# Stochastic thermodynamics of Langevin systems under time-delayed feedback control. II. Nonequilibrium steady-state fluctuations

M. L. Rosinberg,[1,*] G. Tarjus,[1] and T. Munakata[2]

[1]*Laboratoire de Physique Théorique de la Matière Condensée, Université Pierre et Marie Curie, CNRS UMR 7600, 4 place Jussieu, 75252 Paris Cedex 05, France*

[2]*Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan*

(Received 19 December 2016; published 21 February 2017)

This paper is the second in a series devoted to the study of Langevin systems subjected to a continuous time-delayed feedback control. The goal of our previous paper [Phys. Rev. E **91**, 042114 (2015)] was to derive second-law-like inequalities that provide bounds to the average extracted work. Here we study stochastic fluctuations of time-integrated observables such as the heat exchanged with the environment, the extracted work, or the (apparent) entropy production. We use a path-integral formalism and focus on the long-time behavior in the stationary cooling regime, stressing the role of rare events. This is illustrated by a detailed analytical and numerical study of a Langevin harmonic oscillator driven by a linear feedback.

DOI: 10.1103/PhysRevE.95.022123

## I. INTRODUCTION

This paper is part of an ongoing effort to include the effect of time delay in the thermodynamic description of small stochastic systems subjected to a continuous feedback control. Time delay is now recognized to play an essential role in many physical, biological, and information systems and also occurs very frequently in experimental setups. Moreover, within the last two decades, including a delay between the detection and the control operation has emerged as an important feedback strategy for controlling transport or stabilizing irregular motion in classical or quantum systems, especially in the presence of noise (see, e.g., the collection of papers in Refs. [1–3]). Accordingly, there is much interest in the mathematical and control theory literature for exploring the plethora of complex phenomena produced by the combination of time delay and noise.

On the other hand, it is much less common to analyze time-delayed feedback loops from the perspective of energetic and information exchanges, which is the main focus of the emerging fields of stochastic and information thermodynamics [4,5]. One reason is the non-Markovian nature of the dynamics, which makes the theoretical description more challenging (for instance, one cannot resort to a spectral approach using Fokker-Planck operators). This is not an impossible task, though, and in a previous work [6], hereafter referred to as paper I, we have initiated a theoretical study of an underdamped Langevin equation that models the motion of a nanomechanical resonator in contact with a thermal reservoir and subjected to a time-delayed, position-dependent force. The role of the control force is to damp thermal fluctuations and to maintain the resonator in a nonequilibrium steady state (NESS) where its average (configurational or kinetic) temperature is much smaller than the temperature of the environment. Heat is thus permanently extracted from the bath and converted into work, which means that the feedback control operates as an autonomous Maxwell's demon. We then derived a series of second-law-like inequalities that provide bounds to the average extracted work. One of these bounds, obtained by (formally) time reversing the feedback, is intimately related to the non-Markovian character of the dynamics.

However, fluctuations dominate at the nanoscale [7], and it is not sufficient to merely describe observables by their typical value. It is also important to study the large-deviation statistics that characterizes the fluctuations at long times. This is the purpose of the present work where we extend the study of paper I by considering the nonequilibrium fluctuations of three time-integrated thermodynamic quantities, the heat, the work, and a so-called "apparent" entropy production (to be defined below). These observables have the same average value in the stationary state, but their fluctuations may differ because of the unbounded growth of temporal boundary terms. As it turns out, these fluctuations are very dependent on the time delay, and this issue is the central theme of this work.

The paper is organized as follows. We first review in Sec. II some basic facts about the model, the observables, and the calculation of the large deviation rate functions. Then, in Sec. III, we introduce two different conjugate dynamics and use them to derive two expressions of the dissipated heat as a ratio of path probabilities. This allows us to express the path-integral representations of the cumulant generating functions in three different ways, which will play an important role in our study. Section IV, which is the central part of the paper, is devoted to a detailed numerical and analytical study of the large-deviation statistics for a harmonic oscillator driven by a linear feedback. The main objective of the theoretical analysis is to explain the intriguing effect of the delay on the probability distributions of the observables in the long-time limit. Special attention is paid to the behavior of the corresponding scaled cumulant generating functions and to the connection between rare fluctuations of the temporal boundary terms and the asymptotic behavior of the conjugate dynamics. We finally derive two stationary-state fluctuation theorems for the work performed by the feedback force. Summary and closing remarks are presented in Sec. V. Some additional but important pieces of information are given in two appendices. In particular, Appendix B offers a complete analytical study of the fluctuations in the (Markovian) small-delay limit where the

*mlr@lptmc.jussieu.fr

 

feedback generates an additional viscous damping and the so-called "molecular refrigerator" model studied in Refs. [8–10] is recovered.

We have tried to make the present paper self-contained as much as possible. However, we warn the reader that some analytical developments rely strongly on paper I, in particular on Sec. V B 2.

## II. MODEL AND OBSERVABLES

As in paper I, we consider an underdamped Brownian particle of mass $m$ immersed in a thermal environment with viscous damping $\gamma$ and temperature $T$. The dynamical evolution is governed by the one-dimensional Langevin equation

$$m\dot{v}_t = -\gamma v_t + F(x_t) + F_{fb}(t) + \sqrt{2\gamma T}\xi_t, \tag{1}$$

where $v_t = \dot{x}_t, F(x) = -dV(x)/dx$ is a conservative force, and $\xi_t$ is a zero-mean Gaussian white noise with unit variance (throughout the paper, temperatures and entropies are measured in units of the Boltzmann constant $k_B$). $F_{fb}(t)$ is the feedback control force, which depends on the position of the particle at time $t-\tau$:

$$F_{fb}(t) = F_{fb}(x_{t-\tau}), \tag{2}$$

where $\tau > 0$ is the time delay. This model is intended to describe an autonomous feedback process in which the instantaneous state of the system (here the position of the Brownian particle) is continuously monitored with perfect accuracy, but some time is needed to implement the control. Clearly, $\tau$ must be smaller than any relaxation time in the system for the control to be efficient. We stress that it is the stochastic force $F_{fb}(t)$ that makes the system's dynamics non-Markovian and not the interaction with the environment.

Our goal is to study the fluctuations of a time-integrated observable $\mathcal{A}_t$ such as the work done by the feedback force or the heat exchanged with the environment during the time interval $[0,t]$, assuming that the system has reached a nonequilibrium steady state (NESS). As discussed in paper I, this requires to properly choose the parameters of the feedback loop, such as the delay or the feedback gain. In fact, multiple NESSs may exist, which is a remarkable feature of time-delayed systems (see, e.g., Fig. 2 below). Moreover, we will focus on regions of the parameter space where the feedback controller acts as a Maxwell's demon who permanently extracts heat from the environment and uses it as work to maintain the system at a temperature smaller than $T$.

The time-integrated work and dissipated heat are defined as

$$\mathcal{W}_t[\mathbf{X},\mathbf{Y}] = \int_0^t dt'\, F_{fb}(x_{t'-\tau}) \circ v_{t'}, \tag{3}$$

and

$$\begin{aligned}\mathcal{Q}_t[\mathbf{X},\mathbf{Y}] &= \int_0^t dt'\, [\gamma v_{t'} - \sqrt{2\gamma T}\xi_{t'}] \circ v_{t'} \\ &= -\int_0^t dt'\, [m\dot{v}_{t'} - F(x_{t'}) - F_{fb}(x_{t'-\tau})] \circ v_{t'},\end{aligned} \tag{4}$$

where the integrals are interpreted with the Stratonovich prescription. These are standard definitions of work and heat in stochastic thermodynamics [4,11], except for the fact that the delay makes the two observables depending on both $\mathbf{X}$, the system trajectory in phase space in the time interval $[0,t]$, and $\mathbf{Y}$, the trajectory in the previous interval $[-\tau,0]$ [we here assume that $t \geqslant \tau$ so that $\mathbf{x}_i \equiv (x_0,v_0) \equiv \mathbf{y}_f$]. This of course is a source of complication for the theoretical description, although one may suspect that the dependence on $\mathbf{Y}$ does not play a major role at long times. From now on, we will drop the functional dependence of the observables on $\mathbf{X}$ and $\mathbf{Y}$ to simplify the notation. (There are a few other differences with the notations used in paper I: the time window is now $[0,t]$ instead of $[-\mathcal{T},\mathcal{T}]$ and the time-integrated observables are denoted by calligraphic uppercase symbols, e.g., $\mathcal{W}_t$ instead of $w$.)

In the following, we will also consider the fluctuations of the trajectory-dependent functional (dubbed as an "apparent" entropy production)

$$\Sigma_t = \Sigma_t^m + \ln\frac{p_0(\mathbf{x}_i)}{p_1(\mathbf{x}_f)}, \tag{5}$$

where $\Sigma_t^m = \beta\mathcal{Q}_t$ [$\beta = (k_BT)^{-1}$] is the entropy change in the medium, $\mathbf{x}_f \equiv (x_t,v_t)$, and $p_0(\mathbf{x}), p_1(\mathbf{x})$ are arbitrary normalized distributions. In the steady state, the natural choice for these distributions is $p_0(\mathbf{x}) = p_1(\mathbf{x}) = p_{st}(\mathbf{x})$, and an observer unaware of the existence of the feedback control would regard $\Sigma_t$ as the total stochastic entropy production (EP) in the time interval $[0,t]$ [12]. However, $\Sigma_t$ is negative on average in the cooling regime, in apparent violation of the second law, and more generally does not obey an integral fluctuation theorem (IFT), $\langle e^{-\Sigma_t}\rangle \neq 1$. Another, but more complicated, trajectory-dependent functional that may quantify the entropy production in the system was introduced in paper I. This functional does satisfy an IFT.

It is important to notice that the three fluctuating quantities $\beta\mathcal{W}_t, \beta\mathcal{Q}_t$, and $\Sigma_t$ have the same expectation value in the stationary state:

$$\langle\beta\mathcal{W}_t\rangle_{st} = \langle\beta\mathcal{Q}_t\rangle_{st} = \langle\Sigma_t\rangle_{st}. \tag{6}$$

Moreover, $\mathcal{W}_t$ and $\mathcal{Q}_t$ are related via the first law that expresses the conservation of energy at the microscopic level [11],

$$\mathcal{Q}_t = \mathcal{W}_t - \Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f), \tag{7}$$

where

$$\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f) = \tfrac{1}{2}m\left(v_t^2 - v_0^2\right) + V(x_t) - V(x_0) \tag{8}$$

is the change in the internal energy of the system after the time $t$. Accordingly, the fluctuations of $\mathcal{W}_t, \mathcal{Q}_t$, and $\Delta\mathcal{U}$ are not independent.

We are interested in the long-time behavior of the stationary probability distribution functions (pdfs) $P_{st}(\mathcal{A}_t)$, where $\mathcal{A}_t$ stands for either $\beta\mathcal{W}_t$ or $\beta\mathcal{Q}_t$ or $\Sigma_t$. As $t\to\infty$, we expect these pdfs to acquire the scaling form

$$P_{st}(\mathcal{A}_t = at) \sim e^{-I(a)t}, \tag{9}$$

where $I(a) \equiv -\lim_{t\to\infty}(1/t)\ln P_{st}(\mathcal{A}_t = at)$ is the large deviation rate function (LDF) that is used to characterize the statistics of exponentially rare events [13]. As usual, to obtain the rate function, we introduce the moment generating or

characteristic function

$$Z_A(\lambda,t) = \langle e^{-\lambda \mathcal{A}_t} \rangle_{st} \tag{10}$$

and the corresponding scaled cumulant generating function (SCGF)

$$\mu_A(\lambda) \equiv \lim_{t\to\infty} \frac{1}{t} \ln \langle e^{-\lambda \mathcal{A}_t} \rangle_{st}, \tag{11}$$

whose behavior away from $\lambda = 0$ encodes information about rare trajectories contributing to the tails of the pdf. For generic values of $\lambda$, one expects $\mu_A(\lambda)$ to be the same function $\mu(\lambda)$ for $\beta\mathcal{W}_t, \beta\mathcal{Q}_t$, and $\Sigma_t$ since the three observables differ only by temporal boundary terms like $\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)$ or $\ln p_0(\mathbf{x}_i)/p_1(\mathbf{x}_f)$. This amounts to assuming that the generating functions behave asymptotically as

$$Z_A(\lambda,t) \sim g_A(\lambda) e^{\mu(\lambda) t}, \tag{12}$$

where the dependence on the observable $\mathcal{A}_t$ is included in the subleading factor $g_A(\lambda)$ that results from the average over the initial and final states (in the present case, the initial "state" involves the whole trajectory $\mathbf{Y}$). The LDFs $I(a)$ are then obtained via the Legendre transform

$$I(a) = -\lambda^* a - \mu(\lambda^*), \tag{13}$$

with the saddle point $\lambda^*(a)$ being the root of $\mu'(\lambda^*) = -a$ [13]. However, Eq. (13) breaks down when $g_A(\lambda)$ has singularities in the region of the saddle-point integration due to rare but large fluctuations of the boundary terms. Although such terms typically do not grow with time, they may indeed fluctuate to order $t$ when the potential $V(x)$ is unbounded, which is the situation considered here. The leading contribution to the LDF then comes from the singularity, which induces an exponential tail in the pdf. This issue is now well documented in the literature, both theoretically [14–24] and experimentally [25–27]. In consequence, while the three observables $\beta\mathcal{W}_t, \beta\mathcal{Q}_t$, and $\Sigma_t$ have the same expectation value, their LDFs may differ. In some circumstances, large fluctuations of the boundary terms may even induce a discontinuity of the SCGF at $\lambda = 1$, as pointed out recently [28]: the asymptotic expression (12) is then no longer valid and $\mu_A(1) \neq \mu(1)$. We shall see later that this is very much dependent on the time delay.

The main difficulty we are facing in the present study is that no analytical methods are currently known to compute the SCGFs. If the dynamics were Markovian, one would determine the largest eigenvalue of the appropriate Fokker-Planck operator [13]. But there is no such operator in the presence of delay (except in the small-$\tau$ limit where Markovianity is recovered), and one has to rely on numerical simulations or to focus on a linear dynamics for which the calculation of $\mu(\lambda)$ can be carried out by going to the frequency domain. However, even in this case, the expression of the prefactors $g_A(\lambda)$ remains out of reach for generic values of $\lambda$.

## III. CONJUGATE DYNAMICS AND GENERATING FUNCTIONS

### A. Conjugate dynamics and dissipated heat

We begin our study by recalling two expressions for the heat dissipated along a trajectory that will play a significant role in the following. We stress that these relations are valid for trajectories of arbitrary duration. There is no need to take the limit $t \to \infty$.

The first relation is obtained by introducing a modified or "conjugate" Langevin dynamics in which the sign of the viscous damping is flipped:

$$m\dot{v}_t = \gamma v_t + F(x_t) + F_{fb}(x_{t-\tau}) + \sqrt{2\gamma T}\xi_t. \tag{14}$$

This readily yields [28]

$$\beta\mathcal{Q}_t = \ln \frac{\mathcal{P}[\mathbf{X}|\mathbf{Y}]}{\hat{\mathcal{P}}[\mathbf{X}|\mathbf{Y}]} - \frac{\gamma}{m} t, \tag{15}$$

where $\mathcal{P}[\mathbf{X}|\mathbf{Y}]$ and $\hat{\mathcal{P}}[\mathbf{X}|\mathbf{Y}]$ are the conditional probabilities of realizing the trajectory $\mathbf{X}$ with the original and conjugate dynamics, respectively, given the trajectory $\mathbf{Y}$ (and thus the initial value $\mathbf{x}_i = \mathbf{y}_f$). These two probabilities can be expressed in terms of Onsager-Machlup (OM) action functionals [29],

$$\mathcal{P}[\mathbf{X}|\mathbf{Y}] \propto e^{\frac{\gamma}{2m}t}\, e^{-\beta\mathcal{S}[\mathbf{X},\mathbf{Y}]} \tag{16a}$$

$$\hat{\mathcal{P}}[\mathbf{X}|\mathbf{Y}] \propto e^{-\frac{\gamma}{2m}t}\, e^{-\beta\hat{\mathcal{S}}[\mathbf{X},\mathbf{Y}]}, \tag{16b}$$

where

$$\mathcal{S}[\mathbf{X},\mathbf{Y}] = \frac{1}{4\gamma}\int_0^t dt' \,[m\dot{v}_{t'} + \gamma v_{t'} - F(x_{t'}) - F_{fb}(x_{t'-\tau})]^2, \tag{17a}$$

$$\hat{\mathcal{S}}[\mathbf{X},\mathbf{Y}] = \frac{1}{4\gamma}\int_0^t dt' \,[m\dot{v}_{t'} - \gamma v_{t'} - F(x_{t'}) - F_{fb}(x_{t'-\tau})]^2, \tag{17b}$$

and the exponential factors $e^{\pm\frac{\gamma}{2m}t}$ come from the Jacobians of the transformations $\xi(t) \to x(t)$ associated with the two Langevin dynamics (see Ref. [30] or the supplemental material of Ref. [28] for a derivation). As usual, the continuous-time integrals in Eqs. (17) are interpreted as the limit of discrete sums, as discussed, for instance, in the Appendix B of Ref. [31]. We recall that there is no need to specify the interpretation (Ito versus Stratonovitch) of the stochastic calculus as long as $m \neq 0$. Hereafter, the *hat* symbol will refer to quantities associated with the $\gamma \to -\gamma$ conjugate dynamics (14).

From Eq. (15), one immediately obtains an IFT for the dissipated heat [28]:

$$\langle e^{-\beta\mathcal{Q}_t} \rangle = e^{\frac{\gamma}{m}t}. \tag{18}$$

In particular, this implies at long times that

$$\mu_Q(1) = \frac{\gamma}{m}. \tag{19}$$

Moreover, the average dissipated heat satisfies $\langle \beta\mathcal{Q}_t \rangle \geqslant -(\gamma/m)t$ by Jensen's inequality. This bound is trivial, though, and can be directly obtained by averaging Eq. (4), which yields

$$\langle \mathcal{Q}_t \rangle = \frac{\gamma}{m}\int_0^t dt' [T_v(t') - T], \tag{20}$$

where $T_v(t) = m\langle v_t^2 \rangle \geqslant 0$ is the effective temperature of the momentum degree of freedom.

The second relation is obtained by performing the time-reversal operation normally associated with the microscopic

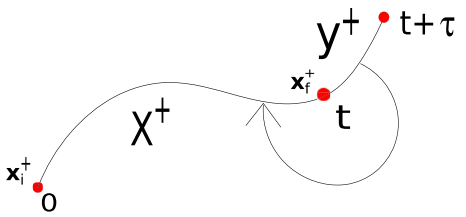


FIG. 1. Time-reversed paths $\mathbf{X}^\dagger$ and $\mathbf{Y}^\dagger$. When the dynamics is governed by the acausal Langevin equation (21), the feedback force depends on the *future* state of the system, as schematically represented by the arrowed line.

reversibility condition [32]. The key point is that $\mathcal{Q}_t$ is no longer an odd quantity under time reversal because of the time delay. To recover this symmetry, one must also flip $\tau$ into $-\tau$ and introduce another conjugate dynamics defined by the *acausal* Langevin equation:

$$m\dot{v}_t = -\gamma v_t + F(x_t) + F_{fb}(x_{t+\tau}) + \sqrt{2\gamma T}\xi_t. \tag{21}$$

The usual local detailed balance equation is then generalized as [6,33]

$$\beta\mathcal{Q}_t = \ln\frac{\mathcal{P}[\mathbf{X}|\mathbf{Y}]}{\widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger]} - \ln\frac{\mathcal{J}_t}{\widetilde{\mathcal{J}}[\mathbf{X}]}, \tag{22}$$

where $\widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger]$ is the probability of realizing the time-reversed trajectory $\mathbf{X}^\dagger$ with the conjugate dynamics (hereafter represented by the *tilde* symbol), given the initial value $\mathbf{x}_i^\dagger \equiv \mathbf{x}^\dagger(t=0) = (x_t, -v_t)$ and the trajectory $\mathbf{Y}^\dagger$. Note that $\mathbf{Y}^\dagger$ denotes the time-reversed path in the time interval $[t, t+\tau]$, so that its initial point is $\mathbf{x}_f^\dagger \equiv \mathbf{x}^\dagger(t) = (x_0, -v_0)$, as shown schematically in Fig. 1.

Therefore, the probability weight of $\mathbf{X}^\dagger$ must be conditioned on both the initial value $\mathbf{x}_i^\dagger$ and the future trajectory $\mathbf{Y}^\dagger$. $\widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger]$ is then expressed as

$$\widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger] \propto \widetilde{\mathcal{J}}[\mathbf{X}]e^{-\beta\widetilde{\mathcal{S}}[\mathbf{X}^\dagger,\mathbf{Y}^\dagger]}, \tag{23}$$

where

$$\widetilde{\mathcal{S}}[\mathbf{X},\mathbf{Y}] = \frac{1}{4\gamma}\int_0^t dt'\,[m\dot{v}_{t'} + \gamma v_{t'} - F(x_{t'}) - F_{fb}(x_{t'+\tau})]^2 \tag{24}$$

and $\widetilde{\mathcal{J}}[\mathbf{X}]$ is the Jacobian of the transformation $\xi(t) \to x(t)$ associated with Eq. (21). As shown in paper I, $\widetilde{\mathcal{J}}[\mathbf{X}]$ is a nontrivial functional of the path in general, but it becomes a path-independent quantity $\widetilde{\mathcal{J}}_t$ like the Jacobian $\mathcal{J}_t$ associated with Eq. (1) in the case of a linear dynamics. (In this work, we use the notation $\widetilde{\mathcal{J}}_t$ and $\mathcal{J}_t$ instead of $\widetilde{\mathcal{J}}$ and $\mathcal{J}$ to emphasize that these quantities depend on the duration of the trajectory.)

Note that the two OM actions $\widetilde{\mathcal{S}}[\mathbf{X},\mathbf{Y}]$ and $\hat{\mathcal{S}}[\mathbf{X},\mathbf{Y}]$ are related by time inversion:

$$\widetilde{\mathcal{S}}[\mathbf{X},\mathbf{Y}] = \hat{\mathcal{S}}[\mathbf{X}^\dagger,\mathbf{Y}^\dagger]. \tag{25}$$

There is also an IFT associated with Eq. (22), but it involves a more complicated path functional [see Eqs. (74)–(79) in paper I] that plays no role in the following. Let us just recall the corresponding second-law-like inequality for the heat flow in the NESS [6,33]:

$$\langle\beta\dot{\mathcal{Q}}\rangle_{st} \geqslant -\dot{\mathcal{S}}_{\mathcal{J}}, \tag{26}$$

where

$$\dot{\mathcal{S}}_{\mathcal{J}} \equiv \lim_{t\to\infty}\frac{1}{t}\left\langle \ln\frac{\mathcal{J}_t}{\widetilde{\mathcal{J}}[\mathbf{X}]}\right\rangle_{st}. \tag{27}$$

This bound is in general different from the trivial bound $\langle\beta\dot{\mathcal{Q}}\rangle_{st} \geqslant -(\gamma/m)$ obtained from Eq. (20) (see, e.g., Fig. 8 below).

### B. Generating functions in the NESS

We now focus on the steady-state regime and drop the suffix "st" in all expressions hereafter to shorten the notation. Our objective in this section is to express the generating functions $Z_A(\lambda,t)$ in three different ways by exploiting expressions (15) and (22) of the dissipated heat. We start from the definition (10), which we write more explicitly as

$$Z_A(\lambda,t) = \iint d\mathbf{x}_i\, d\mathbf{x}_f \int d\mathbb{P}[\mathbf{Y}] \int_{\mathbf{x}_i}^{\mathbf{x}_f} \mathcal{D}\mathbf{X}\, e^{-\lambda\mathcal{A}_t}\mathcal{P}[\mathbf{X}|\mathbf{Y}], \tag{28}$$

where $\int d\mathbb{P}[\mathbf{Y}]\cdots$ is a shorthand notation for $\int d\mathbf{y}_i\, p(\mathbf{y}_i)\int_{\mathbf{y}_i}^{\mathbf{x}_i}\mathcal{D}\mathbf{Y}\,\mathcal{P}[\mathbf{Y}|\mathbf{y}_i]\cdots$ and $\mathbf{y}_i \equiv (x_{-\tau}, v_{-\tau})$ [hence $\int d\mathbb{P}[\mathbf{Y}] = p(\mathbf{x}_i)$]. Since the three observables differ only by temporal boundary terms which are functions of $\mathbf{x}_i$ and $\mathbf{x}_f$, we single out one of them, $\beta\mathcal{W}_t$, and define the $\lambda$-dependent quantity:

$$\mathcal{K}_\lambda[\mathbf{x}_f,t|\mathbf{Y}] = \int_{\mathbf{x}_i}^{\mathbf{x}_f}\mathcal{D}\mathbf{X}\, e^{-\lambda\beta\mathcal{W}_t}\mathcal{P}[\mathbf{X}|\mathbf{Y}]. \tag{29}$$

(The choice of $\beta\mathcal{W}_t$ instead of $\beta\mathcal{Q}_t$ or $\Sigma_t$ will be justified *a posteriori* in Sec. IV B 1.) Loosely speaking, $\mathcal{K}_\lambda[\mathbf{x}_f,t|\mathbf{Y}]$ is a kind of biased transition probability from $\mathbf{x}_i$ to $\mathbf{x}_f$. This allows us to reexpress the three generating functions as

$$Z_A(\lambda,t) = \iint d\mathbf{x}_i d\mathbf{x}_f\; f_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f)\int d\mathbb{P}[\mathbf{Y}]\,\mathcal{K}_\lambda[\mathbf{x}_f,t|\mathbf{Y}], \tag{30}$$

where

$$f_{W,\lambda}(\mathbf{x}_i,\mathbf{x}_f) = 1, \tag{31a}$$

$$f_{Q,\lambda}(\mathbf{x}_i,\mathbf{x}_f) = e^{\lambda\beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)}, \tag{31b}$$

$$f_{\Sigma,\lambda}(\mathbf{x}_i,\mathbf{x}_f) = e^{\lambda[\beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)+\ln p(\mathbf{x}_f)/p(\mathbf{x}_i)]}, \tag{31c}$$

and we have used the first law (7) to define $f_{Q,\lambda}$ and $f_{\Sigma,\lambda}$.

We now use Eq. (15) to replace the path probability $\mathcal{P}[\mathbf{X}|\mathbf{Y}]$ by $\hat{\mathcal{P}}[\mathbf{X}|\mathbf{Y}]$ in Eq. (28). We then define

$$\hat{\mathcal{K}}_\lambda[\mathbf{x}_f,t|\mathbf{Y}] = e^{\frac{\gamma}{m}t}\int_{\mathbf{x}_i}^{\mathbf{x}_f}\mathcal{D}\mathbf{X}\, e^{(1-\lambda)\beta\mathcal{W}_t}\hat{\mathcal{P}}[\mathbf{X}|\mathbf{Y}], \tag{32}$$

which leads to

$$Z_A(\lambda,t) = \iint d\mathbf{x}_i d\mathbf{x}_f\,\hat{f}_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f)\int d\mathbb{P}[\mathbf{Y}]\,\hat{\mathcal{K}}_\lambda[\mathbf{x}_f,t|\mathbf{Y}] \tag{33}$$

with

$$\hat{f}_{W,\lambda}(\mathbf{x}_i,\mathbf{x}_f) = e^{-\beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)}, \tag{34a}$$

$$\hat{f}_{Q,\lambda}(\mathbf{x}_i,\mathbf{x}_f) = e^{(\lambda-1)\beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)}, \tag{34b}$$

$$\hat{f}_{\Sigma,\lambda}(\mathbf{x}_i,\mathbf{x}_f) = e^{(\lambda-1)\beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)+\lambda\ln\frac{p(\mathbf{x}_f)}{p(\mathbf{x}_i)}}. \tag{34c}$$

Likewise, we can use Eq. (22) to replace $\mathcal{P}[\mathbf{X}|\mathbf{Y}]$ by $\widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger]$. In this case, it is convenient to change the path integral over $\mathbf{X}$ in Eq. (28) into an integral over $\mathbf{X}^\dagger$, and define

$$\widetilde{\mathcal{K}}_\lambda(\mathbf{x}_f^\dagger,t|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger) = \int_{\mathbf{x}_i^\dagger}^{\mathbf{x}_f^\dagger}\mathcal{D}\mathbf{X}^\dagger\,\frac{\mathcal{J}_t}{\widetilde{\mathcal{J}}[\mathbf{X}]}e^{(1-\lambda)\beta\mathcal{W}_t}\widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger,\mathbf{Y}^\dagger], \tag{35}$$

which leads to

$$Z_A(\lambda,t) = \iint d\mathbf{x}_i^\dagger\,d\mathbf{x}_f^\dagger\;\tilde{f}_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f)\int d\mathbb{P}[\mathbf{Y}]\;\widetilde{\mathcal{K}}_\lambda(\mathbf{x}_f^\dagger,t|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger) \tag{36}$$

with

$$\widetilde{f}_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f) = \hat{f}_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f). \tag{37}$$

At first glance, it might seem that we have gained nothing by replacing Eq. (30) by two other expressions of the generating functions that are even more complicated. This is true for a generic value of $\lambda$. But the interesting feature of Eqs. (33) and (36) is the special role played by $\lambda = 1$. This makes these two equations well suited to infer the asymptotic behavior of the quantities $Z_A(1,t) = \langle e^{-\mathcal{A}_t}\rangle$, thus revealing the occurrence of large statistical fluctuations originating from temporal boundary terms. However, this requires us to first determine whether or not the conjugate Langevin equations (14) and (21) admit a stationary solution. Although this analysis can be done in a rather general framework, it is quite delicate in the case of the acausal dynamics (21), and it is more illuminating to focus on a specific case, the linear model studied in the next section. We shall thus return to this issue later. A detailed discussion is presented in Appendix A.

## IV. TIME-DELAYED LANGEVIN HARMONIC OSCILLATOR

To be concrete, we now consider the time-delayed linear Langevin equation

$$m\dot{v}_t = -\gamma v_t - kx_t + k'x_{t-\tau} + \sqrt{2\gamma T}\xi_t, \tag{38}$$

which is conveniently rewritten in a dimensionless form as

$$\dot{v}_t = -\frac{1}{Q_0}v_t - x_t + \frac{g}{Q_0}x_{t-\tau} + \xi_t \tag{39}$$

by taking the inverse angular resonance frequency $\omega_0^{-1} = \sqrt{m/k}$ as the unit of time and $x_c = k^{-1}(2\gamma\omega_0 T)^{1/2}$ as the unit of position [34]. In this equation, $Q_0 = \omega_0\tau_0$ denotes the intrinsic quality factor of the oscillator ($\tau_0 = m/\gamma$ is the viscous relaxation time), and $g = k'/(\gamma\omega_0) = (k'/k)Q_0$ represents the gain of the feedback loop. The dynamics of the system is thus fully characterized by the three independent dimensionless parameters $Q_0, g$, and $\tau$. The gain $g$ is usually the control variable in feedback-cooling experimental setups (see, e.g., Ref. [35]).

In these reduced units, the fluctuating work and heat (normalized by $k_BT$) are given by

$$\beta\mathcal{W}_t = \frac{2g}{Q_0^2}\int_0^t dt'\,x_{t'-\tau}\circ v_{t'} \tag{40}$$

and

$$\beta\mathcal{Q}_t = \beta\mathcal{W}_t - \tfrac{1}{Q_0}\big(x_f^2 - x_i^2 + v_f^2 - v_i^2\big). \tag{41}$$

These are quadratic functionals of the noise, and therefore the corresponding probabilities are not Gaussian. To obtain the expression of the EP functional $\Sigma_t$ defined by Eq. (5), we use the expression of the stationary pdf derived in paper I,

$$p(\mathbf{x}) = \frac{1}{2\pi[\langle x^2\rangle\langle v^2\rangle]^{1/2}}e^{-\frac{x^2}{2\langle x^2\rangle}-\frac{v^2}{2\langle v^2\rangle}}, \tag{42}$$

where the mean square position and velocity are expressed in terms of the configurational and kinetic temperatures $T_x$ and $T_v$: $\langle x^2\rangle = (Q_0/2)T_x/T$ and $\langle v^2\rangle = (Q_0/2)T_v/T$. These two effective temperatures are given by Eqs. (113) and (114) in paper I, respectively. We recall that $T_x$ is the temperature commonly measured in experiments involving nanomechanical devices [36–38] whereas $T_v$ determines the heat flow (and thus the extracted work) in the stationary state, according to

$$\beta\dot{\mathcal{Q}} = \frac{1}{Q_0}\left(\frac{T_v}{T} - 1\right), \tag{43}$$

i.e., $\beta\dot{\mathcal{Q}} = (\gamma/m)(T_v/T - 1)$ in original units [see Eq. (20)]. Since $T_x \neq T_v$ in general, the system does not obey the standard equipartition theorem (cf. the discussion in paper I). Inserting Eq. (42) into Eq. (5) then yields

$$\beta\Sigma_t = \beta\mathcal{W}_t + \frac{1}{Q_0}\left[\frac{T-T_x}{T_x}\big(x_t^2 - x_0^2\big) + \frac{T-T_v}{T_v}\big(v_t^2 - v_0^2\big)\right]. \tag{44}$$

Since we are dealing with a linear dynamics with Gaussian noise, all stationary path probabilities are Gaussian distributions [39], and the calculation of the generating functions $Z_A(\lambda,t)$ from Eqs. (30) and (31) amounts to computing Gaussian path integrals. For $t$ finite, however, this calculation cannot be carried out analytically for essentially two reasons. The first is that the Euler-Lagrange equation for the optimal trajectory is a forward-backward delay differential equation that has no closed-form solution in general (it can be solved only by a perturbative expansion in powers of $g$). The second reason is that the explicit expression of $\mathcal{P}[\mathbf{Y}]$ is unknown [40], so that the average over initial conditions cannot be performed.

Things become simpler in the long-time limit as one can use the Fourier transform to obtain an analytical expression of the SCGF $\mu(\lambda)$; see Eq. (53) below (but, as already stressed, the value at $\lambda = 1$ requires special care). The LDFs are then obtained via the Legendre transform in Eq. (13). However, since the behavior of the prefactors $g_A(\lambda)$ is unknown and singularities may occur, additional assumptions are needed. Useful insight on this issue is gained by inspecting the small-$\tau$ limit of the Langevin equation, which corresponds to the Markovian model originally considered in Refs. [8,9]. The

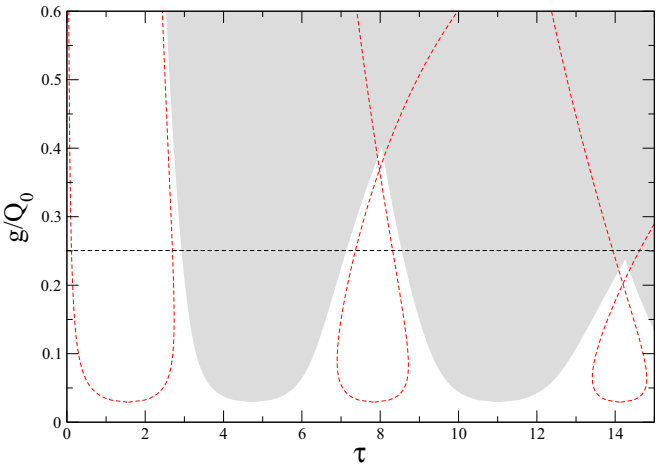


FIG. 2. Stability diagram of the feedback-controlled oscillator for $Q_0 = 34.2$ (the time unit is the inverse angular resonance frequency $\omega_0^{-1}$). The oscillator is unstable inside the shaded regions. The acausal response function $\widetilde{\chi}(s)$ has all its poles located in the r.h.s. of the complex $s$ plane inside the regions delimited by the dashed red lines and two poles in the l.h.s. outside these regions.

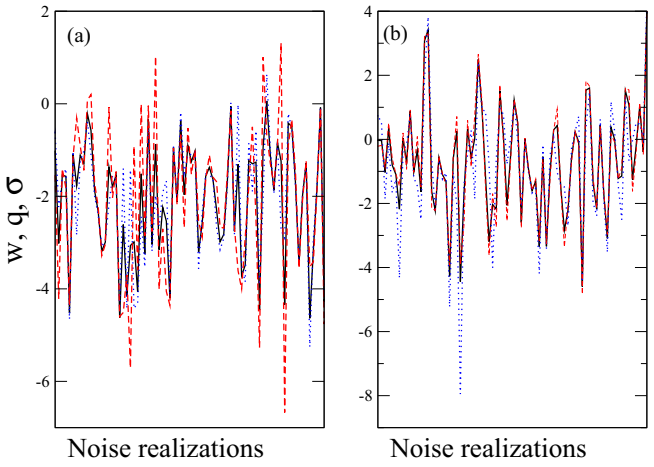


FIG. 3. Stochastic fluctuations of $w = \beta\mathcal{W}_t/t$ (solid black line), $q = \beta\mathcal{Q}_t/t$ (dotted blue line), and $\sigma = \Sigma_t/t$ (dashed red line) for $Q_0 = 34.2\, g/Q_0 = 0.25, \tau = 7.6$ (a) and $\tau = 8.4$ (b). The figure shows the results obtained for an observation time $t = 100$ and 75 independent noise realizations. Lines are a guide for the eyes.

feedback then generates an additional viscous damping, and a complete analytical description of the fluctuations is possible, as detailed in Appendix B. (A first, but incomplete, analysis was performed by two of the present authors in Refs. [10].) This study, together with the additional pieces of information gathered from the direct numerical simulation of Eq. (38), will eventually allow us to propose a global scenario.

In order to give the reader a foretaste of the puzzle that must be resolved, we first present some data obtained from numerical simulations of the Langevin equation (39). The theoretical interpretation is postponed to Sec. IV B.

### A. Numerical study

Although we have studied the model for various values of the dimensionless parameters $Q_0$ and $g$, we here present only numerical results obtained for $Q_0 = 34.2$ and $g/Q_0 = 0.25$. We have chosen this set of parameters for several reasons. In the first place, the value of $Q_0$ corresponds to an actual experimental system: the AFM microcantilever used in the experiments of Ref. [41], which is characterized by a resonance period $2\pi/\omega_0 = 116\ \mu$s and a viscous relaxation time $\tau_0 = 632\ \mu$s. In the second place, the feedback-controlled oscillator has an interesting dynamical behavior for $g/Q_0 = 0.25$, as shown in Fig. 2 (see also Fig. 11 in paper I). Specifically, a stationary state can be reached in two stability lobes $0 < \tau < \tau_1$ and $\tau_2 < \tau < \tau_3$, with $\tau_1 \approx 2.93$, $\tau_2 \approx 7.13$, $\tau_3 \approx 8.55$. For intermediate values of $\tau$ or $\tau > \tau_3$, there is no stationary state. In the third place, the probability distributions have a nontrivial behavior as a function of $\tau$, which vividly illustrates the role of rare events due to boundary temporal terms.

To begin, we show in Fig. 3 an example of the sample-to-sample fluctuations of $\beta\mathcal{W}_t, \beta\mathcal{Q}_t$, and $\Sigma_t$ for two values of $\tau$ chosen in the second stability lobe $\tau_2 < \tau < \tau_3$ (the behavior is qualitatively similar in the first lobe). The observation time is $t = 100$ and the Langevin equation (39) was solved by using Heun's method [42] with a time step $\Delta t = 5 \times 10^{-4}$. As expected, the fluctuations of the three observables are strongly correlated. But, remarkably, the contribution of the temporal boundary terms is still non-negligible despite the long observation time. In particular, they contribute differently to the observables depending on the value of $\tau$: for $\tau = 7.6$ (resp. $\tau = 8.4$) it is $\Sigma_t$ (resp. $\mathcal{Q}_t$) that exhibits the largest fluctuations. Note that the delay is significantly smaller than the viscous relaxation time $\tau_0 = Q_0/\omega_0 = 34.2$ in both cases, and that the system operates in the cooling regime: $T_x/T \approx 0.42$, $T_v/T \approx 0.36$, $\langle\beta\dot{\mathcal{Q}}\rangle \approx -0.019$ for $\tau = 7.6$, and $T_x/T \approx 0.72$, $T_v/T \approx 0.84$, $\langle\beta\dot{\mathcal{Q}}\rangle \approx -0.005$ for $\tau = 8.4$.

To get a more quantitative picture, the corresponding stationary pdfs are shown in Figs. 4 and 5. These plots clearly

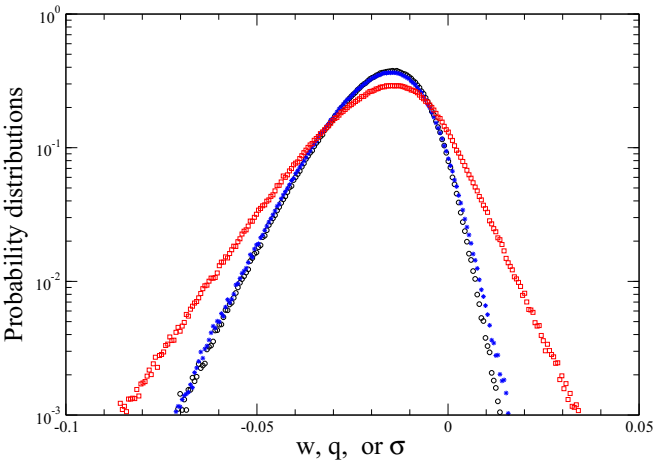


FIG. 4. Probability distribution functions $P(\beta\mathcal{W}_t)$ (black circles), $P(\beta\mathcal{Q}_t)$ (blue stars), and $P(\Sigma_t)$ (red squares) plotted against the scaled variable $a = \mathcal{A}_t/t$ ($a = w$, $q$, or $\sigma$) for $Q_0 = 34.2$, $g/Q_0 = 0.25$, and $\tau = 7.6$. The observation time is $t = 100$. Symbols represent numerical data obtained by solving the Langevin equation (39) for $2 \times 10^6$ realizations of the noise.

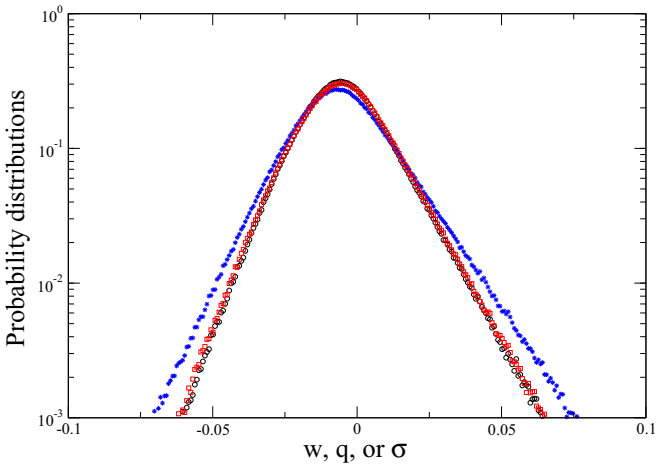


FIG. 5. Same as Fig. 4 for $\tau = 8.4$.

confirm the main feature suggested by Fig. 3: $P(\Sigma_t = \sigma t)$ for $\tau = 7.6$ and $P(\beta \mathcal{Q}_t = qt)$ for $\tau = 8.4$ differ markedly from $P(\beta \mathcal{W}_t = wt)$. Of course, these results should be interpreted with care since it is notoriously difficult to sample rare fluctuations. However, we expect that the picture emerging from Figs. 4 and 5 would not change qualitatively at larger times. Moreover, it is consistent with the exact analytical analysis performed in Appendix B in the small-$\tau$ limit and in the associated Markovian model. This will be rationalized in the next subsection.

The corresponding estimates of the SCGFs $\mu_A(\lambda)$ are plotted in Fig. 6. One noticeable feature is the distinct behavior of $\mu_\Sigma(\lambda)$ for $\tau = 7.6$ and of $\mu_Q(\lambda)$ for $\tau = 8.4$ in the vicinity of $\lambda = 1$. However, it is also manifest that finite-time and/or finite-sample-size effects are significant. In particular, $\mu_\Sigma(\lambda)$ for $\tau = 7.6$ widely differs from the two other SCGFs for $\lambda \lessapprox -1$ and varies linearly with $\lambda$ for $\lambda \lessapprox -1.5$, which is presumably a numerical artifact, as discussed in a more general context in Ref. [43].

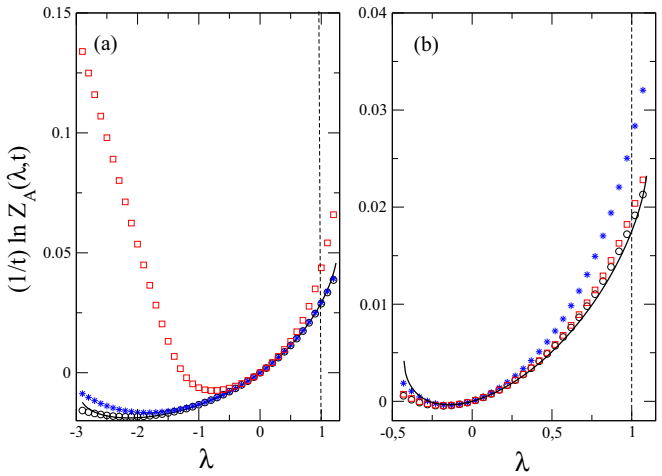


FIG. 6. Numerical estimates of $\mu_A(\lambda)$ using $t = 100$ and $2 \times 10^6$ realizations of the noise for $\tau = 7.6$ (a) and $\tau = 8.4$ (b): $\mu_W$ (black circles), $\mu_Q$ (blue stars), and $\mu_\Sigma$ (red squares). The solid black line represents the theoretical SCGF $\mu(\lambda)$ given by Eq. (53) in the interval $(\lambda_{\min}, \lambda_{\max})$ in which this quantity is real.

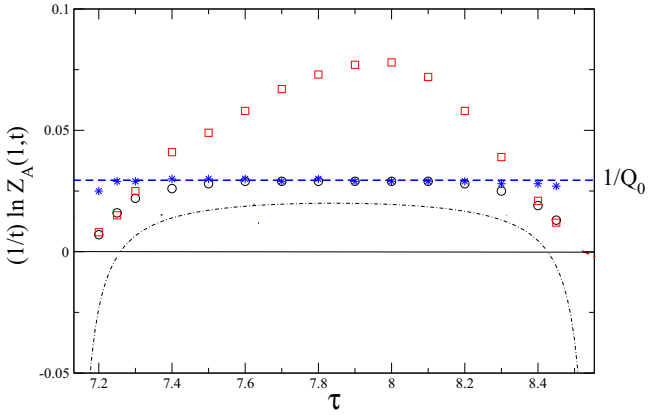


FIG. 7. Numerical estimates of $\mu_A(1) = \lim_{t\to\infty}(1/t)\ln\langle e^{-\mathcal{A}_t}\rangle$ as a function of $\tau$ in the second stability lobe: $\mu_W(1)$ (black circles), $\mu_Q(1)$ (blue stars), and $\mu_\Sigma(1)$ (red squares). The dashed-dotted black line is the average extracted work rate $\langle \beta \dot{\mathcal{W}}_{\rm ext}\rangle = -\langle \beta \dot{\mathcal{Q}}\rangle$, which is positive for $7.26 \leqslant \tau \leqslant 8.43$.

Finally, we focus on the special value $\lambda = 1$ and show in Fig. 7 the numerical estimates of $\mu_A(1) = \lim_{t\to\infty}(1/t)\ln\langle e^{-\mathcal{A}_t}\rangle$ in the whole stability range $7.13 < \tau < 8.55$. For information, we also show the average extracted work rate $\langle \beta \dot{\mathcal{W}}_{\rm ext}\rangle = -\langle \beta \dot{\mathcal{Q}}\rangle$. We first observe that $\mu_Q(1) \approx 1/Q_0$ ($\gamma/m$ in original units) independently of the value of $\tau$. This is indeed what the IFT (18) tells us. In contrast, both $\mu_W(1)$ and $\mu_\Sigma(1)$ display a nontrivial behavior with $\tau$: $\mu_W(1)$ is equal to $1/Q_0$ only in the subinterval $7.37 \lessapprox \tau \lessapprox 8.32$, whereas $\mu_\Sigma(1)$ varies nonmonotonically with a maximum around $\tau \approx 7.9$. This clearly calls for a theoretical explanation.

### B. Theoretical analysis

We now present a theoretical scenario that explains why the three observables, which differ only by temporal boundary terms, have different fluctuations in the long-time limit and why this behavior depends on the delay.

#### *1. Calculation of the (boundary-independent) SCGF*

We first calculate the SCGF $\mu(\lambda)$ defined by the asymptotic formula (12), assuming that boundary terms depending on $\mathbf{x}_i, \mathbf{x}_f$, or $\mathbf{Y}$ play no role in the long-time limit. (We recall that the trajectory $\mathbf{Y}$ is of duration $\tau$.) However, one should keep in mind that the actual value of $\mu_A(1)$ may differ from $\mu(1)$.

In order to compute $\mu(\lambda)$, we impose periodic boundary conditions on the trajectory $\mathbf{X}$ and expand $x_t$ in discrete Fourier series (see, e.g., Refs. [44–48] for similar calculations),

$$x(t) = \sum_{n=-\infty}^{\infty} x_n e^{-i\omega_n t}, \tag{45}$$

with inverse transform

$$x_n = \frac{1}{t}\int_0^t ds\, x(s) e^{i\omega_n s}, \tag{46}$$

where $\omega_n = 2\pi n/t$ and $x_n \equiv x(\omega_n)$. In the limit $t \to \infty$, the standard Fourier transform is recovered.

After inserting into Eq. (40) and neglecting the contribution coming from $\mathbf{Y}$, we obtain

$$\frac{1}{t}\beta\mathcal{W}_t \sim \frac{2g}{Q_0^2}\sum_{n=-\infty}^{\infty}(i\omega_n)x_n x_{-n}e^{i\omega_n\tau}$$
$$\sim -\frac{4g}{Q_0^2}\sum_{n=1}^{\infty}\omega_n x_n x_{-n}\sin(\omega_n\tau). \quad (47)$$

(Of course, this expression is also valid for $\beta\mathcal{Q}_t$ or $\Sigma_t$ since the contribution of the boundary terms are neglected in this calculation.) We then use the linearity of the Langevin equation to replace $x_n$ by the frequency component $\xi_n$ of the noise

$$x_n = \chi(\omega_n)\xi_n, \quad (48)$$

where

$$\chi(\omega) = \left[-\omega^2 - \frac{i\omega}{Q_0} + 1 - \frac{g}{Q_0}e^{i\omega\tau}\right]^{-1} \quad (49)$$

is the Fourier transform of the response function of the time-delayed oscillator. Hence,

$$\langle e^{-\lambda\beta\mathcal{W}_t}\rangle \sim \prod_{n=1}^{\infty}\int d\xi_n\, P(\xi_n)e^{\frac{4\lambda gt}{Q_0^2}\xi_n\xi_{-n}\omega_n\sin(\omega_n\tau)|\chi(\omega_n)|^2} \quad (50)$$

with

$$\mathcal{P}[\xi_n] = \frac{t}{\pi}e^{-t\xi_n\xi_{-n}}. \quad (51)$$

The Gaussian integration over $\xi_n$ gives

$$\langle e^{-\lambda\beta\mathcal{W}_t}\rangle \sim \prod_{n=1}^{\infty}\left[1 - \frac{4\lambda g}{Q_0^2}\omega_n\sin(\omega_n\tau)|\chi(\omega_n)|^2\right]^{-1}, \quad (52)$$

and we finally obtain

$$\mu(\lambda) = -\lim_{t\to\infty}\frac{1}{t}\sum_{n=1}^{\infty}\ln\left[1 - \frac{4\lambda g}{Q_0^2}\omega_n\sin(\omega_n\tau)|\chi(\omega_n)|^2\right]$$
$$= -\frac{1}{2\pi}\int_0^{\infty}d\omega\,\ln\left[1 - \frac{4\lambda g}{Q_0^2}\omega\sin(\omega\tau)|\chi(\omega)|^2\right], \quad (53)$$

where the summation over $n$ has been replaced by an integral over $\omega$ as $t\to\infty$. For a generic value of $\lambda$, the integral must be computed numerically, and the result is real as long as the argument of the logarithm stays positive for all values of $\omega$. Accordingly, $\mu(\lambda)$ is defined in an open domain $(\lambda_{\min},\lambda_{\max})$, with $\lambda_{\min}$ and $\lambda_{\max}$ determined by the minimum and maximum values of the function $f(\omega) = (4g/Q_0^2)\omega\sin(\omega\tau)|\chi(\omega)|^2$. The derivative $\mu'(\lambda)$ diverges at the boundaries, so that the corresponding Legendre transform is asymptotically linear [13].

As regards fluctuation relations, we readily notice from Eq. (53) that $\mu(1-\lambda)\neq\mu(\lambda)$, which implies that the observables do not satisfy a conventional stationary-state fluctuation theorem (SSFT) of the Gallavotti-Cohen type [49–51]: $\lim_{t\to\infty}(1/t)\ln[P(\mathcal{A}_t = at)/P(\mathcal{A}_t = -at)] = a$. On the other hand, alternative SSFTs can be obtained by changing $\gamma$ into $-\gamma$ or $\tau$ into $-\tau$. We will say more about this in Sec. B 4.

How does Eq. (53) compare with the numerical estimates of the SCGFs $\mu_A(\lambda)$ shown in Fig. 6? We see that the agreement is very good for $\mu_W(\lambda)$, although there are still small discrepancies, in particular for $\tau = 8.4$ and the most negative values of $\lambda$. These small deviations will be used to infer the numerical value of the prefactor $g_W(\lambda)$ and build a better approximation of the pdf $P(\beta\mathcal{W}_t = wt)$ [see Eq. (71) below]. Much more significant are the differences with $\mu_\Sigma(\lambda)$ for $\tau = 7.6$ and with $\mu_Q(\lambda)$ for $\tau = 8.4$ in the vicinity of $\lambda = 1$ [leaving aside the spurious linear behavior of $\mu_\Sigma(\lambda)$ for $\tau = 7.6$ and $\lambda \lesssim -1.5$].

Let us investigate this issue in more detail by computing $\mu(1)$. To this aim, we first rewrite Eq. (53) as

$$\mu(\lambda) = \frac{1}{2\pi}\int_0^{\infty}d\omega\,\ln\frac{H_\lambda(\omega)}{H_0(\omega)}, \quad (54)$$

where

$$H_\lambda(\omega)^{-1} \equiv |\chi(\omega)|^{-2} - \frac{4\lambda g}{Q_0^2}\omega\sin(\omega\tau)$$
$$= \left[-\omega^2 + 1 - \frac{g}{Q_0}\cos(\omega\tau)\right]^2$$
$$+ \frac{1}{Q_0^2}[\omega^2 + 2g(1-2\lambda)\omega\sin(\omega\tau) + g^2\sin^2(\omega\tau)] \quad (55)$$

and $H_0(\omega) = |\chi(\omega)|^2$. This immediately shows that

$$H_1(\omega)^{-1} = \left[-\omega^2 + 1 - \frac{g}{Q_0}\cos(\omega\tau)\right]^2 + \frac{1}{Q_0^2}[\omega - g\sin(\omega\tau)]^2$$
$$\equiv |\widetilde{\chi}(\omega)|^{-2}, \quad (56)$$

where

$$\widetilde{\chi}(\omega) \equiv \chi(\omega)|_{\tau\to-\tau} = \left[-\omega^2 - \frac{i\omega}{Q_0} + 1 - \frac{g}{Q_0}e^{-i\omega\tau}\right]^{-1} \quad (57)$$

is the response function of the acausal Langevin equation in the frequency domain. This allows us to express $\mu(1)$ as

$$\mu(1) = \frac{1}{2\pi}\int_0^{+\infty}d\omega\,\ln\frac{|\widetilde{\chi}(\omega)|^2}{|\chi(\omega)|^2} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}d\omega\,\ln\frac{\widetilde{\chi}(\omega)}{\chi(\omega)}, \quad (58)$$

where we have used the fact that the imaginary parts of $\widetilde{\chi}(\omega)$ and $\chi(\omega)$ are odd functions of $\omega$ to eliminate the modulus [52]. We can then compute the integral over $\omega$ by using Cauchy's residue theorem, which requires to locate the poles of $\widetilde{\chi}(\omega)$ in the complex frequency plane [they are not restricted to be in the lower half plane, in contrast with the poles of the causal response function $\chi(\omega)$]. Fortunately, this nontrivial task has already been accomplished in paper I in order to calculate the quantity $\dot{\mathcal{S}}_{\mathcal{J}} \equiv \lim_{t\to\infty}(1/t)\ln\mathcal{J}_t/\widetilde{\mathcal{J}}_t$ involved in the second-law-like inequality (26) obtained from time reversal (we recall that the Jacobian $\widetilde{\mathcal{J}}[\mathbf{X}]$ becomes a path-independent quantity $\widetilde{\mathcal{J}}_t$ when the dynamics is linear [6]). Specifically, it was shown in paper I [Eq. (155)] that

$$\dot{\mathcal{S}}_{\mathcal{J}} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}ds\,\ln\frac{\widetilde{\chi}(s)}{\chi_{g=0}(s)}, \quad (59)$$

where $s = \sigma - i\omega$ is the Laplace complex variable. (From now on, we will mostly work with the Laplace variable in order to directly use the results obtained in paper I, but for simplicity

we will keep the same notation for the response functions.) We thus reexpress Eq. (58) as

$$\mu(1) = \frac{1}{2\pi i}\int_{0-i\infty}^{0+i\infty} ds\ \ln\frac{\widetilde{\chi}(s)}{\chi(s)}, \tag{60}$$

where

$$\chi(s) = \left[s^2 + \frac{s}{Q_0} + 1 - \frac{g}{Q_0}e^{-s\tau}\right]^{-1} \tag{61}$$

and $\widetilde{\chi}(s) \equiv \chi(s)|_{\tau\to-\tau}$. Comparing Eq. (60) to Eq. (59), one may notice two differences: first, one has $\chi(s)$ in the denominator of the logarithm instead of $\chi_{g=0}(s)$, and, second, the integration is performed along the imaginary axis $\mathrm{Re}(s) = 0$ in the complex $s$ plane (since the frequency $\omega$ is real). On the other hand, as was painstakingly discussed in paper I, the Bromwich contour in Eq. (59) [i.e., the value of $c$ crucially depends on the location of the poles of $\widetilde{\chi}(s)$]. The first difference turns out to be irrelevant because all the poles of $\chi(s)$ are located in the left-hand-side (l.h.s.) of the complex $s$ plane. Hence [53],

$$\frac{1}{2\pi i}\int_{0-i\infty}^{0+i\infty} ds\ \ln\frac{\chi(s)}{\chi_{g=0}(s)} = 0. \tag{62}$$

On the other hand, the fact that $c = 0$ in Eq. (60) is relevant in two circumstances:

(1) When all the poles of $\widetilde{\chi}(s)$ lie on the right-hand-side (r.h.s.) of the complex $s$ plane. Then, by using an integration contour similar to the one in Fig. 4 of paper I (with a large semicircle on the l.h.s.), the only singularities inside the contour are the two poles of $\chi_{g=0}(s), s_0^{\pm} = 1/(2Q_0)[-1 \pm i\sqrt{4Q_0^2 - 1}]$. Cauchy's residue theorem then gives the simple result $\mu(1) = 1/Q_0$. This differs from $\dot{S}_{\mathcal{J}}$ because this latter quantity is obtained by also including two poles of $\widetilde{\chi}(s)$ inside the contour in order to avoid the branch cuts of the logarithm. Indeed, as shown in paper I, there must be two, and only two, poles of $\widetilde{\chi}(s)$ on the left side of the integration line $\mathrm{Re}(s) = c$.

(2) When $\widetilde{\chi}(s)$ has more than two poles on the l.h.s. In this case, all these poles contribute to Eq. (60) whereas only the two poles with the smallest real part contribute to $\dot{S}_{\mathcal{J}}$.

To sum up, three different cases may occur:

(a) $\mu(1) = 1/Q_0 \neq \dot{S}_{\mathcal{J}}$ when all the poles of $\widetilde{\chi}(s)$ lie on the r.h.s of the complex $s$ plane

(b) $\mu(1) = \dot{S}_{\mathcal{J}} \neq 1/Q_0$ when only two poles lie on the l.h.s of the complex $s$ plane

(c) $\mu(1) \neq \dot{S}_{\mathcal{J}} \neq 1/Q_0$ when more than two poles lie on the l.h.s of the complex $s$ plane [54].

This calculation of $\mu(1)$, combined with the analysis performed in Appendix A, allows us to elucidate the intriguing dependence of $\mu_W(1)$ and $\mu_\Sigma(1)$ on $\tau$ exhibited in Fig. 7, and more generally the behavior of the SCGFs in the vicinity of $\lambda = 1$ observed in Fig. 6. What is done in Appendix A is first to relate the behavior of the conjugate $\gamma \to -\gamma$ "hat" and $\tau \to -\tau$ "tilde" dynamics to the pole structure of $\widetilde{\chi}(s)$. Then, in a second time, Eqs. (33) and (36) derived in Sec. III B are used to deduce the values of $\mu_W(1)$ and $\mu_\Sigma(1)$. Specifically, it is shown that a stationary state exists with the hat dynamics when all the poles of $\widetilde{\chi}(s)$ are on the r.h.s of the complex $s$ plane [case (a) above] and with the tilde dynamics when two

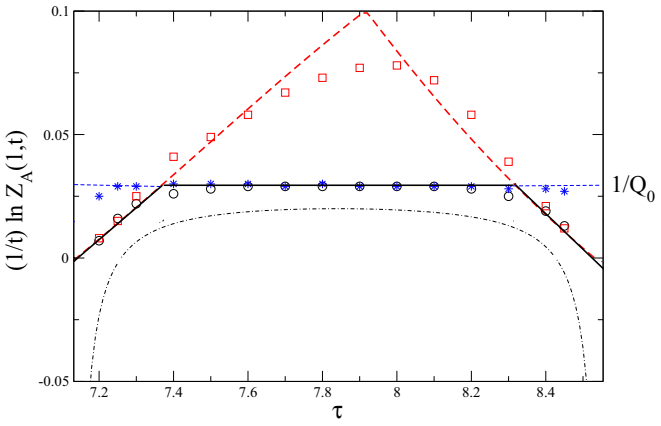


FIG. 8. Comparison between the numerical estimates of $\mu_A(1) = \lim_{t\to\infty}(1/t)\ln\langle e^{-\mathcal{A}_t}\rangle$ displayed in Fig. 7 and the values of $\mu(1)$ (solid black line) and $\dot{\mathcal{S}}_J$ (dashed red line) computed from Eqs. (60) and (59), respectively. One has $\mu(1) = \dot{\mathcal{S}}_J \leqslant 1/Q_0$ for $\tau \leqslant \tau_{p,1} \approx 7.37$ and $\tau \geqslant \tau_{p,2} \approx 8.32$, and $\mu(1) = 1/Q_0 \leqslant \dot{\mathcal{S}}_J$ for $\tau_{p,1} \leqslant \tau \leqslant \tau_{p,2}$. Note that $\dot{\mathcal{S}}_{\mathcal{J}}$ is a tighter bound to the extracted work rate (dashed-dotted black line) than $1/Q_0$ for $\tau < \tau_{p,1}$ and $\tau > \tau_{p,2}$.

(and only two) poles are on the l.h.s [case (b)]. (When there are more than two poles on the l.h.s. [case (c)], a stationary state never exists.) With our present choice for the quality factor $Q_0$ and the feedback gain $g$ ($Q_0 = 34.2$ and $g/Q_0 = 0.25$), we find that case (a) is realized for $\tau_{p,1} < \tau < \tau_{p,2}$, with $\tau_{p,1} \approx 7.37, \tau_{p,2} \approx 8.32$, and case (b) is realized for $\tau < \tau_{p,1}$ and $\tau > \tau_{p,2}$ [for other values of $g$, the boundary between cases (a) and (b) is indicated by the dashed red lines in Fig. 2].

The analysis in Appendix A shows that

$$\mu_W(1) = \mu(1) \tag{63}$$

in *both* cases, i.e., in the whole stability lobe, as illustrated by the solid black line in Fig. 8, whereas

$$\mu_\Sigma(1) = \dot{S}_{\mathcal{J}} = \mu(1) \quad \text{for} \quad \tau < \tau_{p,1} \quad \text{and} \quad \tau > \tau_{p,2}. \tag{64}$$

For $\tau_{p,1} \leqslant \tau \leqslant \tau_{p,2}$, the theoretical analysis indicates only that

$$\mu_\Sigma(1) \neq \mu(1) = 1/Q_0. \tag{65}$$

The latter relation comes from the divergence of the prefactor $g_\Sigma(1)$ [cf. Eq. (A9)]. In addition, there is strong evidence from the numerical data displayed in Fig. 7 that $\mu_\Sigma(1)$ is equal to $\dot{S}_{\mathcal{J}}$ for *all* values of $\tau$, as illustrated by the dashed red line in Fig. 8 [55]. This implies that $\mu_\Sigma(\lambda)$ is discontinuous at $\lambda = 1$ for $\tau_{p,1} \leqslant \tau \leqslant \tau_{p,2}$, which is consistent with the behavior of $(1/t)\ln Z_\Sigma(\lambda,t)$ in the vicinity of $\lambda = 1$ observed in Fig. 6 for $\tau = 7.6$. Similarly, since

$$\mu_Q(1) = 1/Q_0 \neq \mu(1) = \dot{S}_{\mathcal{J}} \quad \text{for } \tau < \tau_{p,1} \text{ and } \tau > \tau_{p,2}, \tag{66}$$

$\mu_Q(\lambda)$ is discontinuous at $\lambda = 1$, which is also consistent with the behavior of $(1/t)\ln Z_Q(\lambda,t)$ observed in Fig. 6 for $\tau = 8.4$.

More precisely, inspired by the exact boundary layer analysis performed in Appendix B in the small-$\tau$ limit and in the associated Markovian model (see in particular Fig. 16

in that Appendix), we conjecture that

$$Z_\Sigma(\lambda,t)e^{-\mu(\lambda)t} \sim e^{(\dot{S}_\mathcal{J}-1/Q_0)t} \quad \text{for} \quad \tau_{p,1} < \tau < \tau_{p,2},$$
$$Z_Q(\lambda,t)e^{-\mu(\lambda)t} \sim e^{(1/Q_0-\dot{S}_\mathcal{J})t} \quad \text{for} \quad \tau < \tau_{p,1} \text{ and } \tau > \tau_{p,2} \tag{67}$$

as $t \to \infty$ and $\lambda \to 1$. Clearly, this anomalous behavior of the two SCGFs can be ascribed to the unbounded (but different) growth of the temporal boundary terms, $\Sigma_t - \beta\mathcal{W}_t = \ln p(\mathbf{x}_i)/p(\mathbf{x}_f) - \beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)$ and $\beta\mathcal{Q}_t - \beta\mathcal{W}_t = \beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)$. In contrast, $\mu_W(\lambda)$ is always equal $\mu(\lambda)$ and is therefore a continuous function of $\lambda$, which is the reason why we have treated $\mathcal{W}_t$ differently from $\mathcal{Q}_t$ and $\Sigma_t$ in Sec. III B.

The analysis performed in Appendix A also gives us some partial information about the values of the prefactors for $\lambda = 1$ when these quantities are finite. This is an interesting outcome since, as we have already pointed out, we are unable to compute the prefactors in general.

For $\tau_{p,1} < \tau < \tau_{p,2}$, after replacing the stationary distributions by their Gaussian expressions in Eq. (A8), we obtain

$$g_W(1) = \frac{T^2}{[(T-T_x)(T+\hat{T}_x)]^{1/2}[(T-T_v)(T+\hat{T}_v)]^{1/2}}, \tag{68}$$

where $\hat{T}_x$ and $\hat{T}_v$ are the steady-state effective temperatures associated with the hat dynamics. [Recall that $g_Q(1) = 1$ and $g_\Sigma(1)$ diverges in this case.]

For $\tau < \tau_{p,1}$ and $\tau > \tau_{p,2}$, the information is more limited since we cannot compute $g_W(1)$ and $g_\Sigma(1)$ separately [while $g_Q(1)$ diverges]. On the other hand, from Eqs. (A15), the ratio of these two prefactors is expected to be

$$\frac{g_W(1)}{g_\Sigma(1)} = \frac{T^2}{[T(T_x+\widetilde{T}_x)-T_x\widetilde{T}_x]^{1/2}[T(T_v+\widetilde{T}_v)-T_v\widetilde{T}_v]^{1/2}}, \tag{69}$$

where $\widetilde{T}_x$ and $\widetilde{T}_v$ are the steady-state effective temperatures associated with the tilde dynamics. The variations of $\hat{T}_v$ and $\widetilde{T}_v$ with $\tau$ are shown in Fig. 15 in Appendix A. It is worth noting that $\hat{T}_v$ and $\widetilde{T}_v$ are larger than $T_v$ in the stationary cooling regime where $T_v < T$.

### 2. *Calculation of $I(w)$*

We now compute the large deviation rate functions and start with $I(w)$. Our basic assumption is that the prefactor $g_W(\lambda)$ has no singularity *whatever* the value of $\lambda$ (and not only for $\lambda = 1$ as discussed above). This is supported by the exact analytical calculations in the Markovian limit reported in Appendix B and is also in line with the exact behavior observed in other (Markovian) nonequilibrium models [14,17,20,24] and checked experimentally [27,56]. Consequently, the LDF $I(w)$ is always given by the Legendre transform $I(w) = -\lambda^* w - \mu(\lambda^*)$ with $\mu'(\lambda^*) + w = 0$. From Eqs. (54)–(55), this amounts to solving numerically the equation

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\ \omega \sin(\omega\tau) H_{\lambda^*}(\omega) = -\frac{Q_0^2}{2g} w \tag{70}$$

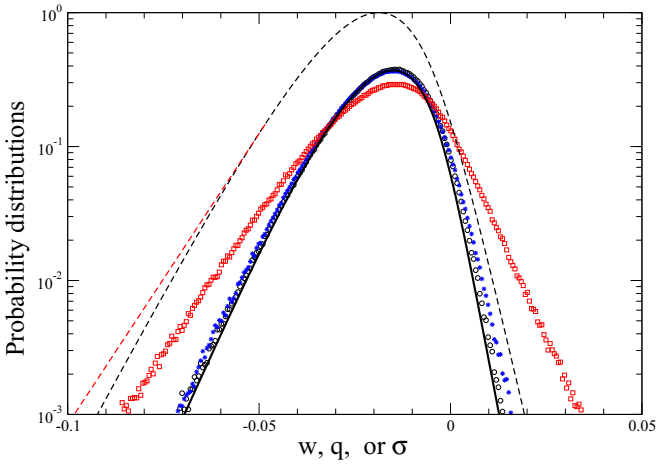


FIG. 9. Probability distribution functions $P(\mathcal{A}_t = at)$ for $\tau = 7.6$. The symbols are the data obtained from the numerical simulations (see Fig. 4), the dashed black line is the large-deviation form $e^{-I(w)t}$, and the solid black line is the semiempirical asymptotic expression given by Eq. (71). The dashed red line on the l.h.s. for $\sigma \lessapprox -0.048$ is the curve $e^{-I_1(\sigma)t}$ obtained from Eq. (73).

so as to obtain the saddle point $\lambda^*$ as a function of $w$. This leads to the curves $e^{-I(w)t}$ shown in Figs. 9 and 10 as dashed black lines.

From these figures, however, it is clear that the large deviation form does not properly describe $P(\mathcal{W}_t = wt)$ for $t = 100$, in particular in Fig. 10, where the slopes on the r.h.s. are quite different. This can be traced back to finite-time corrections which can be computed by using a standard expansion around the saddle point (see, e.g., Refs. [10,21]), assuming again the absence of any singularity in $g_W(\lambda)$. This yields

$$P(\mathcal{W}_t = wt) \approx \frac{g_W[\lambda^*(w)]}{\sqrt{2\pi\mu''[\lambda^*(w)]t}} e^{-I(w)t}. \tag{71}$$

Although the analytical expression of $g_W(\lambda)$ for generic values of $\lambda$ is unknown, a semiempirical estimate can be obtained from Fig. 6, assuming that the very small deviations between

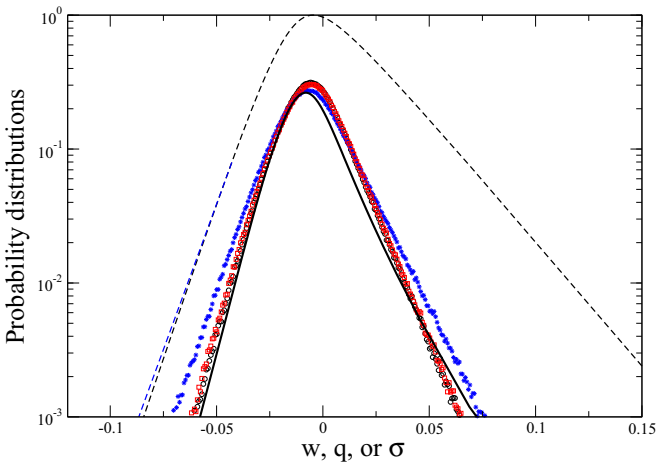


FIG. 10. Same as Fig. 9 for $\tau = 8.4$. The dashed blue line on the l.h.s. for $q \lessapprox -0.042$ is the curve $e^{-I_1(q)t}$ obtained from Eq. (73).

$\mu_W(\lambda)$ and $\mu(\lambda)$ are due to neglecting the prefactor. We thus compute the prefactor as $g_W(\lambda) \approx Z_W(\lambda,t)e^{-\mu(\lambda)t}$, where $Z_W(\lambda,t)$ is obtained from the numerical simulations, and insert the result into Eq. (71) [57]. As shown by the solid black lines in Figs. 9 and 10, this procedure leads to a much better description of the numerical data. We take this as indirect but convincing evidence that our theoretical analysis of the work fluctuations is well sounded. The remaining discrepancies observed for $\tau = 8.4$ may be attributed to statistical uncertainty due to the limited sampling.

#### 3. *Calculation of $I(q)$ and $I(\sigma)$*

The calculation of the LDFs $I(q)$ and $I(\sigma)$ is more challenging because we can no longer assume that the prefactors $g_Q(\lambda)$ and $g_\Sigma(\lambda)$ have no singularities. In particular, we already know from the preceding discussion that $\lambda = 1$ is a pole of $g_Q(\lambda)$ for $\tau = 8.4$ [as $\mu_Q(1) = 1/Q_0 \neq \mu(1)$] and a pole of $g_\Sigma(\lambda)$ for $\tau = 7.6$ [as $\mu_\Sigma(1) = \dot{S}_\mathcal{J} \neq \mu(1)$]. In addition, the exact calculation of the generating functions $Z_Q(\lambda,t)$ and $Z_\Sigma(\lambda,t)$ in the small-$\tau$ limit and in the associated Markovian model shows that other pole singularities appear when performing the stationary average over the *initial* state $\mathbf{x}_i$ [see Eqs. (B19b) and (B19c)]. These poles, due again to rare but large fluctuations of the temporal boundary terms, occur for $\lambda < 0$ and lead to an exponential tail in the r.h.s of the pdfs [58]. (In contrast, the poles at $\lambda = 1$ occur when performing the average over the *final* state $\mathbf{x}_f$ and lead to an exponential tail in the l.h.s. of the LDFs.) We then expect that these rare events are responsible, together with finite-time corrections, for the fact that the slopes of $P(\Sigma_t = \sigma t)$ in Fig. 9 and of $P(\mathcal{Q}_t = qt)$ in Fig. 10 are not correctly described by the Legendre transform of $\mu(\lambda)$.

Unfortunately, we have no way to determine analytically *all* the poles of $g_Q(\lambda)$ and $g_\Sigma(\lambda)$ for an arbitrary value of $\tau$. The best we can do is to describe how the pole at $\lambda = 1$ (when it exists) modifies the LDFs $I(q)$ and $I(\sigma)$. To this aim, we compute the special value of $q$ or $\sigma$ for which the saddle point $\lambda^*$ reaches 1. According to Eq. (53), it is given by

$$a^* = -\mu'(1) = -\frac{g}{\pi Q_0^2}\int_{-\infty}^{\infty} d\omega\, \omega \sin(\omega\tau)|\widetilde{\chi}(\omega)|^2, \quad (72)$$

where $a^*$ stands for either $q^*$ or $\sigma^*$ and we have used $\widetilde{\chi}$ in place of $\chi$. When this corresponds to a pole in the prefactor (depending on the observable and on the value of $\tau$), the LDF becomes linear for $a < a^*$ and is given by

$$I_1(a) = -[\mu(1) - a]. \quad (73)$$

This leads to the modified asymptotic behaviors $P(\Sigma_t = \sigma t) \sim e^{-I_1(\sigma)t}$ and $P(\mathcal{Q}_t = qt) \sim e^{-I_1(q)t}$ shown in Figs. 9 and 10, respectively. We see that the slopes on the l.h.s. are now in much better agreement with the numerical simulations.

#### 4. *Two stationary-state fluctuation theorems (SSFTs)*

To end our study, we now examine the status of the conventional fluctuation relation for the work $\mathcal{W}_t$ and state two alternative relations that hold in the long-time limit.

As we have already mentioned, the SCGF $\mu(\lambda)$, whose expression is given by Eqs. (53) or (54), does not possess the symmetry $\mu(1-\lambda) = \mu(\lambda)$ that would lead to a conventional SSFT expressing the symmetry around 0 of the pdf of an observable $\mathcal{A}_t$ at large times. This is strikingly illustrated by Fig. 11 where we plot the symmetry function $f(w) = I(-w) - I(w) = \lim_{t\to\infty} \frac{1}{t} \ln \frac{P(\mathcal{W}_t = wt)}{P(\mathcal{W}_t = -wt)}$ for $w \geqslant 0$ [with $f(-w) = -f(w)$]. We see that the SSFT symmetry $f(w) = w$ is violated for all values of $w$. On the one hand, one has $f(w) < 0$ for small positive values of $w$ since the average work rate is negative in the cooling regime (as can be seen in Figs. 4 and 5, the probability of having a negative event $\beta\mathcal{W}_t = -wt$ is indeed larger than the probability of having a positive event $\beta\mathcal{W}_t = +wt$). On the other hand, large fluctuations are described at the level of the large deviation function by $I(w) \sim -\lambda_{\max} w$ for $w < 0$ and $I(w) \sim -\lambda_{\min} w$ for $w > 0$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the boundaries of the region of convergence of $\mu(\lambda)$ [13] (see Fig. 6 with $\lambda_{\min} \approx -2.94, \lambda_{\max} \approx 1.26$ for $\tau = 7.6$ and $\lambda_{\min} \approx -0.43, \lambda_{\max} \approx 1.10$ for $\tau = 8.4$). This implies that $f(w) \sim (\lambda_{\min} + \lambda_{\max})w$. As can be seen in Fig. 11, the symmetry function $f(w)$ smoothly interpolates between these two regimes of small and large fluctuations. The remarkable feature is that the second fluctuation regime is quite different for $\tau = 7.6$ and $\tau = 8.4$ as $\lambda_{\min} + \lambda_{\max} < 0$ in the first case and $\lambda_{\min} + \lambda_{\max} > 0$ in the second one. We emphasize that this striking effect of the time delay cannot be attributed to the influence of temporal boundary terms since we focus here only on the fluctuations of the work.

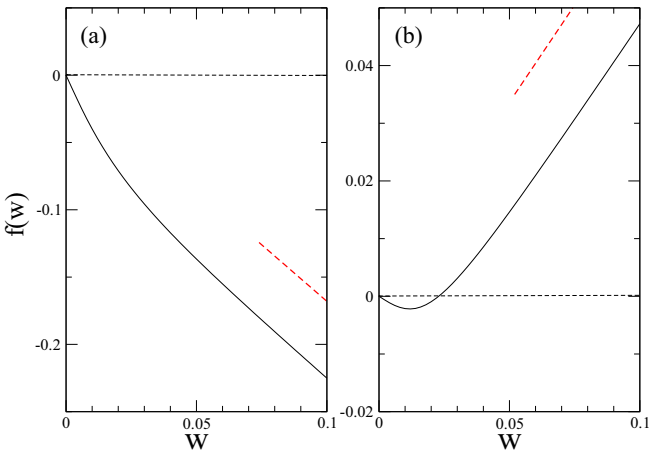


FIG. 11. Symmetry function for the work fluctuations $f(w) = I(-w) - I(w)$ for $\tau = 7.6$ (a) and $\tau = 8.4$ (b). The dashed red lines represent the asymptotic regime of large fluctuations $f(w) \sim (\lambda_{\min} + \lambda_{\max})w$ (see text).

Whereas the standard symmetry $\mu(1-\lambda) = \mu(\lambda)$ does not hold, it is easily seen from the definition of the function $H_\lambda(\omega)$ [Eq. (55)] that

$$\hat{H}_{\lambda-1}(\omega) = H_\lambda(\omega), \quad (74a)$$

$$\widetilde{H}_{1-\lambda}(\omega) = H_\lambda(\omega), \quad (74b)$$

where $\hat{H}_\lambda(\omega) \equiv H_\lambda(\omega)|_{\gamma\to-\gamma}$ and $\widetilde{H}_\lambda(\omega) \equiv H_\lambda(\omega)|_{\tau\to-\tau}$ [we remind the reader that $\hat{\chi}(\omega)^{-1} = -\omega^2 + i\omega/Q_0 + 1 - (g/Q_0)e^{i\omega\tau}$ and $\widetilde{\chi}(\omega)^{-1} = -\omega^2 - i\omega/Q_0 + 1 - (g/Q_0)e^{-i\omega\tau}$ in dimensionless units]. We then deduce from Eq. (54) the two

symmetry relations

$$\hat{\mu}(\lambda - 1) = \mu(\lambda) - \mu(1), \tag{75a}$$

$$\widetilde{\mu}(1 - \lambda) = \mu(\lambda) - \mu(1), \tag{75b}$$

where $\hat{\mu}(\lambda) \equiv \mu(\lambda)|_{\gamma \to -\gamma}$ and $\widetilde{\mu}(\lambda) \equiv \mu(\lambda)|_{\tau \to -\tau}$.

Now, for a SSFT to hold, a stationary state must also exist with the dynamics associated with the transformation $\gamma \to -\gamma$ or $\tau \to -\tau$. In this case, the corresponding pdfs $\hat{P}(\beta \mathcal{W}_t = wt)$ and $\widetilde{P}(\beta \widetilde{\mathcal{W}}_t = wt)$ are expected to acquire asymptotically the large-deviation forms

$$\hat{P}(\beta \mathcal{W}_t = wt) \sim e^{-\hat{I}(w)t}, \tag{76a}$$

$$\widetilde{P}(\beta \widetilde{\mathcal{W}}_t = wt) \sim e^{-\widetilde{I}(w)t}, \tag{76b}$$

where $\beta \widetilde{\mathcal{W}}_t \equiv \beta \mathcal{W}_t|_{\tau \to -\tau} = (2g)/(Q_0^2) \int_0^t dt' \, x_{t'+\tau} \circ v_{t'}$. Assuming again that boundary terms are irrelevant for the fluctuations of the work at large times, whatever the dynamics, the two LDFs $\hat{I}(w)$ and $\widetilde{I}(w)$ are then given by the Legendre transform of the corresponding SCGFs $\hat{\mu}(\lambda)$ and $\widetilde{\mu}(\lambda)$. From Eqs. (75) and the corresponding saddle-point equations, we then obtain

$$I(w) - \hat{I}(w) = -w - \mu(1), \tag{77a}$$

$$I(w) - \widetilde{I}(-w) = -w - \mu(1), \tag{77b}$$

which yields the two SSFTs

$$\lim_{t \to \infty} \frac{1}{t} \ln \frac{P(\beta \mathcal{W}_t = wt)}{e^{t/Q_0} \hat{P}(\beta \mathcal{W}_t = wt)} = w, \tag{78a}$$

$$\lim_{t \to \infty} \frac{1}{t} \ln \frac{P(\beta \mathcal{W}_t = wt)}{e^{t \dot{S}_{\mathcal{J}}} \widetilde{P}(\beta \widetilde{\mathcal{W}}_t = -wt)} = w. \tag{78b}$$

We stress that the fluctuation relation (78a) holds for $\tau_{p,1} < \tau < \tau_{p,2}$ in the second stability lobe [hence $\mu(1) = 1/Q_0$], whereas relation (78b) holds for $\tau < \tau_{p,1}$ and $\tau > \tau_{p,2}$ [hence $\mu(1) = \dot{S}_{\mathcal{J}} \neq 1/Q_0$]. In fact, since $\widetilde{P}(\beta \widetilde{\mathcal{W}}_t = -wt) \sim \widetilde{P}(\beta \mathcal{W}_t = wt)$ asymptotically, this latter relation can be also reexpressed as

$$\lim_{t \to \infty} \frac{1}{t} \ln \frac{P(\beta \mathcal{W}_t = wt)}{e^{t \dot{S}_{\mathcal{J}}} \widetilde{P}(\beta \mathcal{W}_t = wt)} = w. \tag{79}$$

A numerical check of the two SSFTS is provided in Figs. 12(a) and 12(b) (see below for an explanation of the numerical procedure). The agreement is satisfactory in both cases, taking into account that the exponential factor $e^{-\beta \mathcal{W}_t}$ strongly weights work values in the far left tail of $P(\beta \mathcal{W}_t)$ corresponding to very rare realizations of the process that cannot be properly sampled [60]. As $t$ increases, we expect the curves in Fig. 12 to be peaked more and more around the asymptotic work value $w^* = \lim_{t \to \infty} \int dw \, w P(\beta \mathcal{W}_t) e^{-wt} / \int dw \, P(\beta \mathcal{W}_t) e^{-wt} = -\mu'(1)$ [cf. Eq. (72)], with $w^* \approx -0.048$ in Fig. 12(a) and $w^* \approx -0.236$ in Fig. 12(b). This latter figure illustrates the curious feature that atypical fluctuations become typical when generated by an acausal dynamics! This dynamics (when it leads to a stationary state) then defines the so-called "auxiliary" or "driven" process [61–63] that generates asymptotically the ensemble of paths conditioned on the constraint $\beta \mathcal{W}_t / t = w^*$ [64]. In fact, changing

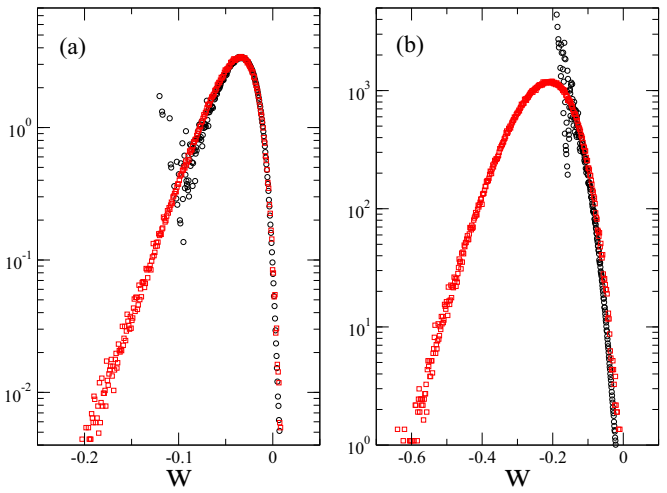


FIG. 12. Verification of the SSFTs (78a) and (79). Here $P(\beta \mathcal{W}_t = wt) e^{-wt}$ (black circles) is compared with (a) $\hat{P}(\mathcal{W}_t = wt) e^{t/Q_0}$ (red squares) for $Q_0 = 34.2$, $g/Q_0 = 0.25$, $\tau = 7.6$, and (b) $\widetilde{P}(\beta \mathcal{W}_t = wt) e^{\dot{S}_{\mathcal{J}} t}$ (red squares) for $Q_0 = 2$, $g/Q_0 = 0.55$, $\tau = 2.5$ ($\dot{S}_{\mathcal{J}} \approx 0.1$). The observation time is $t = 100$.

$\tau$ into $-\tau$ in the stationary cooling regime $T_v < T$ has essentially the same effect as changing $\gamma$ into $-\gamma$, namely, to *enhance* the fluctuations in the system and thus to increase the effective temperatures: for instance, one has $\hat{T}_x \approx 0.792 > T_x \approx 0.420, \hat{T}_v \approx 0.656 > T_v \approx 0.359$, and $\widetilde{T}_x \approx 1.538 > T_x \approx 0.956, \widetilde{T}_v \approx 1.220 > T_v \approx 0.960$ for the two cases represented in Fig. 12 (see also Fig. 15 in Appendix A). But, at the same time, there is *more* work extracted from the bath since in both cases $w^*$ is more negative than the average work, or dissipated heat, rate $(1/Q_0)(T_v/T - 1)$ [cf. Eq. (43)]. This kind of counterintuitive behavior that occurs in the rare fluctuations regime is discussed in Ref. [48] for another model of feedback cooling, where the focus is on the information exchange between the system and the feedback controller.

It is instructive to detail how the numerical data displayed in Fig. 12(b) were obtained. For the $\gamma \to -\gamma$ dynamics, one can directly solve the dimensionless Langevin equation $\dot{v}_t = (1/Q_0) v_t - x_t + (g/Q_0) x_{t-\tau} + \xi_t$ using the standard Euler or Heun's methods. However, these schemes cannot be applied to the acausal Langevin equation $\dot{v}_t = -(1/Q_0) v_t - x_t + (g/Q_0) x_{t+\tau} + \xi_t$. Fortunately, because of the linearity of the equation, there is a strategy for tackling this problem. Indeed, for a given history of the thermal noise $\xi(t)$ over a long time interval $[-t_1, t_2]$, a stationary solution can be approximated as

$$x(t) \approx \int_{-t_1}^{t_2} dt' \widetilde{\chi}(t - t) \xi(t'), \tag{80}$$

where $\widetilde{\chi}(t)$ is the inverse Fourier transform of the acausal response function $\widetilde{\chi}(\omega)$ (see the discussion in Appendix A). If $t_1, t_2 \gg t > 0$ and if $\widetilde{\chi}(t)$ decays sufficiently rapidly for both positive and negative times, Eq. (80) provides a very good approximation of $x(t)$ in the time interval $[0, t]$. In this way, one can generate a representative ensemble of stationary trajectories and estimate the probabilities $\widetilde{P}(\beta \widetilde{\mathcal{W}}_t = -wt)$ or $\widetilde{P}(\beta \mathcal{W}_t = wt)$. [It turns out that the case $Q_0 = 34.2$,

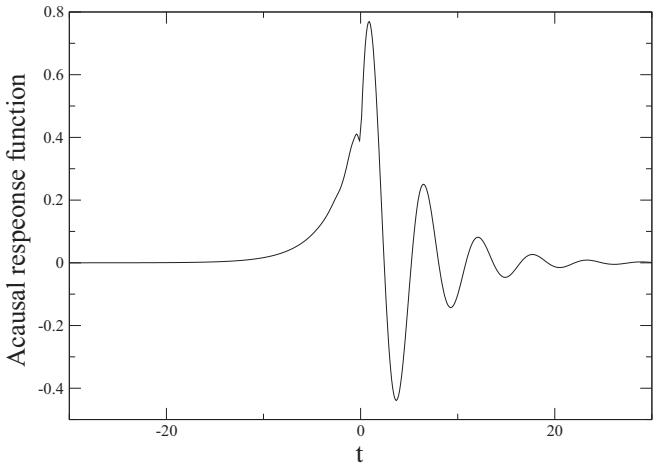


FIG. 13. Acausal response function $\widetilde{\chi}(t)$ for $Q_0 = 2, g/Q_0 = 0.55$ and $\tau = 2.5$. $\widetilde{\chi}(s)$ has two poles on the l.h.s. of the complex $s$ plane and an infinite number of poles on the r.h.s. The poles $s \approx -0.200 \pm 1.12i$ and $s \approx 0.339$ control the behavior of $\widetilde{\chi}(t)$ for $t \geqslant 0$ and for $t \to -\infty$, respectively.

$g/Q_0 = 0.25, \tau = 8.4$ cannot be studied with this method because $\widetilde{\chi}(t)$ decays too slowly for $t > 0$ as the two poles of $\widetilde{\chi}(s)$ on the l.h.s. of the complex $s$ plane have a very small real part. Figure 12(b) thus corresponds to another choice of the parameters for which a stationary state still exists with the acausal dynamics and $\widetilde{\chi}(t)$ decays to 0 rapidly, as shown in Fig. 13.]

Finally, we mention another way to understand the origin of the large fluctuations contributing to $\mathcal{P}(\beta \mathcal{W}_t) e^{-\beta \mathcal{W}_t}$, which is to consider the atypical thermal noise that generates such fluctuations. To this end, we select an atypical stationary trajectory produced by one or the other conjugate process and insert it into the original Langevin equation. The calculation can be readily performed in the frequency domain, which yields, for instance, in the case of the acausal dynamics,

$$\xi_{\rm atyp}(\omega) = \frac{\widetilde{\chi}(\omega)}{\chi(\omega)} \xi(\omega). \tag{81}$$

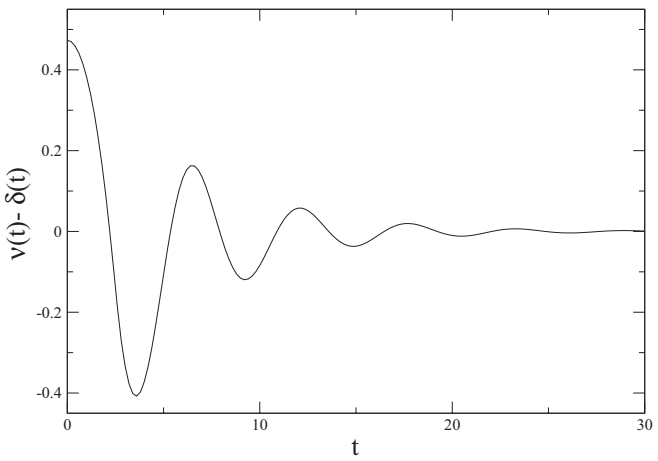


FIG. 14. Autocorrelation function of the atypical colored noise $\xi_{\rm atyp}(t)$ for $Q_0 = 2, g/Q_0 = 0.55$ and $\tau = 2.5$.

The atypical noise is thus colored, with autocorrelation function $\langle \xi_{\rm atyp}(t) \xi_{\rm atyp}(t') \rangle = \nu(t - t')$ given by

$$\begin{aligned} \nu(t) &= \delta(t) + \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \left( \frac{|\widetilde{\chi}(\omega)|^2}{|\chi(\omega)|^2} - 1 \right) e^{-i\omega t} \\ &= \delta(t) + \frac{2g}{\pi Q_0^2} \int_{-\infty}^{+\infty} d\omega \, \omega \sin(\omega\tau) |\widetilde{\chi}(\omega)|^2 e^{-i\omega t} \end{aligned} \tag{82}$$

in dimensionless units. An illustration is provided in Fig. 14 for the same model parameters used in Fig. 12(b) and Fig. 13.

## V. SUMMARY AND CLOSING REMARKS

In this paper we have investigated the nonequilibrium steady-state fluctuations of thermodynamic observables in a Brownian system subjected to a time-delayed feedback control, focusing on the behavior at large times. Our study, based on both analytical and numerical calculations, has revealed that the delay significantly affects the large-deviation statistics of time-integrated thermodynamic observables. In particular, when the state space is unbounded, delay plays a critical role in the occurrence of rare but large fluctuations of temporal boundary terms so that observables with the same typical value exhibit different large deviation rate functions.

Compared to the Markovian case, there is no doubt that the study of time-delayed systems presents some new challenges. From the perspective of stochastic thermodynamics, the most delicate issue is that the behavior of the system under time reversal is modified, which prevents standard fluctuation theorems from being satisfied. Hidden symmetries do exist, but their interpretation is more subtle, as shown in this work, and a complicated analysis of the response function of the conjugate dynamics is required even in the simplest case of a linear dynamics. In fact, it is remarkable that the large-deviation statistics, which in principle is accessible to experiments, cannot be fully elucidated without investigating the unusual properties of an acausal dynamics. Taking into account the ubiquity of time-delayed feedback loops in natural and artificial systems, there is obviously an avenue for future investigations.

## APPENDIX A: CONJUGATE DYNAMICS AND ASYMPTOTIC BEHAVIOR OF $\langle e^{-\beta \mathcal{W}_t} \rangle$ and $\langle e^{-\Sigma_t} \rangle$

In this Appendix, we show how Eqs. (33) and (36) can be used to infer the long-time behavior of the generating functions $Z_W(\lambda, t)$ and $Z_\Sigma(\lambda, t)$ for $\lambda = 1$. [On the other hand, we know from Eq. (18) that $Z_Q(1, t) = e^{(\gamma/m)t}$ at all times.] For concreteness, we restrict the discussion to the case of the linear Langevin equation (38) considered in Sec. IV. The following equations are thus expressed in terms of dimensionless parameters. For instance, the exponential factor $e^{(\gamma/m)t}$ becomes $e^{t/Q_0}$. A similar analysis has been performed in Ref. [28] in the context of heat flow in harmonic chains.

### 1. Stationary solutions of the conjugate dynamics

The first task is to determine under which conditions a stationary solution of the conjugate Langevin equations Eqs. (14) and (21) exists. (As usual, a solution is called stationary if the $n$-point probability distributions are invariant

under time translation.) In the first case of the so-called "hat" dynamics, the existence of a stationary state means that an arbitrary initial condition is eventually forgotten,

$$x(t) \approx \int_{-\infty}^{t} dt'\, \hat{\chi}(t-t')\xi(t'), \tag{A1}$$

with the response function $\hat{\chi}(t)$ decreasing sufficiently fast (typically exponentially) for $t \to +\infty$. The function $\hat{\chi}(\omega) \equiv \chi(\omega)|_{\gamma\to-\gamma}$ is then the genuine Fourier transform of $\hat{\chi}(t)$, i.e., $\hat{\chi}(\omega) = \int_{-\infty}^{\infty} dt\, e^{i\omega t}\hat{\chi}(t)$. Since $\hat{\chi}(t)$ is causal, this requires that all the poles of $\hat{\chi}(\omega)$ lie in the lower half of the complex $\omega$ plane [equivalently, all the poles of $\hat{\chi}(s) = \int_{-\infty}^{\infty} dt\, e^{-st}\hat{\chi}(t)$ lie in the l.h.s. of the complex Laplace plane $s = \sigma - i\omega$].

The case of Eq. (21) is more subtle because the so-called "tilde" dynamics is acausal. The stationary state, if it exists, must then be independent of both the initial condition in the far past *and* the final condition in the far future. Although this may seem an awkward requirement, this simply means that

$$x(t) \approx \int_{-\infty}^{+\infty} dt'\, \widetilde{\chi}(t-t')\xi(t') \tag{A2}$$

with the acausal response function $\widetilde{\chi}(t)$ decreasing sufficiently fast for both $t \to +\infty$ and $t \to -\infty$ (see Fig. 13). Then $\widetilde{\chi}(\omega) \equiv \chi(\omega)|_{\tau\to-\tau}$ is the Fourier transform of $\widetilde{\chi}(t)$, and conversely. However, as explained in paper I [see Eq. (161) and Appendix E], $\widetilde{\chi}(t)$ is more generally defined as the inverse bilateral Laplace transform of $\widetilde{\chi}(s)$, i.e., $\widetilde{\chi}(t) = 1/(2\pi i) = \int_{c-i\infty}^{c+i\infty} dt\, e^{st}\widetilde{\chi}(s)$, with the *same* Bromwich contour $\mathrm{Re}(s) = c$ as the one used for computing the quantity $\dot{S}_{\mathcal{J}}$. Therefore, for $\widetilde{\chi}(t)$ to be the inverse Fourier transform of $\widetilde{\chi}(\omega = is)$, which corresponds to $c = 0$, the bilateral Laplace transform $\widetilde{\chi}(s)$ must have two and only two poles on the l.h.s. of the complex $s$ plane. [In contrast, the functions $\widetilde{\chi}(t)$ plotted in Figs. 18 and 19 in paper I have no Fourier transform.]

Since $\hat{\chi}(s) = \widetilde{\chi}(-s)$, which is a consequence of the general relation (25) between the OM actions $\hat{S}[\mathbf{X},\mathbf{Y}]$ and $\widetilde{S}[\mathbf{X}^\dagger,\mathbf{Y}^\dagger]$, we may rephrase the conditions for the existence of a stationary state as follows: A stationary solution of Eq. (14) exists when all the poles of $\widetilde{\chi}(s)$ lie in the r.h.s of the complex $s$ plane (case 1), and a stationary solution of Eq. (21) exists when two and only two poles of $\widetilde{\chi}(s)$ lie in the l.h.s. (case 2).

The stationary distributions $\hat{p}(x,v)$ and $\widetilde{p}(x,v)$, when they exist, are bivariate Gaussians characterized by the variances of $x$ and $v$ or, equivalently, by the corresponding effective temperatures which we denote by $\hat{T}_x, \hat{T}_v$ and $\widetilde{T}_x, \widetilde{T}_v$, respectively. By definition, the variances are obtained by integrating the power spectral density over frequency. Therefore, since $|\hat{\chi}(\omega)|^2 = |\widetilde{\chi}(\omega)|^2$, the temperatures $\hat{T}_x$ and $\widetilde{T}_x$ (resp. $\hat{T}_v$ and $\widetilde{T}_v$) are given by the same formulas, i.e., in terms of dimensionless parameters:

$$\frac{2}{Q_0}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,|\hat{\chi}(\omega)|^2 = \begin{cases}\hat{T}_x/T, & \text{in case 1}\\ \widetilde{T}_x/T, & \text{in case 2}\end{cases}, \tag{A3a}$$

$$\frac{2}{Q_0}\int_{\infty}^{\infty}\frac{d\omega}{2\pi}\,\omega^2|\hat{\chi}(\omega)|^2 = \begin{cases}\hat{T}_v/T, & \text{in case 1}\\ \widetilde{T}_v/T, & \text{in case 2}\end{cases}. \tag{A3b}$$

We stress, however, that it is only for $\hat{T}_x$ and $\hat{T}_v$ that one can repeat the calculation performed in Appendix B of paper I and obtain closed-form expressions by solving the linear differential equation obeyed by the stationary time-correlation function $\hat{\phi}(t_2 - t_1) = \langle x(t_1)x(t_2)\rangle$ for $0 \leqslant |t_2 - t_1| \leqslant \tau$. The expressions of $\hat{T}_x$ and $\hat{T}_v$ are then simply obtained by changing $\gamma$ into $-\gamma$ in Eqs. (113)–(114) of paper I. One can check that this is in agreement with the numerical integration of Eqs. (A3) *only* when the stationary state exists, so that $\langle x(t_1)x(t_2)\rangle$ depends only on $t_2 - t_1$ and the calculation in Appendix B of paper I is applicable. Otherwise, one finds negative temperatures.

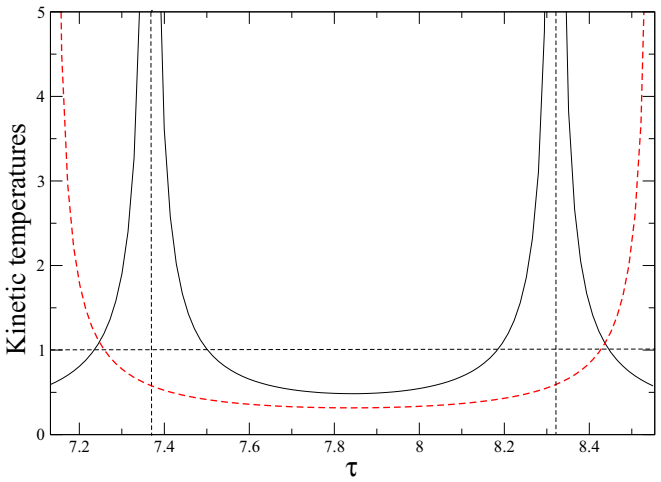


FIG. 15. Kinetic temperatures computed from Eq. (A3b): the black solid line is $\hat{T}_v/T$ for $7.37 < \tau < 8.32$ and $\widetilde{T}_v/T$ for $\tau < 7.37$ and $\tau > 8.32$. The dashed red line is the kinetic temperature $T_v/T$ of the original dynamics, which diverges at the boundaries of the stability region.

As an illustration, we plot in Fig. 15 the kinetic temperatures $\hat{T}_v$ and $\widetilde{T}_v$ computed for $Q_0 = 34.2, g/Q_0 = 0.25$, and when the system operates in the second stability lobe. As predicted by the location of the poles of $\widetilde{\chi}(s)$, a stationary state exists with the hat dynamics for $7.37 < \tau < 8.32$ and with the tilde dynamics for $\tau < 7.37$ or $\tau > 8.32$.

### 2. Asymptotic behavior of $\langle e^{-\beta\mathcal{W}_t}\rangle$ and $\langle e^{-\Sigma_t}\rangle$

We now use the preceding results to predict the long-time behavior of $Z_W(1,t) = \langle e^{-\beta\mathcal{W}_t}\rangle$ and $Z_\Sigma(1,t) = \langle e^{-\Sigma_t}\rangle$ from Eqs. (33) and (36).

#### *a. "Hat" dynamics*

We first consider the $\gamma \to -\gamma$ "hat" dynamics and set $\lambda = 1$ in Eqs. (32)–(34). Then

$$\hat{\mathcal{K}}_1[\mathbf{x}_f,t|\mathbf{Y}] = e^{t/Q_0}\int_{\mathbf{x}_i}^{\mathbf{x}_f}\mathcal{D}\mathbf{X}\,\hat{\mathcal{P}}[\mathbf{X}|\mathbf{Y}] \equiv e^{t/Q_0}\,\hat{p}(\mathbf{x}_f,t|\mathbf{Y}), \tag{A4}$$

where $\hat{p}(\mathbf{x}_f,t|\mathbf{Y})$ may be viewed as a generalized transition probability [if the "hat" process were Markovian, $\hat{p}(\mathbf{x}_f,t|\mathbf{Y})$ would be the standard transition probability $\hat{p}(\mathbf{x}_f,t|\mathbf{x}_i,0)$]. Since $\hat{f}_{Q,1}(\mathbf{x}_i,\mathbf{x}_f) = 1$, we thus have $\int d\mathbf{x}_f\,\hat{\mathcal{K}}_1[\mathbf{x}_f,t|\mathbf{Y}] = e^{t/Q_0}\int d\mathbf{x}_f\,\hat{p}(\mathbf{x}_f,t|\mathbf{Y}) = e^{t/Q_0}$ in Eq. (33), and using $\int d\mathbb{P}[\mathbf{Y}] = p(\mathbf{x}_i)$ we recover the IFT (18), as it must be.

We now assume that the conjugate Langevin equation (14) admits a stationary solution, as discussed above. Initial conditions are then irrelevant in the long-time limit, so that

$$\lim_{t\to\infty} \hat{p}(\mathbf{x}_f,t|\mathbf{Y}) = \hat{p}(\mathbf{x}_f), \tag{A5}$$

where $\hat{p}(\mathbf{x})$ is the corresponding stationary pdf. Equations (33) and (34a) then lead to the asymptotic expression

$$Z_W(1,t) \sim e^{t/Q_0} \int d\mathbf{x}_i\ p(\mathbf{x}_i) e^{\beta\mathcal{U}(\mathbf{x}_i)} \int d\mathbf{x}_f\ \hat{p}(\mathbf{x}_f) e^{-\beta\mathcal{U}(\mathbf{x}_f)}, \tag{A6}$$

which allows us to conclude that

$$\mu_W(1) = \frac{1}{Q_0} \tag{A7}$$

and

$$g_W(1) = \int d\mathbf{x}_i\ p(\mathbf{x}_i) e^{\beta\mathcal{U}(\mathbf{x}_i)} \int d\mathbf{x}_f\ \hat{p}(\mathbf{x}_f) e^{-\beta\mathcal{U}(\mathbf{x}_f)}. \tag{A8}$$

The prefactor is indeed finite as can be checked explicitly by inserting the expression (42) of $p(\mathbf{x})$ and the corresponding expression of $\hat{p}(\mathbf{x})$ (with $T_x$ and $T_v$ replaced by $\hat{T}_x$ and $\hat{T}_v$), and performing the integrations over $\mathbf{x}_i$ and $\mathbf{x}_f$. This yields Eq. (68), showing that $0 < g_W(1) < \infty$ as long as the temperatures $\hat{T}_x$ and $\hat{T}_v$ are positive.

On the other hand, from Eqs. (33) and (34c), we obtain

$$Z_\Sigma(1,t) \sim e^{t/Q_0} \int d\mathbf{x}_i \int d\mathbf{x}_f\ p(\mathbf{x}_f)\hat{p}(\mathbf{x}_f), \tag{A9}$$

which shows that the prefactor diverges. From this, we conclude that $\mu_\Sigma(1) \neq 1/Q_0$, but, unfortunately, we cannot infer the exact value.

#### b. "Tilde" dynamics

We now turn our attention to the $\tau \to -\tau$ "tilde" dynamics (21). Thanks to the linearity of the Langevin equation, the Jacobian $\widetilde{\mathcal{J}}[\mathbf{X}]$ is path independent, and setting $\lambda = 1$ in Eq. (35) yields

$$\widetilde{\mathcal{K}}_1(\mathbf{x}_f^\dagger,t|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger) = \frac{\mathcal{J}_t}{\widetilde{\mathcal{J}}_t} \int_{\mathbf{x}_i^\dagger}^{\mathbf{x}_f^\dagger} \mathcal{D}\mathbf{X}^\dagger\ \widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger,\mathbf{Y}^\dagger]. \tag{A10}$$

At first sight, this resembles Eq. (A4), with the ratio $\mathcal{J}_t/\widetilde{\mathcal{J}}_t$ replacing the exponential factor $e^{t/Q_0}$. There are two features, however, that complicate the asymptotic analysis. The first one is that while we know that $\mathcal{J}_t/\widetilde{\mathcal{J}}_t$ grows exponentially as $e^{\dot{S}_\mathcal{J} t}$, with $\dot{S}_\mathcal{J}$ given by Eq. (59), we do not know the prefactor. The second one is that $\mathbf{Y}^\dagger$ is a trajectory in the time interval $[t,t+\tau]$ (see Fig. 1). Therefore, even when the system relaxes toward a stationary state with the tilde dynamics, $\widetilde{\mathcal{K}}_1(\mathbf{x}_f^\dagger,t|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger)$ still depends on $\mathbf{Y}^\dagger$ in the long-time limit and only the dependence on $\mathbf{x}_i^\dagger$ is lost. Then, asymptotically, the quantity $\int d\mathbb{P}[\mathbf{Y}] \int_{\mathbf{x}_i^\dagger}^{\mathbf{x}_f^\dagger} \mathcal{D}\mathbf{X}^\dagger\ \widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger,\mathbf{Y}^\dagger]$ involves steady-state trajectories $\mathbf{X}^\dagger$ generated by the "tilde" dynamics ending at $\mathbf{x}_f^\dagger = (x_i, -v_i)$ and steady-state trajectories $\mathbf{Y}$ generated by the direct dynamics ending at $\mathbf{x}_i = (x_i, v_i)$. The only dependence is on $x_i$ and $v_i$, and one expects

$$\lim_{t\to\infty} \int d\mathbb{P}[\mathbf{Y}] \int_{\mathbf{x}_i^\dagger}^{\mathbf{x}_f^\dagger} \mathcal{D}\mathbf{X}^\dagger\ \widetilde{\mathcal{P}}[\mathbf{X}^\dagger|\mathbf{x}_i^\dagger,\mathbf{Y}^\dagger] \propto p(\mathbf{x}_i)\widetilde{p}(\mathbf{x}_f^\dagger), \tag{A11}$$

where $\widetilde{p}(\mathbf{x})$ is the stationary pdf of the "tilde" dynamics. The proportionality factor could in principle depend on $x_i, v_i$. However, in the small-$\tau$ limit and the associated Markovian model (see below), this factor is simply equal to 1. In the non-Markovian case, and in the overdamped limit which is simpler to analyze (see e.g. Appendix A in paper I), we have also performed an exact perturbative calculation at the second order in the amplitude of the feedback force. The outcome is again that the prefactor is constant [65]. We therefore consider as most plausible that this is the generic behavior.

As a result, we predict the following asymptotic behavior:

$$\int d\mathbb{P}[\mathbf{Y}] \widetilde{\mathcal{K}}_1(\mathbf{x}_f^\dagger,t|\mathbf{x}_i^\dagger;\mathbf{Y}^\dagger) \sim \kappa\ e^{\dot{S}_\mathcal{J} t} p(\mathbf{x}_i)\widetilde{p}(\mathbf{x}_f^\dagger), \tag{A12}$$

where $\kappa$ is some constant depending on the model parameters for which we have no expression. Fortunately, this is sufficient to infer the asymptotic behavior of $Z_W(1,t)$ and $Z_\Sigma(1,t)$. Indeed, from Eq. (36) we obtain

$$Z_A(1,t) \sim \kappa e^{\dot{S}_\mathcal{J} t} \int d\mathbf{x}_i\ p(\mathbf{x}_i)\widetilde{p}(\mathbf{x}_i) \int d\mathbf{x}_f\ \widetilde{f}_A(\mathbf{x}_i,\mathbf{x}_f), \tag{A13}$$

where we have used the fact that $p(\mathbf{x})$ and $\widetilde{p}(\mathbf{x})$ are even function of $v$ to replace $\mathbf{x}_i^\dagger$ and $\mathbf{x}_f^\dagger$ by $\mathbf{x}_f$ and $\mathbf{x}_i$, respectively. We deduce that

$$\mu_W(1) = \mu_\Sigma(1) = \mu(1) = \dot{S}_\mathcal{J} \tag{A14}$$

and

$$g_W(1) = \kappa \int d\mathbf{x}_i\ p(\mathbf{x}_i)\tilde{p}(\mathbf{x}_i) e^{\beta\mathcal{U}(\mathbf{x}_i)} \int d\mathbf{x}_f\ e^{-\beta\mathcal{U}(\mathbf{x}_f)}, \tag{A15a}$$

$$g_\Sigma(1) = \kappa \int d\mathbf{x}_i\ \tilde{p}(\mathbf{x}_i) \int d\mathbf{x}_f\ p(\mathbf{x}_f) = \kappa. \tag{A15b}$$

Interestingly, the unknown factor $\kappa$ cancels out in the ratio $g_W(1)/g_\Sigma(1)$, which yields Eq. (69). In line with the considerations above, one can check that Eqs. (A14) and (A15) are in agreement with the exact results in the Markovian small-$\tau$ limit for $\gamma > \gamma'$, with $\dot{S}_\mathcal{J} = \gamma'/m$, $g_W(1) = 1 - (\gamma'/\gamma)^2$, and $g_\Sigma(1) = 1$ (see Appendix B 2), as well as in the perturbative calculation for the overdamped limit of the non-Markovian case [65].

On the other hand, Eq. (A13) yields

$$Z_Q(1,t) \sim \kappa e^{\dot{S}_\mathcal{J} t} \int d\mathbf{x}_i\ p(\mathbf{x}_i)\widetilde{p}(\mathbf{x}_i) \int d\mathbf{x}_f, \tag{A16}$$

so that the prefactor diverges. This is expected since $\mu_Q(1) = 1/Q_0 \neq \dot{S}_\mathcal{J}$ when two poles of $\widetilde{\chi}(s)$ lie on the left-hand side of the complex $s$ plane.

## APPENDIX B: SMALL-$\tau$ LIMIT AND MARKOVIAN MODEL

In order to better understand the stationary-state fluctuations in the feedback-cooling model studied in Sec. IV, it is very useful to investigate in detail the Markovian limit obtained

by expanding the feedback force $F_{fb}(t) = k'x(t-\tau)$ at first order in $\tau$. The Langevin equation (38) then reads

$$m\dot{v}_t = -(\gamma+\gamma')v_t - \bar{k}x_t + \sqrt{2\gamma T}\xi_t, \tag{B1}$$

where $\bar{k} = k - k'$ and $\gamma' = k'\tau$. This is precisely the model studied in Refs. [8–10] whose main characteristic is the dependence of the feedback force $F_{fb}(t) = -\gamma' v_t$ on the particle's velocity. Interestingly, this induces features that are similar to those encountered in the original non-Markovian model. The bonus is that the generating functions $Z_A(\lambda,t)$ in the steady state can be computed exactly at *all* times, as shown in this appendix that revisits and extends earlier work by two of the present authors [10]. [Accordingly, to be in line with Ref. [10], we choose to work with Eq. (B1) instead of the dimensionless version.] In passing, we recall that Eq. (B1) also describes a Brownian particle coupled to two thermostats at temperatures $T$ and $T'$ in the limit $T' \to 0$. The quantity of interest in this model is the heat exchanged between the two baths, and the full expression of $Z_Q(\lambda,t)$ for $T' > 0$ was computed in Ref. [16] in the case of a free Brownian particle, i.e., for $\bar{k} = 0$ (see also Ref. [15]). For $\bar{k} > 0, Z_Q(\lambda,t)$ is only known in the long-time limit [45–47].

In Ref. [10], only the generating function $Z_\Sigma(\lambda,t)$ of the entropy production functional $\Sigma_t[\mathbf{X}] = \beta\mathcal{Q}_t[\mathbf{X}] + \ln p(\mathbf{x}_i)/p(\mathbf{x}_f)$ was considered [more precisely, it was the generating function of $\Sigma_t[\mathbf{X}] + (\gamma'/m)t$, which is the quantity called $\Delta S_p$ in Ref. [9]]. Here we generalize this calculation to also include $Z_W(\lambda,t)$ and $Z_Q(\lambda,t)$. In particular, we wish to bring to light some features that were not discussed in Ref. [10] and that are also relevant to the non-Markovian case. Note that in what follows we consider the Markovian model (B1) in its full generality, i.e., with no constraints on $\bar{k}$ and $\gamma'$ (except that they are both positive). The small-$\tau$ limit of Eq. (38) then corresponds to a restricted range of these parameters.

### 1. General expression of the generating functions

The starting point is the path-integral representation of $Z_A(\lambda,t)$ [Eq. (30)], where the dependence on $\mathbf{Y}$ is replaced by a dependence on the initial state $\mathbf{x}_i$ of the trajectory $\mathbf{X}$. This equation becomes

$$Z_A(\lambda,t) = \int d\mathbf{x}_i\; p(\mathbf{x}_i) \int d\mathbf{x}_f\; f_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f)\mathcal{K}_\lambda(\mathbf{x}_f,t|\mathbf{x}_i,0), \tag{B2}$$

where the functions $f_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f)$ are defined in Eqs. (31) and

$$\mathcal{K}_\lambda(\mathbf{x}_f,t|\mathbf{x}_i,0) = \int_{\mathbf{x}_i}^{\mathbf{x}_f} \mathcal{D}\mathbf{X}\, e^{-\lambda\beta\mathcal{W}_t}\mathcal{P}[\mathbf{X}|\mathbf{x}_i] \tag{B3}$$

with $\mathcal{W}_t[\mathbf{X}] = -\gamma'\int_0^t dt'\, v_{t'}^2$. Since the effective damping constant in Eq. (B1) is $\gamma+\gamma'$, the path probability $\mathcal{P}[\mathbf{X}|\mathbf{x}_i]$ can be expressed as

$$\mathcal{P}[\mathbf{X}|\mathbf{x}_i] \propto e^{\frac{\gamma+\gamma'}{2m}t} e^{-\beta\mathcal{S}[\mathbf{X}]} \tag{B4}$$

[see Eq. (16a)], where

$$\mathcal{S}[\mathbf{X}] = \frac{1}{4\gamma}\int_0^t dt'\,[m\dot{v}_{t'} + (\gamma+\gamma')v_{t'} + \bar{k}x_{t'}]^2. \tag{B5}$$

Hence

$$\mathcal{K}_\lambda(\mathbf{x}_f,t|\mathbf{x}_i,0) \propto e^{\frac{\gamma+\gamma'}{2m}t}\int_{\mathbf{x}_i}^{\mathbf{x}_f}\mathcal{D}\mathbf{X}\, e^{-\beta\mathcal{S}_\lambda[\mathbf{X}]}, \tag{B6}$$

where

$$\mathcal{S}_\lambda[\mathbf{X}] \equiv \mathcal{S}[\mathbf{X}] - \lambda\gamma'\int_0^t dt'\, v_{t'}^2. \tag{B7}$$

The crucial feature that distinguishes the small-$\tau$ limit and the associated Markovian model from the full non-Markovian model is that $\mathcal{S}_\lambda[\mathbf{X}]$ can be written as an Onsager-Machlup (OM) action functional for *all* values of $\lambda$. The function $\mathcal{K}_\lambda(\mathbf{x}_f,t|\mathbf{x}_i,0)$ is then a genuine transition probability, which greatly simplifies the calculation of $Z_A(\lambda,t)$ by avoiding the lengthy computation of the path integral over $\mathbf{X}$. Introducing the $\lambda$-dependent friction coefficient

$$\widetilde{\gamma}(\lambda) = [(\gamma+\gamma')^2 - 4\lambda\gamma\gamma']^{1/2}, \tag{B8}$$

we indeed obtain

$$\mathcal{S}_\lambda[\mathbf{X}] \equiv \frac{1}{4\gamma}\int_0^t dt'\,[m\dot{v}_{t'} + \widetilde{\gamma}(\lambda)v_{t'} + \bar{k}x_{t'}]^2 + \frac{\gamma+\gamma'-\widetilde{\gamma}(\lambda)}{4\gamma}\big[\bar{k}\big(x_f^2 - x_i^2\big) + m\big(v_f^2 - v_i^2\big)\big], \tag{B9}$$

and the time-extensive part of this action is the OM functional corresponding to the effective Langevin equation

$$m\dot{v}_t = -\widetilde{\gamma}(\lambda)v_t - \bar{k}x_t + \sqrt{2\gamma T}\xi_t. \tag{B10}$$

Equation (B2) then becomes

$$Z_A(\lambda,t) = e^{\frac{\gamma+\gamma'-\widetilde{\gamma}(\lambda)}{2m}t}\int d\mathbf{x}_i\; p(\mathbf{x}_i)\int d\mathbf{x}_f\, f_{A,\lambda}(\mathbf{x}_i,\mathbf{x}_f) \times e^{-\beta\frac{\gamma+\gamma'-\widetilde{\gamma}(\lambda)}{4\gamma}[\bar{k}(x_f^2-x_i^2)+m(v_f^2-v_i^2)]}\, p_{\widetilde{\gamma}}(\mathbf{x}_f,t|\mathbf{x}_i,0), \tag{B11}$$

where $p_{\widetilde{\gamma}}(\mathbf{x}_f,t|\mathbf{x}_i,0)$ is the transition probability associated with the dynamics (B10). [The extra exponential factor $e^{-\frac{\widetilde{\gamma}(\lambda)}{2m}t}$ in Eq. (B11) comes from the contribution of the effective friction coefficient $\widetilde{\gamma}(\lambda)$ to the Jacobian.] Since $\widetilde{\gamma}(0) = \gamma + \gamma'$ and $f_{A,\lambda=0}(\mathbf{x}_i,\mathbf{x}_f) = 1$, it is readily seen that $Z_A(0,t)$ is properly normalized.

To proceed further, we replace $p(\mathbf{x})$ by its expression in the stationary state [10]

$$p(\mathbf{x}) = \frac{\beta\sqrt{\bar{k}m}}{2\pi}\frac{\gamma+\gamma'}{\gamma}e^{-\beta\frac{\gamma+\gamma'}{2\gamma}[\bar{k}x^2+mv^2]}, \tag{B12}$$

and we compute $p_{\widetilde{\gamma}}(\mathbf{x}_f,t|\mathbf{x}_i,0)$ by using the relation $p_{\widetilde{\gamma}}(\mathbf{x}_f,t|\mathbf{x}_i,0) = p_{\widetilde{\gamma}}(\mathbf{x}_f,t;\mathbf{x}_i,0)/p_{\widetilde{\gamma}}(\mathbf{x}_i)$. The pdf $p_{\widetilde{\gamma}}(\mathbf{x}_i)$ is given by Eq. (B12) with $\gamma+\gamma'$ replaced by $\widetilde{\gamma}(\lambda)$, and $p_{\widetilde{\gamma}}(\mathbf{x}_f,t;\mathbf{x}_i,0)$ is given by the standard formula for the joint probability density of a two-dimensional Ornstein-Uhlenbeck process [66],

$$p_{\widetilde{\gamma}}(\mathbf{x}_f,t;\mathbf{x}_i,0) = \frac{1}{4\pi^2\sqrt{\det\boldsymbol{\Phi}}}e^{-\frac{1}{2}\mathbf{B}^T\boldsymbol{\Phi}^{-1}\mathbf{B}}, \tag{B13}$$

where

$$\mathbf{\Phi}(\lambda,t)=\begin{pmatrix}\phi_{xx}(0,\lambda) & 0 & \phi_{xx}(t,\lambda) & \phi_{xv}(t,\lambda)\\ 0 & \phi_{vv}(0,\lambda) & \phi_{xv}(-t,\lambda) & \phi_{vv}(t,\lambda)\\ \phi_{xx}(t,\lambda) & \phi_{xv}(-t,\lambda) & \phi_{xx}(0,\lambda) & 0\\ \phi_{xv}(t,\lambda) & \phi_{vv}(t,\lambda) & 0 & \phi_{vv}(0,\lambda)\end{pmatrix},$$

and

$$\mathbf{B}\equiv\begin{pmatrix}x_i\\ v_i\\ x_f\\ v_f\end{pmatrix}$$

is the four-dimensional vector representing the initial and final conditions. The functions $\phi_{xx}(t,\lambda)$, $\phi_{xv}(t,\lambda)$, and $\phi_{vv}(t,\lambda)$ are the stationary time-dependent correlation functions associated with Eq. (B10) (see Ref. [10] for the full expressions). In particular, $\phi_{xx}(0,\lambda)=\gamma T/(\widetilde{\gamma}(\lambda)\bar{k})$ and $\phi_{vv}(0,\lambda)=\gamma T/(\widetilde{\gamma}(\lambda)m)$. Plugging all these expressions into Eq. (B11) and carrying out the Gaussian integrals over $x_i$, $v_i$, and $x_f,v_f$, we finally obtain the compact result

$$Z_A(\lambda,t)=\frac{1}{\sqrt{\det(\mathbf{1}+\mathbf{\Phi}\mathbf{L}_A)}}\frac{\gamma+\gamma'}{\widetilde{\gamma}(\lambda)}e^{\mu(\lambda)t}, \tag{B14}$$

where

$$\mu(\lambda)=\frac{1}{2m}[\gamma+\gamma'-\widetilde{\gamma}(\lambda)] \tag{B15}$$

and

$$\mathbf{L}_A(\lambda)=\frac{1}{\gamma T}\begin{pmatrix}\bar{k}h_A^+(\lambda) & 0 & 0 & 0\\ 0 & mh_A^+(\lambda) & 0 & 0\\ 0 & 0 & \bar{k}h_A^-(\lambda) & 0\\ 0 & 0 & 0 & mh_A^-(\lambda)\end{pmatrix}$$

with

$$\begin{aligned}h_W^{\pm}(\lambda)&=m\mu(\lambda),\\ h_Q^{\pm}(\lambda)&=m\mu(\lambda)\pm\lambda\gamma,\\ h_\Sigma^{\pm}(\lambda)&=m\mu(\lambda)\mp\lambda\gamma'.\end{aligned} \tag{B16}$$

It turns out that $\Sigma_t=\beta\mathcal{W}_t+(\gamma'/\gamma)\beta\Delta\mathcal{U}$ in the stationary state, which explains that $h_\Sigma^{\pm}(\lambda)$ is obtained from $h_Q^{\pm}(\lambda)$ by interchanging $\gamma$ and $\gamma'$ and flipping the sign of the last term. Note also that the present definition of $\mu(\lambda)$ differs from that in Ref. [10].

$Z_A(\lambda,t)$ is a complicated function of $\lambda$ and the inverse Fourier transform can be computed only numerically. On the other hand, the long-time limit is readily obtained by noting that the matrix $\mathbf{\Phi}(\lambda,t)$ becomes diagonal when $t\to\infty$, provided $\lambda<\lambda_{\max}=(\gamma+\gamma')^2/(4\gamma\gamma')$ so that $\widetilde{\gamma}(\lambda)$ and thus $\mu(\lambda)$ are real [10]. Then

$$\lim_{t\to\infty}\sqrt{\det(\mathbf{1}+\mathbf{\Phi}\mathbf{L}_A)}=\frac{[\widetilde{\gamma}(\lambda)+h_A^+(\lambda)][\widetilde{\gamma}(\lambda)+h_A^-(\lambda)]}{\widetilde{\gamma}(\lambda)^2}, \tag{B17}$$

which leads to

$$Z_A(\lambda,t)\sim\frac{(\gamma+\gamma')\widetilde{\gamma}(\lambda)}{[\widetilde{\gamma}(\lambda)+h_A^+(\lambda)][\widetilde{\gamma}(\lambda)+h_A^-(\lambda)]}e^{\mu(\lambda)t}. \tag{B18}$$

We can thus identify the three different prefactors as

$$g_W(\lambda)=\frac{4(\gamma+\gamma')\widetilde{\gamma}(\lambda)}{[\gamma+\gamma'+\widetilde{\gamma}(\lambda)]^2}, \tag{B19a}$$

$$g_Q(\lambda)=\frac{4(\gamma+\gamma')\widetilde{\gamma}(\lambda)}{[\gamma+\gamma'+\widetilde{\gamma}(\lambda)]^2-4\lambda^2\gamma^2}, \tag{B19b}$$

$$g_\Sigma(\lambda)=\frac{4(\gamma+\gamma')\widetilde{\gamma}(\lambda)}{[\gamma+\gamma'+\widetilde{\gamma}(\lambda)]^2-4\lambda^2\gamma'^2}. \tag{B19c}$$

One can check that Eq. (B15) is also given by the general expression (54) of $\mu(\lambda)$ in the small-$\tau$ limit. Indeed, the response function $\chi(\omega)$ associated with Eq. (B1) reads

$$\chi(\omega)=[-m\omega^2-i(\gamma+\gamma')\omega+\bar{k}]^{-1}, \tag{B20}$$

and the function $H_\lambda(\omega)$ in Eq. (54) (in the original dimensionful units) is now given by

$$\begin{aligned}H_\lambda(\omega)^{-1}&=|\chi(\omega)|^{-2}-4\lambda\gamma\gamma'\omega^2\\ &=[\bar{k}-m\omega^2]^2+\omega^2[(\gamma+\gamma')^2-4\lambda\gamma\gamma'].\end{aligned} \tag{B21}$$

This can be identified for *all* values of $\lambda<\lambda_{\max}$ with the square modulus of the response function associated with the effective Langevin equation (B10),

$$\chi_\lambda(\omega)\equiv(-m\omega^2-i\widetilde{\gamma}(\lambda)\omega+\bar{k})^{-1}. \tag{B22}$$

Equation (54) then reads

$$\mu(\lambda)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}d\omega\,\ln\frac{\chi_\lambda(\omega)}{\chi_0(\omega)}, \tag{B23}$$

where we have eliminated the modulus since the imaginary part of $\chi(\omega,\lambda)$ is an odd function of $\omega$. The two poles of $\chi_\lambda(\omega)$ lie on the lower-half of the complex $\omega$ plane for all values of the parameters, and by using a contour similar to the one considered in Fig. 4 in paper I (replacing $\omega$ by $is$), one recovers Eq. (B15) from Cauchy's residue theorem. As it must be, Eq. (B15) also agrees with the expression of the SCGF obtained in Ref. [21] when the temperature $T'$ of the second thermostat is set to zero [this is also true for $g_Q(\lambda)$ given by Eq. (B19b)]. Interestingly, $\mu(\lambda)$ and the three prefactors $g_A(\lambda)$ are independent of the spring constant $\bar{k}$.

As discussed in Sec. IV B 1, the value $\lambda=1$ deserves special attention. From Eq. (B8), one obtains $\widetilde{\gamma}(1)=|\gamma-\gamma'|$, so that Eqs. (B15) and (B22) yield

$$\mu(1)=\begin{cases}\frac{\gamma'}{m}, & \text{for } \gamma\geqslant\gamma'\\ \frac{\gamma}{m}, & \text{for } \gamma'\geqslant\gamma\end{cases} \tag{B24}$$

and

$$\chi_1(\omega)=\begin{cases}[-m\omega^2-i(\gamma-\gamma')\omega+\bar{k}]^{-1}\equiv\widetilde{\chi}(\omega), & \text{for } \gamma\geqslant\gamma'\\ [-m\omega^2-i(\gamma'-\gamma)\omega+\bar{k}]^{-1}\equiv\hat{\chi}(\omega), & \text{for } \gamma'\geqslant\gamma\end{cases}, \tag{B25}$$

where $\widetilde{\chi}(\omega)$ and $\hat{\chi}(\omega)=\widetilde{\chi}(-\omega)$ are the response functions obtained from the transformations $\gamma\to-\gamma$ and $\gamma'\to-\gamma'$, respectively, which correspond to the so-called "hat" and "tilde" conjugate dynamics defined in the main text (changing $\tau$ into $-\tau$ in the small-$\tau$ limit is indeed equivalent to flipping the sign of $\gamma'$).

### 2. Fluctuations of the work

We now use Eq. (B18) to investigate how the work $\mathcal{W}_t[\mathbf{X}] = -\gamma' \int_0^t dt'\, v_{t'}^2$ fluctuates in the long-time limit. The key point is that the prefactor $g_W(\lambda)$ defined by Eq. (B19a) has no singularity, so that the LDF $I(w)$ is always given by the Legendre transform of $\mu(\lambda)$, with the saddle point $\lambda^*(w)$ solution of the equation

$$\frac{1}{m}\frac{\gamma\gamma'}{\widetilde{\gamma}(\lambda^*)} + w = 0. \tag{B26}$$

This yields

$$\lambda^*(w) = \frac{\gamma'}{4\gamma}\left[\frac{(\gamma+\gamma')^2}{\gamma'^2} - \frac{\gamma^2}{m^2w^2}\right] \tag{B27}$$

for $w < 0$, whereas there is no solution for $w > 0$. The function $\lambda^*(w)$ decreases monotonically from $\lambda_{\max}$ to $-\infty$ as $w$ increases from $-\infty$ to 0, and the LDF is then given by

$$I(w) = -\frac{(\gamma+\gamma')^2}{\gamma\gamma'}\frac{(w-\bar{w})^2}{4w} \quad \text{for } w < 0, \tag{B28}$$

where $\bar{w} = -\gamma\gamma'/[m(\gamma+\gamma')]$ is the average work rate. (Note that the LDF is defined here with the same sign as in Ref. [13], whereas the opposite convention was adopted in Ref. [10].)

We next consider the long-time behavior of $Z_W(1,t) = \langle e^{-\beta\mathcal{W}_t[\mathbf{X}]}\rangle$ to point out a mistake in Ref. [9]. According to Eq. (18) in that paper, one should have the asymptotic fluctuation relation $\lim_{t\to\infty}\langle e^{-\beta\mathcal{W}_t[\mathbf{X}]+\Delta S_{pu}(t)}\rangle = 1$, where $\Delta S_{pu}(t)$ is the so-called "entropy pumping" contribution, which is equal to $-(\gamma'/m)t$ in the present model (as the feedback force depends linearly on the velocity). On the other hand, the exact asymptotic expression (B18) yields $\lim_{t\to\infty}\langle e^{-\beta\mathcal{W}_t[\mathbf{X}]-\mu(1)t}\rangle = g_W(1)$, which is a different result. First, $\mu(1)$ is equal to $\gamma'/m$ for $\gamma \geqslant \gamma'$ only [cf. Eq. (B24)]. Second, Eq. (B19a) states that $g_W(1) = 1-(\gamma'/\gamma)^2$ for $\gamma \geqslant \gamma'$ and $g_W(1) = 1-(\gamma/\gamma')^2$ for $\gamma' \geqslant \gamma$. In both cases, this is different from 1. The error in Ref. [9] consists in assuming that $\beta\mathcal{W}_t[\mathbf{X}]$ always fluctuates like $\Sigma_t[\mathbf{X}]$ asymptotically because the two observables differ only by a temporal boundary term. However, this term may have large fluctuations of order $t$, as discussed below.

### 3. Fluctuations of the dissipated heat and the entropy production

We now turn our attention to $Z_Q(\lambda,t)$ and $Z_\Sigma(\lambda,t)$. We first notice from Eqs. (B14)–(B16) that the two generating functions are related to one another by interchanging $\gamma$ and $\gamma'$, a symmetry that is not obvious from the mere definition of the observables. Although the long-time behavior of $Z_\Sigma(\lambda,t)$ has already been investigated in Ref. [10], it is worth revisiting this analysis to stress some important points that were left aside.

We know from Eq. (18) that the heat $\mathcal{Q}_t[\mathbf{X}]$ satisfies at all times the IFT

$$\langle e^{-\beta\mathcal{Q}_t}\rangle = e^{\frac{\gamma}{m}t}. \tag{B29}$$

The symmetry $\gamma \leftrightarrow \gamma'$ thus implies that

$$\langle e^{-\Sigma_t}\rangle = e^{\frac{\gamma'}{m}t}, \tag{B30}$$

which is the IFT obtained in Ref. [9] and rederived in Ref. [10]. In the long-time limit, these two relations imply that

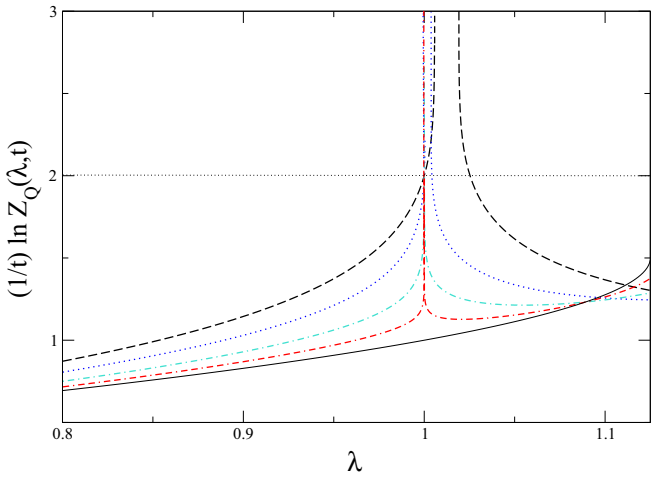


FIG. 16. Behavior of $(1/t)\ln Z_Q(\lambda,t)$ as a function of $\lambda$ in the vicinity of $\lambda = 1$ for $\gamma = 2$ and $\gamma' = 1$ ($m = 1, \bar{k} = 1, T = 1$). From top to bottom: $t = 3,5,10,25$. Observe that $(1/t)\ln Z_Q(1,t) = \gamma/m = 2$ for all values of $t$. The solid black line shows the theoretical SCGF $\mu(\lambda)$ given by Eq. (B15).

$\mu_Q(1) = \gamma/m$ and $\mu_\Sigma(1) = \gamma'/m$. Comparing with Eq. (B24) we thus see that $\mu(1)$ differs from $\mu_Q(1)$ for $\gamma > \gamma'$ and from $\mu_\Sigma(1)$ for $\gamma' > \gamma$. There is no contradiction, however, and the mismatch can be ascribed to rare but large fluctuations of the temporal boundary terms that are not included in the definition (B14) of $\mu(\lambda)$ [and more generally in the calculation that leads to Eq. (54)]. As is clear from Eqs. (B19b) and (B19c), the mathematical consequence is the divergence of the prefactors $g_Q(1)$ for $\gamma \geqslant \gamma'$ and $g_\Sigma(1)$ for $\gamma' \geqslant \gamma$.

To understand more precisely what is going on, let us investigate the behavior of $Z_Q(\lambda,t)$ for finite $t$. [Of course, the same analysis holds for $Z_\Sigma(\lambda,t)$ by changing $\gamma$ into $\gamma'$.] The key observation is that the determinant of the matrix $\mathbf{1} + \mathbf{\Phi}\,\mathbf{L}_Q$ in Eq. (B14) vanishes at $\lambda = \lambda_+(t) > 1$ and that this zero moves towards 1 as $t \to \infty$. The determinant is negative beyond this value but becomes positive again for larger values of $\lambda$. The resulting behavior of $(1/t)\ln Z_Q(\lambda,t)$ is illustrated in Fig. 16. Note that the intermediate region where the determinant is negative and $Z_Q(\lambda,t)$ imaginary shrinks as $t$ increases. As it must be, one has $(1/t)\ln Z_Q(1,t) = \gamma/m$ at all times.

A careful analysis of Eq. (B14) shows that the behavior of $Z_Q(\lambda,t)$ for $t$ large but finite and $\lambda$ close to 1 is described by the boundary-layer expression

$$Z_Q(\lambda,t) \sim \frac{(\gamma+\gamma')(\gamma-\gamma')^2 e^{t\gamma'/m}}{2\gamma^3|1-\lambda|\sqrt{1+B(u,t)}}, \tag{B31}$$

where

$$B(u,t) = \frac{(\gamma+\gamma')(\gamma-\gamma')^2}{\gamma^3}\frac{4km - (\gamma-\gamma')^2\cos(\alpha t/m)}{\alpha^2}u + \frac{(\gamma+\gamma')^2(\gamma-\gamma')^4}{4\gamma^6}u^2, \tag{B32}$$

with the scaling variable $u = (1-\lambda)^{-1}e^{-t(\gamma-\gamma')/m}$ and $\alpha = \sqrt{4\bar{k}m - (\gamma-\gamma')^2}$ (which is here assumed to be real).

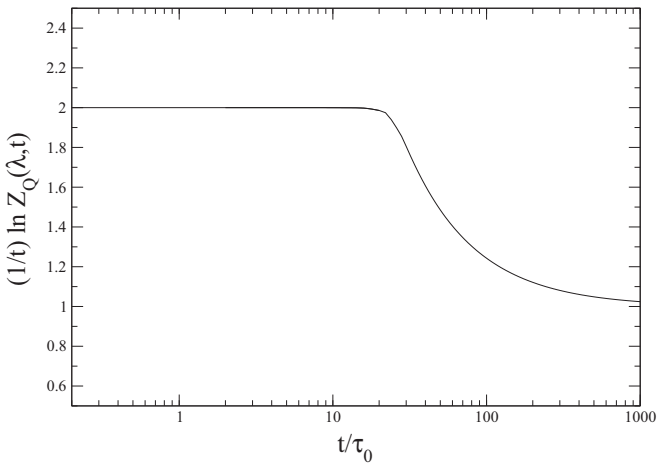


FIG. 17. Behavior of $(1/t)\ln Z_Q(\lambda,t)$ as a function of $t/\tau_0$ for $\gamma = 2, \gamma' = 1$, and $1-\lambda = 10^{-6}$ ($m = 1, \bar{k} = 1, T = 1$). Note the crossover from $\gamma$ to $\gamma'$ around $t/\tau_0 = \frac{\gamma}{\gamma-\gamma'}\ln\frac{1}{1-\lambda} \approx 28$. The crossover time decreases as $\lambda$ moves away from 1.

Accordingly, one has

$$Z_Q(\lambda,t) \sim \frac{(\gamma+\gamma')(\gamma-\gamma')^2}{2\gamma^3|1-\lambda|} e^{\frac{\gamma'}{m}t} \tag{B33}$$

for $u \ll 1$, i.e., $t/\tau_0 \gg \frac{\gamma}{\gamma-\gamma'}\ln\frac{1}{|1-\lambda|}$ (where $\tau_0 = m/\gamma$ is the viscous relaxation time for $\gamma' = 0$), and

$$Z_Q(\lambda,t) \sim e^{\frac{\gamma}{m}t} \tag{B34}$$

for $u \gg 1$, i.e., $t/\tau_0 \ll \frac{\gamma}{\gamma-\gamma'}\ln\frac{1}{|1-\lambda|}$. This crossover behavior, which is reminiscent of a smoothed dynamical first-order transition, is illustrated in Fig. 17.

As it turns out, $g_Q(\lambda)$ has also another pole at $\lambda_- = -(1 + 2\gamma'/\gamma)$, which in contrast with the pole at $\lambda = 1$ exists for both $\gamma \geqslant \gamma'$ and $\gamma' \geqslant \gamma$. More generally, for $t$ finite, $Z_Q(\lambda,t)$ diverges at $\lambda = \lambda_-(t) < \lambda_-$. This singularity moves towards $\lambda_-$ as $t$ increases and is equal to $\lambda_-$ at a *finite* critical time $t_c$. [Alternatively, when regarded as a function of $t, Z_Q(\lambda,t)$ diverges at a certain time $t(\lambda) \leqslant t_c$ for $\lambda \leqslant \lambda_-$.] The behavior of $Z_Q(\lambda,t)$ in the vicinity of $\lambda_-$ is thus different from the behavior in the vicinity of $\lambda = 1$. On the other hand, this kind of behavior is observed in other nonequilibrium models; see, e.g., Ref. [59].

Two comments are in order:

(1) The two poles of $g_Q(\lambda)$ have a different origin, as can be seen by performing the averages over the initial and final conditions in Eq. (B11) separately. The pole at $\lambda = 1$ for $\gamma \geqslant \gamma'$ comes from the average over $\mathbf{x}_f$, whereas the pole at $\lambda = \lambda_-$ comes from the average over $\mathbf{x}_i$. This can also be seen by taking the long-time limit directly in Eq. (B11) using the fact that $p_{\tilde{\gamma}}(\mathbf{x}_f,t|\mathbf{x}_f,0) \to p_{st,\tilde{\gamma}}(\mathbf{x}_f)$ as $t \to \infty$.

(2) These poles are *not* the poles of $g_{\Delta U}(\lambda)$. Indeed, a simple calculation shows that the generating function of $\Delta\mathcal{U}$ behaves asymptotically as

$$Z_{\Delta U}(\lambda,t) \sim \frac{(\gamma+\gamma')^2}{(\gamma+\gamma')^2 - \gamma^2\lambda^2}. \tag{B35}$$

Its domain of definition is thus $[-\frac{\gamma+\gamma'}{\gamma}, \frac{\gamma+\gamma'}{\gamma}]$, which is not the domain of definition of $Z_Q(\lambda,t)$. This results from the fact that the boundary term in Eq. (B11) (for $\mathcal{A}_t = \beta\mathcal{Q}_t$) does not only come from the function $f_{Q,\lambda} = e^{\lambda\beta\Delta\mathcal{U}(\mathbf{x}_i,\mathbf{x}_f)}$. In other words, $\mathcal{W}_t$ and $\Delta\mathcal{U}$ cannot be treated as uncorrelated random variables asymptotically, as is often assumed [17,20,67,68]. As a consequence, the slope of the LDF $I(q)$, which is determined by the poles of $g_Q(\lambda)$ in a certain range of $q$, is not related to the tails of the pdf of $\Delta\mathcal{U}$. Explicitly, we find

(a) For $\gamma' > \gamma$,

$$I(q) = \begin{cases} -\frac{(\gamma+\gamma')^2}{\gamma\gamma'}\frac{(q-\bar{q})^2}{4q} & \text{for } q \leqslant q_1 \\ \frac{\gamma'}{m} + \left(1 + 2\frac{\gamma'}{\gamma}\right)q & \text{for } q \geqslant q_1 \end{cases}, \tag{B36}$$

where $\bar{q} = -\gamma\gamma'/[m(\gamma+\gamma')]$ and $q_1 = -\gamma\gamma'/[m(\gamma+3\gamma')]$ (such that $\lambda^*(q_1) = \lambda_-$).

(b) For $\gamma > \gamma'$,

$$I(q) = \begin{cases} -\frac{\gamma'}{m} - q & \text{for } q \leqslant q_2 \\ -\frac{(\gamma+\gamma')^2}{\gamma\gamma'}\frac{(q-\bar{q})^2}{4q} & \text{for } q_2 \leqslant q \leqslant q_1, \\ \frac{\gamma'}{m} + \left(1 + 2\frac{\gamma'}{\gamma}\right)q & \text{for } q \geqslant q_1 \end{cases} \tag{B37}$$

where $q_2 = -\gamma\gamma'/[m(\gamma-\gamma')]$ [such that $\lambda^*(q_2) = 1$].


[1] F. Atay (ed.), *Complex Time-Delay Systems* (Springer, Berlin, 2010).
[2] W. Just, A. Pelster, M. Schanz, and E. Schöll, Phil. Trans. R. Soc. A **368** 303 (2010).
[3] E. Scholl, S. H. L. Klapp, and P. Hövel (eds.), *Control of Self-Organizing Nonlinear Systems* (Springer, Berlin, 2016).
[4] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[5] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Nat. Phys. **11**, 131 (2015).
[6] M. L. Rosinberg, T. Munakata, and G. Tarjus, Phys. Rev. E **91**, 042114 (2015).
[7] C. Jarzynski, Annu. Rev. Condensed Matter Phys. **2**, 329 (2011).
[8] K. H. Kim and H. Qian, Phys. Rev. Lett. **93**, 120602 (2004).
[9] K. H. Kim and H. Qian, Phys. Rev. E **75**, 022102 (2007).
[10] T. Munakata and M. L. Rosinberg, J. Stat. Mech. (2012) P05010.
[11] K. Sekimoto, Stochastic Energetics, *Lecture Notes in Physics* Vol. 799 (Springer, Berlin, 2010).
[12] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[13] H. Touchette, Phys. Rep. **478**, 1 (2009).
[14] R. van Zon and E. G. D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003); Phys. Rev. E **69**, 056121 (2004).
[15] J. Farago, J. Stat. Phys. **107**, 781 (2002); Physica A **331**, 69 (2004).
[16] P. Visco, J. Stat. Mech. (2006) P06006.
[17] M. Baiesi, T. Jacobs, C. Maes, and N. S. Skantzos, Phys. Rev. E **74**, 021111 (2006).
[18] A. Puglisi, L. Rondoni, and A. Vulpiani, J. Stat. Mech. (2006) P08010.
[19] R. J. Harris, A. Rákos, and G. M. Schütz, Europhys. Lett. **75**, 227 (2006); A. Rákos and R. J. Harris, Stat. Mech. (2008) P05005.

[20] T. Taniguchi and E. G. D. Cohen, J. Stat. Phys. **126**, 1 (2007); **130**, 1 (2008).

[21] S. Sabhapandit, Euro. Phys. Lett. **96**, 20005 (2011); Phys. Rev. E. **85**, 021108 (2012).

[22] T. Nemoto, Phys. Rev. E **85**, 061124 (2012).

[23] J. D. Noh and J.-M. Park, Phys. Rev. Lett **108**, 240603 (2012); J. S. Lee, C. Kwon, and H. Park, Phys. Rev. E **87**, 020104(R) (2013); J. D. Noh, J. Stat. Mech. (2014) P01013.

[24] K. Kim, C. Kwon, and H. Park, Phys. Rev. E **90**, 032117 (2014).

[25] N. Garnier and S. Ciliberto, Phys. Rev. E **71**, 060101(R) (2005); S. Joubaud, N. B. Garnier, and S. Ciliberto, J. Stat. Mech. (2007) P09018; S. Ciliberto, S. Joubaud, and A. Petrosyan, *ibid.* (2010) P12003.

[26] E. Falcon *et al.*, Phys. Rev. Lett. **100**, 064503 (2008).

[27] M. Bonaldi, L. Conti, P. DeGregorio, L. Rondoni, G. Vedovato, A. Vinante, M. Bignotto, M. Cerdonio, P. Falferi, N. Liguori, S. Longo, R. Mezzena, A. Ortolan, G. A. Prodi, F. Salemi, L. Taffarello, S. Vitale, and J. P. Zendri, Phys. Rev. Lett. **103**, 010601 (2009).

[28] M. L. Rosinberg, G. Tarjus, and T. Munakata, Eur. Phys. Lett. **113**, 10007 (2016).

[29] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953); S. Machlup and L. Onsager, *ibid.* **91**, 1512 (1953).

[30] A. Imparato and L. Peliti, Phys. Rev. E **74**, 026106 (2006).

[31] C. Aron, G. Biroli, and L. F. Cugliandolo, J. Stat. Mech. (2010) P11018.

[32] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998); Phys. Rev. E **60**, 2721 (1999); **61**, 2361 (2000).

[33] T. Munakata and M. L. Rosinberg, Phys. Rev. Lett. **112**, 180601 (2014).

[34] S. F. Norrelykke and H. Flyvbjerg, Phys. Rev. E **83**, 041103 (2011).

[35] M. Poot and H. S. J. van der Zant, Phys. Rep. **511**, 273 (2012).

[36] M. Pinard, P. F. Cohadon, T. Briant, and A. Heidmann, Phys. Rev. A **63**, 013808 (2000).

[37] M. Montinaro *et al.*, Appl. Phys. Lett. **101**, 133104 (2012).

[38] M. Poggio, C. L. Degen, H. J. Mamin, and D. Rugar, Phys. Rev. Lett. **99**, 017201 (2007).

[39] In passing, this implies that the stochastic entropy production defined as the relative likelihood of a trajectory and its time reverse [51] is zero.

[40] As discussed in the Appendix A of paper I, we have been able only to compute $\mathcal{P}[\mathbf{Y}]$ in the overdamped limit and for $t \leqslant \tau$.

[41] J. R. Gomez-Solano, L. Bellon, A. Petrosyan, and S. Ciliberto, Euro. Phys. Lett. **89**, 60003 (2010).

[42] R. Mannella, Int. J. Mod. Phys. C **13**, 1177 (2002).

[43] C. M. Rohwer, F. Angeletti, and H. Touchette, Phys. Rev. E **92**, 052104 (2015).

[44] F. Zamponi, F. Bonetto, L. F. Cugliandolo, and J. Kurchan, J. Stat. Mech. (2005) P09013.

[45] A. Kundu, S. Sabhapandit, and A. Dhar, J. Stat. Mech. (2011) P03007.

[46] K. Saito and A. Dhar, Phys. Rev. E **83**, 041121 (2011).

[47] H. C. Fogedby and A. Imparato, J. Stat. Mech. (2012) P04005.

[48] M. L. Rosinberg and J. M. Horowitz, Eur. Phys. Lett. **116**, 10007 (2016).

[49] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).

[50] J. Kurchan, J. Phys. A: Math. Gen. **31**, 3719 (1998).

[51] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).

[52] We could as well replace $\widetilde{\chi}(\omega)$ in Eq. (58) by $\hat{\chi}(\omega) = \widetilde{\chi}(-\omega)$, the response function of the $\gamma \to -\gamma$ "hat" dynamics, since $H_1(\omega)$ is an even function of $\omega$. However, we prefer to use $\widetilde{\chi}(\omega)$ in order to directly exploit the analysis performed in paper I.

[53] Take, for instance, the derivative with respect to $g$, close the integration contour by a large semicircle on the r.h.s. and use Jordan's lemma for integrals of the type $\int dz e^{iz} f(z)$. In passing, note that causality is not a sufficient condition for the integral to vanish. The conditions of Jordan's lemma must also be satisfied. For instance, if one expands $\chi(s)$ at first order in $\tau$, one has $\lim_{|z|\to\infty} |zf(z)| \neq 0$ and the integral is then equal to $g\tau/(2Q_0)$.

[54] Case (c) occurs, for instance, for $Q_0 = 2, g/Q_0 = 0.55$ and $\tau = 8.5$, which corresponds to the second stability lobe in Fig. 2 of paper I. Then $\mu(1) \approx 0.040$ and $\dot{S}_{\mathcal{J}} \approx 0.075$, whereas $1/Q_0 = 1/2$. Note that the feedback control heats the system for these values of the parameters, as can be seen in Figs. 10 and 16 of paper I.

[55] The cusp in $\dot{S}_{\mathcal{J}}$ results from the intricate evolution of the poles of $\widetilde{\chi}(s)$ with $\tau$, as explained in paper I. This nonanalytical behavior is rounded in the numerical estimate of $\mu_\Sigma(1)$, presumably because of finite-time effects.

[56] F. Douarche, S. Joubaud, N. B. Garnier, A. Petrosyan, and S. Ciliberto, Phys. Rev. Lett. **97**, 140603 (2006).

[57] In particular, for $\tau = 7.6$ and $\lambda = 1$, we find $g_W(1) \approx 0.95$, in very good agreement with the prediction of Eq. (68).

[58] As shown in Appendix B for the small-$\tau$ limit and the associated Markovian model, these poles have a different nature than the poles at $\lambda = 1$. They also move with $t$ but lock at a certain value $\lambda_- < 0$ at a finite time. Therefore, there is no boundary layer for large $t$, so that the SCGF is not discontinuous and is equal to $\mu(\lambda_-)$. This kind of behavior is observed in other nonequilibrium models; see, e.g., Ref. [59].

[59] C. Kwon, J. D. Noh, and H. Park, Phys. Rev. E **83**, 061145 (2011).

[60] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).

[61] J. P. Garrahan and I. Lesanovsky, Phys. Rev. Lett. **104**, 160601 (2010).

[62] R. L. Jack and P. Sollich, Prog. Theor. Phys. Supp. **184**, 304 (2010).

[63] R. Chetrite and H. Touchette, Phys. Rev. Lett. **111**, 120601 (2013); Ann. Inst. Poincaré A **16**, 2005 (2015).

[64] Unfortunately, this is the only case where such a driven process can be explicitly defined. For $\lambda \neq 1$, the non-Markovian character of the dynamics makes it impossible to use the spectral approach described in Ref. [63].

[65] The calculation, which involves expanding all quantities, including path probabilities, up to second order in the feedback gain, is long and very technical, and we do not give the details here. They are available upon request.

[66] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, Berlin, 1989).

[67] G. Gradenigo, A. Sarracino, A. Puglisi, and H. Touchette, J. Phys. A: Math. Theor. **46**, 335002 (2013).

[68] H. Touchette and E. G. D. Cohen, Phys. Rev. E **80**, 011114 (2009).