic equations. *SIAM J. Appl. Math.* **50** 167-180
[213] Reinitz J and Vaisnys J R 1990 Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity *J. Theor. Biol.* **145** 295-318
[214] Ribeiro A S 2008 Dynamics and evolution of stochastic bistable gene networks with sensing in fluctuating environments. *Phys. Rev. E* **78** 061902
[215] Richter P H and Eigen M 1974 Diffusion-controlled reaction rates in spheroidal geometry: application to repressor-operator association and membrane bound enzymes. *Biophys. Chem.* **2** 255-263
[216] Roberts E, Be'er S, Bohrer C, Sharma R and Assaf A 2015 Dynamics of simple gene-network motifs subject to extrinsic fluctuations *Phys. Rev. E* **92** 062717
[217] Roma D M, O'Flanagan R A, Ruckenstein A E and Sengupta A M 2005 Optimal path to epigenetic switching *Phys. Rev. E* 011902
[218] Russo G and Slotine J J E 2010 Global convergence of quorum sensing network. *Phys. Rev. E* **82** 041919
[219] Salsa S 2009 *Partial differential equations in action* Springer
[220] Sanchez A, Choubey S and Kondev J 2013 Regulation of noise in gene expression. *Annu. Rev. Biophys.* **42** 469-491
[221] Sasai M and Wolynes P G 2003 Stochastic gene expression as a many-body problem. *Proc. Natl. Acad. Sci.* **100** 2374-2379
[222] Schuss Z 2010 *Theory and applications of stochastic processes: an analytical approach* Springer, New York
[223] Schwabe A, Rybakova K N and Bruggeman F J 2011 Transcription stochasticity of complex gene regulation models. *Biophys J* **103** 1152-1161
[224] Seger J and Brockman H 1987 What is bet-hedging? In *Oxford Surveys in Evolutionary Biology* vol. 4 pp. 182-211 Oxford University Press
[225] Shahrezaei V, Swain PS 2008 Analytical distributions for stochastic gene expression. Proc Natl Acad Sci USA **105**:17256-17261
[226] Sheinman M, Benichou O, Kafri Y and Voituriez R 2012 Classes of fast and specific search mechanisms for proteins on DNA. *Rep. Prog. Phys.* **75** 026601
[227] Shibata T and Fujimoto K 2005 Noisy amplification in ultrasensitive signal transduction *Proc. Natl. Acad. Sci. USA* **102** 331-336

[228] Smiley M W and Proulx S R 2010 Gene expression dynamics in randomly varying environments. *J. Math. Biol.* **61** 231-251
[229] Smits W K, Kuipers O P and Veening J W 2006 Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.* **4** 259-271
[230] Sneppen K 2014 *Models of life: Dynamics and regulation in biological systems.* Cambridge University Press, Cambridge.
[231] Song C, Phenix H, Abed V, Scott M, Ingalls B P, Kaern M, and Perkins T J 2010 Estimating the stochastic bifurcation structure of cellular networks. PLoS Comput. Biol. **6** e1000699
[232] Swain P S and Longtin A 2006 Noise in genetic and neural networks *Chaos* **16** 026101
[233] Swem L R, Swem D L, Wingreen N S and Bassler B L 2008 Deducing receptor signaling parameters from in vivo analysis: LuxN/AI-1 quorum sensing in Vibrio harveyi *Cell* **134** 461-473
[234] Tabareau N, Slotine J J and Pham Q-C 2010 How synchronization protects from noise. *PLoS Comput. Biol.* **6** e1000637
[235] Thattai M and van Oudenaarden A 2001 Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **98** 8614-8619
[236] Thattai M and van Oudenaarden A 2004 Stochastic gene expression in fluctuating environments. *Genetics* **167** 523-530
[237] Thomas P, Popovic N and Grima R 2014 Phenotypic switching in gene regulatory networks *Proc. Natl. Acad. Sci. USA* **111** 6994-6999
[238] Touchette H 2008 The large deviation approach to statistical mechanics. *Phys. Rep.* **478** 1-69
[239] Touchette H 2012 A basic introduction to large deviations: Theory, applications, simulations. arXiv:1106.4146v3
[240] Tsimiring L S, Volfson D and Hasty J 2006 Stochastically driven genetic circuits. *Chaos* **16** 026103
[241] Tsimiring L S 2014 Noise in biology. *Rep. Prog. Phys.* **77** 026601
[242] van Kampen N G 1992 *Stochastic processes in physics and chemistry* North-Holland Amsterdam
[243] Varadhan S R S 1996 Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19** 261-286
[244] Vellela M and Qian H 2009 Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. *J. R. Soc. Interface* **6** 925-940
[245] Visco P, Allen R J, Majumdar S N and Evans M R 2010 Switching and growth for microbial populations in catastrophic responsive environments *Biophys. J.* **98** 1099-1108
[246] von Hippel P H and Berg O G 1989 Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675-678
[247] Walczak A M, Sasai M and Wolynes P G 2005 Self-consistent proteomic field theory of stochastic gene switches. *Biophys. J.* **88** 828-850
[248] Ward J P, King J R, Koerber A J, Williams P, Croft J M and Sockett R E 2001 Mathematical modelling of quorum-sensing in bacteria. *MA J. Math. Appl. Med.* **18**, 263-292
[249] Warmflash A, Bhimalapuram P and Dinner A R 2007 Umbrella sampling for nonequilibrium processes. *J. Chem. Phys.* **127** 154112
[250] Waters C M and Basser B L 2005 Quorum sensing: Cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21** 319-346
[251] Weake V M and Worlkman J L 2010 Inducible gene expression: diverse regulatory mechanisms. *Nat. Rev. Genet.* **11** 426-437
[252] Wei Y, Ng W-L, Cong J and Bassler B L 2012 Ligand and antagonist driven regulation of the Vibrio cholerae quorum-sensing receptor CqsS *Mol. Microbiol.* **83** 1095-1108
[253] Weinan E, Weiqing R and Vanden-Eijnden E 2002 String method for the study of rare events *Phys. Rev. B* **66** 052301
[254] Tran E J and Wente S R Dynamic nuclear pore complexes: Life on the edge 2006 *Cell* **125** 1041-1053
[255] Yang Y M, Austin R H and Cox E C 2006 Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.* **97** 048302
[256] Yu J, Xiao J, Ren X, Lao K and Xie X S 2006 Probing gene expression in live cells, one protein molecule at a time. *Science* **311** 1600-1603
[257] Yvinec R, Zhuge C, Lei J and Mackey M C 2014 Adiabatic reduction of a model of stochastic gene expression with jump Markov process *J. Math. Biol.* **68** 1051-1070
[258] Zeiser S, Franz U, Wittich O and Liebscher V 2008 Simulation of genetic networks modelled by piecewise deterministic Markov processes *IET Syst. Biol.* **2** 113-135
[259] Zhang B W, Jasnow D and Zuckerman D M 2010 The "weighted ensemble" path sampling

method is statistically exact for a broad class of stochastic processes and binning procedures. *J Chem Phys* **132** 054107

[260] Zuckerman M and Woolf T B 2000 Efficient dynamic importance sampling of rare events in one dimension. *Phys. Rev. E* **63** 016702