Path integrals and large deviations in stochastic hybrid systems

Paul C. Bressloff
Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, Utah 84112, USA

Jay M. Newby
Mathematical Biosciences Institute, Ohio State University, Columbus, Ohio 43210, USA

(Received 23 January 2014; revised manuscript received 4 March 2014; published 1 April 2014)

We construct a path-integral representation of solutions to a stochastic hybrid system, consisting of one or more continuous variables evolving according to a piecewise-deterministic dynamics. The differential equations for the continuous variables are coupled to a set of discrete variables that satisfy a continuous-time Markov process, which means that the differential equations are only valid between jumps in the discrete variables. Examples of stochastic hybrid systems arise in biophysical models of stochastic ion channels, motor-driven intracellular transport, gene networks, and stochastic neural networks. We use the path-integral representation to derive a large deviation action principle for a stochastic hybrid system. Minimizing the associated action functional with respect to the set of all trajectories emanating from a metastable state (assuming that such a minimization scheme exists) then determines the most probable paths of escape. Moreover, evaluating the action functional along a most probable path generates the so-called *quasipotential* used in the calculation of mean first passage times. We illustrate the theory by considering the optimal paths of escape from a metastable state in a bistable neural network.

DOI: 10.1103/PhysRevE.89.042701 PACS number(s): 87.19.lj, 87.19.lc, 02.50.Ey, 05.40.−a

I. INTRODUCTION

There are a growing number of problems in biological physics that involve the coupling between a piecewise-deterministic dynamical system and a continuous-time Markov process, which is modeled as a stochastic hybrid process. Often the system of interest exhibits bistability, and one is interested in an associated escape problem. One of the best known examples of a stochastic hybrid system is a conductance-based model of a neuron [1–7]. 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. However, escape problems in stochastic hybrid systems are also important in the study of genetic switches [8,9], motor-driven intracellular transport [10,11], and stochastic neural networks [12]. The last example represents the dynamics of synaptically coupled neuronal populations. The state of each local population is described in terms of two stochastic variables, a continuous synaptic variable and a discrete activity variable. The synaptic variables evolve according to piecewise-deterministic dynamics describing, at the population level, synapses driven by spiking activity. The dynamical equations for the synaptic currents are only valid between jumps in spiking activity, and the latter are described by a continuous-time Markov process whose transition rates depend on the synaptic variables. In the mean-field limit, one recovers standard rate-based neural network models [12].

A general issue regarding metastability in a stochastic dynamical system is how to determine the most probable (optimal) paths of escape. A mathematical approach to addressing this issue in the weak noise limit is *large deviation theory* [13–15]. In order to give a heuristic definition of the latter, consider some random dynamical system in $\mathbb{R}^n$ for which there exists a well defined probability density functional $P_\epsilon[x]$ over the different sample trajectories $\{x(t)\}_0^T$ in a given time interval $[0,T]$. Here ϵ is a small parameter that characterizes the noise level; it could correspond to Ω^{-1} where Ω is the system size, or represent a fast time scale of the dynamics. A large deviation principle for the random paths is that

$$P_\epsilon[x] \sim e^{-\mathcal{A}[x]/\epsilon}, \quad \epsilon \to 0,$$

where $\mathcal{A}[x]$ is known as an action functional. Solving the first passage time problem for escape from a fixed-point attractor of the underlying deterministic system involves finding the most probable paths of escape, which minimize the action functional with respect to the set of all trajectories emanating from the fixed point (under certain additional constraints). Evaluating the action functional along a most probable path from the fixed point to another point x generates a corresponding *quasipotential* $\Phi(x)$ (assuming that the action can be minimized). It follows that any errors in the form of the quasipotential can generate exponentially large errors in the mean first passage time (MFPT) to escape from a metastable state. This explains, for example, why approximating a jump Markov process by a Langevin equation using some form of diffusion approximation can generate large errors in the MFPT, as recently demonstrated for stochastic ion channels [3]

One method for deriving the correct quasipotential is to use a Wentzel-Kramers-Brillouin (WKB) approximation of the (quasi-)steady-state probability density. This method has been applied to master equations [16–21] and more recently to stochastic hybrid systems [3,6,7,9,11,12]. In the former case, one can interpret the WKB equation for the quasipotential in terms of a Hamilton-Jacobi equation, whose corresponding Hamiltonian H is related to the Lagrangian of large deviation theory according to the Legendre transformation $L(x,\dot{x}) = p\dot{x} - H(x,p)$. This can be established by constructing the Doi-Peliti path-integral representation of solutions to the

master equation [22–24], and using the latter to derive a large deviation principle. Thus large deviation theory provides a rigorous foundation for the application and interpretation of WKB methods.

The major issue we wish to address in this paper is how to derive a large deviation principle for stochastic hybrid systems, so that we can interpret the quasipotential obtained using a WKB approximation in terms of an underlying variational problem for optimal paths in the space of stochastic trajectories. In previous work on stochastic ion channels [6] and stochastic neural networks [12], we simply assumed that such a variational principle exists. Rather than using abstract probability theory along the lines of Ref. [13], we will proceed by constructing a path-integral representation of solutions to a stochastic hybrid system. We will then use this to derive a large deviation principle for the stochastic neural network model previously constructed in [12]. Applications to stochastic ion channel models will be developed elsewhere. We begin in Sec. II by reviewing the analogous theory of large deviations in chemical master equations. We then consider the issue of large deviations in stochastic hybrid systems (Sec. III). The path-integral construction of a stochastic hybrid system is presented in Sec. IV, and this is then used to analyze optimal paths in a stochastic neural network model in Sec. V.

II. LARGE DEVIATIONS IN CHEMICAL MASTER EQUATIONS

In order to motivate the issues addressed in this paper regarding stochastic hybrid systems, it is useful to review the corresponding issues within the more familiar context of chemical master equations. For the sake of illustration, consider the birth-death master equation

$$\frac{dP_n(t)}{dt} = \omega_+(n-1)P_{n-1}(t) + \omega_-(n+1)P_{n+1}(t) - [\omega_+(n) + \omega_-(n)]P_n(t) \tag{2.1}$$

for $n = 0, \ldots, N$ with boundary conditions $P_{-1}(t) = P_{N+1}(t) = 0$ and transition rates $\omega_\pm(n,t)$. Multiplying both sides of Eq. (2.1) by n/N and summing over n gives

$$\frac{d\langle n/N\rangle}{dt} = \langle \Omega_+(n/N)\rangle - \langle \Omega_-(n/N)\rangle, \tag{2.2}$$

where $\omega_\pm(n) = N\Omega_\pm(n/N)$, and the brackets $\langle \cdots \rangle$ denote a time-dependent ensemble average over realizations of the stochastic dynamics, that is, $\langle A(n/N)\rangle = \sum_n P_n(t)A(n/N)$ for any function of state $A(n/N)$. If the transition rates in (2.1) are nonlinear functions of n, then there is coupling between different order moments, resulting in a moment closure problem. That is, $\langle \Omega_\pm(n/N)\rangle \neq \Omega_\pm(\langle n\rangle/N)$ for finite N. However, in the thermodynamic limit $N \to \infty$, statistical correlations can be ignored so that one can take the mean-field limit $\langle \Omega_\pm(n/N)\rangle \to \Omega_\pm(\langle n/N\rangle)$. This then yields a deterministic equation for the fraction x of open ion channels:

$$\frac{dx}{dt} = \Omega_+(x) - \Omega_-(x). \tag{2.3}$$

A diffusion approximation of the birth-death master Eq. (2.1) for large but finite N can be obtained by carrying out a Kramers-Moyal or system-size expansion to second order in $1/N$ [25,26].This yields a Fokker-Planck (FP) equation describing the evolution of the probability density of a corresponding continuous stochastic process that is the solution to a stochastic differential equation. A rigorous analysis of the diffusion approximation has been carried out by Kurtz [27]. First, introduce the rescaled variable $x = n/N$ and set $N\Omega_\pm(x) = \omega_\pm(Nx)$. Equation (2.1) can then be rewritten in the form

$$\frac{dp(x,t)}{dt} = N\{\Omega_+(x-1/N)p(x-1/N,t) + \Omega_-(x+1/N)p(x+1/N,t) - [\Omega_+(x) + \Omega_-(x)]p(x,t)\}. \tag{2.4}$$

Treating x, $0 \leqslant x \leqslant 1$, as a continuous variable and Taylor expanding terms on the right-hand side to second order in N^{-1} leads to the FP equation

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[A(x)p(x,t)] + \frac{1}{2N}\frac{\partial^2}{\partial x^2}[B(x)p(x,t)], \tag{2.5}$$

with

$$A(x) = \Omega_+(x) - \Omega_-(x), \quad B(x) = \Omega_+(x) + \Omega_-(x). \tag{2.6}$$

The FP equation takes the form of a conservation equation

$$\frac{\partial p}{\partial t} = -\frac{\partial J}{\partial x}, \tag{2.7}$$

where $J(x,t)$ is the probability flux,

$$J(x,t) = -\frac{1}{2N}\frac{\partial}{\partial x}[B(x)p(x,t)] + A(x)p(x,t). \tag{2.8}$$

The FP equation is supplemented by the no-flux or reflecting boundary conditions at the ends $x = 0,1$ and a normalization condition,

$$J(0,t) = J(1,t) = 0, \quad \int_0^1 p(x,t)dx = 1. \tag{2.9}$$

The FP equation has a unique steady-state solution obtained by setting $J(x,t) = 0$ for all $0 \leqslant x \leqslant 1$. The resulting first-order ODE can be solved to give a steady–state probability density of the form

$$P_{\rm FP}(x) = \mathcal{N}\frac{e^{-N\Psi(x)}}{B(x)}, \tag{2.10}$$

with the so-called quasipotential

$$\Psi(x) \equiv -2\int^x \frac{A(x')}{B(x')}dx' = -2\int^x \frac{\Omega_+(x') - \Omega_-(x')}{\Omega_+(x') + \Omega_-(x')}dx'. \tag{2.11}$$

Here $\mathcal{N}$ is a normalization factor.

It can be shown that in the large-N limit, the steady-state density of the full master Eq. (2.1) is identical in form to (3.15) but with a different quasipotential. As highlighted in the Introduction, one method for constructing the quasipotential of a chemical master equation is to use a WKB approximation of the (quasi)stationary probability density [16–21]. Substituting the stationary solution $P_n = \phi^\epsilon(n/N)$ into (2.1) with

$\epsilon = 1/N \ll 1$ and $x = n/N$ treated as a continuous variable, we have

$$0 = \Omega_+(x - 1/N)\phi^\epsilon(x - 1/N) + \Omega_-(x + 1/N)\phi^\epsilon(x + 1/N) - [\Omega_+(x) + \Omega_-(x)]\phi^\epsilon(x). \quad (2.12)$$

A WKB solution takes the form

$$\phi^\epsilon(x) \sim K(x;\epsilon)e^{-\Phi(x)/\epsilon}, \quad (2.13)$$

with $K(x;\epsilon) \sim \sum_{m=0}^{\infty} \epsilon^m K_m(x)$. Substituting Eq. (2.13) into Eq. (2.12), Taylor expanding with respect to ϵ, and collecting the $O(1)$ terms gives

$$\Omega_+(x)(e^{\Phi'(x)} - 1) + \Omega_-(x)(e^{-\Phi'(x)} - 1) = 0, \quad (2.14)$$

where $\Phi' = d\Phi/dx$. Solving this quadratic equation in $e^{\Phi'}$ shows that there is a unique nontrivial solution

$$\Phi = \int^x \ln \frac{\Omega_-(y)}{\Omega_+(y)} dy. \quad (2.15)$$

The other solution is $\Phi(x) = \text{const}$, which is associated with solutions of the corresponding deterministic dynamics.

Equation (2.14) has the form of a stationary Hamilton-Jacobi equation $H(x,\Phi'(x)) = 0$ for Φ, with Hamiltonian

$$H(x,p) = \sum_{r=\pm} \Omega_r(x)[e^{rp} - 1]. \quad (2.16)$$

This suggests a corresponding classical mechanical interpretation, in which H determines the motion of a particle with position x and conjugate momentum p. A trajectory of the particle is given by the solution of Hamilton's equations

$$\dot{x} = \frac{\partial H}{\partial p} = \sum_{r=\pm 1} r\Omega_r(x)e^{rp}, \quad (2.17)$$

$$\dot{p} = -\frac{\partial H}{\partial x} = \sum_{r=\pm 1} \frac{\partial \Omega_r}{\partial x}(x)[1 - e^{rp}]. \quad (2.18)$$

Introducing the Lagrangian

$$L(x,\dot{x}) = p \cdot \dot{x} - H(x,p), \quad (2.19)$$

it follows that $\Phi(x)$ with $\Phi(\bar{x}) = 0$ corresponds to the classical action evaluated along the least-action trajectory from $\bar{x}$ to x:

$$\Phi(x) = \inf_{x(-\infty)=\bar{x},x(T)=x} \int_{-\infty}^{T} L(x,\dot{x})dt. \quad (2.20)$$

(The lower limit is taken to be at $t = -\infty$, since we are interested in the steady-state solution, and thus $\bar{x}$ corresponds to a stable fixed point of the deterministic system.) Since $p = \Phi'$ everywhere along this path, we have

$$\Phi(x) = \int_{\bar{x}}^{x} p(x')dx', \quad (2.21)$$

with the integral taken along the trajectory. It follows that the leading-order term in the WKB approximation is determined by finding zero-energy solutions $p = p(x)$ such that $H(x,p(x)) = 0$.

An important issue is how to physically interpret the least-action trajectory obtained from the WKB approximation in terms of the underlying stochastic process. It turns out that the least-action path is the most probable (optimal) path from $\bar{x}$ to x. This can be established rigorously using large deviation theory [13,15]. An alternative approach is to construct a Doi-Peliti path-integral representation of solutions to the master equation [22–24], which can then be used to derive a least-action principle [18,19]. One major application of the theory of large fluctuations is the study of rare events such as the escape from a metastable state in the weak-noise limit. For example, suppose that the deterministic Eq. (2.3) exhibits bistability; that is, there exists a pair of stable fixed points at $x = x_\pm$ separated by an unstable fixed point at $x = x_0$. The mean first passage time (MFPT) to escape from the fixed point at x_- can be calculated by imposing an absorbing boundary at x_0, expressing the resulting quasistationary solution in terms of the WKB approximation (including higher-order terms) and then matching the WKB solution with an appropriate inner solution in a neighborhood of the point $x = x_0$. This is necessary since the WKB solution does not satisfy the absorbing boundary condition at the point x_0. It turns out that the MFPT takes the general Arrhenius form [16,18,19,28,29]

$$\tau \sim \frac{\Gamma(x_0,x_-)}{\sqrt{|\Phi''(x_0)|\Phi''(x_-)}} e^{N[\Phi(x_0)-\Phi(x_-)]}, \quad (2.22)$$

where Γ is an appropriate prefactor and $\Phi(x)$ is the previously calculated quasipotential. Moreover, $\Phi(x_0) - \Phi(x_-)$ is the value of the action along the optimal path from x_- to x_0. A similar expression for τ is obtained using the diffusion approximation, except that $\Phi(x)$ is replaced by the quasipotential $\Psi(x)$ and there is a different prefactor. Since $\Phi(x) \neq \Psi(x)$ and N is large, it follows that the diffusion approximation generates exponentially large errors in the MFPT to escape from a metastable state.

The physical interpretation of the least-action trajectories becomes crucial when solving escape problems in higher dimensions, since a metastable state is now surrounded by a nontrivial boundary (rather than a single point) and one needs to determine the relative weighting of optimal paths crossing different points on the boundary; see Fig. 1. However, care must be taken in higher dimensions, since it is possible there is no solution to the variational problem, that is, a quasipotential

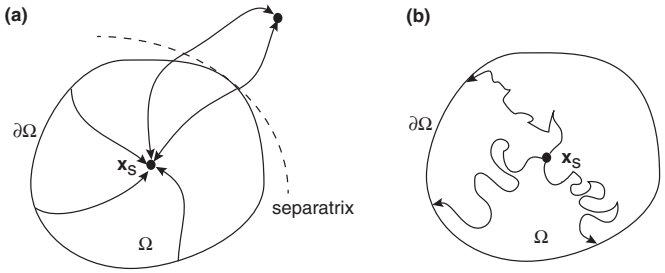

FIG. 1. (a) Deterministic trajectories of a multistable dynamical system. The subset Ω is contained within the basin of attraction of a fixed point $\mathbf{x}_s$. The boundary of the basin of attraction consists of separatrices, which are also solution trajectories. Paths that start in a neighborhood of a separatrix are attracted by different fixed points, depending on whether they begin on the left or right of the separatrix. (b) Random trajectories of the stochastic system. Escape from the domain Ω occurs when a random trajectory hits the boundary $\partial\Omega$.

may not exist. Nevertheless, the WKB method has been applied successfully to master equations in a wide range of applications in biophysics [19,28–30] and population biology [31–33].

III. ONE-DIMENSIONAL STOCHASTIC HYBRID SYSTEM

Let us now turn to a one-dimensional stochastic hybrid system evolving according to the equation

$$\frac{dx}{dt} = \frac{1}{\tau_x} v_n(x), \tag{3.1}$$

where $x \in \mathbb{R}$ is a continuous variable, $n = 0, \ldots, K-1$ is a discrete internal state variable, and the latter evolves according to a jump Markov process $n' \to n$ with transition rates $W_{nn'}(x)/\tau_n$. (For simplicity, we restrict ourselves to a single continuous variable. It is also possible to have a set of discrete variables, but one can always relabel the internal states so that they are effectively indexed by a single integer.) The jump propagator $W_{nn'}(x)dt/\tau_n$ is the probability that the system at x switches from the discrete internal state n' at time t to the state n at time $t + dt$. The resulting stochastic process is an example of a stochastic hybrid system based on a piecewise deterministic process. That is, the transition rates generally depend on x, with the latter coupled to the associated jump Markov process according to Eq. (3.1), which is only defined between jumps, during which $x(t)$ evolves deterministically. Denote the random state of the full model (3.1) at time t by $[X(t),N(t)]$, and introduce the corresponding probability density,

$$p_n(x,t)dx = \text{Prob}\{X(t) \in (x, x+dx), N(t) = n\}, \tag{3.2}$$

given an initial state $x(0) = x_0, n(0) = n_0$. The probability density evolves according to the differential Chapman-Kolmogorov (CK) equation (dropping the explicit dependence on initial conditions)

$$\frac{\partial p}{\partial t} + \frac{1}{\tau_x}\frac{\partial [v_n(x)p_n(x,t)]}{\partial x} = \frac{1}{\tau_n}\sum_{n'=0}^{K-1}[W_{nn'}(x)p_{n'}(x,t) - W_{n'n}(x)p_n(x,t)]. \tag{3.3}$$

We have introduced two time scales in the system, a relaxation time scale τ_x for the x dynamics and a transition time scale τ_n for the jump process.

In many of the listed biophysical applications, the kinetics associated with the jump process are much faster than the relaxation dynamics of x, that is, $\tau_n \ll \tau_x$. Let us fix the time units by setting $\tau_x = 1$ and introduce the small parameter $\epsilon = \tau_n/\tau_x$. We can then rewrite (5.6) in the more compact form

$$\frac{\partial p}{\partial t} = -\frac{\partial [v_n(x)p_n(x,t)]}{\partial x} + \frac{1}{\epsilon}\sum_{n'=0}^{K-1} A_{nn'}(x)p_{n'}(x,t), \tag{3.4}$$

with

$$A_{nn'}(x) = W_{nn'}(x) - \sum_{m=0}^{K-1} W_{mn}(x)\delta_{n',n}.$$

In the limit $\epsilon \to 0$, Eq. (3.1) reduces to the deterministic or mean-field equation

$$\frac{dx}{dt} = \mathcal{V}(x) \equiv \sum_{n=0}^{K-1} v_n(x)\rho_n(x), \tag{3.5}$$

where $\rho_n(x)$ is the unique steady-state distribution satisfying $\sum_{m\in I} A_{nm}(x)\rho_m(x) = 0$. We are assuming that for fixed x, the matrix $A_{n,m}(x)$ is irreducible (which means that there is a nonzero probability of transitioning, possibly in more than one step, from any state to any other state in the jump Markov process), and has a simple zero eigenvalue with corresponding left eigenvector $\mathbf{1}$ whose components are all unity, that is, $\sum_n A_{nm}(x) = 0$ for all m. (The latter follows immediately from the definition of $\mathbf{A}$; we are assuming that there are no other eigenvectors with a zero eigenvalue.) The Perron-Frobenius theorem [34] then ensures that all other eigenvalues are negative and the continuous-time Markov process for fixed x,

$$\frac{dp_n(x,t)}{dt} = \frac{1}{\epsilon}\sum_{m=0}^{K-1} A_{nm}(x)p_m(x,t),$$

has a globally attracting steady state $\rho_n(x)$ such that $p_n(x,t) \to \rho_n(x)$ as $t \to \infty$.

A. Quasi-steady-state diffusion approximation

Now suppose that the system operates in the regime $0 < \epsilon \ll 1$, for which there are typically a large number of transitions between different internal states n, while the variable x hardly changes at all. This suggests that the system rapidly converges to the (quasi) steady state $\rho_n(x)$, which will then be perturbed as x slowly evolves. The resulting perturbations can be analyzed using a quasi-steady-state (QSS) diffusion or adiabatic approximation, in which the CK Eq. (3.4) is approximated by a Fokker-Planck (FP) equation. The QSS approximation was first developed from a probabilistic perspective by Papanicolaou [35]; see also [26]. It has subsequently been applied to a wide range of problems in biology, including cell movement [36,37], wavelike behavior in models of slow axonal transport [38,39], molecular motor-based models of random intermittent search [40], and stochastic neural networks [12]. The QSS reduction proceeds in the following steps.

(1) Decompose the probability density as

$$p_n(x,t) = C(x,t)\rho_n(x) + \epsilon w_n(x,t), \tag{3.6}$$

where $\sum_n p_n(x,t) = C(x,t)$ and $\sum_n w_n(x,t) = 0$. Substituting into (3.4) yields

$$\frac{\partial C}{\partial t}\rho_n(x) + \epsilon\frac{\partial w_n(x,t)}{\partial t} = -v_n\frac{\partial [C(x,t)\rho_n(x) + \epsilon w_n(x,t)]}{\partial x} + \frac{1}{\epsilon}\sum_{n'=0}^{K-1} A_{nn'}(x)[C(x,t)\rho_{n'}(x) + \epsilon w_{n'}(x,t)].$$

Summing both sides with respect to n then gives

$$\frac{\partial C}{\partial t} = -\frac{\partial V(x)C(x,t)}{\partial x} - \epsilon \sum_n v_n \frac{\partial w_n(x,t)}{\partial x}, \tag{3.7}$$

where $V(x) = \sum_m v_m \rho_m(x)$.

(2) Using the equation for C and the fact that $A\rho = 0$, we have

$$\begin{aligned}\epsilon \frac{\partial w_n}{\partial t} &= \sum_{n'} A_{nn'}(x) w_{n'}(x,,t) \\ &\quad - v_n \frac{\partial \rho(x,n)C}{\partial x} + \rho(x,n)\frac{\partial V(x)C}{\partial x} \\ &\quad - \epsilon \sum_m [v_n \delta_{m,n} - \rho_n(x) v_m] \frac{\partial w_m(x,t)}{\partial x}.\end{aligned}$$

(3) Introduce the asymptotic expansion

$$w \sim w^{(0)} + \epsilon w^{(1)} + \epsilon^2 w^{(0)} + \cdots$$

and collect $O(1)$ terms:

$$\begin{aligned}\sum_{n'} A_{nn'}(x) w^{(0)}_{n'}(x,t) &= v_n \frac{\partial \rho_n(x) C(x,t)}{\partial x} \\ &\quad - \rho_n(x) \frac{\partial V(x) C(x,t)}{\partial x}.\end{aligned} \tag{3.8}$$

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

(4) Combining Eqs. (3.8) and (3.7) shows that C evolves according to the FP equation

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial x}(VC) + \epsilon \frac{\partial}{\partial x}\left(D \frac{\partial C}{\partial x} \right), \tag{3.9}$$

with the drift V and diffusion coefficient D given by

$$V(x) = \sum_{m=0}^{K-1} v_m(x)\rho_m(x), \quad D(x) = \sum_{n=0}^{K-1} Z_n(x) v_n(x), \tag{3.10}$$

and $Z(x,n)$ is the unique solution to

$$\sum_m A_{nm}(x) Z_m(x) = [V(x) - v_n(x)]\rho_n(x) \tag{3.11}$$

with $\sum_m Z_m(x) = 0$. Note that we have written the FP Eq. (3.9) in the Ito form—differences between Ito and Stratonovich involve $O(\epsilon)$ corrections to the drift term, which we ignore at the lowest level of approximation.

The FP Eq. (3.9) takes the form of a conservation equation

$$\frac{\partial C}{\partial t} = -\frac{\partial J}{\partial x}, \tag{3.12}$$

where $J(x,t)$ is the probability flux,

$$J(x,t) = -\epsilon D(x) \frac{\partial C(x,t)}{\partial x} + V(x) C(x,t). \tag{3.13}$$

Suppose that the FP equation is restricted to the domain $[x_a, x_b]$ with $x_a < x_- < x_+ < x_b$, and impose the no-flux or reflecting boundary conditions at the ends x_a, x_b:

$$J(x_a,t) = J(x_b,t) = 0. \tag{3.14}$$

The FP equation then has a unique steady-state solution obtained by setting $J(x,t) = 0$ for all $x_a \leqslant x \leqslant x_b$. The resulting first-order ODE can be solved to give a steady-state probability density of the form

$$C(x) = \mathcal{N} e^{-\Phi_{\rm FP}(x)/\epsilon}, \tag{3.15}$$

with the quasipotential

$$\Phi_{\rm FP}(x) = -\int^x \frac{V(y)}{D(y)} dy. \tag{3.16}$$

Here $\mathcal{N}$ is a normalization factor such that $\int_{x_a}^{x_b} C(x)dx = 1$.

B. WKB approximation

The FP equation captures the Gaussian-like fluctuations within the basin of attraction of a fixed point of the mean-field Eq. (3.5). On the other hand, as in the case of chemical master equations (see Sec. II), the diffusion approximation for small ϵ can break down when considering rare transitions between metastable states due to the fact that it generates a different quasipotential to the one obtained using a WKB approximation [3,6,11]. The latter takes the form

$$\phi^\epsilon_n(x) \sim R_n(x) \exp\left(-\frac{\Phi(x)}{\epsilon} \right), \tag{3.17}$$

where $\Phi(x)$ is the quasipotential. Substituting into the time-independent version of Eq. (3.4) yields

$$\begin{aligned}&\sum_{m=0}^{K-1} [A_{nm}(x) + \Phi'(x)\delta_{n,m} v_m(x)] R_m(x) \\ &\quad = \epsilon \frac{d v_n(x) R_n(x)}{dx} + \lambda_0 R_n(x),\end{aligned} \tag{3.18}$$

where $\Phi' = d\Phi/dx$. Introducing the asymptotic expansions $R \sim R^{(0)} + \epsilon R^{(1)}$ and $\Phi \sim \Phi_0 + \epsilon \Phi_1$, and using the fact that $\lambda_0 = O(e^{-L/\epsilon})$, the leading-order equation is

$$\sum_{m=0}^{K-1} A_{nm}(x) R^{(0)}_m(x) = -\Phi'_0(x) v_n(x) R^{(0)}_n(x). \tag{3.19}$$

[Note that since $v_n(x)$ is nonzero almost everywhere for $x \in \Omega$, we can identify $-\Phi'_0$ and $R^{(0)}$ as an eigenpair of the matrix operator $\widehat{A}_{nm}(x) = A_{nm}(x)/v_n(x)$ for fixed x.] Positivity of the probability density ϕ_ϵ requires positivity of the corresponding solution $R^{(0)}$. One positive solution is $R^{(0)} = \rho$, for which $\Phi'_0 = 0$. However, such a solution reflects the quasiequilibrium state around the fixed point x_- and is thus not the correct WKB solution. It can be proven that if $v_n(x)$ for fixed $x \in \Omega$ changes sign as n increases from zero, then there exists one other positive eigenfunction $R^{(0)}$, which can be identified as the appropriate WKB solution [11].

C. Higher-dimensional stochastic hybrid systems

Recently, the above QSS and WKB constructions have been extended to a number of higher-dimensional systems (more than one piecewise deterministic variable), including a stochastic conductance-based model [6] and a stochastic hybrid neural network [12]. In the former case, the continuous variables are membrane voltage and a recovery variable, whereas the single discrete variable represents the number of open sodium ion channels. On the other hand,

the stochastic neural network model consists of a set of $\alpha = 1, \ldots, M$ neuronal populations, and the state of each population is described in terms of two stochastic variables $X_\alpha(t)$ and $N_\alpha(t)$. The continuous variables $X_\alpha(t)$ represent population-averaged synaptic currents, which evolve according to piecewise-deterministic dynamics describing synapses driven by population spiking activity $N_\alpha(t)$. The discrete variables $N_\alpha(t)$ satisfy a continuous-time Markov process whose transition rates depend on the synaptic variables (see Sec. V). For both models, we used WKB to construct an effective Hamiltonian, and assumed that the solutions to the resulting Hamilton's equations generated optimal stochastic paths (paths of maximum likelihood) in the multidimensional phase space. We thus obtained nontrivial results concerning the geometric nature of the paths of escape from a metastable state. Although our conclusions based on WKB appeared consistent with direct Monte Carlo simulations of the full stochastic models, we did not have a variational principle to prove that these paths were indeed optimal paths. Therefore, we now derive such a variational principle for stochastic hybrid systems using a path-integral formulation. The latter differs considerably from the Doi-Peliti path integral for a master equation. We then apply the path-integral formulation to the particular example of a stochastic hybrid neural network in Sec. V—the advantage of this model is that the transition rates are linear in the discrete variables so that various quantities such as the quasipotential can be calculated explicitly. Elsewhere we will consider applications to more complicated models such as stochastic ion channels with nonlinear transition rates, where the construction of a Hamiltonian is less straightforward.

IV. PATH-INTEGRAL FORMULATION

For ease of notation, we will derive the path-integral formulation for a one-dimensional stochastic hybrid system evolving according to the CK equation (3.4); the generalization to higher dimensions is then straightforward, at least formally. We first discretize time by dividing a given interval $[0,T]$ into N equal subintervals of size Δt such that $T = N\Delta t$ and set $x_j = x(j\Delta t), n_j = n(j\Delta t)$. The conditional probability density for $x_1, \ldots, x_N$ given x_0 and a particular realization of the stochastic discrete variables $n_j, j = 0, \ldots, N-1$, is

$$P(x_1, \ldots, x_N | x_0, n_0, \ldots, n_{N-1}) = \prod_{j=0}^{N-1} \delta(x_{j+1} - x_j - v_{n_j}(x_j)\Delta t).$$

Inserting the Fourier representation of the Dirac δ function gives

$$P(x_1, \ldots, x_N | x_0, n_0, n_1, \ldots n_{N-1}) = \prod_{j=0}^{N-1} \left[\int_{-\infty}^{\infty} e^{-ip_j[x_{j+1} - x_j - v_{n_j}(x_j)\Delta t]} \frac{dp_j}{2\pi} \right] \equiv \prod_{j=0}^{N-1} \left[\int_{-\infty}^{\infty} H_{n_j}(x_{j+1}, x_j, p_j) \frac{dp_j}{2\pi} \right].$$

On averaging with respect to the intermediate states $n_j, j = 1, N-1$, we have

$$P(x_1, \ldots, x_N | x_0, n_0) = \left[\prod_{j=0}^{N-1} \int_{-\infty}^{\infty} \frac{dp_j}{2\pi} \right] \sum_{n_1, \ldots, n_{N-1}} \prod_{j=0}^{N-1} T_{n_{j+1}, n_j}(x_j) H_{n_j}(x_{j+1}, x_j, p_j),$$

where

$$T_{n_{j+1}, n_j}(x_j) \sim A_{n_{j+1}, n_j}(x_j) \frac{\Delta t}{\epsilon} + \delta_{n_{j+1}, n_j} \left(1 - \sum_m A_{m, n_j}(x_j) \frac{\Delta t}{\epsilon} \right) + o(\Delta t) = \left(\delta_{n_{j+1}, n_j} + A_{n_{j+1}, n_j}(x_j) \frac{\Delta t}{\epsilon} \right).$$

Consider the eigenvalue equation

$$\sum_m [A_{nm}(x) + q\delta_{n,m} v_m(x)] R_m^{(s)}(x,q) = \lambda_s(x,q) R_n^{(s)}(x,q), \tag{4.1}$$

and let $\xi_m^{(s)}$ be the adjoint eigenvector. Inserting multiple copies of the identity

$$\sum_s \xi_m^{(s)}(x,q) R_n^{(s)}(x,q) = \delta_{m,n},$$

for all x, q, with $q = q_j$ at the jth time step, we have

$$T_{n_{j+1} n_j}(x_j) H_{n_j}(x_{j+1}, x_j, p_j) \sim \sum_{s_j, m} R_{n_{j+1}}^{(s_j)}(x_j, q_j) \xi_m^{(s_j)}(x_j, q_j) \left(\delta_{n_j, m} + A_{m n_j}(x_j) \frac{\Delta t}{\epsilon} \right) H_{n_j}(x_{j+1}, x_j, p_j)$$

$$= \sum_{s_j} \left(1 + [\lambda_{s_j}(x_j, q_j) - q_j v_{n_j}(x_j)] \frac{\Delta t}{\epsilon} \right) R_{n_{j+1}}^{(s_j)}(x_j, q_j) \xi_{n_j}^{(s)}(x_j, q_j) H_{n_j}(x_{j+1}, x_j, p_j)$$

$$= \sum_{s_j} \left(1 + [\lambda_{s_j}(x_j, q_j) - q_j v_{n_j}(x_j)] \frac{\Delta t}{\epsilon} \right) e^{-ip_j[x_{j+1} - x_j - v_{n_j}(x_j)\Delta t]} R_{n_{j+1}}^{(s_j)}(x_j, q_j) \xi_{n_j}^{(s_j)}(x_j, q_j)$$

$$\sim \sum_{s_j} \exp\left(\left[\lambda_{s_j}(x_j,q_j) - q_j v_{n_j}(x_j)\right]\frac{\Delta t}{\epsilon} - i p_j[x_{j+1} - x_j - v_{n_j}(x_j)\Delta t]\right) R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_{n_j}(x_j,q_j)$$

$$= \sum_{s_j} \exp\left(\left[\lambda_{s_j}(x_j,q_j) - q_j\frac{x_{j+1}-x_j}{\Delta t}\right]\frac{\Delta t}{\epsilon}\right)\exp\left(\left[i\epsilon p_j v_{n_j}(x_j) - q_j v_{n_j}(x_j)\right]\frac{\Delta t}{\epsilon}\right) R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_{n_j}(x_j,q_j), \quad (4.2)$$

to leading order in $O(\Delta x,\Delta t)$. Let us now introduce the probability density

$$P(x_N,n_N|x_0,n_0) = \prod_{j=1}^{N-1}\int_{-\infty}^{\infty} dx_j\, P(x_1,\ldots,x_N,n_N|x_0,n_0). \quad (4.3)$$

Substituting for P using Eqs. (4.1) and (4.2) leads to

$$P(x_N,n_N|x_0,n_0) = \left[\prod_{j=1}^{N-1}\int_{-\infty}^{\infty} dx_j\right]\left[\prod_{j=0}^{N-1}\int_{-\infty}^{\infty}\frac{dp_j}{2\pi}\right]\sum_{n_1,\ldots,n_{N-1}}\sum_{s_0,\ldots,s_{N-1}}\left[\prod_{j=0}^{N-1} R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_{n_j}(x_j,q_j)\right]$$
$$\times \exp\left(\sum_j\left[\lambda_{s_j}(x_j,q_j) - q_j\frac{x_{j+1}-x_j}{\Delta t}\right]\frac{\Delta t}{\epsilon}\right)\exp\left([i\epsilon p_j v_{n_j}(x_j) - q_j v_{n_j}(x_j)]\frac{\Delta t}{\epsilon}\right). \quad (4.4)$$

By inserting the eigenfunction products and using the Fourier representation of the Dirac δ function, we have introduced sums over the discrete labels s_j and new phase variables p_j. However, this representation allows us to derive a large deviation principle in the limit $\epsilon \to 0$. First, note that the discretized path integral is independent of the q_j. Therefore, we are free to set $q_j = i\epsilon p_j$ for all j, thus eliminating the final exponential factor. This choice means that we can perform the summations with respect to the intermediate discrete states n_j using the orthogonality relation

$$\sum_n R^{(s)}_n(x_j,q_{j-1})\xi^{(s')}_n(x_{j+1},q_j) = \delta_{s,s'} + O(\Delta x,\Delta q).$$

We thus obtain the result that $s_j = s$ for all j, which means that we can then take the continuum limit of Eq. (5.31) to obtain the following path integral from $x(0) = x_0$ to $x(\tau) = x$ [after performing the change of variables $i\epsilon p_j \to p_j$ (complex contour deformation)]:

$$P(x,n,\tau|x_0,n_0,0) = \sum_s \int_{x(0)=x_0}^{x(\tau)=x}\exp\left(-\frac{1}{\epsilon}\int_0^\tau [p\dot{x} - \lambda_s(x,p)]dt\right)$$
$$\times R^{(s)}_n(x,p(\tau))\xi^{(s)}_{n_0}(x_0,p(0))\mathcal{D}[p]\mathcal{D}[x]. \quad (4.5)$$

Applying the Perron-Frobenius theorem [34] to the linear operator on the left-hand side of Eq. (4.1) shows that there exists a real, simple Perron eigenvalue labeled by $s = 0$, say, such that $\lambda_0 > \mathrm{Re}(\lambda_s)$ for all $s > 0$. It follows that, in the limit $\epsilon \to 0$, the largest contributions to the path integral (4.5) and the most likely paths in phase space (x,p) are obtained by restricting the sum over s to $s = 0$. Also note that the factor $R^{(0)}_n(x,p(\tau))\xi^{(0)}_{n_0}(x_0,p(0))$ in Eq. (4.5) essentially projects on to stochastic trajectories that start in the discrete state n_0 and terminate in the discrete state n. We will ignore any restrictions on these discrete states and simply consider the probability density [for fixed $x(0) = x_0$]

$$P(x,t) = \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} D[x]D[p]e^{-S[x,p]/\epsilon}, \quad (4.6)$$

with the action

$$S[x,p] = \int_0^\tau [p\dot{x} - \lambda_0(x,p)]\,dt. \quad (4.7)$$

We now have a classical variational problem, in which the Perron eigenvalue $\lambda_0(x,p)$ is identified as a Hamiltonian and the most probable path is the solution to Hamilton's equations

$$\dot{x} = \frac{\partial\mathcal{H}}{\partial p}, \quad \dot{p} = -\frac{\partial\mathcal{H}}{\partial x}, \quad \mathcal{H}(x,p) = \lambda_0(x,p). \quad (4.8)$$

One could also formulate the least action principle in terms of the corresponding Lagrangian, through the *Legendre transform* of $\lambda_0(x,p)$:

$$L[x,\dot{x}] \equiv p\dot{x} - \lambda_0(x,p), \quad (4.9)$$

with $p = p(x,\dot{x})$ given implicitly by the equation $\dot{x} = \partial\mathcal{H}/\partial p$. Hamilton's equations are equivalent to the Euler-Lagrange equations for L.

Now suppose that we have a higher-dimensional stochastic hybrid system with M continuous variables x_α, $\alpha = 1,\ldots M$, and a single discrete variable $n = 0,\ldots,K-1$. (It is straightforward to extend to the case of several discrete variables as occurs in the stochastic neural network model of Sec. V.) The multivariate CK equation takes the form

$$\frac{\partial p}{\partial t} = -\sum_{\alpha=1}^{M}\frac{\partial}{\partial x_\alpha}[v_\alpha(\mathbf{x},n)p(\mathbf{x},n,t)] + \frac{1}{\epsilon}\sum_m A(n,m;\mathbf{x})p(\mathbf{x},m,t). \quad (4.10)$$

The drift "velocities" $v_\alpha(\mathbf{x},n)$ for fixed n represent the piecewise-deterministic dynamics according to

$$\tau \frac{du_\alpha}{dt} = v_\alpha(\mathbf{x},n), \quad \alpha = 1,\ldots,M, \tag{4.11}$$

Following along identical lines to the one-dimensional case, we can derive a path-integral representation of the solution to Eq. (4.10):

$$\begin{aligned} &p(\mathbf{x},n,\tau|\mathbf{x}_0,n_0,0) \\ &= \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} \mathcal{D}[\mathbf{p}]\mathcal{D}[\mathbf{x}] \exp\left(-\frac{1}{\epsilon}S[\mathbf{x},\mathbf{p}]\right) \\ &\quad \times R^{(0)}(\mathbf{x},\mathbf{p}(\tau),n)\xi^{(0)}(\mathbf{x}_0,\mathbf{p}(0),n_0), \end{aligned} \tag{4.12}$$

with action

$$S[\mathbf{x},\mathbf{p}] = \int_0^\tau \left[\sum_{\alpha=1}^M p_\alpha \dot{x}_\alpha - \lambda_0(\mathbf{x},\mathbf{p})\right] dt. \tag{4.13}$$

Here λ_0 is the Perron eigenvalue of the following linear operator equation [cf. Eq. (4.1)]:

$$\begin{aligned} &\sum_m A(n,m;\mathbf{x})R^{(0)}(\mathbf{x},\mathbf{p},m) \\ &= [\lambda_0(\mathbf{x},\mathbf{p}) - \sum_{\alpha=1}^M p_\alpha v_\alpha(\mathbf{x},n)]R^{(0)}(\mathbf{x},\mathbf{p},n), \end{aligned} \tag{4.14}$$

and $\xi^{(0)}$ is the adjoint eigenvector. Suppose that the underlying deterministic system (4.11) has a unique stable fixed point $\mathbf{x}_*$. The quasipotential of the corresponding stationary density can then be obtained by finding zero-energy solutions of Hamilton's equations

$$\dot{\mathbf{x}} = \nabla_{\mathbf{p}}\mathcal{H}(\mathbf{x},\mathbf{p}), \quad \dot{\mathbf{p}} = -\nabla_{\mathbf{x}}\mathcal{H}(\mathbf{x},\mathbf{p}), \tag{4.15}$$

with $\mathbf{x} = (x,y), \mathbf{p} = (p_x,p_y)$. If such a solution can be found, then we can construct a quasipotential Φ by identifying it as the action along a zero-energy solution curve $\mathbf{x}(t)$. That is,

$$\frac{d\Phi}{dt} \equiv \sum_{\alpha=1}^M \frac{\partial\Phi}{\partial x_\alpha}\frac{dx_\alpha}{dt} = \sum_{\alpha=1}^M p_\alpha \frac{dx_\alpha}{dt}, \tag{4.16}$$

with $p_\alpha = \partial\Phi/\partial x_\alpha$.

V. APPLICATION TO A STOCHASTIC HYBRID NEURAL NETWORK

In order to illustrate the path-integral method, we will consider a stochastic hybrid neural network model that generalizes the so-called neural master equation [32,41–43] by incorporating synaptic dynamics. (A detailed derivation of the model can be found in [12].) The master equation formulation assumes that noise in neural networks arises intrinsically as a collective population effect, and describes the stochastic dynamics as a continuous-time Markov process. Neurons are partitioned into a set of M local homogeneous populations labeled $\alpha = 1,\ldots,M$, each consisting of K neurons. The state of each population at time t is specified by the number $N_\alpha(t)$ of active neurons in a sliding window $(t,t+\Delta t]$, and transition rates between the discrete states are chosen so that standard rate-based models are obtained in the mean-field limit, where statistical correlations can be ignored. There are two versions of the neural master equation, which can be distinguished by the size of the sliding window width Δt. (Note that the stochastic models are keeping track of *changes* in population activity.) One version assumes that each population operates close to an asynchronous state for large K [32,42], so that one-step changes in population activity occur relatively slowly. Hence one can set $\Delta t = 1$ and take K to be large but finite. The other version of the neural master equation assumes that population activity is approximately characterized by a Poisson process [41,43]. In order to maintain a one-step jump Markov process, it is necessary to take the limits $\Delta t \to 0, K \to \infty$ such that $K\Delta t = 1$. Thus one considers the number of active neurons in an infinite background sea of inactive neurons, which is reasonable if the networks are in low activity states.

One way to link the two versions of the neural master equation is to extend the Doi-Peliti path-integral representation of chemical master equations [22–24] to the neural case; the difference between the two versions then reduces to a different choice of scaling of the underlying action functional [42]. Buice *et al.* [41,43] used diagrammatic perturbations methods (Feynman graphs) to generate a truncated moment hierarchy based on factorial moments, and thus determined corrections to mean-field theory involving coupling to two-point and higher-order cumulants. They also used renormalization-group methods to derive scaling laws for statistical correlations close to criticality, that is, close to a bifurcation point of the underlying deterministic model [41]. On the other hand, Bressloff [32,42] showed how the path-integral representation of the master equation can be used to investigate large deviations or rare event statistics underlying escape from the basin of attraction of a metastable state, following along analogous lines to previous work on large deviations in chemical master equations [18,19,29].

One limitation of both versions of the neural master equation is that they neglect the dynamics of synaptic currents. The latter could be particularly significant if the time scale τ of synaptic dynamics is smaller than the window width Δt. Therefore, we recently extended the Buice *et al.* neural master equation by formulating the network population dynamics in terms of the following stochastic hybrid system [12]. Consider a set of M homogeneous populations labeled $\alpha = 1,\ldots,M$, with K neurons in each population. [A straightforward generalization would be for each population to consist of $O(K)$ neurons.] The output activity of each population is taken to be a discrete stochastic variable $A_\alpha(t)$ given by

$$A_\alpha(t) = \frac{N_\alpha(t)}{K\Delta t}, \tag{5.1}$$

where $N_\alpha(t)$ is the number of neurons in the αth population that fired in the time interval $[t-\Delta t,t]$, and Δt is the width of a sliding window that counts spikes. The discrete stochastic variables $N_\alpha(t)$ are taken to evolve according to a one-step jump Markov process:

$$N_\alpha(t) \overset{\omega_+/\tau_a}{\to} N_\alpha(t)+1, \quad N_\alpha(t) \overset{\omega_-/\tau_a}{\to} N_\alpha(t)-1, \tag{5.2}$$

with corresponding transition rates

$$\omega_+ = K\Delta t\, F(X_\alpha), \quad \omega_- = N_\alpha. \tag{5.3}$$

Here F is a sigmoid firing rate or gain function

$$F(x) = \frac{F_0}{1 + e^{-\gamma(x-\kappa)}}, \tag{5.4}$$

where γ,κ correspond to the gain and threshold, respectively, and $X_\alpha(t)$ is the effective synaptic current into the αth population, which evolves (for exponential synapses) according to

$$\tau\, dX_\alpha(t) = \left[-X_\alpha(t) + \frac{1}{K\Delta t} \sum_{\beta=1}^{M} w_{\alpha\beta} N_\beta(t) \right] dt. \tag{5.5}$$

In the following, we take the thermodynamic limit $K \to \infty$, $\Delta t \to 0$ such that $K\Delta t = 1$ is fixed. Our model then reduces to the Buice *et al.* [41,43] version of the neural master equation in the limit $\tau \to 0$; see below. The resulting stochastic process defined by Eqs. (5.1)–(5.5) is clearly an example of a stochastic hybrid system based on a piecewise deterministic process. That is, the transition rate ω_+ depend on X_α, with the latter itself coupled to the associated jump Markov according to Eq. (5.5), which is only defined between jumps, during which $X_\alpha(t)$ evolves deterministically. It is important to note that the time constant τ_a cannot be identified directly with membrane or synaptic time constants. Instead, it determines the relaxation rate of a local population to the instantaneous firing rate.

Introduce the probability density

$$\begin{aligned} &\text{Prob}\{X_\alpha(t) \in (x_\alpha, x_\alpha + dx), N_\alpha(t) = n_\alpha; \alpha = 1, \ldots, M\} \\ &\quad = p(\mathbf{x},\mathbf{n},t|\mathbf{x}_0,\mathbf{n}_0,0)d\mathbf{x}, \end{aligned}$$

with $\mathbf{x} = (x_1, \ldots, x_M)$ and $\mathbf{n} = (n_1, \ldots, n_M)$. It follows from Eqs. (5.1)–(5.5) that the probability density evolves according to the differential Chapman-Kolmogorov (CK) equation (dropping the explicit dependence on initial conditions)

$$\begin{aligned} &\frac{\partial p}{\partial t} + \frac{1}{\tau} \sum_\alpha \frac{\partial [v_\alpha(\mathbf{x},\mathbf{n}) p(\mathbf{x},\mathbf{n},t)]}{\partial x_\alpha} \\ &\quad = \frac{1}{\tau_a} \sum_\alpha \left(\mathbb{T}_\alpha^{+1} - 1 \right) [n_\alpha p(\mathbf{x},\mathbf{n},t)] \\ &\qquad + \frac{1}{\tau_a} \sum_\alpha \left(\mathbb{T}_\alpha^{-1} - 1 \right) [v(x_\alpha) p(\mathbf{x},\mathbf{n},t)], \end{aligned} \tag{5.6}$$

with

$$v_\alpha(\mathbf{x},\mathbf{n}) = -x_\alpha + \sum_\beta w_{\alpha\beta} n_\beta \tag{5.7}$$

and $\mathbb{T}_\alpha^\pm$ are translation operators: $\mathbb{T}_\alpha^{\pm 1} f(\mathbf{n}) = f(\mathbf{n}_{\alpha\pm})$ for any function f with $\mathbf{n}_{\alpha\pm}$ denoting the configuration with n_α replaced by $n_\alpha \pm 1$. Equation (5.6) can be re-expressed in the more compact form [cf. Eq. (4.10)]

$$\begin{aligned} \frac{\partial p}{\partial t} &= -\frac{1}{\tau} \sum_{\alpha=1}^{M} \frac{\partial}{\partial x_\alpha} [v_\alpha(\mathbf{x},\mathbf{n}) p(\mathbf{x},\mathbf{n},t)] \\ &\quad + \frac{1}{\tau_a} \sum_{\mathbf{m}} W(\mathbf{n},\mathbf{m};\mathbf{x}) p(\mathbf{x},\mathbf{m},t). \end{aligned} \tag{5.8}$$

The drift velocities $v_\alpha(\mathbf{x},\mathbf{n})$ for fixed $\mathbf{n}$ represent the piecewise-deterministic synaptic dynamics according to

$$\tau \frac{dx_\alpha}{dt} = v_\alpha(\mathbf{x},\mathbf{n}), \quad \alpha = 1, \ldots, M. \tag{5.9}$$

The Perron-Frobenius theorem [34] ensures that the continuous-time Markov process for fixed $\mathbf{x}$,

$$\frac{dp(\mathbf{x},\mathbf{n},t)}{dt} = \frac{1}{\tau_a} \sum_{\mathbf{m}} W(\mathbf{n},\mathbf{m};\mathbf{x}) p(\mathbf{x},\mathbf{m},t),$$

has a globally attracting steady state $\rho(\mathbf{x},\mathbf{n})$ such that $p(\mathbf{x},\mathbf{n},t) \to \rho(\mathbf{x},\mathbf{n})$ as $t \to \infty$. For the given model, the steady-state solution $\rho(\mathbf{x},\mathbf{n})$ of Eq. (5.6) can be factorized as $\rho(\mathbf{x},\mathbf{n}) = \prod_{\beta=1}^M \rho_0(x_\beta,n_\beta)$ with

$$0 = \sum_{\alpha=1}^M \left[\prod_{\beta\neq\alpha} \rho_0(x_\beta,n_\beta) \right] [J(x_\alpha,n_\alpha+1) - J(x_\alpha,n_\alpha)],$$

where

$$J(x,n) = n\rho_0(x,n) - F(x)\rho_0(x,n-1).$$

Since $\rho_0(x,-1) \equiv 0$, it follows that $J(x,n) = 0$ for all n and x. Hence

$$\rho_0(x,n) = \rho_0(x,0) \prod_{m=1}^n \frac{F(x)}{m} = \rho_0(x,0) \frac{F(x)^n}{n!}, \tag{5.10}$$

so that the corresponding normalized density is a Poisson process with rate $F(x)$:

$$\rho_0(x,n) = e^{-F(x)} \frac{F(x)^n}{n!}. \tag{5.11}$$

There are two time scales in the CK Eq. (5.8), the synaptic time constant τ and the time constant τ_a, which characterize the relaxation rate of population activity. In the limit $\tau \to 0$, Eq. (5.5) reduces to the neural master equation of Buice *et al.* [41,43]. First, note that the synaptic variables $X_\alpha(t)$ are eliminated by setting $v_\alpha = 0$, that is, $X_\alpha(t) = \sum_\beta w_{\alpha\beta} N_\beta(t)$. This then leads to a pure birth-death process for the discrete variables $N_\alpha(t)$. That is, let $P(\mathbf{n},t) = \text{Prob}[\mathbf{N}(t) = \mathbf{n}]$ denote the probability that the network of interacting populations has configuration $\mathbf{n} = (n_1,n_2,\ldots,n_M)$ at time $t, t > 0$, given some initial distribution $P(\mathbf{n},0)$. The probability distribution then evolves according to the birth-death master equation [41–43]

$$\begin{aligned} \frac{dP(\mathbf{n},t)}{dt} &= \sum_\alpha \left\{ (\mathbb{T}_\alpha - 1) \left[\Omega_\alpha^-(\mathbf{n}) P(\mathbf{n},t) \right] \right. \\ &\quad \left. + \left(\mathbb{T}_\alpha^{-1} - 1 \right) \left[\Omega_\alpha^+(\mathbf{n}) P(\mathbf{n},t) \right] \right\}, \end{aligned} \tag{5.12}$$

where

$$\Omega_\alpha^+(\mathbf{n}) = \frac{1}{\tau_a} F\left(\sum_\beta w_{\alpha\beta} n_\beta \right), \quad \Omega_\alpha^-(\mathbf{n}) = \frac{n_\alpha}{\tau_a}. \tag{5.13}$$

It can be shown that the network operates in a Poisson-like regime in which the rates of the Poisson process are stochastic variables whose means evolve according to the activity-based mean-field equation of Wilson and Cowan [44]

$$\tau_\alpha \frac{da_\alpha}{dt} = -a_\alpha(t) + F\left(\sum_\beta w_{\alpha\beta} a_\beta(t) \right). \tag{5.14}$$

On the other hand, if $\tau_a \to 0$ for fixed τ, then we obtain deterministic voltage or current-based mean-field equations

$$\tau \frac{dx_\alpha}{dt} = \sum_{\mathbf{n}} v_\alpha(\mathbf{x}(t),\mathbf{n})\rho(\mathbf{x}(t),\mathbf{n}) = -x_\alpha(t) + \sum_{\beta=1}^{M} w_{\alpha\beta} \sum_{\mathbf{n}} n_\beta \rho(\mathbf{x}(t),\mathbf{n}). \quad (5.15)$$

Since $\rho(\mathbf{x},\mathbf{n})$ is given by a product of independent Poisson processes with rates $F(x_\alpha)$, consistent with the operating regime of the Buice *et al.* master equation [41,43], it follows that

$$\langle n_\beta \rangle = F(x_\beta), \quad (5.16)$$

and Eq. (5.15) reduces to the standard voltage or current-based activity equation of Amari [45]

$$\tau \frac{dx_\alpha}{dt} = -x_\alpha(t) + \sum_{\beta=1}^{M} w_{\alpha\beta} F(x_\beta). \quad (5.17)$$

Note that the limit $\tau_a \to 0$ is analogous to the slow synapse approximation used by Ermentrout [46] to reduce deterministic conductance-based neuron models to voltage-based rate models. Here we are interested in the regime $0 < \tau_a/\tau \equiv \epsilon \ll 1$, for which there are typically a large number of transitions between different firing states $\mathbf{n}$, while the synaptic currents $\mathbf{x}$ hardly change at all. This suggests that the system rapidly converges to the (quasi) steady state $\rho(\mathbf{x},\mathbf{n})$, which will then be perturbed as $\mathbf{x}$ slowly evolves. The resulting perturbations can be analyzed using a quasi-steady-state (QSS) diffusion approximation along the lines of Sec. IIIA, in which the CK Eq. (5.8) is approximated by a Fokker-Planck equation [12]. However, as we have already highlighted, the diffusion approximation can break down when considering an escape problem in the weak noise (small ϵ) limit, It is then necessary to use some form of large deviation theory such as the path-integral formulation of Sec. IV. We will illustrate these issues by considering bistability in a one-population and then a multipopulation neural network.

A. Quasipotential in one-population model

In order to illustrate the limitations of the diffusion approximation, consider the simple case of a single recurrent population ($M = 1$) and set $x_1 = x, n_1 = n, p(x_1,n_1,t) = p_n(x,t)$. The CK equation for p_n takes the form

$$\frac{\partial p_n}{\partial t} + \frac{\partial [v_n(x)p_n(x,t)]}{\partial x} = \frac{1}{\epsilon} \sum_m A_{nm}(x) p_m(x,t), \quad (5.18)$$

with drift term

$$v_n(x) = -x + wn, \quad (5.19)$$

and tridiagonal transition matrix

$$A_{n,n-1}(x) = F(x), \quad A_{nn}(x) = -F(x) - n, \quad A_{n,n+1}(x) = n + 1. \quad (5.20)$$

Following the discussion at the end of Sec. II, we expect the finite-time behavior of the stochastic population for $\epsilon \ll 1$ to be characterized by small perturbations about the stable steady state $\rho_n(x)$ of the underlying jump Markov process, with x treated as a constant over time scales comparable to the relaxation time of the birth-death process. The steady-state density is given by the Poisson distribution (5.11) and the mean-field equation obtained in the $\epsilon \to 0$ limit is

$$\frac{dx}{dt} = \sum_{n=0}^{\infty} v_n(x)\rho_n(x) = -x + wF(x) \equiv -\frac{d\Psi}{dx}. \quad (5.21)$$

The sigmoid function $F(x)$ given by Eq. (5.4) is a bounded, monotonically increasing function of x with $F(x) \to F_0$ as $x \to \infty$ and $F(x) \to 0$ as $x \to -\infty$. Moreover, $F'(x) = \gamma F_0/\{4\cosh^2[\gamma(x-\kappa)/2]\}$ so that $F(x)$ has a maximum slope at $x = \kappa$ given by $\gamma F_0/4$. It follows that the function $-x + wF(x)$ only has one zero if $w\gamma F_0 < 4$ and this corresponds to a stable fixed point. On the other hand, if $w\gamma F_0 > 4$ then, for a range of values of the threshold κ, $[\kappa_1,\kappa_2]$, there exists a pair of stable fixed points $x_\pm$ separated by an unstable fixed point x_0 (bistability). A stable-unstable pair vanishes via a saddle-node bifurcation at $\kappa = \kappa_1$ and $\kappa = \kappa_2$. This can also be seen graphically by plotting the potential function $\Psi(x)$, whose minima and maxima correspond to stable and unstable fixed points of the mean-field equation. An example of the bistable case is shown in Fig. 2.

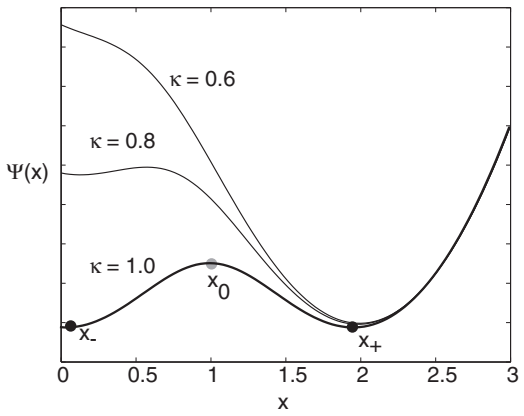

FIG. 2. Bistable potential Ψ for the deterministic network satisfying $\dot{x} = -x + F(x) = -d\Psi/dx$, with F given by the sigmoid (5.4) for $\gamma = 4$, $\kappa = 1.0$, and $F_0 = 2$. There exist two stable fixed points $x_\pm$ separated by an unstable fixed point x_0. As the threshold κ is reduced the network switches to a monostable regime via a saddle-node bifurcation.

Let us first calculate the quasipotential using the QSS diffusion approximation of Sec. IIIA. Substituting Eqs. (5.11), (5.19), and (5.20) into Eqs. (3.10) and (3.11), we find that [47]

$$Z_n(x) = w[n - F(x)]\rho_n(x),$$

so that

$$D(x) = w \sum_n [n - F(x)](-x + wn)\rho_n(x) = w\{w\langle n^2\rangle - [x + wF(x)]\langle n\rangle + xF(x)\} = w\{w[F(x) + F(x)^2] - [x + wF(x)]F(x) + xF(x)\} = w^2 F(x). \quad (5.22)$$

We have used the fact that the mean and variance of the Poisson distribution (5.11) are both $F(x)$. It follows from Eq. (3.16) that the FP quasipotential is

$$\Phi_{\rm FP}(x) = -\int^{x} \frac{V(y)}{D(y)}dy = -\int^{u} \frac{-y + wF(y)}{w^2 F(y)}dy. \quad (5.23)$$

As we now show, this differs significantly from the more accurate estimate of the quasipotential obtained from the path-integral formulation of Sec. IV. The latter shows there exists a Hamiltonian H that can be identified with the Perron eigenvalue $\lambda_0(x,p)$, which is the unique nonzero solution of the eigenvalue equation

$$\sum_m [A_{nm}(x) + q\delta_{n,m}v_m(x)]R_m^{(0)}(x,q) = \lambda_0(x,q)R_n^{(0)}(x,q), \quad (5.24)$$

for which $R_n^{(0)}$ is positive for all n. Substituting the explicit expressions for $v_n(x)$ and $\mathbf{A}$, we have $R_n^{(0)}(x,p) = \psi_n(x,p)$, where

$$\begin{aligned} &F(x)\psi_{n-1}(x,p) - [\lambda_0 + F(x) + n]\psi_n(x,p) \\ &\quad + (n+1)\psi_{n+1}(x,p) \\ &= -p(-x + wn)\psi_n(x,p). \end{aligned}$$

Since the eigenvector associated with the Perron eigenvalue is positive, we consider the trial solution

$$\psi_n(x,p) = \frac{\Lambda(x,p)^n}{n!}, \quad (5.25)$$

which yields the following equation relating Λ and p:

$$\left[\frac{F(x)}{\Lambda} - 1\right] n + \Lambda - F(x) - \lambda_0 = -p(-x + wn).$$

We now collect terms independent of n and linear in n, respectively, to obtain the pair of equations

$$p = -\frac{1}{w}\left[\frac{F(x)}{\Lambda} - 1\right], \quad \Lambda = F(x) + px + \lambda_0.$$

Eliminating Λ from these equations gives

$$H(x,p) \equiv \lambda_0(x,p) = pw\frac{F(x)}{1 - wp} - px, \quad (5.26)$$

where we identify H as the Hamiltonian.

Suppose that the deterministic Eq. (5.21) has a unique stable fixed point at x_*. The corresponding steady-state density of the stochastic network is obtained by taking the initial time $t \to -\infty$ and $x(-\infty) = x_*$ in Eq. (4.6). Optimal paths are then given by zero-energy solutions $H(x,p) = 0$ of Eq. (5.26). One such solution is $p = 0$ for which λ_0 corresponds to the zero Perron eigenvalue of the matrix $\mathbf{A}$. Moreover, $R_m^{(0)}(x,0) = \rho(x,n)$ and $\xi_m^{(0)} = 1$, since they correspond to the right and left eigenvectors of the matrix $\mathbf{A}$, and the equation $\dot{x} = \partial_p H$ reduces to the mean-field Eq. (5.21). There is also one nontrivial zero-energy solution given by

$$p = -\mu(x) \equiv \frac{1}{w}\left[1 - \frac{wF(x)}{x}\right]. \quad (5.27)$$

Finally, evaluating the action along the nontrivial zero-energy solution determines the quasipotential:

$$\Phi_0(x) = -\int^{x} \mu(y)dy = -\int^{x} \frac{1}{w}\left[\frac{wF(y)}{y} - 1\right]dy. \quad (5.28)$$

The quasipotential $\Phi_0(x)$ for the network operating in a bistable regime is plotted in Fig. 3 and compared with the quasipotential based on the QSS diffusion approximation. It can be seen that they differ significantly over a wide range of values of x, thus resulting in exponential errors when calculating the MFPT to escape from a metastable state. Note that it is also possible to derive the quasipotential Φ_0 using a WKB approximation [12], but the latter does not have an obvious Hamiltonian structure. Hence the use of path integrals or large deviation theory is crucial in order to obtain the correct variational principle; this becomes even more significant in the case of the multipopulation model (see below).

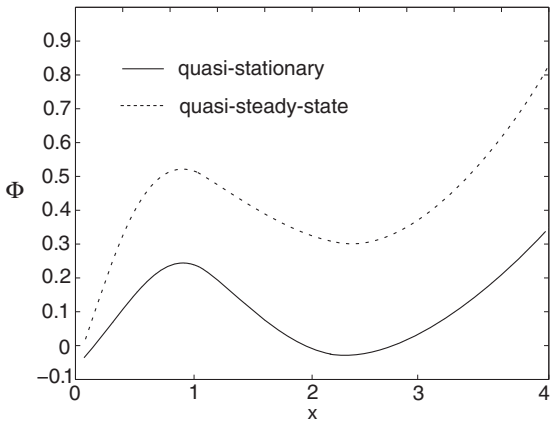

FIG. 3. Comparison of the quasipotentials $\Phi_0(x)$ and $\Phi_{\rm FP}(x)$ obtained using the quasistationary approximation and the QSS diffusion approximation, respectively. Parameter values are chosen so that deterministic network is bistable: $F_0 = 2$, $\gamma = 4$, $\kappa = 1$, and $w = 1.15$.

Another useful feature of the path-integral representation (4.6) is that it provides a direct method for deriving the Gaussian approximation underlying the QSS reduction of Sec. III. Performing the rescaling $p \to ip/\epsilon$ in the action (4.7) with λ_0 given by Eq. (5.26) gives the path integral

$$P(x,t) = \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} D[x]D[p]\exp\left(-\int_0^{\tau} ip\left[\dot{x} + x - \frac{wF(x)}{1 - i\epsilon wp}\right]dt\right). \quad (5.29)$$

The Gaussian approximation involves Taylor expanding the Lagrangian to first order in ϵ, which yields a quadratic in p:

$$P(x,t) = \int_{x(0)=x_0}^{x(\tau)=x} D[x]D[p]\exp\left(\int_0^{\tau}\{ip[\dot{x} + x - wF(x)] - \epsilon p^2 w^2 F(x)\}dt\right).$$

We can now perform the integration over p either directly or by returning to the discretized path integral, Taylor expanding to second order in p_j, and then performing the Gaussian integration with respect to p_j before taking the continuum limit:

$$P(x,t)=\int_{x(0)=x_0}^{x(\tau)=x} D[x]\exp\left(-\int_0^\tau \frac{[\dot{x}+x-wF(x)]^2}{4\epsilon w^2 F(x)}dt\right).$$

This path integral is identical in form to the Onsager-Machlup path-integral representation [48] of solutions to the FP equation (3.9) obtained using a QSS reduction, whose corresponding Langevin equation is

$$dX=-X+wF(X)+\sqrt{2\epsilon D(X)}dW(t), \tag{5.30}$$

with $D(x)=w^2F(x)$. Since there is no additional Jacobian factor in the Onsager-Machlup path integral, it follows that the Langevin equation is of the Ito form.

B. Variational principle and optimal paths for a multipopulation model

The path-integral formulation of solutions to the multipopulation CK equation (5.8) again yields a variational principle, with an action generated from a Hamiltonian $H(\mathbf{x},\mathbf{p})$ corresponding to the Perron eigenvalue of the following linear equation [cf. Eq. (5.24)]:

$$\sum_{\mathbf{m}} W(\mathbf{n},\mathbf{m};\mathbf{x})R^{(0)}(\mathbf{x},\mathbf{p},\mathbf{m}) = \left[\lambda_0(\mathbf{x},\mathbf{p})-\sum_{\alpha=1}^M p_\alpha v_\alpha(\mathbf{x},\mathbf{n})\right]R^{(0)}(\mathbf{x},\mathbf{p},\mathbf{n}), \tag{5.31}$$

and $\xi^{(0)}$ is the adjoint eigenvector. Equation (5.31) can be solved for the Perron eigenvalue using the ansatz

$$R^{(0)}(\mathbf{x},\mathbf{p},\mathbf{n})=\prod_{\alpha=1}^M \frac{\Lambda_\alpha(\mathbf{x},\mathbf{p})^{n_\alpha}}{n_\alpha!}. \tag{5.32}$$

Substituting into Eq. (5.31) and using the explicit expressions for $\mathbf{A}$ and v_α, we find that

$$\sum_{\alpha=1}^M\left(\left[\frac{F(x_\alpha)}{\Lambda_\alpha}-1\right]n_\alpha+\Lambda_\alpha-F(x_\alpha)\right)-\lambda_0 = -\sum_{\alpha=1}^M p_\alpha\left[-x_\alpha+\sum_\beta w_{\alpha\beta}n_\beta\right]. \tag{5.33}$$

Collecting terms in n_α for each α yields

$$\frac{F(x_\alpha)}{\Lambda_\alpha}-1=-\sum_{\beta=1}^M p_\beta w_{\beta\alpha}, \tag{5.34}$$

and collecting terms independent of all n_α gives

$$\lambda_0=\sum_{\alpha=1}^M[\Lambda_\alpha-F(x_\alpha)-x_\alpha p_\alpha]. \tag{5.35}$$

Solving for each Λ_α in terms of $\mathbf{p}$, we have

$$\lambda_0(\mathbf{x},\mathbf{p})\equiv\sum_{\alpha=1}^M\left[\frac{F(x_\alpha)}{1-\sum_{\beta=1}^M p_\beta w_{\beta\alpha}}-x_\alpha p_\alpha-F(x_\alpha)\right], \tag{5.36}$$

which we identify as the Hamiltonian H.

Suppose that the underlying deterministic system (5.17) has a unique stable fixed $\mathbf{x}_*$. The quasipotential of the corresponding stationary density can then be obtained by finding zero-energy solutions of Hamilton's equations

$$\dot{\mathbf{x}}=\nabla_{\mathbf{p}}\mathcal{H}(\mathbf{x},\mathbf{p}),\quad \dot{\mathbf{p}}=-\nabla_{\mathbf{x}}\mathcal{H}(\mathbf{x},\mathbf{p}), \tag{5.37}$$

with $\mathbf{x}=(x,y),\mathbf{p}=(p_x,p_y)$. Substituting for $\mathcal{H}$, Hamilton's equations have the explicit form

$$\frac{dx_\alpha}{dt}=-x_\alpha+\sum_\beta\frac{w_{\alpha\beta}F(x_\alpha)}{1-\sum_{\gamma=1}^M p_\gamma w_{\gamma\alpha}}, \tag{5.38a}$$

$$\frac{dp_\alpha}{dt}=p_\alpha-\frac{F'(x_\alpha)}{1-\sum_{\gamma=1}^M p_\gamma w_{\gamma\alpha}}+F'(x_\alpha). \tag{5.38b}$$

The quasipotential Φ can now be identified as the action along a zero-energy solution curve $\mathbf{x}(t)$. That is,

$$\frac{d\Phi}{dt}\equiv\sum_{\alpha=1}^M\frac{\partial\Phi}{\partial x_\alpha}\frac{dx_\alpha}{dt}=\sum_{\alpha=1}^M p_\alpha\frac{dx_\alpha}{dt}, \tag{5.39}$$

and we can make the identification $p_\alpha=\partial\Phi/\partial x_\alpha$. Note that one zero-energy solution is $\mathbf{p}=\mathbf{0}$, for which λ_0 corresponds to the zero Perron eigenvalue of the matrix $\mathbf{W}$; see Eq. (5.31). It follows that $R^{(0)}(\mathbf{x},\mathbf{0},\mathbf{n})=\rho(\mathbf{x},\mathbf{n})$ and $\xi^{(0)}(\mathbf{x},\mathbf{0},\mathbf{n})=1$, since they correspond to the right and left eigenvectors of the matrix $\mathbf{W}$, respectively. Moreover, Eq. (5.38a) reduces to the mean-field Eq. (5.17) and $\Phi=0$.

In our previous paper [12], we derived the quasipotential for a two-population model ($M=2$) using WKB methods, and showed that Φ satisfied a Hamilton-Jacobi equation, which is recovered by setting $\lambda_0=0$ and $p_\alpha=\partial\Phi/\partial x_\alpha$ in Eq. (5.31). We also assumed that the solutions of the corresponding Hamilton's equations generated optimal stochastic paths (paths of maximum likelihood) in the phase space of the two-population model. Although our conclusions based on WKB appeared consistent with Monte Carlo simulations of the full stochastic system (5.6), we did not have a variational principle to prove that these paths were indeed optimal paths. Moreover, our particular choice of Hamiltonian was not uniquely determined using the WKB method. The path-integral formulation of Sec. IV has allowed us to derive a variational problem from first principles, and to establish that the characteristic paths of the Hamiltonian constructed in Ref. [12] are indeed optimal paths. In Figs. 4 and 5 we show a few results from our previous analysis of a bistable two-population model [12], in order to illustrate the important role that optimal paths play in higher-dimensional escape problems. Figure 4(a) shows optimal paths originating from each of the stable fixed points. If a trajectory crosses the separatrix away from the saddle, it is most likely to cross the separatrix above the saddle when starting from the left fixed point and below the saddle when starting from the right

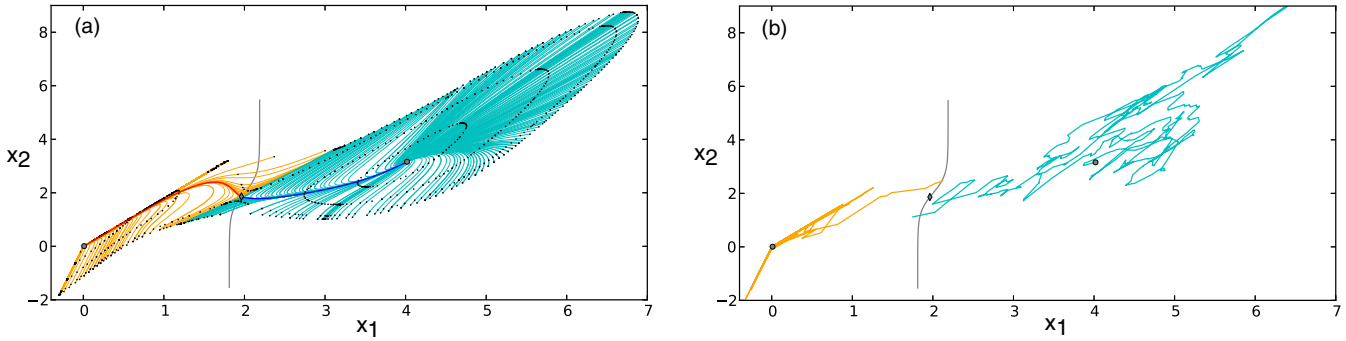

FIG. 4. (Color online) Two-population network with an excitatory population ($\alpha = 1$) and an inhibitory population ($\alpha = 2$). Parameters are chosen so that the network is bistable: $F_0 = 1$, $\gamma = 3$, $\kappa = 2$, $w_{11} = 5$, $w_{12} = -1$, $w_{21} = 9$, and $w_{22} = -6$. (a) Characteristic paths of maximum likelihood emerging from a pair of stable fixed points separated by a saddle. Paths originating from the left (right) stable fixed point are shown in orange (or light gray) and cyan (or dark gray), respectively, with the paths connecting to the saddle shown as thicker horizontal curves. The gray vertical curve is the separatrix Σ. Level curves of constant Φ are shown as black dots. (b) Sample trajectories of the stochastic network using the Gillespie algorithm with $\epsilon = 0.05$ and $N\Delta t = 1$. (The maximum likelihood paths are independent of ϵ.)

fixed point; see also Figs. 4(b) and Fig. 5(a). The distribution of exit points along the separatrix is shown for optimal paths originating from either metastable state (square symbols show the histogram for exit from the left well and, likewise, "o" symbols for the right well). Each density function is peaked away from the saddle point, showing a phenomena known as saddle-point avoidance [49,50].

As in the one-population model, we can use the multipopulation path integral (4.12) to derive a Gaussian approximation of the stochastic hybrid system, equivalent to the one obtained using the more complicated QSS reduction. Again we will ignore the factor $R^{(0)}(\mathbf{x},\mathbf{p}(\tau),\mathbf{n})\xi^{(0)}(\mathbf{x}_0,\mathbf{p}(0),\mathbf{n}_0)$ in Eq. (4.12), which projects on to stochastic trajectories that start in the discrete state $\mathbf{n}_0$ and terminate in the discrete state $\mathbf{n}$. Therefore, we will consider the probability density [for fixed $\mathbf{x}(0) = \mathbf{x}_0$]

$$P(\mathbf{x},t) = \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} D[\mathbf{x}]D[\mathbf{p}]e^{-S[\mathbf{x},\mathbf{p}]/\epsilon}, \tag{5.40}$$

with the action defined by Eq. (4.13). Performing the rescaling $\mathbf{p} \to i\mathbf{p}/\epsilon$ then gives

$$P(x,t) = \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} D[\mathbf{x}]D[\mathbf{p}] \exp\left(-\int_0^\tau i\sum_\alpha p_\alpha\left[\dot{x}_\alpha + x_\alpha - \sum_\beta \frac{w_{\alpha\beta}F(x_\beta)}{1 - i\epsilon\sum_\gamma w_{\gamma\beta}p_\gamma}\right]dt\right).$$

The Gaussian approximation involves Taylor expanding the Lagrangian to first order in ϵ, which yields a quadratic in p:

$$P(x,t) = \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} D[\mathbf{x}]D[\mathbf{p}] \exp\left(\int_0^\tau\left[i\sum_\alpha p_\alpha\left(\dot{x}_\alpha + x_\alpha - \sum_\beta w_{\alpha\beta}F(x_\beta)\right) - \epsilon\sum_{\alpha,\gamma} p_\alpha \mathcal{Q}_{\alpha\gamma}(\mathbf{x})p_\gamma\right]dt\right),$$

where $\mathcal{Q}_{\alpha\gamma}(\mathbf{x}) = \sum_\beta w_{\alpha\beta}F(x_\beta)w_{\gamma\beta}$. Performing the Gaussian integration along similar lines to the one-population model yields the multivariate Onsager-Machlup path integral

$$P(\mathbf{x},t) = \int D[\mathbf{x}]e^{-\mathcal{A}[\mathbf{x}]/\epsilon},$$

with action functional

$$\mathcal{A}[\mathbf{x}] = \frac{1}{4}\int_0^\tau \sum_{\alpha,\beta}\{\dot{x}_\alpha(t) - V_\alpha[\mathbf{x}(t)]\}\mathcal{Q}^{-1}_{\alpha\beta}(\mathbf{x}) \times \{\dot{x}_\beta(t) - V_\beta[\mathbf{x}(t)]\}dt, \tag{5.41}$$

where $V_\alpha(\mathbf{x}) = -x_\alpha + \sum_\beta w_{\alpha\beta}F(x_\beta)$. The corresponding Ito Langevin equation is

$$dX_\alpha(t) = V_\alpha(\mathbf{X})dt + \sqrt{2\epsilon}\sum_\beta w_{\alpha\beta}\sqrt{F(x_\beta)}dW_\beta(t), \tag{5.42}$$

where the $W_\alpha(t)$ are independent Wiener processes.

VI. DISCUSSION

In conclusion, we have constructed a path-integral representation of solutions to a stochastic hybrid system, and used this to derive a large deviation principle. In particular, we have shown that optimal paths of the stochastic dynamics are given by solutions to a Hamiltonian dynamical system, whose Hamiltonian can be identified with the Perron eigenvalue of an appropriately defined linear operator. In this paper, we applied the analysis to a stochastic hybrid neural network and showed that, for this particular system, the action along an optimal path is equal to the quasipotential derived using WKB methods. We illustrated the theory by considering escape from a metastable state in a bistable neural network.

Finally, it is important to emphasize that the variational principle derived using path integrals is not simply an alternative to WKB methods, but subsumes the latter in a more general theory. It is a common feature of many Markov processes under

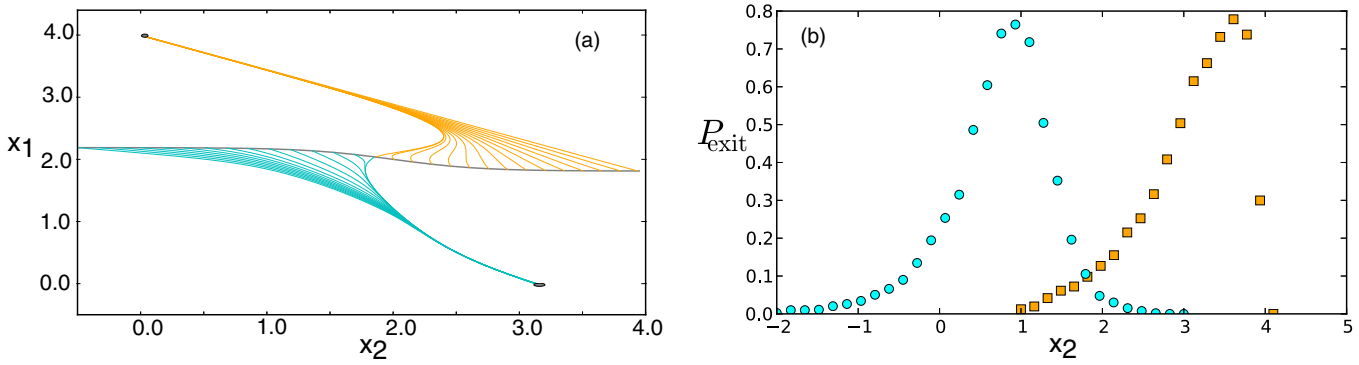

FIG. 5. (Color online) (a) Maximum-likelihood trajectories crossing the separatrix. (b) The probability density for the exit point (x_2 coordinate) where the separatrix is crossed by an exiting trajectory. Results are obtained by 10^2 Monte Carlo simulation with the same parameters as used in Fig. 4, with $\epsilon = 0.08$. The square symbols show trajectories from the left well and "o" symbols show trajectories from the right well.

weak noise conditions that the WKB approximation results in a static Hamilton-Jacobi equation for the quasipotential. This includes processes for which detailed balance does not hold. In terms of the WKB analysis, one of the key differences between a birth-death or continuous Markov processes and a hybrid process is that the latter does not result in a uniquely defined Hamiltonian. The path-integral derivation of the WKB approximation resolves this ambiguity and yields a variational formulation of the quasipotential. That is, the quasipotential is defined as the minimum of all possible paths satisfying the Hamiltonian dynamical system. Recently, numerical algorithms have been developed for solving static Hamilton-Jacobi equations that take advantage of the variational formulation (e.g., an ordered upwind method [51] and the geometric minimum action method [52]). The path-integral formulation presented here allows these algorithms to be adapted for use in analyzing hybrid stochastic processes.

ACKNOWLEDGMENTS

P.C.B. was supported by the National Science Foundation (DMS-1120327) and J.M.N. by the Mathematical Biosciences Institute and the National Science Foundation under Grant No. DMS 0931642.

[1] R. F. Fox and Y. N. Lu, Phys. Rev. E **49**, 3421 (1994).
[2] C. C. Chow and J. A. White, Biophys. J. **71**, 3013 (1996).
[3] J. P. Keener and J. M. Newby, Phy. Rev. E **84**, 011918 (2011).
[4] J. H. Goldwyn and E. Shea-Brown, PLoS Comput. Biol. **7**, e1002247 (2011).
[5] E. Buckwar and M. G. Riedler, J. Math. Biol. **63**, 1051 (2011).
[6] J. M. Newby, P. C. Bressloff, and J. P. Keener, Phys. Rev. Lett. **111**, 128101 (2013).
[7] P. C. Bressloff and J. M. Newby, Phys. Biol. **11** 016006 (2014).
[8] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[9] J. M. Newby, Phys. Biol. **9**, 026002 (2012).
[10] J. M. Newby and P. C. Bressloff, Phys. Biol. **7**, 036004 (2010).
[11] J. M. Newby and J. P. Keener, SIAM Multiscale Model. Simul. **9**, 735 (2011).
[12] P. C. Bressloff and J. M. Newby, SIAM Appl. Dyn. Syst. **12**, 1394 (2013).
[13] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1998).
[14] A. Faggionato, D. Gabrielli, and M. Crivellari, Markov Process. Relat. Fields **16**, 497 (2010).
[15] H. Touchette, Phys. Rep. **478**, 1 (2009).
[16] P. Hanggi, H. Grabert, P. Talkner, and H. Thomas, Phys. Rev. A **29**, 371 (1984).
[17] C. Knessl, B. J. Matkowsky, Z. Schuss, and C. Tier, SIAM J. Appl. Math. **45**, 1006 (1985).
[18] M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, J. Chem. Phys. A **100**, 5735 (1994).
[19] V. Elgart and A. Kamenev, Phys. Rev. E **70**, 041106 (2004).
[20] C. R. Doering, K. V. Sargsyan, L. M. Sander, and E. Vanden-Eijnden, J. Phys.: Condens. Matter **19**, 065145 (2007).
[21] M. Vellela and H. Qian, J. R. Soc. Interface **6**, 925 (2009).
[22] M. Do, J. Phys. A **9**, 1465 (1976).
[23] M. Doi, J. Phys. A **9**, 1479 (1976).
[24] L. Peliti, J. Phys. **46**, 1469 (1985).
[25] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[26] C. W. Gardiner, *Handbook of Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).
[27] T. G. Kurtz, Math. Prog. Stud. **5**, 67 (1976).
[28] R. Hinch and S. J. Chapman, Eur. J. Appl. Math. **16**, 427 (2005).
[29] C. Escudero and A. Kamanev, Phys. Rev. E **79**, 041149 (2009).
[30] X. Xing and L. Radzihovsky, Phys. Rev. E **71**, 011802 (2005).
[31] A. Kamenev and B. Meerson, Phys. Rev. E **77**, 061107 (2008).
[32] A. Dobrinevski and E. Frey, Phys. Rev. E **85**, 051903 (2012).
[33] O. Ovaskainen and B. Meerson, Trends Ecol. Evol. **25**, 643 (2010).
[34] The Perron-Frobenius theorem asserts that a real square matrix with positive entries has a unique largest real eigenvalue (the Perron eigenvalue) and that the corresponding eigenvector has strictly positive components. The theorem can also be extended to matrices with non-negative entries, provided that the matrix

is irreducible. However, there can now be complex eigenvalues with the same absolute value as the Perron eigenvalue. In the case of a transition matrix, the Perron eigenvalue is zero. Strictly speaking, the Perron-Frobenius theorem applies to finite-dimensional matrices, so we will assume that it still holds for **A** in cases where the number of discrete states $K \to \infty$.

[35] G. C. Papanicolaou, Bull. Am. Math. Soc. **81**, 330 (1975).

[36] H. Othmer, S. Dunbar, and W. Alt, J. Math. Biol. **26**, 263 (1988).

[37] T. Hillen and H. Othmer, SIAM J. Appl. Math. **61**, 751 (2000).

[38] M. C. Reed, S. Venakides, and J. J. Blum, SIAM J. Appl. Math. **50**, 167 (1990).

[39] A. Friedman and G. Craciun, J. Math. Biol. **51**, 217 (2005).

[40] J. M. Newby and P. C. Bressloff, Bull. Math. Biol. **72**, 1840 (2010).

[41] M. A. Buice and J. D. Cowan, Phys. Rev. E **75**, 051919 (2007).

[42] P. C. Bressloff, SIAM J. Appl. Math. **70**, 1488 (2009).

[43] M. Buice, J. D. Cowan, and C. C. Chow, Neural Comput. **22**, 377 (2010).

[44] H. R. Wilson and J. D. Cowan, Kybernetik **13**, 55 (1973).

[45] S.-i. Amari, Biol. Cybern. **27**, 77 (1977).

[46] G. B. Ermentrout, Neural Comput. **6**, 679 (1994).

[47] Note that the solution $Z_n(x)$ to Eq. (3.11) is nonunique, since $\rho_n(x)$ is a null vector of the linear operator on the left-hand side of (3.11). In our previous paper [12], we took $Z_n(x) = wn\rho_n(x)$, which does not satisfy the constraint $\sum_n Z_n(x) = 0$ for all x. In order to satisfy this constraint, we require $Z_n(x) = w[n - F(x)]\rho_n(x)$ as $\langle n \rangle = F(x)$. The resulting qusaipotential $\Phi_{\rm FP}(x)$ thus differs from the one constructed in [12]; however, it still differs from the WKB quasipotential.

[48] R. Graham and T. Tel, J. Stat. Phys. **35**, 729 (1984).

[49] R. S. Maier and D. L. Stein, SIAM J. Appl. Math. **57**, 752 (1997).

[50] Z. Schuss, *Theory and Applications of Stochastic Processes: An Analytical Approach*, Applied Mathematical Sciences Vol. 170 (Springer, New York, 2010).

[51] M. K. Cameron, Physica D **241**, 1532 (2012).

[52] M. Heymann and E. Vanden-Eijnden, Commun. Pure Appl. Math. **61**, 1052 (2008).