Isolating intrinsic noise sources in a stochastic genetic switch

Jay M Newby

Mathematical Institute, University of Oxford, 24-29 St. Giles’, Oxford OX1 3LB, UK
E-mail: newby@maths.ox.ac.uk

Received 17 October 2011
Accepted for publication 2 February 2012
Published 3 April 2012
Online at stacks.iop.org/PhysBio/9/026002

Abstract

The stochastic mutual repressor model is analysed using perturbation methods. This simple model of a gene circuit consists of two genes and three promotor states. Either of the two protein products can dimerize, forming a repressor molecule that binds to the promotor of the other gene. When the repressor is bound to a promotor, the corresponding gene is not transcribed and no protein is produced. Either one of the promotors can be repressed at any given time or both can be unrepressed, leaving three possible promotor states. This model is analysed in its bistable regime in which the deterministic limit exhibits two stable fixed points and an unstable saddle, and the case of small noise is considered. On small timescales, the stochastic process fluctuates near one of the stable fixed points, and on large timescales, a metastable transition can occur, where fluctuations drive the system past the unstable saddle to the other stable fixed point. To explore how different intrinsic noise sources affect these transitions, fluctuations in protein production and degradation are eliminated, leaving fluctuations in the promotor state as the only source of noise in the system. The process without protein noise is then compared to the process with weak protein noise using perturbation methods and Monte Carlo simulations. It is found that some significant differences in the random process emerge when the intrinsic noise source is removed.

1. Introduction

Random molecular interactions can have a profound effect on gene expression. Because the expression of a gene can be regulated by a single promotor, and because the number of mRNA copies and protein molecules is often small, deterministic models of gene expression can miss important behaviours. A deterministic model might show multiple possible stable behaviours, any of which can be realized depending on the initial conditions of the system. Different stable behaviours that depend on initial conditions allow for variability in response and adaptation to environmental conditions [1].

Although in some cases, noise from multiple sources can push the behaviour far from the deterministic model, here we focus on situation where the system fluctuates close to the deterministic trajectory (i.e. weak noise). Of particular interest is behaviour predicted by a stochastic model that is qualitatively different from its deterministic counterpart [2], even if the fluctuations are small.

Several interesting questions emerge when including stochastic effects in a model of gene expression. For example, what are the different sources of fluctuations affecting a gene circuit? Can noise be harnessed for useful purpose, and if so, what new functions can noise bring to the gene-regulation toolbox? One way in which noise can induce qualitatively different behaviour is by a rare sequence of random events that pushes the system from one of the stable deterministic behaviours to another, one that would never be realized in the deterministic model without changing the initial conditions. For example, if the deterministic model is bistable, fluctuations can cause the protein concentration to shift between the different metastable protein concentrations. This happens when fluctuations push the system past the unstable fixed point that separates two stable fixed points.

While often times a spontaneous change in gene expression might be harmful, it might also be beneficial. For example, in certain types of bacteria, a few individuals within a population enter a slow-growth state in order to resist exposure to antibiotics. In a developing organism, a population
of differentiating cells might first randomly choose between two or more expression profiles during their development and then later segregate into distinct groups by chemotaxis. In both examples, switching between metastable states leads to mixed populations of phenotypic expression [3]. This leads to the question of how cells coordinate and regulate different sources of biochemical fluctuations, or noise, to function within a genetic circuit.

In many cases, the genes within a given circuit are turned on and off by regulator proteins, which are often the gene products of the circuit. If a gene is switched on, its DNA is transcribed into one or more mRNA copies that are then translated into several protein molecules. Typically, the protein products form complexes with each other or with other proteins that bind to regulatory DNA sequences, or operators, to alter the expression state of a gene. For example, a repressor binds to an operator which blocks the promotor—the region of DNA that a polymerase protein binds to before transcribing the gene—so that the gene is turned off and no mRNA is transcribed. This feedback enables a cell to regulate gene expression, and often multiple genes interact within groups to form gene circuits.

Understanding how different noise sources affect the behaviour of a gene circuit and comparing this with how the circuit behaves with multiple noise sources is essential for understanding how a cell can use different sources of noise productively. Fluctuations arising from the biochemical reactions involving the DNA, mRNA and proteins are commonly classified as ‘intrinsic’ noise [4]. One important source of intrinsic noise is the chemical reactions governing mRNA transcription, protein translation and degradation of both mRNA and protein product. This type of noise is common among many of the biochemical reactions within a cell, and its effect is reduced as the number of reacting species within a given volume grows large. Another source of intrinsic noise is the fluctuating expression state of the genes within the circuit. Typically, there is only one or two copies of a gene within a cell, which means that thermal fluctuations that affect regulatory-protein activity have a significant effect on mRNA production. Here, we consider the situation where transitions in the behaviour of a gene circuit are primarily driven by fluctuations in the on/off state of its promoters and examine the effect of removing all other sources of noise.

Stochastic gene circuits are typically modelled using a discrete Markov process, which tracks the random number of mRNA and/or proteins along with the state of one or more promoters [5–9] (but see also [10]). Monte Carlo simulations using the Gillespie algorithm [33] can be used to generate exact realizations of the random process. The process can also be described by its probability density function, which satisfies a system of linear ordinary differential equations known as the Master equation. The dimension of the Master equation is the number of possible states the system can occupy and can be quite large, leading to the problem of dimensionality when analysing the Master equation directly. However, for the problem considered here, the full solution to the Master equation is not necessary in order to understand metastable transitions.

Biological considerations motivate the following question. What percentage of a population of cells will exhibit a metastable transition within a given timeframe? If a spontaneous transition is harmful to the cell, one expects that reaction rates and protein/DNA interactions should evolve so that transition times are likely to be much larger than the lifetime of the cell. On the other hand, if a spontaneous transition is functional, transition times should be tuned to achieve the desired population in which the transition occurs. In either case, the key quantity of interest is the distribution of transition times between metastable states.

Except for a few special cases, exact results, even for the mean transition time, are not possible and approximation techniques or Monte Carlo simulations must be used. However, because rare events typically involve long simulation times, Monte Carlo simulations are computationally expensive to perform, leaving perturbation analysis ideally suited for the task. Past studies of metastable transitions, where perturbation methods were applied to the Master equation, have used a simplifying assumption so that the state of the promotor is not accounted for explicitly [6, 7]. The assumption is that proteins are produced in ‘bursts’ during which one or more mRNA copies are translated to rapidly produce many proteins. In these models, production bursts occur as instantaneous jumps, with a predefined distribution determining the number of proteins produced during a given burst. More recently, Assaf and co-workers analysed a model where the on/off state of a single stochastic promotor is accounted for explicitly, and mRNA copies are produced stochastically at a certain rate when the promotor is turned on [11]. However, the case where the model contains an arbitrary number of promotors or promotor states has not been addressed, and as we show in this paper, accounting for even just three promotor states is nontrivial. Similar asymptotic methods have also been developed to study metastable transitions in continuous Markov processes [12–17], but cannot be applied to a discrete chemical reaction system because continuous approximations—such as the system-size expansion or Kramers–Moyal expansion—of discrete Markov processes do not, in general, accurately capture transition times [18]. Another source of difficulty arises from isolating promotor state fluctuations as the only source of noise: the resulting state space of the Markov process is both continuous and discrete.

After removing all sources of intrinsic noise except for the fluctuating promotor, the protein levels change deterministically and continuously, and the promotor’s state jumps at exponentially distributed random times. The random jumps in the promotor’s state makes the protein levels appear random, even though they are only responding deterministically to changes in the promotor state. Such random processes are sometimes called hybrid systems, or piecewise deterministic [19, 20]. Here, we refer to it as the quasi-deterministic (QD) process because we are taking part of the randomly fluctuating discrete state of the system (the number of protein molecules) and replacing it with a deterministically changing continuous state. Recently, we have developed asymptotic methods—similar to methods for the Master equation or the Fokker–Planck equation—to study
metastable transitions in Markov processes with both discrete and continuous state spaces [21, 22]. However, these methods cannot account for two or more continuous state variables, and therefore cannot be applied to a gene circuit model with more than one protein products.

In this paper, we develop new perturbation methods so that we can study metastable transitions in genetic circuits driven by promoter fluctuations. These methods are based on previous theory developed for one-dimensional velocity jump processes and are generalized to account for the multiple continuous states representing multiple protein products of the genetic circuit. They also fit within a larger framework of methods to study metastable transitions in continuous Markov processes [17] and in certain discrete jump Markov processes [23–29, 6, 7, 11]. For illustration, we use a simple model known as the ‘mutual repressor’ model [30], a gene circuit model with two genes, two promotors and three promoter states. Although our example considers only three promoter states, the methods presented are general and can account for an arbitrary number of promoter states. For a range of parameter values, the deterministic limit of the mutual repressor model is bistable, having two stable fixed points separated by an unstable saddle point. For the stochastic model, the deterministic forces create two confining wells surrounding each stable fixed point, separated by a stability barrier along the separatrix that contains the unstable saddle. This geometric interpretation—called the ‘stability landscape’—is useful because of its intuitive appeal and is defined as the logarithm of the stationary probability density function. We approximate the first exit time density for the random process to escape over the stability barrier from one of the stability wells to the other. Using this model, we seek to answer the following question. How does the random process change when protein noise is removed, leaving the state of the promotor as the only source of randomness? That is, are there any qualitative differences in the behaviour of the system without other sources of intrinsic noise?

The paper is organized as follows. In section 2, the mutual repressor model is presented along with its reduction to the QD process, and then in section 3, perturbation methods for estimating the exit time density are applied to the QD process. For comparison, the stability landscape is also computed for the full process that includes fluctuations in protein production/degradation. Finally, results are presented in section 4, and the QD process is compared to the full process, using analytical/numerical approximations and Monte Carlo simulations.

2. Mutual repressor model

The mutual repressor model [30] is a hypothetical gene circuit consisting of a single promotor driving the expression of two genes: N and M. Each protein product can dimerize and bind to the promotor to repress the expression of the other. When no dimer is bound to the promotor, both genes are expressed equally. Thus, the two promotors can (we refer to the state of both promotors as the ‘promotor state’) be in one of three states: N bound to a repressor (O_0), neither gene repressed (O_1), or M bound to a repressor (O_2). Note that for simplicity we ignore the state where both genes are repressed. Let the number of protein products of genes N and M be n and m, respectively. It is assumed that the mRNA and protein production steps can be combined into a single protein production rate and that the dimerization reaction is fast so that it can be taken to be in quasi-steady state. We then have the following transition between the three promotor states:

\[
O_0 \xrightarrow{\rho} O_1 \xrightarrow{\beta \delta} O_2.
\]  

(2.1)

where \( \kappa \) is a rate and \( \beta \) is a nondimensional dissociation constant. Protein N (M) is produced at a rate \( \alpha \) while the promotor is in states \( O_{0,0} \) (\( O_{1,2} \)), and both proteins are degraded at a rate \( \delta \) in all three promotor states.

Introduce the probability density function, \( p_k(n, m, t) \), where \( k = 0, 1, 2 \) is the promotor state. It satisfies the Master equation

\[
\frac{d}{dt} p(n, m, t) = [ \mathbb{A} + \mathbb{W} ] p,
\]  

(2.2)

where

\[
p(n, m, t) \equiv (p_0(n, m, t) \quad p_1(n, m, t) \quad p_2(n, m, t))^T,
\]

(2.3)

and

\[
\mathbb{A} = \kappa \begin{bmatrix}
-\beta & n(n-1) & 0 \\
\beta & -n(n-1) - m(m-1) & \beta \\
0 & m(m-1) & -\beta
\end{bmatrix}
\]  

(2.4)

is the transition-rate matrix for the promotor state. The diagonal matrix \( \mathbb{W} \), describing protein production/degradation, has elements

\[
W_0 = D + \alpha (E^{-}_n - 1),
\]

(2.5)

\[
W_1 = D + \alpha (E^{-}_n + E^{-}_m - 2),
\]

(2.6)

\[
W_2 = D + \alpha (E^{-}_m - 1)
\]

(2.7)

with

\[
D = \delta [ (E^+_n - 1)n + (E^+_m - 1)m ].
\]

(2.8)

The shift operators \( E^+_j \) are defined according to

\[
E^+_j f(j) = f(j+1).
\]

(2.9)

We now introduce the nondimensional/rescaled variables \( t \to t\delta, x = ny, y = my \), where \( \gamma = \delta / \alpha \) is the average protein number when the promotor is switched on. Then, the Master equation (2.2) for the rescaled probability density, \( p(n, m, t) \to \rho(x, y, t) \), becomes

\[
\frac{d}{dt} \rho(x, y, t) = \left[ -\frac{1}{\epsilon} A_y + \frac{1}{\gamma} W \right] \rho,
\]

(2.10)

with dimensionless parameters \( b = \frac{\beta \delta}{\alpha} \) and \( \epsilon = \frac{\delta}{\alpha \gamma} \). The matrices are given by

\[
A_y = \begin{bmatrix}
-\beta & x(x+y) & 0 \\
b & -x(x+y) - y(y+y) & b \\
0 & y(y+y) & -b
\end{bmatrix}
\]

(2.11)

and

\[
W = (e^{\delta x} - 1)x + (e^{\delta y} - 1)yI + \text{diag}(e^{-\delta x} - 1, e^{-\delta x} + e^{-\delta y} - 2, e^{-\delta y} - 1).
\]

(2.12)
and replace the shift operators $E$ leaving the lower basin of attraction to reach the separatrix. The blue curve shows a stochastic trajectory and the green circles show the stable fixed points, the red circle shows the unstable saddle. The green curve shows nullclines and the grey curve shows the $x$-nullcline. 

The operators $e^{\pm \delta x}$ and $e^{\pm \delta y}$ are defined in terms of Taylor series expansions (in small $\gamma$),

$$e^{\pm \delta x} f(x) = \sum_{n=0}^{\infty} \frac{(\pm \gamma)^n}{n!} \frac{\partial^n f}{\partial x^n} = f(x \pm \gamma),$$

and replace the shift operators $e^{\pm \delta y}$.

Assume that $\gamma \ll 1$ is a small parameter, so that there is a large average number of proteins, and assume also that the parameter $\varepsilon \ll 1$ is small, which reflects rapid switching between promotor states compared to the rate of protein production/degradation. Because we have two small parameters in our system, $\gamma$ and $\varepsilon$, when pursuing an asymptotic solution, we must carefully consider how the limits $\varepsilon \to 0$ and $\gamma \to 0$ are taken, or more practically, how large $\varepsilon$ is compared to $\gamma$. The fluctuations in the promotor state are controlled by $\gamma$, and in the limit $\gamma \to 0$, the transitions are infinitely fast so that the promotor behaves deterministically. The fluctuations in protein levels are controlled by $\varepsilon$, and in the limit $\gamma \to 0$ the protein production/degradation behaves deterministically. Since we are concerned primarily with rare transitions driven by promotor fluctuations and not by fluctuations in the protein production/degradation reaction, we assume that $\gamma \ll \varepsilon$ (i.e. $\frac{\gamma^2}{\varepsilon^2} \ll 1$).

Taking both limits, $\varepsilon \to 0$ and $\gamma \to 0$, yields the fully deterministic dynamics

$$\dot{x} = f(x, y), \quad \dot{y} = f(y, x),$$

where

$$f(x, y) = \frac{1}{1 + \frac{2}{3} \frac{x^2}{y^3} + \frac{2}{3} \frac{y^2}{x^3}} - x.$$  

Note the symmetry in the problem: it is unchanged if we exchange $x \leftrightarrow y$. Dynamically, the system is bistable for $0 < b < b_c$. At $b = b_c = 4/9$ there is a saddle-node bifurcation, and for $b > b_c$ there is a single stable fixed point. We consider only the bistable case. In figure 1 the nullclines and fixed points are shown.

The two stable fixed points are located near the corners and the unstable saddle point is located along the separatrix. Arrows show the eigenvectors of the Jacobian with their direction determined by the sign of the eigenvalues, all of which are real. A stochastic trajectory that starts at the lower stable fixed point remains nearby for a long period of time until a rare sequence of jumps carries it to the separatrix. Because the separatrix is the stable manifold, trajectories are most likely to exit near the unstable saddle point.

To remove protein noise from the system, consider the limit $\gamma \to 0$, with $\varepsilon \ll 1$ fixed, so that the protein production/degradation process is deterministic within each promotor state, while the promotor state remains random. The Master equation (2.10) converges to the differential Chapman–Kolmogorov (CK) equation

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}(F \rho) - \frac{\partial}{\partial y}(G \rho) + \frac{1}{\varepsilon} A \rho, \quad (x, y) \in (0, 1)^2,$$

where

$$F(x, y) = \text{diag}(1 - x, 1 - x, -x),$$

$$G(x, y) = \text{diag}(-y, 1 - y, 1 - y),$$

and

$$A = \begin{bmatrix} -b & x^2 & 0 \\ b & -x^2 & y^2 \\ 0 & y^2 & -b \end{bmatrix}.$$  

For more on the CK equation and its derivation see [32].

3. Perturbation-based exit time theory for rare events

The focus of the remaining analysis is to obtain an accurate approximation of the first exit time density function (FETD) for the QD process to evolve from one metastable state to other. To obtain the FETD, we supplement an absorbing boundary condition to the governing equation along the separatrix,

$$\Gamma = \{(x, y) \in (0, 1)^2 : x = y\},$$

of the deterministic dynamics, which is the barrier the process must surmount in order to transition to the other metastable state. For the CK equation (2.16), the absorbing boundary condition is

$$\rho_n(x, y) = 0, \quad (x, y) \in \Gamma,$$

for all $n = 0, 1, 2$ such that $(F_{n, n}, G_{n, n}) \cdot \hat{n} < 0$, where $\hat{n}$ is the unit vector normal to $\Gamma$ pointing into the domain. Then, the domain for the QD process with the absorbing boundary is given by

$$\mathcal{D} = (0, 1)^2 \cap \{x \leq y\}.$$  

Note that the choice of the lower triangular region (instead of the upper triangular region with $x \geq y$) is arbitrary due to the symmetry in the problem.

To see how the absorbing boundary on $\Gamma$ sets up the exit time problem, define $T$ to be the random time at which the separatrix is reached for the first time, given that the process starts at the stable fixed point $(x_0, y_0) \in \mathcal{D}$. Consider the survival probability

$$S(t) \equiv \sum_{n=0}^{3} \int_{\mathcal{D}} \rho_n(x, y, t) dA,$$
which is the probability that $t < T$. The FETD (or probability density function for $T$) is then

$$\mathcal{F}(t) = -\frac{dS}{dt}. \quad (3.5)$$

The FETD for the QD process can be approximated using perturbation methods as follows. Suppose we have a CK equation of the form

$$\partial_t \rho = -\mathcal{L}_\epsilon \rho,$$  \quad (3.6)

where $\mathcal{L}_\epsilon$ is a linear operator acting on the continuous and discrete state variables of the density function. In the case of the CK equation (2.16), we have

$$\mathcal{L}_\epsilon = \frac{\partial}{\partial x} (F \rho) + \frac{\partial}{\partial y} (G \rho) - \frac{1}{\epsilon} A \rho, \quad (x, y) \in \mathcal{D}. \quad (3.7)$$

Assume that $\mathcal{L}_\epsilon$ has a complete set of eigenfunctions, $\{\phi_i\}$, so that the solution can be written as

$$\rho(x, y, t) = \sum_{j=0}^{\infty} c_j \phi_j(x, y) e^{-\lambda_j t}, \quad (3.8)$$

for some constants $c_j$, and that all of the eigenvalues, $\lambda_j$, are non-negative. Assume further that if we impose a reflecting boundary condition on $\Gamma$ then the principal eigenvalue, $\lambda_0$, is the only zero eigenvalue (i.e. $\lambda_0 = 0$) and $\phi_0$ is the stationary density of the process (after appropriate normalization). Furthermore, we assume that the stationary density is exponentially small on the boundary. Then, if instead we place an absorbing boundary condition on $\Gamma$, the stationary density no longer exists and the principal eigenvalue is exponentially small in $\epsilon$, while the remaining eigenvalues are much larger so that the solution resembles the stationary density after some initial transients [12]. It is this difference in timescales that we exploit to approximate the FETD.

A universal feature of the FETD for rare events is its exponential form, which follows from the separation of timescales; that is, it is approximately exponentially distributed because the time dependence is $e^{-\lambda_0 t}$, for $\lambda_0 t \gg 1$. Indeed, the FETD (3.5) is

$$\mathcal{F}(t) \sim \lambda_0 e^{-\lambda_0 t}, \quad \text{for } \lambda_0 t \gg 1. \quad (3.9)$$

Thus, for large times the FETD is completely characterized by the principal eigenvalue, $\lambda_0$. The mean exit time is simply $1/\lambda_0$, which means that the eigenvalue also has the physical interpretation of the rate at which metastable transitions occur.

To obtain an approximation of $\lambda_0$, we use a spectral projection method that makes use of the adjoint operator $\mathcal{L}^*_\epsilon$. Consider the adjoint eigenfunctions $\{\xi_j\}, j = 0, 1, \ldots$ satisfying

$$\mathcal{L}^*_\epsilon \xi_j = \lambda_j \xi_j, \quad (3.10)$$

and $\langle \phi_i, \eta_j \rangle = \delta_{i,j}$ so that the two sets of eigenfunctions are biorthogonal. Now suppose that $\phi_0$ is well approximated by the stationary density, $\bar{\rho}$. By the application of the divergence theorem, the adjoint operator is such that

$$\langle \bar{\rho}, \mathcal{L}^*_\epsilon \xi_0 \rangle = \langle \mathcal{L}_\epsilon \bar{\rho}, \xi_0 \rangle + \oint_{\mathcal{D}} \xi_0^* (F - G) \bar{\rho} \, ds. \quad (3.11)$$

where the boundary contribution is nonzero because $\bar{\rho}$ does not satisfy the absorbing boundary condition. Then, since $\mathcal{L} \bar{\rho} = 0$, the principal eigenvalue is

$$\lambda_0 = \frac{\int_{\mathcal{D}} \xi_0^* (F - G) \bar{\rho} \, ds}{\int_{\mathcal{D}} \xi_0^* \bar{\rho} \, ds}. \quad (3.12)$$

In the remainder of this section, we approximate $\bar{\rho}$ and $\xi_0$, which are then used in (3.12) to approximate $\lambda_0$.

### 3.1. Quasi-stationary distribution

In this section, we obtain an approximation to the stationary density, $\bar{\rho}(x, y)$, using a WKBo approximation method. We begin by illustrating the procedure for the QD process. That is, we seek an approximation of the solution to the equation

$$\left[ \frac{1}{\epsilon} - \frac{\partial}{\partial x} F - \frac{\partial}{\partial y} G \right] \bar{\rho}(x, y) = 0. \quad (3.13)$$

Consider the ansatz

$$\bar{\rho}(x, y) = (r_0(x, y) + \epsilon r_1(x, y)) \exp \left[ \frac{1}{\epsilon} \Phi(x, y) + k(x, y) \right], \quad (3.14)$$

where $r_{0,1}(x, y)$ are 3-vectors and both $\Phi(x, y)$ and $k(x, y)$ are scalar functions. Note that in other studies of gene regulation models where similar methods are used, the small parameter in the exponential is $\gamma$. This difference in scaling arises from the assumption that the metastable transitions are driven by fluctuations in the promotor state and not by fluctuations in the production or degradation of protein. Substituting (3.14) into (3.13) and collecting leading-order terms in $\epsilon$ yields

$$[A + p_1 F + p_2 G] r_0 = 0, \quad (3.15)$$

where

$$p_1 \equiv \frac{\partial \Phi}{\partial x}, \quad p_2 \equiv \frac{\partial \Phi}{\partial y}. \quad (3.16)$$

Note that if the CK equations were scalar valued rather than matrix valued (as is the case for existing WKBo theory for exit time problems), the leading-order equation (analogous to (3.15)) obtained after applying the WKBo ansatz can be decoupled to obtain two separate equations for $\Phi(x, y)$ and the term analogous to $r_0$.

There are few interesting implications that arise because (3.15) is matrix valued. First, the vector $r_0$ (up to a normalization factor) is simply the nullspace of the matrix $M = [A + p_1 F + p_2 G]$, and we assume that it is normalized so that $\sum_{i=0}^{\infty} (r_0)_i = 1$. Using theorem 3.1 in [22], we can provide necessary and sufficient conditions for $r_0$ to be unique and positive. For any fixed $(x, y, p_1, p_2)$, there exists a unique vector $r_0 > 0$, satisfying (3.15) if and only if the diagonal matrix $H \equiv p_1 F + p_2 G$ is such that at least two of its elements have opposite signs. That is, there exist $i, j$, with $i \neq j$, such that $H_{ii} H_{jj} < 0$. It is interesting to speculate that if the solution $(p_1(x, y), p_2(x, y))$ to (3.15) is substituted into $M$ then this requirement satisfies for all $(x, y) \in (0, \infty)^2$.

However, this is not necessarily the case, which means that the stationary density is restricted to a subdomain where $r_0(x, y) > 0$. It is obvious that the protein levels must be bounded within the domain $(x, y) \in (0, 1)^2$ when protein
production/degradation is deterministic. That is, while a gene remains in the un-repressed state, the protein level tends towards the mean value \((n, m = 1/\gamma \text{ or } x, y = 1)\), but never exceeds it since protein levels do not fluctuate. However, it is not as obvious that the total amount of protein is further bounded so that \(1 < x + y < 2\), which means that the domain is further restricted to the upper triangular portion of the unit square. Consequently, once a trajectory enters this domain, it remains for all the time and cannot escape. To show this, we simply need to look at the rate of protein production/degradation for each state normal to the line \(y = 1 - x\). This gives the rate for each of the promoter states \((x, on, both\ on, y\ on)\) when both protein levels satisfy \(y = 1 - x\). These rates are given by the diagonal components of the matrix \(F(x) + G(1 - x) = \text{diag}(0, 1, 0)\), where \(F\) and \(G\) are defined in (2.17). It is evident that when no repressor is bound and both proteins are produced, the flux across this line is in the positive direction, and when one repressor is bound, there is no flux across this line.

Although we cannot decouple (3.15) to obtain an equation for \(r_0\) that does not depend on \(\Phi\), we can obtain a nonlinear partial differential equation for \(\Phi\) that is independent of \(r_0\) by taking the determinant of \(M\) to obtain
\[
\mathcal{H}(x, y, p_1, p_2) \equiv \det[A(x, y) + p_1 F(x, y) + p_2 G(x, y)] = 0.
\]
(3.17)

The function \(\mathcal{H}\) is referred to as the Hamiltonian for the system (see Appendix D for an explicit formula), due to the similarity to classical Hamiltonian dynamics. An implicit assumption,
\[
\frac{\partial p_1}{\partial y} - \frac{\partial p_2}{\partial x} = 0,
\]
(3.18)
is present to ensure that \((p_1, p_2)\) is the gradient of the scalar field \(\Phi\). The above PDE can be reduced by the method of characteristics to a system of ordinary differential equations (see [31] p 360),
\[
\begin{align*}
\dot{x} &= \frac{\partial \mathcal{H}}{\partial p_1}, \\
\dot{y} &= \frac{\partial \mathcal{H}}{\partial p_2}, \\
\dot{p}_1 &= -\frac{\partial \mathcal{H}}{\partial x}, \\
\dot{p}_2 &= -\frac{\partial \mathcal{H}}{\partial y},
\end{align*}
\]
(3.19)
where each variable is parameterized by \(t\) (which should not be confused with physical time). The above system of ordinary differential equations is supplemented with an equation for the stability landscape
\[
\Phi \equiv \frac{\partial \mathcal{H}}{\partial x} + \frac{\partial \Phi}{\partial y} = p_1 \frac{\partial \mathcal{H}}{\partial p_1} + p_2 \frac{\partial \mathcal{H}}{\partial p_2},
\]
(3.20)
and solutions specify \(\Phi\) along the curves \((x(t), y(t))\), called rays. A family of rays is defined by specifying Cauchy data
\[
x(0) = x_0(\theta), \quad y(0) = y_0(\theta), \quad p_1(0) = p_{1,0}(\theta), \quad p_2(0) = p_{2,0}(\theta),
\]
(3.21)
along a curve parameterized by \(\theta\).

One of the difficulties of using the method of characteristics on a nonlinear PDE like (3.17) is determining appropriate Cauchy data. At the stable fixed point, the value of each of the variables is known (i.e. \(p_1 = p_2 = 0\) and \(x = x_*, y = y_*\)); however, data at a single point cannot hope to generate a family of rays. Therefore, data must be specified on a small ellipse surrounding the fixed point. Expanding \(\Phi\) in a Taylor series around the fixed point yields the quadratic form
\[
\Phi(x, y) \approx \frac{1}{2} r^T Z r, \quad r = \left( x - x_*, y - y_* \right),
\]
(3.22)
as its leading-order term, where \(Z\) is the Hessian matrix,
\[
Z \equiv \begin{bmatrix} \frac{\partial^2 \Phi}{\partial x^2} & \frac{\partial^2 \Phi}{\partial x \partial y} \\ \frac{\partial^2 \Phi}{\partial x \partial y} & \frac{\partial^2 \Phi}{\partial y^2} \end{bmatrix}.
\]
(3.23)
Cauchy data are specified on the ellipse
\[
\frac{1}{2} r(\theta) Z r(\theta)^T = \omega,
\]
(3.24)
for some suitably small \(\omega \ll 1\). In practice, \(\omega\) must be small enough to generate accurate numerical results, but large enough so that trajectories can be generated to cover the domain. On the elliptical contour, the initial values for \(p_1, p_2\) are
\[
\begin{bmatrix} p_{1,0}(\theta) \\ p_{2,0}(\theta) \end{bmatrix} = Z \begin{bmatrix} x_0(\theta) - x_* \\ y_0(\theta) - y_* \end{bmatrix}.
\]
(3.25)
It can be shown [17] that the Hessian matrix is the solution to the algebraic Riccati equation
\[
ZBZ + ZC + C^T Z = 0,
\]
(3.26)
where
\[
B = \begin{bmatrix} \frac{\partial^2 \mathcal{H}}{\partial p_1 \partial p_1} & \frac{\partial^2 \mathcal{H}}{\partial p_1 \partial p_2} \\ \frac{\partial^2 \mathcal{H}}{\partial p_2 \partial p_1} & \frac{\partial^2 \mathcal{H}}{\partial p_2 \partial p_2} \end{bmatrix}, \quad C = \begin{bmatrix} \frac{\partial^2 \mathcal{H}}{\partial x \partial p_1} & \frac{\partial^2 \mathcal{H}}{\partial y \partial p_1} \\ \frac{\partial^2 \mathcal{H}}{\partial x \partial p_2} & \frac{\partial^2 \mathcal{H}}{\partial y \partial p_2} \end{bmatrix},
\]
(3.27)
evaluated at \(p_1 = p_2 = 0, x = x_*\) and \(y = y_*\). This equation can be transformed into a linear problem (in order to actually solve it) by making the substitution \(Q = Z^{-1}\) to obtain
\[
B + CQ + QC^T = 0.
\]
(3.28)

3.1.1. An equation for \(k(x, y)\). An equation for the scalar function \(k(x, y)\) is found by substituting (3.14) into (3.13) and keeping second-order terms in \(\epsilon\) to obtain
\[
[A + p_1 F + p_2 G] r_1 = \frac{\partial}{\partial x} (F r_0) + \frac{\partial}{\partial y} (G r_0)
\]
\[
- \left( \frac{\partial k}{\partial x} F - \frac{\partial k}{\partial y} G \right) r_0.
\]
(3.29)
For solutions \(r_1\) to exist, the Fredholm alternative theorem requires that for all \(I\) satisfying
\[
I^T [A + p_1 F + p_2 G] r_1 = 0,
\]
(3.30)
we must have that
\[
I^T \left[ \frac{\partial}{\partial x} (F r_0) + \frac{\partial}{\partial y} (G r_0) - \left( \frac{\partial k}{\partial x} F - \frac{\partial k}{\partial y} G \right) r_0 \right] = 0.
\]
(3.31)
Note that because \(r_0\) spans the right nullspace of \(A + p_1 F + p_2 G\), then the left nullspace is also one dimensional. After rewriting (3.31), we have the following PDE for \(k\):
\[
\frac{\partial k}{\partial x} (t^2 F r_0) + \frac{\partial k}{\partial y} (t^2 G r_0) = I^T \left( \frac{\partial}{\partial x} (F r_0) + \frac{\partial}{\partial y} (G r_0) \right).
\]
(3.32)
Although the solution to this equation can be formulated by the method of characteristics, it requires values of the
vectors $\mathbf{r}_0$ and $\mathbf{l}$, which in turn require the solution to the ray equations (3.19). Since rays must be integrated numerically in most cases, solving (3.32) along its own characteristics is impractical. Instead, we solve (3.32) numerically along the characteristic curves of (3.19). We leave the details of this to Appendix A.

3.1.2. Stability landscape with protein noise. The above analysis can be repeated to obtain a Hamiltonian system for the full/rescaled process (with protein noise), but a choice must be made for how the limits $\epsilon \to 0$ and $\gamma \to 0$ are taken. First consider the equation for the stationary density, $\tilde{\rho}$, of the full process (2.10):

$$\left[ \frac{1}{\epsilon} A + \frac{1}{\gamma} \mathbf{F} \right] \tilde{\rho}(x, y) = 0. \tag{3.33}$$

Here, the domain is the cone $0 < x < y < \infty$. As before, the stationary density is assumed to have the form

$$\tilde{\rho}(x, y) = \mathbf{r}(x, y) \exp \left[ \frac{1}{\epsilon} \Phi(x, y) \right], \tag{3.34}$$

where $\mathbf{r}$ is a 3-vector and $\Phi$ is a scalar function representing the stability landscape. Note that we have ignored higher order terms here because we seek only the stability landscape for comparison to the QD process. Substituting (3.34) into (3.33) does not lead to any meaningful equation at leading order unless we make an assumption about how the limit $\gamma \to 0$ is taken. There are two relevant cases: $\gamma = o(\epsilon)$ and $\gamma = O(\epsilon)$. In the former case, we recover the QD result (3.15), and in the latter case, collecting terms of leading order in $\epsilon$, with $\gamma = \varphi \epsilon$, yields

$$\left[ A + \frac{1}{\varphi} \mathbf{H} \right] \mathbf{r} = 0, \tag{3.35}$$

where $\mathbf{H} \equiv \text{diag}(H_0(x, y, p_1, p_2), H_1(x, y, p_1, p_2), H_2(x, y, p_1, p_2))$ and

$$H_0(x, y, p_1, p_2) \equiv h(x, y, p_1, p_2) + \frac{e^{-\varphi p_1} - 1}{\varphi}, \tag{3.36}$$

$$H_1(x, y, p_1, p_2) \equiv h(x, y, p_1, p_2) + \frac{e^{-\varphi p_1} - 1}{\varphi} + \frac{e^{-\varphi p_2} - 1}{\varphi}, \tag{3.37}$$

$$H_2(x, y, p_1, p_2) \equiv h(x, y, p_1, p_2) + \frac{e^{-\varphi p_1} - 1}{\varphi}, \tag{3.38}$$

$$h(x, y, p_1, p_2) \equiv \left( \frac{e^{-\varphi p_1} - 1}{\varphi} \right) x + \left( \frac{e^{-\varphi p_2} - 1}{\varphi} \right) y. \tag{3.39}$$

Thus, the Hamiltonian for the full process is

$$\mathcal{H}(x, y, p_1, p_2) \equiv \det [A(x, y) + \mathbf{H}(x, y, p_1, p_2)], \tag{3.40}$$

which we refer to as the full Hamiltonian. To obtain the stability landscape, we use the method of characteristics as outlined in the previous section.

The differences between the full process and the QD process are nicely illustrated by comparing their associated Hamiltonians. Note that the full Hamiltonian (3.40) is a transcendental function of $p_1$ and $p_2$, whereas the Hamiltonian for the QD process (3.17) is a cubic polynomial in $p_1$ and $p_2$. One can view this as a Taylor series expansion of the full Hamiltonian about $(p_1, p_2) = (0, 0)$. For this reason, the QD process—as an approximation for the full process with a small amount of protein noise—is only valid within a neighbourhood of a deterministic fixed point.

An example of numerical integration (for details regarding numerics see appendix C) of the ray equations (3.19) for the QD (3.17) and full (3.40) Hamiltonian is shown in figure 2. The QD rays are shown above the separatrix for comparison. Note that the QD rays are contained within a triangular domain, while the rays from the full Hamiltonian cover the entire domain. This is due to the domain restriction that occurs when removing the protein fluctuations from the process.

3.2. Adjoint eigenfunction

Up to terms that are exponentially small in $\epsilon$, the adjoint eigenfunction satisfies

$$\left[ \frac{1}{\epsilon} \mathbf{A}^T + F \frac{\partial}{\partial x} + G \frac{\partial}{\partial y} \right] \xi_0 = 0, \tag{3.41}$$

along with the boundary condition

$$\xi_n(x, y) = 0, \quad (x, y) \in \Gamma, \tag{3.42}$$

for all $n = 0, 1, 2$ such that $(F_{n,n}, G_{n,n}) \cdot \mathbf{n} > 0$. Note that we write the components of $\xi_0$ as $\xi_n$, $n = 0, 1, 2$ for notational convenience. To make things easier, we change coordinates to

$$\tau = x + y - 1, \quad \sigma = x - y \tag{3.43}$$

so that

$$x = \frac{1}{2} (1 + \tau + \sigma), \quad y = \frac{1}{2} (1 + \tau - \sigma). \tag{3.44}$$

This transforms the absorbing boundary, $x - y = 0$, to the vertical line, $\sigma = 0$. Then, (3.41) becomes

$$\left[ \frac{1}{\epsilon} \mathbf{A}^T (\tau, \sigma) + \hat{F}(\tau) \frac{\partial}{\partial \tau} + \hat{G}(\sigma) \frac{\partial}{\partial \sigma} \right] \xi_0(\tau, \sigma) = 0. \tag{3.45}$$
where
\[
\hat{A}(\tau, \sigma) \equiv A(x(\tau, \sigma), y(\tau, \sigma)),
\]
\[
\hat{F}(\tau) \equiv F + G = \text{diag}(\tau, 1 - \tau, -\tau),
\]
\[
\hat{G}(\sigma) \equiv F - G = \text{diag}(1 - \sigma, -\sigma, -(1 + \sigma)),
\]
where \(A, F \) and \(G \) are defined in (2.17) and (2.18). The absorbing boundary condition is then
\[
\varepsilon \tilde{e}_2(\tau, 0) = 0, \quad \text{for} \quad \tau \in (-1, 1).
\]
The approximation of the adjoint eigenfunction proceeds using singular perturbation methods, along the lines of [21]. Three solutions are found which are valid in different regions of the domain: an outer solution, a boundary-layer solution for \(\tau \) small, and a layer solution in the \(\tau \) large régime. For convenience, we have defined functions on the boundary
\[
\tilde{e}_2(\tau, 0) = 0, \quad \text{for} \quad \tau \in (-1, 1).
\]

### 3.3. Principal eigenvalue

We now have all of the components necessary to approximate the principal eigenvalue, using formula (3.12). First, we examine the term in the denominator, where we approximate the adjoint eigenfunction with the outer approximation, \(\hat{e}_0 \sim 1 \) (see appendix B) and the (unnormalized) stationary density with (3.14) (the higher order term \(\hat{r}_1 \) can be ignored). Then, the term in the denominator of (3.12) is simply the normalization factor for the stationary density, which can be approximated using Laplace’s method to obtain
\[
\int_D \exp \left[ -\frac{1}{\epsilon} \Phi(x, y) - k(x, y) \right] \, d\lambda \sim \frac{2\pi \epsilon}{\sqrt{\text{det}(Z(x_\ast, y_\ast))}},
\]
where \(Z \) is the Hessian matrix (3.23) of \(\Phi \) and \((x_\ast, y_\ast)\) is the stable fixed point where the process begins. Note that we have used the fact that \(k(x_\ast, y_\ast) = \Phi(x_\ast, y_\ast) = 0 \) and that \(r_0 \) is normalized so that its entries sum to 1.

The term in the numerator of (3.12) requires the approximation (B.33) of the adjoint eigenfunction on the absorbing boundary. This integral can also be approximated using Laplace’s method to obtain
\[
\int_{\Gamma} \hat{e}_0^T \hat{F} \, d\sigma \sim \frac{e b \sqrt{\pi} e^{-k(x_u, y_u)}}{b \sqrt{\pi} - \epsilon \sqrt{2\mu_1^{(0)}(\tau_u)}} \sqrt{2\mu_1^{(0)}(\tau_u)} \Phi''(\tau_u)
\]
\[
\times \varepsilon \epsilon \hat{G}(0) \tilde{r}_0(\tau) \exp \left[ -\frac{1}{\epsilon} \Phi(x_u, y_u) \right].
\]

For convenience, we have defined functions on the boundary in the variable \(\tau \) with
\[
f_0(\tau) = r_0(x(\tau, 0), y(\tau, 0))
\]
\[
\hat{f}(\tau) = \Phi(x(\tau, 0), y(\tau, 0))
\]
\[
\hat{k}(\tau) = k(x(\tau, 0), y(\tau, 0)),
\]
where \(x(\tau, \sigma) \) and \(y(\tau, \sigma) \) are defined in (3.44).

Although the quantities \(\Phi(x_u, y_u)\) and \(k(x_u, y_u)\) must be computed numerically, the remaining unknown terms can be computed analytically by exploiting the reflection symmetry of the problem. Along \(\Gamma\), we have that \(x = y \) and \(p_1 = p_2 \) so that the equation (3.15) for \(\Phi \) and \(r_0 \) can be written as
\[
[\hat{A}(\tau, 0) - \mu \hat{f}(\tau)]\hat{r}_0(\tau) = 0,
\]
where we have defined
\[
\mu(\tau) \equiv -p_1(x(\tau, 0), y(\tau, 0)) = -p_2(x(\tau, 0), y(\tau, 0)).
\]

The stability landscape function on the boundary is then
\[
\hat{f}(\tau) = \Phi(x_u, y_u) - \int_{\tau}^{\tau + \epsilon} \mu(\tau') \, d\tau'.
\]
The above is just an eigenvalue problem with three possible solutions, one of which can be excluded because there is a zero eigenvalue corresponding to the nullspace of \(\hat{A}\). It can be shown [22] that if the diagonal elements of \(\hat{F}(\tau)\) are such that at least two have opposite signs then only one of the remaining two eigenvalues has a corresponding positive eigenvector \(\tilde{r}_0(\tau)\) must have positive elements), making the solution to (3.55) unique. It turns out that this is only true for \(\tau \in (0, 1) \) because of the domain restriction caused by removing protein fluctuations. The result is
\[
\mu(\tau) = -\frac{\tau^2 + 2\tau^2 + 2(\tau - 1) + \tau}{2(\tau - 1)},
\]
\[
\tilde{r}_0(\tau) = \left( \frac{1}{\tau (1 - \tau), \tau, \frac{1}{\tau} (1 - \tau) \right)^T.
\]
We also have that
\[
\hat{f}''(\tau) = -\mu''(\tau) = -\frac{\tau^4 + 2\tau^3 + (3 + 2b)\tau^2 - 4b\tau + 2b}{2\tau^2(1 - \tau)^2}.
\]
And finally, using (B.9), (3.48) and (3.59) we obtain
\[
\tilde{z}^T \hat{G}(0) \tilde{r}_0(\tau) = \frac{1}{b} (1 - \tau).
\]
Combining these components together, we have the final result that
\[
\lambda_0 \sim \frac{1}{\sqrt{2\pi}} \left[ \frac{1 - \tau_u}{b - \epsilon \sqrt{2\mu_1^{(0)}(\tau_u)}} \Phi''(\tau_u) \right]
\]
\[
\times \exp \left[ -\frac{1}{\epsilon} \Phi(x_u, y_u) \right].
\]
Where \(\tau_u = x_u + y_u - 1\) and \(\mu_1^{(0)}(\tau)\) is given by (B.22). As expected, the eigenvalue is exponentially small in \(\epsilon\), which means that \(\Phi(x_u, y_u)\) (the height of the stability landscape at the unstable fixed point) must be approximated as accurately as possible.

The remaining terms are often referred to as the ‘prefactor’ and—except for the quantity \(k(x_u, y_u)\) —they represent properties local to the fixed points. The remaining term in the prefactor depends on \(k(x, y)\), which depends on properties of the process not local to the fixed points and must be computed numerically.

Finally, we point out that the eigenvalue approximation (3.62) takes the typical form, sometimes called the Kramers rate, for the escape rate from a multidimensional potential well (with a single unstable fixed point on the boundary of the
Figure 3. Level curves of the QD (blue) and full process (orange) stability landscape function $\Phi$. Black lines show the domain boundary for the QD process, and the dashed line shows the boundary of positive protein levels. (a) The QD result is compared to the full process with $\varphi = 0.1$ so that the protein noise is weak compared to promoter noise. (b) Same as (a) but with $\varphi = 1$ so that the protein noise strength is comparable to promoter noise. Parameter values are the same as figure 2.

potential well). That is, the principal eigenvalue takes a similar general form for different types of Markov processes, known for some discrete jump Markov processes [23–29, 6, 7, 11] and for continuous Markov processes [17]. In each case, there is a function in the prefactor analogous to $k(x, y)$, a square root term in the prefactor describing the curvature of the potential at the stable and unstable fixed points and an exponentially small term (in $\epsilon$) that depends upon the height of the potential well. This places the current results for Markov processes containing a hybrid discrete/continuous state space within the context of existing theory.

4. Results

We now use the results of the previous section to explore how removing protein production/degradation noise affects the random process. In particular, we examine the stability landscape and the metastable transition times. First, in figure 3 the numerical solutions to the ray equations (3.19) are used to generate level curves of the stability landscape function, $\Phi$, for both the QD (3.17) and the full (3.40) Hamiltonians. For presentation, the level curves are shown in the $(x + y, x − y)$ plane, with the separatrix along the left edge $(x − y = 0)$ of each frame. In figure 3(a), level curves for the full process are shown for $\varphi = 0.1$ so that the protein noise is small compared to promoter noise. Recall that $\varphi$ controls the strength of protein noise relative to the strength of promoter noise so that there is no protein noise in the limit $\varphi \to 0$ (see section 3.1.2). The resulting curves match closely in a neighbourhood of the stable fixed point and extend out toward the unstable saddle point. As expected, the level curves begin to diverge farther away from the either fixed point they are. In figure 3(b), the strength of the protein noise is increased, with $\varphi = 1$, and in this case, the stability landscape of the two processes are quite different, matching only near the fixed points of the deterministic dynamics.

As a result of the domain restriction effect, level curves of the full process extend into regions the QD process is excluded from. This is because protein levels can only cross above the line $y = 1 − x$, not below it, when there is no protein noise. This is also true of the lines $x = 1$ and $y = 1$. The domain restriction effect is eliminated when protein noise is added back into the process, even if it is very small compared to promoter noise. For the model considered here, the effect is of no serious consequence for metastable transitions as the restricted domain still contains all three fixed points. However, a model of a more complex gene circuit might be significantly affected by removing protein noise—especially if this restricts the domain for the protein levels in such a way as to generate qualitatively different behaviour, which would imply a nontrivial contribution of protein noise, no matter how negligible it may be.

Several stochastic trajectories, generated with the Gillespie algorithm, are shown in figure 4. Each pane shows a trajectory that escapes the lower-right stable fixed point for a different value of $\varphi$. In the first pane (top left), we set $\varphi = 1$, and the discrete nature of the process is visually apparent. As $\varphi$ decreases, the trajectories appear less ‘discrete’ (more like Brownian motion), but still exhibit sharp transitions due to the fluctuating promotor state. Finally, in the last pane (lower right), a QD trajectory is shown that represents the limit $\varphi \to 0$. For each nonzero value of $\varphi$, the trajectory can travel outside the restricted domain (by crossing the solid black lines), whereas the QD trajectory cannot.

Because of the symmetry in the problem, we obtained analytical results for various quantities on the separatrix, including the shape stability landscape. We can use these results to obtain an analytical approximation of the probability density for the position along the separatrix a trajectory passes through as it transitions from one basin of attraction to another. Using the results of section 3.3 and appendix B, the stationary density along the separatrix is given by

$$p_{\text{ext}}(\tau) \sim \frac{\bar{r}_0(\tau) e^{-k(\tau)}}{\int_0^1 e^{-k(\tau)} \exp \left[ -\frac{1}{\tau} \Phi(\tau) \right] d\tau},$$

(4.1)

where we remind the reader that $\tau = x + y - 1$ and $x = y$ along the separatrix. The only term that cannot be obtained analytically is $k(\tau)$, which can be ignored as a first approximation. For simplicity, we also average over the promotor state to get the scalar marginal probability density for
Figure 4. Stochastic trajectories generated using the Gillespie algorithm for different values of $\phi$. Each trajectory starts at the lower-right stable fixed point and terminates once it reaches the absorbing boundary (dashed line). Trajectories are more likely to exit near the unstable saddle located on the absorbing boundary. As in figure 3, the black lines represent the boundary of the restricted domain for the QD process. For each simulation, we set $b = 0.15$ and $\epsilon = 0.01$.

Figure 5. The density of exit points along the separatrix as a function of $\tau = x + y - 1$, with $x = y$. The black curve shows the analytical approximation for the QD ($\phi = 0$) process, and the symbols show histograms from $10^4$ Monte Carlo simulations for different values of $\phi$. The exit point. Then, using Laplace’s method, the exit density is

$$P_{\text{exit}}(\tau) \sim \sqrt{\Phi''(\tau_u)} \exp \left[ -\frac{1}{\epsilon} (\Phi(\tau) - \Phi(\tau_u)) \right],$$

where $\Phi(\tau)$ is given by

$$\Phi(\tau) = -\frac{\tau^2}{4} - \frac{3\tau}{2} - b \log(\tau) - 2 \log(1 - \tau).$$

The QD exit density approximation is shown in figure 5 along with two histograms obtained from Monte Carlo simulations. While the histogram for $\phi = 0.1$ is close to the QD approximation, it is evident that some trajectories pass through the separatrix in the interval $(-1, 0)$, which is impossible without protein noise.

The FETD for a trajectory starting from a stable fixed point to reach the separatrix is asymptotically exponential in the large time limit, and the timescale is determined by the principal eigenvalue, $\lambda_0$ (see equation (3.62)). We can then approximate the mean exit time with $T \sim 1/\lambda_0$. In figure 6 the mean exit time is shown on a log scale as a function of $1/\epsilon$ along with results from Monte Carlo simulations.

Note that the approximation and the Monte Carlo simulations are asymptotically linear as $1/\epsilon \rightarrow \infty$. For the approximation, the slope of this line is determined by the height of the stability barrier, $\Phi(x_u, y_u)$, while the prefactor affects the vertical shift. From the Monte Carlo results (symbols with grey lines), we see that the mean exit time converges to the approximation as $\phi \rightarrow 0$. However, it is clear that the slope of the analytical curve is slightly different than that of the Monte Carlo results even when $\phi$ is small. Thus, we may think of the mean exit time approximation for the QD process as an asymptotic approximation of the full process for $\phi \ll 1$, provided that $\epsilon$ is also small but not too small.

5. Discussion

Understanding how different noise sources affect the dynamics of a gene circuit is essential to understand how different regulatory components interact to produce the complex variety of environmental responses and behaviours. Even if one excludes extrinsic noise sources—such as environmental and organism-to-organism variations—there are several sources of intrinsic noise, such as fluctuations in translation, transcription and the conformational state of DNA regulatory units.

The behaviour we are interested in understanding is a transition from one metastable state to another. Each metastable state corresponds to the stable steady-state solutions of the underlying deterministic system. The bistable mutual repressor model has two identical stable steady states separated by an unstable saddle node. On small timescales, the protein levels fluctuate near one of the two stable steady states. On large timescales, fluctuations cause a metastable
In this section, we show how to solve equation (3) by award no KUK-C1-013-4 made by King Abdullah University of Science and Technology (KAUST).

**Appendix A. Computing \( k(x, y) \)**

In this section, we show how to solve equation (3.32) for \( k(x, y) \) numerically along the characteristics defined by (3.19). First, differentiating (3.32) along characteristics yields

\[
\dot{k} = \frac{\partial k}{\partial x} \dot{x} + \frac{\partial k}{\partial y} \dot{y} = \frac{\partial k}{\partial x} \dot{p}_1 + \frac{\partial k}{\partial y} \dot{p}_2.
\]  

(A.1)

Using the fact that \((I^T G r_0)x - (I^T F r_0)y = 0\) along characteristics, we can define

\[
h(x, y) \equiv \frac{\partial x}{\partial y} I^T F r_0 = \frac{\partial x}{\partial y} I^T G r_0.
\]  

(A.2)

Then, after combining (3.32) and (A.1) we have that

\[
\dot{k} = h(x, y) I^T \left[ \left( \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} \right) r_0 + F \frac{\partial r_0}{\partial x} + G \frac{\partial r_0}{\partial y} \right].
\]  

(A.3)

The above requires values of \( \frac{\partial r_0}{\partial x} \) and \( \frac{\partial r_0}{\partial y} \), which are not provided by the system (3.19). To obtain these, a formula is needed to relate the Hessian matrix, \( Z(x, y) \), of \( \Phi(y, x) \) to \( \nabla r_0 \). Then, the Hessian matrix can be computed by expanding the system of ray equations (3.19).

First, differentiate both sides of equation (3.15) to obtain

\[
[A + p_1 F + p_2 G] \nabla r_0 = -(\nabla A + \nabla (p_1 F) + \nabla (p_2 G)) r_0.
\]  

(A.4)

The Fredholm alternative theorem requires that

\[
I^T (\nabla A + \nabla (p_1 F) + \nabla (p_2 G)) r_0 = I^T (\nabla M) r_0 = 0.
\]  

(A.5)

which is always true since \( M r_0 = M^* I = 0 \) and

\[
0 = \nabla (I^T M r_0) \quad \text{(A.6)}
\]

\[= (\nabla I^T) M r_0 + I^T (\nabla M) r_0 + I^T M (\nabla r_0) \quad \text{(A.7)}
\]

\[= I^T (\nabla M) r_0. \quad \text{(A.8)}
\]

The general solution to (A.4) is

\[
\nabla r_0 = -M^* \nabla M r_0 + \alpha r_0, \quad \text{(A.9)}
\]

where \( M^* \) is the pseudoinverse of the matrix \( M \) and \( \alpha \) is an unknown constant. Since the vector \( r_0 \) is normalized so that its entries sum to 1, it follows that \( \sum_n (\nabla r_0)_n = 0 \). Summing over both sides of equation (A.9) then yields

\[
\alpha = \sum_{n=0}^2 (M^* \nabla M r_0)_n. \quad \text{(A.10)}
\]

Thus, we have that

\[
\nabla r_0 = z - \sum_{n=0}^2 z_n r_0, \quad z = -M^* \nabla M r_0. \quad \text{(A.11)}
\]
Equation (A.11) gives a relationship between $Z, r_0$ and $\nabla r_0$. To obtain the Hessian matrix, $Z(x, y)$, away from the fixed point, the ray equations are extended to include the variables

\[
x_j = \frac{\partial x}{\partial u_j}, \quad y_j = \frac{\partial y}{\partial u_j},
\]

\[
p^1_j = \frac{\partial p_1}{\partial u_j}, \quad p^2_j = \frac{\partial p_2}{\partial u_j},
\]

for $j = 1, 2$. A good choice for the current problem is to take $u_1 = x_0(\theta)$ and $u_2 = y_0(\theta)$, where $(x_0(\theta), y_0(\theta))$ is a point on the initial curve defined by (3.24). The Hessian matrix is then obtained using

\[
\begin{bmatrix}
p^1_1 & p^2_1 \\
p^1_2 & p^2_2
\end{bmatrix} = Z \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.
\]

As long as the matrix on the RHS is invertible, the matrix $Z$ can be obtained along characteristics and $k$ can be integrated numerically using (A.3) and (A.4). The dynamics for the extended variables (A.12) is given by

\[
v_j = J(x, y, p_1, p_2) v_j,
\]

where $v_j \equiv (x_j, y_j, p^1_j, p^2_j)^T$ and $J(x, y)$ is the Jacobian matrix for the system (3.19). One can choose different variables $u_j$ with which to extend the system based on what works in practice, and the only thing one must change are the initial conditions. For our choice, the initial conditions are

\[
x_j(0) = \delta_{i,j}, \quad y_j(0) = \delta_{i,j},
\]

\[
p^1_j(0) = Z_{j,1}(x_*, y_*), \quad p^2_j(0) = Z_{j,2}(x_*, y_*),
\]

where the matrix $Z(x_*, y_*)$ is the solution to (3.26).

**Appendix B. Singular perturbation analysis of the adjoint eigenfunction**

In this section, we present the singular perturbation analysis of the adjoint eigenfunction, which satisfies equation (3.41). Before proceeding, it is convenient to make the following definitions. In the rest of this section, we make frequent use of the eigenvectors (right eigenvectors $\psi_n$ and left eigenvectors $\eta_n$) and eigenvalues, $\mu$ and satisfying

\[
\hat{A}(\sigma, \tau) \psi_n(\sigma, \tau) = \mu_n(\sigma, \tau) \hat{G}(\sigma) \psi_n(\sigma, \tau)
\]

\[
\hat{A}(\sigma, \tau)^T \eta_n(\sigma, \tau) = \mu_n(\sigma, \tau) \hat{G}(\sigma) \eta_n(\sigma, \tau).
\]

We normalize the two sets of eigenvectors (which are biorthogonal) so that $\eta^T \hat{G} \psi = \delta_{i,j}$. Note that because the matrices $\hat{A}$ and $\hat{G}$ are functions of $(\sigma, \tau)$, so are the eigenpairs. It is easily shown that one of the eigenvalues is zero for all values of $(\sigma, \tau)$, which we set to $\mu_0 = 0$. The right eigenvector, $\psi_0$, is then given by the nullspace of the matrix $A$:

\[
\psi_0(\sigma, \tau) = \frac{1}{2(\sigma^2 + (\tau + 1)^2) + 2b} \left( \begin{array}{c} 1 + \tau + \sigma^2 \\ 4b \\ 1 + \tau - \sigma^2 \end{array} \right)^T.
\]

Furthermore, the corresponding left eigenvector is simply

\[
\eta_0 = 1 \equiv (1, 1, 1)^T.
\]

It is convenient to define distinct notation for the eigenpairs evaluated on $\Gamma$, with

\[
\hat{\psi}_n(\tau) \equiv \psi_n(\tau, 0), \quad \hat{\eta}_n(\tau) \equiv \eta_n(\tau, 0), \quad \hat{\mu}_n(\tau) \equiv \mu_n(\tau, 0).
\]

At the boundary one of the eigenvalues, $\mu_1$ say, vanishes, and the eigenspace for the zero eigenvalue is degenerate (i.e. there are two zero eigenvalues but the nullspace is one dimensional), which means that $\hat{\mu}_1 = 0, \hat{\psi}_1 = \hat{\psi}_0$ and $\hat{\eta}_1 = 1$.

The approximation of the adjoint eigenfunction proceeds using singular perturbation methods, along the lines of [21]. Three solutions are found which are valid in different regions of the domain: an outer solution, a boundary-layer solution for the $O(\epsilon)$ strip near the absorbing boundary and a transition-layer solution in the $O(\epsilon^{1/2})$ overlap region between the other two.

Away from the boundary, the exact solution (that does not satisfy the boundary condition (3.49)) is

\[
\xi_{\text{out}} = 1.
\]

To obtain a uniform asymptotic approximation that also satisfies (3.49), a boundary-layer solution is needed. Consider the stretched variable $z = \sigma/\epsilon$. To leading order in $\epsilon$, the boundary-layer solutions, $\xi_{\text{bl}}(\tau, z)$, satisfies

\[
\hat{G}(0) \frac{\partial \xi_{\text{bl}}}{\partial z} + \hat{A}(\tau, 0)^T \xi_{\text{bl}} = 0,
\]

where $\hat{G}(0) = \text{diag}(1, 0, -1)$. The solution has the form

\[
\xi_{\text{bl}}(\tau, z) = c_0 \mathbf{1} + c_1 (\zeta + z \mathbf{1}) + c_2 \hat{\eta}_2 e^{-\hat{\mu}_2 z},
\]

where $\hat{\eta}_2, \hat{\mu}_2$ is the only eigenpair (on the boundary) with a nonzero eigenvalue. However, the eigenvalue, $\hat{\mu}_2(\tau)$, is negative for all values of $\tau \in (-1, 1)$, and in order to obtain a bounded solution in the limit $z \to \infty$ we set $c_2 = 0$. The vector $\xi$ is the generalized left eigenvector satisfying $\hat{A}(\tau, 0)^T \xi = \hat{G}(0) \mathbf{1}$ and is given by

\[
\xi = (-1/b, 0, 1/b)^T.
\]

At the boundary, the solution is

\[
\xi_{\text{bl}}(\tau, 0) = c_0 \mathbf{1} + c_1 \zeta,
\]

and the boundary condition (3.49) requires

\[
c_0 + \frac{c_1}{b} = 0,
\]

so that $c_1 = -bc_0$. Thus, up to a single unknown constant, $c_0$, which must be determined by matching, the boundary-layer solution is

\[
\xi_{\text{bl}}(\tau, z) = c_0 (\mathbf{1} - b(\zeta + z \mathbf{1})).
\]

Because $\xi_{\text{bl}}(\tau, z)$ is unbounded in the limit $z \to \infty$, it is not possible to match it to the outer solution, $\xi_{\text{out}}$. We can think of the term, $\zeta + z \mathbf{1}$, in the boundary-layer solution is a truncated Taylor series expansion of the true solution around $z = 0$. To match the boundary-layer and outer solutions, a transition-layer solution is required for the strip of width $O(\sqrt{\epsilon})$ along the boundary.

Consider the stretched coordinate $s = \sigma/\sqrt{\epsilon}$. Keeping terms to $O(\epsilon^{1/2})$, the transition-layer solution, $\xi_{\text{tl}}(\tau, s)$, satisfies

\[
\sqrt{\epsilon} \hat{G}(0) \frac{\partial \xi_{\text{tl}}}{\partial s} + \hat{A}(\tau, 0)^T \xi_{\text{tl}} = 0.
\]
It is less clear how to truncate the above equation to obtain a leading-order transition-layer solution. Because we must match the outer solution, \( I \), to the boundary-layer solution that has the generalized eigenvector \( \xi \), we try a solution of the form
\[
\xi_0(t, s) = \xi_0(t, s) I + a_1(t, s) \chi(t, s),
\]
(B.14)
where
\[
\chi(t, s) = \frac{1}{v(t, \sqrt{\epsilon s})} (1 - \eta_1(t, \sqrt{\epsilon s})),
\]
(B.15)
and \( a_{0,1} \) are unknown scalar functions. In the limit \( \sigma \to 0 \), the deterministic flux across the boundary,
\[
v(t, \sigma) = f(x(t, \sigma), y(t, \sigma)) - f(y(t, \sigma), x(t, \sigma)),
\]
(B.16)
(with \( f(x, y) \) given by (2.15)) vanishes and the eigenvector \( \eta_1 \to 1 \). That is, the eigenvalue \( \tilde{\mu}_1 \), corresponding to the eigenvector \( \eta_1 \), vanishes on the boundary. Furthermore, it can be shown that
\[
\lim_{s \to 0} \chi(t, s) = -\frac{\tilde{\mu}_1(t)(t)}{\tilde{\mu}_1(t)(t)} \xi,
\]
(B.17)
where
\[
\tilde{v}^{(\sigma)}(t) = \frac{\partial}{\partial \sigma} v(t, 0), \quad \tilde{\mu}_1(t)(t) = \frac{\partial}{\partial \sigma} \mu_1(t, 0).
\]
(B.18)
Substituting (B.14) into (B.13) yields
\[
\sqrt{\epsilon} \tilde{G}(0) \left( \frac{\partial \tilde{a}_1}{\partial s} + \tilde{a}_1 \tilde{\chi} + \sqrt{\epsilon} \tilde{a}_1 \frac{\partial \tilde{\chi}}{\partial \sigma} \right) + a_1 \tilde{\chi}(t, 0) \chi = 0.
\]
(B.19)
To obtain the unknown functions \( a_{0,1}(t, s) \) we project (B.19) with the right eigenvectors, \( \tilde{\psi}_n(t, s) \), \( n = 0, 1 \). After applying these projections (using the fact that \( \tilde{\psi}_0 \tilde{G} \chi = 1 \), \( \tilde{\psi}_0 \tilde{A}^T = 0 \), \( \tilde{\psi}_0^T \tilde{G} 1 = v(t, s) \), and \( \tilde{\psi}_0^T (\tilde{G} 1 = 0) \) and collecting leading-order terms in \( \epsilon \), we obtain
\[
\frac{\partial a_1}{\partial s} = v(t, \sqrt{\epsilon s}) \frac{\partial a_0}{\partial s}
\]
(B.20)
\[
\frac{\partial a_1}{\partial s} + s \tilde{\mu}_1(t)(t) a_1 = 0.
\]
(B.21)
where
\[
\tilde{\mu}_1^{(\sigma)}(t) = -\frac{b(t^2 + 2b - 1)}{(t + 1)^2}.
\]
(B.22)
It turns out that \( \tilde{\mu}_1^{(\sigma)} \) is related to the curvature of the stability landscape normal to the separatrix; that is, if we define
\[
\phi(t, \sigma) = \phi(x(t, \sigma), y(t, \sigma)),
\]
(B.23)
then
\[
\tilde{\mu}_1^{(\sigma)}(t) = -\frac{\partial^2}{\partial \sigma^2} \phi(t, 0).
\]
(B.24)
At \( r_\alpha = \sqrt{1 - 2b} \), the curvature vanishes and changes sign but is always negative (with \( \tilde{\mu}_1^{(\sigma)}(r_\alpha) > 0 \)) at the unstable fixed point, \( (r_\alpha, 0) \). Divide the separatrix \( (\sigma = 0 \) and \( -1 < t < 1 \) into three regions: \( -1 < t < 0 \), \( 0 < t < r_\alpha \), and \( r_\alpha < t < 1 \). The first region is ignored because it is in part of the domain \( D \) excluded from the stationary density function (see section 3.1). The second region contains the unstable fixed point, and the third we can ignore as only extremely rare trajectories cross the separatrix in this region. Up to an unknown constant, the solutions to (B.20) and (B.21) are
\[
a_0(t, s) \sim -\tilde{\mu}_1(t)(t) \int_0^t \exp \left[ -\frac{1}{2} \tilde{\mu}_1(t)(t) u^2 \right] du,
\]
(B.25)
\[
a_1(t, s) \sim \tilde{a} \exp \left[ -\frac{1}{2} \tilde{\mu}_1(t)(s)^2 \right].
\]
(B.26)
The transition-layer solution is then
\[
\xi_0(t, s) = \tilde{a} \left( -\frac{\tilde{\mu}_1^{(\sigma)}(t)}{\sqrt{1/2 \tilde{\mu}_1^{(\sigma)}(t)}} \int_0^t \exp \left[ -\frac{1}{2} \tilde{\mu}_1^{(\sigma)}(t) u^2 \right] \right) \chi + \xi z 1.
\]
(B.27)
The three solutions can now be matched. First, matching the transition-layer solution to the boundary-layer solution is done using the Van–Dyke rule. In terms of the boundary-layer variable, \( z \), the transition-layer solution is
\[
\xi_0(t, e^{1/3} z) = -\tilde{a} \tilde{\mu}_1^{(\sigma)}(t) (z + 1).
\]
(B.28)
Matching terms with the boundary-layer solution yields
\[
\tilde{a} = bc_0 \tilde{\mu}_1^{(\sigma)}(t).
\]
(B.29)
The composite boundary/transition-layer solution is then
\[
\xi_{b/t}(t, s) = c_0 \left[ 1 - \frac{b}{e^{1/2} \int_0^t \sqrt{1/2 \tilde{\mu}_1^{(\sigma)}}(t)} \exp \left[ -\frac{1}{2} \tilde{\mu}_1^{(\sigma)}(t) s^2 \right] \chi(t, s) \right].
\]
(B.30)
The final unknown constant, \( c_0 \), is determined by matching to the outer solution so that
\[
\lim_{s \to 0} \xi_{b/t}(t, s) = 1,
\]
(B.31)
which implies that
\[
c_0 = -\frac{\sqrt{2b \tilde{\mu}_1^{(\sigma)}}(t)}{b \sqrt{\pi} - \sqrt{2e \tilde{\mu}_1^{(\sigma)}}(t)}.
\]
(B.32)
In order to evaluate the term in the numerator of the eigenvalue formula (3.12), we require the adjoint eigenfunction evaluated on the boundary (in a neighbourhood of the unstable fixed point), which is
\[
\xi_0(t, 0) \sim \frac{\sqrt{2b \tilde{\mu}_1^{(\sigma)}}(t)}{b \sqrt{\pi} - \sqrt{2e \tilde{\mu}_1^{(\sigma)}}(t)} (-1 + b\xi),
\]
(B.33)
Appendix C. Numerical simulations

In this section, we summarize the numerical methods and tools used throughout the paper. Most numerical work is performed in Python, using the Numpy/Scipy package. For more computationally expensive tasks, we use Scipy’s Weave package to include functions written in C, which allows us to use the GNU Scientific Library for numerical Integration of the ray equations (3.19), and for random number generators used in Monte Carlo simulations.

There are a few notable observations regarding integration of the ray equations. First, characteristic projections, \((x(t), y(t))\), have a tendency to ‘stick’ together along certain trajectories, peeling off one at a time (see figure 2). To adequately cover the domain with rays, a shooting method must be used to select points on the Cauchy data. For more details on this see [12]. We found that the simplest method was to use the secant method (we use the ‘brentq’ function in the Scipy.integrate package) to minimize the Euclidian distance between the final value of \((x(t), y(t))\) along the separatix and the saddle node. This method is convenient since it does not require the knowledge of the Hessian matrix, \(Z(x, y)\). Second, the value of \(\omega\) used to generate Cauchy data must be chosen small enough to get accurate results. However, we found that if it is chosen too small, rays are no longer able to cover the domain, and more importantly, we could no longer generate a ray that reaches the unstable fixed point on the separatix.

For the multiperessor model, the trajectory connecting the stable fixed point to the saddle is one of the curves along which characteristics tend to stick to each other. Suppose that \(\theta_m\) is the point on the Cauchy data (3.21) that generates the ray that connects the fixed points. Then small perturbations \(\theta = \theta_m + \delta \theta, |\delta \theta| \ll 1\), cause the characteristic, \((x(t), y(t))\), to diverge sharply away from the saddle. This not only makes it difficult to compute \(\theta_m\) but also creates difficulties for computing \(k(x, y)\) (see section 3.1.1). Since the expanded set of ray equations (A.15) track the derivatives of \(x, y, p_1\) and \(p_2\) with respect to the point on the Cauchy data, which, for values of \(\theta\) near \(\theta_m\), becomes very large as the ray approaches the saddle point. As the expanded variables become very large, computing \(Z\) using equation (A.14) is unstable. Furthermore, this effect becomes worse as the initial value, \(\omega = \Phi(0)\), goes to zero.

Appendix D. Quasi-deterministic Hamiltonian

\[
\mathcal{H}(x, y, p_1, p_2) = -(x-1)^2p_3^2 - y(y-1)^2p_3^2 + (x-1) \\
\times (2x + y - 3xy - 1)p_4^2p_2 + (y-1)(x+2y - 3xy - 1) \\
\times p_1p_2^2 - (x-1)(x^2 + xy^2 + 2bx - b)p_2^2 - (y-1) \\
\times (x^2 + y^2 + 2by - b)p_2^2 \\
- x^3y^3 + (x+y^2 + xy^2 + 3b(x+y) + 4bxy + 2b)p_1p_2 \\
- b(x^3 + x^2 + xy + bx - b)p_1 \\
- b(x^2y + y^3 - y^2 + by - b)p_2. \\
\]

References

[1] Kaern M, Elston T C, Blake W and Collins J J 2005 Stochasticity in gene expression: from theories to phenotypes Nature Rev. Genet. 6 451–64
[2] 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. 36 413–34
[3] Eldar A and Elowitz M B 2010 Functional roles for noise in genetic circuits Nature 467 167–73
[4] Thattai M and van Oudenaarden A 2001 Intrinsinc noise in gene regulatory networks Proc. Natl Acad. Sci. USA 98 8614–9
[5] McAdams H H and Arkin A 1997 Stochastic mechanisms in gene expression Proc. Natl Acad. Sci. USA 94 814–9
[6] Aurell E and Speepen K 2002 Epigenetics as a first exit problem Phys. Rev. Lett. 88 048101
[7] Roma D M, O’Flanagan R A, Ruckenstei A E, Sengupta A M and Mukhopadhyay R 2005 Optimal path to epigenetic switching Phys. Rev. E 71
[8] Paulsson J 2005 Models of stochastic gene expression Phys. Life Rev. 2 157–75
[9] Horn J E M, Schultz D, Innocentini G C P, Wang J, Walczak A M, Onuchic J N and Wolynes P G 2005 Self-regulating gene: an exact solution Phys. Rev. E 72 051907
[10] 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
[11] Assaf M, Roberts E and Luthey-Schulten Z 2011 Determining the stability of genetic switches: explicitly accounting for mRNA noise Phys. Rev. Lett. 106 248102
[12] Ludwig D 1975 Persistence of dynamical systems under random perturbations SIAM Rev. 17 605–40
[13] Matkowsky B J, Schuss Z and Tier C 1983 Diffusion across characteristic boundaries with critical points SIAM J. Appl. Math. 43 673–95
[14] Talkner P 1987 Mean first passage time and the lifetime of a metastable state Z. Phys. B 68 201–7
[15] Nach N, Klosek M M, Matkowsky B J and Schuss Z 1990 A direct approach to the exit problem SIAM J. Appl. Math. 50 595–627
[16] Maier R S and Stein D L 1997 Limiting exit location distributions in the stochastic exit problem SIAM J. Appl. Math. 57 752–90
[17] Schuss Z 2010 Theory and Applications of Stochastic Processes: An Analytical Approach (Applied Mathematical Sciences vol 170) (New York: Springer)
[18] Doering C, Sargsyan K and Sander L 2005 Extinction times for birth-death processes: exact results, continuum asymptotics, and the failure of the Fokker–Planck approximation Multiscale Model. Simul. 3 283–99
[19] Zeiser S, Franz U, Mueller J and Liebscher V 2009 Hybrid modeling of noise reduction by a negatively autoregulated system Bull. Math. Biol. 71 1006–24
[20] Zeiser S, Franz U and Liebscher V 2010 Autocatalytic genetic networks modeled by piecewise-deterministic Markov processes J. Math. Biol. 60 207–46
[21] Keener J P and Newby J M 2011 Perturbation analysis of spontaneous action potential initiation by stochastic ion channels Phys. Rev. E 84 011918
[22] Newby J M and Keener J P 2011 An asymptotic analysis of the spatially inhomogeneous velocity-jump process Multiscale Model. Simul. 9 735–65
[23] Hanggi P, Grabert H, Talkner P and Thomas H 1984 Bistable systems: Master equation versus Fokker–Planck modeling Phys. Rev. A 29 371–8
[24] Dykman M I, Mori E, Ross J and Hunt P M 1994 Large fluctuations and optimal paths in chemical kinetics J. Chem. Phys. 100 5735–50
[25] Hinch R and Chapman S J 2005 Exponentially slow transitions on a Markov chain: the frequency of calcium sparks Eur. J. Appl. Math. 16 427–46
[26] Vellela M and Qian H 2007 A quasistationary analysis of a stochastic chemical reaction: Keizer’s paradox Bull. Math. Biol. 69 1727–46
[27] Escudero C and Kamenev A 2009 Switching rates of multistep reactions Phys. Rev. E 79 041149
[28] Metzner P, Schütte C and Vanden-Eijnden E 2009 Transition path theory for Markov jump processes Multiscale Model. Simul. 7 1192–219
[29] Bressloff P C 2010 Metastable states and quasicycles in a stochastic Wilson–Cowan model of neuronal population dynamics Phys. Rev. E 82 051903
[30] Kepler T B and Elston T C 2001 Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations Biophys. J. 81 3116–36
[31] Ockendon J, Howison S, Lacey A and Movchan A 2003 Applied Partial Differential Equations (Oxford: Oxford University Press) (Revised edition)
[32] Gardiner C W 1983 Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences vol 13 (Berlin: Springer)
[33] Gillespie D T 1977 Exact stochastic simulation of coupled chemical reactions J. Phys. Chem. 81 2340–61