Noise-induced switching and extinction in systems with delay

Ira B. Schwartz, Lora Billings, Thomas W. Carr, and M. I. Dykman

Noise Induced Switching and Extinction in Systems with Delay

Ira B. Schwartz,[1] Lora Billings,[2] Thomas W. Carr,[3] and M. I. Dykman[4]

[1] *US Naval Research Laboratory,Code 6792,Nonlinear System*
Dynamics Section,Plasma Physics Division,Washington, DC 20375
[2] *Department of Mathematical Sciences, Montclair State University, Montclair, NJ 07043*
[3] *Department of Mathematics, Southern Methodist University, Dallas, TX 75275*
[4] *Department of Physics and Astronomy,Michigan State University,East Lansing, MI 48824*

We consider the rates of noise-induced switching between the stable states of dissipative dynamical systems with delay and also the rates of noise-induced extinction, where such systems model population dynamics. We study a class of systems where the evolution depends on the dynamical variables at a preceding time with a fixed time delay, which we call hard delay. For weak noise, the rates of inter-attractor switching and extinction are exponentially small. Finding these rates to logarithmic accuracy is reduced to variational problems. The solutions of the variational problems give the most probable paths followed in switching or extinction. We show that the equations for the most probable paths are acausal and formulate the appropriate boundary conditions. Explicit results are obtained for small delay compared to the relaxation rate. We also develop a direct variational method to find the rates. We find that the analytical results agree well with the numerical simulations for both switching and extinction rates.

I. INTRODUCTION

Many physical systems of current interest, including population systems, are characterized by delay. The evolution of such systems depends not only on the values of their dynamical variables at a given time, but also on the values these variables took at previous times. Often there is a single time delay fixed to a certain value, which we call a hard-delay case. Well-known examples are provided by optical systems such as ring lasers or lasers with external cavities, where the delay is determined by the duration of light propagation along a certain path, cf. [1–8]. Another example is Josephson junctions coupled to shunting transmission lines [9, 10]. Examples in biology include various systems, such as neural networks and genetic networks with delay [11, 12], and others. In population models, [13–16], the delay can be related to temporary immunity in disease propagation or the time between conception and birth. Evidence of the pervasiveness of delays across science disciplines can be seen in several reviews, e.g., [17]. A qualitative stability analysis of delayed systems can be found in Refs. 18–20.

An important role in the systems mentioned above is played by noise. Delayed dissipation and fluctuations of systems coupled to a thermal bath have been discussed starting from the mid-60s [21, 22], and several interesting delay-related features of fluctuations have been found, see [23–29] and references therein. This work focused primarily on the delay described by a retarded friction force of the form of an integral of the velocity over the time preceding the observation time; such forces naturally arise in systems linearly coupled to a thermal bath of harmonic oscillators as well as from hydrodynamic memory effects [30–32]. Systems with hard delay have different physics behind them. In the analysis of such systems, much attention has been paid to describing the system behavior by an effective Fokker-Planck equation, cf. [33–36].

A complete description of fluctuations in the presence of delay is complicated by the fact that the system phase space is infinite dimensional. Even without noise, to predict the future one has to know not just the instantaneous values of the dynamical variables, but their values in the past on a finite or possibly infinite interval of time. This limits the applicability of the description of the dynamics in terms of the probability distribution in a finite-dimensional space.

We will be primarily interested in fluctuations in dynamical systems with hard delay that have one or more simple stationary states. Without noise such systems asymptotically approach one of the stable states depending on initial history. If the noise is weak, on average it causes small fluctuations about the occupied stable state. However, occasionally there occur large outbursts of noise that drive the system far from the stable state in the space of dynamical variables or can result in switching from the occupied stable state to another stable state.

An important qualitative outcome of large noise outbursts, which plays a critical role in understanding population dynamics and more generally, reaction systems, is extinction of one of the species. To the best of our knowledge, there is no analytical theory of weak-noise induced extinction in the presence of delay. When a (generally, multi-species) population is described by a dynamical system, extinction corresponds to reaching a stationary state on the boundary of an attraction basin where one of the dynamical variables is zero; this variable describes the population of the species that goes extinct. In our analysis, we will use the term "extinction" in this meaning.

The idea of our approach goes back to Feynman [37], who noticed that, in noise driven systems, even though the noise is random, each noise realization leads to a certain system trajectory. Therefore, the probability density of realizations of system trajectories is determined by the probability density of realizations of the noise trajectories. To find the rate of occurrence of a rare event one

has to look for the most probable realization of the noise trajectories that bring the system to the desired state.

In what follows we formulate the variational problem for the most probable path followed by the system on its way to a desired state. Its solution gives the exponents of the rates of rare events. The possibility to find the most probable path in the presence of a delay is rooted in the fact that, prior to the rare event, the system spends a long time performing small fluctuations about the occupied stable state. This time is much longer than the delay, if the noise is weak on average. This alleviates the problem of initial conditions for trajectories in systems with delay. Our formulation includes the analysis of switching between the stable states as well as extinction of a species. We show that these two cases should be analyzed differently.

The significant effect of delay on extinction is physically very clear. Indeed if, as a result of a fluctuation, the species has gone extinct at a given instant of time, prior to this time the species population was nonzero. Therefore the population can come back into existence. The idea can be illustrated by a simple example from the delayed population growth problem: if you sacrifice all of the chickens but keep the eggs, you will still have chickens after the eggs hatch.

Previously we discussed the rates of rare events for a one-dimensional particle with inertia and a retarded friction force described by an integral of the velocity over time with a weighting factor [38]. We formulated a variational problem for the most probable path followed in switching between stable states in this case. In Ref. [39] we obtained a numerical solution for the most probable switching path in a model system with hard delay and compared the results with Monte-Carlo simulations, with adjusted parameters.

In contrast to the numerical work on a particular system with hard delay, here we are interested in the general features of the effect of hard delay on the rates of rare events. This includes the possibility of developing a perturbation theory in the delay. It turns out that, because of the difference in the character of the delay, the predictions of the perturbation theory we develop here are qualitatively different from those in systems studied earlier [38]. We also consider the problem of the extinction rate in the presence of delay, which requires a significant extension of both the analytical theory and the Monte Carlo simulation technique, as it is necessary to consider large rare fluctuations induced by a singular multiplicative noise. The other range of questions we are addressing here concerns the onset of scaling behavior of the switching and extinction rate. Yet another general question is whether delay increases or decreases the rates of switching and extinction.

The layout of the paper is as follows. In Sec. II, we formulate a dynamical model that describes fluctuations in systems with hard delay. We discuss the stability of stationary states in the presence of delay and consider small-amplitude fluctuations about the stable states due to weak on average noise. In Sec. III we formulate a variational approach to the problem of reaching a state in the space of dynamical variables, which is remote from the initially occupied stable state. We extend this formulation to consider the rates of switching between stable states and the problem of extinction. In Sec. IV we find explicit solutions for the rates of switching and extinction near bifurcation points, including the critical exponents that describe the scaling of the rates with the distance to the bifurcation point. We develop a perturbation theory in the ratio of the delay to the relaxation time of the system. We also develop a direct variational method for the analysis of the rates. In Sec. V we compare our theory of the switching rates with detailed parameter-free numerical simulations, whereas in Sec. VI we compare the theory and numerical simulations for the problem of extinction. We finish the paper with a discussion in Sec. VII.

II. MODEL OF A DELAYED NOISE-DRIVEN SYSTEM

We consider a system with dynamical variables $\mathbf{q} = (q_1, q_2, \ldots, q_N)$ and with delay time τ. Its stochastic dynamics are described by the Langevin equation

$$\dot{\mathbf{q}}(t) = \mathbf{K}\big(\mathbf{q}(t), \mathbf{q}(t-\tau)\big) + \hat{G}(\mathbf{q}(t))\mathbf{f}(t). \qquad (1)$$

Here, $\mathbf{K} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ defines the noise-free evolution, whereas $\mathbf{f}(t) \equiv (f_1(t), f_2(t), \ldots, f_M(t))$ is the M-dimensional noise vector with zero-mean components $f_m(t)$, and $\hat{G}(\mathbf{q}(t))$ is an $N \times M$ matrix such that $M \leq N$. We assume the noise to be Gaussian, stationary, and weak on average, but generally not δ-correlated in time; also, the components $f_n(t)$ can be cross-correlated. For brevity, we use notations $\mathbf{q}_\tau \equiv \mathbf{q}(t-\tau)$ and $\mathbf{q}_{-\tau} \equiv \mathbf{q}(t+\tau)$.

We assume that the noise-free equation with delay $\dot{\mathbf{q}} = \mathbf{K}(\mathbf{q}, \mathbf{q}_\tau)$ has a stationary stable state (attractor) $\mathbf{q}_A$, near where the system is initially located, and a saddle point $\mathbf{q}_S$. The stationary states satisfy $\mathbf{K}(\mathbf{q}_A, \mathbf{q}_A) = \mathbf{K}(\mathbf{q}_S, \mathbf{q}_S) = \mathbf{0}$.

The stability of each stationary state is given by the linearized equations of motion about that state. They have the form

$$\dot{\mathbf{X}}(t) = \hat{\mathcal{K}}^{(1)}\mathbf{X}(t) + \hat{\mathcal{K}}^{(2)}\mathbf{X}(t-\tau),$$
$$\mathcal{K}^{(1)}_{nn'} = \partial K_n / \partial q_{n'}, \qquad \mathcal{K}^{(2)}_{nn'} = \partial K_n / \partial (q_\tau)_{n'}. \quad (2)$$

The partial derivatives of $\mathbf{K}(\mathbf{q}, \mathbf{q}_\tau)$ that determine the matrices $\hat{\mathcal{K}}^{(1)}$ and $\hat{\mathcal{K}}^{(2)}$ are evaluated at $\mathbf{q} = \mathbf{q}_\tau = \mathbf{q}_A$ or $\mathbf{q} = \mathbf{q}_\tau = \mathbf{q}_S$; respectively, $\mathbf{X} = \mathbf{q} - \mathbf{q}_A$ or $\mathbf{X} = \mathbf{q} - \mathbf{q}_S$. Seeking solution of Eq. (2) in the form $\mathbf{X}(t) = e^{\alpha t}\mathbf{v}$ leads one to the eigenvalue problem

$$h(\alpha, \tau) = \det\left[\alpha\mathcal{I} - \hat{\mathcal{K}}^{(1)} - \exp(-\alpha\tau)\hat{\mathcal{K}}^{(2)}\right] = 0, \quad (3)$$

where $\mathcal{I}$ is the identity matrix.

A. Stability in the case of small delay

From here on, we assume that in the zero delay case, $\tau = 0$, the attractor has all eigenvalues α_i with negative real part. The saddle has only one positive real eigenvalue (associated with an unstable direction in the space of dynamical variables), while the rest of the eigenvalue spectrum lies in the left hand side of the complex plane. In general, delays have the tendency to destabilize existing attractors. We will assume that the attractors are not destroyed. This is the case when delays are small. However, often attractors remain stable even when delay is on the order of the relaxation time. Moreover, chaotic behavior emerges in remote regions of phase space, which the systems does not encounter unless initially prepared there. In this paper we do not study the effects of chaos.

Stability of the small delay case can be discerned directly from Eq. (3). The result for small delay is proven rigorously in Bellman and Cook [18]. It is shown in Ref. 18 and 40 that there is a finite range of the values of τ such that in this range the roots of the characteristic equation $h(\alpha, \tau) = 0$ are either close to the roots of $h(\alpha, 0) = 0$ or that the perturbed roots have negative real parts which are larger in the absolute value than those satisfying $h(\alpha, 0) = 0$. This indicates that small delays do not change the asymptotic stability of the attractors in the absence of delays.

It was also proven in Ref. 18 that simple roots of $h(\alpha, \tau)$ change continuously with τ given that $h(\alpha, \tau)$ is analytic in α in a certain finite range around the corresponding roots for $\tau = 0$ and $\partial_\tau h(\alpha, \tau)$ is continuous. This means that, if motion near the saddle is characterized by one positive root for $\tau = 0$, it will still have one positive root in a certain range of small τ.

We will assume that, for small τ, no new asymptotically stable states emerge compared to the case $\tau = 0$. Of primary interest to us will be two cases. One is the case where the system has two attractors along with the saddle point $\mathbf{q}_S$. The basins of attraction can be built in a standard way by moving away from attractors in negative time (while holding the system near the attractor for time τ). These basins are separated by the separating manifold which contains $\mathbf{q}_S$. For $\tau = 0$ the system moves away from $\mathbf{q}_S$ in positive time along the eigenvector $\mathbf{e}_u$ that corresponds to the positive eigenvalue α. For $\mathbf{q}$ close to $\mathbf{q}_S$, depending on the sign of $(\mathbf{q} - \mathbf{q}_S) \cdot \mathbf{e}_u$ the system will approach one or the other attractor. This is also true for small τ given that the system stays near $\mathbf{q}_S$ for time τ, since the system is characterized by only one positive eigenvalue near $\mathbf{q}_S$; the trajectories from $\mathbf{q}_S$ to the attractors will just slightly change.

The other case is where the system has one attractor and a saddle point, but is constrained to stay in the space of the dynamical variables on the one side of a hyperplane that goes through the saddle point normal to $\mathbf{e}_u$. Here, one of the dynamical variables can be thought of as the population of a certain species in a continuous limit. It coincides with $(\mathbf{q} - \mathbf{q}_S) \cdot \mathbf{e}_u$ for small $|\mathbf{q} - \mathbf{q}_S|$).

By construction, it can always be made nonnegative.

B. Small-amplitude fluctuations about the stable states

The stationary zero-mean Gaussian noise $\mathbf{f}(t)$ is characterized by its time correlation functions $\phi_{nm}(t)$ and their Fourier transforms $\Phi_{nm}(\omega)$,

$$\phi_{nm}(t) = \langle f_n(t) f_m(0) \rangle,$$
$$\Phi_{nm}(\omega) = \frac{1}{2} \int_{-\infty}^{\infty} dt \exp(i\omega t) \phi_{nm}(t). \tag{4}$$

Functions Φ_{nm} determine the power spectrum of the noise and can be often directly accessed in an experiment. The noise intensity D can be defined as twice the maximal eigenvalue of the matrix $\hat{\Phi}(\omega)$ with matrix elements Φ_{nm}. Since the noise $\mathbf{f}(t)$ is stationary, matrix $\hat{\phi}(t)$ with matrix elements $\phi_{nm}(t)$ has an obvious property $\hat{\phi}^\dagger(t) = \hat{\phi}(-t)$.

For small noise intensity, over a characteristic relaxation time the system approaches an attractor and then mostly performs small-amplitude fluctuations about it. For a given attractor $\mathbf{q}_A$ these fluctuations can be described by linearizing equation of motion (1) about $\mathbf{q}_A$ and assuming that $\mathbf{q}(t) \to \mathbf{q}_A$ for $t \to -\infty$. Changing to Fourier components $\delta\mathbf{q}(\omega) = \int_{-\infty}^{\infty} dt \exp(i\omega t)[\mathbf{q}(t) - \mathbf{q}_A]$ and $\mathbf{f}(\omega) = \int_{-\infty}^{\infty} dt \exp(i\omega t)\mathbf{f}(t)$, we find

$$\delta\mathbf{q}(\omega) = \hat{\mathcal{G}}(\omega)\mathbf{f}(\omega), \tag{5}$$
$$\hat{\mathcal{G}}(\omega) = -\left(i\omega\mathcal{I} + \hat{\mathcal{K}}^{(1)} + \hat{\mathcal{K}}^{(2)}e^{i\omega\tau}\right)^{-1}\hat{G}(\mathbf{q}_A),$$

where the matrix elements of matrices $\hat{\mathcal{K}}^{(1,2)}$ are given by Eq. (3).

From Eq. (5), the correlator of small-amplitude fluctuations of the delayed system is

$$\langle \delta q_n(\omega) \delta q_m(\omega') \rangle = 4\pi[\hat{\mathcal{G}}(\omega)\hat{\Phi}(\omega)\hat{\mathcal{G}}^\dagger(\omega)]_{nm}\delta(\omega + \omega'). \tag{6}$$

Equations (5) and (6) fully describe small noise-induced fluctuations about the attractor in the presence of delay. The fluctuations are Gaussian, as in the case where there is no delay. Their probability distribution is determined by the matrix with matrix elements (6). This distribution and the power spectrum of the fluctuations explicitly depend on the delay time τ.

III. VARIATIONAL PROBLEM FOR LARGE RARE FLUCTUATIONS

Even though the noise is weak on average, occasionally there occur large outbursts of the noise that can drive the system far away from the attractor and also lead to switching between coexisting attractors. The analysis of the probability distribution far away from an attractor and of the switching rates in systems with hard delay

and multiplicative noise can be done by extending the analysis for systems driven by additive noise in the absence of delay or with the delay due to linear coupling to a thermal reservoir [38, 41]. The theorems about the stationary states facilitate this analysis in the case of small hard delay. We will not consider fluctuations to the areas of phase space where, because of the delay, there emerge new stationary or periodic states or dynamical chaos.

The key ideas are that, first, prior to a large fluctuation, for a long time the system is performing small-amplitude fluctuations about the initially occupied attractor. Second, as mentioned in the Introduction, even though the noise trajectories themselves are random, each noise trajectory leads to a system trajectory, which is uniquely defined by equation of motion (1) [37]. Third, since the probability densities of different noise trajectories are exponentially different, in a fluctuation to a given state the system most likely follows a well-defined trajectory, which is called the optimal trajectory and which corresponds to the most probable appropriate noise realization. For systems without delay optimal trajectories followed in interstate switching have been observed in experiments [4, 42, 43]. The actual trajectories form a narrow tube centered at the optimal trajectory. This is also true for trajectories followed in a rare fluctuation to any state far from the attractor. The probability of the realization of such a tube of trajectories gives the stationary probability density of reaching this state.

We now apply these ideas to the problem at hand. The probability density of a realization of Gaussian noise $\mathbf{f}(t)$ is given by the probability density functional $\mathcal{P}_{\mathbf{f}}[\mathbf{f}(t)] = \exp\left(-\mathcal{R}_{\mathbf{f}}/D\right)$ [37], where

$$\mathcal{R}_{\mathbf{f}}[\mathbf{f}(t)] = \frac{1}{4} \int_{-\infty}^{\infty} dt\, dt'\, \mathbf{f}(t)\hat{\mathcal{F}}(t - t')\mathbf{f}(t'), \qquad (7)$$

$\hat{\mathcal{F}}(t)$ is the inverse of the pair correlator of $\mathbf{f}(t)$, $\int dt' \hat{\mathcal{F}}(t - t')\hat{\phi}(t' - t'') = 2D\mathcal{I}\delta(t - t'')$, and D is the noise intensity. Clearly, D drops out from the expression for $\mathcal{P}_{\mathbf{f}}$ in terms of $\hat{\phi}(t)$; however, it is convenient to have it written explicitly for bookkeeping purposes.

Following the arguments given above, to find the optimal trajectory for reaching a state $\mathbf{q}_0$ we have to solve the variational problem of minimizing $\mathcal{R}_{\mathbf{f}}[\mathbf{f}(t)]$ (and thus maximizing the value of the probability density functional $\mathcal{P}_{\mathbf{f}}[\mathbf{f}(t)]$) with the constraint that the system is at the attractor, $\mathbf{q}(t) \to \mathbf{q}_A$ for $t \to -\infty$, whereas it is found at $\mathbf{q}_0$ for a given $t = t_0$. This means we have to minimize the functional

$$\mathcal{R}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}] = \mathcal{R}_{\mathbf{f}}[\mathbf{f}(t)]$$
$$+ \int_{-\infty}^{\infty} dt \boldsymbol{\chi}(t)[\dot{\mathbf{q}}(t) - \mathbf{K}(\mathbf{q}, \mathbf{q}_\tau) - \hat{G}(\mathbf{q})\mathbf{f}(t)]. \qquad (8)$$

Here, $\boldsymbol{\chi}(t)$ is the Lagrange multiplier. It relates the trajectory of the system to the noise trajectory. The minimum is taken with respect to trajectories that go from $\mathbf{q}(t) \to \mathbf{q}_A, \mathbf{f}(t) \to 0, \boldsymbol{\chi}(t) \to 0$ for $t \to -\infty$ (the optimal noise realization is $\mathbf{f} \to \mathbf{0}$ where the system is at the attractor for $t \to -\infty$), whereas for a given t_0 we have $\mathbf{q}(t_0) = \mathbf{q}_0$.

To logarithmic accuracy, the probability density $\rho(\mathbf{q}_0)$ to be at point $\mathbf{q}_0$ is given by the value of $\mathcal{P}$ for the appropriate noise realization,

$$\rho(\mathbf{q}_0) = \text{const} \times \exp[-R(\mathbf{q}_0)/D],$$
$$R(\mathbf{q}_0) = \min \mathcal{R}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}], \qquad (9)$$

where the constant weakly depends on the noise intensity D.

Equation (9) can be compared to the Boltzmann distribution, with D playing the role of temperature (more precisely, $k_B T$) and $R(\mathbf{q}_0)$ playing the role of an effective "energy" in the state $\mathbf{q}_0$. We emphasize that this is not the real energy of the considered noise-driven delayed dissipative system, which would be hard to define. For white-noise driven systems without delay the analog of $R(\mathbf{q})$ is sometimes called a nonequilibrium potential or a quasipotential [44, 45]. In Eq. (9) it is assumed that $\exp[-R(\mathbf{q}_0)/D] \ll 1$.

The variational equations for the trajectories $\mathbf{f}(t), \mathbf{q}(t), \boldsymbol{\chi}(t)$ that minimize functional $\mathcal{R}$ are obtained in a straightforward way. One of them, $\delta\mathcal{R}/\delta\boldsymbol{\chi}(t) = 0$, is equation of motion (1). The condition of extremum with respect to $\mathbf{f}(t)$ gives

$$\frac{\delta\mathcal{R}}{\delta\mathbf{f}(t)} = \frac{1}{2} \int_{-\infty}^{\infty} dt'(\hat{\mathcal{F}}(t - t')\mathbf{f}(t') - \hat{G}^\dagger(\mathbf{q}(t))\boldsymbol{\chi}(t)) = 0. \qquad (10)$$

whereas the condition of extremum with respect to $\mathbf{q}(t)$ gives

$$\frac{\delta\mathcal{R}}{\delta\mathbf{q}(t)} = -\dot{\boldsymbol{\chi}}(t) - \partial_{\mathbf{q}(t)}\Big[\boldsymbol{\chi}(t)\mathbf{K}\big(\mathbf{q}(t), \mathbf{q}(t - \tau)\big) + \boldsymbol{\chi}(t + \tau)\mathbf{K}\big(\mathbf{q}(t + \tau), \mathbf{q}(t)\big)\Big] - \partial_{\mathbf{q}(t)}\Big(\boldsymbol{\chi}(t)\hat{G}\big(\mathbf{q}(t)\big)\mathbf{f}(t)\Big) = 0. \qquad (11)$$

Here, differentiation of $\mathbf{K}\big(\mathbf{q}(t), \mathbf{q}(t - \tau)\big)$ and $\mathbf{K}\big(\mathbf{q}(t + \tau), \mathbf{q}(t)\big)$ over $\mathbf{q}(t)$ is done keeping $\mathbf{q}(t \pm \tau)$ constant.

Function $\boldsymbol{\chi}(t)$ has simple physical meaning. It determines the change of $R(q_0)$ if the system is additionally

driven by a regular time-dependent force. This change is linear in the force. Therefore $\boldsymbol{\chi}(t)$ gives the linear in the force change of the logarithm of the probability distribution and has been called the logarithmic susceptibility [46]. It can be directly measured in the experiment. Such function is also well-known in the optimal control theory for systems without delay and without noise, in particular in the linear quadratic optimal control problems, where it determines the optimal control field.

In the considered problem with delay, Eq. (11) for $\boldsymbol{\chi}(t)$ is acausal, the value of $\boldsymbol{\chi}(t)$ depends on the values of the dynamical variables on the trajectory at time $t + \tau$. This is a remarkable generic feature of the optimal paths in systems with delay. It is a consequence of the non-locality in time of the variational functional $\mathcal{R}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}]$. We emphasize that the dynamical equation (1) is causal. The acausality in the problem of optimal fluctuations emerges, because we consider global dynamics where both the initial and the final states are specified. The evolution of the system coordinate $\mathbf{q}(t)$ is causal, but the optimal force needed to drive the system to the final state depends on the global trajectory.

Since on physical grounds it is clear that the large outburst of noise $\mathbf{f}(t)$ that causes a large rare fluctuation should decay for $t \to \pm\infty$, Eq. (10) has an obvious solution

$$\mathbf{f}(t) = D^{-1} \int_{-\infty}^{\infty} dt' \hat{\phi}(t - t') \hat{G}^{\dagger}(\mathbf{q}(t')) \boldsymbol{\chi}(t'). \quad (12)$$

It is also clear on physical grounds that the fate of the noise-driven system after it has reached the target $\mathbf{q}_0$ does not affect the exponent of the probability distribution, and this exponent is determined by the probability density to reach $\mathbf{q}_0$ for the first time. This provides the remaining boundary condition for the variational problem (8) and (9). For $t > t_0$ we "disconnect" the system from the noise, $\boldsymbol{\chi}(t) = 0$ for $t > t_0$. This is the same boundary conditions as in the absence of delay [41].

The functional $\mathcal{R}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}]$ can be associated with mechanical action of an auxiliary dynamical noise-free system. The analogy becomes even more clear if $\mathbf{f}$ is eliminated using Eq. (12) and $\mathcal{R}$ is written in the form

$$\mathcal{R} = \int dt \boldsymbol{\chi}(t) \dot{\mathbf{q}}(t) - \int dt \mathcal{H}[\mathbf{q}, \boldsymbol{\chi}],$$
$$\mathcal{H}[\mathbf{q}, \boldsymbol{\chi}] = \boldsymbol{\chi}(t) \mathbf{K}[\mathbf{q}(t), \mathbf{q}(t - \tau)]$$
$$+ \frac{1}{2D} \int dt' \boldsymbol{\chi}(t) \hat{G}(\mathbf{q}(t)) \hat{\phi}(t - t') \hat{G}^{\dagger}(\mathbf{q}(t')) \boldsymbol{\chi}(t'). \quad (13)$$

One can interpret $\mathcal{H}[\mathbf{q}, \boldsymbol{\chi}]$ as a nonlocal in time Hamiltonian of the auxiliary system; in what follows we call it effective Hamiltonian. Equations (1) and (11) with $\mathbf{f}$ expressed in terms of $\boldsymbol{\chi}$ become generalized Hamiltonian equations, with $\mathbf{q}$ and $\boldsymbol{\chi}$ playing the roles of the coordinate and momentum of the auxiliary system.

A. Switching rate

Of significant interest to us is the switching rate in a system with hard delay. For switching to occur, after the noise outburst decays, the system has to be in the basin of attraction to the other attractor or on its boundary. Since the noise decays for $t \to \infty$, the system should approach a stationary state, which is either the attractor or the saddle point. From Eq. (12), the decay of the noise for $t \to \infty$ requires that either the Lagrange multiplier $\boldsymbol{\chi}(t)$ also decays for $t \to \infty$, or that $\hat{G}^{\dagger}(\mathbf{q}(t)) \to 0$, i.e., that $\hat{G}^{\dagger}(\mathbf{q}(t)) = 0$ at the corresponding stationary state. This latter case will be discussed in the next subsection. Here we assume that $\hat{G}^{\dagger}(\mathbf{q}(t))$ remains finite. Then near a stationary state for $t \to \infty$ we have $|\boldsymbol{\chi}(t)|, |\mathbf{f}(t)| \to 0$. We now linearize Eq. (11) near a stationary state,

$$\dot{\boldsymbol{\chi}}(t) \approx -\boldsymbol{\chi}(t) \hat{\mathcal{K}}^{(1)\dagger} - \boldsymbol{\chi}(t + \tau) \hat{\mathcal{K}}^{(2)\dagger}, \quad (14)$$

where matrices $\hat{\mathcal{K}}^{(1,2)}$ characterize the motion of the system near a stationary state in the absence of noise and are defined by Eq. (2).

By comparing Eq. (14) with the characteristic equation for linearized motion of the system near a stationary state in the absence of noise (3), we see that the roots of the characteristic equation for $\boldsymbol{\chi}(t)$ are equal to the complex-conjugate roots of the characteristic equation for the noise-free system taken with the opposite sign. As a consequence, by moving along the optimal path the system may not approach an attractor, as $\boldsymbol{\chi}(t)$ will be exponentially increasing there. However, the system may approach the saddle point, where one of the roots of the characteristic equation for noise-free motion is positive, and the corresponding root for Eq. (14) is negative. This shows that the endpoint for the optimal trajectory leading to switching is the saddle point, similar to the case of systems with no delay [41]. As the system approaches this point, $\boldsymbol{\chi}(t)$ evolves along the vector $\mathbf{e}_u$ that corresponds to the unstable eigenvector of the linearized noise-free motion.

To logarithmic accuracy, the rate of switching from attractor $\mathbf{q}_{A_j}$ to attractor $\mathbf{q}_{A_{j'}}$ $(j, j' = 1, 2)$ is

$$W_{jj'} = \text{const} \exp(-R_j/D), \quad (15)$$

where $R_j = \min \mathcal{R}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}]$. The minimum is calculated for trajectories that start from the jth attractor with coordinate $\mathbf{q}_{A_j}$ for $t \to -\infty$ and approach the saddle point $\mathbf{q}_S$ for $t \to \infty$. One can draw an analogy between expression (15) and the familiar Arrhenius expression for the reaction rate in thermal equilibrium [47] by again associating D with $k_B T$, in which case R_j can be associated with the activation energy. In what follows we therefore call R_j the effective activation energy for switching.

B. Extinction rate

As mentioned above, an important class of problems described by systems with delay are problems of population dynamics. For example, in the problem of infection transfer, the delay can be the latent period after which an infected person becomes contagious. At the mean-field (fluctuation-free) level dynamics of large populations is often modeled by equations of the type (1) without the noise term, cf. Refs. 15, 16, and 48.

We will assume that the population that goes extinct is described by the dynamical variable q_1. In the absence of noise, extinction occurs where the system stays on the invariant hyperplane $q_1 = 0$. The system trajectories are limited to the halfspace $q_1 \geq 0$. There is no force normal to the hyperplane, $K_1 = 0$ for $q_1 = 0$. The stationary state $\mathbf{q}_S$ lies on the hyperplane $q_1 = 0$. This is a saddle point. The variable q_1 lies in the unstable direction transverse to the hyperplane. For motion in the hyperplane $q_1 = 0$, point $\mathbf{q}_S$ is an attractor.

The effect of noise on extinction is a delicate issue. One has to distinguish two major sources of the noise. The noise in the population system can come from external sources, like weather fluctuations, but it also comes from the very fact that populations are discrete and in an elementary event (birth, death, infection) they change by integral numbers rather than continuously. The elementary events themselves happen at random, and in the simplest models are described by rates. For small fluctuations this description can be mapped onto a stochastic equation of the type of (1), but large rare fluctuations are not adequately described by this approach, cf. Refs. 49 and 50. In this paper we will assume that the noise comes from external sources and will consider the problem of extinction of a certain population group, for example, disease extinction.

Of interest is the situation where, once the population has gone extinct, it is not re-created by the noise. For example, if the infectious disease has been extinct, it does not re-emerge due to weather fluctuations. Then, $G_{1n} = 0$ for $q_1 = 0$, and in particular $G_{1n}(\mathbf{q}_S) = 0$. Quite generally, near $\mathbf{q}_S$ (and near the whole plane $q_1 = 0$) the appropriate components of the random force intensity $G_{1n}G_{m1}$ should be linear in q_1; this is also the case in the models where the noise is used to describe the effect of population discreteness [51]. Then $G_{1n} \propto q_1^{1/2}$. This behavior can be also understood if one thinks that the noises acting on different individuals are independent, and the resulting noise is a sum of these noises, with intensity proportional to the population, i.e., to q_1.

The rate of extinction is determined by the probability per unit time of reaching the hyperplane $q_1 = 0$ and staying there for the delay time τ. In this case there is no regular force that would drive the system away from this hyperplane, as $K_1\big(\mathbf{q}(t), \mathbf{q}(t-\tau)\big) = 0$ for $q_1(t) = q_1(t-\tau) = 0$. Also, there is no noise that would drive the system away from the hyperplane. In the considered model the rate of extinction is much smaller than τ^{-1}.

In the above definition, the probability per unit time is coarse-grained over time $\gtrsim \tau$.

Even though in the mean-field picture the state $\mathbf{q}_S$ is similar to a saddle point, the vanishing of the components $G_{1n}(\mathbf{q})$ for $q_1 = 0$ leads to a major difference of the extinction problem from the switching problem. Indeed, as seen from Eq. (12), the optimal random force decays to zero for $t \to \infty$ if $q_1 \to 0$ even where $\chi_1(t)$ remains finite. And in fact, as we now show, $\chi_1(t)$ does remain finite for $t \to \infty$.

We first note that the variational equations for $\mathbf{q}(t), \boldsymbol{\chi}(t), \mathbf{f}(t)$ have a stationary solution with $\mathbf{q} \to \mathbf{q}_S, \mathbf{f} \to \mathbf{0}$, but with $|\boldsymbol{\chi}| > 0$. Indeed, if $\mathbf{q}(t) \to \mathbf{q}_S$ and $\mathbf{f}(t) \to \mathbf{0}$ for $t \to \infty$, then $\partial_{q_1} G_{1n} \propto q_1^{-1/2}$ diverges. Then one can see using Eq. (12) that, for $t \to \infty$, Eq. (11) has a stationary solution $\chi_{n>1}(t) \to 0$, whereas χ_1 is given by equation

$$
\chi_{1S} = -2D\left(\mathcal{K}_{11}^{(1)} + \mathcal{K}_{11}^{(2)}\right)
$$
$$
\times \left\{ \partial_{q_1}\left[\hat{G}(\mathbf{q}) \int dt\, \hat{\phi}(t) \hat{G}^\dagger(\mathbf{q}) \right]_{11} \right\}^{-1}_{\mathbf{q} \to \mathbf{q}_S}. \quad (16)
$$

One can now consider the hypersurface $(q_1 = 0, \chi_{n>1} = 0)$, which contains the state $(\mathbf{q}_S, \boldsymbol{\chi} = \mathbf{0})$. As seen from Eq. (12), on this hypersurface $\mathbf{f} = 0$, since $G_{1n} = 0$. Therefore the system motion is described by noise-free equation $\dot{\mathbf{q}} = \mathbf{K}$, and thus the system that starts with $q_1 = 0$ will asymptotically approach $\mathbf{q}_S$. Also, since $K_1 = 0$ on the hyperplane $q_1 = 0$, from Eq. (11) $\chi_{n>1}(t) = 0$ for $t > 0$, if $\chi_{n>1}(0) = 0$. One can also see that $\dot{\chi}_1/\chi_1 < 0$ for $\mathbf{q}$ close to $\mathbf{q}_S$, i.e., χ_1 asymptotically goes to zero for $t \to \infty$. Thus the hypersurface $(q_1 = 0, \chi_{n>1} = 0)$ is the stable manifold of the state $(\mathbf{q}_S, \boldsymbol{\chi} = \mathbf{0})$. Since this manifold does not contain the state $\mathbf{q}_A$, there are no optimal trajectories that would take the system from $\mathbf{q}_A$ to $\mathbf{q}_S$ with $\chi_1 = 0$.

This means that, in the extinction problem, the optimal trajectory approaches the extinction hyperplane $q_1 = 0$ with a nonzero generalized momentum χ_1 in the direction normal to the hyperplane. This is similar to what happens in a model white-noise driven system without delay [51]. With this boundary condition, the rate of extinction becomes

$$
W_e = \text{const} \times \exp(-R_e/D), \quad (17)
$$

where $R_e = \min \mathcal{R}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}]$ and the minimum is calculated for trajectories that start from the stable state $\mathbf{q}_A$ for $t \to -\infty$ and approach the extinction state $\mathbf{q}_S, \chi_{n>1} = 0, \chi_1 = \chi_{1S}$ for $t \to \infty$. We will call R_e the effective activation energy of extinction.

IV. SPECIAL CASES

A. White-noise driven systems

The general formulation simplifies in the case where the noise $\mathbf{f}(t)$ is white. Without loss of generality we can

set the noise correlation matrix to be proportional to the identity matrix, $\hat{\phi}(t) = 2D\mathcal{I}\delta(t)$; the difference between the noise component amplitudes can be incorporated into matrix $\hat{G}$ in the equation of motion of the system (1). The relation (12) between $\mathbf{f}(t)$ and the Lagrange multiplier $\boldsymbol{\chi}(t)$ becomes local in time, $\mathbf{f}(t) = 2\hat{G}^\dagger\big(\mathbf{q}(t)\big)\boldsymbol{\chi}(t)$. The Hamiltonian formulation of the variational problem for optimal paths is particularly convenient in this case, as the effective Hamiltonian in Eq. (13) becomes

$$\mathcal{H}[\mathbf{q}, \boldsymbol{\chi}] = \left[\hat{G}^\dagger\big(\mathbf{q}(t)\big)\boldsymbol{\chi}(t)\right]^2 + \boldsymbol{\chi}(t)\mathbf{K}\big(\mathbf{q}(t), \mathbf{q}(t-\tau)\big). \tag{18}$$

Because $\mathcal{H}[\mathbf{q}, \boldsymbol{\chi}]$ has delay, the equations of motion for $\mathbf{q}(t)$ and $\boldsymbol{\chi}(t)$ remain acausal.

In the extinction problem, it is important that the random force vanishes at the extinction state. Therefore we consider the white noise in the Ito sense, $\langle \hat{G}\big(\mathbf{q}(t)\big)\mathbf{f}(t)\rangle = \mathbf{0}$.

For white-noise driven systems without delay, a convenient approach to the analysis of large rare fluctuations is based on writing the Fokker-Planck equation for the probability density $\rho(\mathbf{q})$ and solving this equation in the WKB approximation [45]. It gives the same result as the functional approach described above and is often more convenient when one is calculating the prefactor in $\rho(\mathbf{q})$. However, the WKB method cannot be immediately extended to systems with delay, because there is no known closed-form Fokker-Planck equation for the probability density of a transition from $\mathbf{q}'$ at time t' to $\mathbf{q}$ at time t, cf. [36], which is a consequence of the phase space of the system being infinite-dimensional. Our approach is free from this limitation. Moreover, it applies to non-white noise and can be extended to non-Gaussian noise as well.

B. Perturbation theory

Because the equations for optimal paths in systems with delay are acausal, solving them is more complicated than for systems without delay. However, if the delay τ is small, we can consider its effect by perturbation theory, assuming that the variational trajectories $\mathbf{q}(t), \boldsymbol{\chi}(t)$, and $\mathbf{f}(t)$ for $\tau > 0$ are close to the corresponding trajectories for $\tau = 0$. The overall effect of the delay on the rates of rare events will not necessarily be small since, in the expressions for the rates, the corrections to the effective action R [see Eqs. (9), (15), and (17)] are divided by a small factor, the noise intensity.

We assume that the nonlinear function $\boldsymbol{K}$ is smooth, so that if the trajectory $\mathbf{q}(t-\tau) \equiv \mathbf{q}_\tau$ is close to $\mathbf{q}(t)$, the change of $\mathbf{K}$ is small, $||\boldsymbol{K}(\mathbf{q}, \mathbf{q}) - \boldsymbol{K}(\mathbf{q}, \mathbf{q}_\tau)|| \lesssim \varkappa ||\mathbf{q} - \mathbf{q}_\tau||$, where $\varkappa$ is a positive constant. We then write the variational functional as

$$\mathcal{R}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}] = \mathcal{R}^{(0)}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}] + \mathcal{R}^{(1)}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}], \tag{19}$$

where

$$\mathcal{R}^{(0)}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}] = \frac{1}{4}\int \mathbf{f}(t)\hat{\mathcal{F}}(t-t')\mathbf{f}(t')dtdt'$$
$$+ \int \boldsymbol{\chi}(t)[\dot{\mathbf{q}} - \mathbf{K}(\mathbf{q}, \mathbf{q}) - \hat{G}(\mathbf{q})\mathbf{f}(t)]dt,$$
$$\mathcal{R}^{(1)}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}] = \int \boldsymbol{\chi}(t)[\mathbf{K}(\mathbf{q}, \mathbf{q}) - \mathbf{K}(\mathbf{q}, \mathbf{q}_\tau)]dt. \tag{20}$$

The first-order correction to $\min \mathcal{R}$ can be found by evaluating the value of $\mathcal{R}^{(1)}$ along the variational trajectory $\mathbf{q}^{(0)}(t), \boldsymbol{\chi}^{(0)}(t), \mathbf{f}^{(0)}(t)$ that minimizes $\mathcal{R}^{(0)}$. For example, the effective activation energy R_j of switching from the jth attractor, $\mathbf{q}_A \equiv \mathbf{q}_{A_j}$, is $R_j \approx R_j^{(0)} + R_j^{(1)}$, where $R_j^{(0)} = \min \mathcal{R}^{(0)}[\mathbf{q}, \mathbf{f}, \boldsymbol{\chi}]$, and to the first order in the delay

$$R_j^{(1)} \approx \mathcal{R}^{(1)}[\mathbf{q}^{(0)}, \mathbf{f}^{(0)}, \boldsymbol{\chi}^{(0)}]$$
$$\approx \tau \int dt \boldsymbol{\chi}^{(0)}(t)[(\dot{\mathbf{q}}\,\partial_{\mathbf{q}'})\mathbf{K}(\mathbf{q}, \mathbf{q}')]_{\mathbf{q}=\mathbf{q}'=\mathbf{q}^{(0)}(t)}. \tag{21}$$

We note that the first line here gives the correction to R_j not only in the case of small delay τ, but also where the term that contains $\mathbf{q}(t-\tau)$ in $\mathbf{K}$ has a small coefficient. Similar expressions describe corrections $R_e^{(1)}$ and $R^{(1)}(\mathbf{q})$ to the activation energies of extinction and reaching a state $\mathbf{q}$ far from the attractor $\mathbf{q}_{A_j}$.

The procedure can be extended to obtain higher-order correction in τ. In particular, second-order corrections to the switching rate were obtained earlier for an underdamped system with distributed delay [38], where the first-order correction is zero.

Sometimes corrections $R_j^{(1)}$ or $R_e^{(1)}$ diverge, which indicates the fragility effect [52–54]: a small (but finite) perturbation leads to an exponentially large change of the rate of the rare event that is independent of the perturbation. Formally, the divergence occurs, because $R^{(1)}$ accumulates as $t \to \infty$, which happens near the saddle point in the problem of switching or near the extinction state in the problem of extinction.

An important and nontrivial observation based on Eq. (20) is that delay does *not* lead to the fragility. Since $\mathbf{K}(\mathbf{q}, \mathbf{q}) - \mathbf{K}(\mathbf{q}, \mathbf{q}_\tau)$ exponentially decays as the system approaches the saddle point in the switching problem or the extinction state in the extinction problem, generally the value of the functional $\mathcal{R}^{(1)}$ does not diverge for small $\tau > 0$.

C. Direct variational method

An approximate method that may be helpful in finding the effective activation energies is the direct variational method. We assume that the matrix $\hat{G}(\mathbf{q})\hat{G}^\dagger(\mathbf{q})$ is non-degenerate. By construction, it is symmetric, since the matrix $\hat{G}(\mathbf{q})$ is real. In the simple case of white noise, using matrix $\hat{G}(\mathbf{q})\hat{G}^\dagger(\mathbf{q})$ one can directly express $\boldsymbol{\chi}(t)$ in

terms of $\mathbf{q}(t)$,

$$\boldsymbol{\chi}(t) = \frac{1}{2} \left[\hat{G}(\mathbf{q}(t)) \hat{G}^{\dagger}(\mathbf{q}(t)) \right]^{-1} [\dot{\mathbf{q}} - \mathbf{K}(\mathbf{q}(t), \mathbf{q}(t-\tau))]. \tag{22}$$

With this solution, the variational functional $\mathcal{R}$ becomes

$$\mathcal{R} = \frac{1}{4} \int dt [\dot{\mathbf{q}} - \mathbf{K}(\mathbf{q}(t), \mathbf{q}(t-\tau))] \left[\hat{G}(\mathbf{q}(t)) \hat{G}^{\dagger}(\mathbf{q}(t)) \right]^{-1}$$
$$[\dot{\mathbf{q}} - \mathbf{K}(\mathbf{q}(t), \mathbf{q}(t-\tau))] \tag{23}$$

This functional could be also obtained directly from the functional $\mathcal{R}_{\mathbf{f}}[\mathbf{f}]$, Eq. (7), by substituting $\mathbf{f}(t)$ in terms of $\mathbf{q}(t)$ from the equation of motion (1). In the problem of reaching a given state $\mathbf{q}$ starting from $\mathbf{q}_A$ one should integrate in Eq. (23) from $t \to -\infty$ to $t = 0$, whereas in the problem of switching or extinction (reaching $\mathbf{q}_{\mathcal{S}}$) the upper limit of integration over time is $t \to \infty$. We note that, in the first case, $\boldsymbol{\chi}(t)$ as given by Eq. (22) is discontinuous at $t = 0$, as $\boldsymbol{\chi}(t) = \mathbf{0}$ for $t > 0$, in agreement with the general condition on $\boldsymbol{\chi}(t)$ discussed in Sec. III.

By construction, functional (23) is non-negative. To minimize it one can use a direct variational method by taking a trial trajectory $\mathbf{q}(t)$ that goes from $\mathbf{q}_A$ to the target state and minimizing $\mathcal{R}$ with respect to the parameters of the trajectory. We illustrate this approach below.

D. Vicinity of a bifurcation point

The general analysis significantly simplifies in the important parameter range where the system is close to a bifurcation point. We will consider two types of bifurcation points: a saddle-node bifurcation, where the attractor merges with a saddle point, and a very similar transcritical bifurcation commonly considered in the extinction problem, where the attractor merges with the extinction state. Near a bifurcation point the motion is slowed down, there emerges a soft mode (a slow variable) [55, 56]. Therefore the effect of delay is suppressed, generally speaking. As the system approaches a bifurcation point the delay time τ becomes increasingly smaller than the relaxation time of the system. Moreover, the noise becomes effectively white, as the noise correlation time is small compared to the relaxation time, too. However, there is a parameter range where the delay time exceeds the relaxation time of the fast variables, and yet has not become too short compared to the relaxation time of the slow variable. It is this range that we will discuss.

Quite generally, after the "fast" variables have approached their stable state for a given value of the slow variable q, the equation of motion for this slow variable has the form

$$\dot{q}(t) = K(q(t), q(t-\tau)) + G(q(t))f(t),$$
$$K(q, q) = -\eta + q^2, \qquad \langle f(t)f(t') \rangle = 2D\delta(t-t'). \tag{24}$$

Parameter η gives the distance to the saddle-node bifurcation point in the absence of delay. For $\eta > 0$ the system has a stable state at $q_A = -\eta^{1/2}$ and a saddle at $q_{\mathcal{S}} = \eta^{1/2}$. For $\eta = 0$ these states merge, and for $\eta < 0$ there are no stationary states in the range of small $|q|$. Close to the bifurcation point $|\eta| \ll 1$. For a transcritical bifurcation one has $K(q, q) = rq - q^2$, in which case for $r > 0$ the system has a stable state $q_A = r$ and a saddle point $q_{\mathcal{S}} = 0$, whereas for $r < 0$ in the range $q \geq 0$ the system has only a stable stationary state $q = 0$. By changing $q \to q + r/2, r \to 2\eta^{1/2}, K \to -K$ the equation of motion near a transcritical bifurcation can be brought into the same form as the equation of motion near a saddle-node bifurcation point. Therefore in what follows we use Eq. (24) to consider both cases.

In the analysis of the switching problem, where $G(q)$ is generally non-singular, taking into account that the characteristic scale of q is small, one can set $G(q) = 1$. Then quite generally, to the first order in τ we have the following expressions for the optimal trajectory and the activation energy of switching:

$$q_{\mathrm{sw}}^{(0)}(t) = \eta^{1/2} \frac{e^{2t\sqrt{\eta}} - 1}{e^{2t\sqrt{\eta}} + 1},$$
$$R_{\mathrm{sw}} \approx \frac{4}{3}\eta^{3/2} - \tau \int_{-\sqrt{\eta}}^{\sqrt{\eta}} dq K(q, q) \left[\partial_{q'} K(q, q') \right]_{q'=q}, \tag{25}$$

here we used Eq. (21) and took into account that $\chi_{\mathrm{sw}}^{(0)}(t) = -K(q_{\mathrm{sw}}^{(0)}(t), q_{\mathrm{sw}}^{(0)}(t))$. Interestingly, depending on the form of the K-function, the scaling of the τ-dependent correction with the distance to the bifurcation point η can be different from that of the main term. If $K(q(t), q(t-\tau))$ is linear in $q(t-\tau) - q(t)$, the scaling of the correction and the main term is the same, $\eta^{3/2}$; if the expansion of K starts with $[q(t-\tau) - q(t)]^2$, the correction to R_{sw} dies out with decreasing η as $\eta^{5/2}$.

For the extinction problem, as explained above, $G(q) = (\eta^{1/2} - q)^{1/2}$. In this case we obtain the same trajectory in the absence of delay as in the problem of switching, $q_{\mathrm{e}}^{(0)}(t) = q_{\mathrm{sw}}^{(0)}(t)$, whereas the parameter $\chi^{(0)}$ for the extinction problem is $\chi_{\mathrm{e}}^{(0)}(t) = -K(q_{\mathrm{sw}}^{(0)}(t), q_{\mathrm{sw}}^{(0)}(t))/G^2(q_{\mathrm{e}}^{(0)}(t))$. Then, with account taken of Eq. (21),

$$R_{\mathrm{e}} \approx 2\eta - \tau \int_{-\sqrt{\eta}}^{\sqrt{\eta}} dq \frac{K(q, q)}{\sqrt{\eta} - q} \left[\partial_{q'} K(q, q') \right]_{q'=q}. \tag{26}$$

The leading-order term in the activation energy (26) scales as $r^2 \propto \eta$ with the distance to the transcritical bifurcation point.

The simple expressions (25) and (26) give the activation energies for the switching and extinction problems in the explicit form, to the lowest order in delay. They are particularly important in terms of applications, since the switching rates near bifurcation points are comparatively large and are easier to access in the experiment.

V. SWITCHING INDUCED BY ADDITIVE WHITE NOISE IN A MODEL SYSTEM

In this and the following section, we apply the general results to simple dynamical model systems driven by additive and multiplicative noise. We compare the results of the small-τ perturbation theory with the results of variational calculations as well as Monte Carlo simulations. Each data point in the Monte Carlo figures is the mean value taken over 2000 simulations. The heights of the symbols serve as the error bars. The numerical algorithms are described in Appendix.

We consider two model systems with delayed dynamics. Both are described by one dynamical variable q. The systems are assumed to be driven by white Gaussian noise. In this section we assume that the noise is additive, $G(q) = 1$. The difference between the models is in the form of the delay in the regular force, which is equal to K_{attr} or K_{rpl},

$$K_{\text{rpl}}\big(q(t), q(t-\tau)\big) = -\gamma q(t) - q^2(t) + q(t-\tau),$$
$$K_{\text{attr}}\big(q(t), q(t-\tau)\big) = -\gamma q(t-\tau) - q^2(t) + q(t), \quad (27)$$

with $0 < \gamma < 1$. For the both forms of the force, in the absence of noise and delay the system has a stable state $q_A = 1 - \gamma$ and a saddle point $q_S = 0$. We use subscripts "rpl" and "attr" to indicate that, in the corresponding model, delay is in the term that describes, respectively, repulsion and attraction to the saddle point $q = 0$. In the presence of delay, for K_{attr} the characteristic equation (3) for the saddle point along with a positive root $\alpha > 0$ has also a negative root $\alpha \approx \tau^{-1} \log \gamma\tau$ for $\gamma\tau \ll 1$. This negative root does not change the results on the switching rate for small $\gamma\tau$, the optimal path corresponds to the generalized momentum $\chi(t)$ varying along the eigenvector that corresponds to the positive root.

For $\tau = 0$ the effective Hamiltonian $\mathcal{H}[q, \chi] \equiv \mathcal{H}^{(0)}[q, \chi]$, see Eq. (13), for the both models becomes local in time,

$$\mathcal{H}^{(0)}[q, \chi] = \chi K(q, q) + \chi^2, \quad K(q, q) = (1 - \gamma)q - q^2. \quad (28)$$

Since the value of $\mathcal{H}^{(0)}$ is conserved along the zero-delay variational trajectory $q^{(0)}(t), \chi^{(0)}(t)$, and this trajectory starts for $t \to -\infty$ from $q \to q_A, \chi \to 0$, we have $\mathcal{H}^{(0)} = 0$. Then the zero-delay Hamiltonian trajectory is

$$q^{(0)}(t) = q_A/[1 + \exp(q_A t)],$$
$$\chi^{(0)}(t) = -q_A^2 \exp(q_A t)/[1 + \exp(q_A t)]^2. \quad (29)$$

Clearly, $q^{(0)}(t) \to q_S, \chi^{(0)}(t) \to 0$ for $t \to \infty$, so that Eq. (29) indeed describes the most probable path followed in switching.

From Eq. (13), the delay-induced correction to the effective Hamiltonian $\mathcal{H}^{(1)}[q, \chi] = \chi(t)[K\big(q(t, q(t-\tau)\big) - K\big(q(t), q(t)\big)]$ for the two models becomes, respectively, $\mathcal{H}^{(1)}_{\text{rpl}} = -\chi(t)[q(t) - q(t-\tau)]$ and $\mathcal{H}^{(1)}_{\text{attr}} = \gamma\chi(t)[q(t) - q(t-\tau)]$. The first-order correction to the switching activation energy is given by $-\int \mathcal{H}^{(1)}[q^{(0)}(t), \chi^{(0)}(t)]dt$, cf. Eq. (21).

From Eqs. (20) and (21) we find the zeroth-order activation energy $R_{\text{sw}}^{(0)}$ of switching from the stable state q_A and the delay-induced corrections $R_{\text{sw}}^{(1)}$ for the two models as

$$R_{\text{sw}} \approx R_{\text{sw}}^{(0)} + R_{\text{sw}}^{(1)}, \qquad R_{\text{sw}}^{(0)} = \frac{1}{6}(1 - \gamma)^3,$$
$$R_{\text{sw,rpl}}^{(1)} = \tau R_{\text{sw}}^{(0)}, \qquad R_{\text{sw,attr}}^{(1)} = -\gamma\tau R_{\text{sw}}^{(0)}. \quad (30)$$

Equation (30) shows that, where the delay is in the term that pushes the system away from the saddle point, the delay leads to an increase of the activation energy of switching from the stable state and thus to a decrease in the switching rate. In the opposite case, where the delay is in the term that pushes the system towards the saddle point, the effect is opposite. This can be understood by noticing that the delay-induced change of K is $\propto q(t-\tau) - q(t)$. The latter quantity is positive when the system moves from q_A to q_S. The corresponding change of the regular force has to be overcome by the noise as it drives the system to q_S. In the model with K_{rpl} a stronger noise is needed than without delay, whereas in the model with K_{attr} the needed noise is weaker. The corresponding noise realizations are exponentially less/more probable.

For the models (27), the general approach to finding a correction to the activation energy to first order in τ can be checked by expanding the delayed term in Eq. (27) as $q(t - \tau) \approx q(t) - \tau\dot{q}(t)$. Then the equations of motion for the cases of the delay in the repulsing and attracting terms become, respectively, $(1 + \tau)\dot{q}(t) = K\big(q(t), q(t)\big) + f(t)$ and $(1 - \gamma\tau)\dot{q}(t) = K\big(q(t), q(t)\big) + f(t)$. Dividing these equations by $1 + \tau$ and $1 - \gamma\tau$, respectively, we reduce them to the standard Langevin form with no delay and with the rescaled regular force and the noise intensity. One can easily see that the resulting activation energy has the form (30). However, this method runs into obvious problems if one tries to find higher-order corrections in τ. For example, if $f(t)$ is white noise, the derivative of $\dot{q}$ over time is ill-defined. In contrast, the perturbation theory of Sec. IV B allows finding higher-order corrections in τ directly.

In Fig. 1 we present the τ dependences of the activation energy R and switching times T_{sw} for the models (27). The analytical results (lines) calculated by the first order perturbation theory (30), are compared with the results of the direct variational method and Monte Carlo simulations. In the variational method, we used a trajectory of the form $q_A/[1 + \exp(\lambda t)]$ and minimized R with respect to parameter λ. The variational results for switching activation energies are in excellent agreement with our perturbation theory in the range of small to moderately small τ, which is of interest for the present paper. The results of Monte Carlo simulations in Fig. 1 (b) are discussed in the next section.
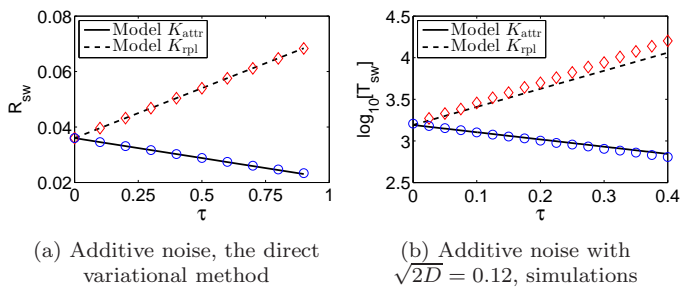
(a) Additive noise, the direct variational method

(b) Additive noise with $\sqrt{2D} = 0.12$, simulations

FIG. 1: The activation energy of switching R_{sw} (a) and the mean switching time (dwell time) T_{sw} (b) as functions of the delay τ for switching due to additive white noise in the models (27). The first order perturbation theory in τ is shown as solid and dashed lines. Numerical calculations in (a) were done using the direct variational method (circles and diamonds), while (b) presents the results of the Monte Carlo simulations (circles and diamonds). Parameter γ is set at $\gamma = 0.4$.

A. Monte Carlo simulations

In this section, we compare the theoretical results with numerical simulations for switching. For the models (27), we looked for the mean switching time T_{sw} as the white additive noise drives the system out of the basin of attraction of q_A to $q < q_{\mathcal{S}}$ (in our system $q_A > q_{\mathcal{S}}$). This is the dwell time of staying in the attraction basin. To insure that a trajectory doesn't drift back to the original basin we require that it goes well past $q_{\mathcal{S}}$, so that the probability of returning to the basin of attraction is exponentially small. More specifically, we define the switching to have happened when $q < -0.2$ and find that this threshold is consistent with the no-return requirement for the noise intensities and the delay times we studied. We then calculate T_{sw} as the mean first passage time (MFPT) to $q = -0.2$ given that the system starts from q_A.

(a) Model K_{attr}, additive noise

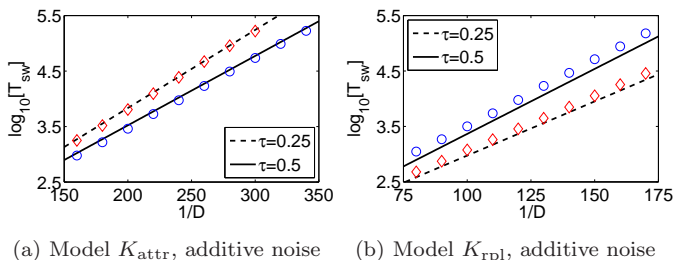(b) Model K_{rpl}, additive noise

FIG. 2: Log base 10 of the mean switching time vs. reciprocal noise intensity for the models given by Eq. (27) with $\gamma = 0.4$. The solid and dashed lines represent the theory from Eqs. (30) and (31), while the data points are the results of the simulations

Switching events are Poisson-distributed, and the switching time is simply the inverse of the switching rate. Therefore from Eq. (15)

$$T_{\mathrm{sw}} = c_{\mathrm{sw}} \exp(R_{\mathrm{sw}}/D), \qquad c_{\mathrm{sw}} \approx 2\pi/(1-\gamma), \quad (31)$$

where R_{sw} is the activation energy of switching in Eq. (30). For the pre-factor c_{sw} we used the result of the Kramer's theory [57] in which there is no delay, since the major effect of the delay for weak noise is in the logarithm of the MFPT, R_{sw}/D.

Figure 1 (b) shows the results of the Monte Carlo simulations for the logarithm of T_{sw} as a function of delay τ. There are no free parameters in the plot. In the model where the regular force is K_{attr}, so that the delay is in the term that pushes the system toward the saddle point, there is excellent agreement between the theory and numerical simulations throughout the whole considered range of τ. For the model with K_{rpl}, where delay pushes the system away from the saddle point, the theory and simulations are very close for $\tau < 0.2$. For larger τ the discrepancy is somewhat larger, which indicates that, for this model, higher-order delay-induced corrections become important. Clearly, the delay has opposite effect in the two models: it destabilizes the system in model K_{attr}, leading to the decrease of T_{sw}, and stabilizes the system in model K_{rpl}, in agreement with the arguments given above.

A major prediction of the theory is the exponential dependence of the mean switching time on the reciprocal noise intensity D. In Fig. 2, we plot the logarithm of the mean switching time as a function of D^{-1} for both models in Eq. (27) obtained for the delays $\tau = 0.25$ and $\tau = 0.5$. Here, too, we have excellent agreement of the theory and simulations.

VI. EXTINCTION IN A MODEL SYSTEM

We now consider the time to extinction. We will use the same models of noise-free motion, Eq. (27). Without delay and noise, the equation of motion for these models $\dot{q} = K(q, q)$ with $K = (1-\gamma)q - q^2$, cf. Eq. (27), can be thought of as a logistic equation with linear feedback. Such an equation is often used to describe continuous population model. In the population dynamics context, the state $q_{\mathcal{S}} = 0$ corresponds to the extinction state. As argued in Sec. III B, of interest in this case is multiplicative noise. We assume the noise to be δ-correlated and choose the factor $G(q)$ that determines the noise strength in the standard for population dynamics form $G(q) = \sqrt{q}$. As mentioned previously, we understand the noise in the Ito sense, in this case.

From Eq. (18), the effective Hamiltonian that describes the most likely trajectory of the system followed in extinction is a sum of the term $\mathcal{H}^{(0)}$ without delay, which is the same for the both dynamical models and, depending on the model, the terms $\mathcal{H}^{(1)}_{\mathrm{rpl}}$ or $\mathcal{H}^{(1)}_{\mathrm{attr}}$ that come from the delay,

$$\mathcal{H}^{(0)}[q, \chi] = q\chi^2 + \chi K(q, q),$$
$$\mathcal{H}^{(1)}_{\mathrm{rpl}}[q, \chi] = \chi(t)[q(t-\tau) - q(t)],$$
$$\mathcal{H}^{(1)}_{\mathrm{attr}}[q, \chi] = -\gamma\chi(t)[q(t-\tau) - q(t)]. \quad (32)$$

As discussed in Sec. III B, and with account taken of Eq. (16), the Hamiltonian trajectory leading to extinction satisfies boundary conditions $q(t) \to q_A = 1 - \gamma, \chi(t) \to 0$ for $t \to -\infty$, and $q(t) \to q_S, \chi(t) \to -(1-\gamma)$ for $t \to \infty$.

If we disregard delay and use $\mathcal{H}^{(0)}$ as the Hamiltonian, the Hamiltonian trajectory $q^{(0)}(t)$ is described by Eq. (29). From condition $\mathcal{H}^{(0)} = 0$ we have $\chi^{(0)}(t) = q^{(0)}(t) - (1-\gamma)$. Using these expressions we obtain from Eqs. (21) and (32) the activation energy of extinction R_e as a sum of the term that describes switching in the absence of delay $R_e^{(0)}$ and a correction of the first order in the delay τ. We denote the corrections for the models (27) as $R_{e,rpl}^{(1)}$ and $R_{e,attr}^{(1)}$,

$$R_e \approx R_e^{(0)} + R_e^{(1)}, \qquad R_e^{(0)} = \frac{1}{2}(1-\gamma)^2,$$
$$R_{e,rpl}^{(1)} \approx \tau R_e^{(0)}, \qquad R_{e,rpl}^{(1)} \approx -\gamma \tau R_e^{(0)} \qquad (33)$$

This result can be again independently checked by expanding $q(t - \tau)$ in Eq. (27) to the first order in τ, as discussed above.
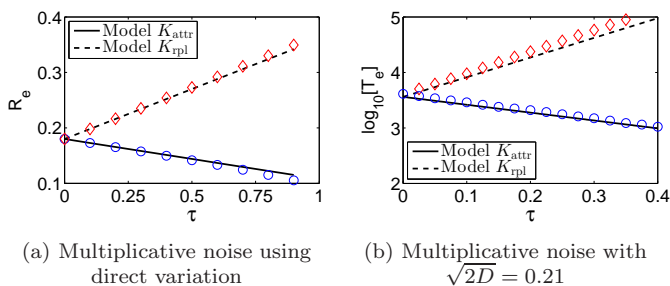

(a) Multiplicative noise using direct variation

(b) Multiplicative noise with $\sqrt{2D} = 0.21$

FIG. 3: The activation energy of extinction R_e (a) and the mean extinction time T_e (b) as functions of the delay τ for extinction due to multiplicative white noise in the models (27). The first order perturbation theory in τ is shown as solid and dashed lines. Numerical calculations in (a) were done using the direct variational method (circles and diamonds), while (b) presents the results of the Monte Carlo simulations (circles and diamonds). Parameter γ is set at $\gamma = 0.4$.

We performed a direct variational calculation of R_e using the same variational trajectory as in the switching problem. The results are shown in Fig. 3(a). They are in excellent agreement with the perturbation theory (33) in the studied range of τ. As seen from (see Fig. 1 (a)) and Fig. 3 (a), the comparison is equally good for the switching and the extinction problems.

A. Monte Carlo simulations

For numerical simulations based on the Ito calculus, we used the Milstein method [58], which consistently takes into account that $\langle G(q(t))f(t)\rangle = 0$, see Appendix. To find the mean time of extinction in the presence of delay,

one has to make sure that, after the system has reached the extinction state in a simulation, it will not leave it. The possibility to leave is particularly clear for the model (27) with K of the form of K_{rpl}. Indeed, if the system has been brought by the noise to $q = 0$ and after that the noise becomes equal to zero (or just very small), the system will move away from $q = 0$ toward the attractor, because it is driven by the force $q(t - \tau) > 0$.

From the above argument, one cannot use the MFPT to reach $q = 0$ as the measure of the reciprocal extinction rate. The system has to stay at $q = 0$ for the time equal at least to the delay time. This makes simulations significantly different from the conventional MFPT simulations.

To make a quantitative comparison of the theory and the simulations we used the prefactor in the mean time to extinction T_e, which was calculated in the absence of delay. If there is no delay and the noise is white, T_e is the MFPT for reaching the extinction state from a point q in the vicinity of the attractor q_A. The equation for the MFPT $T(q)$ in our model reads $Dq\partial_q^2 T + K(q,q)\partial_q T = -1$, cf. Ref. [59]. The solution is

$$T(q) = D^{-1} \int_0^q dq_1 \int_{q_1}^\infty dq_2 (1/q_2) \exp\{[V(q_1) - V(q_2)/D]\}$$

with $V(q) = \frac{1}{2}q^2 - (1-\gamma)q$. For q near q_A, the integrand in the integral over q_2 has a maximum near q_A and is Gaussian near the maximum. The integrand in the integral over q_1 has a maximum at $q_1 = 0$, but it is not Gaussian near the maximum. This is why the result for the prefactor differs from that in the switching problem. Expanding the exponent in this integrand to a linear term in q_1 near $q_1 = 0$, we obtain $T_e = c_e \exp[R_e^{(0)}/D]$ with $c_e = (2\pi D)^{1/2}/(1-\gamma)^2$. In plotting the theoretical data we used

$$T_e = c_e \exp[(R_e/D]$$

with R_e given by Eq. (33) and the above value of c_e, which refers to $\tau = 0$.

We first compare the simulations and the perturbation theory of the mean extinction time T_e as a function of the delay τ. The results are shown in Fig. 3 (b). As in the additive noise case, we see slight disagreement for model K_{rpl} as the delay gets large, but excellent agreement for $\tau < 0.2$. In contrast, model K_{attr} shows excellent agreement for the whole range of τ we have explored. Again, the theoretical curves have no free parameters.

We also considered the effect of noise intensity on the mean extinction times, which is depicted in Fig. 4. Here again we see the characteristic $1/D$ dependence of $\log T_e$, and the results of the theory and the simulations are in excellent agreement for the both models.

VII. CONCLUSIONS

We have considered dynamical systems driven by weak on average noise and studied the effect of delay on small

(a) Model K_{attr},
multiplicative noise
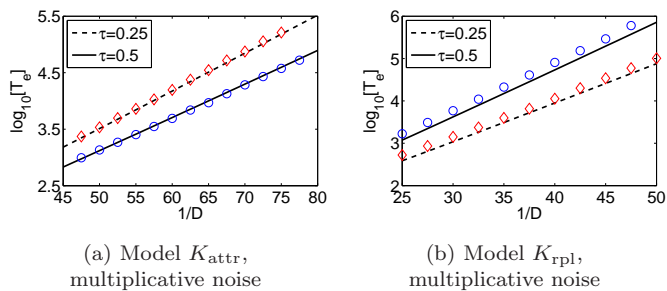
(b) Model K_{rpl},
multiplicative noise

FIG. 4: Log base 10 of the mean extinction time vs. inverse
noise intensity for the models of delayed dynamics in
Eq. (27); $\gamma = 0.4$. The lines represent the theory, Eq. (33),
while the data points are the results of the simulations.

fluctuations about the stable states and on the probabil-
ities of large rare fluctuations. Of central interest was
the effect of delay on the rates of noise-induced switch-
ing between coexisting stable states and noise-induced
extinction understood as reaching a saddle point on the
boundary of the basin of attraction of a stable stationary
state; at this saddle point one of the dynamical variables
(q_1) is zero; we assumed that, even in the presence of
the noise, once this variable has become equal to zero it
remains equal to zero.

In our analysis delay was incorporated into the effective
force that drives the system in the absence of noise. This
force depends not only on the instantaneous values of
the dynamical variables, but also on the values of these
variables a certain time τ earlier. We assumed that the
delay does not destabilize the attractors in the field of
this force compared to the case $\tau = 0$. More generally,
we assumed that it does not lead to new stationary or
periodic states and/or dynamical chaos in the part of the
phase space that contains the $\tau = 0$-attractors and saddle
points as well as the noise-free and the most probable
fluctuational trajectories between the attractors and the
saddle points.

We were interested in fluctuations induced by Gaus-
sian noise. The noise could be additive, in which case
the random force is independent of the dynamical vari-
ables, or multiplicative, where the random force depends
on the instantaneous values of the dynamical variables.
The typical intensity of the noise D was the small param-
eter of the theory. In the analysis of the extinction prob-
lem the noise has to be multiplicative, with its strength
going to zero at $q_1 = 0$, to prevent re-emergence of the
population once it has gone extinct

Using the smallness of D, we showed that small de-
lay does not qualitatively change small-amplitude fluc-
tuations about the stable states. We have obtained the
spectrum of small-amplitude fluctuations.

The analysis of large rare fluctuations was reduced to
a variational problem. The solution of this problem gives
the logarithm of the probability distribution on the tail
and also the exponents of the rates of switching and ex-
tinction. The rates are described by an activation-type

law, with their logarithm being proportional to the recip-
rocal noise intensity D^{-1}. The solution of the variational
problem can be associated with the activation energy of
the corresponding transition (switching or extinction).
The extreme trajectories of the variational problem give
the most probable paths followed by a fluctuation to a
given state or in switching between the stable states or
in extinction. An important part of the formulation is
the boundary conditions for the extreme trajectories. We
found these conditions for systems with delay and, in par-
ticular, found a significant difference in the form of these
conditions for the problems of switching and extinction.

In the presence of delay the variational equations for
the extreme trajectories are acausal: they contain the
past and future values of the dynamical variables, which
are time-shifted by τ. This is somewhat reminiscent
of the situation with variational trajectories in the in-
stanton problems for tunneling with dissipation, cf. [60].
However, in our case the trajectories go in real time.
They are accessible to direct observation in the exper-
iment, as in systems with no delay, cf. [42, 43]. Numer-
ical evaluation of the extreme trajectories based on the
acausal equations of motion requires special tools, since
shooting methods generally do not work. An approach to
the problem which does not rely on the shooting method
was proposed in Ref. 39.

For small delay τ compared to the relaxation time of
the system, the delay-induced corrections to the activa-
tion energies are linear in τ. This is strongly different
from the result for inertial systems with delayed friction
force [38], where the correction to the switching activa-
tion energy studied there was found to be quadratic in
small delay. In our system, the corrections exponentially
strongly affect the rates of switching and extinction, since
they are in the exponents of the expressions for the rates
and are multiplied by a large factor D^{-1}.

Our results also show that, depending on the form of
the equations of motion, delay can increase or decrease
the corresponding rates. We tested this conclusion us-
ing a simple nonlinear model. We found that, in this
model, the results of the perturbation theory in τ agree
in a broad parameter range with the results of the direct
variational method that we employed.

We studied the switching and extinction rates in the
important parameter range where the system is close to
the saddle-node or transcritical bifurcation points. Be-
cause the system slows in the vicinity of equilibria near
bifurcations in this range, it is sufficient to look for the
linear in τ corrections to the activation energies. We
found that both the leading term in the activation en-
ergies and the delay-induced corrections scale as powers
of the distance to the bifurcation point. The exponents
can be the same, or the correction can decrease faster
than the leading-order term as the system approaches
the bifurcation point. The exponents of the leading-order
terms and of the corrections are different in the problems
of switching and extinction.

We carried out careful numerical simulations to test

the theory. A potential pitfall in such simulations is that one has to make sure that the system indeed has switched to another state or indeed has gone extinct. In the presence of delay, it means that one has to check that the delayed force will not pull the system back into the domain of attraction of the initially occupied state. The results of the simulations for the nonlinear models that we employed are in excellent quantitative agreement with the theory, with no adjustable parameters. This refers to both the activation dependence of the rate of the transitions on the noise intensity D and to the dependence of the effective activation energy on τ. In the studied models, this dependence appeared to be linear in a comparatively broad range. However, we found that the range where simulations and the perturbative analytical predictions coincide is narrower where the delay is in the term that repels the system from the saddle point and thus drives it back to the attractor even if the system is already behind the saddle point(in the problem of switching) or has reached the extinction state (in the extinction problem).

Appendix A: Numerical Algorithm

We now describe the numerical algorithms used for a one-dimensional version of Eq. (1) (cf. [58]),

$$\dot{q}(t) = K\big(q(t), q(t-\tau)\big) + G(q(t))f(t). \quad \text{(A1)}$$

We define $f(t)$ as Gaussian noise with intensity $2D$. When computing, we redefine the noise term as $g(q(t))W(t) = G(q(t))f(t)$, with $W(t)$ being Gaussian noise with a standard deviation of unity and $g(q(t)) = \sqrt{2D}\,G(q(t))$.

For additive noise, there is no difference between using Ito or Stratonovich calculus. However, the analysis of extinction requires the Ito formulation. In addition, one has to take into account that, in the extinction problem, $G(q)$ has a singularity at the final point $q = 0$. Our simulations are based on the Milstein method. The Ito representation for the Milstein method is defined as

$$q_{n+1} = q_n + K_n h + g_n \delta W_n + \frac{1}{2} g_n g_n'[(\delta W_n)^2 - h] \quad \text{(A2)}$$

for a step size h and variable notations $q_n = q(t_n)$, $K_n = K\big(q(t_n), q(t_n - \tau)\big)$, $g_n = g(q(t_n))$, $g_n' = dg(q_n)/dq_n$, $\delta W_n \approx \sqrt{h}\,u$, where u is a uniformly distributed random number between 0 and 1.

For the additive noise, $g_n' = 0$ and the algorithm simplifies to

$$q_{n+1} = q_n + K_n h + g_n \sqrt{h}\,u. \quad \text{(A3)}$$

To increase accuracy, we approximated K_n using an improved Euler approach

$$q_{n+1} = q_n + \frac{(K_n + K_{n+1})h}{2} + g_n \sqrt{h}\,u, \quad \text{(A4)}$$

using the approximation $q(t_n + h) \approx q_n + K_n h + g_n \sqrt{h}\,u$. We assume $h < \tau$ and that a pre-history time series of $q(t)$ are stored in increments of τ/h so that the delay term can be quickly accessed.

To take care of the singularity of g_n' at $q_n = 0$, we use the derivative-free Milstein approximation

$$q_{n+1} = q_n + K_n h + g_n \sqrt{h}\,u$$
$$+ \frac{(g_{n+1} - g_n)}{2\sqrt{h}} \left[(\sqrt{h}\,u)^2 - h \right], \quad \text{(A5)}$$

where, by construction, $g_{n+1} = g(q(t_{n+1}))$.

[1] K. Ikeda, K. Otsuka, and K. Matsumoto, Progr. Theor. Phys. Suppl. **99**, 295 (1989).

[2] F. Arecchi, G. Giacomelli, A. Lapucci, and R. Meucci, Phys. Rev. A **43**, 4997 (1991).

[3] T. Heil, I. Fischer, W. Elsasser, J. Mulet, and C. Mirasso, Phys. Rev. Letts. **86**, 795 (2001); T. Heil, I. Fischer, W. Elsasser, B. Krauskopf, K. Green, and A. Gavrielides, Phys. Rev. E **67** (2003).

[4] W. Ray, W. S. Lam, P. N. Guzdar, and R. Roy, Phys. Rev. E **73**, 026219 (2006).

[5] A. L. Franz, R. Roy, L. B. Shaw, and I. B. Schwartz, Phys. Rev. Lett. **99**, 053905 (2007).

[6] A. Uchida, K. Amano, M. Inoue, K. Hirano, S. Naito, H. Someya, I. Oowada, T. Kurashige, M. Shiki, S. Yoshimori, K. Yoshimura, and P. Davis, Nat. Photon. **2**, 728 (2008).

[7] M. Soriano and J. Garcia-Ojalvo, Rev. Mod. Phys. **85**, 421 (2013).

[8] A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto, ArXiv e-prints (2014), arXiv:1407.2871 [quant-ph].

[9] D. Esteve, M. H. Devoret, and J. M. Martinis, Phys. Rev. B **34**, 158 (1986).

[10] H. Grabert and S. Linkwitz, Phys. Rev. A **37**, 963 (1988).

[11] C. M. Marcus and R. M. Westervelt, Phys. Rev. A **39**, 347 (1989).

[12] C. Gupta, J. M. López, R. Azencott, M. R. Bennett, K. Josić, and W. Ott, J. Chem. Phys. **140**, 204108 (2014).

[13] J. M. Cushing, *Integrodifferential Equations and Delay Models in Populations Dynamics.*
Lecture Notes in Biomathematics, vol. 20. (Springer-Verlag, Berlin, 1977).

[14] M. L. Taylor and T. W. Carr, J. Math. Biol. **59**, 841 (2009).

[15] K. B. Blyuss and Y. N. Kyrychko, Bull. Math. Biol. **72**, 490 (2010).

[16] Y. Wang, F. Brauer, J. H. Wu, and J. M. Heffernan, J. Math. Anal. Appl. **414**, 514 (2014).

[17] T. Erneux, *Applied Delay Differentail Equations* (Spinger, 2009).

[18] R. E. Bellman and K. L. Cooke, *Differential-Difference Equations* (Academic Press, New York, 1963).

[19] A. Krasnovkii, *Stabiliity of motion and translation* (Stanford University Press, Stanford, 1963).

[20] J. K. Hale, *Theory of Functional Differential Equations* (Springer-Verlag, New York, 1977).

[21] H. Mori, Progr. Theor. Phys. **33**, 423 (1965).

[22] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).

[23] R. F. Grote and J. T. Hynes, J. Chem. Phys. **73**, 2715 (1980).

[24] E. Pollak, J. Chem. Phys. **85**, 865 (1986).

[25] P. Reimann, Chem. Phys. **268**, 337 (2001).

[26] L. S. Tsimring and A. Pikovsky, Phys. Rev. Lett. **87**, 250602 (2001).

[27] D. Goulding, S. Melnik, D. Curtin, T. Piwonski, J. Houlihan, J. Gleeson, and G. Huyet, Phys. Rev. E **76**, 031128 (2007).

[28] A. Dhar and K. Wagh, EPL **79**, 60003 (2007).

[29] C. Maes, S. Safaverdi, P. Visco, and F. van Wijland, Phys. Rev. E **87**, 022125 (2013).

[30] T. Franosch, M. Grimm, M. Belushkin, F. M. Mor, G. Foffi, L. Forro, and S. Jeney, Nature **478**, 85 (2011).

[31] S. Kheifets, A. Simha, K. Melin, T. C. Li, and M. G. Raizen, Science **343**, 1493 (2014).

[32] A. Donev, T. G. Fai, and E. Vanden-Eijnden, J.Stat. Mech. , P04004 (2014).

[33] S. Guillouzic, I. L'Heureux, and A. Longtin, Phys. Rev. E **59**, 3970 (1999).

[34] S. Guillouzic, I. L'Heureux, and A. Longtin, Phys. Rev. E **61**, 4906 (2000).

[35] T. D. Frank, Phys. Rev. E **66**, 011914 (2002).

[36] T. D. Frank, Phys. Rev. E **71**, 031106 (2005).

[37] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New-York, 1965).

[38] M. I. Dykman and I. B. Schwartz, Phys. Rev. E **86**, 031145 (2012).

[39] B. S. Lindley and I. B. Schwartz, Physica D **255**, 22 (2013).

[40] M. Lakshmanan and D. V. Senthilkumar, *Dynamics of Nonlinear Time-Delay Systems* (Springer, Berlin, 2013).

[41] M. I. Dykman, Phys. Rev. A **42**, 2020 (1990).

[42] J. Hales, A. Zhukov, R. Roy, and M. I. Dykman, Phys. Rev. Lett. **85**, 78 (2000).

[43] H. B. Chan, M. I. Dykman, and C. Stambaugh, Phys. Rev. Lett. **100**, 130602 (2008).

[44] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer-Verlag, N.Y. 1998).

[45] R. Graham, in *Springer Tracts in Modern Physics*, **66**, 1 (1973).

[46] V. N. Smelyanskiy, M. I. Dykman, H. Rabitz, and B. E. Vugmeister, Phys. Rev. Lett. **79**, 3113 (1997).

[47] S. Arrhenius, Z. Phys. Chem. **4**, 226 (1889).

[48] M. L. Taylor and T. W. Carr, J. Math. Biol. **59**, 841 (2009).

[49] C. R. Doering, K. V. Sargsyan, and L. M. Sander, Multiscale Model. Simul. **3**, 283 (2005).

[50] M. Assaf, A. Kamenev, and B. Meerson, Phys. Rev. E **79**, 011127 (2009).

[51] O. A. van Herwaarden and J. Grasman, J. Math. Biol. **33**, 581 (1995).

[52] M. Marthaler and M. I. Dykman, Phys. Rev. A **73**, 042108 (2006).

[53] M. Khasin and M. I. Dykman, Phys. Rev. Lett. **103**, 068101 (2009).

[54] M. Khasin, B. Meerson, E. Khain, and L. M. Sander, Phys. Rev. Lett. **109**, 138104 (2012).

[55] V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer-Verlag, New York, 1988).

[56] J. Guckenheimer and P. Holmes, *Nonlinear Oscillators, Dynamical Systems and Bifurcations of Vector Fields* (Springer-Verlag, New York, 1997).

[57] H. Kramers, Physica (Utrecht) **7**, 284 (1940).

[58] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations.* (Springer, Berlin, 1999).

[59] H. Risken, *The Fokker-Planck Equation*, 2nd ed. (Springer, Berlin, 1996).

[60] A. O. Caldeira and A. J. Leggett, Ann. Phys. (N.Y.) **149**, 374 (1983).