



# Nonequilibrium Markov Processes Conditioned on Large Deviations

Raphaël Chetrite, Hugo Touchette

## ► To cite this version:

Raphaël Chetrite, Hugo Touchette. Nonequilibrium Markov Processes Conditioned on Large Deviations. Annales de l'Institut Henri Poincaré, 2015, pp.1. 10.1007/s00023-014-0375-8 . hal-01144629

**HAL Id: hal-01144629**

**<https://hal.science/hal-01144629>**

Submitted on 22 Apr 2015

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

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

# Nonequilibrium Markov processes conditioned on large deviations

Raphaël Chetrite

*Laboratoire J. A. Dieudonné, UMR CNRS 7351,  
Université de Nice Sophia Antipolis, Nice 06108, France*

Hugo Touchette

*National Institute for Theoretical Physics (NITheP), Stellenbosch 7600, South Africa and  
Institute of Theoretical Physics, Stellenbosch University, Stellenbosch 7600, South Africa*

(Dated: September 10, 2014)

We consider the problem of conditioning a Markov process on a rare event and of representing this conditioned process by a conditioning-free process, called the effective or driven process. The basic assumption is that the rare event used in the conditioning is a large-deviation-type event, characterized by a convex rate function. Under this assumption, we construct the driven process via a generalization of Doob's  $h$ -transform, used in the context of bridge processes, and show that this process is equivalent to the conditioned process in the long-time limit. The notion of equivalence that we consider is based on the logarithmic equivalence of path measures, and implies that the two processes have the same typical states. In constructing the driven process, we also prove equivalence with the so-called exponential tilting of the Markov process, often used with importance sampling to simulate rare events and giving rise, from the point of view of statistical mechanics, to a nonequilibrium version of the canonical ensemble. Other links between our results and the topics of bridge processes, quasi-stationary distributions, stochastic control, and conditional limit theorems are mentioned.

Keywords: Markov processes, large deviations, conditioning, nonequilibrium processes, microcanonical and canonical ensembles

# CONTENTS

I. Introduction	3
II. Notations and definitions	7
A. Homogeneous Markov processes	7
B. Pure jump processes and diffusions	9
C. Conditioning observables	10
D. Large deviation principle	11
E. Nonequilibrium path ensembles	12
F. Process equivalence	13
III. Non-conservative tilted process	14
A. Definition	14
B. Spectral elements	15
C. Marginal canonical density	16
IV. Generalized Doob transform	18
A. Definition	18
B. Historical conditioning of Doob	20
V. Driven Markov process	24
A. Definition	24
B. Equivalence with the canonical path ensemble	25
C. Equivalence with the microcanonical path ensemble	27
D. Invariant density	30
E. Reversibility properties	31
F. Identities and constraints	32
VI. Applications	34
A. Extensive Brownian bridge	34
B. Ornstein-Uhlenbeck process	35
C. Quasi-stationary distributions	36
A. Derivation of the tilted generator	37
1. Pure jump processes	37
2. Diffusion processes	38
B. Change of measure for the generalized Doob transform	39
C. Squared field for diffusion processes	39
D. Generator of the canonical path measure	39
E. Markov chains	40
Acknowledgments	41
References	41

## I. INTRODUCTION

We treat in this paper the problem of conditioning a Markov process  $X_t$  on a rare event  $\mathcal{A}_T$  defined on the time interval  $[0, T]$ , and of representing this conditioned Markov process in terms of a conditioning-free Markov process  $Y_t$ , called the *effective* or *driven process*, having the same typical states as the conditioned process in the stationary limit  $T \rightarrow \infty$ . More abstractly, this means that we are looking for a Markov process  $Y_t$  such that

$$X_t|\mathcal{A}_T \cong Y_t, \quad (1)$$

where  $X_t|\mathcal{A}_T$  stands for the conditioned process and  $\cong$  is an asymptotic notion of process equivalence, related to the equivalence of ensembles in statistical physics, which we will come to define in a precise way below. Under some conditions on  $X_t$ , and for a certain class of large-deviation-type events  $\mathcal{A}_T$ , we will show that  $Y_t$  exists and is unique, and will construct its generator explicitly.

This problem can be considered as a generalization of Doob's work on Markov conditioning [1, 2] and also finds its source, from a more applied perspective, in many fundamental and seemingly unrelated problems of probability theory, stochastic simulations, optimal control theory, and nonequilibrium statistical mechanics. These are briefly discussed next to set the context of our work:

**Conditioned Markov processes:** Doob was the first historically to consider conditioning of Markov processes, starting with the Wiener process conditioned on leaving the interval  $[0, \ell]$  at the boundary  $\{\ell\}$  [1, 2]. In solving this problem, he introduced a transformation of the Wiener process, now referred to as *Doob's h-transform*, which was later adapted under the same name to deal with other conditionings of stochastic processes, including the Brownian bridge [3], Gaussian bridges [4–6], and the Schrödinger bridge [7–11], obtained by conditioning a process on reaching a certain target distribution in time as opposed to a target point. Doob's transform also appears prominently in the theory of quasi-stationary distributions [12–16], which describes in the simplest case the conditioning of a process never to reach an absorbing state.

We discuss some of these historical examples in Sec. IV to explain how Doob's original transform relates to the large deviation conditioning considered here. Following this section, we will see that the construction of the driven process  $Y_t$  also gives rise to a process transformation, which is however different from Doob's transform because of the time-integrated character of the conditioning  $\mathcal{A}_T$  considered.

**Gibbs conditioning and conditional limit theorems:** Let  $X_1, \dots, X_n$  be a sequence of independent and identically distributed random variables with common distribution  $P(x)$  and let  $S_n$  denote their sample mean:

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (2)$$

A conditional limit theorem for this sequence refers to the distribution of  $X_1$  obtained in the limit  $n \rightarrow \infty$  when the whole sequence  $X_1, \dots, X_n$  is conditioned on  $S_n$  being in a certain interval or on  $S_n$  assuming a certain value. In the latter case, it is known that, under some conditions on  $P(x)$ ,

$$\lim_{n \rightarrow \infty} P\{X_1 = x | S_n = s\} = \frac{P(x) e^{kx}}{W(k)} \equiv P_k(x), \quad (3)$$

where  $k$  is a real parameter related to the conditioning value  $s$  and  $W(k)$  is the generating function of  $P(x)$  normalizing the so-called *exponentially tilted* distribution  $P_k(x)$ ; see [17–20] for details.

This asymptotic conditioning of a sequence of random variables is sometimes referred to as *Gibbs conditioning* [21] because of its similarity with the construction of the microcanonical ensemble of statistical mechanics, further discussed below. Other limit theorems can be obtained by considering sub-sequences of  $X_1, \dots, X_n$  instead of  $X_1$ , as above (see [22–24]), or by assuming that the  $X_i$ 's form a Markov chain instead of being independent [25, 26].

This paper came partly as an attempt to generalize these results to general Markov processes and, in particular, to continuous-time processes. The essential step needed to arrive at these results is the derivation of the driven process; the conditional limit theorems that follow from this process will be discussed in a future publication.

**Rare event simulations:** Many numerical methods used for determining rare event probabilities are based on the idea of importance sampling, whereby the underlying distribution  $P$  of a random variable or process is modified to a target distribution  $Q$  putting more weight on the rare events to be sampled [27]. A particularly useful and general distribution commonly used in this context is the exponentially tilted distribution  $P_k$  mentioned earlier, which is also known as the *exponential family* or *Esscher transform* of  $P$  [28]. Such a distribution can be generalized to sequences of random variables, as well as paths of stochastic processes (as a path measure), and corresponds, from the point of view of statistical mechanics, to the canonical ensemble distribution describing a thermodynamic system coupled to a heat bath with inverse temperature  $\beta = -k$ .

This link with statistical mechanical ensembles is discussed in more detail below. For the conditioning problem treated here, we make contact with  $P_k$  by using this distribution as an intermediate step to construct the driven process  $Y_t$ , as explained in Secs. III and V. An interesting byproduct of this construction is that we can interpret  $Y_t$  as a modified Markov process that asymptotically realizes (in a sense to be made precise below) the exponential tilting of  $X_t$ .

A further link with rare event sampling is established in that the semi-group or propagator of  $Y_t$  is deeply related to Feynman-Kac functionals, which underlie cloning [29–31] and genealogical [32] methods also used for sampling rare events. In fact, we will see in Sec. II that the driven process  $Y_t$  is essentially a normalized version of a non-conservative process, whose generator is the so-called tilted generator of large deviations, and whose dominant eigenvalue (when it exists) is the so-called scaled cumulant generating function – the main quantity obtained by cloning methods [29–31].

**Stochastic control and large deviations:** The generalization of Doob's transform that we will discuss in Sec. IV has been considered by Fleming and Sheu in their work on control representations of Feynman-Kac-type partial differential equations (PDEs) [33–36]. The problem here is to consider a linear operator of the form  $L + V(x)$ , where  $L$  is the generator of a Markov process, and to provide a stochastic representation of the solution  $\phi(x, t)$  of the backward PDE

$$\frac{\partial \phi}{\partial t} + (L + V)\phi = 0, \quad t \leq T \quad (4)$$

with final condition  $\phi(x, T) = \Phi(x)$ . The Feynman-Kac formula [37–39] provides, as is well known, a stochastic representation of  $\phi(x, t)$  in terms of the expectation

$$\phi(x, t) = \mathbb{E}[\Phi(X_T) e^{\int_t^T V(X_s) ds} | X_t = x]. \quad (5)$$

The idea of Fleming and Sheu is to consider, instead of  $\phi$ , the *logarithm* or *Hopf-Cole transform*  $I = -\ln \phi$ , which solves the Hamilton-Jacobi-like PDE,

$$\frac{\partial I}{\partial t} + (HI) - V(x) = 0, \quad (6)$$

where  $(HI) = -e^I (Le^{-I})$ , and to find a controlled process  $X_t^u$  with generator  $L^u$ , so as to rewrite

(6) as a dynamic programming equation:

$$\frac{\partial I}{\partial t} + \min_u \{(L^u I)(x) + k_V(x, u)\} = 0, \quad (7)$$

where  $k_V(x, u)$  is some cost function that depends on  $V$ , the system's state, and the controller's state. In this form, they show that  $I$  represents the value function of the control problem, involving a Lagrangian dual to the Hamiltonian  $H$ ; see [40] for a more detailed description.

These results have been applied by Fleming and his collaborators to give control representations of various distributions related to exit problems [33–36], dominant eigenvalues of linear operators [41–43], and optimal solutions of sensitive risk problems [44–46], which aim at minimizing functionals having the exponential form of (5). What is interesting in all these problems is that the generator  $L^u$  of the optimally-controlled process is given by a Doob transformation similar to the one we use to construct the conditioned process. In their work, Fleming *et al.* do not interpret this transformation as a conditioning, but as an optimal change of measure between the controlled and reference processes. Such a change of measure has also been studied in physics more recently by Nemoto and Sasa [47–49]. We will discuss these links in more details in a future publication.

**Fluctuation paths and fluctuation dynamics:** It well known that rare transitions in dynamical systems perturbed by a small noise are effected by special trajectories known as reaction paths, fluctuation paths, most probable paths or instantons; see [50] for a review. These paths are described mathematically by the Freidlin-Wentzell theory of large deviations [51], and are fundamental for characterizing many noise-activated (escape-type) processes arising in chemical reactions, biological processes, magnetic systems, and glassy systems [52–54].

The concept of fluctuation path is specific to the low-noise limit: for processes with arbitrary random perturbations, there is generally not a single fluctuation path giving rise to a rare event, but many different fluctuation paths leading to the same event, giving rise to what we call a *fluctuation dynamics*. The driven process that we construct in this paper is a specific example of such a fluctuation dynamics: it describes the effective dynamics of  $X_t$  as this process is seen to fluctuate away from its typical behavior to ‘reach’ the event  $\mathcal{A}_T$ . Consequently, it can be used to simulate or sample this fluctuation in an efficient way, bringing yet another connection with rare event simulations. This will be made clearer as we come to define this process in Sec. V.

**Statistical ensembles for nonequilibrium systems:** The problem of defining or extending equilibrium statistical ensembles, such as the microcanonical and canonical ensembles, to nonequilibrium systems has a long history in physics. It was revived recently by Evans [55–57], who proposed to derive the transition rates of a system driven by external forces in a stationary nonequilibrium state by conditioning the transition rates of the same system when it is not driven, that is, when it is in an equilibrium state with transition rates satisfying detailed balance. Underlying this proposal is the interesting idea that nonequilibrium systems driven in steady states could be seen as equilibrium systems in which the driving is effected by a conditioning. This means, for example, that a driven nonequilibrium system having a given stationary particle current could be thought of, physically, as being equivalent to a non-driven equilibrium system in which this current appears as a fluctuation.

The validity of this idea needs to be tested using examples of driven physical systems for which nonequilibrium stationary solutions can be obtained explicitly and be compared with conditionings of their equilibrium solutions. Our goal here is not to provide such a test, but to formalize the problem in a clear, mathematical way as a Markov conditioning problem based on large deviations. This leads us to define in a natural way a nonequilibrium generalization of the microcanonical ensemble for trajectories or paths of Markov processes, as well as a nonequilibrium version of the canonical ensemble, which is a path version of the exponentially tilted measure  $P_k$ .

The latter ensemble has been used recently with transition path sampling [58–61] to simulate rare trajectories of nonequilibrium systems associated with glassy phases and dynamical phase

transitions; see [62] for a recent review. In this context, the exponentially tilted distribution  $P_k$  is referred to as the *biased*, *tilted* or *s-ensemble*, the latter name stemming from the fact that the symbol  $s$  is used instead of  $k$  [62–66]. These simulations follow exactly the idea of importance sampling mentioned earlier: they re-weight the statistics of the trajectories or paths of a system in an exponential way so as to reveal, in a typical way, trajectories responsible for certain states or phases that are atypical in the original system. In Sec. V, we will give conditions that ensure that this exponential re-weighting is equivalent to a large deviation conditioning – in other words, we will give conditions ensuring that the path canonical ensemble is equivalent to the path microcanonical ensemble.

The connection with the driven process is established from this equivalence by showing that the canonical ensemble can be realized by a Markov process in the long-time limit. Some results on this canonical-Markov connection were obtained by Jack and Sollich [64] for a class of jump processes and by Garrahan and Lesanovsky [67] for dissipative quantum systems (see also [68–71]). Here we extend these results to general Markov processes, including diffusions, and relate them explicitly to the conditioning problem.

These connections and applications will not be discussed further in the paper, but should hopefully become clearer as we define the driven process and study its properties in the next sections. The main steps leading to this process are summarized in [72]; here we provide the full derivation of this process and discuss, as mentioned, its link with Doob’s results. We also discuss new results related to constraints satisfied by the driven process, as well as special cases of these results for Markov chains, jump processes, and pure diffusions.

The plan of the paper is as follows. In Sec. II, we define the class of general Markov processes and conditioning events (or observables) that we consider, and introduce there various mathematical concepts (Markov semi-groups, Markov generators, path measures) used throughout the paper. We also define in that section the path versions of the microcanonical and canonical ensembles, corresponding respectively to the conditioning and exponential tilting of  $X_t$ , and introduce all the elements of large deviation theory needed to define and study our class of rare event conditioning. We then proceed to construct the driven process  $Y_t$  and prove its equivalence with the conditioned process  $X_t|\mathcal{A}_T$  in three steps. Firstly, we construct in Sec. III a non-conservative process from which various spectral elements, related to the large deviation conditioning, are obtained. Secondly, we study in Sec. IV the generalization of Doob’s transform needed to construct  $Y_t$ , and show how it relates to the original transform considered by Doob. Thirdly, we use the generalized transform to define in Sec. V the driven process proper, and show that it is equivalent to the conditioned process by appealing to general results about ensemble equivalence.

Our main results are contained in Sec. V. Their novelty, compared to previous works, resides in that we treat the equivalence of the driven and conditioned processes explicitly via path versions of the canonical and microcanonical ensembles, derive precise conditions for this equivalence to hold, and express all of our results in the general language of Markov generators, which can be used to describe jump processes, diffusions, or mixed processes, depending on the physical application considered. New properties of the driven process, including constraint rules satisfied by its transition rates or generator, are also discussed in that section. Section VI finally presents some applications of our results for diffusions, to show how the driven process is obtained in practice, and for absorbing Markov chains, to make a connection with quasi-stationary distributions. The specialization of our results to Markov chains is summarized in the Appendices, which also collect various technical steps needed for proving our results.

## II. NOTATIONS AND DEFINITIONS

We define in this section the class of Markov processes and observables of these processes that we use to define the rare event conditioning problem. Markov processes are widely used as models of stochastic systems, for example, in the context of financial time series [73], biological processes [74], and chemical reactions [52–54]. In physics, they are also used as a general framework for modeling systems driven in nonequilibrium steady states by noise and external forces [52–54], such as interacting particle systems coupled to different particle and energy reservoirs, which have been studied actively in the mathematics and physics literature recently [75–79]. For general introductions to Markov processes and their applications in physics, see [52–54, 80–82]; for references on the mathematics of these processes, see [3, 38, 39, 83, 84].

### A. Homogeneous Markov processes

We consider a homogeneous continuous-time Markov process  $X_t$ , with  $t \in \mathbb{R}_+$ , taking values in some space  $\mathcal{E}$ , which, for concreteness, is assumed to be  $\mathbb{R}^d$  or a counting space.<sup>1</sup> The dynamics of  $X_t$  is described by a transition kernel  $P_t(x, dy)$  giving the conditional probability that  $X_{t+t'} \in dy$  given that  $X_{t'} = x$  with  $t \geq 0$ . This kernel satisfies the Chapman-Kolmogorov equation

$$\int_{\mathcal{E}} P_{t'}(x, dy) P_t(y, dz) = P_{t'+t}(x, dz) \quad (8)$$

for all  $(x, y) \in \mathcal{E}^2$ , and is homogeneous in the sense that it depends only on the time difference  $t$  between  $X_{t+t'}$  and  $X_{t'}$ . Here, and in the following,  $dy$  stands for the Lebesgue measure or the counting measure, depending on  $\mathcal{E}$ .

To ensure that  $X_t$  is well behaved, we assume that it admits càdlàg<sup>2</sup> paths as a function of time for every initial condition  $X_0 = x \in \mathcal{E}$ . We further assume that

$$\int_{\mathcal{E}} P_t(x, dy) = 1, \quad (9)$$

so that the probability is conserved at all times. This property is also expressed in the literature by saying that  $X_t$  is *conservative*, *honest*, *stochastically complete* or *strictly Markovian*, and only means physically that there is no killing or creation of probability. Although  $X_t$  is assumed to be conservative, we will introduce later a non-conservative process as an intermediate mathematical step to construct the driven process. In what follows, it will be clear when we are dealing with a conservative or non-conservative process. Moreover, it should be clear that the word ‘conservative’ is not intended here to mean that energy is conserved.

Mathematically, the transition kernel can be thought of as a positive linear operator<sup>3</sup> acting on the space of bounded measurable functions  $f$  on  $\mathcal{E}$  according to

$$(P_t f)(x) \equiv \int_{\mathcal{E}} P_t(x, dy) f(y) \equiv \mathbb{E}_x[f(X_t)] \quad (10)$$

for all  $x \in \mathcal{E}$ , where  $\mathbb{E}_x[\cdot]$  denotes the expectation with initial condition  $X_0 = x$ . In many cases, it is more convenient to give a local specification of the action of  $P_t$  via its *generator*  $L$  according to

$$\partial_t \mathbb{E}_x[f(X_t)] = \mathbb{E}_x[(Lf)(X_t)], \quad (11)$$

<sup>1</sup> In probability theory,  $\mathcal{E}$  is most often taken to be a so-called Polish (metric, separable and complete) topological space.

<sup>2</sup> From the French ‘continue à droite, limite à gauche’: right continuous with left limit.

<sup>3</sup> This operator is positive in the Perron-Frobenius sense, that is,  $(P_t f) \geq 0$  for all  $f \geq 0$ .



where  $(Lf)$  denotes the application of  $L$  on  $f$ . Formally, this is equivalent to the representation

$$P_t = e^{tL}, \quad (12)$$

and the forward and backward Kolmogorov equation, given by

$$\partial_t P_t = P_t L = L P_t, \quad P_0 = I, \quad (13)$$

where  $I$  is the identity operator. For  $P_t(x, dy)$  to be conservative, the generator must obey the relation  $(L1) = 0$ , where 1 is the constant function equal to 1 on  $\mathcal{E}$ .

In the following, we will appeal to a different characterization of  $X_t$  based on the path probability measure  $d\mathbb{P}_{L, \mu_0, T}(\omega)$  representing, roughly speaking, the probability of a trajectory or sample path  $\{X_t(\omega)\}_{t=0}^T$  over the time interval  $[0, T]$ , with  $X_0(\omega)$  chosen according to the initial measure  $\mu_0$ . Technically, the space of such paths is defined as the so-called *Skorohod space*  $D([0, T], \mathcal{E})$  of càdlàg functions on  $\mathcal{E}$ , while  $d\mathbb{P}_{L, \mu_0, T}(\omega)$  is defined in terms of expectations having the form

$$\mathbb{E}_{\mu_0}[C] = \int C(\omega) d\mathbb{P}_{L, \mu_0, T}(\omega), \quad (14)$$

where  $C$  is any bounded measurable functional of the path  $\{X_t(\omega)\}_{t=0}^T$ , and  $\mathbb{E}_{\mu_0}$  now denotes the expectation with initial measure  $\mu_0$ . As usual, this expectation can be simplified to completely characterize  $\mathbb{P}_{L, \mu_0, T}$  by considering so-called *cylinder functions*,

$$C(\omega) = C(X_0(\omega), X_{t_1}(\omega), \dots, X_{t_{n-1}}(\omega), X_T(\omega)), \quad (15)$$

involving  $X_t$  over a finite sequence of times  $0 \leq t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq T$  instead of the whole interval  $[0, T]$ . At this level, the path probability measure becomes a joint probability distribution over these times, given in terms of  $L$  by

$$\mathbb{P}_{L, \mu_0, T}(dx_0, \dots, dx_n) = \mu_0(dx_0) e^{t_1 L}(x_0, dx_1) e^{(t_2 - t_1)L}(x_1, dx_2) \dots e^{(T - t_{n-1})L}(x_{n-1}, dx_n), \quad (16)$$

where the exponentials refer to the operator of (12).

One important probability measure obtained from the path measure is the marginal  $\mu_t$  of  $X_t$ , associated with the single-time cylinder expectation,

$$\mathbb{E}_{\mu_0}[C(X_t)] = \int_{\mathcal{E}} C(y) \mu_t(dy). \quad (17)$$

This measure is also obtained by ‘propagating’ the initial measure  $\mu_0$  according to (16):

$$\mu_t(dy) = \int_{\mathcal{E}} \mu_0(dx_0) e^{tL}(x_0, dy). \quad (18)$$

It then follows from the Kolmogorov equation (13) that

$$\partial_t \mu_t(x) = (L^\dagger \mu_t)(x), \quad (19)$$

where  $L^\dagger$  is the formal adjoint of  $L$  with respect to the Lebesgue or counting measure. In physics, this equation is referred to as the *Master equation* in the context of jump processes or the *Fokker-Planck equation* in the context of diffusions.

The time-independent probability measure  $\mu_{\text{inv}}$  satisfying

$$(L^\dagger \mu_{\text{inv}}) = 0 \quad (20)$$

is called the *invariant measure* when it exists. Furthermore, one says that the process  $X_t$  is an *equilibrium process* (with respect to  $\mu_{\text{inv}}$ ) if its transition kernel satisfies the *detailed balance condition*,

$$\mu_{\text{inv}}(dx)P_t(x, dy) = \mu_{\text{inv}}(dy)P_t(y, dx) \quad (21)$$

for all  $(x, y) \in \mathcal{E}^2$ . In the case where  $\mu_{\text{inv}}$  has the density  $\rho_{\text{inv}}(x) \equiv \mu_{\text{inv}}(dx)/dx$  with respect to the Lebesgue or counting measure, this condition can be expressed as the following operator identity for the generator:

$$\rho_{\text{inv}}L\rho_{\text{inv}}^{-1} = L^\dagger, \quad (22)$$

which is equivalent to saying that  $L$  is self-adjoint with respect to  $\mu_{\text{inv}}$ . If the process  $X_t$  does not satisfy this condition, then it is referred to in physics as a *nonequilibrium* Markov process. Here we follow this terminology and consider both equilibrium and nonequilibrium processes.

## B. Pure jump processes and diffusions

Two important types of Markov processes will be used in the paper to illustrate our results, namely, pure jump processes and diffusions. In continuous time and continuous space, all Markov processes consist of a superposition of these two processes, combined possibly with deterministic motion [3, 85, 86]. The case of discrete-time Markov chains is discussed in Appendix E.

A homogeneous Markov process  $X_t$  is a pure jump process if the probability that  $X_t$  undergoes one jump during the time interval  $[t, t + dt]$  is proportional to  $dt$ , while the probability of undergoing more than one jump is infinitesimal with  $dt$ .<sup>4</sup> To describe these jumps, it is usual to introduce the bounded intensity or *escape rate* function  $\lambda(x)$ , such that  $\lambda(x)dt + o(dt)$  is the probability that  $X_t$  undergoes a jump during  $[t, t + dt]$  starting from the state  $X_t = x$ . When a jump occurs,  $X(t + dt)$  is then distributed with the kernel  $T(x, dy)$ , so that the overall transition rate is

$$W(x, dy) \equiv \lambda(x) T(x, dy) \quad (23)$$

for  $(x, y) \in \mathcal{E}^2$ . Over a time interval  $[0, T]$ , the path of such a process can thus be represented by the sequence of visited states in  $\mathcal{E}$ , together with the sequence of waiting times in those states, so that the space of paths is  $[\mathcal{E} \times (0, \infty)]^{\mathbb{N}}$ .

Under some regularity conditions (see [39, 83]), one can show that this process possesses a generator, given by

$$(Lf)(x) = \int_{\mathcal{E}} W(x, dy)[f(y) - f(x)] \quad (24)$$

for all bounded, measurable function  $f$  defined on  $\mathcal{E}$  and all  $x \in \mathcal{E}$ . In terms of transition rates, the condition of detailed balance with respect to some invariant measure  $\mu_{\text{inv}}$  is expressed as

$$\mu_{\text{inv}}(dx)W(x, dy) = \mu_{\text{inv}}(dy)W(y, dx) \quad (25)$$

for all  $(x, y) \in \mathcal{E}^2$ .

Pure diffusions driven by Gaussian white noise have, contrary to jump processes, continuous sample paths and are best described not in terms of transition rates, but in terms of stochastic differential equations (SDEs). For  $\mathcal{E} = \mathbb{R}^d$ , these have the general form

$$dX_t = F(X_t)dt + \sum_{\alpha} \sigma_{\alpha}(X_t) \circ dW_{\alpha}(t), \quad (26)$$

<sup>4</sup> In a countable space, one can show that all Markov processes with right continuous paths are of this type, a property which is not true in a general space [85, 86].

where  $F$  and  $\sigma_\alpha$  are smooth vector fields on  $\mathbb{R}^d$ , called respectively the *drift* and *diffusion coefficient*, and  $W_\alpha$  are independent Wiener processes (in arbitrary number, so that the range of  $\alpha$  is left unspecified). The symbol  $\circ$  denotes the Stratonovich (midpoint) convention used for interpreting the SDE; the Itô convention can also be used with the appropriate changes.

In the Stratonovich convention, the explicit form of the generator is

$$L = F \cdot \nabla + \frac{1}{2} \sum_{\alpha} (\sigma_{\alpha} \cdot \nabla)^2 = \hat{F} \cdot \nabla + \frac{1}{2} \nabla D \nabla, \quad (27)$$

where

$$\hat{F}(x) = F(x) - \frac{1}{2} \sum_{\alpha} (\nabla \cdot \sigma_{\alpha})(x) \sigma_{\alpha}(x) \quad (28)$$

is the so-called modified drift and

$$D^{ij}(x) = \sum_{\alpha} \sigma_{\alpha}^i(x) \sigma_{\alpha}^j(x) \quad (29)$$

is the covariance matrix involving the components of  $\sigma_{\alpha}$ . The notation  $\nabla D \nabla$  in (27) is a shorthand for the operator

$$\nabla D \nabla = \sum_{i,j} \frac{\partial}{\partial x_i} D^{ij}(x) \frac{\partial}{\partial x_j}, \quad (30)$$

which is also sometimes expressed as  $\nabla \cdot (D \nabla)$  or in terms of a matrix trace as  $\text{tr } D \nabla^2$ . With these notations, the condition of detailed balance for an invariant measure  $\mu_{\text{inv}}(dx)$ , with density  $\rho_{\text{inv}}(x)$  with respect to the Lebesgue measure,<sup>5</sup> is equivalent to

$$\hat{F} = \frac{D}{2} \nabla \ln \rho_{\text{inv}}. \quad (31)$$

Similar results can be obtained for the Itô interpretation. Obviously, the need to distinguish the two interpretations arises only if the diffusion fields  $\sigma_{\alpha}$  depend on  $x \in \mathcal{E}$ . If these fields are constant, then the Stratonovich and Itô interpretations yield the same results with  $\hat{F} = F$  and  $\nabla D \nabla = D \nabla^2$ .

### C. Conditioning observables

Having defined the class of stochastic processes of interest, we now define the class of events  $\mathcal{A}_T$  used to condition these processes. The idea is to consider a random variable or *observable*  $A_T$ , taken to be a real function of the paths of  $X_t$  over the time interval  $[0, T]$ , and to condition  $X_t$  on a general measurable event of the form  $\mathcal{A}_T = \{A_T \in B\}$  with  $B \subset \mathbb{R}$ . This means, more precisely, that we condition  $X_t$  on the subset

$$\mathcal{A}_T = \{\omega \in D([0, T], \mathcal{E}) : A_T(\omega) \in B\} \quad (32)$$

of sample paths satisfying the constraint that  $A_T \in B$ . In the following, we will consider the smallest event possible, that is, the ‘atomic’ event  $\{A_T = a\}$ , representing the set of paths for which  $A_T$  is contained in the infinitesimal interval  $[a, a + da]$  or, more formally, the set of paths such that  $A_T(\omega) = a$ . General conditionings of the form  $\{A_T \in B\}$  can be treated by integration over  $a$ . We

<sup>5</sup> This density exists, for example, when the conditions of Hormander’s Theorem are satisfied [87, 88].

then write  $X_t|A_T = a$  to mean that the process  $X_t$  is conditioned on the basic event  $\{A_T = a\}$ . Formally, we can also study this conditioning by considering path probability densities instead of paths measures, as done in [72].

Mathematically, the observable  $A_T$  is assumed to be non-anticipating, in the sense that it is adapted to the natural ( $\sigma$ -algebra) filtration  $\mathcal{F}_T = \sigma\{X_t(\omega) : 0 \leq t \leq T\}$  of the process up to time  $T$ . Physically, we also demand that  $A_T$  depend only on  $X_t$  and its transitions or displacements. For a pure jump process, this means that we consider a general observable of the form

$$A_T = \frac{1}{T} \int_0^T f(X_t) dt + \frac{1}{T} \sum_{0 \leq t \leq T: \Delta X_t \neq 0} g(X_{t-}, X_{t+}), \quad (33)$$

where  $f : \mathcal{E} \rightarrow \mathbb{R}$ ,  $g : \mathcal{E}^2 \rightarrow \mathbb{R}$ , and  $X_{t-}$  and  $X_{t+}$  denote, respectively, the state of  $X_t$  before and after a jump at time  $t$ . The discrete sum over the jumps of the process is well defined, since we suppose that  $X_t$  has a finite number of jumps in  $[0, T]$  with probability one.

The class of observables  $A_T$  defined by  $f$  and  $g$  includes many random variables of mathematical interest, such as the number of jumps over  $[0, T]$ , obtained with  $f = 0$  and  $g = 1$ , or the occupation time in some set  $\Delta$ , obtained with  $f(x) = \mathbb{1}_\Delta(x)$  and  $g = 0$ , with  $\mathbb{1}_\Delta$  the characteristic function of the set  $\Delta$ . From a physical point of view, it also includes many interesting quantities, including the fluctuating entropy production [89], particle and energy currents [78], the so-called activity [65, 66, 90, 91], which is essentially the number of jumps, in addition to work- and heat-related quantities defined for systems in contact with heat reservoirs and driven by external forces [92, 93].

For a pure diffusion process  $X_t \in \mathbb{R}^d$ , the appropriate generalization of the observable above is

$$A_T = \frac{1}{T} \int_0^T f(X_t) dt + \frac{1}{T} \int_0^T \sum_{i=1}^d g^i(X_t) \circ dX_t^i, \quad (34)$$

where  $f : \mathcal{E} \rightarrow \mathbb{R}$ ,  $g : \mathcal{E} \rightarrow \mathbb{R}^d$ ,  $\circ$  denotes as before the Stratonovich product, and  $g^i$  and  $X_t^i$  are the components of  $g$  and  $X_t$ , respectively. This class of ‘diffusive’ observables defined by the function  $f$  and the vector field  $g$  also includes many random variables of mathematical and physical interest, including occupation times, empirical distributions, empirical currents or flows, the fluctuating entropy production [89], and work and heat quantities [94]. For example, the empirical density of  $X_t$ , which represents the fraction of time spent at  $x$ , is obtained formally by choosing  $f(y) = \delta(y - x)$  and  $g = 0$ , while the empirical current, recently considered in the physics literature [95, 96], is defined, also formally, with  $f = 0$  and  $g(y) = \delta(y - x)$ .

The consideration of diffusions and current-type observables of the form (34) involving a stochastic integral is one of the main contributions of this paper, generalizing previous results obtained by Jack and Sollich [64] for jump processes, Garrahan and Lesanovsky [67] for dissipative quantum systems, and by Borkar *et al.* [26, 97] for Markov chains.

#### D. Large deviation principle

As mentioned in the introduction, the conditioning event  $\mathcal{A}_T$  must have the property of being atypical with respect to the measure of  $X_t$ , otherwise the conditioning should have no effect on this process in the asymptotic limit  $T \rightarrow \infty$ . Here we assume that  $\{A_T = a\}$  is *exponentially* rare with  $T$  with respect to the measure  $\mathbb{P}_{L, \mu_0, T}$  of  $X_t$ , which means that we define this rare event as a *large deviation event*. This exponential decay of probabilities applies to many systems and observables of physical and mathematical interest, and is defined in a precise way as follows. The random variable  $A_T$  is said to satisfy a *large deviation principle* (LDP) with respect to  $\mathbb{P}_{L, \mu_0, T}$  if there exists a lower

semi-continuous function  $I$  such that

$$\limsup_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{L, \mu_0, T} \{A_T \in C\} \geq \inf_{a \in C} I(a) \quad (35)$$

for any closed sets  $C$  and

$$\liminf_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{L, \mu_0, T} \{A_T \in O\} \leq \inf_{a \in O} I(a) \quad (36)$$

for any open sets  $O$  [21, 98, 99]. The function  $I$  is called the *rate function*.

The basic assumption of our work is that the function  $I$  exists and is different from 0 or  $\infty$ . If the process  $X_t$  is ergodic, then an LDP for the class of observables  $A_T$  defined above holds, at least formally, as these observables can be obtained by contraction from the so-called level 2.5 of large deviations concerned with the empirical density and empirical current. This level has been studied formally in [90, 95, 96], and rigorously for jump processes with finite space in [100] and countable space in [101]. The observable  $A_T$  can also satisfy an LDP if the process  $X_t$  is not ergodic; in this case, however, the existence of the LDP must be proved on a process by process basis and may depend on the initial condition of the process considered.

Formally, the existence of the LDP is equivalent to assuming that

$$\lim_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{L, \mu_0, T} \{A_T \in [a, a + da]\} = I(a), \quad (37)$$

so that the measure  $\mathbb{P}_{L, \mu_0, T} \{A_T \in [a, a + da]\}$  decays exponentially with  $T$ , as mentioned. The fact that this decay is in general not exactly, but only approximately exponential is often expressed by writing

$$\mathbb{P}_{L, \mu_0, T} \{A_T \in [a, a + da]\} \asymp e^{-TI(a)} da, \quad (38)$$

where the approximation  $\asymp$  is defined according to the large deviation limit (37) [50, 99]. We will see in the next subsection that this exponential approximation, referred to in information theory as the *logarithmic* equivalence [19], sets a natural scale for defining two processes as being equivalent in the stationary limit  $T \rightarrow \infty$ .

### E. Nonequilibrium path ensembles

We now have the all the notations needed to define our problem of large deviation conditioning. At the level of path measures, the conditioned process  $X_t|A_T = a$  is defined by the path measure

$$d\mathbb{P}_{a, \mu_0, T}^{\text{micro}}(\omega) \equiv d\mathbb{P}_{L, \mu_0, T} \{\omega | A_T = a\}, \quad (39)$$

which is a pathwise conditioning of the reference measure  $\mathbb{P}_{L, \mu_0, T}$  of  $X_t$  on the value  $A_T = a$  after the time  $T$ . By Bayes's Theorem, this is equal to

$$d\mathbb{P}_{a, \mu_0, T}^{\text{micro}}(\omega) = \frac{d\mathbb{P}_{L, \mu_0, T}(d\omega)}{\mathbb{P}_{L, \mu_0, T} \{A_T = a\}} \mathbb{1}_{[a, a+da]}(A_T(\omega)), \quad (40)$$

where  $\mathbb{1}_\Delta(x)$  is, as before, the indicator (or characteristic) function of the set  $\Delta$ . We refer to this measure as the *path microcanonical ensemble* (superscript micro) [55–57] because it is effectively a path generalization of the microcanonical ensemble of equilibrium statistical mechanics, in which the microscopic configurations of a system are conditioned or constrained to have a certain energy value. This energy is here replaced by the general observable  $A_T$ .

Our goal for the rest of the paper is to show that the microcanonical measure can be expressed or realized in the limit  $T \rightarrow \infty$  by a conservative Markov process, called the *driven process*. This process will be constructed, as mentioned in the introduction, indirectly via another path measure, known as the *exponential tilting* of  $d\mathbb{P}_{L,\mu_0,T}(\omega)$ :

$$d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(\omega) \equiv \frac{e^{TkA_T(\omega)} d\mathbb{P}_{L,\mu_0,T}(\omega)}{\mathbb{E}_{\mu_0}[e^{kTA_T}]}, \quad (41)$$

where  $k \in \mathbb{R}$ . In mathematics, this measure is also referred to as a *penalization* or a *Feynman-Kac transform* of  $\mathbb{P}_{L,\mu_0,T}$  [102], in addition to the names ‘exponential family’ and ‘Esscher transform’ mentioned in the introduction. In physics, it is referred, as also mentioned, to as the biased, twisted, or *s*-ensemble, the latter name arising again because the letter *s* is often used in place of *k* [62–66]. We use the name ‘canonical ensemble’ (superscript cano) because this measure is a path generalization of the well-known canonical ensemble of equilibrium statistical. From this analogy, we can interpret *k* as the analog of a (negative) inverse temperature and the normalization factor  $\mathbb{E}_{\mu_0}[e^{kTA_T}]$  as the analog of the partition function.

The plan for deriving the driven process is to define a process  $Y_t$  via a generalization of Doob’s transform and to show that its path measure is equivalent in the asymptotic limit to the path canonical ensemble. Following this result, we will then use established results about ensemble equivalence to show that the canonical path ensemble is equivalent to the microcanonical path ensemble, so as to finally obtain the result announced in (1). The notion of measure or process equivalence underlying these results, denoted by  $\cong$  in (1), is defined next.

## F. Process equivalence

Let  $\mathbb{P}_T$  and  $\mathbb{Q}_T$  be two path measures associated with a Markov process over the time interval  $[0, T]$ , such that  $\mathbb{P}_T$  is absolutely continuous with respect to  $\mathbb{Q}_T$ , so that the Radon-Nikodym derivative  $d\mathbb{P}_T/d\mathbb{Q}_T$  exists. We say that  $\mathbb{P}_T$  and  $\mathbb{Q}_T$  are *asymptotically equivalent* if

$$\lim_{T \rightarrow \infty} \frac{1}{T} \ln \frac{d\mathbb{P}_T}{d\mathbb{Q}_T}(\omega) = 0 \quad (42)$$

almost everywhere with respect to both  $\mathbb{P}_T$  and  $\mathbb{Q}_T$ . In this case, we also say that the Markov process  $X_t$  defined by  $\mathbb{P}_T$  and the different Markov process  $Y_t$  defined by  $\mathbb{Q}_T$  are asymptotically equivalent, and denote this property by  $X_t \cong Y_t$  as in (1).

This notion of process equivalence is borrowed from equilibrium statistical mechanics [103–105] and can be interpreted in two ways. Mathematically, it implies that  $\mathbb{P}_T$  and  $\mathbb{Q}_T$  are logarithmically equivalent for most paths, that is,

$$d\mathbb{P}_T(\omega) \asymp d\mathbb{Q}_T(\omega) \quad (43)$$

for almost all  $\omega$  with respect to  $\mathbb{P}_T$  or  $\mathbb{Q}_T$ . This is a generalization of the so-called *asymptotic equipartition property* of information theory [19], which states that the probability of sequences generated by an ergodic discrete source is approximately (i.e., logarithmically) constant for almost all sequences [19]. Here we have that, although  $\mathbb{P}_T$  and  $\mathbb{Q}_T$  may be different measures, they are approximately equal in the limit  $T \rightarrow \infty$  for almost all path with respect to these measures.

In a more concrete way, the asymptotic equivalence of  $\mathbb{P}_T$  and  $\mathbb{Q}_T$  also implies that an observable satisfying LDPs with respect to these measures concentrate on the same values for both measures in the limit  $T \rightarrow \infty$ . In other words, the two measures lead to the same typical or ergodic states of (dynamic) observables in the long-time limit. A more precise statement of this result based

on the LDP will be given when we come to proving explicitly the equivalence of the driven and conditioned processes. For now, the only important point to keep in mind is that the typical properties of two processes  $X_t$  and  $Y_t$  such that  $X_t \cong Y_t$  are essentially the same. This is a useful notion of equivalence when considering nonequilibrium systems, which is a direct generalization of the notion of equivalence used for equilibrium systems. For the latter systems, typical values of (static) observables are simply called *equilibrium states*.

### III. NON-CONSERVATIVE TILTED PROCESS

We discuss in this section the properties of a non-conservative process associated with the canonical path measure (41) and, more precisely, with the numerator of that measure. This process is important as it allows us to obtain a number of important quantities related to the large deviations of  $A_T$ , in addition to give some clues as to how the driven process will be constructed.

#### A. Definition

We consider as before a Markov process  $X_t$  with path measure  $\mathbb{P}_{L,\mu_0,T}$  and an observable  $A_T$  defined as in (33) or (34) according to the type (jump process or diffusion, respectively) of  $X_t$ . From the path measure of  $X_t$ , we define a new path measure by

$$d\mathbb{P}_{\mathcal{L}_k,\mu_0,T}(\omega) \equiv d\mathbb{P}_{L,\mu_0,T}(\omega) e^{kTA_T(\omega)}, \quad (44)$$

which corresponds to the numerator of the path canonical ensemble  $d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}$ , defined in (41). As suggested by the notation, the new measure  $d\mathbb{P}_{\mathcal{L}_k,\mu_0,T}$  defines a Markov process of generator  $\mathcal{L}_k$ , which we call the *non-conservative tilted process*. This process is Markovian in the sense that

$$\mathbb{E}_{\mu_0}[e^{kTA_T}C] = \int_{\mathcal{E}^{n+1}} C(x_0, \dots, x_n) \mu_0(dx_0) e^{t_1 \mathcal{L}_k(x_0, dx_1)} \dots e^{(T-t_{n-1}) \mathcal{L}_k(x_{n-1}, dx_n)}, \quad (45)$$

for any cylinder functional  $C$  (15), and is non-conservative because  $(\mathcal{L}_k 1) \neq 0$  in general.

The class of observables defined by (33) and (34) can be characterized in the context of this result as the largest class of random variables for which the Markov property above holds. The proof of this property cannot be given for arbitrary Markov processes, but is relatively straightforward when considering jump processes and diffusions. In each case, the proof of (45) and the form of the so-called *tilted generator*  $\mathcal{L}_k$  follow by applying Girsanov's Theorem and the Feynman-Kac formula, as shown in Appendix A 1 for jump processes and Appendix A 2 for diffusions. The result in the first case is

$$(\mathcal{L}_k h)(x) = \left( \int_{\mathcal{E}} W(x, dy) [e^{kg(x,y)} h(y) - h(x)] \right) + kf(x)h(x) \quad (46)$$

for all function  $h$  on  $\mathcal{E}$  and all  $x \in \mathcal{E}$ , where  $f$  and  $g$  are defined as in (33). This can be written more compactly as

$$\mathcal{L}_k = W e^{kg} - (W1) + kf, \quad (47)$$

where the first term is understood as the Hadamard (component-wise) product  $W(x, dy)e^{kg(x,y)}$  and  $kf$  is a diagonal operator  $k(x)f(x)\delta(x-y)$ . In the case of diffusions, we obtain instead

$$\mathcal{L}_k = \hat{F} \cdot (\nabla + kg) + \frac{1}{2}(\nabla + kg)D(\nabla + kg) + kf, \quad (48)$$

where  $f$  and  $g$  are the functions appearing in (34), while  $\hat{F}$  and  $D$  are defined as in (28) and (29), respectively. The double product involving  $D$  is defined as in (30).



## B. Spectral elements

The operator  $\mathcal{L}_k$  defined in (46) or in (48) is a Perron-Frobenius operator or, more precisely, a *Metzler operator* with negative ‘diagonal’ part [106]. The extension of the Perron-Frobenius Theorem to infinite-dimensional, compact operators is ruled by the Krein-Rutman Theorem [107]. For differential elliptic operators having the form (48), this theorem can be applied on compact and smooth domains with Dirichlet boundary conditions.

We denote by  $\Lambda_k$  the real dominant (or principal) eigenvalue of  $\mathcal{L}_k$  and by  $r_k$  its associated ‘right’ eigenfunction, defined by

$$\mathcal{L}_k r_k = \Lambda_k r_k. \quad (49)$$

We also denote by  $l_k$  its ‘left’ eigenfunction, defined by

$$\mathcal{L}_k^\dagger l_k = \Lambda_k l_k, \quad (50)$$

where  $\mathcal{L}_k^\dagger$  is the dual of  $\mathcal{L}_k$  with respect to the Lebesgue or counting measure. These eigenfunctions are defined, as usual, up to multiplicative constants, set here by imposing the following normalization conditions:

$$\int_{\mathcal{E}} l_k(x) dx = 1 \quad \text{and} \quad \int_{\mathcal{E}} l_k(x) r_k(x) dx = 1. \quad (51)$$

For the remaining, we also assume that the initial measure  $\mu_0$  of  $X_t$  is such that

$$\int_{\mathcal{E}} \mu_0(dx) r_k(x) < \infty, \quad (52)$$

and that there is a gap  $\Delta_k$  between the first two largest eigenvalues resulting from the Perron-Frobenius Theorem. Under these assumptions, the semi-group generated by  $\mathcal{L}_k$  admits the asymptotic expansion

$$e^{t\mathcal{L}_k}(x, y) = e^{t\Lambda_k} [r_k(x)l_k(y) + O(e^{-t\Delta_k})] \quad (53)$$

as  $t \rightarrow \infty$ . Applying this result to the Feynman-Kac formula

$$\mathbb{E}_{\mu_0}[e^{kTA_T} \delta(X_T - y)] = \int_{\mathcal{E}} \mu_0(dx_0) e^{T\mathcal{L}_k}(x_0, y), \quad (54)$$

obtained by integrating (45) with  $C = \delta(X_T - y)$ , yields

$$\mathbb{E}_{\mu_0}[e^{kTA_T} \delta(X_T - y)] = e^{T\Lambda_k} \int_{\mathcal{E}} \mu_0(dx_0) [r_k(x_0)l_k(y) + O(e^{-t\Delta_k})]. \quad (55)$$

From this relation, we then deduce the following representations of the spectral elements  $\Lambda_k$ ,  $r_k$ , and  $l_k$ ; a further representation for the product  $r_k l_k$  will be discussed in the next subsection.

- Dominant eigenvalue  $\Lambda_k$ :

$$\Lambda_k = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E}_{\mu_0}[e^{kTA_T}] \quad (56)$$

for all  $\mu_0$  such that (52) is satisfied.



- Right eigenfunction  $r_k$ :

$$r_k(x_0) = \lim_{T \rightarrow \infty} e^{-T\Lambda_k} \mathbb{E}_{x_0}[e^{kTA_T}] \quad (57)$$

for all initial condition  $x_0$ .

- Left eigenfunction  $l_k$ :

$$l_k(y) = \lim_{T \rightarrow \infty} \frac{\mathbb{E}_{\mu_0}[e^{kTA_T} \delta(X_T - y)]}{\mathbb{E}_{\mu_0}[e^{kTA_T}]} \quad (58)$$

for all  $\mu_0$  such that (52) is satisfied.

With these results, we can already build a path measure from  $d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}$ , which is asymptotically equivalent to the canonical path measure. Indeed, it is clear from (56) that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \ln \left( e^{-T\Lambda_k} \frac{d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}}{d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}} \right) = 0 \quad (59)$$

almost everywhere, so that

$$d\mathbb{P}_{k, \mu_0, T}^{\text{cano}} \asymp e^{-T\Lambda_k} d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}. \quad (60)$$

We will see in the next section how to integrate the constant term  $e^{-T\Lambda_k}$  into a Markovian measure so as to obtain a Markov process which is conservative and equivalent to the canonical ensemble. For now, we close this subsection with two remarks:

- The right-hand side of (56) is known in large deviation theory as the *scaled cumulant generating function* (SCGF) of  $A_T$ . The rate function  $I$  can be obtained from this function by using the Gärtner-Ellis Theorem [21, 98, 99], which states (in its simplest form) that, if  $\Lambda_k$  is differentiable, then  $A_T$  satisfies the LDP with rate function  $I$  given by the Legendre-Fenchel transform of  $\Lambda_k$ :

$$I(a) = \sup_k \{ka - \Lambda_k\}. \quad (61)$$

For pure jump processes on a finite space, the differentiability of  $\Lambda_k$  follows from the implicit function theorem and the fact that  $\Lambda_k$  is a simple zero of the characteristic polynomial. In general,  $A_T$  can also satisfy an LDP when  $\Lambda_k$  is nondifferentiable; in this case, the rate function  $I$  is either convex with affine parts or nonconvex; see Sec. 4.4 of [50] for more details.

- The cloning simulation methods [29–31] mentioned in the introduction can be interpreted as algorithms that generate the non-conservative process  $\mathcal{L}_k$  and obtain the SCGF  $\Lambda_k$  by estimating the rate of growth or decay of its (non-normalized) measure, identified as  $\mathbb{E}_{\mu_0}[e^{kTA_T}]$ . An alternative method for simulating large deviations is transition path sampling, which attempts to directly sample paths according to  $\mathbb{P}_{k, \mu_0, T}^{\text{cano}}$  [58–61].

### C. Marginal canonical density

Equation (58) can be reformulated in terms the canonical path measure as

$$l_k(y) = \lim_{T \rightarrow \infty} \int d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}(\omega) \delta(X_T(\omega) - y). \quad (62)$$

This gives a physical interpretation of the left eigenfunction as the limit, when  $T$  is large, of the marginal probability density function of the canonical ensemble at the final time  $t = T$ . If we calculate this marginal for a  $t \in [0, T[$  and let  $t \rightarrow \infty$  after taking  $T \rightarrow \infty$ , we obtain instead

$$l_k(y)r_k(y) = \lim_{t \rightarrow \infty} \lim_{T \rightarrow \infty} \int d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(\omega) \delta(X_t(\omega) - y). \quad (63)$$

The product  $r_k l_k$  is thus the large-time marginal probability density of the canonical process taken over the infinite time interval. We will see in Sec. V that the same product corresponds to the invariant density of the driven process.

To prove (63), take  $C = \delta(X_t - y)$  with  $t < T$  in (45), and integrate to obtain

$$\mathbb{E}_{\mu_0}[e^{kTA_T} \delta(X_t - y)] = \int_{\mathcal{E}} \mu_0(dx_0) e^{t\mathcal{L}_k}(x_0, y) (e^{(T-t)\mathcal{L}_k} 1)(y). \quad (64)$$

Now take the limit  $T \rightarrow \infty$  to obtain

$$\lim_{T \rightarrow \infty} e^{-T\Lambda_k} \mathbb{E}_{\mu_0}[e^{kTA_T} \delta(X_t - y)] = \int_{\mathcal{E}} \mu_0(dx_0) e^{t\mathcal{L}_k}(x_0, y) e^{-t\Lambda_k} r_k(y), \quad (65)$$

which can be rewritten with (55) as

$$\lim_{T \rightarrow \infty} \frac{\mathbb{E}_{\mu_0}[e^{kTA_T} \delta(X_t - y)]}{\mathbb{E}_{\mu_0}[e^{kTA_T}]} = \frac{\int_{\mathcal{E}} \mu_0(dx_0) e^{t\mathcal{L}_k}(x_0, y) e^{-t\Lambda_k} r_k(y)}{\int_{\mathcal{E}} \mu_0(dx_0) r_k(x_0)} \quad (66)$$

assuming (52). Finally, take the limit  $t \rightarrow \infty$  to obtain

$$\lim_{t \rightarrow \infty} \lim_{T \rightarrow \infty} \frac{\mathbb{E}_{\mu_0}[e^{kTA_T} \delta(X_t - y)]}{\mathbb{E}_{\mu_0}[e^{kTA_T}]} = l_k(y)r_k(y), \quad (67)$$

which can be rewritten with the canonical measure as (63). A similar proof applies to (62); see Appendix B of [108] for a related discussion of these results.

The result of (63) can actually be generalized in the following way: instead of taking  $t \in [0, T[$  and letting  $t \rightarrow \infty$  after  $T \rightarrow \infty$ , we can scale  $t$  with  $T$  by choosing  $t = c(T)$  such that

$$\lim_{T \rightarrow \infty} c(T) = \infty \quad \text{and} \quad \lim_{T \rightarrow \infty} T - c(T) = \infty. \quad (68)$$

In this case, it is easy to see from (65)-(67) that we obtain the same result, namely,

$$l_k(y)r_k(y) = \lim_{T \rightarrow \infty} \int d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(\omega) \delta(X_{c(T)}(\omega) - y). \quad (69)$$

In particular, we can take  $c(T) = (1 - \epsilon)T$  with  $0 < \epsilon < 1$  to get  $t$  as close as possible to  $T$ , without reaching  $T$ . This will be used later when considering the equivalence of the driven process with the canonical path measure.

Note that there is no contradiction between (62) and (63), since for  $t \leq T$ ,

$$\begin{aligned} \int d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(\omega) \delta(X_t(\omega) - y) &= \frac{\int_{\mathcal{E}} \mu_0(dx_0) e^{t\mathcal{L}_k}(x_0, y) (e^{(T-t)\mathcal{L}_k} 1)(y)}{\int_{\mathcal{E}} \mu_0(dx_0) (e^{T\mathcal{L}_k} 1)(x_0)} \\ &\neq \frac{\int_{\mathcal{E}} \mu_0(dx_0) e^{t\mathcal{L}_k}(x_0, y)}{\int_{\mathcal{E}} \mu_0(dx_0) (e^{t\mathcal{L}_k} 1)(x_0)} = \int d\mathbb{P}_{k,\mu_0,t}^{\text{cano}}(\omega) \delta(X_t(\omega) - y). \end{aligned} \quad (70)$$

The fact that the left-most and right-most terms are not equal arises because the canonical measure is defined globally (via  $A_T$ ) for the whole time interval  $[0, T]$ , so that the marginal of the canonical measure at time  $t$  depends on times after  $t$ , as well as the end-time  $T$ . We will study in more details the source of this property in Sec. V when proving that the canonical path measure is a non-homogeneous Markov process that explicitly depends on  $t$  and  $T$ .

#### IV. GENERALIZED DOOB TRANSFORM

We define in this section the generalized Doob transform that will be used in the next section to define the driven process. We also review the conditioning problem considered by Doob in order to understand whether the case of large deviation conditioning can be analyzed within Doob's approach. Two examples will be considered: first, the original problem of Doob involving the conditioning on leaving a domain via its boundary and, second, a 'punctual' conditioning at a deterministic time. In each case, we will see that the generator of the process realizing the conditioning is a particular case of Doob's transform, but that the random variable underlying the conditioning is, in general, different from the random variables  $A_T$  defined before.

##### A. Definition

Let  $h$  be a strictly positive function on  $\mathcal{E}$  and  $f$  an arbitrary function on the same space. We call the *generalized Doob transform* of the process  $X_t$  with generator  $L$  the new process with generator

$$L^{h,f} \equiv h^{-1}Lh - f. \quad (71)$$

In this expression,  $h^{-1}Lh$  must be understood as the composition of three operators: the multiplication operator by  $h^{-1}$ , the operator  $L$  itself, and the multiplication operator by  $h$ . Moreover, the term  $f$  represents the multiplication operator by  $f$ , so that the application of  $L^{h,f}$  on some function  $r$  yields

$$(L^{h,f}r)(x) = h^{-1}(x)(Lhr)(x) - f(x)r(x). \quad (72)$$

We prove in Appendix B that the generalized Doob transform of  $L$  is indeed the generator of a Markov process, whose path measure  $\mathbb{P}_{L^{h,f},\mu_0,T}$  is absolutely continuous with respect to the path measure  $\mathbb{P}_{L,\mu_0,T}$  of  $X_t$  and whose Radon-Nikodym derivative is explicitly given by

$$\frac{d\mathbb{P}_{L^{h,f},\mu_0,T}}{d\mathbb{P}_{L,\mu_0,T}}(\omega) = h^{-1}(X_0) \exp\left(-\int_0^T f(X_t) dt\right) h(X_T). \quad (73)$$

In the following, we will also use time-dependent functions  $h_t$  and  $f_t$  to transform  $L$  [109]. In this case, the generalized Doob transform is a non-homogeneous process with path measure given by

$$\frac{d\mathbb{P}_{L^{h,f},\mu_0,T}}{d\mathbb{P}_{L,\mu_0,T}}(\omega) = h_0^{-1}(X_0) \exp\left(-\int_0^T (f_t + h_t^{-1}\partial_t h_t)(X_t) dt\right) h_T(X_T). \quad (74)$$

It is important to note that the transformed process with generator  $L^{h,f}$  is Markovian but not necessarily conservative, which means that its dominant eigenvalue is not necessarily zero. If we require conservation (zero dominant eigenvalue), it is sufficient that we choose  $f = h^{-1}(Lh)$ , in which case (71) becomes

$$L^h = h^{-1}Lh - h^{-1}(Lh), \quad (75)$$

while (73) reduces to

$$\frac{d\mathbb{P}_{L^h, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}}(\omega) = h^{-1}(X_0) \exp \left( - \int_0^T h^{-1}(X_t) (Lh)(X_t) dt \right) h(X_T). \quad (76)$$

Moreover, in the time-dependent case, (74) becomes

$$\frac{d\mathbb{P}_{L^h, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}}(\omega) = h_0^{-1}(X_0) \exp \left( - \int_0^T dt (h_t^{-1}(Lh_t) + h_t^{-1} \partial_t h_t)(X_t) \right) h_T(X_T). \quad (77)$$

Specializing to specific processes, it is easy to see that the generalized Doob transform of a pure jump process with transition rates  $W(x, dy)$  is also a pure jump process with modified transition rates

$$W^h(x, dy) = h^{-1}(x) W(x, dy) h(y) \quad (78)$$

for all  $(x, y) \in \mathcal{E}^2$ . Similarly, it can be shown that the generator of the generalized Doob transform of a diffusion with generator  $L$  is

$$L^h = L + (\nabla \ln h) D \nabla, \quad (79)$$

where the product involving  $D$  is interpreted, as before, according to (30). The generalized Doob transformed process is thus a diffusion with the same noise as the original diffusion, but with a modified drift

$$F^h = F + D \nabla \ln h. \quad (80)$$

The proof of this result is given in Appendix C and follows by re-expressing the generalized Doob transform of (75) as

$$L^h = L + h^{-1} \Gamma(h, \cdot), \quad (81)$$

where  $\Gamma$  is the so-called ‘squared field’ operator,<sup>6</sup> which is a symmetric bilinear operator defined for all  $f$  and  $g$  on  $\mathcal{E}$  as

$$\Gamma(f, g) \equiv (Lfg) - f(Lg) - (Lf)g. \quad (82)$$

Mathematical properties and applications of the generalized Doob transform have been studied by Kunita [110], Itô and Watanabe [111], Fleming and collaborators (see [40] and references cited therein), and have been revisited recently by Palmowski and Rolski [112] and Diaconis and Miclo [113]. From the point of view of probability theory, the Radon-Nikodym derivative associated with this transform is an example of exponential martingale. The generalized Doob transform also has interesting applications in physics: it appears in the stochastic mechanics of Nelson [114] and underlies, as shown in [109], the classical fluctuation-dissipation relations of near-equilibrium systems [54, 115–117], and recent generalizations of these relations obtained for nonequilibrium systems [91, 118–124]. The work of [109] shows moreover that the exponential martingale (76) verifies a non-perturbative general version of these relations, which also include the so-called fluctuation relations of Jarzynski [125] and Gallavotti-Cohen [126–128].

---

<sup>6</sup> From the French ‘opérateur carré du champs’.

## B. Historical conditioning of Doob

The transform considered by Doob is a particular case of the generalized transform (71), obtained for the constant function  $f(x) \equiv \lambda$  and for a so-called  $\lambda$ -*excessive* function  $h$  verifying  $Lh \leq \lambda h$ . For these functions, the Doob transformed process is a non-conservative process of generator

$$L^{h,\lambda} = h^{-1}Lh - \lambda, \quad (83)$$

and path measure

$$\frac{d\mathbb{P}_{L^{h,\lambda},\mu_0,T}}{d\mathbb{P}_{L,\mu_0,T}}(\omega) = h^{-1}(X_0)e^{-T\lambda}h(X_T). \quad (84)$$

When  $(Lh) = \lambda$ ,  $h$  is said to be  $\lambda$ -invariant. If we also have  $\lambda = 0$ , then  $h$  is called a *harmonic* function [1–3], and the process described by  $L^h = h^{-1}Lh$  is conservative with path measure

$$\frac{d\mathbb{P}_{L^h,\mu_0,T}}{d\mathbb{P}_{L,\mu_0,T}}(\omega) = h^{-1}(X_0)h(X_T). \quad (85)$$

In the time-dependent case, the harmonic condition  $Lh = 0$  is replaced by

$$(\partial_t + L_t)h_t = 0, \quad (86)$$

which yields, following (74) and (77),

$$\frac{d\mathbb{P}_{L^h,\mu_0,T}}{d\mathbb{P}_{L,\mu_0,T}}(\omega) = h_0^{-1}(X_0)h_T(X_T). \quad (87)$$

In this case,  $h_t$  is said to be *space-time harmonic* [1–3]. Applications of these transforms have appeared since Doob's work in the context of various conditionings of Brownian motion, including the Gaussian and Schrödinger bridges mentioned in the introduction, in addition to non-colliding random walks related to Dyson's Brownian motion and random matrices [129–131].

The original problem considered by Doob, leading to  $L^h$ , is to condition a Markov process  $X_t$  started at  $X_0 = x_0$  to exit a certain domain  $D$  via a subset of its boundary  $\partial D$ . To be more precise, assume that the boundary of  $D$  can be decomposed as  $\partial D = B \cup C$  with  $B \cap C = \emptyset$ , and condition the process to exit  $D$  via  $B$ . In this case, the path measure of the conditioned process can be written as

$$d\mathbb{P}_{x_0,T}\{\omega|\mathcal{B}\} = d\mathbb{P}_{L,x_0,T}(\omega) \frac{\mathbb{P}_{L,x_0}\{\mathcal{B}|\mathcal{F}_T\}}{\mathbb{P}_{L,x_0}\{\mathcal{B}\}}, \quad (88)$$

where  $\mathbb{P}_{L,x_0,T}$  is the path measure of the process started at  $x_0$ ,  $\mathcal{B} = \{\tau_B \leq \tau_C\}$  is the conditioning event expressed in terms of the exit times,

$$\tau_B \equiv \inf\{t : X_t \in B\}, \quad \tau_C \equiv \inf\{t : X_t \in C\}, \quad (89)$$

and  $\mathcal{F}_T = \sigma\{X_t(\omega) : 0 \leq t \leq T\}$  is the natural filtration of the process up to the time  $T$ .

The conditional path measure (88) is similar to the microcanonical path measure (39) and can be put in the form

$$d\mathbb{P}_{x_0,T}\{\omega|\mathcal{B}\} = d\mathbb{P}_{L,x_0,T}(\omega) M_{[0,T]}, \quad (90)$$

where

$$M_{[0,T]} = \frac{\mathbb{P}_{L,x_0}\{\mathcal{B}|\mathcal{F}_T\}}{\mathbb{P}_{L,x_0}\{\mathcal{B}\}} \quad (91)$$

to emphasize that it is a ‘reweighing’ or ‘penalization’ [102] of the original measure of the process with the *weighting function*  $M_{[0,T]}$ . To show that this reweighing gives rise to a Doob transform for the exit problem, let

$$h(x) = P_x\{\tau_B \leq \tau_C\}. \quad (92)$$

This function is harmonic, since  $(Lh) = 0$  by Dynkin’s formula [3]. Moreover, using the strong Markov property, we can use this function to express the weighting function (91) as

$$M_{[0,T]} = \frac{h(X_{\min(T, \tau_B, \tau_C)})}{h(X_0)}. \quad (93)$$

For  $T \leq \min(\tau_B, \tau_C)$ , we therefore obtain

$$d\mathbb{P}_{x_0, T}\{\omega | \mathcal{B}\} = h^{-1}(x_0) d\mathbb{P}_{L, x, T}(\omega) h(X_T). \quad (94)$$

which has the form of (85). The next example provides a simple application of this result.

**Example 1.** Consider the Brownian or Wiener motion  $W_t$  conditioned on exiting the set  $A = \{0, \ell\}$  via  $B = \{\ell\}$ . The solution of  $(Lh) = 0$  with the boundary conditions  $h(0) = 0$  and  $h(\ell) = 1$  gives the harmonic function  $h(x) = x/\ell$ , which implies from (80) that the drift of the conditioned process is  $F^h(x) = 1/x$ . The conditioned process is thus the Bessel process:

$$dX_t = \frac{1}{X_t} dt + dW_t. \quad (95)$$

Note that the drift of the conditioned process is independent of  $\ell$ , which means by taking  $\ell \rightarrow \infty$  that the Bessel process is also the Wiener process conditioned never to return at the origin. This is expected physically, as  $F^h$  is a repulsive force at the origin which prevents the process from approaching this point.

As a variation of Doob’s problem, consider the conditioning event

$$\mathcal{B}_T = \{X_T \in B_T\}, \quad (96)$$

where  $B_T$  is a subset of  $\mathcal{E}$  that can depend on  $T$ . This event is a particular case of  $A_T$  obtained with  $f = 0$  and  $g = 1$ , so that  $A_T = X_T/T$  assuming  $X_0 = 0$ . Its associated weighting function takes the form

$$M_{[0, T']} = \frac{\mathbb{P}_{L, x_0}\{\mathcal{B}_T | \mathcal{F}_{T'}\}}{\mathbb{P}_{L, x_0}\{\mathcal{B}_T\}} = \frac{\mathbb{P}_{L, x_0}\{\mathcal{B}_T | X_{T'}\}}{\mathbb{P}_{L, x_0}\{\mathcal{B}_T\}}, \quad (97)$$

for  $T' \leq T$ . Defining the function

$$h_{T'}(X_{T'}) \equiv \mathbb{P}_{L, x_0}\{\mathcal{B}_T | X_{T'}\} = \int_{B_T} P_{T-T'}(X_{T'}, dy), \quad (98)$$

we then have

$$M_{[0, T']} = \frac{h_{T'}(X_{T'})}{h_0(X_0)}. \quad (99)$$

Moreover, from the backward Kolmogorov equation, we find that  $h$  is space-time harmonic, as in (86). Therefore, the path measure of  $X_t$  conditioned on  $\mathcal{B}_T$  also takes the form of a Doob transform,

$$d\mathbb{P}_{x_0, T'}\{\omega | \mathcal{B}_T\} = d\mathbb{P}_{L, x_0, T'}(\omega) h_0^{-1}(X_0) h_{T'}(X_{T'}), \quad (100)$$

but now involves a time-dependent space-time harmonic function.<sup>7</sup>

The next two examples apply this type of punctual conditioning to define bridge versions of the Wiener motion and the Ornstein-Uhlenbeck process.

**Example 2 (Brownian bridge).** Let  $W_t|W_T = 0$  be the Wiener motion  $W_t$  conditioned on reaching 0 at time  $T$ . The Kolmogorov equation, which is the classical diffusion equation with  $L = \Delta/2$ , yields the Gaussian transition density of  $W_t$  as the space-time harmonic function:

$$h_t(x) = e^{(T-t)L}(x, 0) = \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{x^2}{2(T-t)}\right), \quad 0 \leq t < T. \quad (101)$$

From (80) and (98), we then obtain, as expected, that  $W_t|W_T = 0$  is the Brownian bridge evolving according to

$$dX_t = -\frac{X_t}{T-t}dt + dW_t \quad (102)$$

for  $0 \leq t < T$ . The limit  $T \rightarrow \infty$  recovers the Wiener process itself as the conditioned process.

**Example 3 (Ornstein-Uhlenbeck bridge).** Consider now the Ornstein-Uhlenbeck process,

$$dX_t = -\gamma X_t dt + \sigma dW_t \quad (103)$$

with  $\gamma > 0$  and  $\sigma > 0$ , conditioned on the event  $X_T = Ta$ . Using the propagator of this process, given explicitly by

$$P_t(x, y) = \sqrt{\frac{\gamma}{\pi\sigma^2(1-e^{-2\gamma t})}} \exp\left(-\frac{\gamma}{\sigma^2} \frac{(y - e^{-\gamma t}x)^2}{1-e^{-2\gamma t}}\right), \quad (104)$$

we obtain from (98),

$$h_t(x) = e^{(T-t)L}(x, 0) = \sqrt{\frac{\gamma}{\pi\sigma^2(1-e^{-2\gamma(T-t)})}} \exp\left(-\frac{\gamma}{\sigma^2} \frac{(x - Ta e^{\gamma(T-t)})^2}{e^{2\gamma(T-t)} - 1}\right). \quad (105)$$

With (80), we then conclude that  $X_t|X_T = aT$  is the non-homogeneous diffusion

$$dX_t = -\gamma X_t dt + F_T(X_t, t)dt + \sigma dW_t, \quad 0 \leq t < T, \quad (106)$$

with added time-dependent drift

$$F_T(x, t) = -2\gamma \frac{x - Ta e^{\gamma(T-t)}}{e^{2\gamma(T-t)} - 1}. \quad (107)$$

The relation between this drift and the conditioning is interesting. Since

$$\lim_{t \rightarrow T} F_T(x, t) = \begin{cases} \infty & x < aT \\ \gamma aT & x = aT \\ -\infty & x > aT, \end{cases} \quad (108)$$

points away from the target  $x = aT$  are infinitely attracted towards this point as  $t \rightarrow T$ , which leads  $X_t$  to reach  $X_T = aT$ . This attraction, however, is all concentrated near the final time  $T$ , as shown in Fig. 3, so that the conditioning  $X_T = aT$  affects the Ornstein-Uhlenbeck process mostly

<sup>7</sup> Note that the probability of  $\mathcal{B}_T$  can vanish as  $T' \rightarrow \infty$ , for example, if  $X_t$  is transient. In this case, (98) vanishes as  $T' \rightarrow \infty$ , so that (100) becomes singular in this limit.

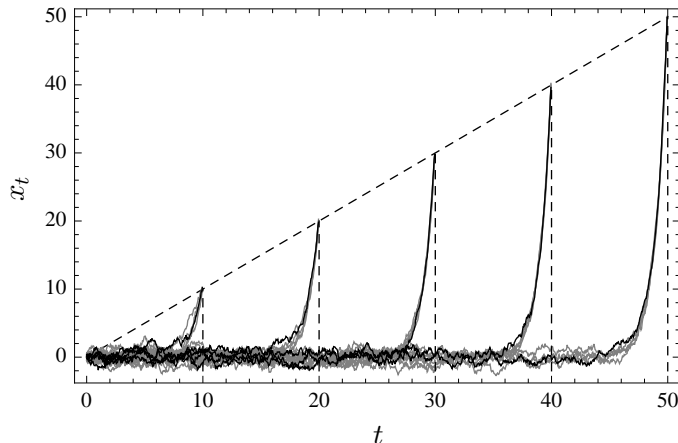


FIG. 1. Sample paths  $\{x_t\}_{t=0}^T$  of the Ornstein-Uhlenbeck process conditioned on the final point  $X_T = aT$  for  $T \in \{10, 20, 30, 40, 50\}$ . Parameters:  $\gamma = 1$ ,  $\sigma = 1$ ,  $a = 1$ . Black curves highlight one of five sample paths generated for each  $T$ . The conditioning mostly affects, as clearly seen, the dynamics only near the final time  $T$ , over a constant time-scale, inferred from (107), to be roughly given by  $1/\gamma$ .

at the boundary of the time interval  $[0, T]$  and marginally in the interior of this interval. In this case, taking the limit  $T \rightarrow \infty$  pushes the whole effect of the conditioning to infinity, so that care must be taken when interpreting this limit. It is clear here that we cannot conclude that, because  $F_\infty(x, t) = 0$  for  $t < \infty$ , the conditioned process is the Ornstein-Uhlenbeck process itself.

This boundary behavior of the conditioning will be discussed later in the paper. Interestingly, this behavior does not arise for the Wiener motion, obtained with  $\gamma = 0$  and  $\sigma = 1$ . In this case, the conditioned process is

$$dX_t = -\frac{X_t - Ta}{T - t}dt + dW_t \quad (109)$$

and converges to

$$dX_t = a dt + dW_t \quad (110)$$

in the limit  $T \rightarrow \infty$ . Thus, the conditioning  $X_T = aT$  is effected by an added drift  $a$ , which affects the dynamics of the process over the complete interval  $[0, T]$ .

We return at this point to our original problem of representing in terms of a conservative Markov process the microcanonical path measure  $\mathbb{P}_{a, \mu_0, T}^{\text{micro}}$  associated with the large deviation conditioning  $X_t|A_T = a$ . Following the preceding examples, the obvious question arises as to whether this measure can be obtained from a ‘normal’ Doob transform involving a suitably-chosen function  $h$ . The answer is, no, for essentially two reasons:

- Since  $A_T$  depends on the whole time interval  $[0, T]$  and not, as in the examples above, on a ‘punctual’ random time  $\tau \leq T$  or a deterministic time  $T$ , the weighting function associated with the large deviation conditioning (40) cannot be expressed exactly and as simply as in (93) or (99). What must be considered for this type of conditioning is an *approximate* and *asymptotic* form of equivalence, which essentially neglects the boundary terms  $h(X_0)$  and  $h(X_T)$ , as well as sub-exponential terms in  $T$ .
- There does not seem to be a way to prove the equivalence of the microcanonical path measure with a Markov measure starting directly from the definitions of the former measure, the



associated weighing function, and the conditioning observable  $A_T$ . Here, we prove this equivalence indirectly via the use of the canonical path measure.

These points are discussed in more details in the next section.

## V. DRIVEN MARKOV PROCESS

We now come to the main point of this paper, which is to define a Markov process via the generalized Doob transform and prove its asymptotic equivalence with the conditioned process  $X_t|A_T = a$ . This equivalence is obtained, as just mentioned, by first proving the asymptotic equivalence of the path measure of the driven process with the canonical path measure, and by then proving the equivalence of the latter measure with the microcanonical path measure using known results about ensemble equivalence. Following these results, we discuss interesting properties of the driven process related to its reversibility and constraints satisfied by its transition rates (in the case of jump processes) or drift (in the case of diffusion). Some of these properties were announced in [72]; here we provide their full proofs in addition to derive new results concerning the reversibility of the driven process. Our main contribution is to treat the equivalence of the canonical and microcanonical path ensembles explicitly and derive conditions for this equivalence to hold. In previous works, the conditioned process is often assumed to be equivalent with the driven process and, in some cases, wrongly interpreted as the canonical path ensemble.

### A. Definition

We define the *driven process*  $Y_t$  by applying the generalized Doob transform to the generator  $\mathcal{L}_k$  of the non-conservative process considered in Sec. III, using for  $h$  the right eigenfunction  $r_k$ , which is strictly positive on  $\mathcal{E}$  by Perron-Frobenius. We denote the resulting generator of  $Y_t$  by  $L_k$ , so that in the notation of the generalized Doob transform (75), we have

$$L_k \equiv \mathcal{L}_k^{r_k} = r_k^{-1} \mathcal{L}_k r_k - r_k^{-1} (\mathcal{L}_k r_k). \quad (111)$$

Although the tilted generator  $\mathcal{L}_k$  is not conservative,  $L_k$  is since  $(L_k 1) = 0$ . Moreover, we infer from (73) that the path measure of this new process is related to the path measure of the non-conservative process by

$$\frac{d\mathbb{P}_{L_k, \mu_0, T}}{d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}} = r_k^{-1}(X_0) e^{-T\Lambda_k} r_k(X_T), \quad (112)$$

which means, using (44), that it is related to the path measure of the original (conservative) process by

$$\frac{d\mathbb{P}_{L_k, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}} = \frac{d\mathbb{P}_{L_k, \mu_0, T}}{d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}} \frac{d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}} = r_k^{-1}(X_0) e^{-T\Lambda_k} e^{kTA_T} r_k(X_T). \quad (113)$$

The existence and form of  $L_k$  is the main result of this paper. Following the expressions of the tilted generator (46) and (48),  $L_k$  can also be re-expressed as

$$L_k = r_k^{-1} \mathcal{L}_k|_{f=0} r_k + kf - \Lambda_k \quad (114)$$

to make the dependence on  $f$  more explicit. We deduce from (46) and this result that the driven process associated with the a pure jump process remains a pure jump process described by the modified rates

$$W_k(x, dy) = r_k^{-1}(x) W(x, dy) e^{kg(x,y)} r_k(y) \quad (115)$$

for all  $(x, y) \in \mathcal{E}^2$ . For a pure diffusion  $X_t$  described by the SDE (26), the driven process  $Y_t$  is a diffusion with the same noise as  $X_t$ , but with the following modified drift:

$$F_k = F + D(kg + \nabla \ln r_k). \quad (116)$$

The proof of this result follows by explicitly calculating

$$h^{-1} \mathcal{L}_k h = \hat{F} \cdot (\nabla + kg + \nabla \ln h) + \frac{1}{2} (\nabla + kg + \nabla \ln h) D (\nabla + kg + \nabla \ln h) + kf \quad (117)$$

for  $h > 0$  on  $\mathcal{E}$ , so as to obtain

$$\mathcal{L}_k^h = \hat{F} \cdot \nabla + \frac{1}{2} \nabla D \nabla + (kg + \nabla \ln h) D \nabla = L + (kg + \nabla \ln h) D \nabla. \quad (118)$$

Applying this formula to  $h = r_k > 0$ , we obtain from (27) that  $L_k$  is the generator of a diffusion with the same diffusion fields  $\sigma_\alpha$  as  $X_t$ , but with the modified drift given in (116). Note that this result carries an implicit dependence (via  $r_k$ ) on the two functions  $f$  and  $g$  defining the observable  $A_T$ , in addition to the explicit dependence on  $g$ .

## B. Equivalence with the canonical path ensemble

The relations (41), (44) and (113) lead together to

$$\frac{d\mathbb{P}_{L_k, \mu_0, T}}{d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}} = \frac{d\mathbb{P}_{L_k, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}} \frac{d\mathbb{P}_{L, \mu_0, T}}{d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}} = r_k^{-1}(X_0) r_k(X_T) e^{-T\Lambda_k} \mathbb{E}_{\mu_0}[e^{kTA_T}]. \quad (119)$$

From the limit (56) associating the SCGF with  $\Lambda_k$ , we therefore obtain

$$\lim_{T \rightarrow \infty} \frac{1}{T} \ln \frac{d\mathbb{P}_{L_k, \mu_0, T}}{d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}}(\omega) = 0 \quad (120)$$

for all paths, which shows that the path measure of the driven process is asymptotically equivalent to the canonical path measure. This means, as explained before, that the two path measures are logarithmically equivalent,

$$d\mathbb{P}_{\mathcal{L}_k^T, \mu_0, T} \asymp d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}, \quad (121)$$

so that, although they are not equal, their differences are sub-exponential in  $T$  for almost all paths.

From this result, it is possible to show, with additional conditions, that the typical values of observables satisfying LDPs with respect to these measures are the same.<sup>8</sup> However, because of the specific form of the canonical path ensemble, we can actually prove a stronger form of equivalence between this path measure and that of the driven process, which implies not only that observables have the same typical values, but also the same large deviations.

This strong form of equivalence follows by noting that the canonical path ensemble represents a time-dependent Markov process. This is an important result, which does not seem to have been noticed before. The meaning of this is that, despite the global normalization factor  $\mathbb{E}_{\mu_0}[e^{kTA_T}]$ , the canonical measure defined in (41) is the path measure of a non-homogeneous Markov process characterized by a time-dependent generator, denoted by  $L_{k, T}^{\text{cano}}$ .<sup>9</sup> The derivation of this generator is presented in Appendix D; the result is

$$L_{k, T}^{\text{cano}} \equiv \mathcal{L}_k^{h_{t, T}} = h_{t, T}^{-1} \mathcal{L}_k h_{t, T} - h_{t, T}^{-1} (\mathcal{L}_k h_{t, T}) \quad (122)$$

<sup>8</sup> H. Touchette, in preparation, 2014.

<sup>9</sup> Time-dependent generators arise when considering probability kernels  $P_s^t$  that depend on the times  $s$  and  $t$  between two transitions, and not just the time difference  $t - s$ , as considered in (8).

for all  $t \in [0, T]$ , where

$$h_{t,T}(x) = (e^{(T-t)\mathcal{L}_k} 1)(x) \quad (123)$$

is space-time harmonic with respect to  $\mathcal{L}_k$  (see Appendix D). Thus we see that the canonical measure is the generalized Doob transform of  $\mathcal{L}_k$  obtained with a time-dependent function  $h_{t,T}$  involving  $\mathcal{L}_k$  itself. At the level of path measures, we then have

$$d\mathbb{P}_{k,\mu_0,T}^{\text{cano}} = d\mathbb{P}_{L_{k,\cdot,T}^{\text{cano}},\mu_0,T}, \quad (124)$$

a result which should be understood in the sense of (16), with  $L$  replaced by the time-dependent generator  $L_{k,t,T}^{\text{cano}}$  and the normal exponential replaced by a time-ordered exponential [109].

To relate this result to the driven process, note that  $(e^{(T-t)\mathcal{L}_k} 1)$  becomes proportional to  $r_k$  as  $T \rightarrow \infty$ , so that

$$\lim_{T \rightarrow \infty} L_{k,t,T}^{\text{cano}} = (\mathcal{L}_k)^{r_k} \equiv L_k. \quad (125)$$

Thus, although the process described by  $d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}$  over the time interval  $[0, T]$  is non-homogeneous for  $T < \infty$ , it becomes homogeneous inside this time interval as the final time  $T$  diverges. Moreover, it converges in this limit to the driven process itself, which is by definition a homogeneous process. This holds for all  $t \in [0, T[$  in the limit  $T \rightarrow \infty$ ; for the final time  $t = T$ , we obtain instead

$$\lim_{T \rightarrow \infty} L_{k,T,T}^{\text{cano}} = \mathcal{L}_k - (\mathcal{L}_k 1). \quad (126)$$

Consequently, the convergence of the canonical process towards the driven process applies only in  $[0, T[$ ; at the boundary of this time interval, the canonical process converges to a different homogeneous process with generator (126). This explains from the point of view of generators why we obtain two different limits for the marginal canonical density at  $t < T$  and  $t = T$ , as seen in Sec. III.

This difference between the ‘interior’ (or ‘bulk’) and ‘boundary’ regimes of a process is an important feature of our theory. In a sense, this theory can only characterize the ‘interior’ of a process (exponentially tilted or conditioned), since we push the boundary to infinity, so to speak, and consider large deviation events that arise entirely from the ‘interior’ regime. Given that the canonical and driven processes are the same in this ‘interior’ regime, the large deviations of  $A_T$  or any other observable satisfying an LDP must therefore also be the same for both processes.

To be more precise, consider an observable  $B_T$  and assume that this observable satisfies an LDP with respect to the canonical path measure with rate function

$$I_k(b) \equiv \lim_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{k,\mu_0,T}^{\text{cano}} \{B_T \in [b, b + db]\}. \quad (127)$$

Let us write this LDP as

$$I_k(b) = \lim_{T \rightarrow \infty} \lim_{\epsilon \rightarrow 0^+} -\frac{1}{T} \ln \mathbb{P}_{k,\mu_0,T}^{\text{cano}} \{B_{(1-\epsilon)T} \in [b, b + db]\}. \quad (128)$$

If we assume that the fluctuations of  $B_T$  arise from the combined effect of canonical fluctuations of  $X_t$  over the whole interval  $[0, T]$  and not just the end interval  $[(1-\epsilon)T, T]$ , we can invert the limits on  $T$  and  $\epsilon$  to obtain

$$\begin{aligned} I_k(b) &= \lim_{\epsilon \rightarrow 0^+} \lim_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{k,\mu_0,T}^{\text{cano}} \{B_{(1-\epsilon)T} \in db\} \\ &= \lim_{\epsilon \rightarrow 0^+} \lim_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{k,\mu_0,T}^{\text{cano}}|_{[0,(1-\epsilon)T]} \{B_{(1-\epsilon)T} \in db\}, \end{aligned} \quad (129)$$

where  $\mathbb{P}_{k,\mu_0,T}^{\text{cano}}|_{[0,(1-\epsilon)T]}$  represents the projection of  $\mathbb{P}_{k,\mu_0,T}^{\text{cano}}$  on  $[0, (1-\epsilon)T]$ , which is different in general from  $\mathbb{P}_{k,\mu_0,(1-\epsilon)T}^{\text{cano}}$ . We know from our discussion above that this projection converges in the limit  $T \rightarrow \infty$  to the path measure of the driven process. If we further assume that this convergence carries over to  $B_{(1-\epsilon)T}$ , we can then write

$$\begin{aligned} I_k(b) &= \lim_{\epsilon \rightarrow 0^+} \lim_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{L_k, \mu_0} \{B_{(1-\epsilon)T} \in db\} \\ &= \lim_{\epsilon \rightarrow 0^+} (1-\epsilon) \lim_{T \rightarrow \infty} -\frac{1}{(1-\epsilon)T} \ln \mathbb{P}_{L_k, \mu_0} \{B_{(1-\epsilon)T} \in db\} \\ &= \lim_{T \rightarrow \infty} -\frac{1}{T} \ln \mathbb{P}_{L_k, \mu_0} \{B_T \in db\}. \end{aligned} \quad (130)$$

Consequently, the LDP for  $B_T$  in the canonical path ensemble implies an LDP for this random variable with respect the driven process with the *same* rate function.

This reasoning is valid, as stressed above, if the large deviations of  $B_T$  and  $B_{(1-\epsilon)T}$  are the same in the canonical path ensemble, that is, if these large deviations arise from the ‘interior’ part of the measure and not from the boundary interval  $[(1-\epsilon)T, T]$ . In most cases of interest, this is verified, although there are pathological cases for which the large deviations actually arise at the boundary. The asymptotic limit of the Ornstein-Uhlenbeck process with  $X_T = aT$ , discussed in Sec. IV, is such a case, which we will come back to in Sec. VI.

### C. Equivalence with the microcanonical path ensemble

We now come back to the problem of characterizing  $X_t|A_T = a$  as a Markov process by showing that the canonical and microcanonical path measures are asymptotically equivalent. This second level of equivalence is weaker than the equivalence between the canonical path measure and the driven process, for the simple reason that the microcanonical and canonical path measures have different supports. Moreover, the fact that  $A_T$  does not fluctuate in the microcanonical path ensemble (by definition of the conditioning) but does, generally, in the canonical path ensemble shows that the large deviation properties of observables cannot be the same in general in the two ensembles. However – and this is the crucial observation for the problem of conditioning – they can have the same typical values of observables, under conditions related to the convexity of the rate function  $I(a)$  [105, 132]. Moreover, the same conditions imply that the microcanonical and canonical path measures are asymptotically equivalent in the logarithmic sense. We discuss each of these levels of equivalence next, beginning with the level based on typical values.

As before, we assume that the conditioning observable  $A_T$  satisfies the LDP with respect to the path measure  $\mathbb{P}_{L,\mu_0,T}$  of the reference process  $X_t$  with rate function  $I(a)$ . We then consider an observable  $B_T$  and assume that it satisfies an LDP with respect to the microcanonical path measure  $\mathbb{P}_{a,\mu_0,T}^{\text{micro}}$  with rate function  $J^a$ , as well as an LDP with respect to the canonical path measure  $\mathbb{P}_{k,\mu_0,T}^{\text{cano}}$  with rate function  $J_k$ . We denote the set of global minima of  $I^a$  by  $\mathcal{B}^a$  and the global minima of  $I_k$  by  $\mathcal{B}_k$ . Since rate functions vanish at their global minimizers [50, 99], we can also write

$$\mathcal{B}^a = \{b : J^a(b) = 0\}, \quad \mathcal{B}_k = \{b : J_k(b) = 0\}. \quad (131)$$

These zeros are called *concentration points* in large deviation theory [50], since they correspond to the values of  $B_T$  at which the microcanonical or canonical measure does not decay exponentially with  $T$ . If these sets are singleton sets, then their unique element correspond to a typical value of  $B_T$  in the sense of the ergodic theorem [50, 99]. For example, if  $\mathcal{B}^a = \{b^*\}$  for a given value  $a$  of  $A_T$ , then  $B_T \rightarrow b^*$  as  $T \rightarrow \infty$  with probability 1 with respect to the microcanonical path measure  $\mathbb{P}_{a,\mu_0,T}^{\text{micro}}$ . A similar result can obviously be stated for the canonical ensemble.

The equivalence problem in this context is to determine pairs  $(a, k)$  for which  $B_T$  has the same typical value or, more generally, for which  $\mathcal{B}^a = \mathcal{B}_k$ . Such pairs turn out to be determined by the convexity properties of  $I(a)$ . Denote by  $\partial I(a)$  the subdifferential of  $I$  at  $a$ . Except possibly at boundary points,  $I$  is convex at  $a$  if  $\partial I(a) \neq \emptyset$ , and is conversely nonconvex at  $a$  if  $\partial I(a) = \emptyset$  [133]. With these notations, we have [105, 132]:

- If  $I$  convex at  $a$ , then  $\mathcal{B}^a = \mathcal{B}_k$  for all  $k \in \partial I(a)$ .
- If  $I$  is nonconvex at  $a$ , then  $\mathcal{B}^a \cap \mathcal{B}_k = \emptyset$  for all  $k \in \mathbb{R}$ . Thus, in this case, there is no  $k \in \mathbb{R}$  such that  $\mathcal{B}^a = \mathcal{B}_k$ .

The proof of these results, found in [105], relies on the following general relationship between the rate functions  $J_k$  and  $J^a$ , which follows directly from the definitions of the microcanonical and canonical ensembles:

$$J_k(b) = \inf_a \{J^a(b) + I(a) + \Lambda_k - ka\}. \quad (132)$$

The idea of the proof is to relate the zeros of the two sides of (132), which define  $\mathcal{B}_k$  and  $\mathcal{B}^a$ , by noting that  $I(a) \geq ka - \Lambda_k$  with equality if and only if  $I(a)$  is convex; see [105, 132] for the details.

A remarkable property of the microcanonical and canonical measures is that the convexity of  $I(a)$  not only determines the equality of  $\mathcal{B}^a$  and  $\mathcal{B}_k$  for general observables, but also the logarithmic equivalence of these measures. This brings us to the second level of equivalence, expressed by the following results:

- If  $I$  is convex at  $a$ , then for all  $k \in \partial I(a)$ ,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \ln \frac{d\mathbb{P}_{a, \mu_0, T}^{\text{micro}}}{d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}}(\omega) = 0, \quad (133)$$

almost everywhere with respect to  $\mathbb{P}_{a, \mu_0, T}^{\text{micro}}$  and  $\mathbb{P}_{k, \mu_0, T}^{\text{cano}}$ .

- If  $I$  is nonconvex at  $a$ , then there is no  $k \in \mathbb{R}$  for which the limit above vanishes.

The proof of these results also follows from the definitions of the microcanonical and canonical measures; see [105].

Our problem of large deviation conditioning can now be solved by linking all the results that we have obtained. To recapitulate:

1. **Driven-canonical measure equivalence:** Assuming the existence of  $\Lambda_k$ ,  $l_k$ , and  $r_k$ , that the conditions (51) and (52) are satisfied, and that the spectrum of  $\mathcal{L}_k$  has a gap, we have that the driven process obtained from the generalized Doob transform (111) is such that

$$d\mathbb{P}_{L_k, \mu, T} \asymp d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}. \quad (134)$$

2. **Driven-canonical observable equivalence:** Any observable  $B_T$  satisfying an LDP with respect to the canonical path measure also satisfies an LDP with respect to the law of the driven process with the *same* rate function, provided that these LDPs are not related to boundary effects. In this case, the large deviations – and by consequence the concentration points – of  $B_T$  are the same for both the canonical and driven processes.
3. **Canonical-microcanonical measure equivalence:** If  $I(a)$  is convex, then

$$d\mathbb{P}_{k, \mu_0, T}^{\text{cano}}(\omega) \asymp d\mathbb{P}_{a, \mu_0, T}^{\text{micro}}(\omega) \quad (135)$$

for all  $k \in \partial I(a)$ , almost everywhere with respect to both measures.

4. **Canonical-microcanonical observable equivalence:**  $B_T$  has in general different rate functions in the canonical and microcanonical path ensembles; however, its concentration points are the same in both ensembles when  $I(a)$  is convex.<sup>10</sup>

We reach two conclusions from these results. The first, obtained by combining (134) and (135), is that if  $I$  is convex at the conditioning value  $a$ , then

$$d\mathbb{P}_{a,\mu_0,T}^{\text{micro}}(\omega) \asymp d\mathbb{P}_{L_k,\mu,T}(\omega) \quad (136)$$

almost everywhere with respect to both measures for all  $k \in \partial I(a)$ . At the level of processes, we therefore write

$$X_t|A_T = a \cong Y_t, \quad (137)$$

where  $Y_t$  is the driven process with generator  $L_k$  such that  $k \in \partial I(a)$ . The second conclusion, obtained from the points 3 and 4 above, is that  $X_t|A_T = a$  and  $Y_t$  have the same typical values of observables, provided that these observables concentrate in a large deviation sense in the long-time limit and that  $k \in \partial I(a)$ . It is in this sense that we say that the conditioned process  $X_t|A_T = a$  is *realized* or *represented* by the driven process  $Y_t$ : the two processes may (and will in general) have different fluctuation properties, but they have the same typical or concentration properties in the stationary limit when  $I(a)$  is convex. In a more physical but looser sense, we can picture them as describing the same long-time stochastic dynamics.

The next subsections discuss further properties of the driven process playing an important role for describing nonequilibrium systems. We list next several remarks that relate more specifically to its equivalence with the conditioned process:

- If  $A_T$  has a unique concentration point  $a^*$ , then it should be expected that

$$X_t|A_T = a^* \cong X_t, \quad (138)$$

since  $A_T \rightarrow a^*$  in the limit  $T \rightarrow \infty$ , so that this value is ‘naturally’ realized by  $X_t$ . This follows from our results by noting that  $0 \in \partial I(a^*)$ ,  $\Lambda_0 = 0$  and  $r_0 = 1$  up to a constant, so that  $L_{k=0} = L$  in general, and  $F_{k=0} = F$  for diffusions. Hence, conditioning on a typical value of the process does not modify it in the asymptotic limit.

- The conditioning  $A_T = a$  is realized by the driven process as a typical value of  $A_T$  in the stationary limit. That is,  $A_T \rightarrow a$  as  $T \rightarrow \infty$  with probability 1 with respect to the law of  $Y_t$ . This follows simply by taking  $B_T = A_T$ .
- The equivalence of  $\mathcal{B}^a$  and  $\mathcal{B}_k$  also implies, in the case where these sets are singleton sets, that bounded functions  $C(B_T)$  have the same expectation in the driven and conditioned processes as  $T \rightarrow \infty$ . In other words, equivalence of concentration points also implies, in the case of unique concentration points, equality of expectations.
- If  $I(a)$  is convex and differentiable, then  $\partial I(a) = \{I'(a)\}$ , so that the value  $k$  achieving equivalence is selected by the condition  $k = I'(a)$ . In the case where  $I(a)$  is strictly convex, we also have by Legendre duality that  $k$  is such that  $\Lambda'_k = a$  [50]. These results are large deviation analogs of the thermodynamic relations connecting, respectively, the temperature with the derivative of the entropy and the energy with the derivative of the free energy [50].

<sup>10</sup> Equilibrium systems also have, in general, different fluctuations in the microcanonical and canonical ensembles, but have the same equilibrium states when the two ensembles are equivalent.

- Since equivalence is for all  $k \in \partial I(a)$ , there is possibly more than one driven process realizing the typical states of a conditioned process. This interesting result should arise whenever  $I(a)$  has exposed (convex) corners at which  $\partial I(a)$  is not a singleton; see [134] for an example.
- Conversely to the above remark, there can be conditionings  $X_t|A_T = a$  that admit no driven process if  $I$  is nonconvex at  $a$ . We conjecture that such a case of nonequivalent processes arises whenever  $X_t$  is not ergodic and switch between ‘phases’ that cannot be represented by a single, homogeneous Markov process.
- The driven process is a priori not unique: since boundary terms in path measures are negligible at the level of the logarithmic equivalence, one could apply an extra generalized Doob transform to  $L_k$  by choosing a function  $h > 0$  such that  $(L_k h) = 0$ . From the definition (111) of  $L_k$ , this is equivalent to  $(\mathcal{L}_k r_k h) = \Lambda_k r_k h$ . Since  $\Lambda_k$  is non-degenerate,  $h$  must therefore be a multiplicative constant having no effect on the driven process.
- In the case of diffusions with constant noise power  $\sigma(x) = \sigma$ , the low-noise limit  $\sigma \rightarrow 0$  yields for the driven process a deterministic differential equation for the (unique) fluctuation path characterizing the conditioning  $A_T = a$ . This can be used to recover known results from the Freidlin-Wentzell theory of fluctuation paths and instantons for noise-perturbed SDEs [51].
- The conditions leading to the equivalence of  $X_t|A_T = a$  and  $Y_t$  prevents many processes from being treated within our theory. Examples include Lévy processes for which there are in general no LDPs ( $I(a) = 0$  everywhere or  $\Lambda_k = \infty$ ), processes for which the LDP for  $A_T$  may have a scaling or ‘speed’ in  $T$  different from  $T$ , as illustrated in the next section, in addition to processes with ‘condensation’ transitions for which either  $\Lambda_k \rightarrow \infty$ ,  $\mathcal{L}_k$  is gapless or the condition (52) is not satisfied; see [64, 135–137] for examples.

#### D. Invariant density

The driven process has an invariant density on  $\mathcal{E}$  corresponding to

$$\rho_k(x) = l_k(x)r_k(x), \quad (139)$$

which is normalized following (51). This is proved directly from the definition (111) of the generator of the driven process, whose dual is

$$L_k^\dagger = r_k \mathcal{L}_k^\dagger r_k^{-1} - r_k^{-1}(\mathcal{L}_k r_k), \quad (140)$$

so that

$$(L_k^\dagger l_k r_k) = r_k(\mathcal{L}_k^\dagger l_k) - l_k(\mathcal{L}_k r_k) = \Lambda_k l_k r_k - \Lambda_k l_k r_k = 0. \quad (141)$$

If the driven process is ergodic, this invariant density is also the (unique) stationary density, in the sense that

$$\rho_k(y) = \lim_{t \rightarrow \infty} \lim_{T \rightarrow \infty} \int_{\mathcal{E}} d\mathbb{P}_{L_k, \mu_0, T}(\omega) \delta(Y_t(\omega) - y) \quad (142)$$

and

$$\rho_k(y) = \lim_{T \rightarrow \infty} \int_{\mathcal{E}} d\mathbb{P}_{L_k, \mu_0, T}(\omega) \delta(Y_T(\omega) - y). \quad (143)$$



In this case, the stationary density is therefore the same independently of the time interval  $[0, T]$  considered, contrary to the canonical ensemble measure which gives two different results for the two limit above; see again (62) and (63).

Note that, if the original process  $X_t$  is ergodic with invariant density  $\rho_{\text{inv}}$ , then  $\rho_{k=0} = \rho_{\text{inv}}$  because  $l_0 = \rho_{\text{inv}}$  and  $r_k = 1$ . Moreover, if  $\mathcal{L}_k$  is self-dual (hermitian), then  $l_k = r_k \equiv \psi_k$  so that  $\rho_k = \psi_k^2$ , in a clear analogy with quantum mechanics.

### E. Reversibility properties

It is interesting physically to describe the class of conditioning observables  $A_T$  for which the driven process is either reversible (equilibrium) or non-reversible (nonequilibrium). We study this problem here by deriving a functional equation involving  $f$  and  $g$ , whose solution provides a necessary and sufficient condition for  $\rho_k$  to be a reversible stationary density. This equation is hard to solve in general; a simpler form is obtained by assuming that the reference Markov process  $X_t$  is reversible, which leads us to study the following question: Under what conditioning is the driven process  $Y_t$  reversible given that  $X_t$  is reversible?

To answer this question, we first consider pure jump processes. For all  $(x, y) \in \mathcal{E}^2$ , the relation (115) for the driven transition rates implies

$$\frac{W_k(x, y)}{W_k(y, x)} = \left( \frac{r_k(y)}{r_k(x)} \right)^2 \frac{W(x, y)}{W(y, x)} e^{k[g(x, y) - g(y, x)]}. \quad (144)$$

Therefore, the driven process is reversible with respect to its invariant density  $\rho_k$  if and only if the ratio above can be written as  $\rho_k(y)/\rho_k(x)$ , which yields, with the expression of  $\rho_k$  shown in (139), a non-trivial functional equation for  $f$  and  $g$ .

We can simplify this equation by assuming that the reference process is reversible, as in (25). The ratio (144) then becomes

$$\frac{W_k(x, y)}{W_k(y, x)} = \left( \frac{r_k(y)}{r_k(x)} \right)^2 \frac{\rho_{\text{inv}}(y)}{\rho_{\text{inv}}(x)} e^{k[g(x, y) - g(y, x)]}. \quad (145)$$

For the driven process to remain reversible, it is thus sufficient that there exists a ‘potential’ function  $h$  on  $\mathcal{E}$  such that

$$g(x, y) - g(y, x) = h(y) - h(x) \quad (146)$$

for all  $(x, y) \in \mathcal{E}^2$ . In this case, we also have, if the invariant density  $\rho_k$  of the driven process is unique, that  $\rho_k$  is proportional to  $r_k^2 \rho_{\text{inv}} e^{kh}$ . This condition on  $g$  is verified, in particular, if  $g$  is symmetric,  $g(x, y) = g(y, x)$ .<sup>11</sup> Accordingly, conditioning observables  $A_T$ , such as the activity, that depend on the jumps of  $X_t$ , but not on the ‘direction’ of these jumps do not modify the reversibility of  $X_t$ . The same is true if  $A_T$  does not depend on the jumps of the process, that is, if  $g = 0$  and the conditioning only involves of an integral of  $f(X_t)$  in time.

These results translate for diffusions as follows. The driven process is reversible with respect to the invariant density  $\rho_k$  if and only if its modified drift

$$\hat{F}_k = \hat{F} + D(kg + \nabla \ln r_k), \quad (147)$$

obtained from (80), satisfies

$$\hat{F}_k = \frac{D}{2} \nabla \ln \rho_k, \quad (148)$$

<sup>11</sup> This slightly corrects the claim made in [64] that the driven process is reversible ‘only if the bias is also time-reversal symmetric:  $g(x, y) = g(y, x)$ ’.



which is equivalent to

$$\hat{F} + D \left( kg + \frac{1}{2} \nabla \left( \ln \frac{r_k}{l_k} \right) \right) = 0. \quad (149)$$

This is a functional equation involving  $f$  and  $g$ , via  $r_k$  and  $l_k$ , which is also difficult to solve in general. We can simplify it, as before, by assuming that the reference process  $X_t$  is reversible with respect to  $\rho_{\text{inv}}$ , as in (31), in which case

$$\hat{F}_k = \frac{D}{2} \nabla \ln(\rho_{\text{inv}} r_k^2) + Dkg. \quad (150)$$

A particular solution of this equation is obtained if  $g$  is gradient,  $g = \nabla h/2$ . Then the driven diffusion is a reversible diffusion with respect to the invariant density  $\rho_k$ , which is moreover proportional to  $r_k^2 \rho_{\text{inv}} e^{kh}$ . In particular, if  $g = 0$  and  $D$  is constant, then the driven process is a reversible diffusion with drift given by

$$F_k = F + D \nabla \ln r_k. \quad (151)$$

We thus see for diffusions that conditioning observables  $A_T$  that do not depend on the transitions of  $X_t$  ( $g = 0$ ) or depend on these transitions but via a gradient perturbation  $g$  do not modify the reversibility of  $X_t$ .

## F. Identities and constraints

It was found in [138–140] that the driven process admits in many cases certain invariant quantities that constrain its transition rates. These constraints arise very generally and very simply from our results. From (115), we can write

$$\begin{aligned} W_k(x, dy)W(y, dx) &= r_k(x)^{-2} W_k(y, dx)W(x, dy) r_k(y)^2 e^{k[g(x,y)-g(y,x)]} \\ W_k(x, dy)W_k(y, dx) &= W(x, dy)W(y, dx) e^{k[g(x,y)+g(y,x)]}, \end{aligned} \quad (152)$$

which are the most general identities that can be obtained for the transition rates of the driven process. If  $g$  is a symmetric function, they reduce to

$$W_k(x, dy)W(y, dx) = r_k(x)^{-2} W_k(y, dx)W(x, dy) r_k(y)^2, \quad (153)$$

whereas if  $g$  is antisymmetric or  $g = 0$ , we find

$$W_k(x, dy)W_k(y, dx) = W(x, dy)W(y, dx) \quad (154)$$

for all  $(x, y) \in \mathcal{E}^2$ . The latter result is referred to in [138, 140] as the ‘product constraint’. An example of symmetric observable is the so-called activity, which is proportional to the number of jumps occurring in a jump process, whereas an example of antisymmetric observable is the current, which assigns opposite signs to a jump and its reversal.

These constraints on the transition rates also imply constraints on the escape rate  $\lambda = (W1)$  of a jump process. Integrating (115) with respect to  $y$ , keeping in mind that  $r_k$  is the right eigenfunction of  $\mathcal{L}_k$  associated with  $\Lambda_k$ , we obtain the following relation between the escape rates of the reference process and those of the driven process:

$$\lambda_k(x) = \lambda(x) - kf(x) + \Lambda_k. \quad (155)$$

In the case  $f = 0$ , this yields

$$\lambda_k(x) = \lambda(x) + \Lambda_k \quad (156)$$

for all  $x \in \mathcal{E}$ , which implies

$$\lambda_k(x) - \lambda_k(y) = \lambda(x) - \lambda(y) \quad (157)$$

for all  $x \in \mathcal{E}$ , a result referred to as the ‘exit rate constraint’ in [138, 140].

Diffusive analogs of these constraints can be derived from our results. In the case where the covariance matrix  $D$  is invertible, (116) implies by taking its exterior derivative that

$$d(D^{-1}F_k - D^{-1}F - kg) = 0, \quad (158)$$

where all vectors are interpreted as 1-forms. We conclude from this result that the 1-form associated with  $D^{-1}F_k - D^{-1}F - kg$  is closed. In two and three dimensions, this implies

$$\nabla \times (D^{-1}F_k) = \nabla \times (D^{-1}F) + k\nabla \times g \quad (159)$$

and thus

$$\nabla \times (D^{-1}F_k) = \nabla \times (D^{-1}F) \quad (160)$$

if  $g$  is gradient. For diffusions on the circle, (158) only implies that  $F_k - F - kg$  is the derivative of a periodic function.<sup>12</sup>

These results can be interpreted physically as circulation constraints, showing that the non-reversibility of the driven process, measured by the circulation of its drift, is directly related to the non-reversibility of the reference process and the non-gradient character of  $g$ . This connection with non-reversible dynamics can be emphasized slightly by rewriting (159) in terms of the stationary probability current, defined by

$$J_{\rho_{\text{inv}}} = \hat{F}\rho_{\text{inv}} - \frac{D}{2}\nabla\rho_{\text{inv}}, \quad (161)$$

or the so-called probability velocity

$$V_{\rho_{\text{inv}}} \equiv \frac{J_{\rho_{\text{inv}}}}{\rho_{\text{inv}}} = \hat{F} - \frac{D}{2}\nabla \ln \rho_{\text{inv}}. \quad (162)$$

Both vector fields are zero for reversible (equilibrium) diffusions. In terms of  $V$ , we then obtain

$$\nabla \times (D^{-1}V_{\rho_k}) = \nabla \times (D^{-1}V_{\rho_{\text{inv}}}) + k\nabla \times g, \quad (163)$$

where  $V_{\rho_k}$  is the probability velocity of the driven process with invariant density  $\rho_k$ . A similar result applies in higher dimensions by replacing the rotational with the exterior derivative.

---

<sup>12</sup> Periodic functions on the circle are not necessarily the derivative of periodic functions: consider, for example, the constant function.

## VI. APPLICATIONS

We study in this section three applications of our results for the Brownian motion, the Ornstein-Uhlenbeck process, and the problem of quasi-stationary distributions. The applications are simple: they are there only to illustrate the different steps needed to obtain the driven process and the effect of boundary dynamics. In the case of quasi-stationary distributions, we also want to show how our results recover known results obtained by a different approach.

For more complex applications involving many-particle dynamics, such as the totally asymmetric exclusion process and the zero-range process, see [64, 136, 141–143]. Applications for diffusions can be found in [139, 144], whereas applications for quantum systems can be found in [67–71]. One interesting aspect of many-particle dynamics is that current-type conditionings have the generic effect of producing long-range interactions between particles at the level of the stationary distribution of the driven process [55, 64, 141–143].

In future publications, we will discuss in more details some of the connections mentioned in the introduction, in particular those relating to conditional limit theorems and optimal control theory, in addition to tackle other applications of our results, including the case of diffusions conditioned on occupation measures,<sup>13</sup> which is relevant for studying metastable states and quasi-stationary distributions. We will also study the low-noise (Freidlin-Wentzell) large deviation limit [51], and develop numerical techniques for obtaining the spectral elements used to construct the driven process.

### A. Extensive Brownian bridge

We revisit Example 3 about the Wiener process conditioned on reaching the point  $X_T = aT$ . This observable is a particular case of  $A_T$  obtained with  $f = 0$ ,  $g = 1$ , and  $X_0 = 0$ .

From the Gaussian propagator of the Wiener process, we find

$$\mathbb{P}_x\{X_T/T \in [a, a + da]\} \asymp e^{-TI(a)} da, \quad (164)$$

with  $I(a) = a^2/2$ , as well as

$$\lim_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E}_x[e^{kX_T}] = \frac{k^2}{2}. \quad (165)$$

The latter result is equal to the dominant eigenvalue  $\Lambda_k$ , as can be verified from the expression of the tilted generator

$$\mathcal{L}_k = \frac{1}{2} \left( \frac{d}{dx} + k \right)^2 = \frac{1}{2} \frac{d^2}{dx^2} + k \frac{d}{dx} + \frac{k^2}{2}, \quad (166)$$

obtained from (48). In fact, in this case, we have  $r_k(x) = l_k(x) = 1$ . From (116), we thus find that the drift of the driven process, equivalent to the conditioned process, is  $F_k = k$ . To re-express this drift as a function of the conditioning  $X_T/T = a$ , we use  $I'(a) = a = k$  to obtain  $F_{k(a)} = a$ . This shows that the process equivalent to the Brownian motion conditioned with  $X_T = aT$  is the drifted Brownian motion  $W_T + at$ , as found previously in (110). Equivalence is for all  $a \in \mathbb{R}$ , since  $I(a)$  is convex. Moreover, since the typical value of  $X_T/T$  is 0, we have  $X_t|X_T = 0 \cong X_t$ ; that is, the Brownian bridge bridged at  $T \rightarrow \infty$  is asymptotically equivalent to the Wiener process.

There is a subtlety involved in this calculation, in that  $r_k = l_k = 1$  are not normalizable. To circumvent this problem, it seems possible to consider the problem on a compact domain of  $\mathcal{E} = \mathbb{R}$ ,

<sup>13</sup> F. Angeletti, H. Touchette, in preparation, 2014.

such as the interval  $[-\ell, \ell]$ , to obtain a gapped spectrum with normalizable eigenfunctions, and then take the limit  $\ell \rightarrow \infty$ . This is a common procedure used in physics – for example, in quantum mechanics to deal with the free particle.

### B. Ornstein-Uhlenbeck process

Consider the Ornstein-Uhlenbeck process, defined in (103), with the conditioning observable

$$A_T = \frac{1}{T} \int_0^T X_t dt, \quad (167)$$

which corresponds to the choice  $f(x) = x$  and  $g = 0$ . The spectral elements of this observable are easily found to be  $r_k(x) = e^{kx/\gamma}$  and  $\Lambda_k = \sigma^2 k^2 / (2\gamma^2)$ . From the expression of  $r_k$  and (116), we then find that the effective drift of the driven process is

$$F_k(x) = -\gamma x + \frac{\sigma^2 k}{\gamma}. \quad (168)$$

With the rate function

$$I(a) = \sup_k \{ka - \Lambda_k\} = \frac{\gamma^2 a^2}{2\sigma^2}, \quad (169)$$

we then find

$$F_{k(a)}(x) = -\gamma x + \frac{a}{\gamma}. \quad (170)$$

Hence, the conditioning only adds a constant drift to the process, which ensures that  $X_T/T \rightarrow a$  as  $T \rightarrow \infty$ . Naturally, since the typical value of  $A_T$  is 0 in the original Ornstein-Uhlenbeck process, conditioning on  $X_T/T = 0$  yields the same process with  $F_{k=0}(x) = -\gamma x$ .

If instead of choosing the linear observable (167), we choose

$$A_T = \frac{1}{T} \int_0^T X_t^2 dt, \quad (171)$$

the same steps can be followed to obtain

$$F_{k(a)}(x) = -\frac{\sigma^2}{2a} x. \quad (172)$$

In this case, the conditioning keeps the linear force of the Ornstein-Uhlenbeck process, but changes its friction coefficient to match the variance of the process with the value of  $A_T$ .

To close this example, let us revisit the conditioning  $A_T = X_T/T = a$ , studied in Example 3, which corresponds to the choice  $f = 0$  and  $g = 1$ . We know from our previous discussion of this example that the driven process cannot describe this conditioning because the latter does not affect the ‘interior’ dynamics of the process in the asymptotic limit  $T \rightarrow \infty$ . Let us see how this arises in our theory. From the exact form of the propagator (104), we find that

$$P\{X_T/T \in [a, a + da]\} \asymp e^{-T^2 \gamma a^2 / \sigma^2}, \quad (173)$$

so that  $I(a) = \infty$  if we take the large deviation limit with the scale  $T$ , as in (37). In this case, we can formally take  $\Lambda_k = 0$  and  $r_k(x) = e^{-kx}$  for the spectral elements, which is consistent with  $I(a) = \infty$ , to obtain  $F_{k(a)}(x) = -\gamma x$ . This, as we know from Example 3, is the correct interior dynamics produced by the conditioning, but it is not the complete dynamics that actually realizes the conditioning. The problem here is that the large deviation is a boundary effect in time – it can be seen, physically, as a temporal analog of a ‘condensation’ – which prevents us from exchanging the two limits in (129). Consequently, though we can formally define the driven process, it is not equivalent with the conditioned process.

### C. Quasi-stationary distributions

A classical problem in the theory of absorbing processes and quasi-stationary distributions is to condition a Markov chain never to escape some subset of its state space. We want to briefly show in this subsection that the solution of this problem, obtained classically by defining a new Markov chain restricted on the subset of interest [15], can be recovered from our results (summarized for Markov chains in Appendix E) by taking the limit  $k \rightarrow \infty$ .

To define the problem, let  $\{X_i\}_{i=0}^\infty$  be a Markov chain with homogeneous transition matrix  $M$ . For a subset  $\mathcal{E}_1$  of  $\mathcal{E}$ , we consider the conditioning event

$$\mathcal{B} = \{\tau_1 > N\}, \quad (174)$$

where  $\tau_1$  is the exit time from  $\mathcal{E}_1$  defined by

$$\tau_1 = \inf\{n : X_n \notin \mathcal{E}_1\}, \quad (175)$$

assuming  $X_0 \in \mathcal{E}_1$ . This means that we are conditioning the Markov chain on leaving  $\mathcal{E}_1$  (or on being ‘killed’ outside  $\mathcal{E}_1$ ) only after the time  $N$ .

Within our theory, this conditioning is effected by considering the observable

$$A_N = \frac{1}{N} \sum_{i=0}^{N-1} \mathbb{1}_{\mathcal{E}_1}(X_i), \quad (176)$$

or its symmetrized version

$$A'_N = \frac{1}{N} \sum_{i=0}^{N-1} \frac{\mathbb{1}_{\mathcal{E}_1}(X_i) + \mathbb{1}_{\mathcal{E}_1}(X_{i+1})}{2} \quad (177)$$

for which we have  $\mathbb{1}_{\mathcal{B}} = \mathbb{1}_{A_{N+1}=1} = \mathbb{1}_{A'_N=1}$ . The second observable leads to the tilted matrix

$$\mathcal{M}_k(x, y) \equiv M(x, y) \exp \left[ \frac{k}{2} (\mathbb{1}_{\mathcal{E}_1}(x) + \mathbb{1}_{\mathcal{E}_1}(y)) \right]. \quad (178)$$

Given the conditioning  $A'_N = 1$ , we must then choose  $k \in \partial I(1)$ , where  $I$  is the rate function of  $A'_N$ . The form of  $\partial I(1)$  depends in general on the Markov chain considered. However, it is always the case that  $\infty \in \partial I(1)$ , since  $I$  is defined on  $[0, 1]$ , so that the conditioning can be achieved in general by taking the limit  $k \rightarrow \infty$ .

To see that this limit recovers the correct result, define the matrix

$$M'(x, y) \equiv \lim_{k \rightarrow \infty} e^{-k} \mathcal{M}_k(x, y). \quad (179)$$

Then

$$M'(x, y) = M(x, y) \mathbb{1}_{\mathcal{E}_1}(x) \mathbb{1}_{\mathcal{E}_1}(y) = \begin{cases} M(x, y) & x, y \in \mathcal{E}_1 \\ 0 & \text{otherwise} \end{cases} \quad (180)$$

represents the restriction of the Markov chain on  $\mathcal{E}_1$ . Denoting by  $\lambda'$  the dominant eigenvalue of  $M'$  and by  $r'$  its associated right eigenvector, we infer

$$\lambda' = \lim_{k \rightarrow \infty} e^{-k} \Lambda_k, \quad r' = \lim_{k \rightarrow \infty} r_k, \quad (181)$$

where  $\Lambda_k$  and  $r_k$  are the corresponding elements of  $\mathcal{M}_k$ .

According to our theory, the effective Markov chain resulting from the asymptotic conditioning  $A'_N = a$  is given by the generalized Doob transform

$$M_k(x, y) = \frac{1}{\Lambda_k r_k(x)} \mathcal{M}_k(x, y) r_k(y). \quad (182)$$

Taking the limit, we then obtain

$$\lim_{k \rightarrow \infty} M_k = \frac{1}{\lambda' r'} M' r', \quad (183)$$

which is the known result characterizing a Markov chain conditioned on eternally staying in  $\mathcal{E}_1$  [15]. In this context, it can be proved that

$$\ln \lambda' = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathbb{P}_{x, M, N} \{ \tau_1 > N \}, \quad (184)$$

where  $X_0 = x \in \mathcal{E}_1$ , so that  $\lambda'$  represents the *survival rate* at which the chain stays in  $\mathcal{E}_1$ , while the left eigenvector

$$l'(y) = \lim_{N \rightarrow \infty} \mathbb{P}_{x, M, N} \{ X_N = y | \tau_1 > N \} \quad (185)$$

represents the quasi-stationary density of the chain as it stays in  $\mathcal{E}_1$ . This last result corresponds to our result (62), and is known as the *Yaglom limit* of the process [15]. Taking the distribution at a time  $n < N$  before the conditioning, we obtain instead

$$l'(y) r'(y) = \lim_{n \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{P}_{x, M, N} \{ X_n = y | \tau_1 > N \}, \quad (186)$$

in agreement with (63).

For recent surveys on quasi-stationary distributions, see [14–16]; for applications in the context of large deviations, see [145, 146]; finally, see Bauer and Cornu [147] for a study of the effect of quasi-stationary conditioning on cycle affinities of finite-state jump processes.

## Appendix A: Derivation of the tilted generator

### 1. Pure jump processes

To derive the form of the tilted generator  $\mathcal{L}_k$  in the case of jump processes, we consider the conservative Markov generator  $G_k$  defined by

$$(G_k h)(x) = \int_{\mathcal{E}} W(x, dy) e^{kg(x, y)} [h(y) - h(x)] \quad (A1)$$

for all  $x \in \mathcal{E}$ . This generator is only a normalization factor away from  $\mathcal{L}_k$ , since

$$G_k = \mathcal{L}_k - (\mathcal{L}_k 1), \quad (A2)$$

which means that  $G_k$  and  $\mathcal{L}_k$  differ only in their diagonal elements.

The process described by  $G_k$  is a jump process with transition rates  $W(x, dy) e^{kg(x, y)}$ . The fact that  $e^{kg(x, y)}$  is strictly positive implies that the measure  $W(x, dy) e^{kg(x, y)}$  and the original measure  $W(x, dy)$  are absolutely continuous, which only means in this context that the  $G_k$  and  $L$  processes have the same set of allowed jumps  $x \rightarrow y$ . In this case, we can use Girsanov's Theorem, as applied

to jump processes [76], to obtain the Radon-Nikodym of the paths measure of the  $G_k$  process with respect to the path measure of the  $L$  process:

$$\frac{d\mathbb{P}_{G_k, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}}(\omega) = \exp \left( k \sum_{0 \leq t \leq T: \Delta X_t \neq 0} g(X_{t-}, X_{t+}) - \int_0^T dt [(We^{kg}1) - (W1)](X_t) \right). \quad (\text{A3})$$

Combining this result with the Feynman-Kac formula, we then arrive at

$$\begin{aligned} \frac{d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}} &= \frac{d\mathbb{P}_{\mathcal{L}_k, \mu_0, T}}{d\mathbb{P}_{G_k, \mu_0, T}} \frac{d\mathbb{P}_{G_k, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}} \\ &= \exp \left( k \sum_{\Delta X_t \neq 0} g(X_{t-}, X_{t+}) - \int_0^T dt [(We^{kg}1) - (W1)](X_t) + \int_0^T (\mathcal{L}_k 1)(X_t) dt \right), \end{aligned}$$

which yields (44) given the expression (46) of  $\mathcal{L}_k$ .

## 2. Diffusion processes

Given the generator  $L$  (27), we introduce a new Markov generator

$$\mathcal{L} = L + a \cdot \nabla + b, \quad (\text{A4})$$

involving the arbitrary vector field  $a$  and scalar field  $b$  on  $\mathcal{E}$ . Combining the Cameron-Martin-Girsanov Theorem and the Feynman-Kac formula [37–39], it can be shown that  $\mathcal{L}$  induces with the initial measure  $\mu_0$  a path measure  $\mathbb{P}_{\mathcal{L}, \mu_0, T}$ , which is absolutely continuous with respect to the path measure  $\mathbb{P}_{L, \mu_0, T}$ , and whose Radon-Nikodym derivative with respect to the latter measure is

$$\frac{d\mathbb{P}_{\mathcal{L}, \mu_0, T}}{d\mathbb{P}_{L, \mu_0, T}}(\omega) = e^{R_T(\omega)}, \quad (\text{A5})$$

where

$$\begin{aligned} R_T &= \int_0^T D^{-1}(X_t) a(X_t) \circ dX_t \\ &\quad + \int_0^T \left( b(X_t) - D^{-1}(X_t) a(X_t) \left( \hat{F} + \frac{a}{2} \right)(X_t) - \frac{1}{2} (\nabla \cdot a)(X_t) \right) dt. \end{aligned} \quad (\text{A6})$$

This is a generalization of the Cameron-Martin-Girsanov Theorem for non-conservative processes with  $b \neq 0$ . In the particular case where  $L$  is the generator of the Wiener process  $W_t$ , and there is no  $b$  perturbation, we recover the classical result

$$R_T = \int_0^T \left( a(W_t) dW_t - \frac{a(W_t)^2}{2} dt \right), \quad (\text{A7})$$

written here in the Itô convention [3]. In our case, we obtain the expression of  $\mathcal{L}_k$  from this general result by equating (A5) with (44) to obtain  $R_T = kT A_T$ , which is solved given (A6) for  $a = kDg$  and

$$b = kg \cdot \left( \hat{F} + k \frac{Dg}{2} \right) + \frac{k}{2} \nabla \cdot (Dg) + kf. \quad (\text{A8})$$

The expression of  $\mathcal{L}_k$  is therefore

$$\begin{aligned} \mathcal{L}_k &= L + kDg \nabla + kg \cdot \left( \hat{F} + k \frac{Dg}{2} \right) + \frac{k}{2} \nabla \cdot (Dg) + kf \\ &= \hat{F} \cdot (\nabla + kg) + (\nabla + kg) \frac{D}{2} (\nabla + kg) + kf. \end{aligned} \quad (\text{A9})$$

## Appendix B: Change of measure for the generalized Doob transform

To prove (73), it is sufficient to show that

$$\mathbb{E}_{L^{h,f},\mu_0}[C] = \mathbb{E}_{L,\mu_0} \left[ C h^{-1}(X_0) e^{-\int_0^T f(X_t)dt} h(X_T) \right] \quad (\text{B1})$$

for any cylinder function  $C$ . In this expression, the generators indicate with respect to which measure the expectation is taken. To arrive at this result, note first that

$$e^{tL^{h,f}}(x, dy) = h^{-1}(x) e^{t(L-f)}(x, dy) h(y) \quad (\text{B2})$$

for all  $(x, y) \in \mathcal{E}^2$ . Next, replace  $t$  by  $t - s$  and use the Feynman-Kac formula to express the exponential semi-group as an expectation, so that

$$e^{(t-s)L^{h,f}}(x, dy) = h^{-1}(x) \mathbb{E}_x \left[ e^{-\int_s^t f(X_u)du} \delta(X_t - y) dy \right] h(y) \quad (\text{B3})$$

for all  $(x, y) \in \mathcal{E}^2$ . Finally, expand (B1) and use (B3) iteratively to obtain

$$\begin{aligned} \mathbb{E}_{L^{h,f},\mu_0}[C] &= \int_{\mathcal{E}^{n+1}} C(x_0, \dots, x_n) \mu_0(dx_0) h^{-1}(x_0) \mathbb{E}_{x_0} [e^{-\int_0^{t_1} f(X_t)dt} \delta(X_{t_1} - x_1) dx_1] \dots \\ &\quad \dots \mathbb{E}_{x_{n-1}} [e^{-\int_{t_n}^T f(X_t)dt} \delta(X_T - x_n) dx_n] h(x_n), \end{aligned} \quad (\text{B4})$$

which, by multiple integration, is equal to (B1).

## Appendix C: Squared field for diffusion processes

Let  $L$  be the generator of the general diffusion defined in (27). The application of this generator on the product of two arbitrary functions  $f$  and  $g$  on  $\mathcal{E}$  yields

$$\begin{aligned} (Lf g) &= f \hat{F} \cdot \nabla g + g \hat{F} \cdot \nabla f + \frac{1}{2} \nabla (g D \nabla f + f D \nabla g) \\ &= f \hat{F} \cdot \nabla g + g \hat{F} \cdot \nabla f + \frac{g}{2} \nabla D (\nabla f) + \nabla f D \nabla g + \frac{f}{2} \nabla D (\nabla g) \\ &= f(Lg) + (Lf)g + \nabla f D \nabla g. \end{aligned} \quad (\text{C1})$$

Comparing with the definition (82) of the squared field  $\Gamma(f, g)$ , we find

$$\Gamma(f, g) = \nabla f D \nabla g. \quad (\text{C2})$$

Putting this result into (81) with  $f = \ln h$ , we then find (79), which represents the generator of a diffusion process with the same noise fields  $\sigma_\alpha$  as the diffusion described by  $L$ , but with the modified drift given in (80).

## Appendix D: Generator of the canonical path measure

We derive here the time-dependent generator associated with the canonical path measure. To this end, we consider this measure on the cylinder events  $\{X_0 = x_0, X_{t_1} = x_1, \dots, X_{t_n} = x_n\}$  with  $0 \leq t_1 \leq \dots \leq t_n \leq T$  to obtain, following (45),

$$d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(x_0, \dots, x_n) = \frac{\mu_0(dx_0) e^{t_1 \mathcal{L}_k}(x_0, dx_1) \dots e^{(t_n - t_{n-1}) \mathcal{L}_k}(x_{n-1}, dx_n) (e^{(T-t_n) \mathcal{L}_k} 1)(x_n)}{\mathbb{E}_{\mu_0}[e^{T k A_T}]}, \quad (\text{D1})$$



with the normalization added according to (41). Therefore,

$$d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(x_n|x_0,\dots,x_{n-1}) \equiv \frac{d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(x_0,\dots,x_n)}{d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(x_0,\dots,x_{n-1})} = \frac{e^{(t_n-t_{n-1})\mathcal{L}_k}(x_{n-1},dx_n)(e^{(T-t_n)\mathcal{L}_k}1)(x_n)}{(e^{(T-t_{n-1})\mathcal{L}_k}1)(x_{n-1})} \quad (\text{D2})$$

which shows that the conditional measure  $d\mathbb{P}_{k,\mu_0,T}^{\text{cano}}(x_n|x_0,\dots,x_{n-1})$  is Markovian, since it does not depend on all the previous points  $x_0,\dots,x_{n-1}$  but only on  $x_{n-1}$ . However, it is non-homogeneous, since it explicitly depends on  $t_{n-1}$ ,  $t_n$ , and  $T$ .

To derive the generator  $L_{k,t,T}^{\text{cano}}$  of this Markovian measure, we introduce the positive function

$$h_{t,T}(x) = (e^{(T-t)\mathcal{L}_k}1)(x) \quad (\text{D3})$$

to write the transition probability associated with (D2) as

$$P_{k,s,T}^{\text{cano},t}(x,dy) = \frac{1}{h_{s,T}(x)} e^{(t-s)\mathcal{L}_k}(x,dy) h_{t,T}(y). \quad (\text{D4})$$

Noting that  $h_{t,T}$  solves the backward differential equation

$$(\partial_t + \mathcal{L}_k)h_{t,T} = 0, \quad h_{T,T} = 1, \quad (\text{D5})$$

we then have that  $h_{t,T}$  is space-time harmonic with respect to  $\mathcal{L}_k$ , which implies from (86)-(87) that the canonical measure is the Doob transform of  $\mathcal{L}_k$  with the function  $h_{t,T}$  involving  $\mathcal{L}_k$  itself. This means explicitly that,

$$L_{k,t,T}^{\text{cano}} = (\mathcal{L}_k)^{(\exp((T-t)\mathcal{L}_k)1)}. \quad (\text{D6})$$

This result is valid for  $t < T$ , but also for  $t = T$  which yields

$$L_{k,T,T}^{\text{cano}} = \mathcal{L}_k - (\mathcal{L}_k 1). \quad (\text{D7})$$

In the limit  $T \rightarrow \infty$ ,  $L_{k,t,T}^{\text{cano}}$  becomes homogeneous; however, the limit is different for  $t < T$  and  $t = T$ , as shown in (125) and (126).

## Appendix E: Markov chains

We briefly re-express in this last section our main results for the simpler case of homogeneous Markov chains. In this context, the generalized Doob transform seems to have appeared for the first time in the work of Miller [148].

The sequence  $X_0, X_1, \dots, X_N$  of random variables is a homogeneous Markov chain if its joint measure is given by

$$d\mathbb{P}_{M,\mu_0,N}(x_1,\dots,x_N) = \mu_0(dx_0)M(x_0,dx_1)\dots M(x_{N-1},dx_N), \quad (\text{E1})$$

where  $M(x,dy)$  is the transition matrix and  $\mu_0$  is the initial measure for  $X_0$ . The generalized Doob transform of the Markov chain is defined as

$$M^h(x,dy) = \frac{1}{(Mh)(x)} M(x,dy)h(y), \quad (\text{E2})$$

where  $h$  is a strictly positive function on  $\mathcal{E}$ . This transformed matrix remains a stochastic matrix, as shown in [148], which can be used to define a discrete-time path measure  $d\mathbb{P}_{M^h,\mu_0,N}$ , whose Radon-Nikodym derivative with respect to the original Markov chain is

$$\frac{d\mathbb{P}_{M^h,\mu_0,N}}{d\mathbb{P}_{M,\mu_0,N}}(x_0,\dots,x_N) = \frac{1}{(Mh)(x_0)} \exp\left(\sum_{i=0}^{N-1} \ln \frac{h(x_{i+1})}{(Mh)(x_i)}\right) h(x_N). \quad (\text{E3})$$

This follows by re-expressing in discrete time the proof presented in Appendix B.

Consider now the observable

$$A_N = \frac{1}{N} \sum_{i=0}^{N-1} g(X_i, X_{i+1}), \quad (\text{E4})$$

where  $g : \mathcal{E}^2 \rightarrow \mathbb{R}$ . The tilted generator  $\mathcal{L}_k$  is replaced for this observable by the *tilted matrix*

$$\mathcal{M}_k(x, dy) = M(x, dy) e^{kg(x,y)}. \quad (\text{E5})$$

The particular observable

$$A_N = \frac{1}{N} \sum_{i=0}^{N-1} f(X_i) \quad (\text{E6})$$

is covered by this result simply by taking  $g(x, y) = f(x)$ , so that we do not have to consider additive and two-point observables separately.

The dominant (Perron-Frobenius) eigenvalue of  $\mathcal{M}_k$  is denote by  $\zeta_k$  and leads to the following result for the SCGF:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathbb{E}_{\mu_0}[e^{kN A_N}] = \ln \zeta_k \equiv \Lambda_k. \quad (\text{E7})$$

Denoting by  $r_k$  the right Perron-Frobenius eigenvector of  $\mathcal{M}_k$ , we can show, similarly to our previous results, that the Markov chain  $\{X_i\}$  conditioned on  $A_N = a$  is asymptotically equivalent to a Markov chain described by the following transition matrix:

$$M_k(x, dy) = \mathcal{M}_k^{r_k}(x, dy) = \frac{1}{\zeta_k r_k(x)} \mathcal{M}_k(x, dy) r_k(y). \quad (\text{E8})$$

The stationary density of this driven process is the same as in the continuous-time case, namely,  $\rho_k(x) = l_k(x) r_k(x)$ , where  $l_k(x)$  is the left eigenvector of  $\mathcal{M}_k$  associated with  $\zeta_k$ . Moreover, all our results about the reversibility of this density apply with minor changes.

## ACKNOWLEDGMENTS

We would like to thank Patrick Cattiaux, Mike Evans, Rosemary J. Harris, and Vivien Lecomte for useful discussions on this paper. We are also grateful for the hospitality and support of the Laboratoire J. A. Dieudonné, the Kavli Institute for Theoretical Physics, and the Galileo Galilei Institute for Theoretical Physics, where parts of this work were developed and written. Further financial support was received from the ANR STOSYMAP (ANR-2011-BS01-015), the National Research Foundation of South Africa (CSUR 13090934303) and Stellenbosch University (project funding for new appointee).

- 
- [1] J. L. Doob, “Conditional Brownian motion and the boundary limits of harmonic functions,” [Bull. Soc. Math. Fr. \*\*85\*\*, 431–458 \(1957\)](#).
  - [2] J. L. Doob, *Classical Potential Theory and Its Probabilistic Counterpart* (Springer, New York, 1984).
  - [3] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales* (Cambridge University Press, Cambridge, 2000).

- [4] F. Baudoin, “Conditioned stochastic differential equations: Theory, examples and application to finance,” *Stoch. Proc. Appl.* **100**, 109–145 (2002).
- [5] D. Gasbarra, T. Sottinen, and E. Valkeila, “Gaussian bridges,” in *Stochastic Analysis and Applications*, Abel Symposia, Vol. 2, edited by F. Espen Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal, and T. Zhang (Springer, Berlin, 2007) pp. 361–382.
- [6] T. Sottinen and A. Yazigi, “Generalized Gaussian bridges,” *Stoch. Proc. Appl.* **124**, 3084–3105 (2014).
- [7] E. Schrödinger, “Über die Umkehrung der Naturgesetze,” *Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl.*, 144–153 (1931).
- [8] E. Schrödinger, “Sur la théorie relativiste de l’électron et l’interprétation de la mécanique quantique,” *Ann. Inst. Henri Poincaré* **2**, 269–319 (1932).
- [9] B. Jamison, “The Markov processes of Schrödinger,” *Z. Wahrsch. Verw. Gebiete* **32**, 323–331 (1975).
- [10] J. C. Zambrini, “Stochastic mechanics according to Schrödinger,” *Phys. Rev. A* **33**, 1532–1548 (1986).
- [11] R. Aebi, *Schrödinger Diffusion Processes* (Birkhäuser, Basel, 1996).
- [12] J. N. Darroch and E. Seneta, “On quasi-stationary distributions in absorbing discrete-time finite Markov chains,” *J. Appl. Prob.* **2**, 88–100 (1965).
- [13] J. N. Darroch and E. Seneta, “On quasi-stationary distributions in absorbing continuous-time finite Markov chains,” *J. Appl. Prob.* **4**, 192–196 (1967).
- [14] D. Villemonais, “Quasi-stationary distributions and population processes,” *Prob. Surveys* **9**, 340–410 (2012).
- [15] P. Collet, S. Martínez, and J. San Martín, *Quasi-Stationary Distributions* (Springer, New York, 2013).
- [16] E. A. van Doorn and P. K. Pollett, “Quasi-stationary distributions for discrete-state models,” *Eur. J. Op. Res.* **230**, 1–14 (2013).
- [17] O. A. Vasicek, “A conditional law of large numbers,” *Ann. Prob.* **8**, 142–147 (1980).
- [18] J. M. van Campenhout and T. M. Cover, “Maximum entropy and conditional probability,” *IEEE Trans. Info. Th.* **27**, 483–489 (1981).
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley, New York, 1991).
- [20] I. Csiszar, “Sanov property, generalized  $I$ -projection and a conditional limit theorem,” *Ann. Prob.* **12**, 768–793 (1984).
- [21] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. (Springer, New York, 1998).
- [22] P. Diaconis and D. Freedman, “A dozen de Finetti-style results in search of a theory,” *Ann. Inst. Henri Poincaré B: Prob. Stat.* **23**, 397–423 (1987).
- [23] P. Diaconis and D. A. Freedman, “Conditional limit theorems for exponential families and finite versions of de Finetti’s theorem,” *J. Theoret. Prob.* **1**, 381–410 (1988).
- [24] A. Dembo and O. Zeitouni, “Refinements of the Gibbs conditioning principle,” *Prob. Th. Rel. Fields* **104**, 1–14 (1996).
- [25] I. Csiszár, T. M. Cover, and B.-S. Choi, “Conditional limit theorems under Markov conditioning,” *IEEE Trans. Info. Th.* **33**, 788–801 (1987).
- [26] V. S. Borkar, S. Juneja, and A. A. Kherani, “Performance analysis conditioned on rare events: An adaptive simulation scheme,” *Commun. Info. Syst.* **3**, 259–278 (2004).
- [27] J. A. Bucklew, *Introduction to Rare Event Simulation* (Springer, New York, 2004).
- [28] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II (Wiley, New York, 1970).
- [29] C. Giardinà, J. Kurchan, and L. Peliti, “Direct evaluation of large-deviation functions,” *Phys. Rev. Lett.* **96**, 120603 (2006).
- [30] V. Lecomte and J. Tailleur, “A numerical approach to large deviations in continuous time,” *J. Stat. Mech.* **2007**, P03004 (2007).
- [31] J. Tailleur and V. Lecomte, “Simulation of large deviation functions using population dynamics,” in *Modeling and Simulation of New Materials: Proceedings of Modeling and Simulation of New Materials*, Vol. 1091, edited by J. Marro, P. L. Garrido, and P. I. Hurtado (AIP, Melville, NY, 2009) pp. 212–219.
- [32] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Probability and Its Applications (Springer, New York, 2004).
- [33] W. H. Fleming, “Exit probabilities and optimal stochastic control,” *Appl. Math. Optim.* **4**, 329–346 (1978).
- [34] W. H. Fleming, “Logarithmic transformations and stochastic control,” in *Advances in Filtering and Optimal Stochastic Control*, Lecture Notes in Control and Information Sciences, Vol. 42, edited by

- W. H. Fleming and L. G. Gorostiza (Springer, New York, 1982) pp. 131–141.
- [35] W. H. Fleming, “A stochastic control approach to some large deviations problems,” in *Recent Mathematical Methods in Dynamic Programming*, Vol. 1119, edited by I. C. Dolcetto, W. H. Fleming, and T. Zolezzi (Springer, 1985) pp. 52–66.
  - [36] W. H. Fleming and S.-J. Sheu, “Stochastic variational formula for fundamental solutions of parabolic pde,” *Appl. Math. Optim.* **13**, 193–204 (1985).
  - [37] M. Kac, “On distributions of certain Wiener functionals,” *Trans. Amer. Math. Soc.* **65**, 1–13 (1949).
  - [38] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes* (Springer, New York, 1979).
  - [39] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd ed. (Springer, Berlin, 1999).
  - [40] W. H. Fleming and H. M. Soner, *Controlled Markov Processes And Viscosity Solutions*, Stochastic Modelling and Applied Probability, Vol. 25 (Springer, New York, 2006).
  - [41] S.-J. Sheu, “Stochastic control and principal eigenvalue,” *Stochastics* **11**, 191–211 (1984).
  - [42] W. H. Fleming, S. J. Sheub, and H. M. Soner, “A remark on the large deviations of an ergodic Markov process,” *Stochastics* **22**, 187–199 (1987).
  - [43] W. H. Fleming and S.-J. Sheu, “Asymptotics for the principal eigenvalue and eigenfunction of a nearly first-order operator with large potential,” *Ann. Prob.* **25**, 1953–1994 (1997).
  - [44] W. H. Fleming and W. M. McEneaney, “Risk sensitive optimal control and differential games,” in *Stochastic Theory and Adaptive Control*, Lecture Notes in Control and Information Sciences, Vol. 184, edited by T. E. Duncan and B. Pasik-Duncan (Springer, New York, 1992) pp. 185–197.
  - [45] W. Fleming and W. McEneaney, “Risk-sensitive control on an infinite time horizon,” *SIAM J. Cont. Opt.* **33**, 1881–1915 (1995).
  - [46] P. Dupuis and W. McEneaney, “Risk-sensitive and robust escape criteria,” *SIAM J. Cont. Opt.* **35**, 2021–2049 (1997).
  - [47] T. Nemoto and S.-I. Sasa, “Thermodynamic formula for the cumulant generating function of time-averaged current,” *Phys. Rev. E* **84**, 061113 (2011).
  - [48] T. Nemoto and S.-I. Sasa, “Variational formula for experimental determination of high-order correlations of current fluctuations in driven systems,” *Phys. Rev. E* **83**, 030105 (2011).
  - [49] T. Nemoto and S.-I. Sasa, “Computation of large deviation statistics via iterative measurement-and-feedback procedure,” *Phys. Rev. Lett.* **112**, 090602 (2014).
  - [50] H. Touchette, “The large deviation approach to statistical mechanics,” *Phys. Rep.* **478**, 1–69 (2009).
  - [51] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Grundlehren der Mathematischen Wissenschaften, Vol. 260 (Springer, New York, 1984).
  - [52] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, 2nd ed., Springer Series in Synergetics, Vol. 13 (Springer, New York, 1985).
  - [53] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
  - [54] H. Risken, *The Fokker-Planck equation: Methods of solution and applications*, 3rd ed. (Springer, Berlin, 1996).
  - [55] R. M. L. Evans, “Rules for transition rates in nonequilibrium steady states,” *Phys. Rev. Lett.* **92**, 150601 (2004).
  - [56] R. M. L. Evans, “Detailed balance has a counterpart in non-equilibrium steady states,” *J. Phys. A: Math. Gen.* **38**, 293–313 (2005).
  - [57] R. M. L. Evans, “Statistical physics of shear flow: a non-equilibrium problem,” *Contemp. Phys.* **51**, 413–427 (2010).
  - [58] C. Dellago, P. G. Bolhuis, and P. L. Geissler, “Transition path sampling,” *Adv. Chem. Phys.* **123**, 1–78 (2003).
  - [59] C. Dellago, P. G. Bolhuis, and P. L. Geissler, “Transition path sampling methods,” in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, Lecture Notes in Physics, Vol. 703, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer, New York, 2006).
  - [60] C. Dellago and P. Bolhuis, “Transition path sampling and other advanced simulation techniques for rare events,” in *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, Advances in Polymer Science, Vol. 221, edited by C. Holm and K. Kremer (Springer, Berlin, 2009) pp. 167–233.
  - [61] E. Vanden-Eijnden, “Transition path theory,” in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, Lecture Notes in Physics, Vol. 703, edited by M. Ferrario,

- G. Ciccotti, and K. Binder (Springer, 2006) pp. 453–493.
- [62] D. Chandler and J. P. Garrahan, “Dynamics on the way to forming glass: Bubbles in space-time,” *Ann. Rev. Chem. Phys.* **61**, 191–217 (2010).
  - [63] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, “Dynamic order-disorder in atomistic models of structural glass formers,” *Science* **323**, 1309–1313 (2009).
  - [64] R. L. Jack and P. Sollich, “Large deviations and ensembles of trajectories in stochastic models,” *Prog. Theoret. Phys. Suppl.* **184**, 304–317 (2010).
  - [65] V. Lecomte, C. Appert-Rolland, and F. van Wijland, “Chaotic properties of systems with Markov dynamics,” *Phys. Rev. Lett.* **95**, 010601 (2005).
  - [66] V. Lecomte, C. Appert-Rolland, and F. van Wijland, “Thermodynamic formalism for systems with Markov dynamics,” *J. Stat. Phys.* **127**, 51–106 (2007).
  - [67] J. P. Garrahan and I. Lesanovsky, “Thermodynamics of quantum jump trajectories,” *Phys. Rev. Lett.* **104**, 160601 (2010).
  - [68] J. P. Garrahan, A. D. Armour, and I. Lesanovsky, “Quantum trajectory phase transitions in the micromaser,” *Phys. Rev. E* **84**, 021115 (2011).
  - [69] C. Ates, B. Olmos, J. P. Garrahan, and I. Lesanovsky, “Dynamical phases and intermittency of the dissipative quantum ising model,” *Phys. Rev. A* **85**, 043620 (2012).
  - [70] S. Genway, J. P. Garrahan, I. Lesanovsky, and A. D. Armour, “Phase transitions in trajectories of a superconducting single-electron transistor coupled to a resonator,” *Phys. Rev. E* **85**, 051122 (2012).
  - [71] J. M. Hickey, S. Genway, I. Lesanovsky, and J. P. Garrahan, “Thermodynamics of quadrature trajectories in open quantum systems,” *Phys. Rev. A* **86**, 063824 (2012).
  - [72] R. Chetrite and H. Touchette, “Nonequilibrium microcanonical and canonical ensembles and their equivalence,” *Phys. Rev. Lett.* **111**, 120601 (2013).
  - [73] I. Karatzas and S. Shreve, *Methods of Mathematical Finance*, Stochastic Modelling and Applied Probability, Vol. 39 (Springer, 1998).
  - [74] H. C. Berg, *Random Walks in Biology* (Princeton University Press, Princeton, 1993).
  - [75] H. Spohn, *Large Scale Dynamics of Interacting Particles* (Springer Verlag, Heidelberg, 1991).
  - [76] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*, Grundlehren der mathematischen Wissenschaften, Vol. 320 (Springer-Verlag, Berlin, 1999).
  - [77] T. M. Liggett, *Interacting Particle Systems*, Classics in Mathematics (Springer, New York, 2004).
  - [78] B. Derrida, “Non-equilibrium steady states: Fluctuations and large deviations of the density and of the current,” *J. Stat. Mech.* **2007**, P07023 (2007).
  - [79] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, “Stochastic interacting particle systems out of equilibrium,” *J. Stat. Mech.* **2007**, P07014 (2007).
  - [80] K. Jacobs, *Stochastic Processes for Physicists: Understanding Noisy Systems* (Cambridge University Press, Cambridge, 2010).
  - [81] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).
  - [82] E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton University Press, Princeton, 1967).
  - [83] W. J. Anderson, *Continuous-Time Markov Chains: An Applications-Oriented Approach*, Springer Series in Statistics (Springer-Verlag, 1991).
  - [84] K. L. Chung and J. B. Walsh, *Markov Processes, Brownian Motion, and Time Symmetry*, 2nd ed. (Springer, Berlin, 2005).
  - [85] D. Applebaum, *Lévy Processes and Stochastic Calculus*, 2nd ed. (Cambridge University Press, Cambridge, 2009).
  - [86] K. Sato, *Lévy processes and Infinite Divisibility*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 1999).
  - [87] L. Hörmander, “Hypoelliptic second order differential equations,” *Acta. Math.* **119**, 147–171 (1967).
  - [88] P. Malliavin, “Stochastic calculus of variations and hypoelliptic operators,” in *Proc. Inter. Symp. Stoch. Diff. Equations, Kyoto 1976* (Wiley, New York, 1978) pp. 195–263.
  - [89] J. L. Lebowitz and H. Spohn, “A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics,” *J. Stat. Phys.* **95**, 333–365 (1999).
  - [90] C. Maes and K. Netočný, “Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states,” *Europhys. Lett.* **82** (2008).
  - [91] M. Baiesi, C. Maes, and B. Wynants, “Fluctuations and response of nonequilibrium states,” *Phys.*



- Rev. Lett. **103**, 010602 (2009).
- [92] C. Jarzynski, “Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach,” *Phys. Rev. E* **56**, 5018–5035 (1997).
  - [93] G. E. Crooks, “Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems,” *J. Stat. Phys.* **90**, 1481–1487 (1998).
  - [94] K. Sekimoto, *Stochastic Energetics*, Lect. Notes. Phys., Vol. 799 (Springer, New York, 2010).
  - [95] C. Maes, K. Netočný, and B. Wynants, “Steady state statistics of driven diffusions,” *Physica A* **387**, 2675–2689 (2008).
  - [96] V. Chernyak, M. Chertkov, S. Malinin, and R. Teodorescu, “Non-equilibrium thermodynamics and topology of currents,” *J. Stat. Phys.* **137**, 109–147 (2009).
  - [97] A. Basu, T. Bhattacharyya, and V. S. Borkar, “A learning algorithm for risk-sensitive cost,” *Maths. Op. Res.* **33**, 880–898 (2008).
  - [98] F. den Hollander, *Large Deviations*, Fields Institute Monograph (Amer. Math. Soc., Providence, R.I., 2000).
  - [99] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer, New York, 1985).
  - [100] A. De La Fortelle, “Large deviation principle for Markov chains in continuous time,” *Prob. Info. Trans.* **37**, 120–139 (2001).
  - [101] L. Bertini, A. Faggionato, and D. Gabrielli, “Large deviations of the empirical flow for continuous time markov chains,” (2012), [arXiv:1210.2004](#).
  - [102] B. Roynette and M. Yor, *Penalising Brownian Paths* (Springer, New York, 2009).
  - [103] J. T. Lewis, C.-E. Pfister, and G. W. Sullivan, “The equivalence of ensembles for lattice systems: Some examples and a counterexample,” *J. Stat. Phys.* **77**, 397–419 (1994).
  - [104] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, “Entropy, concentration of probability and conditional limit theorem,” *Markov Proc. Rel. Fields* **1**, 319–386 (1995).
  - [105] H. Touchette, “Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels,” (2014), [arXiv:1403.6608](#).
  - [106] C. R. MacCluer, “The many proofs and applications of Perron’s theorem,” *SIAM Rev.* **42**, 487–498 (2000).
  - [107] M. G. Krein and M. A. Rutman, “Linear operators leaving a cone in a Banach space,” *Am. Math. Soc. Transl.* **26**, 128 (1950).
  - [108] V. Lecomte, *Thermodynamique des histoires et fluctuations hors d’équilibre*, Ph.D. thesis, Université Paris VII (2007).
  - [109] R. Chetrite and S. Gupta, “Two refreshing views of fluctuation theorems through kinematics elements and exponential martingale,” *J. Stat. Phys.* **143**, 543–584 (2011), [10.1007/s10955-011-0184-0](#).
  - [110] H. Kunita, “Absolute continuity of markov processes and generators,” *Nagoya Mathematical Journal* **36**, 1–26 (1969).
  - [111] K. Itô and S. Watanabe, “Transformation of markov processes by multiplicative functionals,” *Ann. Inst. Fourier* **15**, 13–30 (1965).
  - [112] Z. Palmowski and T. Rolski, “A technique for exponential change of measure for Markov processes,” *Bernoulli* **8**, 767–785 (2002).
  - [113] P. Diaconis and L. Miclo, “On characterizations of Metropolis type algorithms in continuous time,” *ALEA Lat. Am. J. Probab. Math. Stat.* **6**, 199–238 (2009).
  - [114] P. A. Meyer and W. A. Zheng, “Construction de processus de Nelson réversibles,” in *Séminaire de Probabilités XIX 1983/84*, Lecture Notes in Mathematics, Vol. 1123, edited by J. Azéma and M. Yor (Springer Berlin Heidelberg, 1985) pp. 12–26.
  - [115] W. Bernard and H. B. Callen, “Irreversible thermodynamics of nonlinear processes and noise in driven systems,” *Rev. Mod. Phys.* **31**, 1017–1044 (1959).
  - [116] H. B. Callen and T. A. Welton, “Irreversibility and generalized noise,” *Phys. Rev.* **83**, 34–40 (1951).
  - [117] R. Kubo, “The fluctuation-dissipation theorem,” *Rep. Prog. Phys.* **29**, 255 (1966).
  - [118] E. Lippiello, F. Corberi, A. Sarracino, and M. Zannetti, “Nonlinear response and fluctuation-dissipation relations,” *Phys. Rev. E* **78**, 041120 (2008).
  - [119] T. Speck and U. Seifert, “Restoring a fluctuation-dissipation theorem in a nonequilibrium steady state,” *Europhys. Lett.* **74**, 391 (2006).
  - [120] M. Baiesi, C. Maes, and B. Wynants, “Nonequilibrium linear response for Markov dynamics I: Jump processes and overdamped diffusions,” *J. Stat. Phys.* **137**, 1094–1116 (2009).

- [121] M. Baiesi, E. Boksenbojm, C. Maes, and B. Wynants, “Nonequilibrium linear response for Markov dynamics II: Inertial dynamics,” *J. Stat. Phys.* **139**, 492–505 (2010).
- [122] U. Seifert and T. Speck, “Fluctuation-dissipation theorem in nonequilibrium steady states,” *Europhys. Lett.* **89**, 10007 (2010).
- [123] R. Chetrite, G. Falkovich, and K. Gawędzki, “Fluctuation relations in simple examples of non-equilibrium steady states,” *J. Stat. Mech.* **2008**, P08005 (2008).
- [124] R. Chetrite and K. Gawędzki, “Eulerian and Lagrangian pictures of non-equilibrium diffusions,” *J. Stat. Phys.* **137**, 890–916 (2009).
- [125] C. Jarzynski, “Nonequilibrium equality for free energy differences,” *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
- [126] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, “Probability of second law violations in shearing steady states,” *Phys. Rev. Lett.* **71**, 2401–2404 (1993).
- [127] G. Gallavotti and E. G. D. Cohen, “Dynamical ensembles in nonequilibrium statistical mechanics,” *Phys. Rev. Lett.* **74**, 2694–2697 (1995).
- [128] G. Gallavotti and E. G. D. Cohen, “Dynamical ensembles in stationary states,” *J. Stat. Phys.* **80**, 931–970 (1995).
- [129] F. J. Dyson, “A Brownian motion model for the eigenvalues of a random matrix,” *J. Math. Phys.* **3**, 1191–1198 (1962).
- [130] D. J. Grabiner, “Brownian motion in a Weyl chamber, non-colliding particles, and random matrices,” *Ann. Inst. Henri Poincaré B: Prob. Stat.* **35**, 177–204 (1999).
- [131] N. O’Connell, “Random matrices, non-colliding processes and queues,” in *Séminaire de Probabilités XXXVI*, Lecture Notes in Mathematics, Vol. 1801, edited by J. Azéma, M. Émery, M. Ledoux, and M. Yor (Springer, Berlin, 2003) pp. 165–182.
- [132] H. Touchette, “Ensemble equivalence for general many-body systems,” *Europhys. Lett.* **96**, 50010 (2011).
- [133] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970).
- [134] T. R. Gingrich, S. Vaikuntanathan, and P. L. Geissler, “Heterogeneity-induced large deviations in activity and (in some cases) entropy production,” (2014), [arXiv:1406.3311](#).
- [135] N. Merhav and Y. Kafri, “Bose-Einstein condensation in large deviations with applications to information systems,” *J. Stat. Mech.* **2010**, P02011 (2010).
- [136] R. J. Harris, V. Popkov, and G. M. Schütz, “Dynamics of instantaneous condensation in the ZRP conditioned on an atypical current,” *Entropy* **15**, 5065–5083 (2013).
- [137] J. Szavits-Nossan, M. R. Evans, and S. N. Majumdar, “Constraint-driven condensation in large fluctuations of linear statistics,” *Phys. Rev. Lett.* **112**, 020602 (2014).
- [138] A. Baule and R. M. L. Evans, “Invariant quantities in shear flow,” *Phys. Rev. Lett.* **101**, 240601 (2008).
- [139] A. Simha, R. M. L. Evans, and A. Baule, “Properties of a nonequilibrium heat bath,” *Phys. Rev. E* **77**, 031117 (2008).
- [140] A. Baule and R. M. L. Evans, “Nonequilibrium statistical mechanics of shear flow: invariant quantities and current relations,” *J. Stat. Mech.* **2010**, P03030 (2010).
- [141] V. Popkov, G. M. Schütz, and D. Simon, “ASEP on a ring conditioned on enhanced flux,” *J. Stat. Mech.* **2010**, P10007 (2010).
- [142] V. Popkov and G. Schütz, “Transition probabilities and dynamic structure function in the ASEP conditioned on strong flux,” *J. Stat. Phys.* **142**, 627–639 (2011).
- [143] R. L. Jack and P. Sollich, “Large deviations of the dynamical activity in the East model: Analysing structure in biased trajectories,” *J. Phys. A: Math. Theor.* **47**, 015003 (2014).
- [144] M. Knežević and R. M. L. Evans, “Numerical comparison of a constrained path ensemble and a driven quasisteady state,” *Phys. Rev. E* **89**, 012132 (2014).
- [145] J. Chen, H. Li, and S. Jian, “Some limit theorems for absorbing Markov processes,” *J. Phys. A: Math. Theor.* **45**, 345003 (2012).
- [146] J. Chen and X. Deng, “Large deviations and related problems for absorbing Markov chains,” *Stoch. Proc. Appl.* **123**, 2398–2418 (2013).
- [147] M. Bauer and F. Cornu, “Affinity and fluctuations in a mesoscopic noria,” (2014), [arXiv:1402.2422](#).
- [148] H. D. Miller, “A convexity property in the theory of random variables defined on a finite Markov chain,” *Ann. Math. Stat.* **32**, 1260–1270 (1961).