A variational method for analyzing limit cycle oscillations in stochastic hybrid systems

Paul C. Bressloff, and James MacLaurin

Citation: Chaos **28**, 063105 (2018); doi: 10.1063/1.5027077
View online: https://doi.org/10.1063/1.5027077
View Table of Contents: http://aip.scitation.org/toc/cha/28/6
Published by the American Institute of Physics

A variational method for analyzing limit cycle oscillations in stochastic hybrid systems

Paul C. Bressloff and James MacLaurin
Department of Mathematics, University of Utah, Salt Lake City, Utah 84112, USA

(Received 27 February 2018; accepted 18 May 2018; published online 5 June 2018)

Many systems in biology can be modeled through ordinary differential equations, which are piecewise continuous, and switch between different states according to a Markov jump process known as a stochastic hybrid system or piecewise deterministic Markov process (PDMP). In the fast switching limit, the dynamics converges to a deterministic ODE. In this paper, we develop a phase reduction method for stochastic hybrid systems that support a stable limit cycle in the deterministic limit. A classic example is the Morris-Lecar model of a neuron, where the switching Markov process is the number of open ion channels and the continuous process is the membrane voltage. We outline a variational principle for the phase reduction, yielding an exact analytic expression for the resulting phase dynamics. We demonstrate that this decomposition is accurate over timescales that are exponential in the switching rate ϵ^{-1}. That is, we show that for a constant C, the probability that the expected time to leave an $O(a)$ neighborhood of the limit cycle is less than T scales as $T\exp(-Ca/\epsilon)$. *Published by AIP Publishing.* https://doi.org/10.1063/1.5027077

Oscillations abound in nature, from the beating of the heart to the genetic circadian clock that synchronizes with the day-night cycle. However, oscillations are often subject to stochastic fluctuations, which in extreme cases can lead to heart failure or severe jet lag, for example. A subject of intense study has been how to determine the corresponding fluctuations in the amplitude and phase of an oscillator, both at the single and population levels. This theory plays a crucial role in understanding how coupled or noise-driven stochastic oscillators can synchronize, and when oscillators fail. In this paper, we develop the first systematic study of stochastic oscillators of a particular form, in which the dynamics is said to be piecewise deterministic. This means that the state of the system evolves deterministically except at a sequence of random times where the deterministic dynamics switches to a different mode. The major results of our work are as follows: (i) deriving a stochastic phase equation for a hybrid oscillator and (ii) obtaining strong exponential bounds on the size of amplitude fluctuations. The former provides a framework for studying phase synchronization in populations of oscillators, whereas the latter is crucial for determining the time-scale over which the notion of a phase oscillator can be maintained. We illustrate the theory using the example of a neuron whose voltage depends on how many ion channels in its membrane are open. Jumps in the dynamics occur whenever one of the channels opens or closes. However, there are many other applications in the natural world, including gene and brain networks.

I. INTRODUCTION

There is a growing class 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 Γ.[1] The resulting stochastic hybrid system is known as a piecewise deterministic Markov process (PDMP).[2] (A more general type of stochastic hybrid system occurs when the continuous process is itself stochastic.) One important example is given by membrane voltage fluctuations arising from the stochastic opening and closing of ion channels.[3–13] 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 thermodynamic limit that the number of ion channels goes to infinity, one can apply the law of large numbers and recover classical Hodgkin-Huxley type ordinary differential equations (ODEs). However, finite-size effects can result in the noise-induced spontaneous firing of a neuron due to channel fluctuations. Another major example of a stochastic hybrid system occurs within the context of gene regulatory networks. Now the continuous variables are the concentrations of protein products (and possibly mRNAs) and the discrete variables represent the various activation/inactivation states of the genes.[14–20] Yet another example is given by a recent stochastic formulation of synaptically coupled neural networks that has a mathematical structure analogous to regulatory gene networks.[21]

In the above examples, one often 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 relaxation dynamics then $t\epsilon$ is the characteristic time-scale of the Markov chain for some small positive dimensionless parameter ϵ. If the Markov chain is ergodic, then in the fast switching or adiabatic limit $\epsilon \to 0$, one obtains a deterministic dynamical

1054-1500/2018/28(6)/063105/18/$30.00

system in which one averages the piecewise dynamics with respect to the corresponding unique stationary distribution. In the case of gene regulatory networks, the switching on and off of a gene is due to the binding/unbinding of regulatory proteins (transcription factors) to gene promoter sites. Hence, in the fast switching limit, the binding/unbinding reactions are much faster than the rates of synthesis and degradation. (This is often assumed in the studies of stochastic gene expression, which typically focus on the effects of fluctuations in protein numbers.) On the other hand, in single-neuron models, fast switching means that ion channels open and close much faster than the voltage evolves. This is certainly the case for Na^+ ion channels.

Suppose that the deterministic dynamical system obtained in the adiabatic limit $\epsilon \to 0$ exhibits some non-trivial dynamics such as bistability or a limit cycle oscillation. This raises the general issue of determining how the dynamics is affected by switching in the weak noise regime, $0 < \epsilon \ll 1$. In the case of bistability, a variety of methods have been developed to explore noise-induced transitions and metastability in PDMPs, including rigorous large deviation theory,[22–24] WKB approximations and matched asymptotics,[6,10,11,18] and path-integrals.[25] On the other hand, as far as we are aware, there has been very little numerical or analytical work on limit cycle oscillations in PDMPs. A few notable exceptions are Refs. 26–28. However, none of these studies develop a fundamental theory of stochastic limit cycle oscillations in PDMPs analogous to phase reduction methods for stochastic differential equations (SDEs).

Regarding the latter, suppose that a deterministic smooth dynamical system $\dot{x} = F(x)$, $x \in \mathbb{R}^d$ supports a limit cycle $x(t) = \Phi(\theta(t))$ of period Δ_0, where $\theta(t)$ is a uniformly rotating phase, $\dot{\theta} = \omega_0$ and $\omega_0 = 2\pi/\Delta_0$. The phase is neutrally stable with respect to perturbations along the limit cycle—this reflects invariance of an autonomous dynamical system with respect to time shifts. Now suppose that the dynamical system is perturbed by weak Gaussian noise such that $dX = F(X)dt + \sqrt{2\epsilon}G(X)dW(t)$, where $W(t)$ is a d-dimensional vector of independent Wiener processes. If the noise amplitude ϵ is sufficiently small relative to the rate of attraction to the limit cycle, then deviations transverse to the limit cycle are also small (up to some exponentially large stopping time). This suggests that the definition of a phase variable persists in the stochastic setting, and one can derive a stochastic phase equation. However, there is not a unique way to define the phase, which has led to two complementary methods for obtaining a stochastic phase equation: (i) the method of isochrons[29–34] and (ii) an explicit amplitude-phase decomposition.[35–37] (See also the recent survey by Ashwin *et al.*[38])

Recently, we introduced a variational method for carrying out the amplitude-phase decomposition for SDEs, which yields exact SDEs for the amplitude and phase,[39] equivalent to those obtained in Ref. 37 using the implicit function theorem. Within the variational framework, different choices of phase correspond to different choices of the inner product space $\mathbb{R}^d$. In particular, we took a weighted Euclidean norm, so that the minimization scheme determined the phase by projecting the full solution on to the limit cycle using Floquet vectors. Hence, in a neighborhood of the limit cycle, the phase variable coincided with the isochronal phase.[37] This had the advantage that the amplitude and phase decoupled to leading order. In addition, we used the exact amplitude and phase equations to derive strong exponential bounds on the growth of transverse fluctuations.

In this paper, we develop a variational method for PDMPs that support a limit cycle in the adiabatic limit. We derive an exact equation for the phase, which takes the form of an implicit PDMP. Moreover, we show how the latter can be converted to an explicit PDMP for the phase by performing a perturbation expansion in ϵ and show that the phase decouples from the amplitude to leading order. We also consider an alternative approach to analyzing oscillations in PDMPs, based on first carrying out a quasi-steady-state (QSS) diffusion approximation of the full PDMP to obtain an SDE[40] and then performing a phase reduction. We compare the resulting SDE for the phase with the corresponding SDE obtained by carrying out a QSS reduction of the phase-based PDMP. However, one important limitation of any diffusion approximation is that it tends to generate exponentially large errors when estimating the probability of rare events; rare events contribute to the long-time growth of transverse fluctuations.

One major feature of our variational approach is that it allows us to obtain an exponential bound on the growth of transverse fluctuations. This issue, which is typically ignored in the studies of stochastic phase oscillators, is important since any phase reduction scheme ultimately breaks down over sufficiently long time-scales, since there is a non-zero probability of leaving a bounded neighborhood of the limit cycle, and the notion of phase no longer makes sense. Using our variational method, we show that for a constant C, and all $a \le a_0$ (a_0 being a constant independent of ϵ), the probability that the expected time to leave an $O(a)$ neighborhood of the limit cycle is less than T scales as $T\exp(-Ca/\epsilon)$. An interesting difference between the above bound and the corresponding one obtained for SDEs[39] is that in the latter the bound is of the form $T\exp(-Cba/\epsilon)$, where b is the rate of decay towards the limit cycle. In other words, in the SDE case, the bound is still powerful in the large ϵ case, as long as $b\epsilon^{-1} \gg 1$, i.e., as long as the decay towards the limit cycle dominates the noise. However, this no longer holds in the PDMP case. Now, if ϵ is large, then the most likely way that the system can escape the limit cycle is that in stays in any particular state for too long without jumping and the time that it stays in one state is not particularly affected by b (in most cases).

The organization of the paper is as follows. In Sec. II, we define a stochastic hybrid system or PDMP, discuss the QSS diffusion approximation (see also Appendix A), and apply phase reduction methods to the resulting SDE. In Sec. III, we formulate the variational principle for determining the phase of a stochastic limit cycle in the case of a PDMP and show that the resulting phase equation also takes the form of a PDMP. We illustrate our theory in Sec. IV by considering the stochastic Morris-Lecar model of subthreshold oscillations. Finally, in Sec. V and Appendixes C-E, we obtain an exponential bound on the growth of transverse fluctuations.

II. STOCHASTIC HYBRID LIMIT CYCLE OSCILLATOR

Consider a dynamical system whose states are described by a pair $(x, n) \in \Sigma \times \{0, \ldots, N-1\}$, where x is a continuous variable in a connected bounded domain $\Sigma \subset \mathbb{R}^d$ and n is a discrete stochastic variable taking values in the finite set $\Gamma \equiv \{0, \ldots, N_0 - 1\}$. When the internal state is n, the system evolves according to the ordinary differential equation (ODE)

$$\dot{x} = F_n(x), \tag{2.1}$$

where the vector field $F_n : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth function, locally Lipschitz. We assume that the dynamics of x is confined to the domain Σ so that we have existence and uniqueness of a trajectory for each n. The discrete stochastic variable is taken to evolve according to a homogeneous, continuous-time Markov chain with generator $\mathbf{A}(x)$ for a given x, where

$$A_{nm}(x) = \Lambda_{nm}(x) - \delta_{n,m} \sum_{k \in \Gamma} \Lambda_{kn}(x),$$

and $\mathbf{\Lambda}(x)$ is the transition matrix. We make the further assumption that for each x the chain is irreducible, that is, 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 distribution on Γ with components $\rho_m(x)$, such that

$$\sum_{m \in \Gamma} A_{nm}(x)\rho_m(x) = 0, \quad \forall n \in \Gamma. \tag{2.2}$$

The above stochastic model defines a piecewise deterministic Markov process (PDMP)[2] on $\mathbb{R}^d$. It is also possible to consider generalizations of the continuous process, in which the ODE (2.1) is replaced by a stochastic differential equation (SDE) or even a partial differential equation (PDE). In order to allow for such possibilities, we will refer to all of these processes as examples of a stochastic hybrid system. A useful way to implement a PDMP is as follows, see also Fig. 1. Let us decompose the transition matrix of the Markov chain as

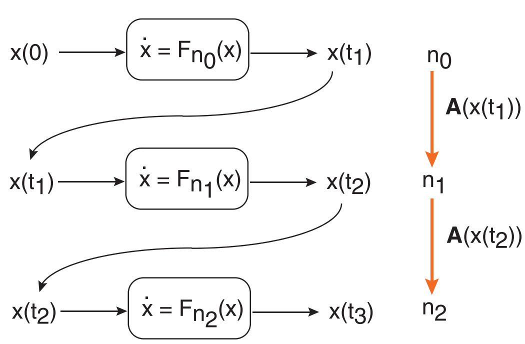

FIG. 1. Schematic diagram of a PDMP for a sequence of jump times $\{t_1, \ldots\}$ and a corresponding of discrete states $\{n_0, n_1, \ldots\}$. See text for details.

$$\Lambda_{nm}(x) = \tilde{\Lambda}_{nm}(x)\lambda_m(x), \tag{2.3}$$

with $\sum_{n \neq m} \tilde{\Lambda}_{nm}(x) = 1$ for all x. Hence, $\lambda_m(x)$ determines the jump times from the state m, whereas $\tilde{\Lambda}_{nm}(x)$ determines the probability distribution that when it jumps the new state is n for $n \neq m$. The hybrid evolution of the system with respect to $x(t)$ and $n(t)$ can then be described as follows. Suppose the system starts at time zero in the state (x_0, n_0). Call $x_0(t)$ the solution of (2.1) with $n = n_0$ such that $x_0(0) = x_0$. Let t_1 be the random variable (stopping time) such that

$$\mathbb{P}(t_1 < t) = 1 - \exp\left(-\int_0^t \lambda_{n_0}(x_0(t'))dt'\right).$$

Then, in the random time interval $s \in [0, t_1)$, the state of the system is $(x_0(s), n_0)$. We draw a value of θ_1 from $\mathbb{P}(t_1 < t)$, choose an internal state $n_1 \in \Gamma$ with probability $\tilde{\Lambda}_{n_1 n_0}(x_0(t_1))$, and call $x_1(t)$ the solution of the following Cauchy problem on $[t_1, \infty)$:

$$\begin{cases} \dot{x}_1(t) = F_{n_1}(x_1(t)), & t \geq \theta_1 \\ x_1(t_1) = x_0(t_1). \end{cases}$$

Iterating this procedure, we construct a sequence of increasing jumping times $(t_k)_{k\geq 0}$ (setting $t_0 = 0$) and a corresponding sequence of internal states $(n_k)_{k\geq 0}$. The evolution $(x(t), n(t))$ is then defined as

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

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.

A. Chapman-Kolmogorov equation

Let $X(t)$ and $N(t)$ denote the stochastic continuous and discrete variables, respectively, at time t, $t > 0$, given the initial conditions $X(0) = x_0, N(0) = n_0$. Introduce the probability density $p_n(x, t|x_0, n_0, 0)$ with

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

It follows that p evolves according to the forward differential Chapman-Kolmogorov (CK) equation[41,42]

$$\frac{\partial p_n}{\partial t} = \mathbb{L}p_n, \tag{2.5}$$

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

$$\mathbb{L}p_n(x, t) = -\nabla \cdot [F_n(x)p_n(x, t)] + \frac{1}{\epsilon}\sum_{m \in \Gamma} A_{nm}(x)p_m(x, t). \tag{2.6}$$

The first term on the right-hand side represents the probability flow associated with the piecewise deterministic dynamics for a given n, whereas the second term represents jumps in the discrete state n. Note that we have rescaled the matrix

A by introducing the dimensionless parameter ϵ, $\epsilon > 0$. This is motivated by the observation that many applications of PDMPs involve a separation of time-scales between the relaxation time for the dynamics of the continuous variables x and the rate of switching between the different discrete states n. The fast switching limit then corresponds to the case $\epsilon \to 0$. Let us now define the averaged vector field $\bar{F} : \mathbb{R} \to \mathbb{R}$ by

$$\bar{F}(x) = \sum_{n\in\Gamma} \rho_n(x) F_n(x). \tag{2.7}$$

It can be shown[23] that, given the assumptions on the matrix **A**, $\bar{F}(x)$ is uniformly Lipschitz. Hence, for all $t \in [0, T]$, the Cauchy problem

$$\begin{cases} \dot{x}(t) = \bar{F}(x(t)) \\ x(0) = x_0 \end{cases} \tag{2.8}$$

has a unique solution for all $n \in \Gamma$. Intuitively speaking, one would expect the stochastic hybrid system (2.1) to reduce to the deterministic dynamical system (2.8) in the fast switching limit $\epsilon \to 0$. That is, for sufficiently small ϵ, the Markov chain undergoes many jumps over a small time interval Δt during which $\Delta x \approx 0$, and thus the relative frequency of each discrete state n is approximately ρ_n. This can be made precise in terms of a Law of Large Numbers for stochastic hybrid systems proven in Ref. 23.

B. Stochastic limit cycle oscillations under the diffusion approximation

For small but non-zero ϵ, one can use perturbation theory to derive lowest order corrections to the deterministic mean field equation, which leads to an SDE with noise amplitude $O(\sqrt{\epsilon})$.[40] More specifically, perturbations of the mean-field Eq. (2.8) can be analyzed using a quasi-steady-state (QSS) diffusion or adiabatic approximation, in which the CK Eq. (2.5) is approximated by a Fokker-Planck (FP) equation for the total density $C(x,t) = \sum_n p_n(x,t)$. The details are presented in Appendix A, and we find that under the Ito representation, the FP equation takes the form

$$\frac{\partial C}{\partial t} = -\nabla \cdot \left[\bar{F}(x)C\right] - \epsilon \nabla \cdot [\mathcal{V}(x)C] + \epsilon \sum_{i,j=1}^{d} \frac{\partial^2 D_{ij}(x)C}{\partial x_i \partial x_j}, \tag{2.9}$$

with the $O(\epsilon)$ correction to the drift, $\mathcal{V}(x)$, and the diffusion matrix $D(x)$ are given by

$$\mathcal{V} = \sum_{n,m} \left\{ (\rho_n F_n)\nabla \cdot (F_m A^{\dagger}_{mn}) - \bar{F}\nabla \cdot (F_m A^{\dagger}_{mn}\rho_n) \right\} \tag{2.10a}$$

and

$$D_{ij} = \sum_{m,n\in\Gamma} \left[F_{m,i} - \bar{F}_i\right] A^{\dagger}_{mn} \rho_n \left[\bar{F}_j - F_{n,j}\right]. \tag{2.10b}$$

In fact, only the symmetric part of $D(x)$ appears in Eq. (2.9) so we will take

$$D_{ij} = -\frac{1}{2} \sum_{m,n\in\Gamma} \left[\bar{F}_i - F_{m,i}\right] \tilde{A}_{mn} \left[\bar{F}_j - F_{n,j}\right], \tag{2.11}$$

where $\tilde{A}_{mn} = A^{\dagger}_{mn}\rho_n + A^{\dagger}_{nm}\rho_m$, i.e., the symmetric part of $A\rho$.

It follows that in the fast switching regime (small ϵ), the deterministic ODE (2.8) can be approximated by the Ito SDE

$$dX = \left[\bar{F}(X) + \epsilon \mathcal{V}(X)\right] dt + \sqrt{2\epsilon}\, G(X) dW(t), \tag{2.12}$$

where ϵ determines the noise strength and $G(X)G^{\top}(X) = D(X)$. Here, $W(t)$ is a vector of uncorrelated Brownian motions in $\mathbb{R}^d$

$$\mathbb{E}\left[W(t)W(t)^{\top}\right] = tI,$$

and I is the $d \times d$ identity matrix.

Now suppose that the unperturbed system (2.8) supports a stable periodic solution $x = \Phi(t)$ with $\Phi(t) = \Phi(t + \Delta_0)$, where $\omega_0 = 2\pi/\Delta_0$ is the natural frequency of the oscillator. In state space, the solution is an isolated attractive trajectory or limit cycle. The dynamics on the limit cycle can be described by a uniformly rotating phase such that

$$\frac{d\theta}{dt} = \omega_0, \tag{2.13}$$

and $x = \Phi(\theta(t))$ with Φ a 2π-periodic function. Note that the phase is neutrally stable with respect to perturbations along the limit cycle—this reflects invariance of an autonomous dynamical system with respect to time shifts. Note that Φ satisfies the equation

$$\omega_0 \frac{d\Phi}{d\theta} = \bar{F}(\Phi(\theta)). \tag{2.14}$$

Differentiating both sides with respect to θ gives

$$\frac{d}{d\theta}\left(\frac{d\Phi}{d\theta}\right) = \omega_0^{-1} \bar{J}(\theta) \cdot \frac{d\Phi}{d\theta}, \tag{2.15}$$

where $\bar{J}$ is the 2π-periodic Jacobian matrix

$$\bar{J}_{jk}(\theta) \equiv \left.\frac{\partial \bar{F}_j}{\partial x_k}\right|_{x=\Phi(\theta)}. \tag{2.16}$$

If the noise amplitude ϵ is sufficiently small relative to the rate of attraction to the limit cycle, then deviations transverse to the limit cycle are also small (up to some exponentially large stopping time). This suggests that the definition of a phase variable persists in the stochastic setting, and one can derive a stochastic phase equation. Here we follow the method of isochrons.[29–34] We only describe the simplest version of the theory, in which $O(\epsilon)$ corrections to the drift term are ignored. The latter arise from transforming between Ito and Stratonovich representations, and coupling between the phase and transverse (amplitude) fluctuations. First, suppose that we stroboscopically observe the unperturbed system at time intervals of length Δ_0. This leads to a Poincare mapping

$$x(t) \to x(t + \bar{\Delta}) \equiv \mathcal{P}(x(t)),$$

for which all points on the limit cycle are fixed points. Choose a point x_* on the limit cycle and consider all points in the vicinity of x^* that are attracted to it under the action of $\mathcal{P}$. They form a $(d-1)$-dimensional hypersurface $\mathcal{I}$ called an isochron,[43–47] crossing the limit cycle at x^*. A unique isochron can be drawn through each point on the limit cycle (at least locally) so the isochrons can be parameterized by the phase, $\mathcal{I}=\mathcal{I}(\theta)$. Finally, the definition of phase is extended by taking all points $x\in\mathcal{I}(\theta)$ to have the same phase, $\Theta(x)=\theta$, which then rotates at the natural frequency ω_0. Hence, for an unperturbed oscillator in the vicinity of the limit cycle, we have

$$\bar{\omega}=\frac{d\Theta(x)}{dt}=\nabla\Theta(x)\cdot\frac{dx}{dt}=\nabla\Theta(x)\cdot\bar{F}(x).$$

Now consider Eq. (2.12) interpreted as a Stratonovich SDE (after dropping $O(\epsilon)$ corrections to the drift) so that the normal rules of calculus apply. Differentiating the isochronal phase using the chain rule gives

$$\begin{aligned}d\Theta&=\nabla\Theta(x)\cdot\left[\bar{F}(X)dt+\sqrt{2\epsilon}G(X)dW(t)\right]\\&=\bar{\omega}dt+\sqrt{2\epsilon}\nabla\Theta(X)\cdot G(X)dW(t).\end{aligned}$$

We now make the further approximation that deviations of X from the limit cycle are ignored on the right-hand side by setting $X(t)=\Phi(\theta(t))$ with Φ as the 2π-periodic solution on the limit cycle. This then yields the closed stochastic phase equation

$$d\theta=\omega_0dt+\sqrt{2\epsilon}\sum_{k,l=1}^{d}R_k(\theta)G_{kl}(\Phi(\theta))dW_l(t), \tag{2.17}$$

where

$$R_k(\theta)=\left.\frac{\partial\Theta}{\partial x_k}\right|_{x=\Phi(\theta)} \tag{2.18}$$

is a 2π-periodic function of θ known as the kth component of the *phase resetting curve* (PRC).[43–47] One way to evaluate the PRC is to exploit the fact that it is the 2π-periodic solution of the linear equation

$$\bar{\omega}\frac{dR(\theta)}{d\theta}=-\bar{J}(\theta)^{\top}\cdot R(\theta), \tag{2.19}$$

under the normalization condition

$$R(\theta)\cdot\frac{d\Phi(\theta)}{d\theta}=1. \tag{2.20}$$

$\bar{J}(\theta)^{\top}$ is the transpose of the Jacobian matrix $\bar{J}(\theta)$.

Finally, we can simplify Eq. (2.17) by noting that the probability law (or statistics) of the sum of stochastic integrals $\sum_{k,l=1}^{d}R_k(\theta)G_{kl}(\Phi(\theta))dW_l(t)$ is identical to the probability law arising from the following single stochastic integral from a single Wiener process $W(t)$, i.e.,

$$d\theta=\omega_0dt+\sqrt{2\epsilon\mathcal{D}(\theta)}dW(t), \tag{2.21}$$

with

$$\begin{aligned}D(\theta)&=\sum_{l=1}^{d}\left(\sum_{k}^{d}R_k(\theta)G_{kl}(\Phi(\theta))\right)\left(\sum_{k'}^{d}R_{k'}(\theta)G_{k'l}(\Phi(\theta))\right)\\&=\sum_{k,k'=1}^{d}R_k(\theta)D_{kk'}(\Phi(\theta))R_{k'}(\theta).\end{aligned} \tag{2.22}$$

The reason that the probability laws of the previous two stochastic processes are identical is that their quadratic variations are identical, i.e.,

$$\begin{aligned}&\left[\int_0^{\cdot}\sqrt{2\epsilon\mathcal{D}(\theta)}dW(t)\right]_s\\&=2\epsilon\int_0^s\mathcal{D}(\theta(t))dt=\left[\int_0^{\cdot}\sqrt{2\epsilon}\sum_{k,l=1}^{d}R_k(\theta)G_{kl}(\Phi(\theta))dW_l(t)\right]_s.\end{aligned}$$

It is a classical result of stochastic analysis that the probability law of a stochastic integral is entirely determined by the above quadratic variation (see, for example, Theorem 4.2 in Ref. 54) One way to understand why this is the case is that a stochastic integral can be characterized as a Brownian motion that has been rescaled in time, with the rescaling determined by the quadratic variation.

The above analysis uses two successive approximations: (i) a diffusion approximation to convert the PDMP to an SDE in the fast switching regime and (ii) a phase reduction of the SDE. Both stages introduce $O(\epsilon)$ corrections to the drift, which we have ignored for ease of presentation. We could also now use the Ito SDE (2.12) to investigate the growth of fluctuations transverse to the limit cycle in the weak noise limit, by applying our recent variational method for analyzing stochastic limit cycle oscillators driven by Gaussian noise.[39] This method yields an implicit stochastic phase equation that is exact even outside the weak noise regime and can be used to derive strong, ϵ-dependent exponential bounds on the growth of transverse fluctuations. However, such a method cannot eliminate the errors introduced by performing the diffusion approximation. This motivates the development of a variational method that can be applied directly to the exact PDMP (2.1). This will yield more accurate bounds on the growth of transverse fluctuations and can also be used to derive an explicit PDMP for the phase.

III. VARIATIONAL PRINCIPLE

In this section, we formulate a variational method for the PDMP (2.1), under the assumption that the latter exhibits a limit cycle oscillation in the fast switching limit. We derive an exact phase equation for the stochastic limit cycle, which now takes the form of an implicit PDMP. Moreover, we show how it can be converted to an explicit PDMP by performing a perturbation expansion in ϵ. Our formulation thus avoids introducing additional errors arising from the diffusion approximation. One potential limitation of any diffusion approximation is that it tends to generate exponentially large errors when estimating the probability of rare events; rare events contribute to the long-time growth of transverse fluctuations.

In order to formulate a variational principle, we fix a particular realization σ_T of the Markov chain up to sometime T, $\sigma_T = \{N(t), 0 \le t \le T\}$. Suppose that there is a finite sequence of jump times $\{t_1, \ldots t_r\}$ within the time interval $(0, T)$ and let n_j be the corresponding discrete state in the interval (t_j, t_{j+1}) with $t_0 = 0$. Introduce the set

$$\mathcal{T} = [0, T] \setminus \cup_{j=1}^{r} \{t_j\}.$$

Analogous to the analysis of SDEs,[39] we wish to decompose the piecewise deterministic solution x_t to the PDMP (2.1) for $t \in \mathcal{T}$ into two components according to

$$x_t = \Phi(\beta_t) + \sqrt{\epsilon} v_t, \tag{3.1}$$

with β_t and v_t corresponding to the phase and amplitude components, respectively. The phase β_t and amplitude v_t evolve according to a PDMP, involving the vector field F_{n_j} in the time intervals (t_j, t_{j+1}), analogous to x_t (see Fig. 1). (It is notationally convenient to switch from $x(t)$ to x_t, etc., in the following.) However, such a decomposition is not unique unless we impose an additional mathematical constraint. We will adapt a variational principle recently introduced to analyze the dynamics of limit cycles with Gaussian noise.[39] In order to construct the variational principle, we first introduce an appropriate weighted norm on $\mathbb{R}^d$, based on a Floquet decomposition.

A. Floquet decomposition and weighted norm

For any $0 \le t$, define $\Pi(t) \in \mathbb{R}^{d\times d}$ to be the following fundamental matrix for the ODE:

$$\frac{dz}{dt} = A(t) z, \tag{3.2}$$

where $A(t) = \bar{J}(\omega_0 t)$. That is, $\Pi(t) := (z_1(t)|z_2(t)|\ldots|z_d(t))$, where $z_i(t)$ satisfies (3.2), and $\{z_i(0)\}_{i=1}^{d}$ is an orthogonal basis for $\mathbb{R}^d$. Floquet Theory states that there exists a diagonal matrix $\mathcal{S} = \mathrm{diag}(\nu_1, \ldots, \nu_d)$ whose diagonal entries are the Floquet characteristic exponents, such that

$$\Pi(t) = P(\omega_0 t) \exp(t\mathcal{S}) P^{-1}(0), \tag{3.3}$$

with $P(\theta)$ being a 2π-periodic matrix whose first column is proportional to $\Phi'(\omega_0 t)$, and $\nu_1 = 0$. That is, $P(\theta)^{-1}\Phi'(\theta) = c_0 \mathbf{e}$ with $\mathbf{e}_j = \delta_{1,j}$ and c_0 being an arbitrary constant. In order to simplify the following notation, we will assume throughout this paper that the Floquet multipliers are real, and hence, $P(\theta)$ is a real matrix. One could readily generalize these results to the case that $\mathcal{S}$ is complex. The limit cycle is taken to be stable, meaning that for a constant $b > 0$, for all $2 \le i \le d$, we have $\nu_i \le -b$. Furthermore, $P^{-1}(\theta)$ exists for all θ, since $\Pi^{-1}(t)$ exists for all t.

The above Floquet decomposition motivates the following weighted inner product: For any $\theta \in \mathbb{R}$, denoting the standard Euclidean dot product on $\mathbb{R}^d$ by $\langle \cdot, \cdot \rangle$

$$\langle x, y \rangle_\theta = \langle P^{-1}(\theta) x, P^{-1}(\theta) y \rangle,$$

and $\|x\|_\theta = \sqrt{\langle x, x \rangle_\theta}$. In the case of SDEs, we showed that this choice of weighting yields a leading order separation of the phase from the amplitude and facilitates strong bounds on the growth of v_t.[39] The former is a consequence of the fact that the matrix $P^{-1}(\theta)$ generates a coordination transformation in which the phase in a neighborhood of the limit cycle coincides with the asymptotic phase defined using isochrons (see also Ref. 37) In particular, one can show that the PRC $R(\theta)$ is related to the tangent vector $\Phi'(\theta)$ according to (see Ref. 39 and Appendix B)

$$P^\top(\theta) R(\theta) = \mathfrak{M}_0^{-1} P^{-1}(\theta) \Phi'(\theta), \tag{3.4}$$

where

$$\mathfrak{M}_0 := \|\Phi'(\theta)\|_\theta^2 = \|P^{-1}(\theta)\Phi'(\theta)\|^2 = c_0^2. \tag{3.5}$$

B. Defining the piecewise deterministic phase using a variational principle

We can now state the variational principle for the stochastic phase: β_t for $t \in \mathcal{T}$ is determined by requiring $\beta_t = a_t(\theta_t)$, where $a_t(\theta_t)$ for a prescribed time dependent weight θ_t is a local minimum of the following variational problem:

$$\inf_{a \in \mathcal{N}(a_t(\theta_t))} \|x_t - \Phi(a)\|_{\theta_t} = \|x_t - \Phi(a_t(\theta_t))\|_{\theta_t}, \quad t \in \mathcal{T}, \tag{3.6}$$

with $\mathcal{N}(a_t(\theta_t))$ denoting a sufficiently small neighborhood of $a_t(\theta_t)$. The minimization scheme is based on the orthogonal projection of the solution on to the limit cycle with respect to the weighted Euclidean norm at some θ_t. We will derive an exact PDMP for β_t (up to some stopping time) by considering the first derivative

$$\mathcal{G}_0(z, a, \theta) := \frac{\partial}{\partial a}\|z - \Phi(a)\|_\theta^2 = -2\langle z - \Phi(a), \Phi'(a)\rangle_\theta. \tag{3.7}$$

At the minimum

$$\mathcal{G}_0(x_t, \beta_t, \theta_t) = 0. \tag{3.8}$$

We stipulate that the location of the weight must coincide with the location of the minimum, i.e., $\beta_t = \theta_t$, so that β_t must satisfy the implicit equation

$$\mathcal{G}(x_t, \beta_t) := \mathcal{G}_0(x_t, \beta_t, \beta_t) = 0. \tag{3.9}$$

It will be seen that, up to a stopping time τ, there exists a unique continuous solution to the above equation. Define $\mathfrak{M}(z, a) \in \mathbb{R}$ according to

$$\begin{aligned}\mathfrak{M}(z, a) &:= \frac{1}{2}\frac{\partial \mathcal{G}(z, a)}{\partial a} \\ &= \frac{1}{2}\frac{\partial \mathcal{G}_0(z, a, \theta)}{\partial a}\bigg|_{\theta = a} + \frac{1}{2}\frac{\partial \mathcal{G}_0(z, a, \theta)}{\partial \theta}\bigg|_{\theta = a} \\ &= \mathfrak{M}_0 - \langle z - \Phi(a), \Phi''(a)\rangle_a \\ &\quad - \left\langle z - \Phi(a), \frac{d}{da}\left\{\left[P(a)P^\top(a)\right]^{-1}\right\}\Phi'(a)\right\rangle.\end{aligned} \tag{3.10}$$

Assume that initially $\mathfrak{M}(u_0, \beta_0) > 0$. We then seek a PDMP for β_t that holds for all times less than the stopping time τ

$$\tau = \inf\{s \geq 0 : \mathfrak{M}(u_s, \beta_s) = 0\}. \tag{3.11}$$

The implicit function theorem guarantees that a unique continuous β_t exists until this time.

In order to derive the PDMP for β_t, we consider the equation

$$\frac{d\mathcal{G}_t}{dt} \equiv \frac{d\mathcal{G}(x_t, \beta_t)}{dt} = 0, \quad t \in \mathcal{T}, \tag{3.12}$$

with x_t evolving according to the PDMP (2.1). From the definition of $\mathcal{G}(x_t, \beta_t)$, it follows that

$$0 = -2\left\langle \frac{dx_t}{dt}, \Phi'(\beta_t) \right\rangle_{\beta_t} + \left.\frac{\partial \mathcal{G}_t}{\partial a}\right|_{a=\beta_t} \frac{d\beta_t}{dt}, \quad t \in \mathcal{T}. \tag{3.13}$$

Rearranging, we find that the phase β_t evolves according to the exact, but implicit, PDMP

$$\frac{d\beta_t}{dt} = \mathfrak{M}(x_t, \beta_t)^{-1} \langle F_n(x_t), \Phi'(\beta_t)\rangle_{\beta_t}, \tag{3.14}$$

with $n = n_j$ for $t \in (t_j, t_{j+1})$. Finally, recalling that the amplitude term v_t satisfies $\sqrt{\epsilon} v_t = x_t - \Phi_{\beta_t}$, we have

$$\begin{aligned} \sqrt{\epsilon}\frac{dv_t}{dt} &= \frac{dx_t}{dt} - \Phi'(\beta_t)\frac{d\beta_t}{dt} \\ &= F_n(x_t) - \mathfrak{M}(x_t, \beta_t)^{-1} \Phi'(\beta_t) \langle F_n(x_t), \Phi'(\beta_t)\rangle_{\beta_t}. \end{aligned} \tag{3.15}$$

C. Weak noise limit

Equation (3.14) is a rigorous, exact implicit equation for the phase β_t. We can derive an explicit equation for β_t by carrying out a perturbation analysis in the weak noise limit, which we refer to as a *linear phase approximation*. Let $0 < \epsilon \ll 1$ and set $x_t = \Phi(\beta_t)$ on the right-hand side of (3.14), that is, $v_t = 0$. Writing $\beta_t \approx \theta_t$, we have the piecewise deterministic phase equation

$$\begin{aligned} \frac{d\theta_t}{dt} &= Z_n(\theta_t) := \mathfrak{M}_0^{-1} \langle F_n(\Phi(\theta_t)), \Phi'(\theta_t)\rangle_\theta, \\ &= \mathfrak{M}_0^{-1} \langle P(\theta_t)^{-1} F_n(\Phi(\theta_t)), P^{-1}(\theta_t)\Phi'(\theta_t)\rangle, \\ &= \mathfrak{M}_0^{-1} \langle F_n(\Phi(\theta_t)), (P(\theta_t)P(\theta_t)^\top)^{-1}\Phi'(\theta_t)\rangle, \\ &= \langle F_n(\Phi(\theta_t)), R(\theta_t)\rangle, \quad n = n_j \quad \text{for } t \in (t_j, t_{j+1}), \\ &= \omega_0 + \langle F_n(\Phi(\theta_t)) - \bar{F}(\Phi(\theta_t)), R(\theta_t)\rangle, \end{aligned} \tag{3.16}$$

after using $\mathfrak{M}(\Phi(\theta), \theta) = \mathfrak{M}_0$ and Eq. (3.4). The last line follows from the observation

$$\begin{aligned} \langle \bar{F}(\Phi(\theta)), R(\theta)\rangle &= \omega_0 \langle \Phi'(\theta), R(\theta)\rangle \\ &= \omega_0 \mathfrak{M}_0^{-1} \|\Phi'(\theta)\|_\theta^2 = \omega_0. \end{aligned}$$

Hence, a phase reduction of the PDMP (2.1) yields a PDMP for the phase θ_t. Of course, analogous to the phase reduction of SDEs, there are errors due to the fact we have ignored $O(\epsilon)$ terms arising from amplitude-phase coupling, see below. As we show numerically in Sec. IV, this leads to deviations of the phase θ_t from the exact variational phase β_t over $O(1/\epsilon)$ timescales. Finally, note that we could now apply a QSS approximation to the phase PDMP (3.16), which would recover the phase SDE (2.21), at least to leading order in the drift.

D. Coupling to the amplitude v

Although neglecting the coupling between the phase and amplitude dynamics by setting $v_t = 0$ yields a closed equation for the phase, it can lead to imprecision at short and intermediate times. Here, we show that taking into account the amplitude coupling only results in $O(\epsilon)$ contributions to the drift, not $O(\sqrt{\epsilon})$. First, setting

$$\mathfrak{R}(v_t, \beta_t) = \mathfrak{M}(\Phi(\beta_t) + \sqrt{\epsilon} v_t, \beta_t)^{-1},$$

and using Eq. (3.10) gives

$$\begin{aligned} \mathfrak{R}(v_t, \beta_t) = \Bigg(&\mathfrak{M}_0 - \sqrt{\epsilon}\langle v_t, \Phi''(\beta_t)\rangle_{\beta_t} \\ &- \sqrt{\epsilon}\left\langle v_t, \frac{d}{da}\left[P(a)P^\top(a)^{-1}\right]\Big|_{a=\beta_t} \Phi'(\beta_t)\right\rangle\Bigg)^{-1}. \end{aligned}$$

Let us define

$$H_n(v, \theta) = \mathfrak{R}(v, \theta)\langle F_n(\Phi(\theta) + \sqrt{\epsilon} v), \Phi'(\theta)\rangle. \tag{3.17}$$

In the phase equation (3.16), we set $v = 0$ and used

$$H_n(0, \theta) = \langle F_n(\Phi(\theta)), R(\theta)\rangle = \mathcal{H}_n(\theta).$$

Suppose that we now include higher-order terms by Taylor expanding $H_n(v, \theta)$ with respect to v. In particular, consider the first derivative

$$\begin{aligned} \frac{\partial H}{\partial v}(0, \theta) \cdot v &= \sqrt{\epsilon}\,\mathfrak{M}_0^{-1} \langle J_n(\theta) \cdot v, \Phi'(\theta)\rangle_\theta + \sqrt{\epsilon}\,\mathfrak{M}_0^{-2} \langle F_n(\Phi(\theta)), \Phi'(\theta)\rangle_\theta \left[\langle v, \Phi''(\theta)\rangle_\theta + \left\langle v, \frac{d}{da}\left[P(a)P^\top(a)^{-1}\right]\Big|_{a=\theta} \Phi'(\theta)\right\rangle\right] \\ &= \sqrt{\epsilon}\mathfrak{M}_0^{-1} \langle J_n(\theta) \cdot v, \Phi'(\theta)\rangle_\theta + \sqrt{\epsilon}\mathfrak{M}_0^{-2} \langle F_n(\Phi(\theta)), \Phi'(\theta)\rangle_\theta \frac{d}{d\theta}\langle v, \Phi'(\theta)\rangle_\theta, \\ &= \sqrt{\epsilon}\mathfrak{M}_0^{-1} \left[\langle \bar{J}(\theta) \cdot v, \Phi'(\theta)\rangle_\theta + \omega_0 \frac{d}{d\theta}\langle v, \Phi'(\theta)\rangle_\theta\right] \\ &\quad + \sqrt{\epsilon}\mathfrak{M}_0^{-1}\left\langle \left[J_n(\theta) - \bar{J}(\theta)\right] \cdot v, \Phi'(\theta)\right\rangle_\theta + \sqrt{\epsilon}\mathfrak{M}_0^{-2}\left\langle \left[F_n(\Phi(\theta)) - \bar{F}(\Phi(\theta))\right], \Phi'(\theta)\right\rangle_\theta \frac{d}{d\theta}\langle v, \Phi'(\theta)\rangle_\theta, \end{aligned}$$

with $J_{n,jk}(\Phi) \equiv \frac{\partial F_{n,j}}{\partial x_k}|_{x=\Phi}$. We have used $\langle \bar{F}(\Phi(\theta)), \Phi'(\theta)\rangle_\theta = \omega_0 \mathfrak{M}_0$. Next, we observe that

$$\begin{aligned}\langle \bar{J}(\theta)\cdot v, \Phi'(\theta)\rangle_\theta &= \langle P^{-1}(\theta)\bar{J}(\theta)\cdot v, P^{-1}(\theta)\Phi'(\theta)\rangle \\ &= \langle \bar{J}(\theta)\cdot v, \left[P(\theta)P^\top(\theta)\right]^{-1}\Phi'(\theta)\rangle \\ &= \mathfrak{M}_0\langle v, \bar{J}(\theta)^\top \cdot R(\theta)\rangle = -\omega_0\mathfrak{M}_0\langle v, R'(\theta)\rangle \\ &= -\omega_0\left\langle v, \frac{d}{d\theta}\left\{\left[P(\theta)P^\top(\theta)\right]^{-1}\Phi'(\theta)\right\}\right\rangle \\ &= -\omega_0\frac{d}{d\theta}\langle v, \Phi'(\theta)\rangle_\theta,\end{aligned}$$

where we have used Eqs. (3.4) and (2.19). We thus have the modified phase equation

$$\begin{aligned}\frac{d\theta}{dt} &= \omega_0 + \left\langle\left[F_n(\Phi(\theta))) - \bar{F}(\Phi(\theta))\right], R(\theta)\right\rangle \\ &\quad + \sqrt{\epsilon}\mathfrak{M}_0^{-1}\left\langle\left[J_n(\theta) - \bar{J}(\theta)\right]\cdot v, \Phi'(\theta)\right\rangle_\theta \\ &\quad + \sqrt{\epsilon}\mathfrak{M}_0^{-2}\left\langle\left[F_n(\Phi(\theta)) - \bar{F}(\Phi(\theta))\right], \Phi'(\theta)\right\rangle_\theta \\ &\quad \times \frac{d}{d\theta}\langle v, \Phi'(\theta)\rangle_\theta. \end{aligned} \tag{3.18}$$

IV. EXAMPLE: STOCHASTIC MORRIS-LECAR MODEL

Deterministic, conductance-based models of a single neuron such as the Hodgkin-Huxley model have been widely used to understand the dynamical mechanisms underlying membrane excitability.[48] These models assume a large population of ion channels so that their effect on membrane conductance can be averaged. As a result, the average fraction of open ion channels modulates the effective ion conductance, which in turn depends on voltage. It is often convenient to consider a simplified planar model of a neuron, which tracks the membrane voltage v and a recovery variable w that represents the fraction of open potassium channels. The advantage of a two-dimensional model is that one can use phase-plane analysis to develop a geometric picture of neuronal spiking. One well-known example is the Morris-Lecar (ML) model.[49] Although this model was originally developed to model Ca^{2+} spikes in molluscs, it has been widely used to study neural excitability for Na^+ spikes,[48] since it exhibits many of the same bifurcation scenarios as more complex models. The ML model has also been used to investigate subthreshold membrane potential oscillations (STOs) due to persistent Na^+ currents.[28,50] For the sake of illustration, we will consider the latter application in this section, following along similar lines to Ref. 28.

Another advantage of the ML model is that it is straightforward to incorporate intrinsic channel noise.[6,10,12] In order to capture the fluctuations in membrane potential from stochastic switching in voltage-gated ion channels, the resulting model includes both discrete jump processes (to represent the opening and closing of Na^+ ion channels) and a two-dimensional continuous-time piecewise process (to represent the membrane potential and recovery variable w). We thus have an explicit example of a PDMP.

A. Deterministic model

First, consider a deterministic version of the ML model[49] consisting of a persistent sodium current (Na^+), a slow potassium current (K^+), a leak current (L), and an applied current (I_{app}). For simplicity, each ion channel is treated as a two-state system that switches between an open and a closed state—the more detailed subunit structure of ion channels is neglected.[7] The membrane voltage v evolves as

$$\begin{aligned}\frac{dv}{dt} &= a_\infty(v)f_{Na}(v) + wf_K(v) + f_L(v) + I_{\text{app}} \\ \frac{dw}{dt} &= (1-w)\alpha_K(v) - w\beta_K,\end{aligned} \tag{4.1}$$

where w is the K^+ gating variable. It is assumed that Na^+ channels are in quasi-steady state $a_\infty(v)$, thus eliminating Na^+ as a variable. For $i = K, Na, L$, let $f_i = g_i(V_i - v)$, where g_i are ion conductances and V_i are reversal potentials. Opening and closing rates of ion channels depend only on membrane potential v are represented by α and β, respectively, so that

$$a_\infty(v) = \frac{\alpha_{Na}(v)}{\alpha_{Na}(v) + \beta_{Na}(v)}. \tag{4.2}$$

For concreteness, take

$$\alpha_i(v) = \beta_i \exp\left(\frac{2(v - v_{i,1})}{v_{i,2}}\right) \quad i = K, Na, \tag{4.3}$$

with $\beta_i, v_{i,1}, v_{i,2}$ constant. Parameters are chosen such that stable oscillations arise for sufficient values of the applied current via a supercritical Hopf bifurcation [see Fig. 2(a)]. This corresponds well to observed behavior of STOs and is not meant to function as a traditional spiking neuron model. Limit cycles in a traditional spiking model often appear via a subcritical Hopf bifurcation. Figures 2(b) and 2(c) show the phase plane of the deterministic system; here, one can see how oscillations arise in the membrane potential $v(t)$ as the applied current is increased.

B. Stochastic model

The deterministic ML model holds under the assumption that the number of ion channels is very large; thus, the ion channel activation can be approximated by the average ionic currents. However, it is known that channel noise does affect membrane potential fluctuations (and thus neural function).[51] In order to account for ion channel fluctuations, we consider a stochastic version of the ML model,[6,10] in which the number N of Na^+ channels is taken to be relatively small. (For simplicity, we ignore fluctuations in the K^+ channels.) Let $n(t)$ be the number of open Na^+ channels at time t, which means that there are $N - n(t)$ closed channels. The voltage and recovery variables then evolve according to the following PDMP:

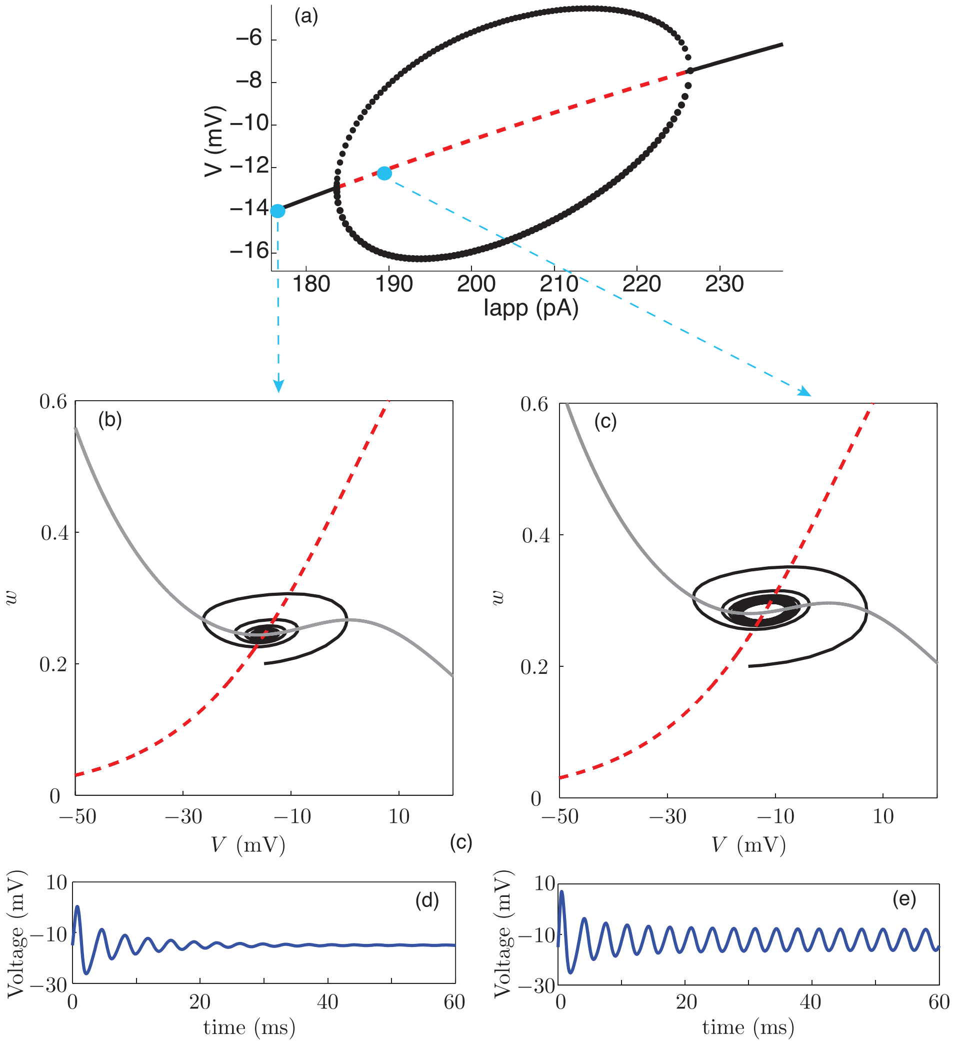

FIG. 2. (a) Bifurcation diagram of the deterministic model. As I_{app} is increased, the system undergoes a supercritical Hopf bifurcation (H) at $I^*_{app} = 183$, which leads to the generation of stable oscillations. The maximum and minimum values of oscillations are plotted as black (solid) curves. Oscillations disappear via another supercritical Hopf bifurcation. (b) and (c) Phase plane diagrams of the deterministic model for (b) $I_{app} = 170$ pA (below the Hopf bifurcation point) and (c) $I_{app} = 190$ pA (above the Hopf bifurcation point). The red (dashed) curve is the w-nullcline and the solid (gray) curve represents the v-nullcline. (d) and (e) Corresponding voltage time courses. *Sodium parameters*: $g_{Na} = 4.4$ mS, $V_{Na} = 55$ mV, $\beta_{Na} = 100\,\text{ms}^{-1}$, $v_{n,1} = -1.2$ mV, and $v_{n,2} = 18$ mV. *Leak parameters*: $g_L = 2$ mS and $V_L = -60$ mV. *Potassium parameters*: $g_K = 8$ mS, $V_K = -84$ mV, $\beta_K = 0.35\,\text{ms}^{-1}$, $v_{k,1} = 2$ mV, and $v_{k,2} = 30$ mV.

$$\begin{aligned}\frac{dv}{dt} &= \frac{n}{N} f_{Na}(v) f_{Na}(v) + w f_K(v) + f_L(v) + I_{app},\\ \frac{dw}{dt} &= (1-w)\alpha_K(v) - w\beta_K.\end{aligned} \tag{4.4}$$

Suppose that individual channels switch between open (O) and closed (C) states via a two-state Markov chain

$$C \underset{\beta_{Na}/\epsilon}{\overset{\alpha_{Na}(v)/\epsilon}{\rightleftharpoons}} O. \tag{4.5}$$

It follows that at the population level, the number of open ion channels evolves according to a birth-death process with

$$\begin{aligned} n \to n-1 \qquad & \omega_n^-(v) = n\beta_{Na},\\ n \to n+1 \qquad & \omega_n^+(v) = (N-n)\alpha_{Na}(v).\end{aligned} \tag{4.6}$$

Note that we have introduced the small parameter ϵ in order to reflect the fact that Na^+ channels open and close much faster than the relaxation dynamics of the system (v, w). The stationary density of the birth-death process is

$$\rho_n(v) = \frac{N!}{n!(N-n)!}\frac{\alpha_{Na}^n(v)\beta_{Na}^{(N-n)}}{(\alpha_{Na}(v)+\beta_{Na})^N}. \tag{4.7}$$

The corresponding CK equation is

$$\begin{aligned}\frac{\partial P_n}{\partial t} &= -\frac{\partial}{\partial v}\left[\left(\frac{n}{N}f_{Na}(v) + w f_K(v) + f_L(v) + I_{app}\right)P_n(v,w,t)\right]\\ &\quad - \frac{\partial}{\partial w}[((1-w)\alpha_K(v) - w\beta_K)P_n(v,w,t)]\\ &\quad + \frac{1}{\epsilon}\left(\omega_{n-1}^+(v)P_{n-1}(v,w,t) + \omega_{n+1}^-(v)P_{n+1}(v,w,t)\right)\\ &\quad - \frac{1}{\epsilon}\left((\omega_n^+(v) + \omega_n^-(v))P_n(v,w,t)\right).\end{aligned} \tag{4.8}$$

Comparison with the general CK equation (2.6) shows that $x = (v, w)$, $\nabla = (\partial_v, \partial_w)^\top$

$$\begin{aligned}F_n(v,w) &:= \begin{pmatrix} f_n(v,w)\\ f(v,w)\end{pmatrix}\\ &= \begin{pmatrix} n f_{Na}(v)/N + w f_K(v) + f_L(v) + I_{app}\\ (1-w)\alpha_K(v) - w\beta_K\end{pmatrix},\end{aligned}$$

and **A** is the tridiagonal generator matrix of the birth-death process.

In Figs. 3 and 4, we show results of numerical simulations for $N = 10, \epsilon = 0.01$ and $N = 10, \epsilon = 0.001$, respectively. In both figures, we compare solutions of the explicit phase equation (3.16) with the exact phase defined using the

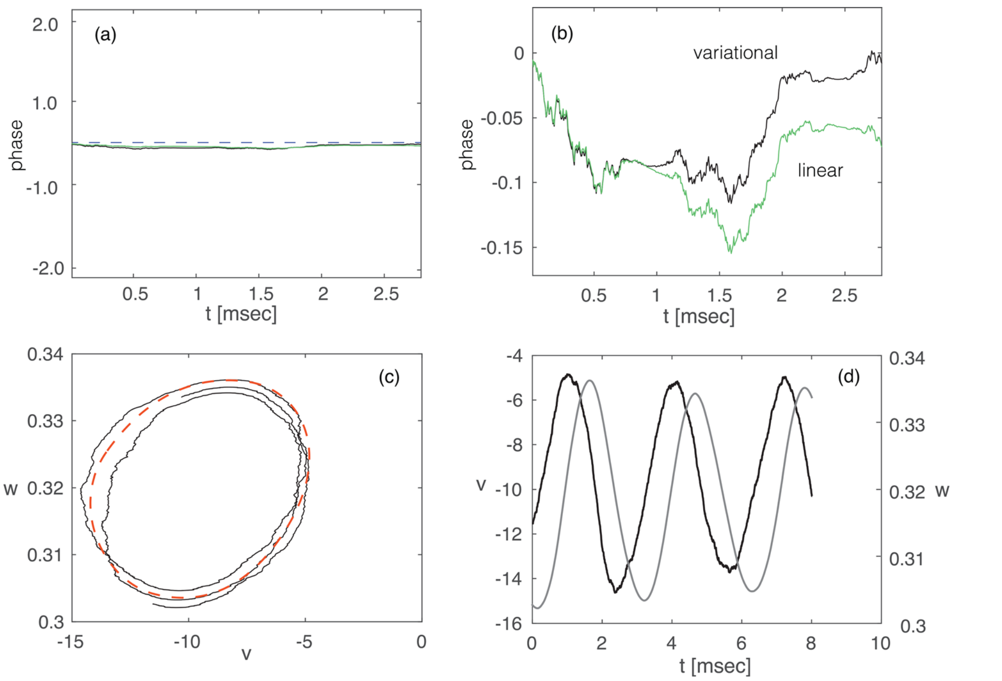

FIG. 3. We simulate the stochastic Morris-Lecar model with $N = 10$ and $\epsilon = 0.01$. (a) and (b) Plot of the linearized phase $\theta_t - t\omega_0$ in green, and the exact variational phase [satisfying (3.9)] $\beta_t - t\omega_0$ in black. On the scale $[-\pi, \pi]$, the two phases are in strong agreement. However, zooming in one can see the phases slowly drift apart as noise accumulates. The diffusive nature of the drift in both phases can be clearly seen, with the typical deviation of the phase from $\omega_0 t$ increasing in time. (b) Stochastic trajectory around limit cycle (dashed curve) in the v and w-plane. The stable attractor of the deterministic limit cycle is quite large, which is why the system can tolerate quite substantial stochastic perturbations. (c) and (d) Corresponding time variations in v (black) and w (gray).

variational principle [see Eq. (3.14)]. We also show the sample trajectories for (v, w). It can be seen that initially the phases are very close, and then very slowly drift apart as noise accumulates. The diffusive nature of the drift in both phases can be clearly seen, with the typical deviation of the phase from $\omega_0 t$ increasing in time.

V. BOUNDING THE NORM OF THE AMPLITUDE TERM

In this section, we obtain a bound for the probability of the difference in amplitude exceeding a certain threshold. That is, we show that there are positive constants C, a_0, such that for all $a \leq a_0$ and a_0 sufficiently small

$$\mathbb{P}(\tau_a \leq T) \leq T \exp\left(-\frac{Ca}{\epsilon}\right), \tag{5.1}$$

where

$$\tau_a = \inf\{t : x_t \notin \mathcal{U}_a\}, \quad \mathcal{U}_a = \{u \in \mathbb{R}^d : \|u - \Phi(\alpha)\|_\alpha \leq a\},$$

and in the above α is the variational phase of u, satisfying $\mathcal{G}(u, \alpha, \alpha) = 0$ [as in (3.9)]. We assume that initially $\|x_t - \Phi(\beta_0)\|_{\beta_0} \leq a/2$. Here, $\mathbb{P}(\tau_a \leq T)$ is the probability that x_t leaves the neighborhood $\mathcal{U}_a$ of the limit cycle over a time interval of length T. Note that a_0 is independent of ϵ but depends on the rate b of attraction to the limit cycle.

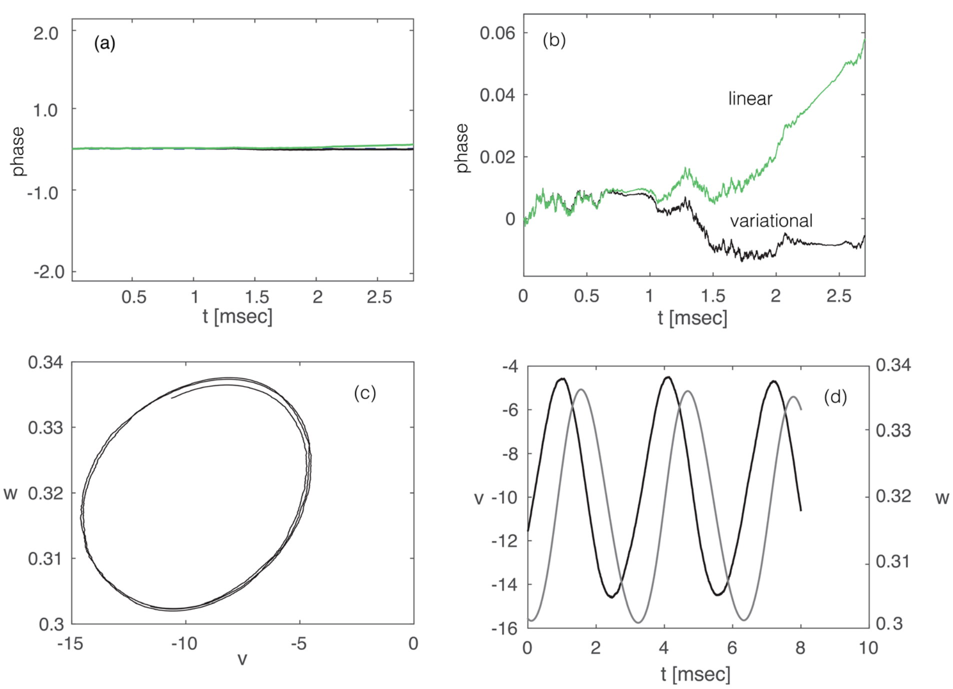

FIG. 4. Same as Fig. 3 except that $\epsilon = 0.001$.

We proceed by first deriving an exact PDMP for the dynamics of $\|x_t - \Phi(\beta_t)\|_{\beta_t}$ (Sec. V A). Then in Sec. V B, we obtain a bound for the probability that the maximum of $\|x_t - \Phi(\beta_t)\|_{\beta_t}$ between successive jumps exceeds $\eta a/2$ with $\eta < 1$. (For concreteness, we take $\eta = 1/4$.) This yields a bound with the same asymptotic order as the right-hand side of Eq. (5.1). However, it is still possible for x_t to leave the domain $\mathcal{U}_a$ due to the accumulative effects of multiple jumps. In Sec. V C and Appendixes C-E, we obtain exponential bounds on the probability of this occurring, which depend on both b and a, thus establishing that if a is sufficiently small, then Eq. (5.1) holds. The bounds on the accumulative growth of the amplitude are derived by decomposing the growth of $\|x_t - \Phi(\beta_t)\|_{\beta_t}$ into the sum of several terms, which we bound individually. More precisely, over the time interval $[t_j, t_{j+1}]$, we take a linear (in the time difference $t_{j+1} - t_j$) approximation to the dynamics of x_t. To leading order, the dynamics of x_t decomposes into the sum of a deterministic part plus a piecewise-constant stochastic part. The deterministic part is stabilizing once $\|x_t - \Phi(\beta_t)\|_{\beta_t} \geq a/2$, due to the assumed linear stability of the limit cycle. We then show that over timescales of $O(b^{-1})$, if the fluctuations remain below $O(a)$, then they will always be dominated by the stabilizing effect of the deterministic component.

For ease of exposition, we take the generator $\mathbf{A}$ of the discrete Markov chain to be independent of x. However, it is possible to extend the analysis to the case of x-dependent rates, as in the case of the Morris-Lecar model.

A. Derivation of PDMP for norm of amplitude term

Let

$$w_t = P(\beta_t)^{-1}(x_t - \Phi(\beta_t)). \tag{5.2}$$

We are going to see that w_t decays towards the limit cycle (to leading order). This is a key reason why we chose the weighted norm $\|\cdot\|_{\beta_t}$ to define the phase. Differentiating with respect to t and using Eq. (B4) gives

$$\frac{dw_t}{dt} = \frac{\dot{\beta}_t}{\omega_0}\Big\{\mathcal{S}w_t - P(\beta_t)^{-1}J(\beta_t)(x_t - \Phi(\beta_t)) \\ -\omega_0 P(\beta_t)^{-1}\Phi'(\beta_t)\Big\} + P(\beta_t)^{-1}F_n(x_t), \tag{5.3}$$

where $\mathcal{S} = \mathrm{diag}(\nu_1, \ldots, \nu_d)$ with ν_j the Floquet characteristic exponents (see Sec. III). Combining this with Eq. (3.14) shows that

$$\frac{dw_t}{dt} = \mathcal{S}w_t + \mathfrak{F}(\beta_t, x_t) + \mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t),$$

where

$$\mathfrak{F}(\beta_t, x_t) = P(\beta_t)^{-1}\bar{F}(x_t) \\ - \mathfrak{M}(x_t, \beta_t)^{-1}\omega_0^{-1}\langle \bar{F}(x_t), \Phi'(\beta_t)\rangle P(\beta_t)^{-1} \\ \times \{\Phi'(\beta_t) + J(\beta_t)(x_t - \Phi(\beta_t))\} \\ + \{\mathfrak{M}(x_t, \beta_t)^{-1}\omega_0^{-1}\langle \bar{F}(x_t), \Phi'(\beta_t)\rangle - 1\}\mathcal{S}w_t$$

and

$$\mathfrak{G}(K_n(x_t), \beta_t, x_t) = \omega_0^{-1}\mathfrak{M}(x_t, \beta_t)^{-1}\langle K_n(x_t), \Phi'(\beta_t)\rangle_{\beta_t} \\ \times \{\mathcal{S}w_t - P(\beta_t)^{-1}J(\beta_t)(x_t - \Phi(\beta_t)) \\ - \omega_0 P(\beta_t)^{-1}\Phi'(\beta_t)\} + P(\beta_t)^{-1}K_n(x, t).$$

This means that

$$\frac{d\|w_t\|^2}{dt} = 2\langle w_t, \mathcal{S}w_t + \mathfrak{F}(\beta_t, x_t) \\ +\mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t)\rangle.$$

Taking square roots,

$$\frac{d\|w_t\|}{dt} = \|w_t\|^{-1}\langle w_t, \mathcal{S}w_t + \mathfrak{F}(\beta_t, x_t) \\ +\mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t)\rangle.$$

It should be noted that the above PDMP is well-defined in the limit as $\|w_t\| \to 0$, since by the Cauchy-Schwarz Inequality

$$|\langle w_t, \mathcal{S}w_t + \mathfrak{F}(\beta_t, x_t) + \mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t)\rangle| \\ \leq \|w_t\|\|-\mathcal{S}w_t + \mathfrak{F}(\beta_t, x_t) + \mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t)\|.$$

Now, by definition of β_t, $\langle w_t, P(\beta_t)^{-1}\Phi'(\beta_t)\rangle = 0$. Since, by assumption, $\langle u, \mathcal{S}u\rangle \leq -(b/\omega_0)\|u\|^2$ for all vectors u such that $\langle u, P(\alpha)^{-1}\Phi'(\alpha)\rangle = 0$ (where α is the variational phase of u), we find that

$$\frac{d\|w_t\|}{dt} \leq -b\|w_t\| + \|w_t\|^{-1}\langle w_t, \mathfrak{F}(\beta_t, x_t) \\ +\mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t)\rangle.$$

B. Bounding fluctuations in $\|w_t\|$ between successive jumps

Our first step is to bound the fluctuations of $\|w_t\|$ between successive jumps, which occur at times t_j, $j \geq 0$. Let

$$C_L = \sup_{x\in\mathcal{U}_a, n\in\Gamma}\{\|w\|^{-1}|\langle w, \mathcal{S}w + \mathfrak{F}(\beta, x) \\ +\mathfrak{G}(F_n(x) - \bar{F}(x), \beta, x)\rangle|\},$$

where β is the variational phase corresponding to x, and w is the remainder term. It is straightforward to show that $C_L < \infty$. Let

$$\hat{\tau}_a = \inf\left\{t_j : t_{j+1} - t_j \geq \frac{a}{8C_L}\right\}. \tag{5.4}$$

It follows from this definition that for all $t_j \leq \tau_a$ and $t_{j-1} \leq \hat{\tau}$,

$$\sup_{t\in[t_{j-1}, t_j]} |\|w_t\| - \|w_{t_{j-1}}\|| \leq \frac{a}{8}. \tag{5.5}$$

Now since the length of the interval between successive jumps is distributed in a Poissonian manner, we have the following bound for the conditional probability:

$$\mathbb{P}\left(t_{j+1}-t_j \geq \frac{a}{8C_L}\middle| n(t_j)=m\right) \leq \exp\left(-\frac{a\lambda_m}{8C_L\epsilon}\right),$$

where λ_m is the rate of the exponential density of switching times form the discrete state m [see Eq. (2.3)]. By assumption, $\inf_{m\in\Gamma}\lambda_m > 0$. We thus see that there exists a positive constant C such that the conditional probability has the uniform bound

$$\mathbb{P}\left(t_{j+1}-t_j \geq \frac{a}{8C_L}\middle| n(t_j)=m\right) \leq \exp\left(-\frac{Ca}{\epsilon}\right). \quad (5.6)$$

Now define $\mathcal{J}$ to be the typical number of jumps that occur over the time interval T, i.e.,

$$\mathcal{J} = \left\lfloor \frac{T}{\epsilon}\sum_{m\in\Gamma}\rho_m\lambda_m \right\rfloor. \quad (5.7)$$

We thus find that

$$\mathbb{P}\left(\text{For some } j \leq \mathcal{J},\ t_{j+1}-t_j \geq \frac{a}{8C_L}\right) \leq \mathcal{J}\exp\left(-\frac{Ca}{\epsilon}\right) \leq T\exp\left(-\frac{Ca}{2\epsilon}\right), \quad (5.8)$$

for ϵ sufficiently small. (We have absorbed other constant factors into C.) We have thus shown that

$$\mathbb{P}(\hat{\tau}_a \leq T) \leq T\exp\left(-\frac{Ca}{\epsilon}\right). \quad (5.9)$$

In order to prove Eq. (5.1), we can now proceed by determining bounds for $\mathbb{P}(\tau_a \leq T)$ given that $T \leq \hat{\tau}_a$ and then show that these bounds are weaker than the right-hand side of Eq. (5.1) when a is sufficiently small. In other words, having looked at changes in $\|w_t\|$ between successive jumps, we turn to the accumulative changes in $\|w_{t_j}\|$ over a sequence of jumps.

C. Bounding the probability of x_t leaving $\mathcal{U}_a$

In the following, we assume that $t\in[0,T]$ with $T \leq \hat{\tau}_a$. We will show that when $x_t \in \mathcal{U}_a$, the deterministic component of the dynamics of $\|w_t\|$ is dominated by the first term i.e., $\|w_t\|^{-1}\langle w_t, \mathcal{S}w_t\rangle$, which is stabilizing. Our analysis will centre on the times when $\|w_t\| \in [\frac{a}{2}, a]$. We can do this because, by the intermediate value theorem, if $x_t \notin \mathcal{U}_a$, then immediately prior to leaving $\mathcal{U}_a$, it must be such that $\|w_t\| \in [\frac{a}{2}, a]$. The reason why we insist on a lower bound for $\|w_t\|$ of $a/2$ is that we require that, with very high probability, the linear decay is sufficiently great to dominate the fluctuations due to the switching. It should be noted that our choice of $a/2$ for the lower bound is not particularly necessary: one could have for example chosen a/X for any real X and obtained comparable results.

Let

$$u_j = (t_{j+1}-t_j)\|w_t\|^{-1}\langle w_{t_j}, \mathfrak{G}(F_{n_j}(x_{t_j}) - \bar{F}(x_{t_j}), \beta_{t_j}, x_{t_j})\rangle.$$

We make the decomposition

$$\begin{aligned}\|w_{t_{k+m}}\| - \|w_{t_k}\| &= \sum_{j=k}^{k+m}(u_j + (\delta t_j)^2 C_j) + \int_{t_k}^{t_{k+m}}\|w_t\|^{-1}(\langle w_t, \mathcal{S}w_t\rangle + \langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle)dt \\ &\leq \sum_{j=k}^{k+m}(u_j + (\delta t_j)^2 C_j) + \int_{t_k}^{t_{k+m}}(-b\|w_t\| + \|w_t\|^{-1}\langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle)dt.\end{aligned} \quad (5.10)$$

Here, C_j is by definition the remainder term for the switching part of $\|w_t\|$. Through Taylor's Theorem

$$C_j = \frac{1}{2}\frac{\partial}{\partial t}\left\{\|w_t\|^{-1}\langle w_t, \mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t)\rangle\right\}\bigg|_{t=\bar{t}_j},$$

for some $\bar{t}_j \in [t_j, t_{j+1}]$. Now let

$$\bar{C}_L = \frac{1}{2}\sup_{x_t\in\mathcal{U}_a}\left|\frac{\partial}{\partial t}\left\{\|w_t\|^{-1}\langle w_t, \mathfrak{G}(F_n(x_t) - \bar{F}(x_t), \beta_t, x_t)\rangle\right\}\right|.$$

Now as long as for all $t \leq t_{k+m}$, $x_t \in \mathcal{U}_a$,

$$|C_j| \leq \bar{C}_L.$$

This means that, as long as $t_{k+m} \leq \tau_a$

$$\|w_{t_{k+m}}\| - \|w_{t_k}\| \leq \sum_{j=k}^{k+m}(u_j + (\delta t_j)^2\bar{C}_L) + \int_{t_k}^{t_{k+m}}(-b\|w_t\| + \|w_t\|^{-1}\langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle)dt. \quad (5.11)$$

In Appendixes C-E, we obtain bounds on each of the individual terms on the right-hand side of Eq. (5.11), and thus establish that if $T \leq \hat{\tau}_a$ then

$$\mathbb{P}(\tau_a \leq T) \leq O(T\exp(-\tilde{C}ba^2\epsilon^{-1}), T\exp(-\hat{C}b\epsilon^{-1})),$$

for constants $\tilde{C}, \hat{C}$. These bounds will be smaller than the bound of Eq. (5.1) if b is sufficiently large.

VI. DISCUSSION

In summary, we have presented the first systematic phase reduction of stochastic hybrid oscillators (PDMPs that support a stable limit cycle in the adiabatic limit). In particular, we adapted a variational principle previously developed for SDEs[39] in order to derive an exact stochastic phase equation, which takes the form of an implicit PDMP. Moreover, we showed how the latter can be converted to an explicit PDMP for the phase by performing a perturbation expansion in ϵ (linear phase approximation), see Eq. (3.16), and that the phase

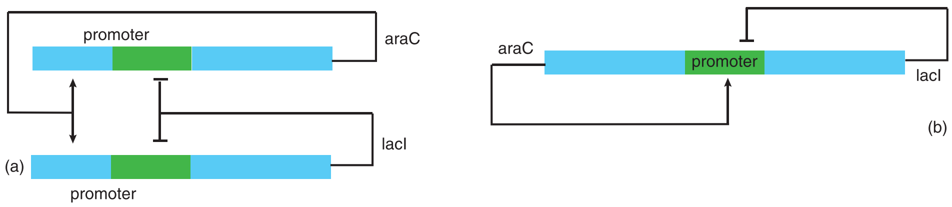

FIG. 5. Dual-feedback gene regulatory network. (a) Two promoter sites. (b) One promoter site.

decouples from the amplitude to leading order. The phase equation (3.16) is in a form consistent with the idea that in the fast switching regime (small ϵ), one can treat $F_n(x) - \bar{F}(x)$ as a small stochastic perturbation of the limit cycle, and thus determine the phase dynamics by projecting the perturbation on to the phase resetting curve $R(\theta)$. Although the linear phase approximation yields an accurate approximation of the exact variational phase over a single cycle, it slowly diffuses away from the latter over longer time-scales due to the effects of higher order terms that couple the amplitude and phase.

More significantly, as with SDEs, the variational formulation itself ultimately breaks down over sufficiently long time-scales, since there is a non-zero probability of leaving a bounded neighborhood of the limit cycle, and the notion of phase no longer makes sense. Hence, it is important to obtain estimates for the probability of escape over a time interval of length T, and how it depends on ϵ, the size a of the neighborhood, and the rate b of attraction to the limit cycle. In light of this, we used probabilistic methods to establish that for a constant C, and all $a \leq a_0$ (a_0 being a constant independent of ϵ), the probability that the time to leave an $O(a)$ neighborhood of the limit cycle is less than T scales as $T \exp(-Ca/\epsilon)$. This result differs significantly from the corresponding bound for SDEs. More precisely, our analysis in Ref. 39 demonstrated that the SDE system stays close to the limit cycle for a very long time if $b\epsilon^{-1}$ is very large: i.e., as long as the rate of attraction to the limit cycle dominates the magnitude of the noise. However, by contrast, with a switching PDMP oscillator, if $b\epsilon^{-1}$ is large, but $\epsilon \geq O(1)$, then the oscillator will in most cases leave any neighborhood of the limit cycle relatively quickly. This is because if $\epsilon \geq O(1)$, then it will typically avoid switching for times of $O(1)$ or greater, and so the system will not 'feel' the stabilizing effect of the averaged system and over this time period can leave the attracting neighborhood of the limit cycle.

Having established a framework for deriving phase equations for stochastic hybrid oscillators, it should now be possible to investigate the synchronization of populations of uncoupled hybrid oscillators subject to common noise. Such noise could either be due to some common external fluctuating input, such as I_{app} in the Morris-Lecar model, or a randomly switching environment in which the discrete variable $N(t)$ is common to all the oscillators. Moreover, the probabilistic approach used to derive exponential bounds on the probability of large transverse fluctuations can be extended to obtain precise bounds on the probability of two synchronized oscillators desynchronizing, and conditions under which two oscillators never desynchronize.

Finally, although we have illustrated our theory using the example of the stochastic Morris-Lecar model of a point neuron with stochastic ion channels, there are several other potential application domains. One notable example is a gene regulatory network with dual feedback, which arises in experimental synthetic biology.[52] This consists of two genes, one whose protein (araC) acts as an activator of both genes and one whose protein (lacI) acts as a repressor of both genes, see Fig. 5(a). This engineered network generates robust oscillations in *Escherichia coli*. Moreover, mathematical modeling of the network has shown that oscillations occur in both the adiabatic and nonadiabatic regimes.[26] One could also consider a simplified version of the model, by taking the pair of genes to share a single promoter site that can be occupied by either activator proteins or repressor proteins but not both, see Fig. 5(b)—the full model has two promoter sites per gene. In either case, if the number of protein molecules is sufficiently large, then the stochastic dynamics evolves according to a PDMP in which protein numbers are the continuous variables, whereas as the states of the promoters are the discrete switching variables.

ACKNOWLEDGMENTS

P.C.B. and J.N.M. were supported by the National Science Foundation (Grant No. DMS-1613048).

APPENDIX A: QSS REDUCTION

The basic steps of the QSS reduction are as follows:

(a) Decompose the probability density as

$$p_n(x,t) = C(x,t)\rho_n(x) + \epsilon w_n(x,t),$$

where $\sum_n p_n(x,t) = C(x,t)$ and $\sum_n w_n(x,t) = 0$. Substituting into Eq. (2.5) yields

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

Summing both sides with respect to n then gives

$$\frac{\partial C}{\partial t} = -\nabla\cdot\left[\bar{F}(x)C\right] - \epsilon\sum_{n\in\Gamma}\nabla\cdot[F_n(x)w_n]. \quad \text{(A1)}$$

(b) Using the equation for C and the fact that $\sum_{m\in\Gamma} A_{nm}(x)\rho_m(x) = 0$, we have

$$\epsilon\frac{\partial w_n}{\partial t} = \sum_{m\in\Gamma} A_{nm}(x)w_m - \nabla\cdot[F_n(x)\rho_n(x)C] + \rho_n(x)\nabla\cdot\left[\bar{F}(x)C\right] - \epsilon\left[\nabla\cdot(F_n(x)\omega_n) - \rho_n(x)\sum_{m\in\Gamma}\nabla\cdot[F_m(x)w_m]\right].$$

(c) Introduce the asymptotic expansion

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

and collect $O(1)$ terms

$$\sum_{m\in\Gamma} A_{nm}(x) w_m^{(0)} = \nabla\cdot[\rho_n(x)F_n(x)C(x,t)] - \rho_n(x)\nabla\cdot\left[\bar{F}(x)C\right].$$

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

$$w_m^{(0)}(x) = \sum_{n\in\Gamma} A^{\dagger}_{mn}(x)\left(\nabla\cdot[\rho_n(x)F_n(x)C(x,t)] - \rho_n(x)\nabla\cdot\left[\bar{F}(x)C\right]\right). \tag{A2}$$

where $\mathbf{A}^{\dagger}$ is the pseudo-inverse of the generator $\mathbf{A}$.

(d) Combining Eqs. (A2) and (A1) shows that C evolves according to the Fokker-Planck (FP) equation

$$\frac{\partial C}{\partial t} = -\nabla\cdot\left[\bar{F}(x)C\right] - \epsilon\sum_{n,m\in\Gamma} A^{\dagger}_{nm}\rho_m\nabla\cdot\left(F_n(x)\nabla\cdot\left[F_m(x)-\bar{F}(x)\right]C\right]\right). \tag{A3}$$

This can be converted to an Ito FP according to

$$\frac{\partial C}{\partial t} = -\nabla\cdot\left[\bar{F}(x)C\right] + \epsilon\nabla\cdot\sum_{n,m}\{(\rho_n F_n)\nabla\cdot(F_m A^{\dagger}_{mn}) - \bar{F}\nabla\cdot(F_m A^{\dagger}_{mn}\rho_n)\}C + \epsilon\sum_{i,j=1}^{d}\frac{\partial^2 D_{ij}(x)C}{\partial x_i\partial x_j}, \tag{A4}$$

where

$$D_{ij}(x) = \sum_{m,n\in\Gamma} F_{m,i}(x)A^{\dagger}_{mn}(x)\rho_n(x)\left[\bar{F}_j(x) - F_{n,j}(x)\right]. \tag{A5}$$

Using the fact that $\sum_m A^{\dagger}_{mn}=0$ and dropping the $O(\epsilon)$ correction to the drift finally yields Eq. (2.10b). Note that one typically has to determine the pseudo-inverse of $\mathbf{A}$ numerically.

For the sake of illustration, we write down the Ito FP equation for $C(v,w,t)=\sum_{n=0}^{N} p_n(v,w,t)$ in the case of the stochastic Morris-Lecar model introduced in Sec. IV B (see also Ref. 28)

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial v}[f_n(v,w)C] - \frac{\partial}{\partial w}[f(v,w)C] - \epsilon\frac{\partial}{\partial v}[\mathcal{V}(v,w)C] + \epsilon\frac{\partial^2 D(v)C}{\partial v^2}, \tag{A6}$$

with

$$\mathcal{V} = \sum_{m,n}\left(\bar{f}(v,w)\frac{\partial}{\partial v}(\rho_n(v)A^{\dagger}_{mn}(v)f_m(v,w) - \rho_n(v)f_n(v,w)\frac{\partial}{\partial v}(A^{\dagger}_{mn}(v)f_m(v,w))\right) \tag{A7a}$$

and

$$\begin{aligned} D &= \sum_{m,n}\left[f_m(v,w) - \bar{f}(v,w))\right]A^{\dagger}_{mn}(v)\rho_n(v)\left[\bar{f}(v,w) - f_n(v,w)\right] \\ &= \sum_{m,n}\left[\frac{m-\langle m\rangle}{N}f_{Na}(v)\right]A^{\dagger}_{mn}(v)\rho_n(v)\left[\frac{\langle n\rangle - n}{N}f_{Na}(v)\right] \\ &= \frac{1}{N}f_{Na}(v)^2 a_{\infty}(v)[1-a_{\infty}(v)]^2. \end{aligned} \tag{A7b}$$

The last line follows from a calculation in Ref. 6.

APPENDIX B: ADJOINT EQUATION FOR THE PHASE RESETTING CURVE

Suppose that $R(\theta)$ is related to the tangent vector $\Phi'(\theta)$ according to Eq. (3.4). We will show that $R(\theta)$ then satisfies the adjoint Eq. (2.19) for the PRC. Differentiating both sides of Eq. (3.4) with respect to θ, we have

$$\mathfrak{M}'P^{\top}R + \mathfrak{M}P^{T}R' + \mathfrak{M}(P^{\top})'R = P^{-1}\Phi'' + (P^{-1})'\Phi', \tag{B1}$$

with

$$\mathfrak{M}' = 2\langle P^{-1}\Phi'' + (P^{-1})'\Phi', P^{-1}\Phi'\rangle.$$

Next, differentiating Eq. (3.3) gives

$$\omega_0 P'(\theta) = \bar{J}(\theta)P(\theta) - P(\theta)\mathcal{S}, \tag{B2}$$

where again $\mathcal{S}=\text{diag}(\nu_1,\ldots,\nu_d)$ with ν_j the Floquet characteristic exponents, which implies that

$$\omega_0(P^{\top}(\theta))' = P^{\top}(\theta)\bar{J}^{\top}(\theta) - \mathcal{S}P^{\top}(\theta) \tag{B3}$$

and

$$\omega_0(P^{-1}(\theta))' = -P^{-1}(\theta)\bar{J}(\theta) + \mathcal{S}P^{-1}(\theta). \tag{B4}$$

We have used the fact that $\mathcal{S}$ is a diagonal matrix and $P^{-1}P' + (P^{-1})'P = 0$ for any square matrix. Substituting these identities in Eq. (B1) yields

$$\mathfrak{M}'P^{\top}R + \mathfrak{M}P^{T}(R' + \omega_0^{-1}\bar{J}^{\top}R) - \omega_0^{-1}\mathfrak{M}\mathcal{S}P^{\top}R = P^{-1}\left[\Phi'' - \omega_0^{-1}\bar{J}\Phi'\right] + \omega_0^{-1}\mathcal{S}P^{-1}\Phi'$$

and

$$\mathfrak{M}' = \langle P^{-1}\left[\Phi'' - \omega_0^{-1}\bar{J}\Phi'\right] + \omega_0^{-1}\mathcal{S}P^{-1}\Phi', P^{-1}\Phi'\rangle.$$

Now note that Φ' satisfies Eq. (2.15) and $\mathcal{S}P^{-1}\Phi'=0$. The latter follows from the condition $P(\theta)^{-1}\Phi'(\theta)=\mathbf{e}$ and $\mathcal{S}\mathbf{e}=\nu_1=0$. It also holds that $\mathfrak{M}'(\theta)=0$. (In fact, for the specific choice of $P(\theta)$, we have $\mathfrak{M}(\theta)=\mathfrak{M}_0=c_0^2\langle\mathbf{e},\mathbf{e}\rangle = c_0^2$.) Finally, from the definition of $R(\theta)$, Eq. (3.4), we deduce that $\mathcal{S}P^{\top}(\theta)R(\theta)=0$ and hence

$$\mathfrak{M}_0 P^{T}(R' + \omega_0^{-1}\bar{J}^{\top}R) = 0. \tag{B5}$$

Since $P^{T}(\theta)$ is non-singular for all θ, R satisfies the adjoint Eq. (2.19) together with the normalization condition (2.20). Hence, $R(\theta)$ can be identified as the classical PRC.[46,47]

APPENDIX C: BOUNDING THE PROBABILITY $\mathbb{P}(\tau_a \leq T)$ (PART I)

In this appendix, we set up the basic framework for deriving bounds on $\mathbb{P}(\tau_a \leq T)$ given that $T \leq \hat{\tau}_a$. The actual bounds are derived in Appendixes D and E. For a positive constant C (that can be inferred from the following analysis), let $\mathfrak{k} = \lfloor \exp(Cb\epsilon^{-1}) \rfloor \in \mathbb{Z}^+$ and define $T = \mathfrak{k}/2b$. (Here, the floor function $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.) For the constant M_b defined further below in (C2), we outline a set of events $\{\mathcal{A}_k^m\}_{1\leq k\leq \mathfrak{k}, 1\leq m\leq M_b}$ and $\{\mathcal{B}_k\}_{1\leq k\leq \mathfrak{k}}$, such that if they all hold, and $\hat{\tau}_a \geq T$, then necessarily

$$\tau_a \geq T. \tag{C1}$$

This will mean that, if we can show that for some positive constant C_2

$$\mathbb{P}((\mathcal{A}_k^m)^c) \leq \exp(-C_2 ba\epsilon^{-1}),$$
$$\mathbb{P}(\mathcal{B}_k^c) \leq \exp(-C_2 b\epsilon^{-1}),$$

then

$$\mathbb{P}(\tau_a \leq T) \leq \sum_{k=1}^{\mathfrak{k}} \left[\mathbb{P}(\mathcal{B}_k^c) + \sum_{m=1}^{M_b} \mathbb{P}((\mathcal{A}_k^m)^c) \right],$$
$$\leq O(\epsilon^{-1} T \exp(-C_2 ba\epsilon^{-1}), T\exp(-C_2 b\epsilon^{-1})).$$

(Here, $\mathcal{A}^c$ is the complementary set of $\mathcal{A}$ so that $\mathbb{P}(\mathcal{A}^c)$ is the probability that the event $\mathcal{A}$ has not occurred.)

We now define the events $\mathcal{A}_k^m$ and $\mathcal{B}_k$, and afterwards we will explain why $\cup_{k=1}^{\mathfrak{k}} \cup_{m=1}^{M_b} \mathcal{A}_k^m$ and $\cup_{k=1}^{\mathfrak{k}} \mathcal{B}_k$ ensures that $\tau_a \geq T$. We are going to define $\mathcal{A}_k^m$ to be a set of events that hold over timescales of $O(b^{-1})$: over this timescale, the stabilizing decay due to the term $\int_t^{t+b^{-1}} -b\|w_s\| ds$ dominates the typical fluctuations due to the switching. Thus, we define M_b to be the typical number of jumps that occur over time intervals of size b^{-1}, i.e.,

$$M_b = \left\lfloor \frac{1}{b\epsilon} \sum_{m\in\Gamma} \rho_m \lambda_m \right\rfloor. \tag{C2}$$

We define $\mathcal{B}_k$ to be the set of events

$$t_{k+M_b} - t_k \geq \frac{1}{2b}, \tag{C3a}$$

$$\bar{C}_L \sum_{j=k}^{k+M_b} (\delta t_j)^2 \leq \frac{a}{32}, \tag{C3b}$$

$$\sup_{x_t\in\mathcal{U}_a : \|w_t\| \in \left[\frac{a}{2}, a\right]} \|w_t\|^{-1} \langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle \leq \frac{ba}{16}, \tag{C3c}$$

and we define $\mathcal{A}_k^m$ to be the set of events

$$\|w_{t_k}\| \in \left[\frac{a}{2}, \frac{5a}{8}\right], \tag{C4a}$$

$$\|w_{t_l}\| \geq \frac{a}{2} \quad \text{for all } \ k \leq l \leq k+m, \tag{C4b}$$

$$\sum_{j=k}^{k+m} u_j \leq \frac{3a}{16}. \tag{C4c}$$

We now explain why the union of the above events ensures that (C1) holds. It suffices to show that if $\cup_{k=1}^{\mathfrak{k}} \mathcal{B}_k$ holds, but there exists $t \in [0, T\wedge \hat{\tau}_a]$ such that

$$\|w_t\| \geq a, \tag{C5}$$

then necessarily there must exist $k \leq \mathfrak{k}$ and $m \leq M_b$ such that $(\mathcal{A}_k^m)^c$ holds.

Now, if (C5) holds, then it follows from (5.5) that there must exist some $t_J \leq t$ such that

$$\|w_{t_J}\| \geq \frac{7a}{8}. \tag{C6}$$

Let $k = \max\left\{ j < J : \|w_{t_j}\| \in \left[\frac{a}{2}, \frac{5a}{8}\right]\right\}$. k exists because successive increments in $\|w_{t_j}\|$ cannot differ by more than $a/8$, thanks to (5.5). Suppose first that $J - k > M_b$. Since, by assumption, $\mathcal{B}_k$ holds, it must be that $t_J - t_k > \frac{1}{2b}$. This means that, writing $l = k + M_b$

$$\int_{t_k}^{t_l} (-b\|w_t\| + \|w_t\|^{-1}\langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle) dt$$
$$< \frac{1}{2b}\left\{ -b \inf_{t\in[t_k,t_l]} \|w_t\| + \sup_{t\in[t_k,t_l]} \|w_t\|^{-1}\langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle \right\}$$
$$\leq \frac{1}{2b}\left\{ -\frac{ba}{2} + \frac{ba}{16}\right\} = \frac{-7a}{32}.$$

It thus follows from (5.10) that

$$\|w_{t_l}\| < \|w_{t_k}\| + \sum_{j=k}^{l}\left(u_j + (\delta t_j)^2 \bar{C}_L\right) - \frac{7a}{32}$$
$$\leq \frac{5a}{8} + \sum_{j=k}^{l} u_j + \frac{a}{32} - \frac{7a}{32}, \tag{C7}$$

using the definition of $\mathcal{B}_k$ and the fact that $\|w_{t_k}\| \leq \frac{5a}{8}$. However, from the definition of k, it must be that $\|w_{t_l}\| > \frac{5a}{8}$. This means that

$$\sum_{j=k}^{l} u_j + \frac{a}{32} - \frac{7a}{32} > 0,$$

which implies that $\sum_{j=k}^{l} u_j > \frac{3a}{16}$. This means that $(\mathcal{A}_k^{M_b})^c$ holds.

Now suppose that $J - k \leq M_b$, and, for a contradiction, that $\mathcal{A}_k^m$ holds. In this case, since [similarly to (C7)]

$$\int_{t_k}^{t} (-b\|w_t\| + \|w_t\|^{-1}\langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle) dt < 0,$$

for all $t \in [t_k, t_J]$,

$$\|w_{t_J}\| \leq \|w_{t_k}\| + \sum_{j=k}^{l} (u_j + (\delta t_j)^2 \bar{C}_L)$$
$$\leq \frac{5a}{8} + \frac{3a}{16} + \frac{a}{32} < \frac{7a}{8}.$$

This contradicts our assumption that $\|w_{t_l}\| \geq \frac{7a}{8}$. We thus conclude that $(\mathcal{A}_k^m)^c$ holds.

To summarize the above argument, we have shown that if $\hat{\tau}_a \geq T$ and the events $\cup_{k=1}^{\mathfrak{k}} \cup_{m=1}^{M_b} \mathcal{A}_k^m$ and $\cup_{k=1}^{\mathfrak{k}} \mathcal{B}_k$ all hold, then $\tau_a \geq T$. It thus suffices for us to bound the probabilities of each event in $\mathcal{A}_k^m$ and $\mathcal{B}_k$. In fact, the event (C3c) will always hold, as long as ab is sufficiently small. This is for the following reasons.

It has already been shown in Ref. 39 that, as long as $\|x_t - \Phi(\beta_t)\|$ is not too great (i.e., if a is sufficiently small)

$$\mathfrak{M}(x_t,\beta_t)^{-1}\omega_0^{-1}\langle \bar{F}(x_t), \Phi'(\beta_t)\rangle - 1 = O(\|x_t - \Phi(\beta_t)\|^2).$$

Since the matrix-norms of $P(\beta_t)$ and $P(\beta_t)^{-1}$ are uniformly bounded for all β_t, we find that

$$(\mathfrak{M}(x_t,\beta_t)^{-1}\omega_0^{-1}\langle \bar{F}(x_t), \Phi'(\beta_t)\rangle - 1)\mathcal{S}w_t = O(\|w_t\|^3).$$

We also have that, since $\omega_0\Phi'(\beta_t) = \bar{F}(\Phi(\beta_t))$

$$\begin{aligned}\mathfrak{F}(x_t,\beta_t) &= P(\beta_t)^{-1}\{\bar{F}(x_t) - \bar{F}(\Phi(\beta_t)) - J(\beta_t)(x_t - \Phi(\beta_t))\} \\ &\quad + O(\|w_t\|^3)\mathbb{P}(\beta_t)^{-1}\{-\bar{F}(\Phi(\beta_t)) \\ &\quad - J(\beta_t)(x_t - \Phi(\beta_t))\} = O(\|w_t\|^2),\end{aligned}$$

through Taylor's Theorem. This means that $|\langle w_t, \mathfrak{F}(\beta_t, x_t)\rangle| \leq C_3\|w_t\|^3$, for some constant C_3.

Finally, note that

$$\mathbb{P}((\mathcal{A}_k^m)^c) \leq \mathbb{P}\left(\sum_{j=k}^{k+m} u_j > \frac{3a}{16}\right). \tag{C8}$$

In Appendixes D and E, we derive the bound

$$\sup_{1\leq m\leq M_b} \mathbb{P}\left(\sum_{j=1}^{m} u_j \geq \frac{3a}{16}\right) \leq \exp\left(-\frac{\hat{C}ba^2}{\epsilon}\right). \tag{C9}$$

The proofs of (C3a) and (C3b) are similar and are omitted. For (C3b), we would find that for a positive constant $\breve{C}$

$$\mathbb{P}\left(\bar{C}_L \sum_{j=k}^{k+M_b} (\delta t_j)^2 > \frac{a}{32}\right) \leq \exp\left(-\frac{\breve{C}\,a}{\epsilon^2}\right),$$

which is of lower order than the other probabilities.

APPENDIX D: BOUNDING THE PROBABILITY $\mathbb{P}(\tau_a \leq T)$ (PART II)

In this appendix, we show how to bound the probability of $\sum_{j=1}^m u_j$ exceeding $\frac{3a}{16}$. Let the scaled transition matrix be $\tilde{\Lambda}$, with elements $(\tilde{\Lambda}_{nm})_{n,m\in\Gamma}$

$$\tilde{\Lambda}_{nm} = \Lambda_{nm}/\lambda_m. \tag{D1}$$

See Eq. (2.3). Let $\mathfrak{P}$ be the Perron projection associated with $\tilde{\Lambda}$, i.e., $\mathfrak{P}$ is the rank 1 matrix with the ith element of each column equal to $\rho_i\lambda_i/\sum_{a\in\Gamma}\rho_a\lambda_a$. We have used the fact that the dominant right eigenvector of $\tilde{\Lambda}$ is the column of $\mathfrak{P}$, and the dominant left eigenvector of $\tilde{\Lambda}$ is $(1,1,\ldots,1)$. It is a consequence of the Perron-Frobenius Theorem[53] that for some positive constant C_W and $\gamma \in (0,1)$, for all $p\in\mathbb{Z}^+$

$$\|\tilde{\Lambda}^p - \mathfrak{P}\| \leq C_W\gamma^p. \tag{D2}$$

In many situations, C_W and γ can be quite optimal, such as when the Markov Chain satisfies the Doeblin Condition or a log-Sobolev Inequality. Refer to Ref. 53 for a more in-depth discussion.

Write $\frac{3a}{16} = z$, and $C = 2C_L^2C_W^2(1-\gamma)^{-1}$. We assume that $\frac{C_2 z^2}{\epsilon^2} \gg 1$. The main result that we prove in this section is that for any positive integer R

$$\mathbb{P}\left(\sum_{j=1}^m u_j \geq z\right) \leq 2\left(\frac{CR\epsilon^2 m}{z^2}\right)^R. \tag{D3}$$

Now by Chebyshev's inequality

$$\mathbb{P}\left(\sum_{j=1}^m u_j \geq z\right) \leq \mathbb{E}\left[\left(\sum_{j=1}^m u_j\right)^{2R}\right] \times (z)^{-2R}.$$

Using the result in Appendix E

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^m u_j\right)^{2R}\right] &= \sum_{1\leq p_i\leq 2R} \mathbb{E}[u_{p_1}\ldots u_{p_R}] \\ &\leq (\epsilon^2 C_W^2 C_L^2 m(1-\gamma)^{-1})^R \\ &\quad \times \left\{1 - (1-\gamma)^{-1}\frac{R}{m}\right\}^{-1}\frac{(2R)!}{R!}.\end{aligned}$$

We use the (very crude) bound $\frac{(2R)!}{R!} \leq (2R)^R$ and we assume that

$$\left\{1 - (1-\gamma)^{-1}\frac{R}{m}\right\}^{-1} \leq 2.$$

Collecting the above bounds, we thus find that

$$\mathbb{P}\left(\sum_{j=1}^m u_j \geq z\right) \leq 2\left(\frac{CR\epsilon^2 m}{z^2}\right)^R, \tag{D4}$$

where $C = 2C_L^2C_W^2(1-\gamma)^{-1}$.

We can find the approximate R that optimizes the above bound by differentiating (i.e., approximating R to be any real number). Upon doing this, we find that the optimal R is approximately given by

$$R_m = \left\lfloor \frac{z^2}{Ce\epsilon^2 m}\right\rfloor. \tag{D5}$$

Hence, we find that

$$\mathbb{P}\left(\sum_{j=1}^m u_j \geq z\right) \leq \exp\left(-\left\lfloor\frac{z^2}{Ce\epsilon^2 m}\right\rfloor\right), \tag{D6}$$

which yields Eq. (D3). Technically, we must take C to be greater than its defined value, to account for the loss of accuracy due to $R_m \in \mathbb{Z}$. When we use this bound in Appendix C, m

ranges from 1 to $M_b = O\left(\frac{1}{b\epsilon}\right)$. We thus find that, for some positive constant $\hat{C}$ that is independent of ϵ and b, Eq. (C9) holds.

APPENDIX E: BOUNDING THE PROBABILITY $\mathbb{P}(\tau_a \leq T)$ (PART III)

In the following lemma, we bound the expectation of the sum $\sum_{j=1}^m u_j$ raised to the power of $2R$. This bound is the key result needed to bound the probability in Appendix D. This bound is useful in the regime $m \gg R$: in this regime the expectation scales as $O(\epsilon^{2R} m^R)$. The main result of this section is as follows:

$$\sum_{1\leq p_j \leq m} \mathbb{E}[u_{p_1}\ldots u_{p_{2R}}] \leq \left(\epsilon^2 C_L^2 m C_W^2 (1-\gamma)^{-1}\right)^R \times \left\{1-(1-\gamma)^{-1}\frac{R}{m}\right\}^{-1} \frac{(2R)!}{R!}. \tag{E1}$$

It follows from a substitution of the definitions that, assuming that $p_j \geq p_{j-1}$,

$$\mathbb{E}[u_{p_1}\ldots u_{p_R}] = \epsilon^{2R} \sum_{a_i \in \Gamma} \tilde{\Lambda}^{p_{2R}-p_{2R-1}}_{a_{2R}a_{2R-1}} \tilde{\Lambda}^{p_{2R-1}-p_{2R-2}}_{a_{2R-1}a_{2R-2}} \ldots \tilde{\Lambda}^{p_2-p_1}_{a_2 a_1} \tilde{u}_{p_{2R}}(a_{2R}) \tilde{u}_{p_{2R-1}}(a_{2R-1})\ldots \tilde{u}_{p_1}(a_1)\rho_{a_1}, \tag{E2}$$

where $\tilde{\Lambda}$ is the scaled transition matrix.

Now define

$$X^{p_{j+1},p_j} := \tilde{\Lambda}^{p_{j+1}-p_j} - \mathfrak{P}, \tag{E3}$$

where we recall from Appendix D that $\mathfrak{P}$ is the rank 1 matrix with the ith element of each column equal to $\rho_i\lambda_i / \sum_{a\in\Gamma}\rho_a\lambda_a$. It follows that

$$\mathbb{E}[u_{p_1}\ldots u_{p_{2R}}] = \epsilon^{2R} \sum_{q=0}^{2R} Y_q, \tag{E4}$$

where Y_q is the sum of terms of the form

$$\sum_{a_i\in\Gamma} Q(2R)_{a_{2R}a_{2R-1}} Q(2R-1)_{a_{2R-1}a_{2R-2}} \ldots Q(1)_{a_2 a_1} \tilde{u}_{p_{2R}}(a_{2R}) \tilde{u}_{p_{2R-1}}(a_{2R-1})\ldots \tilde{u}_{p_1}(a_1)\rho_{a_1},$$

and q of $\{Q(j)\}$ are equal to $\mathfrak{P}$, and the rest of $\{Q(j)\}$ are of the form X^{p_{j+1},p_j}. Now $Y_q = 0$ for all $q > R$. The reason for this is that if $q > R$, then by the pigeon-hole principle there must be some j such that $Q(j) = Q(j+1) = \mathfrak{P}$. It then follows that

$$\begin{aligned}&\sum_{a_i\in\Gamma} Q(2R)_{a_{2R}a_{2R-1}}\ldots Q(j)_{a_j a_{j-1}} Q(j-1)_{a_{j-1}a_{j-2}} \\ &\ldots Q(2)_{a_2 a_1}\tilde{u}_{p_{2R}}(a_{2R})\ldots\tilde{u}_{p_1}(a_1)\rho_{a_1} = g\sum_{a_i\in\Gamma} Q(2R)_{a_{2R}a_{2R-1}} \\ &\ldots Q(j+2)_{a_{j+2}a_{j+1}}\rho_{a_{j+1}}\lambda_{a_{j+1}}\rho_{a_j}\lambda_{a_j} Q(j-1)_{a_{j-1}a_{j-2}} \\ &\ldots Q(2)_{a_2 a_1}\tilde{u}_{p_{2R}}(a_{2R})\ldots\tilde{u}_{p_1}(a_1)\rho_{a_1} = 0,\end{aligned} \tag{E5}$$

where $g = \left(\sum_{a\in\Gamma}\rho_a\lambda_a)\right)^{-2}$, since

$$\sum_{a\in\Gamma} \rho_{a_j}\tilde{u}_{p_{j-1}}(a_j)\lambda_{a_j} = 0. \tag{E6}$$

Now using the Perron bound in (D2) and the Lipschitz bound for u

$$\left|\sum_{a_i\in\Gamma} X^{p_r-p_{r-1}}_{a_r a_{r-1}} X^{p_{r-1}-p_{r-2}}_{a_{r-1}a_{r-2}} \ldots X^{p_2-p_1}_{a_2 a_1}\tilde{u}_{p_r}(a_r)\tilde{u}_{p_{r-1}}(a_{r-1}) \ldots \tilde{u}_{p_1}(a_1)\rho_{a_1}\right| \leq C_W^r \gamma^{p_r-p_0} C_L^r.$$

In the following decomposition, we note that there if there are $j\,\mathfrak{P}$'s and $(2R-j)\,X$'s, then there are at most $m^j/j!$ possible ways of arranging the $\mathfrak{P}$'s and X's. We thus find that

$$\begin{aligned}\sum_{p_j : p_j \leq p_{j+1}} \mathbb{E}[u_{p_1}\ldots u_{p_{2R}}] &\leq \epsilon^{2R} C_W^{2R} C_L^{2R}\left[m^R/R!(1+\gamma+\gamma^2+\cdots)^R + m^{R-1}/(R-1)!(1+\gamma+\gamma^2+\cdots)^{R+1} + \cdots + m(1+\gamma+\gamma^2+\cdots)^{2R-1}\right] \\ &= (\epsilon^2 C_L^2 C_W^2 m(1-\gamma)^{-1})^R/R! \times \left\{1+\frac{R}{m}(1-\gamma)^{-1} + \frac{R(R-1)}{m^2}(1-\gamma)^{-1} + \cdots \frac{R!}{M^R}(1-\gamma)^{-R}\right\} \\ &\leq (\epsilon^2 C_L^2 C_W^2 m(1-\gamma)^{-1})^R \left\{1-(1-\gamma)^{-1}\frac{R}{m}\right\}^{-1}/R!,\end{aligned}$$

assuming that $(1-\gamma)^{-1}\frac{R}{m} < 1$. Now since there are at most $(2R)!$ different ways of choosing the $\{p_r\}$ such that $p_{j-1} \leq p_j$, we find that

$$\sum_{p_j} \mathbb{E}[u_{p_1}\ldots u_{p_{2R}}] \leq (\epsilon^2 C_W^2 C_L^2 m(1-\gamma)^{-1})^R \times \left\{1-(1-\gamma)^{-1}\frac{R}{m}\right\}^{-1}\frac{(2R)!}{R!}.$$

[1] P. C. Bressloff, "Stochastic switching in biology: From genotype to phenotype (Topical Review)," J. Phys. A **50**, 133001 (2017).

[2] M. H. A. Davis, "Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models," J. R. Soc., Ser. B **46**, 353–388 (1984).

[3] 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).

[4] C. C. Chow and J. A. White, "Spontaneous action potentials due to channel fluctuations," Biophys. J. **71**, 3013–3021 (1996).

[5]K. Pakdaman, M. Thieullen, and G. Wainrib, "Fluid limit theorems for stochastic hybrid systems with application to neuron models," Adv. Appl. Probab. **42**, 761–794 (2010).
[6]J. P. Keener and J. M. Newby, "Perturbation analysis of spontaneous action potential initiation by stochastic ion channels," Phys. Rev. E **84**, 011918 (2011).
[7]J. H. Goldwyn and E. Shea-Brown, "The what and where of adding channel noise to the Hodgkin-Huxley equations," PLoS Comput. Biol. **7**, e1002247 (2011).
[8]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).
[9]G. Wainrib, M. Thieullen, and K. Pakdaman, "Reduction of stochastic conductance-based neuron models with time-scales separation," J. Comput. Neurosci. **32**, 327–346 (2012).
[10]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).
[11]P. C. Bressloff and J. M. Newby, "Stochastic hybrid model of spontaneous dendritic NMDA spikes," Phys. Biol. **11**, 016006 (2014).
[12]J. M. Newby, "Spontaneous excitability in the Morris–Lecar model with ion channel noise," SIAM J. Appl. Dyn. Syst. **13**, 1756–1791 (2014).
[13]D. Anderson, G. B. Ermentrout, and P. J. Thomas, "Stochastic representations of ion channel kinetics and exact stochastic simulation of neuronal dynamics," J. Comput. Neurosci. **38**, 67–82 (2015).
[14]T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116–3136 (2001).
[15]R. Karmakar and I. Bose, "Graded and binary responses in stochastic gene expression," Phys. Biol. **1**, 197–204 (2004).
[16]S. Zeiser, U. Franz, O. Wittich, and V. Liebscher, "Simulation of genetic networks modelled by piecewise deterministic Markov processes," IET Syst. Biol. **2**, 113–135 (2008).
[17]M. W. Smiley and S. R. Proulx, "Gene expression dynamics in randomly varying environments," J. Math. Biol. **61**, 231–251 (2010).
[18]J. M. Newby, "Isolating intrinsic noise sources in a stochastic genetic switch," Phys. Biol. **9**, 026002 (2012).
[19]J. M. Newby, "Bistable switching asymptotics for the self regulating gene," J. Phys. A **48**, 185001 (2015).
[20]P. G. Hufton, Y. T. Lin, T. Galla, and A. J. McKane, "Intrinsic noise in systems with switching environments," Phys. Rev. E **93**, 052119 (2016).
[21]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).
[22]Y. Kifer, *Large Deviations and Adiabatic Transitions for Dynamical Systems and Markov Processes in Fully Coupled Averaging*, Memoirs of the American Mathematical Society, Volume 201, Number 944 (AMS, 2009).
[23]A. Faggionato, D. Gabrielli, and M. R. Crivellari, "Averaging and large deviation principles for fully-coupled piecewise deterministic Markov processes and applications to molecular motors," arXiv:0808.1910 (2008).
[24]P. C. Bressloff and O. Faugeras, "On the Hamiltonian structure of large deviations in stochastic hybrid systems," J. Stat. Mech. **2017**, 033206.
[25]P. C. Bressloff and J. M. Newby, "Path-integrals and large deviations in stochastic hybrid systems," Phys. Rev. E **89**, 042701 (2014).
[26]H. Feng, B. Han, and J. Wang, "Landscape and global stabilityn of nonadiabatic and adiabatic oscillations in a gene network," Biophys. J. **102**, 1001–1010 (2012).
[27]D. Labavic, H. Nagel, W. Janke, and H. Meyer-Ortmanns, "Caveats in modeling a common motif in genetic circuits," Phys. Rev. E **87**, 062706 (2013).
[28]H. A. Brooks and P. C. Bressloff, "Quasicycles in the stochastic hybrid Morris-Lecar neural model," Phys. Rev. E **92**, 012704 (2015).
[29]J. N. Teramae and D. Tanaka, "Robustness of the noise-induced phase synchronization in a general class of limit cycle oscillators," Phys. Rev. Lett. **93**, 204103 (2004).
[30]D. S. Goldobin and A. Pikovsky, "Synchronization and desynchronization of self-sustained oscillators by common noise," Phys. Rev. E **71**, 045201 (2005).
[31]H. Nakao, K. Arai, and Y. Kawamura, "Noise-induced synchronization and clustering in ensembles of uncoupled limit cycle oscillators," Phys. Rev. Lett. **98**, 184101 (2007).
[32]K. Yoshimura and K. Ara, "Phase reduction of stochastic limit cycle oscillators," Phys. Rev. Lett. **101**, 154101 (2008).
[33]J. N. Teramae, H. Nakao, and G. B. Ermentrout, "Stochastic phase reduction for a general class of noisy limit cycle oscillators," Phys. Rev. Lett. **102**, 194102 (2009).
[34]G. B. Ermentrout, "Noisy oscillators," in *Stochastic Methods in Neuroscience*, edited by C. R. Laing and G. J. Lord (Oxford University Press, Oxford, 2009).
[35]D. Gonze, J. Halloy, and P. Gaspard, "Biochemical clocks and molecular noise: Theoretical study of robustness factors," J. Chem. Phys. **116**, 10997–11010 (2002).
[36]H. Koeppl, M. Hafner, A. Ganguly, and A. Mehrotra, "Deterministic characterization of phase noise in biomolecular oscillators," Phys. Biol. **8**, 055008 (2011).
[37]M. Bonnin, "Amplitude and phase dynamics of noisy oscillators," Int. J. Circuit Theory Appl. **45**, 636–659 (2017).
[38]P. Ashwin, S. Coombes, and R. Nicks, "Mathematical frameworks for oscillatory network dynamics in neuroscience," J. Math. Neurosci. **6**, 2 (2016).
[39]P. C. Bressloff and J. N. Maclaurin, "Variational method for analyzing stochastic limit cycle oscillators," SIAM J. Appl. Dyn. Syst. (to be published).
[40]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).
[41]C. W. Gardiner, *Handbook of Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).
[42]P. C. Bressloff, *Stochastic Processes in Cell Biology* (Springer, 2014).
[43]A. Winfree, *The Geometry of Biological Time* (Springer-Verlag, New York, 1980).
[44]Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence* (Springer-Verlag, New-York, 1984).
[45]L. Glass and M. C. Mackey, *From Clocks to Chaos* (Princeton University Press, Princeton, 1988).
[46]G. B. Ermentrout, "Type I membranes, phase resetting curves, and synchrony," Neural Comput. **8**, 979 (1996).
[47]E. Brown, J. Moehlis, and P. Holmes, "On the phase reduction and response dynamics of neural oscillator populations," Neural Comput. **16**, 673–715 (2004).
[48]B. Ermentrout and D. Terman, *Mathematical Foundations of Neuroscience* (Springer, New York, 2010).
[49]C. Morris and H. Lecar, "Voltage oscillations in the barnacle giant muscle fiber," J. Biophys. **35**, 193–213 (1981).
[50]J. A. White, T. Budde, and A. R. Kay, "A bifurcation analysis of neuronal subthreshold oscillations," Biophys. J. **69**, 1203–1217 (1995).
[51]J. A. White, J. T. Rubinstein, and A. R. Kay, "Channel noise in neurons," Trends Neurosci. **23**, 131–137 (2000).
[52]J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator," Nature **456**, 516–519 (2008).
[53]L. Saloff-Coste, "Lectures on finite Markov chains," in *Lectures on Probability Theory and Statistics* (Springer, 1997), pp. 301–413.
[54]I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed. (Springer-Verlag, 1991).