



**HAL**  
open science

# On the Hamiltonian structure of large deviations in stochastic hybrid systems

Paul Bressloff, Olivier Faugeras

► **To cite this version:**

Paul Bressloff, Olivier Faugeras. On the Hamiltonian structure of large deviations in stochastic hybrid systems. 2015. hal-01072077v2

**HAL Id: hal-01072077**

**<https://inria.hal.science/hal-01072077v2>**

Preprint submitted on 22 Sep 2015 (v2), last revised 22 Sep 2017 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Public Domain Mark 4.0 International License

# On the Hamiltonian structure of large deviations in stochastic hybrid systems

Paul C. Bressloff<sup>1</sup> and Olivier Faugeras<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, Utah 84112, USA

<sup>2</sup>NeuroMathComp Group, INRIA, Sophia Antipolis, France

E-mail: [bressloff@math.utah.edu](mailto:bressloff@math.utah.edu), [olivier.faugeras@inria.fr](mailto:olivier.faugeras@inria.fr)

## Abstract.

We develop the connection between large deviation theory and more applied approaches to stochastic hybrid systems by highlighting a common underlying Hamiltonian structure. A stochastic hybrid system involves the coupling between a piecewise deterministic dynamical system in  $\mathbb{R}^d$  and a time-homogeneous Markov chain on some discrete space  $\Gamma$ . We assume that the Markov chain on  $\Gamma$  is ergodic, and that the discrete dynamics is much faster than the piecewise deterministic dynamics (separation of time-scales). Using the Perron-Frobenius theorem and the calculus-of-variations, we evaluate the rate function of a large deviation principle in terms of a classical action, whose Hamiltonian is given by the Perron eigenvalue of a  $|\Gamma|$ -dimensional linear equation. The corresponding linear operator depends on the transition rates of the Markov chain and the nonlinear functions of the piecewise deterministic system. The resulting Hamiltonian is identical to one derived using path-integrals and WKB methods. We illustrate the theory by considering the example of stochastic ion channels. Finally, we indicate how the analysis can be extended to a multi-scale stochastic process, in which the slow dynamics is described by a piecewise SDE.

## 1. Introduction

There are a growing number of problems in biology that involve the coupling between a piecewise deterministic dynamical system in  $\mathbb{R}^d$  and a time-homogeneous Markov chain on some discrete space  $\Gamma$ , resulting in a stochastic hybrid system, also known as a piecewise deterministic Markov process (PDMP) [8]. One important example is given by membrane voltage fluctuations in neurons due to the stochastic opening and closing of ion channels [4, 6, 7, 14, 17, 20, 29, 30, 34]. Here the discrete states of the ion channels evolve according to a continuous-time Markov process with voltage-dependent transition rates and, in-between discrete jumps in the ion channel states, the membrane voltage evolves according to a deterministic equation that depends on the current state of the ion channels. In the limit that the number of ion channels goes to infinity, one can apply the law of large numbers and recover classical Hodgkin-Huxley type equations. However, finite-size effects can result in the noise-induced spontaneous firing of a neuron due to channel fluctuations. Other examples of stochastic hybrid systems include genetic switches [21, 25], motor-driven intracellular transport [27, 28], and stochastic neural networks [1].

In many of the above examples, one finds that the transition rates between the discrete states  $n \in \Gamma$  are much faster than the relaxation rates of the piecewise deterministic dynamics for  $x \in \mathbb{R}^d$ . Thus there is a separation of time scales between the discrete and continuous processes, so that if  $t$  is the characteristic time-scale of the Markov chain then  $\epsilon t$  is the characteristic time-scale of the relaxation dynamics for some small positive parameter  $\epsilon$ . Assuming that the Markov chain is ergodic, in the limit  $\epsilon \rightarrow 0$  one obtains a deterministic dynamical system in which one averages the piecewise dynamics with respect to the corresponding unique stationary measure. This then raises the important problem of characterizing how the law of the underlying stochastic process approaches this deterministic limit in the case of weak noise,  $0 < \epsilon \ll 1$ .

A rigorous mathematical approach to addressing the above issue is *large deviation theory*, which has been developed extensively within the context of stochastic differential equations (SDEs) [9, 15, 33]. In particular, consider some random dynamical system in  $\mathbb{R}^d$  for which there exists a well defined probability density functional or law  $P_\epsilon[x]$  over the different sample trajectories  $\{x(t)\}_0^T$  in a given time interval  $[0, T]$ . Here  $\epsilon$  is a small parameter that characterizes the noise level, with  $x(t)$  given by the solution  $x^*(t)$  of some ODE  $\dot{x} = F(x)$  in the limit  $\epsilon \rightarrow 0$ . A large deviation principle (LDP) for the random paths of the SDE over some time interval  $[0, T]$  is

$$P_\epsilon[x] \sim e^{-J_T[x]/\epsilon}, \quad \epsilon \rightarrow 0,$$

where  $J_T[x]$  is known as the rate function and  $J_T[x^*] = 0$ . In the case of SDEs, the rate function can be interpreted as a classical action with corresponding Lagrangian  $L$  [15],

$$J_T[x] = \int_0^T L(x, \dot{x}) dt.$$

Such a Lagrangian formulation is more amenable to explicit calculations. In particular, it can be used to solve various first passage time problems associated with the escape from a fixed point attractor of the underlying deterministic system in the weak noise limit. This involves finding the most probable paths of escape, which minimize the action with respect to the set of all trajectories emanating from the fixed point.

Evaluating the action along a most probable path from the fixed point to another point  $x$  generates a corresponding *quasipotential*  $\Phi(x)$ . From classical variational analysis, it can be shown that the quasipotential satisfies a Hamilton-Jacobi equation  $H(x, \partial_x \Phi) = 0$ , where  $H$  is the Hamiltonian obtained from  $L$  [15] via a Fenchel-Legendre transformation:

$$H(x, p) = \sup_y \{py - L(x, y)\}.$$

The optimal paths of escape correspond to solutions of Hamilton's equations on the zero energy surface ( $H = 0$ ). Interestingly, the same quasipotential is obtained by considering a Wentzel-Kramers-Brillouin (WKB) approximation of the stationary state of the continuous process  $x(t)$ . More recently, large deviation theory has been applied to stochastic hybrid systems [11, 12, 23].

Independently of the developments in large deviation theory, a variety of techniques in applied mathematics and mathematical physics have been used to solve first passage time problems in biological applications of stochastic hybrid systems. These include WKB approximations and matched asymptotics [4, 20, 25, 28, 29], and path-integrals [3]. Although such approaches are less rigorous than large deviation theory, they are more amenable to explicit calculations. In particular, they allow one to calculate the prefactor in Arrhenius-like expressions for mean first passage times, rather than just the leading order exponential behavior governed by the quasipotential. The main aim of this paper is to make explicit the connection between large deviation theory and more applied approaches to stochastic hybrid systems, by highlighting the common underlying Hamiltonian structure.

In §2 we define a stochastic hybrid system and specify our various mathematical assumptions. We then construct a WKB approximation of solutions to the differential Chapman-Kolmogorov equation, which describes the evolution of the probability density function for a stochastic hybrid system (§3). In particular, we show that the resulting quasipotential satisfies a Hamilton-Jacobi equation, whose associated Hamiltonian is given by the Perron eigenvalue of a corresponding matrix equation that depends on the transition rates of the Markov chain and the nonlinear functions of the piecewise deterministic system. This Hamiltonian is identical to the one rigorously derived by Kifer using large deviation theory [23] and by Bressloff and Newby using formal path-integral methods [3, 5]. Further insights into the Hamiltonian structure of large deviations in stochastic hybrid systems are developed in §4 (and appendix A), where we present an alternative derivation of the Hamiltonian that does not require the full machinery of large deviation theory. We take as our starting point an LDP for stochastic hybrid systems due to Faggionato et al. [11, 12], which is equivalent to the LDP of Kifer [23]. (For completeness, we summarize a simplified version of Kifer's formulation in appendix B.) Using the Perron-Frobenius theorem and the calculus-of-variations, we evaluate the LDP rate function in terms of a classical action, whose equations of motion are given by the expected Hamiltonian dynamical system. We then consider some specific examples in order to illustrate our analysis (§5). Finally, in §6, we briefly indicate how the analysis can be extended to a multi-scale stochastic process, in which the slow dynamics is given by a piecewise SDE.

## 2. The model

Consider a one-dimensional stochastic hybrid system also called a piecewise deterministic Markov process (PDMP) [8, 23]. Its states are described by a pair

$(x, n) \in \Omega \times \{1, \dots, K\}$ , where  $x$  is a continuous variable in a connected, bounded domain  $\Omega \subset \mathbb{R}^d$  with a regular boundary  $\partial\Omega$  and  $n$  a discrete internal state variable taking values in  $\Gamma \equiv \{1, \dots, K\}$ . When the internal state is  $n$ , the system evolves according to the ordinary differential equation (ODE)

$$\dot{x} = \mathbf{F}_n(x), \quad (2.1)$$

where the vector field  $\mathbf{F}_n : \Omega \rightarrow \mathbb{R}^d$  is a continuous function, locally Lipschitz. That is, given a compact subset  $\mathcal{K}$  of  $\Omega$ , there exists a positive constant  $K_n(\mathcal{K})$  such that

$$|\mathbf{F}_n(x) - \mathbf{F}_n(y)| \leq K_n(\mathcal{K})|x - y|, \quad \forall x, y \in \Omega. \quad (2.2)$$

We assume that the dynamics of  $x$  is confined to the domain  $\Omega$  so that we have existence and uniqueness of a trajectory for each  $n$ . One final mild constraint is that the vector field does not have identical components anywhere in  $\Omega$ , that is, for any  $x \in \Omega$ , there exists at least one pair  $(n, m) \in \Gamma, n \neq m$  for which  $\mathbf{F}_m(x) \neq \mathbf{F}_n(x)$ .

In order to specify how the system jumps from one internal state to the other for each  $n \in \Gamma$ , the function  $W_{nm}(x)$  defined on  $\Gamma \times \Gamma \times \mathbb{R}$  with the following properties

- (i)  $\forall n, m, W_{nm}(\cdot)$  is  $C^1(\Omega)$
- (ii)  $\forall x, n, W_n(x)$  is a probability measure on  $\Gamma$  such that  $W_{nn}(x) = 0$ .

At each point  $x$  of  $\Omega$  this defines a time-homogeneous Markov chain with state space  $\Gamma$ . We make the further assumption that this chain is irreducible (i.e. that the  $K \times K$  matrix  $\mathbf{W}(x)$  is irreducible) for all  $x$  in  $\Omega$ , meaning that 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. This implies the existence of a unique invariant probability measure on  $\Gamma$ , noted  $\rho(x, n)$ , such that

$$\sum_{n \in \Gamma} \rho(x, n) W_{nm}(x) = \rho(x, m) \quad \forall m \in \Gamma, \forall x \in \Omega$$

or in matrix form:

$$\rho(x, \cdot) \mathbf{W}(x) = \rho(x, \cdot), \quad (2.3)$$

where  $\rho(x, \cdot)$  is the  $K$ -dimensional row vector of coordinates  $\rho(x, n)$ ,  $n \in \Gamma$ . The existence of the unique invariant measure can be seen as a consequence of the well-known Perron-Frobenius theorem<sup>‡</sup>.

The hybrid evolution of the system is described as follows. Suppose the system starts at time zero in the state  $(x_0, n_0)$ . Call  $x_0(t)$  the solution of (2.1) with  $n = n_0$  such that  $x_0(0) = x_0$ . Let  $\theta_1$  be the random variable such that

$$\mathbb{P}(\theta_1 > t) = \exp\left(-\frac{t}{\epsilon}\right).$$

Then in the random time interval  $[0, \theta_1)$  the state of the system is  $(x_0(s), n_0)$ . We draw a value of  $\theta_1$  from the corresponding probability density

$$p(t) = \frac{1}{\epsilon} \exp\left(-\frac{t}{\epsilon}\right).$$

<sup>‡</sup> A finite-dimensional, real irreducible square matrix with non-negative entries has a unique largest positive real eigenvalue (the Perron eigenvalue) and the corresponding eigenvector has strictly positive components [18].

We now choose an internal state  $n_1 \in \Gamma$  with probability  $W_{n_0 n_1}(x_0(\theta_1))$  and call  $x_1(t)$  the solution of the following Cauchy problem on  $[\theta_1, \infty)$ :

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

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

$$(x(t), n(t)) = (x_k(t), n_k) \quad \text{if } \theta_k \leq t < \theta_{k+1}. \quad (2.4)$$

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.

Given the above iterative definition of the stochastic hybrid process, let  $X(t)$  and  $N(t)$  denote the stochastic continuous and discrete variables, respectively, at time  $t$ ,  $t > 0$ , given the initial conditions  $X(0) = x_0, N(0) = n_0$ . Although the evolution of the continuous variable  $X(t)$  or the discrete variable  $N(t)$  is non-Markovian, it can be proven that the joint evolution  $(X(t), N(t))$  is a strong Markov process [8]. The corresponding infinitesimal generator  $L$  is easily shown to be equal to

$$Lg(x, n) = \mathbf{F}_n(x) \cdot \nabla_x g(x, n) + \frac{1}{\epsilon} \sum_{m \in \Gamma} W_{nm}(x)(g(x, m) - g(x, n)),$$

for regular functions  $g : \Omega \times \Gamma \rightarrow \mathbb{R}$ . We rewrite this as

$$Lg(x, n) = \mathbf{F}_n(x) \cdot \nabla_x g(x, n) + \frac{1}{\epsilon} \sum_{m \in \Gamma} A_{nm}(x)g(x, m),$$

where

$$A_{nm}(x) = W_{nm} - \delta_{nm}.$$

Note that  $\sum_{m \in \Gamma} A_{nm} = 0 \quad \forall n \in \Gamma$ , and, in matrix form,

$$\mathbf{A}(x) = \mathbf{W}(x) - \mathbf{I}_K, \quad (2.5)$$

where  $\mathbf{I}_K$  is the  $K \times K$  identity matrix. Note that (2.3) implies

$$\rho(x, \cdot) \mathbf{A}(x) = \mathbf{0}$$

In this paper, we focus on the stochastic dynamics in the limit of fast kinetics, that is,  $\epsilon \rightarrow 0$ .

Let us now define the averaged vector field  $\bar{\mathbf{F}} : \mathbb{R} \rightarrow \mathbb{R}$  by

$$\bar{\mathbf{F}}(x) = \sum_{n \in \Gamma} \rho(x, n) \mathbf{F}_n(x)$$

It can be shown [12] that, given the assumptions on the matrix  $\mathbf{W}$ , the functions  $\rho(x, n)$  on  $\Omega$  belong to  $C^1(\Omega)$  for all  $n \in \Gamma$  and that this implies that  $\bar{\mathbf{F}}(x)$  is locally Lipschitz. Hence, for all  $t \in [0, T]$ , the Cauchy problem

$$\begin{cases} \dot{x}(t) &= \bar{\mathbf{F}}(x(t)) \\ x(0) &= x_0 \end{cases} \quad (2.6)$$

has a unique solution for all  $x_0 \in \Omega$ . Intuitively speaking, one would expect the stochastic hybrid system (2.1) to reduce to the deterministic dynamical system (2.6) in the limit  $\epsilon \rightarrow 0$ . That is, for sufficiently small  $\epsilon$ , the Markov chain undergoes many jumps over a small time interval  $\Delta t$  during which  $\Delta x \approx 0$ , and thus the relative frequency of each discrete state  $n$  is approximately  $\rho(x, n)$ . This can be made precise in terms of a Law of Large Numbers for stochastic hybrid systems proven in [12].

### 3. WKB method and quasipotentials

For notational simplicity, we restrict ourselves in this section to the case  $d = 1$ .

Introduce the probability density  $\rho(x, n, t|x_0, n_0, 0)$  with

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

It follows that  $\rho$  evolves according to the forward differential Chapman-Kolmogorov (CK) equation [3, 16]

$$\frac{\partial \rho}{\partial t} = L_t \rho, \quad (3.1)$$

with (dropping the explicit dependence on initial conditions)

$$L_t \rho(x, n, t) = -\frac{\partial F_n(x) \rho(x, n, t)}{\partial x} + \frac{1}{\epsilon} \sum_{m \in \Gamma} A_{nm}^\top(x) \rho(x, m, t), \quad (3.2)$$

Suppose that the averaged equation (2.6) is bistable with a pair of stable fixed points  $x_\pm$  separated by an unstable fixed point  $x_0$  (see section 4 and Fig. 1 for an explicit example). Assume that the stochastic system is initially at  $x_-$ . On short time scales ( $t \ll 1/\epsilon$ ) the system rapidly converges to a quasistationary solution within the basin of attraction of  $x_-$ , which can be approximated by a Gaussian solution of the reduced FP equation obtained using a quasi-steady-state (QSS) diffusion or adiabatic approximation of the CK equation (3.1) [26]. However, on longer time scales, the survival probability slowly decreases due to rare transitions across  $x_0$  at exponentially small rates, which cannot be calculated accurately using the diffusion approximation. One thus has to work with the full CK equation (3.1) supplemented by an absorbing boundary conditions at  $x_0$ :

$$\rho(x_0, n, t) = 0, \text{ for all } n \in \Sigma, \quad (3.3)$$

where  $\Sigma \subset \Gamma$  is the set of discrete states  $n$  for which  $F_n(x_0) < 0$ . The initial condition is taken to be

$$\rho(x, n, 0) = \delta(x - x_-) \delta_{n, \bar{n}}. \quad (3.4)$$

Let  $T$  denote the (stochastic) first passage time for which the system first reaches  $x_0$ , 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  $x_0$ , that is,

$$\text{Prob}\{t > T\} = S(t) \equiv \int_{-\infty}^{x_0} \sum_{n \in \Gamma} \rho(x, n, t) dx.$$

The first passage time density is then

$$f(t) = -\frac{dS}{dt} = -\int_{-\infty}^{x_0} \sum_{n \in \Gamma} \frac{\partial \rho(x, n, t)}{\partial t} dx. \quad (3.5)$$

Substituting for  $\partial \rho / \partial t$  using the CK equation (3.1) shows that

$$f(t) = \int_{-\infty}^{x_0} \left[ \sum_{n \in \Gamma} \frac{\partial [F_n(x) \rho(x, n, t)]}{\partial x} \right] dx = \sum_{n \in \Gamma} F_n(x_0) \rho(x_0, n, t). \quad (3.6)$$

We have used  $\sum_n A_{nm}^\top(x) = 0$  and  $\lim_{x \rightarrow -\infty} F_n(x) \rho(x, n, t) = 0$ . The first passage time density can thus be interpreted as the probability flux  $J(x, t)$  at the absorbing boundary, since we have the conservation law

$$\sum_{n \in \Gamma} \frac{\partial \rho(x, n, t)}{\partial t} = -\frac{\partial J(x, t)}{\partial x}, \quad J(x, t) = \sum_{n \in \Gamma} F_n(x) \rho(x, n, t). \quad (3.7)$$

Suppose that the solution to the CK equation (3.1) with an absorbing boundary at  $x = x_0$  has the eigenfunction expansion

$$\rho(x, n, t) = \sum_{r \geq 0} C_r e^{-\mu_r t} V_r(x, n), \quad (3.8)$$

where  $\text{Re}[\mu_r] > 0$  for all  $r \geq 0$  and  $(\mu_r, V_r)$  are an eigenpair of the linear operator  $\mathcal{L}$  on the right-hand side of the CK equation (3.1):

$$\mathcal{L}V_r(x, n) \equiv \frac{d}{dx}(F_n(x)V_r(x, n)) - \frac{1}{\epsilon} \sum_{m \in \Gamma} A_{nm}^\top V_r(x, m) = \mu_r V_r(x, n), \quad (3.9)$$

together with the boundary conditions

$$V_r(x_0, n) = 0, \text{ for } n \in \Sigma. \quad (3.10)$$

Furthermore, assume that there exists a simple, real eigenvalue  $\mu_0$  such that  $0 < \mu_0 < \text{Re}[\mu_1] \leq \text{Re}[\mu_2] \leq \dots$  with  $\mu_0 \sim e^{-C/\epsilon}$ , whereas  $\text{Re}[\mu_r] = O(1)$  for  $r > 0$ . It follows that at large times we have the quasistationary approximation

$$\rho(x, n, t) \sim C_0 e^{-\mu_0 t} V_0(x, n). \quad (3.11)$$

Substituting such an approximation into equation (3.6) implies that  $f(t) \sim \mu_0 e^{-\mu_0 t}$  with  $\mu_0^{-1}$  the mean first passage time (MFPT). As shown elsewhere [20, 28], one can determine  $\mu_0$  by first using a WKB approximation to construct a quasistationary solution  $V^\epsilon(x, n)$  for which  $\mathcal{L}V^\epsilon = 0$  and  $V^\epsilon(x_0, n) \sim O(e^{-\tau/\epsilon})$ , and then performing an asymptotic expansion in order to match the quasistationary solution with the solution in a neighborhood of  $x_0$ . The WKB approximation  $V^\epsilon(x)$  takes the form

$$V^\epsilon(x, n) \sim \eta(x, n) \exp\left(-\frac{\Phi(x)}{\epsilon}\right), \quad (3.12)$$

where  $\Phi(x)$  is the quasipotential. Substituting into the time-independent version of equation (3.1) yields

$$\sum_{m \in \Gamma} (A_{nm}^\top(x) + \Phi'(x)\delta_{nm}F_m(x)) \eta(x, m) = \epsilon \frac{dF_n(x)\eta(x, n)}{dx}, \quad (3.13)$$

where  $\Phi' = d\Phi/dx$ . Introducing the asymptotic expansions  $\eta \sim \eta^{(0)} + \epsilon\eta^{(1)} + \dots$  and  $\Phi \sim \Phi_0 + \epsilon\Phi_1 + \dots$ , the leading order equation is

$$\sum_{m \in \Gamma} A_{nm}^\top(x)\eta^{(0)}(x, m) = -\Phi_0'(x)F_n(x)\eta^{(0)}(x, n). \quad (3.14)$$

It is well known that for SDEs in the weak-noise limit, the associated quasipotential satisfies a Hamilton-Jacobi equation [15, 32]. An analogous result also holds for jump Markov processes [10, 19, 24]. We now show that equation (3.14) can be reformulated as a Hamilton-Jacobi equation for the quasipotential  $\Phi_0$  of a stochastic hybrid system. First, we introduce the family of eigenvalue equations

$$\sum_{m \in \Gamma} A_{nm}^\top(x)z_m(x, p) + pF_n(x)z_n(x, p) = \lambda(x, p)z_n(x, p), \quad (3.15)$$

which are parameterized by the pair  $(x, p)$  with  $p$  an auxiliary variable. An application of the Perron-Frobenius theorem along similar lines to Appendix A establishes that, for each fixed  $(x, p)$ , there exists a unique (Perron) eigenvalue whose associated eigenvector is strictly positive. Therefore, in the following we take  $\lambda(x, p)$  to be the Perron eigenvalue and  $z_n(x, p), n = 1, \dots, |\Gamma|$ , to be the corresponding eigenvector.



Comparison of equation (3.14) with (3.15) then shows that we can make the identifications  $\Phi'_0(x) = p$ ,  $\eta^{(0)}(x, m) = z_m(x, p)$  and  $\lambda(x, p) = 0$ . That is, the quasipotential is the solution of the Hamilton-Jacobi equation

$$\lambda(x, \Phi'_0(x)) = 0. \quad (3.16)$$

This is equivalent to finding zero energy solutions of Hamilton's equations

$$\dot{x} = \frac{\partial \lambda(x, p)}{\partial p}, \quad \dot{p} = -\frac{\partial \lambda(x, p)}{\partial x},$$

and identifying  $\Phi_0$  as the action along the resulting solution curve  $(x(t), p(t))$ :

$$\Phi_0(x) = \int_{\bar{x}}^x p(x') dx'.$$

Given the quasi-potential  $\Phi_0$ , the mean first passage time  $\tau$  to escape from the fixed point at  $x_-$  can be calculated by considering higher order terms in the WKB approximation, and using matched asymptotics to deal with the absorbing boundary at  $x_0$ . One finds that  $\tau$  takes the general Arrhenius form [20, 28]

$$\tau \sim \frac{\chi(x_0, x_-)}{\sqrt{|\Phi_0''(x_0)| |\Phi_0''(x_-)|}} e^{[\Phi_0(x_0) - \Phi_0(x_-)]/\epsilon},$$

where  $\chi$  is an appropriate prefactor. Hence, the WKB method provides a powerful calculational tool. On the other hand, there is no *a priori* justification for interpreting the quasipotential and its associated Hamiltonian in terms of an underlying variational problem for optimal paths in the space of stochastic trajectories. This becomes crucial when solving escape problems in higher dimensions, since a metastable state is now surrounded by a non-trivial boundary (rather than a single point) and one needs to determine the relative weighting of optimal paths crossing different points on the boundary.

#### 4. Classical action from a large deviation principle

A powerful methodology for rigorously deriving a variational principle for optimal paths is large deviation theory [15]. The particular application of large deviation theory to stochastic hybrid systems has been developed in considerable detail by Kifer [23], who shows how to derive the action functional with Hamiltonian given by the Perron eigenvalue  $\lambda(x, p)$  from an underlying large deviation principle (LDP). However, the technical nature of the analysis makes it difficult for a non-expert to understand. Therefore, in this section and appendix A we present an alternative derivation of the Hamiltonian system starting from the LDP of Faggionato et al. [11, 12]. The advantage of the latter LDP, which is equivalent to the LDP of Kifer (see [23] and appendix B), is that it is relatively straightforward to write down. As far as we are aware, our derivation of the Hamiltonian action functional is new, and it provides additional insights into the origin and meaning of the ‘‘momentum’’ variable  $p$ . Finally, note that yet another method for deriving the action functional is to use formal path-integral methods [3]. One of the useful features of the path-integral formulation is that it can also be used to implement a variety of approximation schemes, including diffusion approximations and perturbation expansions of moment equations [5].

In order to write down the LDP of Faggionato et al [11, 12], it is first necessary to introduce some notation. Let  $\mathcal{M}_+([0, T])$  denote the space of non-negative finite measures on the interval  $[0, T]$  with a density w.r.t to the Lebesgue measure.

It is convenient to isolate a special subset  $\mathcal{Y}$  of  $C([0, T], \Omega) \times \mathcal{M}_+([0, T])^\Gamma$  as follows. Consider the pairs  $\{(x(t), \psi(t))\}_{t \in [0, T]}$  such that the  $K$ -dimensional vector  $\{\psi(t)\}_{t \in [0, T]}$  is an element of the product space  $\mathcal{M}_+([0, T])^\Gamma$  satisfying the following extra condition

$$\sum_{n \in \Gamma} \psi_n(t) = 1,$$

where we have identified the measures  $\psi_n(t)dt$  with their densities  $\psi_n(t)$ .

A particular realization of the stochastic process,  $\{(x(t), n(t))\}_{t \in [0, T]}$ , then lies in  $\mathcal{Y}$  with

$$\psi_n(t) = 1_{\{n(t)=n\}} \equiv \begin{cases} 1, & \text{if } n(t) = n, \\ 0, & \text{if } n(t) \neq n \end{cases} \quad (4.1)$$

and

$$x(t) = x_0 + \int_0^t \sum_{n \in \Gamma} \psi_n(s) \mathbf{F}_n(x(s)) ds. \quad (4.2)$$

$\mathcal{Y}$  contains both the set of trajectories of the stochastic hybrid system with  $\psi_n(t)$  given by equation (4.1) and  $n(t)$  evolving according to the Markov chain, and the solution  $x^*(t)$  of the averaged system (2.6) for which  $\psi_n = \psi_n^*$  with  $\psi_n^*(t) = \rho(x^*(t), n)$ . It can be proven that  $\mathcal{Y}$  is a compact subspace of  $C([0, T], \Omega) \times \mathcal{M}_+([0, T])^\Gamma$  with topology defined by the metric [12]

$$\begin{aligned} & d(\{(x(t), n(t))\}_{t \in [0, T]}, \{(\tilde{x}(t), \tilde{n}(t))\}_{t \in [0, T]}) \\ &= \sup_{t \in [0, T]} |x(t) - \tilde{x}(t)| + \sum_{n \in \Gamma} \sup_{0 \leq t \leq T} \left| \int_0^t [\psi_n(s) - \tilde{\psi}_n(s)] ds \right|. \end{aligned} \quad (4.3)$$

Finally, we take  $P_{x_0}^\epsilon$  to be the probability density functional or law of the set of trajectories  $\{x(t)\}_{t \in [0, T]} \in C([0, T], \Omega)$ .

The following large deviation principle then holds [11, 12]: Given an element  $\{x(t)\}_{t \in [0, T]} \in C([0, T], \Omega)$ ,

$$\mathbb{P}_{x_0}^\epsilon [\{x(t)\}_{t \in [0, T]}] \sim e^{-J_T(\{x(t)\}_{t \in [0, T]})/\epsilon}. \quad (4.4)$$

where the rate function  $J_{[0, T]} : C([0, T], \Omega) \rightarrow [0, \infty)$  is given by

$$\begin{aligned} & J_T(\{x(t)\}_{t \in [0, T]}) \\ &= \inf_{\{\psi(t)\}_{t \in [0, T]} : \{x(t), \psi(t)\}_{t \in [0, T]} \in \mathcal{Y}_{x_0}} \int_0^T j(x(t), \psi(t)) dt \end{aligned} \quad (4.5)$$

and

$$j(x, \psi) = \sup_{\mathbf{R} \in (0, \infty)^\Gamma} \sum_{(n, m) \in \Gamma \times \Gamma} \psi_n W_{nm}(x) \left[ 1 - \frac{R_m}{R_n} \right]. \quad (4.6)$$

Here the symbol  $\sim$  means asymptotic logarithmic equivalence in the limit  $\epsilon \rightarrow 0$ .

A key idea behind the above LDP is that a slow dynamical process coupled to the fast Markov chain on  $\Gamma$  rapidly samples the different discrete states of  $\Gamma$  according to some non-negative measure  $\psi$ . In the limit  $\epsilon \rightarrow 0$ , one has  $\psi \rightarrow \rho$ , where  $\rho$  is the invariant measure of the Markov chain defined by (2.3). On the other hand, for small but non-zero  $\epsilon$ ,  $\psi$  is itself distributed according to a LDP, whereby one averages the different functions  $\mathbf{F}_n(x)$  over the measure  $\psi$  to determine the dynamics of the slow system. In most biological applications, one is not interested in the internal discrete

state of the system, that is, one only observes the statistical behavior of the continuous variable  $x(t)$ . For example,  $x(t)$  could represent the membrane voltage of a neuron [29] or the position of a molecular motor along a microtubular track [27, 28]. Faggionato et al. [11, 12] explicitly calculated the rate function (4.5) for a restricted class of stochastic hybrid systems, whose stationary density is exactly solvable. One major constraint on this class of model is that the vector field of the piecewise deterministic system has non-vanishing components within a given confinement domain. However, this constraint does not hold for biological systems such as ion channels [4, 20, 29], gene networks [21, 25], and neural networks [1] (Faggionato et al. were motivated by a model of molecular motors that is exactly solvable. Such a solvability condition also breaks down for molecular motors when local chemical signaling is taken into account [27].)

In appendix A we show (without the restrictions of Faggionato et al. [11, 12]) that the rate function  $J_T(\{x(t)\}_{t \in [0, T]})$  of the non-Markov process  $\{x(t)\}_{t \in [0, T]}$  can be written in the form of an action

$$J_T(\{x(t)\}_{t \in [0, T]}) = \int_0^T L(x, \dot{x}) dt, \quad (4.7)$$

with Lagrangian given by

$$L(x, \dot{x}) = \langle \boldsymbol{\mu}(x, \dot{x}), \dot{x} \rangle - \lambda(x(t), \boldsymbol{\mu}(x, \dot{x})),$$

Here  $\lambda(x, \boldsymbol{\mu})$  is the Perron eigenvalue of the linear equation (3.15) (suitably generalized, see Appendix A) for  $p \rightarrow \boldsymbol{\mu}$ , that is,

$$\mathbf{A}(x)\mathbf{R}(x, \boldsymbol{\mu}) + (\mathbf{F}(x)\boldsymbol{\mu}) \circ \mathbf{R}(x, \boldsymbol{\mu}) = \lambda(x, \boldsymbol{\mu})\mathbf{R}(x, \boldsymbol{\mu}). \quad (4.8)$$

Here, for any  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^K$ ,

$$[\mathbf{a} \circ \mathbf{b}]_n \equiv [\text{diag}(\mathbf{a})\mathbf{b}]_n = a_n b_n, \quad n = 1, \dots, K, \quad (4.9)$$

the  $K$ -dimensional vector  $\mathbf{F}(x)\boldsymbol{\mu}$  is the product of the  $K \times d$  matrix  $\mathbf{F}(x)$  whose  $K$  rows are the  $d$ -dimensional vectors  $\mathbf{F}_m(x)$ ,  $m = 1, \dots, K$ , with the  $d$ -dimensional vector  $\boldsymbol{\mu}$ , and  $\boldsymbol{\mu} = \boldsymbol{\mu}(x, \dot{x})$  is the solution of the invertible equation

$$\dot{x} = \sum_{m \in \Gamma} \psi_m(x, \boldsymbol{\mu}) \mathbf{F}_m(x), \quad (4.10)$$

with

$$\psi_m(x, \boldsymbol{\mu}) = z_m(x, \boldsymbol{\mu}) R_m(x, \boldsymbol{\mu}), \quad (4.11)$$

where  $\mathbf{z}$  is the positive eigenvector of the adjoint equation

$$\mathbf{A}^\top(x)\mathbf{z}(x, \boldsymbol{\mu}) + (\mathbf{F}(x)\boldsymbol{\mu}) \circ \mathbf{z}(x, \boldsymbol{\mu}) = \lambda(x, \boldsymbol{\mu})\mathbf{z}(x, \boldsymbol{\mu}) \quad (4.12)$$

under the normalizations  $\sum_n z_n = 1$  and  $\langle \mathbf{z}, \mathbf{R} \rangle = \sum_m z_m(x, \boldsymbol{\mu}) R_m(x, \boldsymbol{\mu}) = 1$ .

Given the Lagrangian  $L$ , we can determine the Hamiltonian  $H$  according to the Fenchel-Legendre transformation

$$H(x, \mathbf{p}) = \sup_{y \in \mathbb{R}^d} [\langle \mathbf{p} - \boldsymbol{\mu}(x, y), y \rangle + \lambda(x, \boldsymbol{\mu}(x, y))].$$

Evaluating the right-hand side yields the equation

$$\mathbf{p} - \boldsymbol{\mu}(x, y) + \frac{\partial \boldsymbol{\mu}}{\partial y} \left[ \frac{\partial \lambda}{\partial \boldsymbol{\mu}} - y \right] = 0 \quad (4.13)$$

with

$$y = \sum_{m \in \Gamma} \psi_m(x, \boldsymbol{\mu}) \mathbf{F}_m(x).$$

Differentiating equation (4.8) with respect to  $\boldsymbol{\mu}$  gives

$$\mathbf{L}(x, \boldsymbol{\mu}) \frac{\partial \mathbf{R}}{\partial \boldsymbol{\mu}} = \mathbf{R} \frac{\partial \lambda}{\partial \boldsymbol{\mu}} - \text{diag}(\mathbf{R}) \mathbf{F}(x), \quad (4.14)$$

with  $\mathbf{L}(x, \boldsymbol{\mu}) = \mathbf{L}(x, \mathbf{p})|_{\mathbf{p}_m = \langle \boldsymbol{\mu}, \mathbf{F}_m \rangle}$ . Since the adjoint of the linear operator on the left-hand side has a one-dimensional null space spanned by  $\mathbf{z}$ , see (4.12), multiplying both sides of the equation by  $\mathbf{z}^\top$ , it follows from the normalization  $\sum_m z_m R_m = 1$  and (4.11) that

$$\frac{\partial \lambda(x, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^\top = \sum_{m \in \Gamma} \psi_m(x, \boldsymbol{\mu}) \mathbf{F}_m(x) = y.$$

Equation (4.13) thus shows that  $\mathbf{p} = \boldsymbol{\mu}$ , and we can identify  $\mathbf{p}$  as the ‘‘conjugate momentum’’ of the Hamiltonian

$$H = \lambda(x, \mathbf{p}), \quad (4.15)$$

where  $\lambda(x, \mathbf{p})$  is the Perron eigenvalue of the linear equation (4.8).

## 5. Examples

In this section we illustrate the Hamiltonian structure of stochastic hybrid systems by considering a few explicit models.

### 5.1. Binary model

We begin with the simple example of a binary stochastic hybrid process in dimension 1 ( $d = 1$ ) with two discrete states  $n = 0, 1$ , which was analyzed in some detail by Faggionato et al [11] using a different method. The latter authors exploited the fact that the model is exactly solvable, in the sense that the stationary density of the corresponding Chapman-Kolmogorov equation can be computed explicitly, and used a fluctuation-dissipation theorem to determine the Hamiltonian and quasipotential. Here we will obtain the same results more directly by calculating the Perron eigenvalue. One biological application of the binary model is to the bidirectional transport of a molecular motor along a one-dimensional microtubular track, in which  $x$  represents the spatial location of the motor on the track and the two discrete states represent the motor moving either towards the  $+$  end or  $-$  end of the track. (In more complex models, the discrete space  $\Gamma$  represents multiple internal conformational states of the motor, each of which has an associated velocity on the track [1].)

Suppose that the continuous variable evolves according to piecewise dynamics on some finite interval  $(a, b)$ ,

$$\dot{x} = F_n(x), \quad n = 0, 1, \quad (5.1)$$

with  $F_0, F_1$  continuous and locally Lipschitz. Suppose that  $F_0, F_1$  are non-vanishing within the interval  $(a, b)$ , and  $F_n(a) \geq 0, F_n(b) \leq 0$  for  $n = 0, 1$ ; the dynamics is then confined to  $(a, b)$ . Denote the transition rates of the two-state Markov chain by  $\omega_\pm(x)$  with

$$\{n = 0\} \xrightleftharpoons[\omega_-(x)]{\omega_+(x)} \{n = 1\}.$$

The stationary measure of the Markov chain is given by

$$\rho(x, 0) = \frac{\omega_+(x)}{\omega_-(x) + \omega_+(x)}, \quad \rho(x, 1) = \frac{\omega_-(x)}{\omega_-(x) + \omega_+(x)}. \quad (5.2)$$

The linear equation (4.8) can be written as the two-dimensional system

$$\begin{pmatrix} -\omega_+(x) + pF_0(x) & \omega_+(x) \\ \omega_-(x) & -\omega_-(x) + pF_1(x) \end{pmatrix} \begin{pmatrix} R_0 \\ R_1 \end{pmatrix} = \lambda \begin{pmatrix} R_0 \\ R_1 \end{pmatrix}. \quad (5.3)$$

The corresponding characteristic equation is

$$0 = \lambda^2 + \lambda[\omega_+(x) + \omega_-(x) - p(F_0(x) + F_1(x))] \\ + (pF_1(x) - \omega_-(x))(pF_0(x) - \omega_+(x)) - \omega_-(x)\omega_+(x).$$

It follows that the Perron eigenvalue is given by

$$\lambda(x, p) = \frac{1}{2} \left[ \Sigma(x, p) + \sqrt{\Sigma(x, p)^2 - 4\gamma(x, p)} \right] \quad (5.4)$$

where

$$\Sigma(x, p) = p(F_0(x) + F_1(x)) - [\omega_+(x) + \omega_-(x)],$$

and

$$\gamma(x, p) = (pF_1(x) - \omega_-(x))(pF_0(x) - \omega_+(x)) - \omega_-(x)\omega_+(x).$$

A little algebra shows that

$$D(x, p) \equiv \Sigma(x, p)^2 - 4\gamma(x, p) = [p(F_0 - F_1) - (\omega_+ - \omega_-)]^2 + \omega_+\omega_- > 0$$

so that as expected  $\lambda$  is real. Hence, from Hamilton's equations

$$\begin{aligned} \dot{x} &= \frac{\partial \lambda(x, p)}{\partial p} \\ &= \frac{F_0(x) + F_1(x)}{2} + \frac{\partial D(x, p)}{\partial p} \frac{1}{2\sqrt{D(x, p)}} \\ &= \frac{F_0(x) + F_1(x)}{2} + \frac{F_0(x) - F_1(x)}{2} \frac{p(F_0 - F_1) - (\omega_+ - \omega_-)}{\sqrt{[p(F_0 - F_1) - (\omega_+ - \omega_-)]^2 + \omega_+\omega_-}} \end{aligned} \quad (5.5)$$

which is the same result as obtained in example 10.4 of Faggionato et al [11]. Moreover, writing

$$\dot{x} = F_0(x)\psi_0(x) + F_1(x)\psi_1(x),$$

we see that

$$\psi_0(x) = \frac{1}{2} \left[ 1 + \frac{p(F_0 - F_1) - (\omega_+ - \omega_-)}{\sqrt{[p(F_0 - F_1) - (\omega_+ - \omega_-)]^2 + \omega_+\omega_-}} \right], \quad (5.6)$$

and

$$\psi_1(x) = \frac{1}{2} \left[ 1 - \frac{p(F_0 - F_1) - (\omega_+ - \omega_-)}{\sqrt{[p(F_0 - F_1) - (\omega_+ - \omega_-)]^2 + \omega_+\omega_-}} \right], \quad (5.7)$$

so that  $\psi_{0,1}(x) \geq 0$  with  $\psi_0(x) + \psi_1(x) = 1$ .

## 5.2. Stochastic ion channels

Another important example of stochastic hybrid systems is a conductance-based model of a neuron, in which the stochastic opening of membrane ion channels generates a stochastic ionic current that drives the membrane voltage. It is then possible that ion channel noise induces spontaneous action potentials (SAPs), which can have a large effect on a neuron's function [4, 6, 7, 14, 17, 20, 29]. If SAPs are too frequent, a neuron cannot reliably perform its computational role. Hence, ion channel noise imposes a fundamental limit on the density of neural tissue. Smaller neurons must function with fewer ion channels, making ion channel fluctuations more significant and more likely to cause a SAP. Here we will consider the simple case of a single type of ion channel, namely, a fast sodium (Na) channel, which was previously analyzed using WKB methods [20]. Let  $x(t)$  denote the membrane voltage of the neuron at time  $t$  and  $N$  be the fixed number of sodium channels. We assume that each channel can either be open ( $O$ ) or closed ( $C$ ), and can switch between each state according to the kinetic scheme



with voltage-dependent transition rates. (A more detailed biophysical model would need to treat each ion channel as a cluster of subunits rather than a single unit. In other words, the Markov chain of events associated with opening and closing of an ion channel would involve transitions between more than two internal states,.) The stochastic membrane voltage is taken to evolve according to the piecewise deterministic equation

$$\frac{dx}{dt} = F_n(x) \equiv \frac{n}{N}f(x) - g(x), \quad (5.9)$$

where

$$f(x) = g_{\text{Na}}(V_{\text{Na}} - x), \quad g(x) = -g_L[V_L - x] - I.$$

Here  $g_{\text{Na}}$  is the maximal conductance of a sodium channel and  $V_{\text{Na}}$  is the corresponding membrane reversal potential. Similarly,  $g_L$  and  $V_L$  are the effective maximal conductance and reversal potential of any other currents, which are assumed to be independent of the opening and closing of ion channels, and  $I$  is an external current. The four quantities ( $g_{\text{Na}}, g_L, V_{\text{Na}}, V_L$ ) are taken to be constants. Since the right-hand side of (5.9) is negative for large  $x$  and positive for small  $x$ , it follows that the voltage  $x$  is confined to some interval  $\Omega = [x_L, x_R]$ . The function  $F_n(x)$  is clearly continuous and locally Lipschitz.

In this example the space  $\Gamma$  of discrete states is the set of integers  $\{n = 0, 1, \dots, N\}$  and the Markov chain is given by a birth-death process:

$$n \xrightleftharpoons[\omega_+(n,x)/\epsilon]{\omega_-(n,x)/\epsilon} n \pm 1, \quad n \xrightarrow{\omega_-(n,x)/\epsilon} n - 1 \quad (5.10)$$

with transition rates

$$\omega_+(x, n) = \alpha(x)(N - n), \quad \omega_-(x, n) = \beta(x)n. \quad (5.11)$$

The small parameter  $\epsilon$  reflects the fact that sodium channels open at a much faster rate than the relaxation dynamics of the voltage [20]. It follows that the matrix  $\mathbf{A}(x)$  for fixed  $x$  is tridiagonal matrix with

$$A_{n-1,n}(x) = \omega_+(x, n-1), \quad A_{nn}(x) = -\omega_+(x, n) - \omega_-(x, n), \quad A_{n+1,n}(x) = \omega_-(x, n+1) \quad (5.12)$$

for  $n = 0, 1, \dots, N$ . It is straightforward to show that the Markov chain is ergodic with unique invariant measure (for fixed  $n$ ) given by

$$\rho(x, n) = \frac{N!}{(N-n)!n!} a(x)^n b(x)^{N-n}, \quad (5.13)$$

with

$$a(x) = \frac{\alpha(x)}{\alpha(x) + \beta(x)}, \quad b(x) = \frac{\beta(x)}{\alpha(x) + \beta(x)}. \quad (5.14)$$

The above stochastic hybrid system satisfies all of the conditions specified in section 2. Hence, the law of large numbers implies that in the mean-field limit  $\epsilon \rightarrow 0$ , we obtain the deterministic kinetic equation

$$\frac{dx}{dt} = \bar{F}(x) \equiv a(x)f(x) - g(x) \quad (5.15)$$

where

$$a(x) = \langle n \rangle / N, \quad \langle n \rangle = \sum_{n=1}^N n \rho(x, n),$$

and  $\rho$  is the stationary density (5.13). One of the features of the averaged model is that it can exhibit bistability for a range of physiologically reasonable parameter values. This is illustrated in Fig. 1, where we plot the deterministic potential  $U(x) = -d\bar{F}/dx$  as a function of  $x$ . Here  $x_-$  represents a resting state of the neuron, whereas  $x_+$  represents an active state; noise-induced transitions from  $x_-$  to  $x_+$  can be interpreted in terms of the initiation of a spontaneous action potential. Elsewhere, WKB methods and matched asymptotics have been used to calculate the MFPT to escape from  $x_-$ . Here, we will focus on the quasipotential and its relation to the Perron eigenvalue

Substituting the explicit expressions for  $\mathbf{A}$  and  $F_n(x)$  into the adjoint equation (4.12), yields the following equation for the Perron eigenvalue  $\lambda$  and the left eigenvector  $\mathbf{z}$ :

$$\begin{aligned} (N-n+1)\alpha(x)z_{n-1} - [\lambda + n\beta(x) + (N-n)\alpha(x)]z_n \\ + (n+1)\beta(x)z_{n+1} = -p \left( \frac{n}{N}f(x) - g(x) \right) z_n \end{aligned} \quad (5.16)$$

Consider the trial solution [4]

$$z_n(x, p) = \frac{\Gamma(x, p)^n}{(N-n)!n!}, \quad (5.17)$$

which yields the following equation relating  $\Gamma$  and  $\lambda$ :

$$\frac{n\alpha}{\Gamma} + \Gamma\beta(N-n) - \lambda - n\beta - (N-n)\alpha = -p \left( \frac{n}{N}f - g \right).$$

Collecting terms independent of  $n$  and terms linear in  $n$  yields the pair of equations

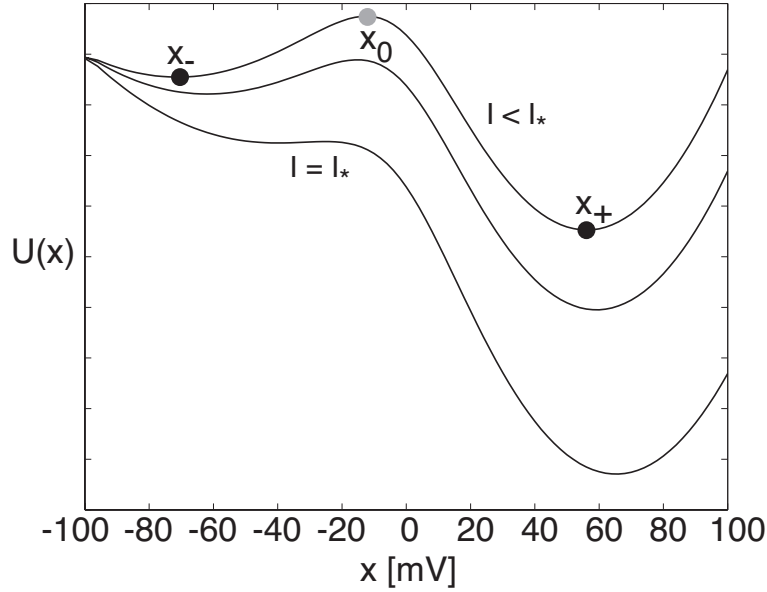
$$p = -\frac{N}{f(x)} \left( \frac{1}{\Gamma(x, p)} + 1 \right) (\alpha(x) - \beta(x)\Gamma(x, p)), \quad (5.18)$$

and

$$\lambda(x, p) = -N(\alpha(x) - \Gamma(x, p)\beta(x)) - pg(x). \quad (5.19)$$

Eliminating  $\Gamma$  from these equation gives

$$p = \frac{1}{f(x)} \left( \frac{N\beta(x)}{\lambda(x, p) + N\alpha(x) + pg(x)} + 1 \right) (\lambda(x, p) + pg(x))$$



**Figure 1.** Plot of deterministic potential  $U(x)$  as a function of voltage  $x$  for different values of the external stimulus current  $I$ . Parameter values are  $V_{\text{Na}} = 120$  mV,  $V_L = -62.3$  mV,  $g_{\text{Na}} = 4.4$  mS/cm<sup>2</sup>,  $g_L = 2.2$  mS/cm<sup>2</sup>, and  $\alpha(x) = \beta \exp[(x - v_1)/v_2]$  with  $\beta = 0.8$  s<sup>-1</sup>,  $v_1 = -1.2$  mV,  $v_2 = 18$  mV.

This yields a quadratic equation for  $\lambda$  of the form

$$\lambda^2 + \sigma(x)\lambda - h(x, p) = 0. \quad (5.20)$$

with

$$\begin{aligned} \sigma(x) &= (2g(x) - f(x)) + N(\alpha(x) + \beta(x)), \\ h(x, p) &= p[-N\beta(x)g(x) + (N\alpha(x) + pg(x))(f(x) - g(x))]. \end{aligned}$$

Given the Perron eigenvalue, we can determine the quasipotential  $\Phi(x)$  by solving the Hamilton-Jacobi equation  $\lambda(x, \partial_x \Phi) = 0$ . Equation (5.20) then yields the reduced Hamilton-Jacobi equation

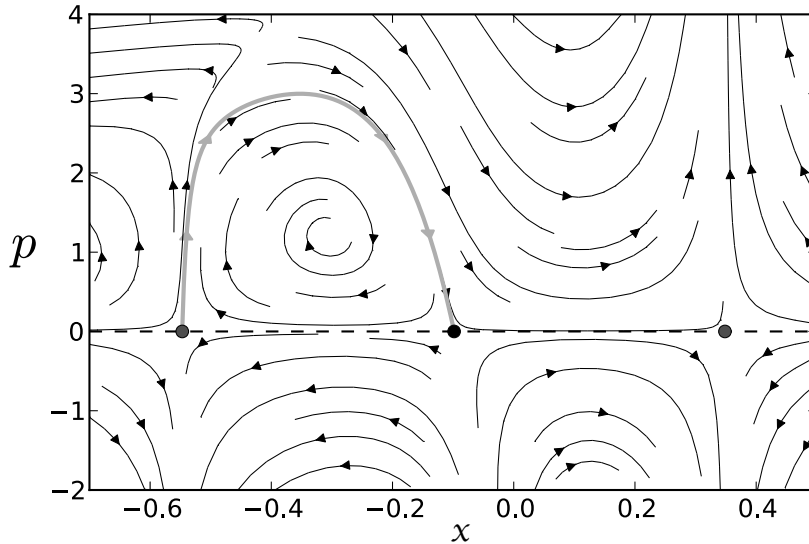
$$h(x, \partial_x \Phi_0) = 0. \quad (5.21)$$

The latter is precisely the equation for the quasipotential previously derived using WKB methods [29]. It has the following pair of solutions for  $\Phi'_0 = \partial_x \Phi_0$ :

$$\Phi'_0 = 0 \text{ and } \Phi'_0(x) = -N \frac{\alpha(x)f(x) - (\alpha(x) + \beta)g(x)}{g(x)(f(x) - g(x))}. \quad (5.22)$$

The trivial solution  $\Phi_0 = \text{constant}$  occurs along deterministic trajectories, which converge to the fixed point, whereas the non-trivial solution for  $\Phi_0(x)$  occurs along the most likely escape trajectories. In Fig. 2 we show solutions to Hamilton's equations in the  $(x, p)$ -plane, highlighting the zero energy maximum likelihood curve linking  $x_-$  and  $x_0$ .





**Figure 2.** Phase portrait of Hamilton's equations of motion for the ion channel model with Hamiltonian given by the Perron eigenvalue (5.19). The zero energy solution representing the maximum likelihood path of escape is shown as the gray curve. Dimensionless units are used with same parameter values as Fig. 1 and  $I = 0$ .

## 6. Multi-scale stochastic process

So far we have assumed that the slow process is piecewise deterministic. However, one of the useful features of taking the Lagrangian LDP [11, 12] as our starting point is that it is relatively straightforward to extend our analysis to the case where the slow process is a piecewise SDE. First, recall that the key idea behind the Faggionato et al. Lagrangian construction is that the slow dynamical process coupled to the fast Markov chain on  $\Gamma$  rapidly samples the different discrete states of  $\Gamma$  according to some non-negative measure  $\psi$ . In order to extend this construction to a piecewise SDE, it is necessary to take account of the fact that there are now two levels of stochasticity. That is, after averaging the transition rates of the drift and variance of the SDE with respect to a given measure  $\psi$ , the resulting system is still stochastic. Since the slow system operates in a weak noise regime, it follows that one can apply an LDP to the slow system for a given  $\psi$ . The LDP for the full system is then obtained by combining the rate function of the slow system with the supremum rate function for  $\psi$ . In this final section, we sketch how to extend the analysis to a piecewise SDE. (Approaching large deviation theory for multi-scale stochastic processes in terms of solutions to an eigenvalue problem has also been considered by Feng and Kurtz [13].) As in section 3, we consider only the case  $d = 1$ .

Consider the piecewise Ito SDE

$$dX(t) = F_n(X) + \sqrt{\epsilon}\sigma_n(X)dW(t), \quad (6.1)$$

where  $n \in \Gamma$  and  $W(t)$  is a Wiener process. The drift term  $F_n(X)$  and diffusion term  $\sigma_n(X)$  are both taken to be Lipschitz. When the SDE is coupled to the fast discrete process on  $\Gamma$ , the stochastic dynamics is described by a differential Chapman-Kolmogorov equation. That is, writing

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

we have

$$\frac{\partial \rho(x, n, t)}{\partial t} = -\frac{\partial}{\partial x} [F_n(x)\rho(x, n, t)] + \frac{\epsilon}{2} \frac{\partial^2}{\partial x^2} [\sigma_n^2(x)\rho(x, n, t)] + \frac{1}{\epsilon} \sum_m A_{nm}^\top(x)\rho(x, m, t). \quad (6.2)$$

We define the measure space  $\mathcal{M}_+([0, T])$  as before, but now modify the definition of the subspace  $\mathcal{Y}_{x_0} \subset C([0, T]) \times \mathcal{M}_+([0, T])^\Gamma$  by taking it to be the set of stochastic trajectories satisfying

$$\begin{aligned} dX(t) &= \sum_{n=1}^K \psi_n F_n(X) + \sqrt{\epsilon} \sqrt{\sum_{n=1}^K \psi_n(t) \sigma_n^2(X)} dW(t) \\ &\equiv F(X, \psi) + \sqrt{\epsilon} \sigma^2(X, \psi) dW(t) \end{aligned} \quad (6.3)$$

for  $\psi \in \mathcal{M}_+([0, T])^\Gamma$ . Such a space contains the set of trajectories of the SDE (6.1) with  $\psi_n(t)$  given by equation (4.1) and  $n(t)$  evolving according to the Markov chain on  $\Gamma$ . Consider a particular realization of the Wiener process  $W(t)$  on  $[0, T]$ , which is independent of  $\{x(t), \psi(t)\}_{[0, T]}$ . For a given realization, one can write down an LDP along identical lines to the case of a piecewise deterministic system, see equation (4.4). Assuming that we can then average with respect to the Wiener process, we obtain an additional contribution to the rate function so that

$$J_T(\{x(t)\}_{t \in [0, T]}) = \inf_{\{\psi(t)\}_{t \in [0, T]}} J_T(\{(x(t), \psi(t))\}_{t \in [0, T]}). \quad (6.4)$$

with

$$J_T(\{(x(t), \psi(t))\}_{t \in [0, T]}) = \int_0^T j(x(t), \psi(t)) dt + A_T(\{(x(t), \psi(t))\}_{t \in [0, T]}) \quad (6.5)$$

for  $j(x, \psi)$  given by equation (4.6) and

$$A_T(\{(x(t), \psi(t))\}_{t \in [0, T]}) = \int_0^T \frac{(\dot{x} - F(x, \psi))^2}{2\sigma^2(x, \psi)} dt. \quad (6.6)$$

Equation (6.6) is the well known action functional for a one-dimensional SDE in the case of a fixed, time-independent  $\psi$  [15]. One can evaluate the rate function (6.4) along similar lines to the case of a piecewise deterministic system (see appendix A.3) to obtain a classical action with Hamiltonian given by the Perron eigenvalue  $\lambda(x, p)$  of the linear equation

$$\sum_m A_{nm}(x) R_m(x, p) + (pF_n(x) + p^2\sigma_n^2/2) R_n(x, p) = \lambda(x, p) R_n(x, p). \quad (6.7)$$

## Acknowledgements

We thank Jay Newby for drawing our attention to the work of Yuri Kifer [22, 23] and producing Fig. 2. The work was conducted while PCB was visiting the NeuroMathComp group where he holds an INRIA International Chair. PCB was also partially supported by the NSF (DMS-1120327). OF was partially supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 269921 (BrainScaleS), no. 318723 (Mathematics), and by the ERC advanced grant NerVi no. 227747.

## Appendix A. Classical action from the Large Deviation Principle

In this appendix we present the details of deriving the classical action from the LDP given by equation (4.4). We first summarize a few useful properties of  $j(x, \psi)$  defined by equation (4.6). Since the double sum in equation (4.6) excludes diagonal terms, we introduce the set  $\Gamma_\Delta = \Gamma \times \Gamma \setminus \Delta$  where  $\Delta$  is the diagonal of  $\Gamma \times \Gamma$ . Let  $c_{nm} = \psi_n W_{nm}(x)$ . Suppose that  $\psi$  is a strictly positive measure,  $\psi_n > 0$  for all  $n \in \Gamma$ . It then follows from the properties of the transition matrix  $\mathbf{W}$  that the mapping  $c : \Gamma_\Delta \rightarrow [0, \infty)$  is *irreducible* in the sense that, for all  $n \neq m \in \Gamma$ , there exists a finite sequence  $n_1, n_2, \dots, n_k$  such that  $n_1 = n$ ,  $n_k = m$  and  $c_{n_i n_{i+1}} > 0$  for  $i = 1, \dots, k-1$ . Define the mapping  $\mathcal{J} : [0, \infty)^{\Gamma_\Delta} \rightarrow \mathbb{R}$  as

$$\mathcal{J}(c) = \sup_{\mathbf{R} \in (0, \infty)^\Gamma} \hat{\mathcal{J}}(c, \mathbf{R}), \quad \hat{\mathcal{J}}(c, \mathbf{R}) = \sum_{(n, m) \in \Gamma_\Delta} c_{nm} (1 - R_m / R_n) \quad (\text{A.1})$$

The following lemma is proven in [12, 23] and follows in part from material found in [31].

*Lemma 1:* The function  $\mathcal{J}$  is convex and continuous and takes its values in  $[0, \infty)$ . Moreover, for each irreducible  $c$ , the supremum on  $[0, \infty)^\Gamma$  of the function  $\hat{\mathcal{J}}(c, \cdot)$  is a maximum and is the unique solution of the set of equations

$$\sum_{m \in \Gamma} c_{nm} \frac{R_m}{R_n} = \sum_{m \in \Gamma} c_{mn} \frac{R_n}{R_m} \quad n \in \Gamma \quad (\text{A.2})$$

under the normalization  $\sum_{n \in \Gamma} R_n = 1$ .

### Appendix A.1. Evaluating the supremum

We proceed by introducing the following Ansatz regarding the solution  $\mathbf{R} = (R_n, n \in \Gamma)$  of the variational problem (4.6) for fixed  $x$  and strictly positive measure  $\psi$ , namely, that it is an eigenvector of the following matrix equation:

$$\mathbf{A}(x)\mathbf{R} + \mathbf{q} \circ \mathbf{R} = \lambda \mathbf{R} \quad (\text{A.3})$$

for some bounded vector  $\mathbf{q} = (q_n, n \in \Gamma)$ . We have used the notation introduced in (4.9).

Note that we are free to shift the vector  $\mathbf{q}$  by a constant since, under the transformation  $q_n \rightarrow q_n - c$ , the eigenvalue shifts by  $\lambda \rightarrow \lambda - c$  and the eigenvector is unchanged. That is, for fixed  $x$ ,

$$\lambda(x, \mathbf{q} - c\mathbf{1}_K) = \lambda(x, \mathbf{q}) - c. \quad (\text{A.4})$$

For simplicity, we choose  $c = q_K$  and take  $\mathbf{Q} = (q_1, q_2, \dots, q_{K-1}, 0)$  so that  $\lambda(x, \mathbf{Q}) = \lambda(x, \mathbf{q}) - q_K$  and  $\mathbf{R} = \mathbf{R}(x, \mathbf{Q})$  are solutions of the matrix equation

$$\mathbf{A}(x)\mathbf{R} + \mathbf{Q} \circ \mathbf{R} = \lambda \mathbf{R} \quad (\text{A.5})$$

There are  $K - 1$  independent variables,  $q_n$ ,  $n = 1, \dots, K - 1$ . One of the crucial features of the above Ansatz is that we can then ensure  $R_n = R_n(x, \mathbf{Q})$  is strictly positive by using the Perron-Frobenius theorem and taking  $\lambda = \lambda(x, \mathbf{Q})$  to be the Perron eigenvalue. Indeed, choosing  $\kappa$  such that

$$\kappa > \max_{n=1, \dots, K-1} \{|q_n|\}, \quad (\text{A.6})$$

it is clear that the new matrix  $\mathbf{A} + \text{diag}(\mathbf{Q}) + \kappa \mathbf{I}_K$  is irreducible and positive. According to the Perron-Frobenius theorem, it has a unique strictly positive eigenvector  $\mathbf{z}$  with the usual normalization  $\sum_m R_m = 1$ , and this eigenvector is also an eigenvector of  $\mathbf{A} + \text{diag}(\mathbf{Q})$  (with a shifted eigenvalue, though).

Applying equation (A.5) to the double sum in equation (4.6), which holds for strictly positive  $\psi$ , we find

$$\begin{aligned} \sum_{m,n=1}^K \psi_n W_{nm}(x) \left[ 1 - \frac{R_m}{R_n} \right] &= \sum_{n=1}^K \psi_n \left[ \sum_{m=1}^K W_{nm}(x) - \frac{1}{R_n} \sum_m (A_{nm}(x) + \delta_{nm}) R_m \right] \\ &= \sum_{n=1}^K \psi_n [-\lambda + Q_n] \\ &= \sum_{n=1}^{K-1} q_n \psi_n - \lambda(x, \mathbf{Q}), \end{aligned}$$

because of (2.5), (A.5), and since  $\sum_{m=1}^K \psi_m = 1$  and  $Q_K = 0$ . In order to ensure that we have found the true supremum with respect to  $\mathbf{z}$ , we use equation (A.5) to show that the supremum satisfies equation (A.2). Its left-hand side is

$$\sum_{m=1}^K \psi_n W_{nm}(x) \frac{R_m}{R_n} = \frac{\psi_n}{R_n} \sum_{m=1}^K W_{nm}(x) R_m = \frac{\psi_n}{R_n} (\lambda R_n - Q_n R_n) + R_n = \psi_n (\lambda - Q_n + 1).$$

The right-hand side reads

$$R_n \sum_{m=1}^K \frac{\psi_m}{R_m} W_{mn}(x) = R_n \sum_{m=1}^K \frac{\psi_m}{R_m} (A_{mn}(x) + \delta_{mn})$$

Dividing the left and right-hand sides by  $R_n$ , we deduce that

$$\psi_n = R_n(x, \mathbf{Q}) z_n(x, \mathbf{Q}) \quad (\text{A.7})$$

for each  $n = 1, \dots, K$ , where  $\mathbf{z} = (z_n, n \in \Gamma)$  is the corresponding unique strictly positive eigenvector of the adjoint linear equation

$$\mathbf{A}(x)^\top \mathbf{z} + \mathbf{Q} \circ \mathbf{z} = \lambda \mathbf{z}. \quad (\text{A.8})$$

such that  $\langle \mathbf{z}, \mathbf{R} \rangle = \mathbf{z}^\top \mathbf{R} = 1$ . Equation (A.7) ensures that  $\psi$  is a strictly positive measure and that  $\sum_{m=1}^K \psi_m = 1$ . Assuming that such a solution exists, the corresponding function  $j(x, \psi)$  is given by

$$j(x, \psi) = \sum_{n=1}^{K-1} q_n \psi_n - \lambda(x, \mathbf{Q}), \quad (\text{A.9})$$

where we have dropped the constant term.

In the given variational problem there are  $K - 1$  independent variables  $\psi_n$ ,  $n = 1, \dots, K - 1$  with  $\psi_K = 1 - \sum_{n=1}^{K-1} \psi_n$ . Similarly, there are  $K - 1$  independent variables  $q_n$ ,  $n = 1, \dots, K - 1$ . Therefore, equation (A.7) determines a mapping between the sets  $\{q_n, n = 1, \dots, K - 1\}$  and  $\{\psi_n, n = 1, \dots, K - 1\}$ . It remains to show that there exists a unique solution  $\mathbf{q}$  for each  $\psi \in \mathcal{M}_+([0, T])^\Gamma$ , that is, the mapping is invertible. Differentiating equation (A.8) with respect to  $q_m$ ,  $m = 1, \dots, K - 1$ , yields the inhomogeneous linear equation

$$\mathbf{L}(x, \mathbf{Q}) \frac{\partial \mathbf{z}}{\partial q_m} \equiv [\mathbf{A}^\top(x) + \text{diag}(\mathbf{Q}) - \lambda \mathbf{I}_K] \frac{\partial \mathbf{z}}{\partial q_m} = \frac{\partial \lambda}{\partial q_m} \mathbf{z} - z_m \mathbf{e}_m, \quad (\text{A.10})$$

where  $(\mathbf{e}_m)_n = \delta_{mn}$ . Multiplying both sides of (A.10) on the left with  $\mathbf{R}^\top$  and using (A.5) we obtain

$$\frac{\partial \lambda(x, \mathbf{Q})}{\partial q_m} = R_m(x, \mathbf{Q}) z_m(x, \mathbf{Q}), \quad m = 1, \dots, K-1. \quad (\text{A.11})$$

Since  $\mathbf{R}$  and  $\mathbf{z}$  are strictly positive,  $\lambda(x, \mathbf{Q})$  is a monotonically increasing function of the  $q_m$ . Moreover, equations (A.3) and (A.8) imply that, in the limit  $q_l \rightarrow \infty$  with all other  $q_m$  fixed,  $R_l, z_l \rightarrow 1$  and  $\partial \lambda / \partial q_l \rightarrow 1$ . On the other hand, if  $q_l \rightarrow -\infty$  then  $R_l, z_l \rightarrow 0$  and the Perron eigenvalue becomes independent of  $q_l$  with  $\partial \lambda / \partial q_l \rightarrow 0$ . Hence, by continuity, for each  $\psi \in \mathcal{M}_+([0, T])^\Gamma$  there exists a vector  $\mathbf{Q}$  such that  $\psi_n = R_n(x, \mathbf{Q}) z_n(x, \mathbf{Q})$  for all  $n = 1, \dots, K-1$ . For such a solution to be unique, the inverse function theorem implies that the Jacobian of the transformation must be invertible. Differentiating equation (A.11) with respect to  $q_n$  shows that the Jacobian is equivalent to the Hessian of  $\lambda$  with respect  $\mathbf{Q}$ , since

$$D_{mn}(x, \mathbf{Q}) \equiv \frac{\partial R_m(x, \mathbf{Q}) z_m(x, \mathbf{Q})}{\partial q_n} = \frac{\partial^2 \lambda(x, \mathbf{Q})}{\partial q_m \partial q_n} \quad (\text{A.12})$$

for all  $m, n = 1, \dots, K-1$ . This also establishes that the Jacobian is a symmetric matrix with real eigenvalues. Invertibility follows from the convexity of the function  $\mathcal{J}(c)$  defined by equation (A.1). That is, differentiating

$$j(x, \psi) = \mathcal{J}(c), \quad c_{nm} = \psi_n W_{nm}(x),$$

with respect to  $\psi$  for fixed  $x$  gives

$$W_{nm}(x) W_{mm'}(x) \frac{\partial^2 \mathcal{J}(c)}{\partial c_{nm} \partial c_{mm'}} = \frac{\partial^2 j(x, \psi)}{\partial \psi_n \partial \psi_m} \quad (\text{A.13})$$

with  $n \neq m'$  and  $m \neq m'$ . On the other hand, differentiating equation (A.9) for  $j(x, \psi)$  with respect to  $\psi_n$  gives

$$\frac{\partial j(x, \psi)}{\partial \psi_n} = q_n + \sum_{j=1}^{K-1} \left[ \psi_j - \frac{\partial \lambda}{\partial q_j} \right] \frac{\partial q_j}{\partial \psi_n} = q_n,$$

because of (A.7) and (A.11). And so

$$\frac{\partial^2 j(x, \psi)}{\partial \psi_n \partial \psi_m} = \frac{\partial q_n}{\partial \psi_m} = [\mathbf{D}^{-1}]_{nm}.$$

Hence, convexity of  $\mathcal{J}(c)$  together with irreducibility of the non-negative transition matrix  $\mathbf{W}$  means that the Jacobian is invertible and positive definite.

In summary, we have shown that for a strictly positive measure  $\psi$  and fixed  $x$ , a unique solution  $q_n = q_n(x, \psi)$  exists for all  $n = 1, \dots, K-1$  and we have solved the first variational problem by identifying  $\mathbf{z}$  with the unique (up to scalar multiplication), strictly positive eigenfunction of the matrix equation (A.5). Now suppose  $\psi$  is a non-negative rather than a strictly positive measure, that is,  $\psi_m = 0$  for at least one state  $m \in \Gamma$ . In this case  $c_{nm} = \psi_n W_{nm}(x)$ ,  $n \neq m$ , is not irreducible and lemma 1 no longer applies. However, as proven by Faggionato et al. [12], the function  $\mathcal{J}(c)$  is continuous with respect to  $c$ . Hence, assuming that the form of the rate function (4.4) still holds (which isn't necessarily true), we can take a sequence of strictly positive measures  $\psi^{(l)}$  on  $\Gamma$  such that  $\psi_n^{(l)} \rightarrow \psi_n$  for each  $n \in \Gamma$ . This implies that (for fixed  $x$ )

$$c[\psi^{(l)}] \rightarrow c[\psi]$$

and

$$j(x, \psi^{(l)}) = \mathcal{J}(c[\psi^{(l)}]) \rightarrow \mathcal{J}(c[\psi]) = j(x, \psi).$$

so that one can extend equation (A.9) to non-negative measures by taking

$$q_n(x, \psi) = \lim_{l \rightarrow \infty} q_n(x, \psi^{(l)}).$$

*Example.* We will illustrate the Perron eigenvalue solution to the supremum variational problem by considering an example for  $K = 2$ . Let us take the transition matrix to be

$$\mathbf{W} = \begin{pmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{pmatrix}$$

Consider the eigenvalue equation

$$\mathbf{W}\mathbf{R} + \mathbf{q} \circ \mathbf{R} = \lambda \mathbf{R},$$

where we have absorbed the diagonal terms  $\sum_{k=1,2} W_{km}$  into the definition of  $q_m$ . The resulting characteristic equation is a quadratic in  $\lambda$  and the leading or Perron eigenvalue is given by

$$\lambda = \frac{q_1 + q_2}{2} + \frac{7}{12} + \frac{1}{2} \sqrt{(q_1 - q_2)^2 - (q_1 - q_2)/3 + 25/36}.$$

It follows that

$$\begin{aligned} \psi_1 &\equiv \frac{\partial \lambda}{\partial q_1} = \frac{1}{2} + f(q_1 - q_2) \\ \psi_2 &\equiv \frac{\partial \lambda}{\partial q_2} = \frac{1}{2} - f(q_1 - q_2) \end{aligned}$$

with

$$f(q) = \frac{1}{4} \frac{2q - 1/3}{\sqrt{q^2 - q/3 + 25/36}}.$$

Note that  $\psi_1 + \psi_2 = 1$  as required. The function  $f(q)$  is a monotonically increasing function of  $q$  with  $f(-\infty) = -1/2$  and  $f(\infty) = 1/2$ . Thus, one can find a unique, finite value of  $q = q_1 - q_2$  for all  $\psi_1 \in (0, 1)$ , that is, for all strictly positive  $\psi$ . In the case of a non-negative  $\psi$  with  $\psi_1 = 0$  or  $\psi_2 = 0$ , we have  $q \rightarrow \pm\infty$ .

#### Appendix A.2. Evaluating the infimum

The next step is to substitute for  $j(x, \psi)$  in the rate function (4.5), which gives

$$J_T(\{x(t)\}_{t \in [0, T]}) = \inf_{\psi: \dot{x} = \sum_{n=1}^K \psi_n \mathbf{F}_n(x)} \int_0^T \left[ \sum_{n=1}^{K-1} q_n(t) \psi_n(t) - \lambda(x(t), \mathbf{Q}(t)) \right] dt, \quad (\text{A.14})$$

with  $q_n(t) = q_n(x(t), \psi(t))$  and  $\sum_{m=1}^K \psi_m = 1$ . In order to solve this variational problem, we introduce a  $d$ -dimensional Lagrange multiplier  $\boldsymbol{\mu}(t)$  and set

$$J_T(\{x(t)\}_{t \in [0, T]}) = \inf_{\psi, \boldsymbol{\mu}} S[x, \boldsymbol{\mu}, \psi], \quad (\text{A.15})$$

where

$$\begin{aligned} S[x, \boldsymbol{\mu}, \Psi] = \int_0^T & \left[ \sum_{n=1}^{K-1} q_n(t) \psi_n(t) - \lambda(x(t), \mathbf{Q}(t)) \right. \\ & \left. + \langle \boldsymbol{\mu}(t), \dot{x} - \sum_{n=1}^{K-1} [\mathbf{F}_n(x) - \mathbf{F}_K(x)] \psi_n(t) - \mathbf{F}_K(x) \rangle \right] dt, \end{aligned} \quad (\text{A.16})$$

and we have imposed the constraint  $\sum_{m=1}^K \psi_m = 1$ . The variational problem can now be expressed in terms of functional derivatives of  $S$ :

$$\frac{\delta S}{\delta \boldsymbol{\mu}(s)} = \dot{x}(s) - \sum_{n=1}^K \mathbf{F}_n(x(s)) \psi_n(s) = 0, \quad (\text{A.17})$$

and

$$\frac{\delta S}{\delta \psi_m(s)} = \sum_n \frac{\partial q_n}{\partial \psi_m} \psi_n(s) + q_m(s) - \sum_{n=1}^{K-1} \frac{\partial \lambda}{\partial q_n} \frac{\partial q_n}{\partial \psi_m} - \langle \boldsymbol{\mu}(s), \mathbf{F}_m(x(s)) - \mathbf{F}_K(x(s)) \rangle = 0 \quad (\text{A.18})$$

for  $m = 1, \dots, K-1$ . Combining with equations (A.7) and (A.11), we obtain the following solution to the variational problem in terms of  $\boldsymbol{\mu}$ :

$$\begin{aligned} q_m &= \langle \boldsymbol{\mu}, \mathbf{F}_m(x) - \mathbf{F}_K(x) \rangle, \\ \psi_m(x, \boldsymbol{\mu}) &= z_m(x, \boldsymbol{\mu}) R_m(x, \boldsymbol{\mu}), \end{aligned} \quad (\text{A.19})$$

for all  $m = 1, \dots, K-1$ , with  $R_n(x, \boldsymbol{\mu}), z_n(x, \boldsymbol{\mu})$  the positive eigenvectors of the matrix equations

$$\mathbf{A}\mathbf{R} + (\mathbf{F}(x)\boldsymbol{\mu}) \circ \mathbf{R} = \lambda \mathbf{R} \quad (\text{A.20})$$

and its adjoint, respectively. The  $K$ -dimensional vector  $\mathbf{F}(x)\boldsymbol{\mu}$  is the product of the  $K \times d$  matrix  $\mathbf{F}(x)$  whose  $K$  rows are the  $d$ -dimensional vectors  $\mathbf{F}_m(x)$ ,  $m = 1, \dots, K$ , with the  $d$ -dimensional vector  $\boldsymbol{\mu}$ . Here

$$\lambda = \lambda(x, \boldsymbol{\mu}) \equiv \lambda(x, \mathbf{Q}) q_m = \langle \boldsymbol{\mu}, \mathbf{F}_m - \mathbf{F}_K \rangle, m = 1, \dots, K-1).$$

The corresponding function  $j(x, \psi)$  becomes

$$j(x, \psi) = \sum_{n=1}^K \psi_n(x, \boldsymbol{\mu}) \langle \boldsymbol{\mu}, \mathbf{F}_n(x) \rangle - \lambda(x, \boldsymbol{\mu}).$$

The final step is to show that for each  $x$ , the equation

$$\dot{x} = \sum_{n=1}^K \mathbf{F}_n(x) \psi_n(x, \boldsymbol{\mu})$$

is invertible so that the function  $\boldsymbol{\mu} = \boldsymbol{\mu}(x, \dot{x})$  exists. From the inverse function theorem we require that

$$\sum_{n=1}^K \frac{\partial \psi_n(x, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \mathbf{F}_n(x) \neq 0, \quad (\text{A.21})$$

for all  $x \in \Omega$ . Note that the left hand side is a scalar. Following along identical lines to the analysis of equation (A.10), we differentiate the linear equation (A.20) with respect to  $\boldsymbol{\mu}$  to give

$$\mathbf{L}(x, \boldsymbol{\mu}) \frac{\partial \mathbf{R}}{\partial \boldsymbol{\mu}} = \mathbf{R} \frac{\partial \lambda}{\partial \boldsymbol{\mu}} - [\mathbf{G}_1 \circ \mathbf{R} \cdots \mathbf{G}_d \circ \mathbf{R}], \quad (\text{A.22})$$

with  $\mathbf{G}_k$ ,  $k = 1, \dots, d$  the  $K$ -dimensional vector of the  $k$ th coordinates of the  $K$  vectors  $\mathbf{F}_m(x)$ ,  $m = 1, \dots, K$ , and  $\mathbf{L}(x, \boldsymbol{\mu}) = \mathbf{L}(x, \mathbf{p})|_{p_m = \langle \boldsymbol{\mu}, \mathbf{F}_m \rangle}$ . Using the same arguments as previously we obtain

$$\frac{\partial \lambda(x, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^\top = \sum_{n=1}^K \mathbf{F}_n(x) z_n(x, \boldsymbol{\mu}) R_n(x, \boldsymbol{\mu}) \quad (\text{A.23})$$

Combining (A.12), (A.19), (A.23) and (A.21) we require

$$\frac{\partial^2 \lambda(x, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}^2} \equiv \sum_{m,n=1}^{K-1} \frac{\partial^2 \lambda(x, \mathbf{Q})}{\partial q_m \partial q_n} \Big|_{q_m = \langle \boldsymbol{\mu}, \mathbf{F}_m(x) - \mathbf{F}_K(x) \rangle} \langle \mathbf{F}_m(x) - \mathbf{F}_K(x), \mathbf{F}_n(x) - \mathbf{F}_K(x) \rangle \neq 0.$$

If we define  $X$  to be the  $(K-1) \times d$  matrix whose rows are the  $d$ -dimensional vectors  $\mathbf{F}_m(x) - \mathbf{F}_K(x)$ ,  $m = 1, \dots, K-1$  we have

$$\frac{\partial^2 \lambda(x, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}^2} = \text{Trace}(\mathbf{D}X X^\top) = \text{Trace}(X^\top \mathbf{D}X) = \sum_{k=1}^d X_k^\top \mathbf{D}X_k$$

The second equality is true because of the properties of the Trace operator. The  $(K-1)$ -dimensional vectors  $X_k$ ,  $k = 1, \dots, d$  are the column vectors of the matrix  $X$ . All terms in the right hand side are positive since the Jacobian  $\mathbf{D}$  has been shown to be positive definite. At least one of the vectors  $X_k$  is non zero since  $\mathbf{F}_m(x) \neq \mathbf{F}_K(x)$  for at least one  $m \neq K$  and the corresponding term in the sum in the right hand side of the last equality is strictly positive.

Finally, from equation (A.6), we require  $\boldsymbol{\mu}$  to be bounded, that is, there exists a  $\kappa$  for which

$$\kappa > \max_{m=1, \dots, K} \{|\langle \boldsymbol{\mu}, \mathbf{F}_m(x) \rangle|\}$$

for all  $x \in \Omega$ .

### Appendix A.3. Extension to a multi-scale process

The above analysis can be extended to the LDP (6.4) for a multi-scale stochastic process. In particular, using equation (A.9) we have

$$J_T(\{x(t)\}_{t \in [0, T]}) = \inf_{\{\psi(t)\}_{t \in [0, T]}} S[x, \psi] \quad (\text{A.24})$$

where

$$S[x, \psi] = \int_0^T \left[ \sum_{n=1}^{K-1} q_n(t) \psi_n(t) - \lambda(x(t), \mathbf{Q}(t)) + \frac{(\dot{x} - \sum_{n=1}^K \psi_n F_n(x))^2}{2 \sum_{n=1}^{K-1} \psi_n \sigma_n^2} \right] dt \quad (\text{A.25})$$

with  $q_n(t) = q_n(x(t), \psi(t))$ ,  $\sum_{m=1}^K \psi_m = 1$  and  $\lambda$  the Perron eigenvalue of equation (A.3). Taking the infimum with respect to  $\psi_k$ ,  $k = 1, \dots, K-1$ , gives

$$0 = \frac{\delta S}{\delta \psi_k} = q_k - \frac{(\dot{x} - \sum_{n=1}^K \psi_n F_n(x)) [F_k(x) - F_K(x)]}{\sum_{n=1}^K \psi_n \sigma_n^2} - \frac{(\dot{x} - \sum_{n=1}^K \psi_n F_n(x))^2 [\sigma_k^2(x) - \sigma_K^2(x)]}{2 [\sum_{n=1}^K \psi_n \sigma_n^2]^2} + \sum_{n=1}^{K-1} \left( \psi_n - \frac{\partial \lambda}{\partial q_n} \right) \frac{\partial q_n}{\partial \psi_k} \quad (\text{A.26})$$

Introducing the new variables

$$\boldsymbol{\mu} = \sum_{n=1}^K \psi_n F_n(x), \quad \sigma^2 = \sum_{n=1}^K \psi_n \sigma_n^2, \quad p = \frac{\dot{x} - \boldsymbol{\mu}}{\sigma^2}, \quad (\text{A.27})$$

and noting that  $\partial \lambda / \partial q_m = \psi_m$ , we have

$$q_k = p [F_k(x) - F_K(x)] + \frac{p^2}{2} [\sigma_k^2(x) - \sigma_K^2(x)] \quad (\text{A.28})$$



and

$$S[x, \psi] = \int_0^T \left[ p \sum_{n=1}^K F_n \psi_n + \frac{p^2}{2} \sum_{n=1}^K \psi_n \sigma_n^2 - \lambda(x, p) + \frac{(\dot{x} - \mu)^2}{2\sigma^2} \right] dt \quad (\text{A.29})$$

$$= \int_0^T \left[ p\mu + \frac{p^2}{2} \sigma^2 - \lambda(x, p) + \frac{p}{2} (\dot{x} - \mu) \right] dt \quad (\text{A.30})$$

$$= \int_0^T [p\dot{x} - \lambda(x, p)] \quad (\text{A.31})$$

with  $\lambda$  the Perron eigenvalue for the linear equation (which is independent of  $\mu$ )

$$\sum_n A_{mn} R_n + (pV_m(x) + p^2\sigma_m^2/2)R_m = \lambda R_m. \quad (\text{A.32})$$

Finally  $\mu$  and  $\sigma^2$  are determined from the identities

$$\mu = \sum_j R_j(x, p) z_j(x, p) V_j(x), \quad \sigma^2 = \sum_j R_j(x, p) z_j(x, p) \sigma_j^2(x). \quad (\text{A.33})$$

## Appendix B. Relation with the work of Kifer

In this appendix, we briefly describe the version of an LDP for stochastic hybrid systems developed in some detail by Kifer [23]. We first note that Kifer considers a more general dynamical system than presented here, namely, a random evolution on  $\Omega \times M \times \Gamma$  with  $M$  a compact Riemannian manifold. We will consider the more restricted case  $M = \emptyset$ . Let  $(n_{x_0, n_0}^\varepsilon(t), x_{x_0, n_0}^\varepsilon(t))$  denote a particular solution of the stochastic hybrid system evolving according to equation (2.4) with initial conditions  $(x_0, n_0)$ . Kifer [23] proves that the Hamiltonian

$$H(x_0, x, \beta) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \left[ \exp \beta \int_0^t F_{n_{x_0, n_0}^\varepsilon(s)}(x) ds \right] \quad (\text{B.1})$$

exists uniformly in  $x_0$ ,  $x \in \bar{\Omega}$  and  $|\beta| \leq b$  for some positive constant  $b$ . Moreover,  $H$  is strictly convex in  $\beta$  and Lipschitz continuous in the other variables, and does not depend on  $n_0$ . The Hamiltonian is then used to define the Lagrangian  $L(x, x', \alpha)$  as the Fenchel-Legendre transform of  $H$ ,

$$L(x, x', \alpha) = \sup_{\beta} (\alpha\beta - H(x, x', \beta)),$$

which is non-negative, (strictly) convex, lower semicontinuous and  $H$  is its inverse Fenchel-Legendre transform. Note the quantities that map onto our Hamiltonian and Lagrangian are  $H(x, \beta) = H(x, x, \beta)$  and  $L(x, \alpha) = L(x, x, \alpha)$ .

Consider  $C_{0T}$  the space of continuous mappings  $[0, T] \rightarrow \Omega$ . For each  $\gamma \in C_{0T}$  note  $\gamma_t = \gamma(t)$ ,  $t \in [0, T]$ . For each absolutely continuous  $\gamma$  its velocity  $\dot{\gamma}$  exists for almost all  $t \in [0, T]$  and  $t \rightarrow \dot{\gamma}_t$  is measurable. Hence one can define the action

$$S_{0T}(\gamma) = \int_0^T L(\gamma_t, \dot{\gamma}_t) dt.$$

It can be shown that  $S_{0T}$  is lower semicontinuous with respect to the *sup* norm on  $C_{0T}$ , which implies that the set  $\Psi_{0T}^a(x) = \{\gamma \in C_{0T} : \gamma_0 = x, S_{0T}(\gamma) \leq a\}$  is a closed set. Introducing the time-rescaled solution  $z_{x_0, n_0}^\varepsilon(t) = x_{x_0, n_0}^\varepsilon(t/\varepsilon)$ , Kifer establishes the existence of an LDP for the family of random paths  $\{z_{x_0, n_0}^\varepsilon\}$ : For all

$a, \delta, \lambda > 0$  and every  $\gamma \in C_{0T}$  such that  $\gamma_0 = x_0$  there exists  $\varepsilon_0 = \varepsilon_0(x_0, \gamma, a, \delta, \lambda)$  such that for  $\varepsilon < \varepsilon_0$ , uniformly in  $n_0 \in \Gamma$ , we have the lower bound on the open sets

$$P(|z_{x_0, n_0}^\varepsilon - \gamma| < \delta) \geq \exp \left\{ -\frac{1}{\varepsilon} (S_{0T}(\gamma) + \lambda) \right\},$$

and the upper bound on the closed sets

$$P(|z_{x_0, n_0}^\varepsilon - \Psi_{0T}^a(x_0)| \geq \delta) \leq \exp \left\{ -\frac{1}{\varepsilon} (a - \lambda) \right\},$$

Note that the initial condition is taken to be sufficiently far from the boundary  $\partial\Omega$  by requiring  $\inf_{z \in \partial\Omega} |x_0 - z| \geq 2KT$ . By the contraction principle one can then obtain an asymptotic estimate of the probability of the rescaled slow process to exit from a given open set  $V \subset \Omega$ . That is, defining  $\tau_{x_0, n_0}^\varepsilon(V)$  to be  $\inf\{t \geq 0 : z_{x_0, n_0}^\varepsilon(t) \notin V\}$ , then

$$\lim_{\varepsilon \rightarrow 0} P(\tau_{x_0, n_0}^\varepsilon(V) < T) = \exp \left\{ -\frac{1}{\varepsilon} \inf\{S_{0t}(\gamma) : \gamma \in C_{0T}, t \in [0, T], \gamma_0 = x_0, \gamma_t \notin V\} \right\}$$

In order to connect with our analysis of the LDP obtained by Faggionato et al [11, 12], we note that the Hamiltonian given by equation (B.1) can be expressed in terms of a Perron eigenvalue. The first step is to define the transition matrix according to

$$P(n_{x_0, n_0}^\varepsilon(t + \Delta) = n \mid n_{x_0, n_0}^\varepsilon(t) = m, x_{x_0, n_0}^\varepsilon(t) = x) = W_{nm}(x)\Delta + o(\Delta)$$

The infinitesimal generator  $\mathcal{L}^{x_0}$  of the continuous time Markov chain defined by the matrix  $\mathbf{W}(x_0)$  is the operator acting on vectors  $z \in \mathbb{R}^{|\Gamma|}$  by the formula

$$(\mathcal{L}^{x_0} z)_k = \sum_{l \in \Gamma} W_{lk}(x_0)(z_l - z_k)$$

It can then be proven [22, 23] that  $H(x_0, x, \beta)$  is the principal or Perron eigenvalue of the matrix operator  $\mathcal{L}^{x_0} + \beta F(x) : \mathbb{R}^{|\Gamma|} \rightarrow \mathbb{R}^{|\Gamma|}$  acting on vectors  $z \in \mathbb{R}^{|\Gamma|}$  by

$$((\mathcal{L}^{x_0} + \beta F(x))z)_k = (\mathcal{L}^{x_0} z)_k + \beta F_k(x)z_k = \sum_{l \in \Gamma} W_{lk}(x_0)(z_l - z_k) + \beta F_k(x)z_k \quad (\text{B.2})$$

The corresponding Lagrangian  $L(x_0, x, \alpha)$  is then given in explicit form by

$$L(x_0, x, \alpha) = \inf_{\psi} \{I_{x_0}(\psi) : \sum_{k \in \Gamma} F_k(x)\psi_k = \alpha\}$$

where  $\psi$  is a probability measure on  $\Gamma$ , and  $I_{x_0}(\psi)$  is defined by

$$I_{x_0}(\psi) = - \inf_{z > 0} \sum_{k \in \Gamma} \psi_k \frac{(\mathcal{L}^{x_0} z)_k}{z_k} = \sup_{z > 0} \sum_{k, l \in \Gamma} \psi_k W_{lk}(x_0) \left(1 - \frac{z_l}{z_k}\right)$$

Note that  $I_x(\psi)$  is precisely the function  $j(x, \psi)$  of equation (4.6),  $L(x, \alpha) = L(x, x, \alpha)$  is equivalent to the definition of the Lagrangian appearing in equation (4.5), and the eigenvalue equation (B.2) for  $x_0 = x$  is equivalent to equation (4.12) with  $\mu = \beta$ .

- [1] P. C. Bressloff and J. M. Newby Metastability in a stochastic neural network modeled as a velocity jump Markov process. SIAM Appl. Dyn. Syst. **12** 1394-1435 (2013).
- [2] P. C. Bressloff and J. Newby. Stochastic models of intracellular transport. Rev. Mod. Phys. **85** 135-196, 2013.
- [3] P. C. Bressloff and J.M. Newby. Path-integrals and large-deviations in stochastic hybrid systems. Phys. Rev. E **89** 042701, 2014.

- [4] P. C. Bressloff and J. M. Newby Stochastic hybrid model of spontaneous dendritic NMDA spikes. *Phys. Biol. Phys. Biol.* **11**, 016006, 2014.
- [5] P. C. Bressloff. Path-integral methods for analyzing the effects of fluctuations in stochastic hybrid neural networks. *J. Math. Neurosci* **5** 33pp. (2015)
- [6] E. Buckwar and M. G. Riedler An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution. *J. Math. Biol.* **63**, 1051–1093, 2011.
- [7] C. C. Chow and J. A. White. Spontaneous action potentials due to channel fluctuations. *Biophys. J.* **71**, 3013–3021, 1996.
- [8] M.H.A. Davis. Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Society, Series B (Methodological)*, 46:353–388, 1984.
- [9] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Applications of Mathematics, vol. 38. Springer, Berlin, 1998.
- [10] M. I. Dykman, E. Mori, J. Ross and P. M. Hunt. Large fluctuations and optimal paths in chemical kinetics. *J. Chem. Phys. A* **100**, 5735–5750, 1994.
- [11] A. Faggionato, D. Gabrielli, and M. Ribezzi Crivellari. Non-equilibrium thermodynamics of piecewise deterministic Markov Processes. *J Stat Phys* 137: 259-304, 2009.
- [12] A. Faggionato, D. Gabrielli and M. Crivellari. Averaging and large deviation principles for fully-coupled piecewise deterministic markov processes and applications to molecular motors. *Markov Processes and Related Fields* **16** 497–548, 2010
- [13] J. Feng and T. G. Kurtz. *Large Deviations for Stochastic Processes*. American Mathematical Society (2006).
- [14] R. F. Fox and Y. N. Lu Emergent collective behavior in large numbers of globally coupled independent stochastic ion channels. *Phys. Rev. E* **49**, 3421–3431 (1994).
- [15] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*. Springer, New York (1998).
- [16] C. W. Gardiner. *Handbook of stochastic methods, 4th edition*. Springer, Berlin, 2009.
- [17] J. H. Goldwyn and E. Shea-Brown. The what and where of adding channel noise to the Hodgkin-Huxley equations. *PLoS Comp. Biol.* **7** (11), e1002247, 2011.
- [18] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes* 3rd ed. Oxford University Press, Oxford, 2001.
- [19] P. Hanggi, H. Grabert, P. Talkner and H. Thomas. Bistable systems: master equation versus Fokker–Planck modeling. *Phys. Rev. A* **29**, 371–378, 1984)
- [20] J. P. Keener and J. M. Newby. Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. *Phy. Rev. E* **84** 011918, 2011.
- [21] T. B. Kepler and T. C. Elston. *Biophys. J.* **81** 3116–3136 (2001).
- [22] Y. Kifer. Principal eigenvalues and equilibrium states corresponding to weakly coupled parabolic systems of PDEs. *J. DÖAnalyse Math.* **59**, 89–102, 1992.
- [23] Y. Kifer. Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging. *Memoirs of the AMS* **201** (944), 2009.
- [24] C. Knessl, B. J. Matkowsky, Z. Schuss and C.Tier. An asymptotic theory of large deviations for Markov jump processes. *SIAM J. Appl. Math.* **46**, 1006–1028, 1985.
- [25] J. M. Newby, Isolating intrinsic noise sources in a stochastic genetic switch. *Phys. Biol.* **9** 026002, 2012.
- [26] J. M. Newby and P. C. Bressloff: Quasi-steady state reduction of molecular-based models of directed intermittent search. *Bull Math Biol* **72**, 1840–1866, 2010.
- [27] J. M. Newby and P. C. Bressloff. Local synaptic signalling enhances the stochastic transport of motor-driven cargo in neurons. *Phys. Biol.* **7**, 036004, 2010.
- [28] J. M. Newby and J. P. Keener An asymptotic analysis of the spatially inhomogeneous velocity-jump process. *SIAM Multiscale Model. Simul.* **9** 735–765, 2011.
- [29] J. M. Newby, P. C. Bressloff and J. P. Keener Breakdown of fast-slow analysis in an excitable system with channel noise. *Phys. Rev. Lett.* **111** 128101, 2013.
- [30] K. Pakdaman, M. Thieullen and G. Wainrib. Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. Appl. Prob.* **42**, 761–794, 2010.
- [31] R.T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, N. J., 1970.
- [32] R. S. Maier, and D. L. Stein. Limiting exit location distribution in the stochastic exit problem, *SIAM J. Appl. Math.* **57**, 752–790, 1997.
- [33] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports* **478** 1–69, 2009.
- [34] G. Wainrib, M. Thieullen and K. Pakdaman. Reduction of stochastic conductance-based neuron models with time-scales separation. *J. Comput. Neurosci.* **32**, 327–46, 2012.