Unified theory of thermodynamics and stochastic thermodynamics for nonlinear Langevin systems driven by non-conservative forces

Mingnan Ding [1], Fei Liu [2] and Xiangjun Xing [1,3,4,*]

[1] *Wilczek Quantum Center, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*

[2] *School of Physics, Beihang University, Beijing, China*

[3] *T.D. Lee Institute, Shanghai Jiao Tong University, Shanghai 200240, China*

[4] *Shanghai Research Center for Quantum Sciences, Shanghai 201315, China*

(Received 3 August 2022; accepted 20 October 2022; published 22 November 2022)

We construct a unified theory of thermodynamics and stochastic thermodynamics for classical nonequilibrium systems driven by non-conservative forces, using the recently developed covariant Ito-Langevin theory. The thermodynamic forces are split into a conservative part and a non-conservative part. Thermodynamic functions are defined using the reference conservative system. Work and heat are partitioned into *excess parts* and *house-keeping parts*, which are due to, respectively, conservative forces and non-conservative forces. Excess entropy production (EP) and house-keeping EP are analogously defined. The splitting of thermodynamic forces is subjected to an arbitrariness resembling a gauge symmetry, with each gauge defining a reference conservative Langevin system. In the special *Gibbs gauge*, the nonequilibrium steady state (NESS) is characterized by Gibbs canonical distribution, the excess heat agrees with that defined by Hatano and Sasa, and the excess EP agrees with that of Glansdorff and Prigogine, i.e., it is the time rate of the second-order variation of system entropy near the NESS. Adopting the Gibbs gauge, and focusing on the excess parts of thermodynamic quantities, a complete analogy between thermodynamics of non-conservative systems and that of conservative systems is established. One important consequence of this analogy is that both the free energy and excess EP are minimized at NESS. Our theory therefore constitutes a statistical foundation both for the steady-state thermodynamics theory due to Sasa and Tasaki and for the stability theory of NESS due to Glansdorff and Prigogine. These results are valid even if the system is far from equilibrium. By studying detailed fluctuation theorem, we find striking differences between systems with symmetric kinetic matrices and those with asymmetric kinetic matrices. For systems with asymmetric kinetic matrices, the total EP is the sum of house-keeping EP, excess EP, and pumped entropy. Entropy pumping is an exchange of entropy between the system and environment without necessarily involving dissipation. In the presence of entropy pumping, the system may behave as either a demon or an antidemon. Fluctuation theorems and work relations are derived both for total work and for excess work. For systems with symmetric kinetic matrices, there is no entropy pumping, yet in the Gibbs gauge, the excess work and house-keeping work each satisfies a separate fluctuation theorem. We illustrate our theory using many concrete examples.

DOI: 10.1103/PhysRevResearch.4.043125

I. INTRODUCTION

When a system, whether large or small, is in contact with a thermal bath and is perturbed by constant external forces, it may converge to an equilibrium state, or to a nonequilibrium steady state (NESS). For the sake of convenience, we shall call the former system *conservative* whereas the latter system *non-conservative*. These terms will be made precise in the setting of nonlinear Langevin dynamics later in this paper. Whilst thermodynamics and stochastic thermodynamics of conservative systems are well understood, the situation of non-conservative systems turn out to be much more complicated. Study of the statistical mechanics and thermodynamics of NESS have a very long and convoluted history, and is still deemed unfinished. A brief review of important ideas and theories up to 2000 is provided in Sec. I of Ref. [1]. We also refer to the readers to classic textbooks [2–7] and papers [8–14] for more details. Both the classical theory of irreversible thermodynamics [2,3] and linear response theory [13] are applicable only for near-equilibrium systems. The least dissipation principle of Onsager and Machlup [11,12] can be used to characterize the most probable path of a fluctuating nonequilibrium system, but not the statistical distribution. Glansdorff and Prigogine [7,15] defined the excess entropy production (EP) as the rate of the second-order variation of system entropy around the NESS, and showed that its positivity characterizes the stability of many NESS systems. Its applicability in the general setting was however disputed

*xxing@sjtu.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

[16]. A promising candidate for thermodynamic theory of NESS is the *steady-state thermodynamics* (SST), initiated by Oono and Paniconi [17], and further developed by Sasa and Tasaki [1], which aims to establish a full analogy between NESS and equilibrium thermodynamics. A crucial step in this endeavor is to introduce the concept of *house-keeping heat*, which must be subtracted from the total heat so that the remaining *excess heat* is finite for a quasistatic transition between different NESS. *The principle of minimum (excess) work* [1,17] emerges as an important conjecture, which may be understood as an analog of second law for nonequilibrium steady systems.

Stochastic thermodynamics [18–22], which combines Markov dynamics with nonequilibrium thermodynamics, constitutes a powerful framework for statistical physics of nonequilibrium systems. It provides the most natural setting for study of various fluctuation theorems [18,21], which are rooted in local detailed balance properties, same as Onsager-Casimir reciprocal symmetry of linear kinetic coefficients [8–10]. In the setting of NESS, the steady-state fluctuation theorem [23,24] supplies an asymptotically exact relation between rate of EP and time-reversal asymmetry of path probability. The *Hatano-Sasa equality* [25] was proved for general Markov processes without detailed balance, and its connections with *excess heat* and "the second law of SST" [1,25] were established for over-damped Brownian dynamics driven by non-conservative forces. Speck and Seifert [26] proved a similar equality for the *house-keeping heat*, which is the difference between the total heat and the excess heat. Chernyak, Chertkov, and Jarzynski [27] studied general non-conservative Langevin models with even variables, and show that Crooks fluctuation theorem [22] can be constructed for two distinct types of EP-like quantities, corresponding to two distinct definitions of time-reversal in the trajectory space. Esposito and Van der Broeck [28–30] showed that, again for systems with even variables, the total EP can be decomposed into two parts, which they call adiabatic EP and nonadiabatic EP, and each of which satisfies a fluctuation theorem. Integration of these fluctuation theorems yield respectively the equalities established in Ref. [25] and in Ref. [26]. About the same time, Ge and Qian [31] clarified the thermodynamic meanings of these components of EP for master equation systems. Later, it was shown independently by Spinney and Ford [32,33] and by Park *et al.* [34] that for general Langevin systems with both even and odd variables, such a decomposition does not work. Alone a different line, Qian [35,36] studied entropy production for Langevin dynamics with even variables and additive noises. He showed that the stability condition of NESS is indeed the positivity of excess entropy production as defined by Glansdorff and Prigogine, i.e., the rate of the second-order variation of system entropy. This explicitly verifies the validity of Glansdorff-Prigogine stability criterion, without assuming that the system is near equilibrium. Additionally, Kim and Qian [37,38] studied underdamped Hamiltonian systems driven by velocity-dependent force, and discovered a novel effect of *entropy pumping*, which describes exchange of entropy between the system and its environment without involving dissipation. This stimulated some further studies on thermodynamics of information-feedback systems in underdamped Hamiltonian dynamics [39,40].

Whilst all these ground-breaking papers, as well as many ensuing papers on related problems [41–47], provide deep insights about thermodynamics and stochastic thermodynamics of Langevin systems driven by non-conservative forces, many important questions remain open or only partially answered. For example, the terms house-keeping and excess have been introduced for heat, work, entropy, and production in many papers and in seemingly different ways [1,7,17,36,43,44], it is not clear whether these definitions are consistent with each other. Whilst most previous theories are formulated for systems with even variables and even control parameters, it is not clear whether and to what extent they can be extended to more general systems with both even and odd variables and parameters. Additionally, the relation between Glansdorff-Prigogine stability criterion and the second law of SST is not clear in general setting. The full correspondence between SST and stochastic thermodynamics has not been worked out. Finally, it is not understood how entropy pumping works in non-Hamiltonian systems. In summary, a general and fully consistent theory of stochastic thermodynamics for non-conservative systems is not yet available.

With all these questions in mind, we set out to construct a unified theory of thermodynamics and stochastic thermodynamics for nonlinear Langevin systems driven by non-conservative forces. During the course, the terms "excess" and "house-keeping" will acquire precise and consistent meanings, and their relations with previous usages will be clarified. Major results about non-conservative systems, previously established for more specific systems [25–28,37], will be rederived in a more general setting and synthesized into a unified perspective. A complete analogy between conservative systems and non-conservative systems will be established. The Glansdorff-Prigogine stability criterion will be rigorously established for NESS of generic systems driven by non-conservative forces, which turns out to be equivalent to the second law of SST. Overall, a firm foundation of SST will be established.

The present paper is the fourth of a sequel dedicated to a unified theory of thermodynamics and stochastic thermodynamics for nonlinear Langevin systems, which is designed to be applicable to classical nonequilibrium systems with continuous state variables in a very broad setting: The system may contain both even and odd variables and control parameters; the noises acting on system are generally multiplicative; the space of system variables may be curved; the interaction between the system and environment may be strong; the forces driving the system may be conservative or non-conservative. The theory is constructed on two basic assumptions: (1) white noises, and (2) detailed balance, which reflects the time-reversal symmetry of microscopic dynamics. The precise definition of detailed balance, is supplied in Sec. II for conservative systems and in Sec. III for non-conservative systems.

In the first paper [48], a covariant formulation of Ito-Langevin dynamics was developed. A salient feature of this formulation is a clear separation of the static aspect, characterized by a generalized potential U, and the kinetic aspect, characterized by a kinetic matrix L^{ij}. This allows a simple and covariant formulation of detailed balance conditions. Special attention is paid to a peculiar term called *spurious drift*, which shows up when the noises are multiplicative, or the state

space is curved. Neglect of spurious drift generally spoils covariance and detailed balance. In the second paper [49], the covariant Langevin dynamics was used to construct a covariant stochastic thermodynamic theory for small systems in contact with a single heat bath. Using concrete examples, it was demonstrated that whenever spurious drift shows up, the conventional theories of stochastic energetics and stochastic thermodynamics must be replaced by the covariant theory. Mathematically spurious drift arises because the differential dx in stochastic differential equations do not transform as regular vector, but according to the exotic Ito's rule. We refer to readers to Refs. [48,49] for more details about spurious drift. In the third paper [50], it was demonstrated that the formalism is applicable to systems that are strongly coupled to their environments, as long as the Hamiltonian of mean force is defined as the fluctuating internal energy. There is no need to change the definitions of other thermodynamic variables.

In all three preceding papers [48–50], it was assumed that the thermodynamic forces acting on the system are conservative, i.e., they can be expressed as negative the gradient of a generalized potential U, which is associated directly with a thermodynamic equilibrium. For fixed system parameters, the system converges to this equilibrium state. There are however many *non-conservative* systems, which converge not to any equilibrium state, but to NESS with positive EP. The thermodynamic forces acting on these systems are said to be non-conservative. A typical example is a polymer chain dragged by a constant force supplied by, e.g., an optical or magnetic tweezer. In the NESS, the polymer moves with a constant speed, and also exhibits internal deformation. It is the purpose of the present paper to develop a theory of stochastic thermodynamics for these systems, by generalizing the recently developed covariant stochastic thermodynamics [49].

Our journey begins with the splitting of thermodynamic forces into a conservative part and a non-conservative part. Work and heat are both decomposed into a house-keeping part due to non-conservative force and an excess part due to conservative force. The house-keeping work is directly dissipated and absorbed by the heat bath (in the form of negative the house-keeping heat), whereas both the excess work and the excess heat are used to transform the system state. These definitions share the same spirit as those of Oono and Paniconi [17], but are made mathematically precise using the language of stochastic thermodynamics. Excess EP and house-keeping EP are also analogously defined. Thermodynamic quantities are defined both at the trajectory level and at the ensemble level. In the steady state, the excess parts of thermodynamic quantities vanish identically, whereas the house-keeping parts are generally nonvanishing. Differentials of thermodynamic quantities can be expressed in terms of only excess work and heat. These differential relations are very much analogous to those in equilibrium thermodynamics.

The splitting of thermodynamic forces can be carried out in an infinite many different ways, each defining a reference conservative system. This is very similar to the gauge symmetry of electromagnetism, hence we will use the term *gauge* throughout the paper. There is a special *Gibbs gauge*, where the NESS is characterized by Gibbs canonical distribution. In this gauge, the excess heat agrees with that defined by Hatano and Sasa. The excess EP agrees with that defined by Glansdorff and Prigogine, i.e., it is the second-order variation of the total EP near the NESS, and is always non-negative. By adopting the Gibbs gauge and focusing on the excess parts of thermodynamic quantities, a complete analogy between thermodynamics of non-conservative systems and that of conservative systems can be established. The NESS minimizes both the free energy and excess EP. For a nonautonomous process, the free energy difference constitutes a lower bound on the excess work that is needed to carry out a transition from one NESS to another. This is precisely the *principle of minimal excess work*, or the *second law of SST* [1], now established for most general nonlinear Langevin systems, arbitrary far from equilibrium. For over-damped Langevin systems with additive noises, these results are consistent with those of Refs. [31,35,36].

Using the detailed fluctuation theorem, we study the total EP and find two broad classes of behaviors. For systems with asymmetric kinetic matrices, i.e., *asymmetric systems*, the EP can be decomposed into a house-keeping part, an excess part, and an intriguing *pumped entropy*. Whilst both excess EP and the total EP are positive, the house-keeping EP and the pumped entropy may be either positive or negative. The pumped entropy describes entropy transfer between the system and the agent who supply the driving forces without necessarily involving heat dissipation, and is peculiar to systems with asymmetric kinetic matrices. Depending on the signs of pumped entropy, the system may behave either as a Maxwell demon, or an antidemon, transforming heat into mechanical energy, or the other way around. Fluctuation theorems and work identities are derived both for processes starting from thermal equilibrium and for processes starting from NESS. The former class of results concerns total work, and are gauge dependent. By continuously varying the gauge, we obtain an entire manifold of physically distinct fluctuation theorems and work identities. The latter class of result concerns only excess work and excess EP. All results are illustrated using multiple concrete examples.

For *symmetric systems*, i.e., systems with symmetric kinetic matrices, there is no entropy pumping, and the total EP can be split into a house-keeping part and an excess part. In the special Gibbs gauge, the NESS obeys Gibbs distribution, whilst both house-keeping EP and excess EP are positive definite. At the trajectory level, house-keeping work and excess work each obeys a fluctuation theorem and work identity. These fluctuation theorems involve adjoint process and adjoint backward process, which are explicitly constructed. The fluctuation theorem for the house-keeping work is a significant refinement of the well-known steady-state fluctuation theorem. Whilst these are essentially rewriting of earlier results [25,27–30] in unified notations and a slightly more general setting, we also supply multiple novel insights about fluctuation theorems in non-conservative systems, together with several concrete examples where various results can be tested.

The mission of the present paper is to synthesize and generalize many significant but seemingly disparate ideas and theories, so as to provide a unified theoretical framework for an important but heterogeneous field. Accordingly, we will take a pedagogical approach, and spend great volume in explaining physical concepts and deriving equations step

by step. Even though we made a great effort to simplify the mathematical derivations, they are still rather complicated, owing to the intrinsic difficulty of nonlinear Langevin dynamics. To help explaining the theory, we will also discuss many concrete model systems. The discussion of these model systems will however be brief. More detailed discussions of these example problems, together with numerical simulations, will be discussed in future publications. We believe that such a style, which is also shared to a less extent by two of the previous papers in the sequel [49,50], fits the nature of the issue we address. We are confident that the present paper constitutes a firm foundation for more detailed studies of many non-conservative Langevin systems, which will be carried out in future.

A few remarks about the applicability of our theory are in order. Firstly even though in Sec. III we assume that the system is in contact with a single heat bath, our theory can be adapted to systems coupled to multiple heat baths with different temperatures, as we demonstrate using a simple example in Sec. V C. Secondly, we always impose natural boundary conditions or periodic boundary conditions to the Fokker-Planck equation, so that all boundary terms vanish identically. There are many realistic systems that are commonly described by other types of boundary conditions, some of which lead to NESS. We believe however these boundary conditions can always be alternatively described in terms of suitable potential energy or non-conservative forces. Indeed, the example in Sec. V C shows that a temperature gradient imposed on two boundaries can be mimicked by non-conservative forces. It is also important to note that many systems, such as chemical reactions, have discrete variables, and hence are better described by master equations. If the system size is not too small, however, a continuous approximation may be appropriate, and our theory becomes applicable. Finally we emphasize that whilst we supply a unified and consistent theory for thermodynamics and stochastic thermodynamics of non-conservative Langevin systems, there are also alternative formulations of theories, see e.g., [31,51–53]. The relations and differences between these theories and ours await further study.

The content of this paper is organized as follows. In Sec. II we briefly review the covariant Langevin dynamics and stochastic thermodynamics developed in the previous papers in the sequel [48,49]. In Sec. III we generalize the theory to include non-conservative forces, and establish a full analogy with the theory of conservative systems. In Sec. V, we illustrate the theory using several examples of systems with asymmetric kinetic matrices. In Sec. VI we specialize to systems with symmetric kinetic matrices. In Sec. VII, we discuss several examples of systems with symmetric kinetic matrices. Finally, in Sec. VIII, we conclude the paper with projection of future researches. In the Appendix, we supply detailed derivation of the detailed fluctuation theorem for general nonlinear Langevin dynamics with non-conservative forces.

II. REVIEW OF PREVIOUS RESULTS

We first briefly review the covariant theories of Langevin dynamics and stochastic thermodynamics developed in previous papers [48–50]. Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be the set of slow variables, λ the control parameter, and $U(\boldsymbol{x}, \lambda)$ the generalized potential. For simplicity, we further assume that the metric tensor is trivial, i.e., $g_{ij} = \delta_{ij}$. Generalization to the case of nontrivial metric tensor is straightforward, as demonstrated in Ref. [49]. The system dynamics is described by the covariant nonlinear Ito-Langevin equation

$$dx^i + (L^{ij}\partial_j U - \partial_j L^{ij})dt = b^{i\alpha} dW_\alpha(t), \tag{2.1}$$

where $dW_\alpha(t)$ are the standard Wiener noises, satisfying $dW_\alpha(t)dW_\beta(t) = \delta_{\alpha\beta}$, $-\partial_j U$ are the thermodynamic forces [54], and $L^{ij}(\boldsymbol{x})$ are kinetic coefficients, both familiar in the classical theory of irreversible thermodynamics. The term $-\partial_j L^{ij}$ is called spurious drift and has been analyzed in detail in two previous papers [48,49]. Equation (2.1) is mathematically equivalent to the more familiar form of nonlinear Langevin equation: $dx^i = F^i dt + b^{i\alpha} dW_\alpha(t)$, yet has the advantages that U, L^{ij} both have clear physical meanings and transform in a simple way under nonlinear transformation of variables. The kinetic matrix $L^{ij}(\boldsymbol{x})$ can be decomposed as

$$L^{ij}(\boldsymbol{x}) = B^{ij}(\boldsymbol{x}) + Q^{ij}(\boldsymbol{x}), \tag{2.2a}$$

where $B^{ij}(\boldsymbol{x})$ and $Q^{ij}(\boldsymbol{x})$ are respectively symmetric and antisymmetric. The product $b^{i\alpha}(\boldsymbol{x})dW_\alpha$ in the r.h.s. of Eq. (2.1) is interpreted in Ito's sense. The amplitudes of noises $b^{i\alpha}(\boldsymbol{x})$ are related to $B^{ij}(\boldsymbol{x})$ via

$$b^{i\alpha} b^{j\alpha} = 2B^{ij} = L^{ij} + L^{ji}. \tag{2.2b}$$

For more detailed explanations, we refer to readers to Ref. [48]. This relation guarantees that the symmetric matrix B^{ij} is semipositive definite. We will however assume that B^{ij} is positive definite for convenience of analysis. Note that B^{ij}, Q^{ij}, U may all depend the control parameter λ. The Fokker-Planck equation (FPE) associated with the Langevin dynamics (2.1) is given by

$$\partial_t p(\boldsymbol{x}, t) = \partial_i L^{ij}(\partial_j + (\partial_j U))p(\boldsymbol{x}, t) = \mathscr{L} p, \tag{2.3}$$

where $\mathscr{L}$ is the Fokker-Planck operator

$$\mathscr{L} \equiv \partial_i L^{ij}(\partial_j + (\partial_j U)). \tag{2.4}$$

We further assume that the following detailed balance conditions [48,49] of B^{ij}, Q^{ij}, U are satisfied:

$$\varepsilon_i B^{ij}(\boldsymbol{x}^*, \lambda^*)\varepsilon_j = B^{ij}(\boldsymbol{x}, \lambda), \tag{2.5a}$$

$$\varepsilon_i Q^{ij}(\boldsymbol{x}^*, \lambda^*)\varepsilon_j = -Q^{ij}(\boldsymbol{x}, \lambda), \tag{2.5b}$$

$$U(\boldsymbol{x}^*, \lambda^*) = U(\boldsymbol{x}, \lambda), \tag{2.5c}$$

where $\boldsymbol{x}^*, \lambda^*$ are respectively the time reversals of $\boldsymbol{x}, \lambda$, and $\int_{\boldsymbol{x}}$ means integration over the space of $\boldsymbol{x}$, as in Refs. [49,50]. It is further assumed that U satisfies

$$\int_{\boldsymbol{x}} e^{-U(\boldsymbol{x})} = 1. \tag{2.5d}$$

The detailed balance conditions (2.5) guarantee that $p^{\rm EQ}(\boldsymbol{x}, \lambda) = e^{-U(\boldsymbol{x},\lambda)}$ can be interpreted as a proper thermal equilibrium state, and the dynamics satisfies the reversibility condition

$$P_\lambda(\boldsymbol{x}_1|\boldsymbol{x}; dt)p_{\rm EQ}(\boldsymbol{x}, \lambda) = P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}_1^*; dt)p_{\rm EQ}(\boldsymbol{x}_1^*, \lambda^*). \tag{2.6}$$

Since B^{ij} is positive definite, for fixed control parameter λ, the system converges to the unique thermal equilibrium state $p^{\rm EQ}(\boldsymbol{x}, \lambda)$.

Equation (2.6) implies that the probability density of a dynamic trajectory $\boldsymbol{x}(t)$ in the forward process (with control parameter λ) is the same as the probability density of the reversed trajectory $\boldsymbol{x}^*(-t)$ in the backward process with control parameter λ^*. This is the precise meaning of *detailed balance*, or *microscopic reversibility* according to Onsager [8–10].

Assuming that the system is in contact with a single heat bath, the generalized potential is related to the Hamiltonian of mean force $H(\boldsymbol{x}, \lambda)$ [50,55–59] via

$$U(\boldsymbol{x}, \lambda) = \beta(H(\boldsymbol{x}, \lambda) - F_{\rm EQ}(\lambda)), \tag{2.7}$$

where $F_{\rm EQ}(\lambda)$ is the *equilibrium free energy* of the system as a function of λ,

$$F_{\rm EQ}(\lambda) = -T \log \int_{\boldsymbol{x}} e^{-\beta H}. \tag{2.8}$$

The time-reversal symmetry (2.5c) is translated into

$$H(\boldsymbol{x}^*, \lambda^*) = H(\boldsymbol{x}, \lambda), \tag{2.9a}$$

$$F(\lambda^*) = F(\lambda). \tag{2.9b}$$

In a previous paper of this sequel [50], it was shown that the generalized potential U is uniquely determined by the Hamiltonian of mean force H but not vice versa. This is because U must satisfy the normalization Eq. (2.5d), and hence contains less information.

The Hamiltonian of mean force is defined as the fluctuating internal energy [49,50]. The nonequilibrium internal energy, entropy, and free energy of the system are defined respectively as

$$E[p] \equiv \int_{\boldsymbol{x}} p(\boldsymbol{x}) H(\boldsymbol{x}, \lambda), \tag{2.10a}$$

$$S[p] \equiv -\int_{\boldsymbol{x}} p(\boldsymbol{x}) \log p(\boldsymbol{x}), \tag{2.10b}$$

$$F[p] \equiv E - TS = \int_{\boldsymbol{x}} p(\boldsymbol{x})(H + T \log p(\boldsymbol{x})). \tag{2.10c}$$

Note that $k_B = 1$, and hence entropy is dimensionless.

Heat and work at trajectory level are defined as

$$đ\mathscr{Q} \equiv d_{\boldsymbol{x}} H(\boldsymbol{x}, \lambda), \tag{2.11a}$$

$$đ\mathscr{W} \equiv d_{\lambda} H(\boldsymbol{x}, \lambda), \tag{2.11b}$$

where $d_{\boldsymbol{x}} H$ and $d_{\lambda} H$ are respectively the differentials of H due to changes of $\boldsymbol{x}$ and of λ. The first law holds at the trajectory level,

$$dH(\boldsymbol{x}, \lambda) = đ\mathscr{Q} + đ\mathscr{W}. \tag{2.11c}$$

Heat and work at the ensemble level are defined as ensemble averages of the corresponding quantities at trajectory level, defined in Eqs. (2.11). In Langevin dynamics, ensemble average means averaging both over noises and over pdf $p(\boldsymbol{x}, t)$. It was proved in Ref. [49] that

$$đQ = \int_{\boldsymbol{x}} \langle d_{\boldsymbol{x}} H \rangle \, p = \int_{\boldsymbol{x}} H \mathscr{L} p \, dt, \tag{2.12a}$$

$$đW = \int_{\boldsymbol{x}} p \, d_{\lambda} H, \tag{2.12b}$$

where $\langle d_{\boldsymbol{x}} H \rangle$ means average of $d_{\boldsymbol{x}} H$ over noises, and $\mathscr{L}$ is defined in Eq. (2.4). The first law also holds at the ensemble level

$$dE = đQ + đW. \tag{2.13}$$

The differential of the nonequilibrium free energy is

$$dF = đW + đQ - T dS. \tag{2.14}$$

It was shown in Ref. [50] that $-\beta đQ$ is the change of conditional entropy of the heat bath given the system state. Hence $dS - \beta đQ$ is the change of the joint entropy of the system and the bath, i.e., the change of total entropy of the universe. The differential of the total entropy is

$$\begin{aligned} dS^{\rm tot} &= dS - \beta đQ = -\int_{\boldsymbol{x}} (\log p + U) dp \\ &= \beta(đW - dF). \end{aligned} \tag{2.15}$$

The rate of total EP can be calculated using Eq. (2.3),

$$\begin{aligned} \Sigma^{\rm tot} &\equiv \frac{dS^{\rm tot}}{dt} = \frac{dS}{dt} - \frac{\beta đQ}{dt} = \beta\left(\frac{đW}{dt} - \frac{dF}{dt}\right) \\ &= \int_{\boldsymbol{x}} [(\partial_i + \partial_i U)p] \frac{B^{ij}}{p} [(\partial_j + \partial_j U)p] \geqslant 0, \end{aligned} \tag{2.16}$$

which is non-negative. The last inequality follows form the positivity of the symmetric matrix B^{ij}.

If λ is fixed, $đW/dt = 0$, Eq. (2.16) says that the rate of EP is always non-negative, and vanishes when the system achieves equilibrium. Hence the equilibrium state is characterized by minimization of EP rate. Additionally, Eq. (2.16) also says that the free energy decreases monotonically as the system evolves, until it achieves equilibrium. Hence the equilibrium state is characterized both by minimization of free energy and by minimization of EP rate. If λ is not fixed, Eq. (2.16) says $đW \geqslant dF$, which imposes a lower bound on the work needed to realize a nonequilibrium process. This result is called the principle of minimal work, and may be taken as an alternative but equivalent representation of the second law.

The probability current is defined as [48]

$$j^i = -L^{ij}(\partial_j + (\partial_j U))p + \partial_j (Q^{ij} p), \tag{2.17}$$

such that the FPE (2.3) can be rewritten into the form of current conservation

$$\partial_t p + \partial_k j^k = 0. \tag{2.18}$$

The probability current can be decomposed into a reversible part $j^i_{\rm R}$ and an irreversible part j^i_{IR} [60,61],

$$j^i = j^i_{\rm R} + j^i_{IR}, \tag{2.19a}$$

$$j^i_{IR} = -B^{ij}(\partial_j + (\partial_j U))p, \tag{2.19b}$$

$$j^i_{\rm R} = -Q^{ij}(\partial_j + (\partial_j U))p + \partial_j (Q^{ij} p), \tag{2.19c}$$

where j^i_{IR} only involves B^{ij}, and $j^i_{\rm R}$ only involves Q^{ij}. If the matrix B^{ij} is nonsingular, we can rewrite Eq. (2.16) in terms

of irreversible probability current,

$$\Sigma^{\rm tot} = \int_{\boldsymbol{x}} j^i_{\rm IR} B^{-1}_{ij} p^{-1} j^j_{\rm IR}, \tag{2.20}$$

where B^{-1}_{ij} is the inverse matrix of B^{ij}.

Let $P_\lambda(\boldsymbol{x}_1|\boldsymbol{x}; dt)$ be the probability density that the system starts from $\boldsymbol{x}$ and evolves to $\boldsymbol{x}_1 = \boldsymbol{x} + d\boldsymbol{x}$ after time dt, with the control parameter fixed at λ. Corresponding to this infinitesimal *forward step* of trajectory, there is a time-reversed infinitesimal step, *the backward step*, where the system starts from $\boldsymbol{x}^*_1$ and evolves to $\boldsymbol{x}^*$ after time dt, with the control parameter fixed at λ^*. The corresponding transition probability density is $P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}^*_1; dt)$. The ratio of $P_\lambda(\boldsymbol{x}_1|\boldsymbol{x}; dt)$ and $P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}^*_1; dt)$ is related to the heat $đ\mathcal{Q}$ absorbed by the system during the forward step via the following formula:

$$\log \frac{P_\lambda(\boldsymbol{x}_1|\boldsymbol{x}; dt)}{P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}^*_1; dt)} = -\beta đ\mathcal{Q} = -d_{\boldsymbol{x}} U. \tag{2.21a}$$

This identity was proved in the Appendix of Ref. [49] for the Langevin dynamics Eq. (2.1) satisfying detailed balance (2.5). Since heat is the energy transfer from the heat bath to the system, and the heat bath is assumed in equilibrium, $-\beta đ\mathcal{Q} = \beta dE_B = d\mathcal{S}_B$ can be understood as the entropy change of the bath during the infinitesimal forward process, where $\mathcal{S}_B$ is the bath entropy conditioned on the system state. Hence Eq. (2.21) can also be rewritten as

$$\log \frac{P_\lambda(\boldsymbol{x}_1|\boldsymbol{x}; dt)}{P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}^*_1; dt)} = d\mathcal{S}_B. \tag{2.21b}$$

Equations (2.21) supply the key connection between EP and time-reversal asymmetry of path probability density, and play a fundamental role in the development of stochastic thermodynamics.

Using Eq. (2.21), both Crooks fluctuation theorem and Jarzynski equality can be established, from which follows the second law inequality (2.16). For details, see Sec. IV of Ref. [49].

III. STOCHASTIC THERMODYNAMICS OF NON-CONSERVATIVE LANGEVIN SYSTEMS

A. Non-conservative thermodynamic forces

The thermodynamic force $-\partial_j U$ in Langevin equation (2.1) is said to be *conservative*, because it is negative the gradient of a generalized potential. We introduce a non-conservative component φ_j of thermodynamic force such that the total force becomes $-\partial_j U + \varphi_j$. The non-conservative force φ_i cannot be written as gradient of a potential. This may be due to two possible reasons: (i) $\partial_i \varphi_j \neq \partial_j \varphi_i$, or (ii) $\partial_i \varphi_j = \partial_j \varphi_i$ but there is a close loop in the $\boldsymbol{x}$ space along which $\oint d\varphi_i \neq 0$. The later arises only if the $\boldsymbol{x}$ space is multiply connected. In the presence of non-conservative force, Eqs. (2.1), (2.3), and (2.4) become

$$dx^i + (L^{ij}(\partial_j U - \varphi_j) - \partial_j L^{ij})dt = b^{i\alpha} dW_\alpha, \tag{3.1}$$

$$\partial_t p = \partial_i L^{ij}(\partial_j + (\partial_j U - \varphi_j))p, \tag{3.2}$$

$$\mathscr{L} \equiv \partial_i L^{ij}(\partial_j + (\partial_j U - \varphi_j)). \tag{3.3}$$

We assume that the non-conservative force φ_i is generated by some reversible mechanism, same as the potential generalized U. Hence we expect that the non-conservative force φ_i satisfies time-reversal symmetry similar to the conservative force $-\partial_i U$. However, since $\partial_i U$ and φ_i appear in the dynamics only through the combination $-\partial_i U + \varphi_i$, the detailed balance conditions should also be imposed on the combination. Hence the *local detailed balance* conditions for non-conservative Langevin systems are

$$\varepsilon_i B^{ij}(\boldsymbol{x}^*, \lambda^*)\varepsilon_j = B^{ij}(\boldsymbol{x}, \lambda), \tag{3.4a}$$

$$\varepsilon_i Q^{ij}(\boldsymbol{x}^*, \lambda^*)\varepsilon_j = -Q^{ij}(\boldsymbol{x}, \lambda), \tag{3.4b}$$

$$\epsilon_i(\partial^*_i U(\boldsymbol{x}^*, \lambda^*) - \varphi_i(\boldsymbol{x}^*, \lambda^*)) = \partial_i U(\boldsymbol{x}, \lambda) - \varphi_i(\boldsymbol{x}, \lambda), \tag{3.4c}$$

$$\int_{\boldsymbol{x}} e^{-U(\boldsymbol{x})} = 1. \tag{3.4d}$$

In the above equations, we use λ to denote all control parameters, which may appear both in U, L^{ij} as well as in φ_i. In the absence of non-conservative force φ_i, Eq. (3.4c) may be integrated to yield Eq. (2.5c). Equations (3.4) are called *local detailed balance conditions* because in the presence of non-conservative force φ_i, we do not expect the existence of an equilibrium state $p_{\rm EQ}$ such that Eq. (2.6) holds. Instead the system converges to a NESS with positive EP, whose study is the central task of the present paper.

The Langevin dynamics (2.1), which is obtained from Eq. (3.1) by setting $\varphi_i = 0$, will be referred to as the *reference conservative Langevin system*. The decomposition of the total thermodynamic force $-\partial_i U + \varphi_i$ into a conservative force $-\partial_i U$ and a non-conservative force φ_i is not unique. We can make the following *gauge transformation* using an arbitrary function $\psi(\boldsymbol{x})$:

$$U \to U + \psi, \tag{3.5a}$$

$$\varphi_i \to \varphi_i + \partial_i \psi, \tag{3.5b}$$

which leaves the total force $-\partial_i U + \varphi_i$, and hence also leaves the Langevin equation (3.1) invariant. (Strictly speaking ψ is not completely arbitrary, because $e^{-U-\psi}$ should also be normalized. This subtlety, however, has no influence on our discussion below.) The essence of this gauge invariance is the arbitrariness of the reference conservative Langevin system. We will discuss this gauge invariance in more detail in Sec. III K.

In some situations, for the sake of simplicity, we prefer to impose time-reversal symmetry on U and on φ_i separately. Equation (3.4c) then should be replaced by the following two conditions:

$$U(\boldsymbol{x}^*, \lambda^*) = U(\boldsymbol{x}, \lambda), \tag{3.6a}$$

$$\epsilon_i \varphi_i(\boldsymbol{x}^*, \lambda^*) = \varphi_i(\boldsymbol{x}, \lambda). \tag{3.6b}$$

It is however important to remember that we are allowed to split the thermodynamic force in a way such that Eq. (3.4c) is valid but Eqs. (3.6) are not.

It is important to note that for a given physical system, it may not be always clear how to define time reversal λ^* of the control parameter. In Sec. VII C, we address a problem where there are two possible ways to impose the time-reversal symmetry on the non-conservative forces. In one way, Eqs. (3.6)

are satisfied, hence the theory has local detailed balance, and the theory constructed in this paper is applicable. Yet in the other way, Eqs. (3.6) are not satisfied, the local detailed balance is lost, and the theory constructed in this paper is not applicable. A general theory of stochastic thermodynamics for non-conservative Langevin systems without local detailed balance is much more complicated, and shall not be addressed in this paper. For some basic aspects of these systems, see Ref. [42].

There seems a common misconception that detailed balance property is sufficient to guarantee the existence of thermal equilibrium. Our discussion above indicates that local detailed balance conditions are not sufficient to guarantee that the system converges to thermal equilibrium. A simple counter-example is a colloid driven by a constant force, which settles down to a NESS with constant velocity and constant entropy production. The Langevin equation clearly has detailed balance. Yet the constant force cannot be expressed as the negative gradient of potential satisfying Eq. (3.4d), hence must be deemed as non-conservative. More interesting examples are given by systems with entropy-pumping, which constantly exchange entropy with their environment. The theory of entropy pumping will be developed in Sec. III H and Sec. III I. Examples of entropy pumping will be given by Sec. V A and Sec. V B. The stationary states in these systems must be understood as nonequilibrium steady state, even though the rate of EP may be (arbitrarily close to) zero.

The probability current and its reversible and irreversible components are obtained from Eqs. (2.19) by replacing $-\partial_j U$ with the total thermodynamic force $-\partial_j U + \varphi_j$:

$$j^i = j^i_{\rm R} + j^i_{IR}, \tag{3.7a}$$

$$j^i_{IR} = -B^{ij}(\partial_j + (\partial_j U) - \varphi_j)p, \tag{3.7b}$$

$$j^i_{\rm R} = -Q^{ij}(\partial_j + (\partial_j U) - \varphi_j)p + \partial_j(Q^{ij}p). \tag{3.7c}$$

The FPE (3.2) can be rewritten into the form of current conservation,

$$\partial_t p + \partial_k j^k = 0. \tag{3.8}$$

Finally, we note that Eqs. (3.1) and (3.2) are both covariant under nonlinear transformation of variables. The transformation rules of L^{ij} and U were discussed in preceding papers [48,49]. The non-conservative force φ_i transform as a covariant vector, same as $\partial_i U$.

B. Alternative parameterizations

The steady state pdf of FPE (3.2) can be written into the following form:

$$p^{\rm SS}(\boldsymbol{x};\lambda) = e^{-U^{\rm G}(\boldsymbol{x};\lambda)}. \tag{3.9}$$

We define the *Gibbs gauge* such that the generalized potential is $U^{\rm G}(\boldsymbol{x};\lambda)$ and the non-conservative force is $\varphi^{\rm G}_i(\boldsymbol{x},\lambda)$, which is determined by the following condition:

$$\partial_i U^{\rm G}(\boldsymbol{x},\lambda) - \varphi^{\rm G}_i(\boldsymbol{x},\lambda) = \partial_i U(\boldsymbol{x},\lambda) - \varphi_i(\boldsymbol{x},\lambda). \tag{3.10}$$

The FPE (3.2) can then be rewritten as

$$\partial_t p = \partial_i L^{ij}\big(\partial_j + \big(\partial_j U^{\rm G} - \varphi^{\rm G}_j\big)\big)p, \tag{3.11}$$

The Hamiltonian of mean force can be analogously defined in the presence of non-conservative force, which leads to [the counterpart of Eq. (2.7)]

$$U^{\rm G}(\boldsymbol{x},\lambda) = \beta H^{\rm G} - \beta F_{\rm SS}, \tag{3.12}$$

where $F_{\rm SS}$ is *the free energy of the steady state*,

$$F_{\rm SS} = -T\log\int_{\boldsymbol{x}} e^{-\beta H^{\rm G}}. \tag{3.13}$$

Hence, Eq. (3.9) can also be written in the Gibbs form

$$p^{\rm SS}(\boldsymbol{x};\lambda) = e^{-\beta H^{\rm G}+\beta F_{\rm SS}}, \tag{3.14}$$

We use the name *Gibbs gauge* because within this gauge, the NESS pdf, Eq. (3.9) or (3.14), has a Gibbs form. This will lead to a complete analogy between the thermodynamic theory of NESS and that of usual equilibrium states. The NESS pdf Eq. (3.9) must satisfy the FPE (3.11) with the left-hand side (l.h.s.) vanishing. This leads to the *Gibbs gauge condition*

$$\partial_i\big(L^{ij}e^{-U^{\rm G}}\varphi^{\rm G}_j\big) = 0, \tag{3.15}$$

which, together with Eq. (3.10), completely determines U^G and φ^G_i.

In the steady state, the probability current is divergenceless, and hence it is possible (here we assume that $\boldsymbol{x}$ space is multidimensional) to parametrize it as

$$j^i_{\rm SS} = \partial_j(\hat{Q}^{ij}e^{-U^{\rm G}}), \tag{3.16}$$

where $\hat{Q}^{ij}$ is another antisymmetric matrix. Following the discussion in Ref. [48], we may define

$$\hat{L}^{ij} = B^{ij} + \hat{Q}^{ij}, \tag{3.17}$$

and rewrite Eqs. (3.1) and (3.2) into

$$dx^i + (\hat{L}^{ij}(\partial_j U^{\rm G}) - \partial_j\hat{L}^{ij})dt = b^{i\alpha}dW_\alpha, \tag{3.18}$$

$$\partial_t p = \partial_i\hat{L}^{ij}(\partial_j + (\partial_j U^{\rm G}))p. \tag{3.19}$$

Using these notations, the probability current is

$$j^i = -\hat{L}^{ij}(\partial_j + (\partial_j U^G))p + \partial_j(\hat{Q}^{ij}p), \tag{3.20}$$

Equations (3.18), (3.19), and (3.20) must be equivalent to Eqs. (3.1), (3.2), and (3.7) with U, φ_i replaced by U^G, φ^R_i. In particular, comparing Eqs. (3.18) and (3.1), we find the following relation between $\hat{Q}^{ij}$ and Q^{ij}:

$$(\hat{Q}^{ij} - Q^{ij})\partial_j U^G - \partial_j(\hat{Q}^{ij} - Q^{ij}) = -L^{ij}\varphi_j. \tag{3.21}$$

As pointed out in Ref. [48], the antisymmetric matrices $\hat{Q}$ and Q are not uniquely determined by the Langevin equation.

If the system has odd variables and parameters, the NESS pdf generally is not invariant under time reversal, which means

$$U^{\rm G}(\boldsymbol{x};\lambda) \neq U^{\rm G}(\boldsymbol{x}^*;\lambda^*), \tag{3.22}$$

$$H^{\rm G}(\boldsymbol{x};\lambda) \neq G^{\rm G}(\boldsymbol{x}^*;\lambda^*), \tag{3.23}$$

$$F_{\rm SS}(\lambda) \neq F_{\rm SS}(\lambda^*). \tag{3.24}$$

Additionally, the matrix $\hat{Q}^{ij}$ generally does not have the same symmetry as Q^{ij},

$$\varepsilon_i \hat{Q}^{ij}(\boldsymbol{x}^*, \lambda^*)\varepsilon_j \neq -\hat{Q}^{ij}(\boldsymbol{x}, \lambda).. \tag{3.25}$$

These (not very nice) features should be kept in mind.

C. Work and heat at trajectory level

For now we assume that the system is in contact with a single heat bath. We shall use the reference conservative system to define various thermodynamic potentials of the non-conservative system. Hence the HMF $H(\boldsymbol{x}, \lambda)$ of the reference conservative system is defined as the fluctuating internal energy, which is related to $U(\boldsymbol{x}, \lambda)$ via Eq. (2.7). The internal energy, entropy, and free energy are defined as Eqs. (2.10). Note that these definitions depend on the choice of gauge.

We define the work at the trajectory level as

$$đ\mathscr{W} \equiv d_\lambda H + f_i \circ dx^i, \tag{3.26}$$

where $\circ$ is the product in Stratonovich's sense,

$$\begin{aligned} f_i(\boldsymbol{x}) \circ dx^i &\equiv f_i(\boldsymbol{x} + d\boldsymbol{x}/2)dx^i, \\ &= f_i(\boldsymbol{x})dx^i + \tfrac{1}{2}(\partial_j f_i)dx^i dx^j, \end{aligned} \tag{3.27}$$

and f_i is related to φ_i via

$$f_i = T\,\varphi_i. \tag{3.28}$$

Demanding the first law of thermodynamics at the trajectory level

$$dH = đ\mathscr{W} + đ\mathscr{Q}, \tag{3.29}$$

the heat at the trajectory level should be defined as

$$đ\mathscr{Q} \equiv d_{\boldsymbol{x}} H - f_i \circ dx^i. \tag{3.30}$$

In the absence of non-conservative forces, these definitions reduce to those of conservative systems, see Eq. (2.11a). If f_i can be written as the gradient of a potential ψ, we can write $đ\mathscr{Q} = d_{\boldsymbol{x}}(H - \psi)$, which is also of the form Eq. (2.11a).

D. Justification of definitions of heat and work

To justify our definitions of work and heat in the presence of non-conservative force, we consider a Hamiltonian system interacting with a heat bath, which is also modeled as a Hamiltonian system. The total Hamiltonian is decomposed into a system part H and a bath part H_{B},

$$H^{\mathrm{tot}} = H(\boldsymbol{x}; \lambda(t)) + H_{\mathrm{B}}(\boldsymbol{x}, \boldsymbol{y}), \tag{3.31}$$

where $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p})$ are the system variables and $\boldsymbol{y} = (\boldsymbol{Q}, \boldsymbol{P})$ are the bath variables. The system is additionally driven by non-conservative force $\boldsymbol{f}$, which cannot be derived from a potential energy. The dynamics of the joint system is described by the Hamiltonian equations

$$\dot{q}^i = \frac{\partial H^{\mathrm{tot}}}{\partial p_i}, \tag{3.32a}$$

$$\dot{p}_i = -\frac{\partial H^{\mathrm{tot}}}{\partial q^i} + f_i, \tag{3.32b}$$

$$\dot{Q}^a = \frac{\partial H^{\mathrm{tot}}}{\partial P_a}, \tag{3.32c}$$

$$\dot{P}_a = -\frac{\partial H^{\mathrm{tot}}}{\partial Q^a}. \tag{3.32d}$$

The work should be defined as the differential of total energy. Using the Hamiltonian equations, we find

$$đ\mathscr{W} \equiv dH^{\mathrm{tot}} = d_\lambda H + f_i dq^i, \tag{3.33}$$

which agrees with our definition Eq. (3.26), since, in the absence of noises, $f_i dq^i$ can be interpreted either as a Stratonovich product or as an Ito product. Now consider the change of system energy,

$$\begin{aligned} dH &= \frac{\partial H}{\partial q^i}dq^i + \frac{\partial H}{\partial p_i}dp_i + \frac{\partial H}{\partial \lambda}d\lambda \\ &= d_{\boldsymbol{x}} H + d_\lambda H, \\ &= d_{\boldsymbol{x}} H - f_i dq^i + d_\lambda H + f_i dq^i. \end{aligned} \tag{3.34}$$

Since the last two terms in right-hand side (r.h.s.) is the work, the first two terms must be the heat, which again agrees with our definition of heat, Eq. (3.30). Equation (3.34) then becomes the first law at the trajectory level,

$$dH = đ\mathscr{Q} + đ\mathscr{W}. \tag{3.35}$$

Subtracting Eq. (3.35) from Eq. (3.33), we obtain

$$d(H^{\mathrm{tot}} - H) = dH_{\mathrm{B}} = -đ\mathscr{Q}. \tag{3.36}$$

Hence heat is negative the energy change of the bath.

To provide further justification for our definitions of work and heat, we consider the unitary limit of the Langevin dynamics where $B^{ij} = 0$. This may be, for example, realized by switching off the coupling between the system and the heat bath. For simplicity, we further assume $\partial_j Q^{ij} = 0$. (This condition is satisfied by all examples we discussed below.) The Langevin equation (3.1) then reduces to

$$dx^i + Q^{ij}(\partial_j U - f_j)dt = 0. \tag{3.37}$$

The differential heat and work can be calculated by using Eq. (3.37) in Eqs. (3.26) and (3.30). Noticing that noise variances vanish and hence the Stratonovich product $f_i \circ dx^i$ can be replaced by usual product $f_i dx^i$, we find that the heat vanishes identically,

$$đ\mathscr{Q} = 0, \tag{3.38}$$

$$đ\mathscr{W} = dH = d_\lambda H + f_i dx^i. \tag{3.39}$$

These results are of course completely expected.

E. Work and heat at ensemble level

Heat and work at ensemble level are defined as averages of the corresponding quantities at trajectory level,

$$đW \equiv \langle\langle đ\mathscr{W} \rangle\rangle = \langle\langle d_\lambda H + f_i \circ dx^i \rangle\rangle, \tag{3.40a}$$

$$đQ \equiv \langle\langle đ\mathscr{Q} \rangle\rangle = \langle\langle d_{\boldsymbol{x}} H - f_i \circ dx^i \rangle\rangle, \tag{3.40b}$$

where the symbol $\langle\langle \cdot \rangle\rangle$ means average both over the Wiener noises dW_α and over the probability distribution $p(\boldsymbol{x}, t)$. When calculating these averages, formula (3.27) should be used for Stratonovich product, $d_{\boldsymbol{x}} H$ should be expanded up to the second order in $d\boldsymbol{x}$, and the Ito-Langevin equation (3.1) should be

used for dx^i. Additionally Ito's formulas [49] must be applied for $dx^i dx^j$,

$$dx^i dx^j = b^{i\alpha} b_{j\beta} dW_\alpha dW_\beta + O(dt^{3/2}) = 2B^{ij} dt + O(dt^{3/2}), \tag{3.41}$$

where ignored terms do not contribute to the continuous time limit. Finally, it should be noted that Wiener noises $dW_\alpha(t)$ are independent of any function of $\boldsymbol{x}(t)$. Carrying out all these steps, it can be shown that for arbitrary function $\psi(\boldsymbol{x})$, we have

$$\langle\langle d_{\boldsymbol{x}} \psi \rangle\rangle = \int_{\boldsymbol{x}} \psi \mathscr{L} p. \tag{3.42}$$

Calculating the averages in Eqs. (3.53), we obtain work and heat at ensemble level,

$$đW = dt \int_{\boldsymbol{x}} [-f_i L^{ij} (\partial_j + \partial_j U - \varphi_j) + Q^{ij} (\partial_i f_j)] p + \int_{\boldsymbol{x}} (d_\lambda H) p, \tag{3.43a}$$

$$đQ = -dt \int_{\boldsymbol{x}} (\partial_i H - f_i) L^{ij} (\partial_j + \partial_j U - \varphi_j) p - dt \int_{\boldsymbol{x}} Q^{ij} (\partial_i f_j) p. \tag{3.43b}$$

Let us now look at differential of internal energy, defined in Eq. (2.10a). The change of Eq. (2.10a) are due to two sources: change of parameter λ, and dynamic evolution of $p(\boldsymbol{x}, \lambda)$ as described by FPE (3.2). We obtain

$$dE = d \int_{\boldsymbol{x}} H p = \int_{\boldsymbol{x}} (p\, d_\lambda H + H \mathscr{L} p\, dt). \tag{3.44}$$

But we may also understand dE as the ensemble average of Eq. (3.29),

$$dE = \langle\langle dH \rangle\rangle = \langle\langle đ\mathscr{W} \rangle\rangle + \langle\langle đ\mathscr{Q} \rangle\rangle = đW + đQ, \tag{3.45}$$

with $đW$ and $đQ$ given by Eqs. (3.53). Equations (3.44) and (3.45) are of course equivalent, as we can directly verify.

Now recall that entropy and free energy are defined in Eqs. (2.10). Taking the differential of Eq. (2.10c), and using the first law at the ensemble level, Eq. (3.45), we obtain the differential of free energy,

$$dF = đW + đQ - T\, dS, \tag{3.46}$$

which may be rewritten into

$$dS - \beta đQ = \beta(đW - dF). \tag{3.47}$$

The l.h.s. is the change of the joint entropy of the system and the heat bath. In classical thermodynamics, it has the interpretation of total EP, and is positive definite. In stochastic thermodynamics with conservative force, Eq. (3.47) is indeed positive, see Eq. (2.16). We will see that in the presence of non-conservative force, things become more complicated.

The differential of system entropy is

$$dS = -d \int_{\boldsymbol{x}} p \log p = -dt \int_{\boldsymbol{x}} (\log p) \mathscr{L} p. \tag{3.48}$$

Combining this with Eq. (3.43b) we obtain

$$dS - \beta đQ = dS + dS_B = dt \int_{\boldsymbol{x}} (\partial_i + (\partial_i U) - \varphi_i) p \frac{B^{ij}}{p} (\partial_j + (\partial_j U) - \varphi_j) p + dt \int_{\boldsymbol{x}} p\, Q^{ij} (\partial_i \varphi_j). \tag{3.49}$$

Whilst the first term in r.h.s. is positive definite, the second term $dt \int_{\boldsymbol{x}} p\, Q^{ij} (\partial_i \varphi_j)$, is not. Unless $Q^{ij}(\partial_i \varphi_j) = 0$, Eq. (3.49) is generally not positive definite, and hence cannot be interpreted as EP. The physical meaning of the second term in r.h.s. of Eq. (3.49) will be discussed in Sec. III H.

F. House-keeping and excess, and analogy with conservative systems

It is convenient to separate the work at the trajectory level into a house-keeping part and an excess part,

$$đ\mathscr{W}^{\rm hk} \equiv f_i \circ dx^i, \tag{3.50a}$$

$$đ\mathscr{W}^{\rm ex} \equiv d_\lambda H = (\partial_\lambda H)\, d\lambda, \tag{3.50b}$$

$$đ\mathscr{W} = đ\mathscr{W}^{\rm hk} + đ\mathscr{W}^{\rm ex}, \tag{3.50c}$$

where the superscript "hk" and "ex" mean respectively "house-keeping" and "excess". The heat can be decomposed in a similar fashion,

$$đ\mathscr{Q}^{\rm hk} \equiv -f_i \circ dx^i = -đ\mathscr{W}^{\rm hk}, \tag{3.51a}$$

$$đ\mathscr{Q}^{\rm ex} \equiv d_{\boldsymbol{x}} H, \tag{3.51b}$$

$$đ\mathscr{Q} = đ\mathscr{Q}^{\rm hk} + đ\mathscr{Q}^{\rm ex}. \tag{3.51c}$$

Note that the excess work (3.50b) and excess heat (3.51b) only involve the HMF $H(\boldsymbol{x}, \lambda)$, and are formally identical to those for conservative systems, see Eqs. (2.11). By contrast, the house-keeping work (3.50a) and heat (3.51a) only involve the non-conservative force, and are equal in magnitude but opposite in sign. The first law Eq. (3.29) can thus be rewritten into an alternative form

$$dH = đ\mathscr{W}^{\rm ex} + đ\mathscr{Q}^{\rm ex}, \tag{3.52}$$

which involves only the excess work and heat. Qualitatively speaking, the house-keeping work is entirely dissipated to the environment in the form of (negative) the house-keeping heat, whereas both the excess work $d_\lambda H$ and the excess heat $d_{\boldsymbol{x}} H$ are spent on transforming the system state. The usage of these terms are in spirit as those in previous papers on steady-state thermodynamics [1,17,25].

We must remember however that the splitting of work and heat is gauge dependent. In another word, a gauge transformation, as given by (3.5), leads to a different splitting. In laboratory, each experimental setup prescribes a particular control scheme of conservative and non-conservative forces, and therefore defines a particular gauge. This experimentally chosen gauge, however, may not be the most convenient one for theoretical study.

House-keeping and excess heat and work at ensemble level are defined as the ensemble averages of the corresponding quantities at the trajectory level,

$$đW^{\rm hk} \equiv \langle\langle đ\mathscr{W}^{\rm hk} \rangle\rangle, \tag{3.53a}$$

$$đQ^{\rm hk} \equiv \langle\langle đ\mathscr{Q}^{\rm hk}\rangle\rangle = -đW^{\rm hk}, \tag{3.53b}$$

$$đW^{\rm ex} \equiv \langle\langle đ\mathscr{W}^{\rm ex}\rangle\rangle, \tag{3.53c}$$

$$đQ^{\rm ex} \equiv \langle\langle đ\mathscr{Q}^{\rm ex}\rangle\rangle, \tag{3.53d}$$

$$đW = đW^{\rm hk} + đW^{\rm ex}, \tag{3.53e}$$

$$đQ = đQ^{\rm hk} + đQ^{\rm ex}. \tag{3.53f}$$

Carrying out these ensemble averages, we find the house-keeping work and heat,

$$đW^{\rm hk} = dt \int_{\boldsymbol{x}} [-f_i L^{ij}(\partial_j + \partial_j U - \varphi_j) + Q^{ij}(\partial_i f_j)]p, \tag{3.54a}$$

$$đQ^{\rm hk} = -đW^{\rm hk}, \tag{3.54b}$$

which are again equal in magnitude and opposite in sign. The excess work and heat at ensemble level are

$$đW^{\rm ex} = \int_{\boldsymbol{x}} (d_\lambda H)p, \tag{3.55a}$$

$$đQ^{\rm ex} = -dt \int_{\boldsymbol{x}} (\partial_i H)L^{ij}(\partial_j + \partial_j U - \varphi_j)p = dt \int_{\boldsymbol{x}} H\,\mathscr{L}p. \tag{3.55b}$$

Note that Eqs. (3.55a) and (3.55b), are formally identical to those of conservative systems, see Eqs. (2.12). One must remember however the Fokker-Planck operators are different for conservative and non-conservative systems.

Using $đQ^{\rm hk} + đW^{\rm hk} = 0$, we can rewrite Eqs. (3.45) and (3.46) as

$$dE = đW^{\rm ex} + đQ^{\rm ex}, \tag{3.56}$$

$$dF = đW^{\rm ex} + đQ^{\rm ex} - T\,dS, \tag{3.57}$$

which again indicates that only the excess work and heat are used to transform the system state, whereas the house-keeping work as being directly dissipated and absorbed by the heat bath. In the steady state, both λ and $p(\boldsymbol{x}, t)$ remain fixed. Hence according to Eqs. (3.55), $đW^{\rm ex}$ and $đQ^{\rm ex}$ both vanish identically. Equation (3.44) was formulated first by Hatano and Sasa in Ref. [25] for over-damped Langevin dynamics with additive noises.

Let us define the *house-keeping EP* and *excess EP* as

$$dS^{\rm hk} \equiv -\beta đQ^{\rm hk} = \beta đW^{\rm hk}, \tag{3.58}$$

$$dS^{\rm ex} \equiv dS - \beta đQ^{\rm ex} = dS - \langle\langle d_{\boldsymbol{x}} U\rangle\rangle. \tag{3.59}$$

Again $dS^{\rm hk}$ may be understood as the entropy change due to heat directly dissipated into the heat bath, whereas $dS^{\rm ex}$ may be understood as the entropy change due to transform of system state. Combining these with Eq. (3.53f) we obtain

$$dS - \beta đQ = dS^{\rm hk} + dS^{\rm ex}. \tag{3.60}$$

Using Eq. (3.57) and (3.59) we also have

$$dS^{\rm ex} = \beta(đW^{\rm ex} - dF). \tag{3.61}$$

Combining Eqs. (3.58) and (3.59) with Eqs. (3.54), (3.55), and (3.48), we obtain the following expressions for house-keeping EP and excess EP:

$$\Sigma^{\rm hk} \equiv \frac{dS^{\rm hk}}{dt} = \int_{\boldsymbol{x}} [-\varphi_i L^{ij}(\partial_j + \partial_j U - \varphi_j) + Q^{ij}(\partial_i \varphi_j)]p, \tag{3.62a}$$

$$\Sigma^{\rm ex} \equiv \frac{dS^{\rm ex}}{dt} = \int_{\boldsymbol{x}} [(\partial_i + (\partial_i U))p]\frac{L^{ij}}{p}[(\partial_j + (\partial_j U) - \varphi_j)p]. \tag{3.62b}$$

Thermodynamic relations (3.52), (3.56), (3.57), and (3.61) for non-conservative systems have identical structure as Eqs. (2.11c), (2.13), (2.14), and (2.15) for conservative systems. A comparison of these relations, which are summarized in the first five rows of Table I, allows us to establish a useful analogy between stochastic thermodynamics of conservative systems and that of non-conservative systems. This analogy suggests the physical picture that the house-keeping parts of work, heat, and EP provide the dissipative background, whereas the excess parts of work, heat, and EP follow the same rules as in the stochastic thermodynamics of conservative systems. Consider now the steady state, where all thermodynamic variables are stationary. For a conservative system, the steady stat is the thermal equilibrium state, and $dS^{\rm tot}$, $đW$, $đQ$ all vanish identically. For a non-conservative system, the steady state is a nonequilibrium state with $đW$, $đQ$ nonvanishing. Yet the excess quantities $dS^{\rm ex}$, $đW^{\rm ex}$, $đQ^{\rm ex}$ all vanish identically. This provides a further analogy between thermodynamics of non-conservative systems and that of conservative systems, which is summarized in the sixth row of Table I.

Note however the analogy is not complete. For conservative systems, Eq. (2.16) tells us that the thermodynamic equilibrium state minimizes both the free energy and the EP. Yet for non-conservative systems, the excess EP as given by Eq. (3.62b) is generically not positive definite. This implies that the excess EP is generally not minimized at NESS. Additionally, Eq. (3.61) tells us that for fixed λ, the excess EP is proportional to the change of free energy. We then also see that in general the free energy is not minimized at NESS neither.

It is tempting to interpret the house-keeping EP as the EP needed to maintain the NESS. Such an interpretation is however inappropriate in general situation, for at least two reasons. Firstly, the house-keeping EP as given by Eq. (3.62a) generically depends on the pdf $p(\boldsymbol{x}, t)$, and hence is not a property of the NESS alone. Secondly, and more seriously, Eq. (3.62a) may even be negative. We will discuss a simple example in Sec. V A.

G. Gibbs gauge, and further analogy with conservative systems

The analogy between thermodynamic theory of conservative systems and that of non-conservative systems can be made complete in the Gibbs gauge, which was defined in

TABLE I. Comparison of thermodynamic relations for conservative systems and for non-conservative systems. Results with † are valid only in the Gibbs gauge.

	Conservative	Non-conservative
Thermodynamic variables	$đ\mathscr{W}, đ\mathscr{Q}, đW, đQ, dS^{\rm tot}$	$đ\mathscr{W}^{\rm ex}, đ\mathscr{Q}^{\rm ex}, đW^{\rm ex}, đQ^{\rm ex}, dS^{\rm ex}$
First law (trajectory)	$dH = đ\mathscr{W} + đ\mathscr{Q}$	$dH = đ\mathscr{W}^{\rm ex} + đ\mathscr{Q}^{\rm ex}$
First law (ensemble)	$dE = đW + đQ$	$dE = đW^{\rm ex} + đQ^{\rm ex}$
Free energy	$dF = đW + đQ - T\,dS$	$dF = đW^{\rm ex} + đQ^{\rm ex} - T\,dS$
EP	$dS^{\rm tot} = \beta(đW - dF) = dS - \beta đQ$	$dS^{\rm ex} = \beta(đW^{\rm ex} - dF) = dS - \beta đQ^{\rm ex}$
Steady state	$dS^{\rm tot} = dS = dE = dF = đW = đQ = 0$	$dS^{\rm ex} = dS = dE = dF = đW^{\rm ex} = đQ^{\rm ex} = 0$
Second law	$dS^{\rm tot} \geqslant 0$	$dS^{\rm ex} \geqslant 0$ †
Free energy	(EQ) Minimum free energy	(NESS) Minimum free energy†
Entropy production	(EQ) Minimum Entropy production	(NESS) Minimum excess EP†
Work	(EQ) Minimum work	(NESS) Minimum excess work†

Sec. III B. In this gauge, we have

$$\begin{aligned}\int_{\boldsymbol{x}} [(\partial_i + \partial_i U^{\rm G})p] L^{ij}\varphi_j^{\rm G} &= \int_{\boldsymbol{x}} (\partial_i p\, e^{U^{\rm G}}) L^{ij} e^{-U^{\rm G}}\varphi_j^{\rm G} \\ &= -\int_{\boldsymbol{x}} (p\, e^{U^{\rm G}}) \partial_i L^{ij} e^{-U^{\rm G}}\varphi_j^{\rm G} \\ &= 0, \end{aligned} \tag{3.63}$$

where in the second equality we integrated by parts, and in the last step we used the Gibbs gauge condition (3.15). Using this, together with Eqs. (3.59), (3.61) and (3.62b), the excess EP can be rewritten as

$$\begin{aligned}\Sigma^{\rm ex} &= \beta\left(\frac{đW^{\rm ex}}{dt} - \frac{dF_{\rm SS}}{dt}\right) = \frac{dS}{dt} - \beta\frac{đQ^{\rm ex}}{dt} \\ &= \int_{\boldsymbol{x}} (\partial_i(\log p + U^{\rm G}))B^{ij} p(\partial_j(\log p + U^{\rm G})) \\ &\geqslant 0, \end{aligned} \tag{3.64}$$

where $F_{\rm SS}$ is the steady-state free energy defined in Eq. (3.13), and in the last inequality we used the positivity of the matrix B^{ij}. Hence the positivity of excess EP can be used as a criterion of the stability of NESS.

The structures of Eqs. (3.64) and (2.16) are identical. Since Eq. (2.16) tells us that equilibrium state minimizes both the rate of EP and free energy, Eq. (3.64) tells us that, in Gibbs gauge, the NESS minimizes both the rate of excess EP and the free energy. Hence NESS in Gibbs gauge is characterized both by *the principle of minimal free energy* and by *the principle of minimal excess dissipation*. Likewise, whilst Eq. (2.16) says that in conservative systems, the minimum work needed to carry out a process from one equilibrium state to another equilibrium state is the free energy difference between these two equilibrium states, Eq. (3.64) tells us that in the Gibbs gauge the *minimum excess work* needed to carry out a process from one NESS to another NESS is the free energy difference between these two steady states. This is precisely *the principle of minimal excess work*, posed as conjecture by Oono and Paniconi [17], as well as by Sasa and Tasaki [1], now proved for arbitrary nonlinear Langevin dynamics within Gibbs gauge. These results are summarized in the last four rows of Table I.

The upshot of Table I is that within the Gibbs gauge, there is a complete analogy between thermodynamic theory of conservative systems and that of non-conservative systems. Specializing to the stationary state, it means that the thermodynamics of equilibrium states is identical to that of steady sates, as long as we only keep track of the excess parts of work, heat, and EP. More specifically, as long as we identify the $H^{\rm G}$ as the system Hamiltonian (of mean force), classical statistical mechanics and thermodynamics can be used to study the static properties of NESS, as well as the transitions between different NESS.

To further demonstrate the analogy, we consider slow variation of the control parameter λ, such that the system transits goes from one NESS to another NESS. In equilibrium thermodynamics, we know that the total EP associated with a process from one equilibrium state to another approaches zero as the speed of the process becomes infinitely slow, i.e., it becomes *quasistatic*. For quasistatic process from NESS to another NESS, the *excess* EP is given by Eq. (3.64). We expect that the pdf deviates only slightly from the instantaneous NESS pdf as given by Eq. (3.9). Hence let us write,

$$p(\boldsymbol{x}, t) = p_{\rm SS}(\boldsymbol{x}, \lambda(t)) + \delta p(\boldsymbol{x}, t). \tag{3.65}$$

Substituting this back into FPE (3.2) and expand up to first order, we obtain

$$\partial_t \delta p + \frac{\partial p_{\rm SS}}{\partial\lambda}\dot{\lambda} = \mathscr{L}\delta p. \tag{3.66}$$

It is easy to see that δp is linear in $\dot{\lambda}$, and hence vanishes as $\dot{\lambda} \to 0$. It then follows that the integrand in Eq. (3.64) is quadratic in $\dot{\lambda}$, and the integrated *excess* EP converges to zero as the process becomes infinitely slow,

$$\Delta S^{\rm ex} = \int_{t_i}^{t_f} \frac{dS^{\rm ex}}{dt} dt \propto \int dt\, \dot{\lambda}^2 \to 0. \tag{3.67}$$

Hence indeed the excess EP of quasistatic transition from one NESS to another is zero, which is analogous to equilibrium thermodynamics.

The relative entropy between p and the NESS $p_{\rm SS}$ is

$$\begin{aligned} D(p||p_{\rm SS}) &= \int_{\boldsymbol{x}} p(\boldsymbol{x}, t) \log\frac{p(\boldsymbol{x}, t)}{p_{\rm SS}(\boldsymbol{x})} \\ &= \int_{\boldsymbol{x}} p(\boldsymbol{x}, t)(\log p(\boldsymbol{x}, t) + U^{\rm G}(\boldsymbol{x}; \lambda)). \end{aligned} \tag{3.68}$$

Recall that the nonequilibrium free energy in Gibbs gauge is

$$\beta F[p(\boldsymbol{x},t)] = \int_{\boldsymbol{x}} p(\boldsymbol{x},t)(\log p(\boldsymbol{x},t) + \beta H^{\mathrm{G}}(\boldsymbol{x};\lambda)), \quad (3.69)$$

and further use Eq. (3.12), we may rewrite Eq. (3.68) as

$$D(p||p_{\mathrm{SS}}) = \beta F[p(\boldsymbol{x},t)] - \beta F_{\mathrm{SS}}, \quad (3.70)$$

whereas F_{SS} is the steady-state free energy defined in Eq. (3.13).

Now assume that the control parameter λ is fixed. Taking the time derivative of Eq. (3.70), we obtain

$$\frac{d}{dt}D(p||p_{\mathrm{SS}}) = \beta\frac{d}{dt}F[p(\boldsymbol{x},t)] = \int_{\boldsymbol{x}} (\log p(\boldsymbol{x},t) + U(\boldsymbol{x};\lambda))\mathscr{L}p(\boldsymbol{x},t). \quad (3.71)$$

Further, using Eq. (3.3) and integrating by parts, and comparing with Eq. (3.62b), we obtain

$$\frac{d}{dt}D(p||p_{\mathrm{SS}}) = \beta\frac{d}{dt}F[p(\boldsymbol{x},t)] = -\Sigma^{\mathrm{ex}}. \quad (3.72)$$

Hence in the Gibbs gauge the excess EP is precisely negative the rate of the relative entropy $D(p||p_{\mathrm{SS}})$. Furthermore, $D(p||p_{\mathrm{SS}})$, or equivalently $F[p(\boldsymbol{x},t)]$, serves as a Lyapunov function for non-conservative Langevin dynamics, because it decreases monotonically, and its rate vanishes identically at the NESS.

H. Entropy production and pumped entropy

For Langevin dynamics described by Eq. (3.1) satisfying detailed balance conditions (3.4), the logarithmic ratio between the transition probability of the forward process, $P_{\lambda}(\boldsymbol{x}_1|\boldsymbol{x};dt)$, and that of the backward process, $P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}_1^*;dt)$, is derived in Eq. (4.6) of Appendix,

$$\log\frac{P_{\lambda}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}_1^*;dt)} = d\mathcal{S}_{\mathrm{B}} + d\mathcal{S}_{\mathrm{P}}, \quad (3.73a)$$

where $d\mathcal{S}_{\mathrm{B}}$ and $d\mathcal{S}_{\mathrm{P}}$ are defined as

$$d\mathcal{S}_{\mathrm{B}} \equiv -\beta đ\mathcal{Q} = -d_{\boldsymbol{x}}U + \varphi_i \circ dx^i, \quad (3.73b)$$

$$d\mathcal{S}_{\mathrm{P}} \equiv -Q^{ij}(\partial_i\varphi_j)dt. \quad (3.73c)$$

Note that the control parameter may also appear in the non-conservative force φ_i. Whilst $d\mathcal{S}_{\mathrm{B}}$ is the change of bath entropy, $d\mathcal{S}_{\mathrm{P}}$ should be interpreted as entropy cost for maintaining the non-conservative force. Alternatively, we may understand $d\mathcal{S}_{\mathrm{P}}$ as the entropy pumped out of the system by the non-conservative force, and hence will be called the *pumped entropy*. Similar terms have been identified and studied sometime ago in Brownian motion driven velocity dependent force [37,38]. The ensemble average of $d\mathcal{S}_{\mathrm{P}}$ is

$$dS_{\mathrm{P}} = \Sigma_{\mathrm{P}}dt = \int_{\boldsymbol{x}} d\mathcal{S}_{\mathrm{P}}\, p = -dt\int_{\boldsymbol{x}} Q^{ij}(\partial_i\varphi_j)p, \quad (3.74)$$

which is precisely negative the extra term appearing in the r.h.s. of Eq. (3.49). The sum $d\mathcal{S}_{\mathrm{B}} + d\mathcal{S}_{\mathrm{P}}$ in the r.h.s. of Eq. (3.73a) can then be understood as the entropy change of the *environment* during the infinitesimal step of evolution in the forward process. Iin the absence of non-conservative force, Eqs. (3.73) reduces to Eqs. (2.21).

To have a better understanding of the pumped entropy, we revisit the unitary limit of the Ito-Langevin dynamics, Eq. (3.37). The corresponding FPE is

$$\begin{aligned}\partial_t p &= \partial_i Q^{ij}(\partial_j + (\partial_j U) - f_j)p \\ &= (\partial_j U)(\partial_i p) - Q^{ij}\partial_i(f_j p),\end{aligned} \quad (3.75)$$

where we have used the assumption $\partial_j Q^{ij} = 0$. Let us calculate the rate of Gibbs-Shannon entropy using Eq. (3.75). Integrating by parts a few times, we obtain

$$\begin{aligned}\frac{d}{dt}S[p] &= -\frac{d}{dt}\int_{\boldsymbol{x}} p\log p = -\int_{\boldsymbol{x}} \log p\,\partial_t p, \\ &= \int_{\boldsymbol{x}} Q^{ij}(\partial_i\varphi_j)p = -\frac{d}{dt}S_{\mathrm{P}} = -\Sigma_{\mathrm{P}}.\end{aligned} \quad (3.76)$$

Let us rewrite this into a more informative form

$$\frac{dS}{dt} + \frac{dS_{\mathrm{P}}}{dt} = 0, \quad (3.77)$$

which says that in the presence of non-conservative force, the system entropy is not conserved, but the sum of S and S_{P} is conserved. Our expectation is that a reversible, unitary dynamics must preserve the total entropy. Hence dS_{P} must be interpreted as the entropy change of the agent who supplies the non-conservative force, or, equivalently, the entropy pumped out of the system by the agent during the dynamic evolution. The associated rate Σ_{P} is then the rate of entropy pumping. Such an interpretation was first proposed by Kim and Qian in the study of Hamiltonian system driven by velocity-dependent forces [37,38]. Note the interesting similarity between the above argument and Landauer's principle, which says that thermodynamic entropy may be reduced at the cost of information entropy, and vice versa. Note also that it is logically inconsistent to associate dS_{P} with entropy change of the heat bath, the system described by Eq. (3.75) is not in contact with any heat bath.

Coming back to the non-conservative Langevin dynamics. The total entropy change of the universe now consists of three parts: entropy change of system, entropy change of the heat bath, and the entropy pumped out of the system by the non-conservative force. Equation (3.49) can now be rewritten into the following form:

$$\Sigma^{\mathrm{tot}} = \frac{dS^{\mathrm{tot}}}{dt} = \frac{dS}{dt} + \frac{dS_B}{dt} + \Sigma_{\mathrm{P}} \quad (3.78a)$$

$$= \int_{\boldsymbol{x}} (\partial_i + (\partial_i U) - \varphi_i)p\frac{B^{ij}}{p}(\partial_j + (\partial_j U) - \varphi_j)p$$

$$\geqslant 0, \quad (3.78b)$$

where the EP is apparently non-negative, and hence is consistent with the second law. Using Eq. (3.7b), we may rewrite Eq. (3.78a) solely in terms of irreversible probability current,

$$\Sigma^{\mathrm{tot}} = \int_{\boldsymbol{x}} j^i_{\mathrm{IR}}B^{-1}_{ij}p^{-1}j^j_{\mathrm{IR}}, \quad (3.78c)$$

which is formally identical to Eq. (2.20). Using Eq. (3.47), we may also rewrite Eq. (3.78a) as

$$\Sigma^{\mathrm{tot}} = \frac{dS^{\mathrm{tot}}}{dt} = \beta\left(\frac{đW}{dt} - \frac{dF}{dt}\right) + \Sigma_{\mathrm{P}} \geqslant 0, \quad (3.78d)$$

which imposes a lower bound on the work needed to carry out a thermodynamic process. Inequality similar to Eq. (3.78d) was derived in the study of information heat engines [62,63], which transform thermodynamic entropy into information entropy and at the same time extract mechanical energy from heat bath.

Note that using Eqs. (3.78a) and (3.60), we can also decompose the EP in a different way,

$$dS^{\rm tot} = dS^{\rm hk} + dS^{\rm ex} + dS_{\rm P}, \tag{3.79a}$$

$$\Sigma^{\rm tot} = \Sigma^{\rm hk} + \Sigma^{\rm ex} + \Sigma_{\rm P}. \tag{3.79b}$$

Whilst we have shown that both $dS^{\rm tot}$ and $dS^{\rm ex}$ are positive definite, $dS^{\rm hk}$ and $dS_{\rm P}$ may be either positive or negative. Hence the total EP is generally not minimized at the NESS.

Decomposition of EP for nonlinear Langevin dynamics with both even and odd variables were also studied in Refs. [32,33]. The results are, however, different from ours.

I. Entropy and energy balance in NESS

NESSs are similar to equilibrium states in the sense that all thermodynamic variables are stationary. Yet, they are also different from equilibrium states in that they have nonvanishing rate of EP. As we have seen above, for systems with asymmetric kinetic matrix, there is an also interesting possibility of entropy pumping, where the system exchanges entropy with environment without dissipation. This leads to several possible scenarios of entropy and energy balance in NESS, which we will discuss now.

Since in NESS, both λ and the pdf are fixed, we have from Eqs. (3.55) that $đW^{\rm ex} = đQ^{\rm ex} = 0$. It then follows from Eq. (3.54b) and (3.53) that

$$đQ = đQ^{\rm hk} = -đW^{\rm ex} = -đW. \tag{3.80}$$

This can in fact be understood as the first law in NESS, since the internal energy is also stationary. Now since $dF = 0$, Eqs. (3.78d) and (3.80) yield

$$\frac{dS_{\rm P}}{dt} \geqslant -\frac{\beta đW}{dt} = \frac{\beta đQ}{dt}, \tag{3.81}$$

where $dS_{\rm P}, đQ, đW$ can be either positive or negative. In the boring scenario, illustrated in Fig. 1(a), pumped entropy vanishes, $dS_{\rm P} = 0$, and we have

$$-\frac{đQ}{dt} = \frac{đW}{dt} \geqslant 0, \tag{3.82}$$

which means that the non-conservative force does positive work, and the energy is constantly dissipated to the heat bath in the form of heat. The conservative force does no work since the parameter λ is fixed. A slightly interesting more scenario is shown in Fig. 1(b), where the work is only partially dissipated to the bath, with the remaining part output as another form of work. This can be realized by coupling two non-conservative forces to the system, one doing positive work, and the other doing negative work. The agent and the system then behave as an engine with nonideal efficiency. An interesting example involving time-dependent protocol is *Brownian gyrator* proposed by Filliger and Reimann [64].

The situation becomes more interesting if the pumped entropy is nonzero. For simplicity, we assume that $dS^{\rm tot}/dt$ is very small, i.e., the NESS is nearly reversible, comparing with dS_P/dt. If $dS_P/dt > 0$, entropy is constantly pumped out of the system by the non-conservative force. Equation (3.81) then becomes

$$-\frac{đQ}{dt} = \frac{đW}{dt} < 0, \tag{3.83}$$

which means that heat flows from the bath to the system at a constant rate, and the non-conservative force does negative work, i.e., energy is constantly extracted out of the system. In this case, as illustrated in Fig. 1(c), the agent acts like a Maxwell demon, who extracts both energy and entropy from the system. This is similar to a Szilard engine, where the entropy extracted from the bath is saved in the memory space of the agent in the form of information entropy. Unlike a Szilard engine, however, here the system does not need any time-dependent protocol to realize entropy pumping. The system described by Langevin dynamics with fixed λ is a *steady-state information engine*.

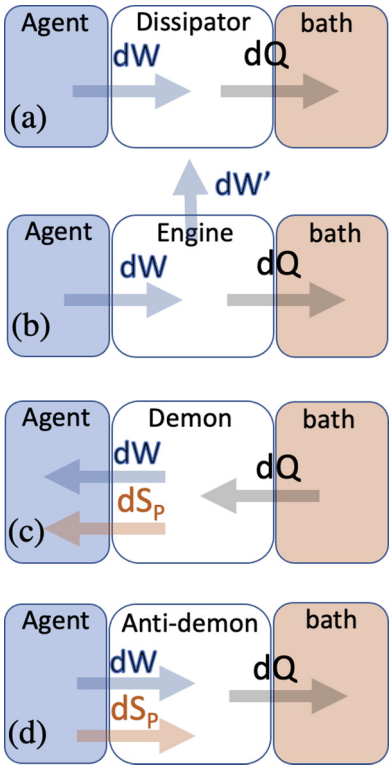

FIG. 1. Four possible scenarios of entropy and energy balance in steady states. (a) The agent supplies work to the system, which is completely dissipated to the heat bath. (b) The agent supplies work to the system, part of which is output as work of another form, the remaining energy is dissipated to the bath. The system then behaves as an engine. One example is to use thermo-electric effect as a battery. (c) The agent extracts both work and entropy from the system, and the system absorbs heat from the bath. The agent behaves as a Maxwell demon constantly measures the system and extracts energy and information entropy. (d) The agent supplies both work and entropy to the system, and the system releases heat to the bath. The agent behaves as a demon running backwards, i.w., an antidemon.

If $dS_P/dt < 0$, Eq. (3.81) then becomes

$$-\frac{đQ}{dt} = \frac{đW}{dt} > 0, \tag{3.84}$$

which means that the agent does positive work, and the associated energy flows to the heat bath in the form of heat, as shown in Fig. 1(d). In this case, entropy is pumped into the system at a constant rate by the non-conservative force. This is the Maxwell demon *running backwards*. In the case of reversed Szilard engine, the agent erases information in its memory via mechanical work. The erased information entropy and the energy are dissipated in the bath.

Examples of entropy pumping will be discussed in Sec. V.

J. Connections with steady-state thermodynamics and Glansdorff-Prigogine stability criterion

Sasa and Tasaki defined a *free energy* for the steady state expressed as a Gibbs distribution, see Eq. (2.33) of Ref. [1], which is precisely the steady-state free energy we defined in Eq. (3.13).

Hatano and Sasa defined excess heat and house-keeping heat in the setting of over-damped Langevin dynamics as

$$đ\mathcal{Q}^{\rm hk} = -f_i^{\rm G} \circ dx^i, = -T\,\varphi_i^{\rm G} \circ dx^i, \tag{3.85}$$

$$đ\mathcal{Q}^{\rm ex} = đ\mathcal{Q} - đ\mathcal{Q}^{\rm ex} = d_{\boldsymbol{x}} H^{\rm G}, \tag{3.86}$$

using our notations (c.f. Eq. 16 of Ref. [25]). But these are precisely our definitions (3.51) specialized in Gibbs gauge. Hence, our theory can be deemed as a generalization of the theory of Hatano and Sasa to covariant nonlinear Langevin dynamics where both system variables and control parameters may have odd components.

Hatano and Sasa [25] proved the following equality for general Markov process with a unique steady state:

$$\langle e^{-\int d_\lambda U^{\rm G}} \rangle = 1, \tag{3.87}$$

where it is assumed that the system starts from the NESS (3.9) in the initial time. As long as the system variables are continuous, the Markov process can be described by our theory. Using Eq. (3.12) we have

$$\begin{aligned} d_\lambda U^{\rm G} &= \beta d_\lambda H^{\rm G} - \beta d_\lambda F_{\rm SS} \\ &= \beta đ W^{\rm ex} - \beta dF_{\rm SS}. \end{aligned} \tag{3.88}$$

Hence the Hatano-Sasa equality (3.87) becomes

$$\langle e^{-W^{\rm ex} + \beta \Delta F_{\rm SS}} \rangle = 1. \tag{3.89}$$

Further using Jensen's inequality, we obtain

$$W^{\rm ex} \geqslant \Delta F_{\rm SS} = F_{\rm SS}(\lambda_f) - F_{\rm SS}(\lambda_i), \tag{3.90}$$

where λ_i and λ_f are respectively the control parameters in the initial and final time of the process, whilst $F_{\rm SS}(\lambda_f)$ and $F_{\rm SS}(\lambda_i)$ are the corresponding steady-state free energies. This result is called "the second law of steady-state thermodynamics" in Refs. [25] and "the second law of time-dependent Markov processes" in Refs. [1]. Since Eq. (3.90) was proved using Eq. (3.87), it is applicable only for processes starting from NESS. Note however in the final time the system is generally not in the NESS.

By contrast, Eq. (3.64) was derived for processes that start from generic nonequilibrium states, with dF the differential of nonequilibrium free energy, defined in Eq. (2.10c). Integrating this inequality over a finite process, we obtain

$$W^{\rm ex} \geqslant \Delta F = F[p(t_f)] - F[p(t_i)], \tag{3.91}$$

where $F[p(t_i)], F[p(t_f)]$ are the nonequilibrium free energy at the initial and final times. Hence the range of applicability of the inequality (3.91) is larger than that of (3.90). For a process starting from NESS, both Eqs. (3.91) and (3.90) are applicable. Yet Eq. (3.91) becomes

$$W^{\rm ex} \geqslant \Delta F = F[p(t_f)] - F_{\rm SS}(\lambda_i). \tag{3.92}$$

Since the NESS minimizes the free energy, we have $F[p(t_f)] \geqslant F_{\rm SS}(\lambda_f)$, and hence Eq. (3.92) is stronger than Eq. (3.90). Consequently, we may call Eq. (3.64) *the strong version of the second law of steady-state thermodynamics*.

Glansdorff and Prigogine [7,15] defined the excess EP as the rate of the second-order variation of system entropy, and showed that for systems near thermal equilibrium, its positivity can be used as a criterion for the stability of NESS. The applicability of Glansdorff-Prigogine criterion for systems far from equilibrium or stochastic systems is however still controversial. Here we will show that the definition of excess EP by Glansdorff and Prigogine is consistent with our definition in Gibbs gauge, and their stability criterion is valid for nonlinear Langevin systems even far away from equilibrium.

We consider the variation δp around the NESS,

$$p = p_{\rm SS} + \delta p, \tag{3.93}$$

and study the resulting variation of the total EP, expanded in terms of δp. Firstly inspecting Eqs. (3.62a) and (3.74) we see that $\Sigma^{\rm hk}$ and $\Sigma_{\rm P}$ are both linear in p, and hence they do not have second-order variation. Secondly, inspecting Eq. (3.64), we see that $\Sigma^{\rm ex}$ has no zeroth- or first-order contribution. Hence near the NESS the total EP can be expanded as

$$\Sigma^{\rm tot} = \Sigma^{\rm tot}_{\rm SS} + \delta^1\Sigma^{\rm tot} + \delta^2\Sigma^{\rm tot} + \cdots, \tag{3.94}$$

$$\delta^1\Sigma^{\rm tot} = \delta^1\Sigma^{\rm hk} + \delta^1\Sigma_{\rm P}, \tag{3.95}$$

$$\delta^2\Sigma^{\rm tot} = \Sigma^{\rm ex}, \tag{3.96}$$

where $\delta^1\Sigma^{\rm tot}$, $\delta^2\Sigma^{\rm tot}$ are respectively of first and second order in δp, and neglected terms are at least of third order in δp. Using Eq. (3.64) we can further obtain the following explicit form of $\Sigma^{\rm ex}$ in terms of δp near NESS,

$$\Sigma^{\rm ex} = \int_{\boldsymbol{x}} (\partial_i e^{U^{\rm G}} \delta p) B^{ij} e^{-U^{\rm G}} (\partial_i e^{U^{\rm G}} \delta p) + \cdots. \tag{3.97}$$

We may further define "affinity" $A_i(\boldsymbol{x})$ via

$$A_i(\boldsymbol{x}) \equiv \partial_i(\log p + U) - \varphi_i, \tag{3.98}$$

so that the total EP (3.78a) can be rewritten as

$$\Sigma^{\rm tot} = \int_{\boldsymbol{x}} A_i(\boldsymbol{x}) j^i_{\rm IR}(\boldsymbol{x}), \tag{3.99}$$

where $j^i_{\rm IR}(\boldsymbol{x})$ is the irreversible probability current given in Eq. (6.4). This result is analogous to the result in irreversible thermodynamics, see Eq. (14.6) of Ref. [3], or the result in

master equation systems, see Eq. (7.6) of Ref. [14]. The difference is of course in previous theories, all system variables are even, and hence the irreversible current is also the total current, whereas in the present, theory, only the irreversible current enters Eq. (3.99). Note that in the NESS, the total current is divergenceless, but the the irreversible current is generally not divergenceless.

Recall that the free energy is minimized at NESS. Using Eq. (3.69), we may expand $F[p]$ near the NESS,

$$F[p] = F_{\rm SS} + \delta^2 F = F_{\rm SS} + T\int_{\boldsymbol{x}} \frac{\delta p^2}{2\, p_{\rm SS}}. \tag{3.100}$$

It is easy to see from Eq. (3.69) that the second-order variation of βF is identical to negative the second-order variation of system entropy,

$$\delta^2 \beta F[p] = -\delta^2 S[p]. \tag{3.101}$$

Combining these results with Eqs. (3.72) and (3.96), we finally obtain

$$\Sigma^{\rm ex} = \delta^2 \Sigma^{\rm tot} = -\frac{d}{dt}\delta^2 \beta F = \frac{d}{dt}\delta^2 S. \tag{3.102}$$

Hence, same as in Glansdorff-Prigogine theory, the excess EP in our theory is also the second-order variation of the system entropy around the NESS. Furthermore, Eq. (3.97) explicitly demonstrates that the excess EP is positive if and only if the matrix B^{ij} is positive, which is in turn equivalent to the stability of the NESS. Hence the positivity of the excess EP can be used as a criterion for stability of NESS. For overdamped Langevin systems with additive noises, these results were established in Refs. [35,36].

An interesting comparison can be made between stochastic thermodynamics and irreversible thermodynamics at this stage. At the level of stochastic thermodynamics, the positivity of excess EP is guaranteed by positivity of the noise covariance matrix B^{ij}, which is always valid. As shown in Sec. V A of Ref. [49], irreversible thermodynamics corresponds to the deterministic limit of stochastic thermodynamics, where the Langevin equations becomes deterministic dynamic equations. It may happen that there is a bifurcation in the loci of minimum of the generalized potential $U(\boldsymbol{r}, \lambda)$, as one tunes the parameter λ. Such a bifurcation is identified as a nonequilibrium phase transition in irreversible thermodynamics. Yet, at the level of stochastic thermodynamics, due to statistical fluctuations, both the steady state pdf and current change smoothly crossing the bifurcation point. Nonequilibrium phase transition can only appear in the thermodynamic limit, where statistical fluctuations are negligible, and irreversible thermodynamics becomes asymptotically exact, except in the vicinity of a critical point.

K. Gauge transformation revisited

For completeness, we summarize the properties of gauge transformation. According to Eqs. (3.5) and (3.28), (2.7), U, H, φ_i, f_i transform as

$$U \to U + \psi, \tag{3.103a}$$
$$H \to H + T\,\psi, \tag{3.103b}$$
$$\varphi_i \to \varphi_i + \partial_i \psi, \tag{3.103c}$$
$$f_i \to f_i + T\,\partial_i \psi. \tag{3.103d}$$

Note that $\psi(\boldsymbol{x}, \lambda)$ generally depend both on $\boldsymbol{x}$ and on λ, and is not invariant under time reversal, i.e., $\psi(\boldsymbol{x}, \lambda) \neq \psi(\boldsymbol{x}^*, \lambda^*)$. Using Eqs. (3.26), (3.29), (3.30), and (3.73c), we further have

$$đ\mathscr{W} \to đ\mathscr{W} + T\, d\psi, \tag{3.104a}$$
$$đH \to đH + T\, d\psi, \tag{3.104b}$$
$$đ\mathcal{Q} \to đ\mathcal{Q}, \tag{3.104c}$$
$$d\mathcal{S}_{\rm P} \to d\mathcal{S}_{\rm P}. \tag{3.104d}$$

Further using Eqs. (3.50) and (3.51), we may find

$$đ\mathscr{W}^{\rm hk} \to đ\mathscr{W}^{\rm hk} + T d_{\boldsymbol{x}}\psi, \tag{3.105a}$$
$$đ\mathscr{W}^{\rm ex} \to đ\mathscr{W}^{\rm ex} + T d_{\lambda}\psi, \tag{3.105b}$$
$$đ\mathcal{Q}^{\rm hk} \to đ\mathcal{Q}^{\rm hk} - T d_{\boldsymbol{x}}\psi, \tag{3.105c}$$
$$đ\mathcal{Q}^{\rm ex} \to đ\mathcal{Q}^{\rm ex} + T d_{\boldsymbol{x}}\psi. \tag{3.105d}$$

Hence even though the heat is gauge invariant, neither of its house-keeping or excess component is gauge invariant.

Since both the heat and pumped entropy are gauge-invariant, the formula Eq. (3.73) is also gauge-invariant.

The transform of thermodynamic quantities at ensemble level can be similarly obtained. In particular, $đQ$, $đS_{\rm P}$, dS, and $dS^{\rm tot}$ are all invariant under gauge transformation, whereas the differential of nonequilibrium free energy and work transform as

$$dF \to dF + T\langle\langle d\psi\rangle\rangle, \tag{3.106a}$$
$$dW \to dW + T\langle\langle d\psi\rangle\rangle. \tag{3.106b}$$

Finally, the house-keeping and excess EP transform as

$$dS^{\rm hk} \to dS^{\rm hk} + T\langle\langle d_{\boldsymbol{x}}\psi\rangle\rangle, \tag{3.107a}$$
$$dS^{\rm ex} \to dS^{\rm ex} - T\langle\langle d_{\boldsymbol{x}}\psi\rangle\rangle. \tag{3.107b}$$

That is, the decomposition Eq. (3.79) of EP is gauge dependent. This is in strong contrast with the decomposition Eq. (3.78a), which is gauge invariant.

Associated with the system Hamiltonian, are the corresponding equilibrium (or steady state) free energy and equilibrium Gibbs distribution,

$$F_{\rm EQ} = -T\log\int_{\boldsymbol{x}} e^{-\beta H}, \tag{3.108}$$
$$p_{\rm EQ}(\boldsymbol{x}, \lambda) = e^{-\beta(H - F_{\rm EQ})}, \tag{3.109}$$

which transform as

$$F_{\rm EQ} \to F_{\rm EQ} - T\log\langle e^{-\psi}\rangle, \tag{3.110}$$
$$p_{\rm EQ} \to \frac{e^{-\psi}}{\langle e^{-\psi}\rangle} p_{\rm EQ}. \tag{3.111}$$

The transformation Eqs. (3.103) is a gauge transform from the perspective of Langevin dynamics, because it preserves the dynamics equation (3.1). Yet, it leads to a physically distinct HMF and Gibbs distribution. Similar features also show up in the gauge transformation of quantum mechanics, where

a gauge transformation leaves the Schrodinger equation invariant, but changes the quantum Hamiltonian.

IV. FLUCTUATION THEOREMS

In this section we will derive two types of fluctuation theorems. The first type concerns the total EP and is applicable to processes that start from and end at thermal equilibrium states, which are invariant under time reversal. The theorem is formulated in terms of total work and pumped entropy, both defined as functional of trajectory. The second type concerns the excess EP, and is applicable to processes that start from and end at NESS, which are generically not invariant under time reversal. The theorem is formulated in terms of excess work. The first type of fluctuation theorems is gauge dependent. Hence as we change the gauge continuously, we obtain a continuous manifold of fluctuation theorems. By contrast, the second type of fluctuation theorem is applicable only in Gibbs gauge. Both types of fluctuations belong to *transient fluctuation theorem*, which hold exactly for processes with finite durations. We do not expect house-keeping EP or pumped entropy obey any transient fluctuation theorem in generic situation.

A. Fluctuation theorem for the total EP

Let us define the forward process and the backward process as follows.

(i) *Forward process.* At $t = t_i$, the system starts with equilibrium state

$$p_{\rm EQ}(\boldsymbol{x}, \lambda_i) = e^{-U(\boldsymbol{x},\lambda_i)} = e^{\beta(F(\lambda_i)-H(\boldsymbol{x},\lambda_i))}. \tag{4.1}$$

where $\lambda_i = \lambda(t_i)$. Because U, H, F all have time-reversal symmetry, i.e., they satisfy Eqs. (2.5c) and (2.9), $p_{\rm EQ}(\boldsymbol{x}, \lambda_i)$ is invariant under simultaneous time-reversal of $\boldsymbol{x}$ and λ_i: $p_{\rm EQ}(\boldsymbol{x}^*, \lambda_i^*) = p_{\rm EQ}(\boldsymbol{x}, \lambda_i)$. The system evolves according to the Langevin dynamics (3.1), with the dynamic protocol given by $\lambda^{\rm F}(t) = \lambda(t)$, $\varphi^{\rm F}(t) = \varphi(\lambda(t))$ with $t \in [t_i, f_f]$. Here the superscript F means Forward, the subscripts i and f mean initial and final. Note that the time dependence of the nonconservative force φ is completely determined by that of the control parameter λ, and hence strictly speaking, we do not need to specify φ separately.

(ii) *Backward process.* At $t = -t_f$, the system starts with equilibrium state

$$p_{\rm EQ}(\boldsymbol{x}, \lambda_f^*) = e^{-U(\boldsymbol{x},\lambda_f^*)} = e^{\beta(F(\lambda_f^*)-H(\boldsymbol{x},\lambda_f^*))}, \tag{4.2}$$

where $\lambda_f = \lambda(t_f)$. Note that $p_{\rm EQ}(\boldsymbol{x}, \lambda_f^*)$ is also invariant under simultaneous time reversal of $\boldsymbol{x}$ and λ_f: $p_{\rm EQ}(\boldsymbol{x}^*, \lambda_f) = p_{\rm EQ}(\boldsymbol{x}, \lambda_f^*)$. The dynamic protocol is given by $\lambda^{\rm B}(t) = \lambda^*(-t)$, $\varphi^{\rm B}(t) = \varphi(\lambda^*(-t))$, with $t \in [-f_f, -t_i]$. Here the superscript B means backward.

As in Ref. [49], we use γ to denote a generic dynamic trajectory (the forward trajectory) of the system $\boldsymbol{x}(t), t \in [t_i, t_f]$, and use $\tilde{\gamma}$ to denote the time-reversed trajectory $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}^*(-t)$ (the backward trajectory). As illustrated in Fig. 2(a), the forward trajectory starts from $\boldsymbol{x}(t_i)$ and ends at $\boldsymbol{x}(t_f)$, whereas the backward trajectory starts from $\tilde{\boldsymbol{x}}(-t_f) = \boldsymbol{x}^*(t_f)$ and ends at $\tilde{\boldsymbol{x}}(-t_i) = \boldsymbol{x}^*(t_i)$. For every infinitesimal step in the forward trajectory, there is a corresponding infinitesimal step in the backward trajectory, as illustrated in Figs. 2(a) and 2(b). In the forward step, the system goes from $\boldsymbol{x}$ to $\boldsymbol{x} + d\boldsymbol{x}$, whereas the control parameter goes from λ to $\lambda + d\lambda$. By contrast, in the backward step, the system goes from $\boldsymbol{x}^* + d\boldsymbol{x}^*$ to $\boldsymbol{x}^*$, whereas the control parameter goes from $\lambda^* + d\lambda^*$ to λ^*. The differential work, heat, and pumped entropy along the infinitesimal forward trajectory, defined in Eqs. (3.26), (3.30), and (3.73c), are the exact opposite of the corresponding quantities along the back trajectory,

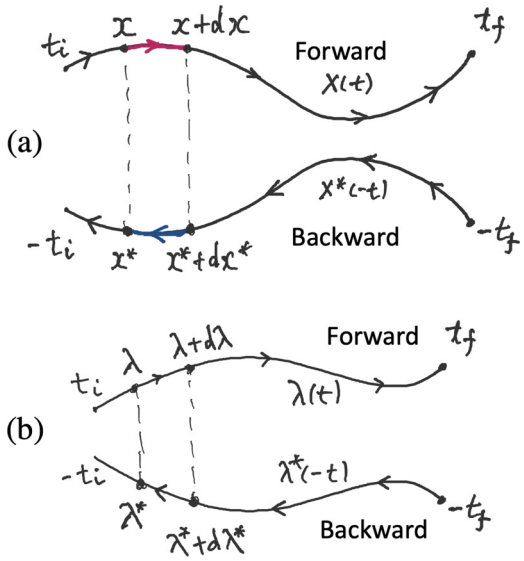

FIG. 2. Top: Forward and backward trajectories. Corresponding to every step $\boldsymbol{x} \to \boldsymbol{x} + d\boldsymbol{x}$ in the forward trajectory (shown in red) there is a step $\boldsymbol{x}^* + d\boldsymbol{x}^* \to \boldsymbol{x}^*$ in the backward trajectory (shown in blue). Bottom: Forward and backward protocols.

$$(đ\mathscr{W})_{\rm B} = -(đ\mathscr{W})_{\rm F}, \tag{4.3}$$

$$(đ\mathscr{Q})_{\rm B} = -(đ\mathscr{Q})_{\rm F}, \tag{4.4}$$

$$(d\mathcal{S}_{\rm P})_{\rm B} = -(d\mathcal{S}_{\rm P})_{\rm F}. \tag{4.5}$$

The first two equalities are obvious. We prove the last equality below

$$\begin{aligned}(d\mathcal{S}_{\rm P})_{\rm B} &= Q^{ij}(\boldsymbol{x}^*, \lambda^*)\partial_i^* \varphi_j(\boldsymbol{x}^*, \lambda^*)dt \\ &= \epsilon_i Q^{ij}(\boldsymbol{x}^*, \lambda^*)\epsilon_j \partial_i \varphi_j(\boldsymbol{x}, \lambda)dt \\ &= -Q^{ij}(\boldsymbol{x}, \lambda)\partial_i \varphi_j(\boldsymbol{x}, \lambda)dt \\ &= -(d\mathcal{S}_{\rm P})_{\rm F},\end{aligned} \tag{4.6}$$

where in the second and third equalities we used respectively the detailed balance conditions (3.6b) and (3.4b).

Now the integrated work, the integrated heat, and the integrated pumped entropy along the forward trajectory can be obtained,

$$\mathscr{W}[\gamma, \lambda^{\rm F}] \equiv \int_\gamma (đ\mathscr{W})_{\rm F} = \int_\gamma (d_\lambda H + f_i \circ dx^i), \tag{4.7}$$

$$\mathscr{Q}[\gamma, \lambda^{\rm F}] \equiv \int_\gamma (đ\mathscr{Q})_{\rm F} = \int_\gamma (d_{\boldsymbol{x}} H - f_i \circ dx^i), \tag{4.8}$$

$$\mathcal{S}_{\rm P}[\gamma, \lambda^{\rm F}] \equiv \int_\gamma (d\mathcal{S}_{\rm P})_{\rm F} = -\int_\gamma Q^{ij}(\partial_i \varphi_j)dt. \tag{4.9}$$

The corresponding quantities along the backward trajectory in the backward process can be similarly calculated,

$$\mathscr{W}[\tilde{\gamma}, \lambda^{\rm B}] \equiv \int_{\gamma} (d\!\!\!^{-}\mathscr{W})_{\rm B} = -\mathscr{W}[\gamma, \lambda^{\rm F}], \tag{4.10a}$$

$$\mathscr{Q}[\tilde{\gamma}, \lambda^{\rm B}] \equiv \int_{\gamma} (d\!\!\!^{-}\mathscr{Q})_{\rm B} = -\mathscr{Q}[\gamma, \lambda^{\rm F}], \tag{4.10b}$$

$$\mathscr{S}_{\rm P}[\tilde{\gamma}, \lambda^{\rm B}] \equiv \int_{\gamma} (d\mathscr{S}_{\rm P})_{\rm B} = -\mathscr{S}_{\rm P}[\gamma, \lambda^{\rm F}]. \tag{4.10c}$$

For each infinitesimal segment of the forward process and the corresponding segment of the backward process, we obtain an equality (3.73), which relates the transition probabilities to change of environmental entropy. Summing up all these equalities, we obtain an equality for the entire trajectory, which relates the conditional path probability of the forward process to that of the backward process,

$$\log \frac{p_{\rm F}[\gamma|\gamma_0]}{p_{\rm B}[\tilde{\gamma}|\tilde{\gamma}_0]} = -\beta \mathscr{Q}[\gamma, \lambda] + \Delta \mathscr{S}_{\rm P}[\gamma, \lambda], \tag{4.11}$$

where $\gamma_0 = \boldsymbol{x}(t_i)$ is the initial state of the forward trajectory, whereas $\tilde{\gamma}_0 = \boldsymbol{x}^*(t_f)$ is the initial state of the backward trajectory. Note that we used λ to represent the dynamic protocol of the forward process $\lambda^{\rm F}(t)$. As carefully explained in Ref. [49], Eq. (4.11), as well as its extensions (4.14) and (4.18), are better understood in their discretized forms.

The unconditional path probabilities of the forward and backward processes can be constructed by multiplying the conditional path probabilities and the pdf of the initial states, given respectively by Eqs. (4.1) and (4.2),

$$p_{\rm F}[\gamma] = p_{\rm F}[\gamma|\gamma_0]\, e^{-U(\boldsymbol{x}(t_i), \lambda_i)}, \tag{4.12}$$

$$\begin{aligned} p_{\rm B}[\tilde{\gamma}] &= p_{\rm B}[\tilde{\gamma}|\tilde{\gamma}_0]\, e^{-U(\boldsymbol{x}^*(t_f), \lambda_f^*)} \\ &= p_{\rm B}[\tilde{\gamma}|\tilde{\gamma}_0]\, e^{-U(\boldsymbol{x}(t_f), \lambda_f)}, \end{aligned} \tag{4.13}$$

where in the last equality we used the detailed balance condition (3.4c) for $U(\boldsymbol{x}, \lambda)$. Taking the ratio of the preceding two equations and further using Eq. (4.11), we obtain

$$\log \frac{p_F[\gamma]}{p_B[\tilde{\gamma}]} = \Delta U[\gamma, \lambda] - \beta \mathscr{Q}[\gamma, \lambda] + \mathscr{S}_{\rm P}[\gamma, \lambda], \tag{4.14}$$

where $\Delta U[\gamma, \lambda]$ is the difference of $U(\boldsymbol{x}, \lambda)$ between the final and initial state of the forward trajectory,

$$\begin{aligned} \Delta U[\gamma, \lambda] &= U(\boldsymbol{x}(t_f), \lambda_f) - U(\boldsymbol{x}(t_i), \lambda_i) \\ &= \beta (H(\boldsymbol{x}(t_f), \lambda_f) - H(\boldsymbol{x}(t_i), \lambda_i) - \Delta F(\lambda)), \end{aligned} \tag{4.15}$$

where in the last step we have used Eq. (2.7), and $\Delta F(\lambda)$ is the difference of equilibrium free energies between the final and the initial states,

$$\beta \Delta F(\lambda) = F(\lambda_f) - F(\lambda_i). \tag{4.16}$$

Now applying the first law of thermodynamics along the forward trajectory we find

$$H(\boldsymbol{x}(t_f), \lambda_f) - H(\boldsymbol{x}(t_i), \lambda_i) = \mathscr{W}[\gamma, \lambda] + \mathscr{Q}[\gamma, \lambda], \tag{4.17}$$

we can relate the unconditional path probability of the forward process to that of the backward process,

$$\log \frac{p_F[\gamma]}{p_B[\tilde{\gamma}]} = \beta \mathscr{W}[\gamma, \lambda] - \beta \Delta F(\lambda) + \mathscr{S}_{\rm P}[\gamma, \lambda]. \tag{4.18}$$

With the understanding that both system and bath are in equilibrium in the initial and final state of the process, the r.h.s. can be understood as the change of total entropy along the trajectory γ in the forward process. The first term is the change of system entropy and bath entropy, whereas the second term is the entropy pumped out of the system by the agent who supplies the non-conservative force. In the absence of entropy pumping, Eq. (4.18) reduces to Eq. (4.33) of Ref. [49].

We can now define the joint pdfs of integrated work and pumped entropy in the forward and backward processes,

$$p_{\rm F}(W, S_{\rm P}) \equiv \int \mathscr{D}[\gamma]\ p_{\rm F}[\gamma]\, \delta(W - \mathscr{W}[\gamma, \lambda]) \delta(S_{\rm P} - \mathscr{S}_{\rm P}[\gamma, \lambda]), \tag{4.19}$$

$$p_{\rm B}(W, S_{\rm P}) \equiv \int \mathscr{D}[\tilde{\gamma}]\ p_{\rm B}[\tilde{\gamma}]\, \delta(W - \mathscr{W}[\tilde{\gamma}, \tilde{\lambda}]) \delta(S_{\rm P} - \mathscr{S}_{\rm P}[\tilde{\gamma}, \tilde{\lambda}]). \tag{4.20}$$

Using Eqs. (4.18) and (4.10), as well as the fact that the functional integral measure is invariant under time reversal: $\mathscr{D}[\gamma] = \mathscr{D}[\tilde{\gamma}]$, we can rewrite Eq. (4.19) into

$$p_{\rm F}(W, S_{\rm P}) = \int \mathscr{D}[\gamma]\ p_{\rm B}[\tilde{\gamma}] \frac{p_{\rm F}[\gamma]}{p_{\rm B}[\tilde{\gamma}]}\, \delta(W - \mathscr{W}[\gamma, \lambda]) \delta(S_{\rm P} - \mathscr{S}_{\rm P}[\gamma, \lambda]) \tag{4.21}$$

$$= \int \mathscr{D}[\tilde{\gamma}]\ p_{\rm B}[\tilde{\gamma}] e^{\beta W[\gamma, \lambda] - \beta \Delta F(\lambda) + S_{\rm P}[\gamma, \lambda]}\, \delta(W - \mathscr{W}[\gamma, \lambda]) \delta(S_{\rm P} - \mathscr{S}_{\rm P}[\gamma, \lambda]) \tag{4.22}$$

$$= e^{\beta W - \beta \Delta F(\lambda) + S_{\rm P}} \int \mathscr{D}[\tilde{\gamma}]\ p_{\rm B}[\tilde{\gamma}]\, \delta(W + \mathscr{W}[\tilde{\gamma}, \tilde{\lambda}]) \delta(S_{\rm P} + \mathscr{S}_{\rm P}[\tilde{\gamma}, \tilde{\lambda}]) \tag{4.23}$$

$$= e^{\beta W - \beta \Delta F(\lambda) + S_{\rm P}} p_{\rm B}(-W, -S_{\rm P}). \tag{4.24}$$

We obtain a fluctuation theorem for joint pdfs of W, $S_{\rm P}$,

$$p_{\rm F}(W, S_{\rm P}) e^{-S_{\rm P} - \beta W + \beta \Delta F} = p_{\rm B}(-W, -S_{\rm P}). \tag{4.25}$$

The total entropy production at the trajectory level is $\Delta S^{\rm tot} \equiv \beta W - \beta \Delta F + S_{\rm P}$, we can obtain from Eq. (4.25) the fluctuation theorem for $\Delta S^{\rm tot}$,

$$p_{\rm F}(\Delta S^{\rm tot}) e^{-\Delta S^{\rm tot}} = p_{\rm B}(-\Delta S^{\rm tot}). \tag{4.26}$$

If we integrate out $\mathcal{S}_{\rm P}$ in both sides, we obtain

$$p_{\rm F}(W) \langle e^{-S_{\rm P}} \rangle_W e^{-\beta W + \beta \Delta F} = p_{\rm B}(-W), \tag{4.27}$$

where $\langle e^{-S_{\rm P}} \rangle_W$ means the conditional average of $e^{-S_{\rm P}}$ given W. If we integrate our W in both sides, we obtain

$$p_{\rm F}(S_{\rm P}) \langle e^{-\beta W} \rangle_{S_{\rm P}} e^{-S_{\rm P} + \beta \Delta F} = p_{\rm B}(-S_{\rm P}), \tag{4.28}$$

where $\langle e^{-\beta W} \rangle_{S_{\rm P}}$ means the conditional average of $e^{-\beta W}$ given $S_{\rm P}$. If we integrate out both W and $S_{\rm P}$, we obtain the generalized Jarzynski equality,

$$\langle e^{-\beta W - \Delta S_{\rm P}} \rangle = e^{-\beta \Delta F}. \tag{4.29}$$

Fluctuation theorem and work identity in the presence of entropy pumping were first derived by Qian and Kim in Ref. [38]. Note however our result Eq. (4.25) is slightly more general than the results of Ref. [38]. Similar fluctuation theorems have also been obtained in the study of information heat engines [62,63].

B. Bochkov-Kuzovlev gauge

Let us now consider a gauge transformation as defined in Eqs. (3.5), which keeps the Langevin equation invariant. In Sec. III K we have already shown that heat, pumped entropy, and the formulas (3.73) and (4.11) are all gauge invariant.

By strong contrast, system Hamiltonian, work, and free energy all change under generic gauge transformation, as demonstrated in Sec. III K. Consequently, the initial pdfs of the forward and backward processes, Eqs. (4.1) and (4.2), also change under the gauge transformation: we are studying different physical processes as we change the definition of system Hamiltonian. It then follows that fluctuation theorems and work identity (4.25), (4.27), (4.28), and (4.29), all change under gauge transformation. As we continuously change of the definition of system Hamiltonian, we obtain a continuous manifold of fluctuation theorems and work identities, which are physically distinct from each other.

An extreme choice of gauge is to let the Hamiltonian H to be completely fixed and independent of λ, i.e., $d_\lambda H = 0$. This implies that the free energy is also fixed, $\Delta F = 0$. The conservative force then remains fixed, with the control parameter only appears in the non-conservative force f_i. We will call this *Bochkov-Kuzovlev (BK) gauge*. In this gauge, the excess work vanishes identically. The total work and heat are then

$$d\!\!\!^{-}\mathscr{W}^{\rm BK} = f_i^{\rm BK} \circ dx^i, \tag{4.30}$$

$$d\!\!\!^{-}\mathscr{Q}^{\rm BK} = d_{\boldsymbol{x}} H^{\rm BK} - f_i^{\rm BK} \circ dx^i. \tag{4.31}$$

Equations (4.25), (4.27), (4.28), and (4.29) then become

$$p_{\rm F}(W, S_{\rm P}) e^{-\beta W - S_{\rm P}} = p_{\rm B}(-W, -S_{\rm P}), \tag{4.32}$$

$$p_{\rm F}(W) \langle e^{-S_{\rm P}} \rangle_W e^{-\beta W} = p_{\rm B}(-W), \tag{4.33}$$

$$p_{\rm F}(S_{\rm P}) \langle e^{-\beta W} \rangle_{S_{\rm P}} e^{-S_{\rm P}} = p_{\rm B}(-S_{\rm P}), \tag{4.34}$$

$$\langle e^{-\beta W - \Delta S_{\rm P}} \rangle = 1. \tag{4.35}$$

Equation (4.35) is called Bochkov-Kuzovlev (BK) equality [65]. Jarzynski studied an example [66] driven by conservative force, and show that BK equality is different from Jarzynski equality, even though these equalities concern the same physical processes.

C. Fluctuation theorem for excess EP

Let us first consider the case where the control parameter is fixed. We define the *adjoint process* (or dual process), with respect to the forward process, using the following relation of path probability distributions:

$$p_{\rm F}[\gamma] = p_{\rm Ad}[\tilde{\gamma}], \tag{4.36}$$

where $\tilde{\gamma} = \boldsymbol{x}^*(-t)$ is the time reversal of the path $\gamma = \boldsymbol{x}(t)$ as defined in Sec. IV A. The adjoint process is generated by a Fokker-Planck operator $\mathscr{L}_{\rm Ad}$, which will be derived below. Since the forward process is stationary, so is the adjoint process. In particular, the adjoint process starts from the NESS associated with its own FPO. Equation (4.36) implies the following relation between two-time joint pdfs:

$$p_{\rm F}(\boldsymbol{x} + d\boldsymbol{x}, dt; \boldsymbol{x}, 0) = p_{\rm Ad}(\boldsymbol{x}^*, dt; \boldsymbol{x}^* + d\boldsymbol{x}^*, 0). \tag{4.37}$$

If we integrating both sides over $\boldsymbol{x}_2$, we obtain a relation between *the steady state pdfs* of the forward and adjoint processes,

$$p_{\rm F}^{\rm SS}(\boldsymbol{x}) = e^{-U(\boldsymbol{x})} = p_{\rm Ad}^{\rm SS}(\boldsymbol{x}^*) = e^{-\bar{U}(\boldsymbol{x}^*)}. \tag{4.38}$$

This leads to the ration between the generalized potentials of two processes, both in Gibbs gauge,

$$U(\boldsymbol{x}) = \bar{U}(\boldsymbol{x}^*). \tag{4.39}$$

Strictly speaking, we should use notations U^G and $\bar{U}^G$ since we are using Gibbs gauge. Throughout this subsection, however, we always adopt Gibbs gauge but neglect the superscript G in order to simplify the notations. Using Markov property, Eq. (4.37) can now be written in terms of one-time pdfs and transition pdfs,

$$\begin{aligned} p_{\rm F}(\boldsymbol{x} + d\boldsymbol{x} | \boldsymbol{x}; dt) e^{-U(\boldsymbol{x})} &= p_{\rm Ad}(\boldsymbol{x} | \boldsymbol{x}^* + d\boldsymbol{x}^*; dt) e^{-\bar{U}(\boldsymbol{x}^* + d\boldsymbol{x}^*)} \\ &= p_{\rm Ad}(\boldsymbol{x} | \boldsymbol{x}^* + d\boldsymbol{x}^*; dt) e^{-U(\boldsymbol{x} + d\boldsymbol{x})}, \end{aligned} \tag{4.40}$$

which can be further written as

$$\log \frac{p_{\rm F}(\boldsymbol{x} + d\boldsymbol{x} | \boldsymbol{x}; dt)}{p_{\rm Ad}(\boldsymbol{x} | \boldsymbol{x}^* + d\boldsymbol{x}^*; dt)} = -d_{\boldsymbol{x}} U(\boldsymbol{x}). \tag{4.41}$$

To find the Fokker-Planck operator for the adjoint process, it is more convenient to use the parametrization of Ito-Langevin equation discussed in Sec. III B, i.e., Eqs. (3.18) and (3.19). Adapting the derivation in Sec. III A of Ref. [48] of this sequel, we can easily show that the Fokker-Planck operator of the adjoint process (which was called the *tilde process* there), is characterized by generalized potential $\bar{U}$ and kinetic matrix $\bar{\tilde{L}}^{ij}$:

$$\bar{\tilde{L}}^{ij} = \bar{B}^{ij} + \bar{\tilde{Q}}^{ij}, \tag{4.42}$$

which are related to those of the forward process via

$$\bar{U}(\boldsymbol{x}, \lambda) \equiv U(\boldsymbol{x}^*, \lambda), \tag{4.43a}$$

$$\bar{B}^{ij}(\boldsymbol{x},\lambda) \equiv \epsilon_i B^{ij}(\boldsymbol{x}^*,\lambda)\epsilon_j = B^{ij}(\boldsymbol{x},\lambda^*), \tag{4.43b}$$

$$\bar{\hat{Q}}^{ij}(\boldsymbol{x},\lambda) \equiv -\epsilon_i \hat{Q}^{ij}(\boldsymbol{x}^*,\lambda)\epsilon_j. \tag{4.43c}$$

Note that Eq. (4.43a) is a rewriting of Eq. (4.39). Note that in the second equality of Eq. (4.43b), we have used the detailed balance property of B^{ij}, Eq. (3.4a). If the forward process has detailed balance and is conservative, the detailed balance properties Eqs. (2.5) must hold (with $\hat{Q}^{ij}$ replacing Q^{ij}). Then we may use Eqs. (2.5) to rewrite the r.h.s. of Eqs. (4.43), which yields

$$\bar{B}^{ij}(\boldsymbol{x},\lambda) \equiv B^{ij}(\boldsymbol{x},\lambda^*), \tag{4.44a}$$

$$\bar{\hat{Q}}^{ij}(\boldsymbol{x},\lambda) \equiv \hat{Q}^{ij}(\boldsymbol{x},\lambda^*), \tag{4.44b}$$

$$\bar{U}(\boldsymbol{x},\lambda) \equiv U(\boldsymbol{x},\lambda^*), \tag{4.44c}$$

which means that the adjoint process is related to the forward process via time reversal of control parameter. This is in fact defined as *the backward process* for the conservative system.

Using the representation (3.18) and (3.19), the probability current is given by Eq. (3.20). In the steady state it becomes Eq. (4.45),

$$j^i_{\rm SS} = \partial_j(\hat{Q}^{ij}e^{-U}), \tag{4.45}$$

The steady-state current of the adjoint process is obtained by replacing $\hat{Q}^{ij}$ and U respectively by $\bar{\hat{Q}}^{ij}$ and $\bar{U}$ in Eq. (4.45),

$$\bar{j}^i_{\rm SS} = \partial_j(\bar{\hat{Q}}^{ij}e^{-\bar{U}}), \tag{4.46}$$

Combining Eqs. (4.45) and (4.46) with Eqs. (4.43), it is easy to prove the following symmetry property:

$$\bar{j}^i_{\rm SS}(\boldsymbol{x}) = -\epsilon_i j^i_{\rm SS}(\boldsymbol{x}^*,\lambda). \tag{4.47}$$

In fact, Eqs. (4.43a) and (4.47) can be deemed as an alternative definition of the adjoint process. For models with only even variables and even control parameter, they reduce to

$$\bar{U}(\boldsymbol{x}) = V(\boldsymbol{x}), \tag{4.48}$$

$$\bar{j}^i_{\rm SS}(\boldsymbol{x}) = -j^i_{\rm SS}(\boldsymbol{x},\lambda). \tag{4.49}$$

Hence the adjoint process and the forward process share the same pdf by opposite probability current in their steady states. This agrees with the definition of dual process in Ref. [27–30].

Now if we take the alternative representation of the forward process in terms of U, L^{ij}, φ_i, as given in Eqs. (3.1), the probability current is given by Eq. (3.7). In the steady state we have

$$j^i_{\rm SS} = L^{ij}\varphi_j e^{-U} + \partial_j Q^{ij} e^{-U}. \tag{4.50}$$

Equating Eqs. (4.50) with (4.45), we recover the relation (3.21). Likewise, the steady-state current of the adjoint process can be obtained from Eq. (4.50) by replacing L^{ij}, φ_i and U with $\bar{L}^{ij}$, $\bar{\varphi}_i$ and $\bar{U}$,

$$\bar{j}^i_{\rm SS} = \bar{L}^{ij}\bar{\varphi}_j e^{-\bar{U}} + \partial_j \bar{Q}^{ij} e^{-\bar{U}}. \tag{4.51}$$

Inserting Eqs. (4.50) and (4.51) into Eq. (4.47), and further using Eqs. (4.43), we obtain

$$\epsilon_i\bar{\varphi}_i(\boldsymbol{x}^*) = -[(\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{Q}(\boldsymbol{x}))^{-1}(\boldsymbol{B}(\boldsymbol{x}) + \boldsymbol{Q}(\boldsymbol{x}))\boldsymbol{\varphi}(\boldsymbol{x})]_j, \tag{4.52}$$

which fully determines the non-conservative force in the adjoint process in terms of that in the forward process. Note that this relation holds only in the Gibbs gauge.

Now consider the case where the control parameter varies with time $\lambda^{\rm F}(t) = \lambda(t)$ in the forward process with $t \in [t_i, t_f]$. The *adjoint backward process* is defined in the time interval $t \in [-t_f, -t_i]$, such that

$$U_{\rm AdB}(\boldsymbol{x},t) = \bar{U}(\boldsymbol{x}^*,-t), \tag{4.53}$$

$$B^{ij}_{\rm AdB}(\boldsymbol{x},t) = \bar{B}^{ij}(\boldsymbol{x}^*,-t), \tag{4.54}$$

$$Q^{ij}_{\rm AdB}(\boldsymbol{x},t) = \bar{Q}^{ij}(\boldsymbol{x}^*,-t). \tag{4.55}$$

Hence, it is the adjoint process with time being further reversed. Furthermore, the forward process starts from the NESS associated with the control parameter λ_i,

$$p_{\rm SS}(\boldsymbol{x},\lambda_i) = e^{-U(\boldsymbol{x},\lambda_i)}, \tag{4.56}$$

whereas the adjoint backward process starts from the NESS with pdf

$$\bar{p}_{\rm SS}(\boldsymbol{x},\lambda_f) = e^{-\bar{U}(\boldsymbol{x},\lambda_f)} = e^{-U(\boldsymbol{x}^*,\lambda_f)}. \tag{4.57}$$

which is also the time reversal of the steady state of the forward process with control parameter fixed at λ_f.

Now for each dynamic trajectory $\gamma = \boldsymbol{x}(t)$ in the forward process, there is a corresponding trajectory $\tilde{\gamma} = \boldsymbol{x}^*(-t)$ in the adjoint backward process. We can divided both trajectories into many infinitesimal steps. For each step, we can apply the relation Eq. (4.41). Summing up all these relations, we obtain

$$\begin{aligned}\log\frac{p_{\rm F}[\gamma|\gamma_0]}{p_{\rm AdB}[\tilde{\gamma}|\tilde{\gamma}_0]} &= -\int_\gamma d_x U(\boldsymbol{x}) \\ &= -\Delta U + \int_\gamma d_\lambda U,\end{aligned} \tag{4.58}$$

where $\gamma_0 = \boldsymbol{x}(t_i)$ and $\tilde{\gamma}_0 = \boldsymbol{x}^*(t_f) = \tilde{\boldsymbol{x}}(-t_f)$ are the initial states of the forward and backward trajectories, whereas ΔU is the total change of U along the forward trajectory,

$$\Delta U[\gamma,\lambda] = U(\boldsymbol{x}(t_f),\lambda_f) - U(\boldsymbol{x}(t_i),\lambda_i) \tag{4.59}$$

We may construct the unconditional path probabilities from the conditional path probabilities. For the forward process we have

$$p_{\rm F}[\gamma] = p_{\rm F}[\gamma|\gamma_0]\, e^{-U(\boldsymbol{x}(t_i),\lambda_i)}, \tag{4.60}$$

whereas for the adjoint backward process we have

$$p_{\rm AdB}[\tilde{\gamma}] = p_{\rm AdB}[\tilde{\gamma}|\tilde{\gamma}_0]\, e^{-U(\boldsymbol{x}(t_f),\lambda_f)} \tag{4.61}$$

Combining Eqs. (4.58) to (4.61), we obtain

$$\begin{aligned}\log\frac{p_{\rm F}[\gamma]}{p_{\rm AdB}[\tilde{\gamma}]} &= \int_\gamma d_\lambda U^G = \beta\int_\gamma d_\lambda(H^G - F_{\rm SS}) \\ &= \beta W^{\rm ex}[\gamma] - \beta\Delta F_{\rm SS},\end{aligned} \tag{4.62}$$

where in the second equality we have used Eq. (3.12), and $F_{\rm SS}$ is the steady-state free energy defined in Eq. (3.13). The r.h.s. of Eq. (4.62) can be understood as the excess EP at the trajectory level.

Now following the same proof as in Sec. IV A, we obtain a fluctuation theorem for the excess work on the forward

process,

$$\frac{p_{\rm F}(W^{\rm ex})}{p_{\rm AdB}(-W^{\rm ex})} = e^{\beta W^{\rm ex} - \beta \Delta F}, \tag{4.63}$$

where $p_{\rm AdB}(-W^{\rm ex})$ is the pdf of the excess work in the adjoint backward process. From this result we easily obtain the Hatano-Sasa equality, Eq. (3.89).

V. EXAMPLES OF ASYMMETRIC SYSTEMS

We discuss several examples of nonlinear Langevin systems driven by non-conservative forces with asymmetric kinetic matrices. The main purpose is to illustrate the descriptive power of our theory. Hence, analyses of these examples will be rather brief. Nonetheless, all presented results can be verified directly using our theory. Some of these examples will be studied in greater detail in future papers.

A. Weakly damped Hamiltonian systems

We consider a Hamiltonian system with Hamiltonian $\boldsymbol{p}^2/2m + V(\boldsymbol{x})$ and driven by a non-conservative force $\boldsymbol{f}$. The system is also in contact with a heat bath. The Langevin equations are

$$d\boldsymbol{r} = \frac{\boldsymbol{p}}{m} dt, \tag{5.1a}$$

$$d\boldsymbol{p} = \left(-\nabla V + \boldsymbol{f} - \gamma \frac{\boldsymbol{p}}{m}\right) dt + \sqrt{2\gamma T}\, d\mathbf{W}, \tag{5.1b}$$

where $-\nabla V$ is the conservative force, whilst $\boldsymbol{f}$ is the non-conservative force, which possibly depends on momentum. The case $\partial_{\boldsymbol{p}} \cdot \boldsymbol{f} \neq 0$ was first studied by Kim and Qian in Ref. [37,38], where the notion of entropy pumping was first discussed. Equations (5.1) can be written into the standard form (3.1) with

$$U = \beta H - \beta F = \beta \left(\frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r}, \lambda) - F \right), \tag{5.2}$$

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{p} \end{pmatrix}, \quad \boldsymbol{\varphi} = \begin{pmatrix} \beta \boldsymbol{f} \\ 0 \end{pmatrix}, \quad \boldsymbol{L} = T \begin{pmatrix} 0 & -\boldsymbol{I} \\ \boldsymbol{I} & \gamma \boldsymbol{I} \end{pmatrix}, \tag{5.3}$$

$$\boldsymbol{B} = T \begin{pmatrix} 0 & 0 \\ 0 & \gamma \boldsymbol{I} \end{pmatrix}, \quad \boldsymbol{Q} = T \begin{pmatrix} 0 & -\boldsymbol{I} \\ \boldsymbol{I} & 0 \end{pmatrix}. \tag{5.4}$$

The pumped entropy (3.73c) is given by

$$d\mathcal{S}_{\rm P} = -(\partial_{\boldsymbol{p}} \cdot \boldsymbol{f})\, dt. \tag{5.5}$$

To avoid confusion, we use $\rho(\boldsymbol{r}, \boldsymbol{p})$ instead of $p(\boldsymbol{r}, \boldsymbol{p})$ to denote the pdf. The total EP can be obtained from Eq. (3.78a),

$$\Sigma^{\rm tot} = \gamma T \int_{\boldsymbol{r},\boldsymbol{p}} \rho(\boldsymbol{r}, \boldsymbol{p}) \left(\partial_{\boldsymbol{p}} \log p + \frac{\beta \boldsymbol{p}}{m} \right)^2, \tag{5.6}$$

which is non-negative, and vanishes if and only if the momentum distribution is Maxwellian,

$$\rho(\boldsymbol{r}, \boldsymbol{p}) = g(\boldsymbol{r})\, e^{-\boldsymbol{p}^2/2mT}. \tag{5.7}$$

It is important to note that neither conservative nor non-conservative force appears in Eq. (5.6). By contrast, the FPE is

$$\partial_t \rho = \left[-\frac{\boldsymbol{p}}{m} \nabla_{\boldsymbol{r}} + \nabla_{\boldsymbol{p}} \cdot (\nabla_{\boldsymbol{r}} V - \boldsymbol{f}) + \frac{\gamma}{m} \nabla_{\boldsymbol{p}} \cdot \boldsymbol{p} + \gamma T \nabla_{\boldsymbol{p}}^2 \right] \rho, \tag{5.8}$$

which does explicitly depend on the forces.

Let us consider the simplest example, where $V = 0$ and $\boldsymbol{f}$ is constant. We impose periodic boundary condition on the coordinates, so that there is no globally defined potential, which generates $\boldsymbol{f}$. The Langevin equations then become linear and can be exactly solved. In particular, the steady state pdf can be easily found,

$$\rho^{\rm SS}(\boldsymbol{x}, \boldsymbol{p}) = \frac{e^{-(\boldsymbol{p} - m\boldsymbol{f}/\gamma)^2/2mT}}{(2\pi m T)^{3/2} \mathscr{V}}, \tag{5.9}$$

where $\mathscr{V}$ is the volume of the coordinate space. Substituting this result into Eq. (5.6) and integrating, we find that the total EP is

$$\Sigma = \frac{\boldsymbol{f}^2}{\gamma T} = \frac{\boldsymbol{f} \cdot \boldsymbol{v}_0}{T}, \tag{5.10}$$

where $\boldsymbol{v}_0 = \boldsymbol{f}/\gamma$ is the average velocity of the particle. The pumped entropy vanishes identically. The excess EP can be obtained from Eq. (3.64),

$$\Sigma^{\rm ex} = \gamma T \int_{\boldsymbol{r},\boldsymbol{p}} \rho(\boldsymbol{r}, \boldsymbol{p}) \left(\partial_{\boldsymbol{p}} \log p + \frac{\beta(\boldsymbol{p} - m\boldsymbol{f}/\gamma)}{m} \right)^2, \tag{5.11}$$

which is non-negative and is minimized at the NESS. The house-keeping EP is then

$$\Sigma^{\rm hk} = \Sigma^{\rm tot} - \Sigma^{\rm ex}, \tag{5.12}$$

which may be either positive or negative.

Let us now look at the over-damped limit of this problem. The Langevin equation becomes

$$d\boldsymbol{r} + \frac{T}{\gamma}(\beta \nabla V - \beta \boldsymbol{f}) dt = \sqrt{\frac{2T}{\gamma}}\, d\boldsymbol{W}, \tag{5.13}$$

which corresponds to the standard form (3.1) with

$$U = \beta V, \quad \boldsymbol{x} = \boldsymbol{r}, \quad \boldsymbol{\varphi} = \beta \boldsymbol{f}, \tag{5.14}$$

$$\boldsymbol{B} = \frac{T}{\gamma} \boldsymbol{I}, \quad \boldsymbol{Q} = 0. \tag{5.15}$$

The EP is

$$\Sigma^{\rm tot}_{\rm od} = \frac{T}{\gamma} \int_{\boldsymbol{r}} \rho(\boldsymbol{x}) (\nabla_{\boldsymbol{r}} \log p - \beta \boldsymbol{f})^2, \tag{5.16}$$

where the subscript od means "over-damped". The square inside the integral can be expanded, and the cross term vanishes identically because $\boldsymbol{f}$ is constant. Hence, we find

$$\Sigma^{\rm tot}_{\rm od} = \frac{\boldsymbol{f}^2}{\gamma T} + \frac{T}{\gamma} \int_{\boldsymbol{r}} \frac{(\nabla \rho)^2}{\rho}. \tag{5.17}$$

In this over-damped theory, the total EP achieves its minimum at the steady state where $\rho = 1/\mathscr{V}$ and $\nabla \rho = 0$.

Inspecting the expressions of EP in under-damped theory, Eq. (5.6) and in over-damped theory, Eq. (5.17), we see that these two functionals are very different. Whereas Eq. (5.6) has a minimum zero, the minimum of Eq. (5.17) has a minimum

value $f^2/\gamma T$. A moment of thought indicates that the dissipation occurs completely in the momentum degree of freedom, and in the over-damped theory, the momentum has already achieved a steady (non-Maxwellian) distribution with finite dissipation. This is why the EP in the over-damped theory has a minimum $f^2/\gamma T$. So is EP minimized at the NESS? Apparently, the answer to this question depends on the level of the description we adopt for the problem. In the under-damped theory, the answer is no, whereas in the over-damped theory, the answer is yes. It is also interesting to comment that whilst in our Fokker-Planck theory, the EP as given by (3.78a) is always non-negative, we may also define EP at the level of dynamic trajectory, as the sum of $-\beta đ\mathcal{Q}$, the pumped entropy, and the change of stochastic entropy. This EP at the trajectory level can take negative values, since there are trajectories where the non-conservative force does negative work.

Let us now consider a slightly more complicated case, where $V = 0$, but the non-conservative force depends on momentum. For simplicity, we only consider one dimensional case, then $f = h'(p)$ can always be written as a derivative. The Fokker-Planck operator can be rewritten into the standard form Eq. (3.1) with

$$U(p) = \beta\left(\frac{p^2}{2m} - \frac{1}{\gamma}h(p)\right) + \mathrm{C}, \tag{5.18}$$

where C is a normalization constant. Hence the steady-state distribution is given by

$$\rho^{\mathrm{SS}}(x,p) = \frac{\exp\left\{\beta\left(\frac{p^2}{2m} - \frac{1}{\gamma}h(p)\right) + \mathrm{C}\right\}}{(2\pi mT)^{3/2}\mathscr{V}}. \tag{5.19}$$

The EP rate and the pumped entropy rate can both be calculated and expressed as average over the steady-state distribution,

$$\Sigma = \frac{\langle f(p)^2\rangle_{\mathrm{SS}}}{\gamma T} \geqslant 0, \tag{5.20}$$

$$\Sigma_{\mathrm{P}} = -\langle f'(p)\rangle_{\mathrm{SS}}, \tag{5.21}$$

$$\frac{\beta đQ}{dt} = -\frac{\beta đW}{dt} = \Sigma - \Sigma_{\mathrm{P}}. \tag{5.22}$$

By choosing appropriate function $f(p)$, we may achieve one of the following three interesting scenarios:

(1) $\Sigma_{\mathrm{P}} = \Sigma > 0$, $đQ/dt = -đW/dt = 0$. In general, the average momentum of Eq. (5.19) is not zero. The system then moves with a steady velocity without any energy input. The total EP is just the entropy pumped out of the system, which is presumably stored as information entropy. The system behaves as an information swimmer [67], which use information entropy to perform steady motion.

(2) $\Sigma_{\mathrm{P}} > 0$, $đQ/dt = -đW/dt > 0$. The system then absorbs heat from the bath, output work, and also entropy is pumped out of the system. The system behaves as a demon who transforms thermal energy into mechanical work.

(3) $\Sigma_{\mathrm{P}} < 0$, $đQ/dt = -đW/dt < 0$. The system behaves as a demon running backwards, transforming information entropy into thermodynamic entropy, as the cost of mechanical work.

More detailed study of this model will be presented in a future paper.

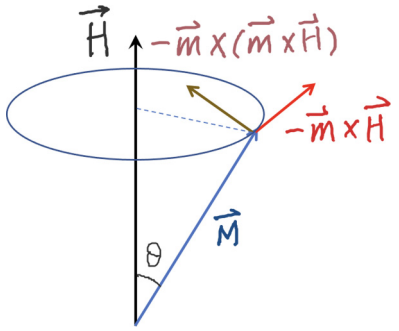

FIG. 3. Schematics of Landau-Lifshitz-Gilbert dynamics.

B. Landau-Lifshitz-Gilbert equation and spin torque

Here we study the dynamics of a magnetic moment inside a uniform magnetic field, possibly driven by a non-conservative force called *spin torque* [68,69]. We start from the Landau-Lifshitz-Gilbert equation (LLG), which describes the dissipative dynamics of magnetic moment $\boldsymbol{m}$ inside an effective magnetic field $\boldsymbol{H}$,

$$\frac{d\boldsymbol{m}}{dt} = -\gamma_0 \boldsymbol{m}\times\boldsymbol{H} - \eta\,\boldsymbol{m}\times(\boldsymbol{m}\times\boldsymbol{H}), \tag{5.23}$$

where $\boldsymbol{m}\times\boldsymbol{H}$ denotes cross product between vectors $\boldsymbol{m}$ and $\boldsymbol{H}$. The first term in r.h.s. gives the precession of the magnetic moment around the magnetic field $\boldsymbol{H}$, whereas the second term describes damping, as illustrated in Fig. 3, with η the damping coefficient. Note that the magnitude $|\boldsymbol{m}|$ is conserved by the dynamics (5.23).

The stochastic version of LLG was derived both in Stratonovich's form [70–72] and in Ito's form [71,72]. Here we use Ito's calculus, which is equivalent but is considerably simpler. We first note that Eq. (5.23) can be written in the form of

$$dm^i + L^{ij}\partial_j U = 0, \tag{5.24a}$$

$$L^{ij} = B^{ij} + Q^{ij}, \tag{5.24b}$$

$$Q^{ij} = T\gamma_0\epsilon^{ijk}m_k, \tag{5.24c}$$

$$B^{ij} = T\eta\left(m^2\delta^{ij} - m^i m^j\right), \tag{5.24d}$$

$$U = -\beta\boldsymbol{m}\cdot\boldsymbol{H}. \tag{5.24e}$$

Note that both $\boldsymbol{m}$ and $\boldsymbol{H}$ are odd under time reversal. It is easy to check that Q^{ij}, B^{ij}, U as given in Eqs. (5.24) satisfy the detailed balance conditions (2.5).

But Eq. (5.24a) is deterministic limit [48] of the covariant Langevin equation (2.1). Adding back the spurious drift and noises, we can recover the full covariant Langevin equation,

$$dm^i + (L^{ij}\partial_j U - \partial_j L^{ij})dt = b^{i\alpha}dW_\alpha(t). \tag{5.25}$$

The matrix $b^{i\alpha}$ is not uniquely determined by Eq. (5.24a). There are many possible which satisfy the condition Eq. (2.2b). The simplest choice is

$$b^{i\alpha} = \frac{\sqrt{2\eta T}}{m}(m^2\delta^{i\alpha} - m^i m^\alpha), \tag{5.26}$$

where $\alpha = 1, 2, 3$, and $m = |\boldsymbol{m}|$ is the magnitude of $\boldsymbol{m}$. Further noticing

$$\partial_j L^{ij} = -2T\eta\, m^i, \tag{5.27}$$

we may rewrite the covariant Langevin equation (5.25) in the following vector form:

$$\begin{aligned} d\boldsymbol{m} =& -\gamma_0 \boldsymbol{m}\times\boldsymbol{H}\,dt - \eta\,\boldsymbol{m}\times(\boldsymbol{m}\times\boldsymbol{H})\,dt \\ & - 2T\eta\,\boldsymbol{m}dt + \sqrt{2\eta T}\,\boldsymbol{m}\times d\boldsymbol{W}, \end{aligned} \tag{5.28}$$

where $d\boldsymbol{W}$ is the 3d vector-valued Wiener noise, and the wedge product $\boldsymbol{m}\times d\boldsymbol{W}$ is interpreted in Ito's sense. Note that $-2T\eta\,\boldsymbol{m}dt$ is the spurious drift.

Using Ito's formula

$$dW_\alpha dW_\beta = \delta_{\alpha\beta}dt, \tag{5.29}$$

and Ito-Langevin equation (5.28), we easily find

$$\begin{aligned} dm^2 &= d(\boldsymbol{m}\cdot\boldsymbol{m}) = 2\boldsymbol{m}\cdot d\boldsymbol{m} + d\boldsymbol{m}\cdot d\boldsymbol{m} \\ &= -4T\eta m^2 dt + 4T\eta m^2 dt = 0. \end{aligned} \tag{5.30}$$

Hence, the magnitude $|\boldsymbol{m}|$ is conserved by the stochastic LLG dynamics. Note that it would not be so if we miss the spurious drift term.

The Fokker-Planck operator associated with the Langevin dynamics Eq. (5.28) can be worked out in a standard way,

$$\begin{aligned} \mathscr{L}_0 =& T\eta(m^2\nabla^2 - m^i m^j\partial_i\partial_j - 2\boldsymbol{m}\cdot\nabla) \\ & + \eta(-m^2\boldsymbol{H}\cdot\nabla + (\boldsymbol{m}\cdot\boldsymbol{H})\boldsymbol{m}\cdot\nabla + 2\boldsymbol{m}\cdot\boldsymbol{H}) \\ & + \gamma_0(\boldsymbol{m}\times\boldsymbol{H})\cdot\nabla. \end{aligned} \tag{5.31}$$

It is easy to verify that this operator has a steady state $e^{\beta\boldsymbol{m}\cdot\boldsymbol{H}}$, which describes the thermodynamic equilibrium.

If the system is not far below the Curie point, the fluctuations of the magnitude of $\boldsymbol{m}$ may not be neglected. The corresponding Langevin equation is usually called *stochastic Landau-Lifshitz-Bloch* equation [73–75], which was formulated in several different formats. Here we again formulate it in the standard form (3.1). To describe the fluctuations of $|\boldsymbol{m}|$ we add a term $a\,(\boldsymbol{m}^2 - \boldsymbol{m}_0^2)^2/2$ to the generalized potential Eq. (5.24e), so that the new generalized potential becomes

$$U = -\beta\boldsymbol{m}\cdot\boldsymbol{H} + \frac{a}{2}\left(\boldsymbol{m}^2 - \boldsymbol{m}_0^2\right)^2. \tag{5.32}$$

We also need to add longitudinal part to the kinetic matrix B^{ij}, which controls the dynamics of $|\boldsymbol{m}|$,

$$B^{ij} = T\eta_\perp(m^2\delta^{ij} - m^i m^j) + T\eta_\parallel m^i m^j, \tag{5.33}$$

$$b^{i\alpha} = \frac{\sqrt{2\eta_\perp T}}{m}(m^2\delta^{i\alpha} - m^i m^\alpha) + \frac{\sqrt{2\eta_\parallel T}}{m}m^i m^\alpha, \tag{5.34}$$

where $\eta_\perp$ and $\eta_\parallel$ are respectively the damping coefficients in directions perpendicular and parallel to $\boldsymbol{m}$. Q^{ij} remains the same as in Eqs. (5.24) The resulting covariant Langevin equation then becomes

$$\begin{aligned} d\boldsymbol{m} =& \gamma_0 T\,(\boldsymbol{m}\times\nabla U)dt \\ & + \eta_\perp T\,\boldsymbol{m}\times(\boldsymbol{m}\times\nabla U)dt + \eta_\parallel T\,\boldsymbol{m}(\boldsymbol{m}\cdot\nabla U)\,dt \\ & - 2T(\eta_\perp - 2\eta_\parallel)\,\boldsymbol{m}dt \\ & + \sqrt{2\eta_\perp T}\,\boldsymbol{m}\times d\boldsymbol{W} + \sqrt{2\eta_\parallel T}\,\boldsymbol{m}\cdot d\boldsymbol{W}, \end{aligned} \tag{5.35}$$

where ∇ means gradient with respect to $\boldsymbol{m}$.

We shall now add a non-conservative force into Eq. (5.28), which has the particular form $\boldsymbol{f} = \boldsymbol{m}\times\boldsymbol{P}$. It describes the torque acting the magnetization due to injection of spin current from outside, and is called *spin torque* [68,69]. The equation with spin torque is obtained from Eq. (5.28) via the replacement

$$\boldsymbol{H} \to \boldsymbol{H} + \boldsymbol{f} = \boldsymbol{H} + \boldsymbol{m}\times\boldsymbol{P}. \tag{5.36}$$

It is clear that $\boldsymbol{f} = \boldsymbol{m}\times\boldsymbol{P}$ cannot be written as the gradient of a potential, and hence is non-conservative. To satisfy detailed balance (3.6b), the vector $\boldsymbol{P}$ must be even under time reversal. This is in fact guaranteed by unit analysis.

The stochastic LLG with spin torque, which may be called Landau-Lifshitz-Gilbert-Slonczewski equation (LLGS), is then

$$\begin{aligned} d\boldsymbol{m} =& -\gamma_0\boldsymbol{m}\times(\boldsymbol{H} + \boldsymbol{m}\times\boldsymbol{P})\,dt \\ & - \eta\,\boldsymbol{m}\times(\boldsymbol{m}\times(\boldsymbol{H} + \boldsymbol{m}\times\boldsymbol{P}))\,dt \\ & - 2T\eta\,\boldsymbol{m}dt + \sqrt{2\eta T}\,\boldsymbol{m}\times d\boldsymbol{W}, \end{aligned} \tag{5.37}$$

For simplicity let us assume that $\boldsymbol{P}$ is independent of $\boldsymbol{m}$. The pumped entropy term can then be easily calculated,

$$d\mathcal{S}_{\rm P} = -Q^{ij}(\partial_i\varphi_j)\,dt = 2\,\gamma_0\boldsymbol{m}\cdot\boldsymbol{P}\,dt. \tag{5.38}$$

Hence in the presence of non-conservative force, the system constantly exchange entropy with its environment.

The FP operator associated with Eq. (5.37) is

$$\mathscr{L} = \mathscr{L}_0 + \delta\mathscr{L}, \tag{5.39}$$

where $\mathscr{L}_0$ is given in Eq. (5.28) and $\delta\mathscr{L}$ is due to the non-conservative force,

$$\begin{aligned} \delta\mathscr{L} =& \gamma_0(-m^2\boldsymbol{P}\cdot\nabla + (\boldsymbol{m}\cdot\boldsymbol{P})\boldsymbol{m}\cdot\nabla + 2\boldsymbol{m}\cdot\boldsymbol{P}) \\ & - \eta m^2(\boldsymbol{m}\times\boldsymbol{P})\cdot\nabla. \end{aligned} \tag{5.40}$$

The total EP can be calculated using (5.24), (5.36) back into Eq. (3.78a),

$$\begin{aligned} \Sigma =& T\eta\langle m^2(\nabla\log p - \beta\boldsymbol{H} - \beta\boldsymbol{f})^2 \\ & - (\boldsymbol{m}\cdot(\nabla\log p - \beta\boldsymbol{H} - \beta\boldsymbol{f}))^2\rangle, \end{aligned} \tag{5.41}$$

where $\langle\cdot\rangle$ means $\int_{\boldsymbol{m}}\cdot\,p(\boldsymbol{m})$. Depending on the magnitude and direction of $\boldsymbol{P}$, the system may behave as either a dissipator, a demon, or an anti-demon. A systematic study of this model will be supplied in a future paper.

C. Temperature gradient

Up to now we have assumed that the system is in contact with a single heat bath above. This assumption is in fact not necessary. Our Langevin equation with non-conservative forces, Eq. (3.1), can be used to describe systems coupled to multiple heat baths with different temperatures. The temperature difference behaves as non-conservative forces after appropriate transformation of variables. This is demonstrated by a toy model of heat transport shown in Fig. 4. The total Hamiltonian is

$$\begin{aligned} H =& \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{p_3^2}{2m_3} \\ & + V_1(q_1 - q_3) + V_2(q_2 - q_3) + V_3(q_3;\lambda), \end{aligned} \tag{5.42}$$

where the control parameter λ is only coupled to q_3. The under-damped Langevin equations are

$$\dot{q}_1 = \frac{\partial H}{\partial p_1} = \frac{p_1}{m_1}, \tag{5.43a}$$

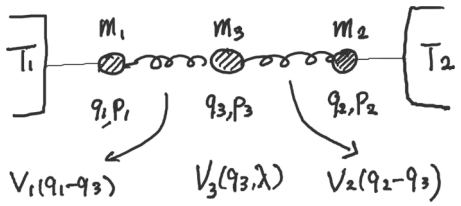

FIG. 4. A toy model of heat transport. An elastic solid consisting of three particles connected by springs. Particles 1 and 2 are respectively in contact with heat bath T_1 and T_2, whereas the particle 3 is not in contact with any heat bath. The control parameter λ only appears in $V_3(q_3;\lambda)$.

$$\dot{q}_2 = \frac{\partial H}{\partial p_2} = \frac{p_2}{m_2}, \tag{5.43b}$$

$$\dot{q}_3 = \frac{\partial H}{\partial p_3} = \frac{p_3}{m_3}, \tag{5.43c}$$

$$\dot{p}_1 = -\gamma_1 \frac{\partial H}{\partial p_1} - \frac{\partial H}{\partial q_1} + \sqrt{2\gamma_1 T_1}\xi_1, \tag{5.43d}$$

$$\dot{p}_2 = -\gamma_2 \frac{\partial H}{\partial p_2} - \frac{\partial H}{\partial q_2} + \sqrt{2\gamma_2 T_2}\xi_2, \tag{5.43e}$$

$$\dot{p}_3 = -\frac{\partial H}{\partial q_3} = -\frac{\partial (V_1 + V_2 + V_3)}{\partial q_3}, \tag{5.43f}$$

where ξ_1, ξ_2 are usual Gaussian white noises

$$\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t-t'), \quad i = 1, 2. \tag{5.44}$$

Stochastic thermodynamics for such system can be developed by generalizing the theory discussed in Ref. [49]. Within this theory the entropy change of environment at the trajectory level is given by

$$d\mathcal{S}^{\rm EV} = -\beta_1 đ\mathcal{Q}_1 - \beta_2 đ\mathcal{Q}_2, \tag{5.45a}$$

$$đ\mathcal{Q}_1 = d_{(q_1,p_1)}H, \tag{5.45b}$$

$$đ\mathcal{Q}_2 = d_{(q_2,p_2)}H, \tag{5.45c}$$

where $đ\mathcal{Q}_1$ and $đ\mathcal{Q}_2$ are respectively the heat released by the heat baths T_1 and T_2. The definition of work remains the same $đ\mathcal{W} = d_\lambda V_3$.

Let us define the generalized potential as

$$\begin{aligned} U = {} & \beta_1 \frac{p_1^2}{2m_1} + \bar{\beta}_1 V_1(q_1 - q_3) \\ & + \beta_2 \frac{p_2^2}{2m_2} + \bar{\beta}_2 V_2(q_2 - q_3) \\ & + \beta_3 \frac{p_3^2}{2m_3} + \beta_3 V_3(q_3;\lambda)), \end{aligned} \tag{5.46a}$$

and define the non-conservative forces as

$$\varphi_{q_1} = (\bar{\beta}_1 - \beta_1)\frac{\partial V_1}{\partial q_1}, \tag{5.46b}$$

$$\varphi_{q_2} = (\bar{\beta}_2 - \beta_2)\frac{\partial V_2}{\partial q_2}, \tag{5.46c}$$

$$\varphi_{q_3} = (\bar{\beta}_1 - \beta_3)\frac{\partial V_1}{\partial q_3} + (\bar{\beta}_2 - \beta_3)\frac{\partial V_2}{\partial q_3}, \tag{5.46d}$$

$$\varphi_{p_1} = \varphi_{p_2} = \varphi_{p_3} = 0. \tag{5.46e}$$

Further defining $\boldsymbol{x} = (q_1, p_1, q_2, p_2, q_3, p_3)$, and the matrix of kinetic coefficients as

$$\boldsymbol{L} = \begin{pmatrix} 0 & -T_1 & 0 & 0 & 0 & 0 \\ T_1 & \gamma_1 T_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -T_2 & 0 & 0 \\ 0 & 0 & T_2 & \gamma_2 T_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -T_3 \\ 0 & 0 & 0 & 0 & T_3 & 0 \end{pmatrix}, \tag{5.47}$$

we can rewrite Eqs. (5.43) into the standard form of non-conservative Langevin equation (3.1),

$$\dot{q}_1 = T_1 \frac{\partial U}{\partial p_1}, \tag{5.48a}$$

$$\dot{q}_2 = T_2 \frac{\partial U}{\partial p_2}, \tag{5.48b}$$

$$\begin{aligned} \dot{q}_3 &= T_3 \frac{\partial U}{\partial p_3}, \\ \dot{p}_1 &= -\gamma_1 T_1 \frac{\partial U}{\partial p_1} - T_1\left(\frac{\partial U}{\partial q_1} - \varphi_{q_1}\right) + \sqrt{2\gamma_1 T_1}\xi_1, \\ \dot{p}_2 &= -\gamma_2 T_2 \frac{\partial U}{\partial p_2} - T_2\left(\frac{\partial U}{\partial q_2} - \varphi_{q_2}\right) + \sqrt{2\gamma_2 T_2}\xi_2, \\ \dot{p}_3 &= -T_3\left(\frac{\partial U}{\partial q_3} - \varphi_{q_3}\right). \end{aligned} \tag{5.48c}$$

Note that the "temperatures" $\bar{\beta}_1, \bar{\beta}_2, \beta_3$ remain arbitrary. They are the gauge parameters, which reflect the arbitrariness we discussed in Sec. III K. It is easy to verify that these equations are mathematically equivalent to Eqs. (5.43).

It is easy to verify that there is no entropy pumping in this problem. Hence the entropy change of the environment associated with Eqs. (5.48) is given by Eq. (3.73b), which we rewrite explicitly

$$d\mathcal{S}^{\rm EV} = -d_{\boldsymbol{x}} U + \varphi_i \circ dx^i. \tag{5.49}$$

Using Eqs. (5.46) it is easy to verify that Eqs. (5.45a) and (5.49) are equivalent. Hence, the total EPs in two representations are identical. Variation of q_3, p_3 makes no contribution to $dS^{\rm EV}$ because these variables do not couple to any heat bath. We will not elaborate further on the details of this problem.

D. Charged particle in magnetic field

Lorentz force cannot be understood as a usual conservative force, which is the gradient of certain potential energy. On the other hand, Lorentz force can neither be understood as usual non-conservative force, since it does not do any work, and does not dissipate any energy. In usual Hamiltonian mechanics, the effects of magnetic field are taken into account by defining canonical momenta, which are linear combination of mechanical momentum and vector potential. Here we show that magnetic field can be taken into account by introducing a new component of the antisymmetric kinetic matrix $\boldsymbol{Q}$.

Consider a charged particle in a uniform magnetic field and also coupled to a damping environment, which is assumed to be uncharged and hence not influenced by the magnetic field.

The Langevin equations can be obtained by incorporating the Lorentz force into Eq. (5.1),

$$d\boldsymbol{r} = \frac{\boldsymbol{p}}{m}dt, \tag{5.50a}$$

$$d\boldsymbol{p} = \left(-\nabla V + q\,\boldsymbol{v}\times\mathbf{H} - \gamma\frac{\boldsymbol{p}}{m}\right)dt + \sqrt{2\gamma T}\,d\mathbf{W}, \tag{5.50b}$$

where $\mathbf{H}$ is the constant magnetic field. These equations can be rewritten into the standard form of covariant Ito-Langevin equations, (3.1), by defining

$$U = \beta\left(\frac{\boldsymbol{p}^2}{2m} - F\right), \quad \boldsymbol{x} = \begin{pmatrix}\boldsymbol{r}\\ \boldsymbol{p}\end{pmatrix}, \quad \boldsymbol{\varphi} = \begin{pmatrix}0\\ 0\end{pmatrix}, \tag{5.51}$$

$$\boldsymbol{B} = T\begin{pmatrix}0 & 0\\ 0 & \gamma\boldsymbol{I}\end{pmatrix}, \quad \boldsymbol{Q} = T\begin{pmatrix}0 & -\boldsymbol{I}\\ \boldsymbol{I} & \boldsymbol{Q}_B\end{pmatrix}, \tag{5.52}$$

where $\boldsymbol{Q}_B$ is a 3×3 matrix with

$$(\boldsymbol{Q}_B)_{ij} = -q\epsilon_{ijk}H_k. \tag{5.53}$$

It is easy to see that $\boldsymbol{Q}$ satisfies the detailed balance condition (3.4b), because magnetic field is odd under time reversal.

VI. SYMMETRIC SYSTEMS

In this section, we consider Langevin systems whose variables and control parameters are all even, i.e., $\epsilon_i = 1$, hence according to Eq. (3.4) the kinetic matrix is symmetric. These systems will be called *symmetric systems*. As we have already known, for these systems, there is no entropy pumping. We will further show that both excess EP and house-keeping EP are positive definite. Additionally, the joint pdf of house-keeping work and excess work obeys a series of fluctuation theorems, from which we can also derive fluctuation theorems and work identities for pdfs of house-keeping work and excess work separately. The fluctuation theorem for the excess work is a special case of the theorem we derived in Sec. IV C. Throughout this section, we will use Gibbs gauge, but will ignore the subscript G.

A. Gibbs gauge and thermodynamics of NESS

For symmetric systems, Eqs. (3.1), (3.2), and (3.3) become

$$dx^i + (\,B^{ij}(\partial_j U - \varphi_j) - \partial_j B^{ij}\,)dt = b^{i\alpha}dW_\alpha, \tag{6.1}$$

$$\partial_t p = \partial_i B^{ij}(\partial_j + (\partial_j U - \varphi_j))p, \tag{6.2}$$

$$\mathscr{L} \equiv \partial_i B^{ij}(\partial_j + (\partial_j U - \varphi_j)). \tag{6.3}$$

The probability current becomes

$$j^i = j^i_{IR} = -B^{ij}(\partial_j + (\partial_j U) - \varphi_j)p. \tag{6.4}$$

Using the Gibbs gauge, the steady-state current is

$$j^i_{\rm SS} = B^{ij}\varphi_j e^{-U}, \tag{6.5}$$

$$\partial_i j^i_{\rm SS} = \partial_i B^{ij}\varphi_j e^{-U} = 0. \tag{6.6}$$

Note that if we reverse the non-conservative force $\varphi_i \to -\varphi_i$ with the generalized potential U fixed, the steady-state current (6.5) changes sign whereas the steady state pdf (3.14) remains the same. This is the meaning of adjoint process, to be discussed shortly below. Note that Eq. (6.6) is the *Gibbs gauge condition* for symmetric systems, which can be obtained from Eq. (3.15) by setting $L^{ij} = B^{ij}$.

Since $Q^{ij} = 0$, the pumped entropy (3.73c) vanishes identically, and Eq. (3.79) reduces to

$$dS^{\rm tot} = dS - \beta đQ = (\Sigma^{\rm ex} + \Sigma^{\rm hk})dt. \tag{6.7}$$

The house-keeping and excess EPs are given by Eqs. (3.62a) and (3.62b), with L^{ij} replaced by B^{ij}. The excess EP is already shown to be positive in Eq. (3.64). Using Eq. (3.63), the house-keeping EP can also be shown to be positive,

$$\Sigma^{\rm hk} = \int_{\boldsymbol{x}} \varphi_i B^{ij} p\,\varphi_j \geqslant 0, \tag{6.8}$$

$$\Sigma^{\rm ex} = \int_{\boldsymbol{x}} (\partial_i(\log p + U))B^{ij}p(\partial_j(\log p + U)) \geqslant 0. \tag{6.9}$$

It was first shown by Esposito and van den Broeck [28–30] that the total EP of symmetric systems can be decomposed into two positive components.

It is important to note that the house-keeping EP as given by (6.8) is generally depends on the pdf, hence minimization of excess EP, which yields the NESS, is different from minimization of the total EP. There is, however, one interesting exception. If both the kinetic coefficients B^{ij} and the non-conservative forces φ_i are independent of the state variables $\boldsymbol{x}$, they may be pulled out of the integral in Eq. (6.8). Then the house-keeping EP is indeed independent of the system state,

$$\Sigma^{\rm hk} \to \varphi_i\varphi_j B^{ij}. \tag{6.10}$$

In this case, the steady state does minimize the total EP. In some systems, it may happen that φ_i and B^{ij} depend only weakly on system state, and hence their fluctuations are negligible. Then the total EP is approximately minimized at the steady state.

B. Fluctuation theorems

For symmetric systems, Eq. (3.73) reduces to

$$\log\frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)} = -d_{\boldsymbol{x}}U + \varphi_i\circ dx^i. \tag{6.11}$$

Equations (4.27) and (4.29) then reduce to the usual Crooks fluctuation theorem and Jarzynski equality,

$$p_{\rm F}(W)e^{-\beta W+\beta\Delta F} = p_{\rm B}(-W), \tag{6.12a}$$

$$\langle e^{-\beta W}\rangle = e^{-\beta\Delta F}, \tag{6.12b}$$

where are applicable to processes starting from and ending at equilibrium states. Because all variables and parameters are even under time reversal, these equilibrium states can also be interpreted either as NESSs. In fact, for symmetric systems, the difference between equilibrium states and NESS can only be seen at the level of probability current, not at the probability distribution. Note also Eqs. (6.11) and (6.12) are valid in arbitrary gauge. Hence the gauge degree of freedom can be used to obtain a continuous manifold of fluctuation theorems and work identities.

To derive fluctuation theorems for excess work and house-keeping work, we need to define four types of dynamic processes, all in Gibbs gauge. For simplicity of notations, we denote non-conservative force as φ.

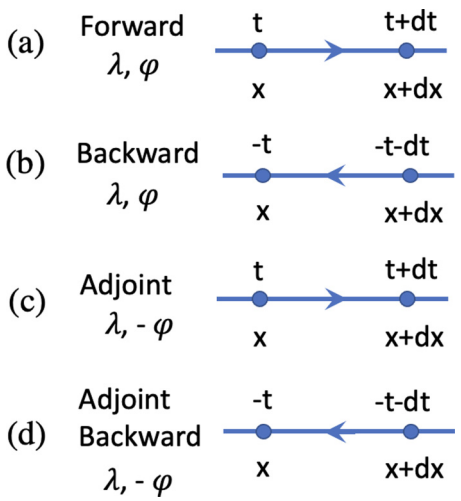

FIG. 5. An infinitesimal step in the forward process, the backward, the adjoint, and the adjoint backward processes. The arrow indicates the propagation of time.

(i) *Forward process.* The system starts from NESS

$$p_{\rm SS}(\boldsymbol{x}, \lambda(t_i)) = e^{-U(\boldsymbol{x},\lambda(t_i))}, \tag{6.13}$$

at $t = t_i$ and the dynamic protocol is $\lambda^{\rm F}(t) = \lambda(t), \varphi^{\rm F}(t) = \varphi(\lambda(t))$ with $t \in [t_i, f_f]$. Here we use the term *protocol* to denote both the control parameter and the non-conservative force.

(ii) *Backward process.* The system starts from NESS

$$p_{\rm SS}(\boldsymbol{x}, \lambda(t_f)) = e^{-U(\boldsymbol{x},\lambda(t_f))}. \tag{6.14}$$

at $t = -t_f$, and the dynamic protocol is $\lambda^{\rm B}(t) = \lambda(-t), \varphi^{\rm B}(t) = \varphi(\lambda(-t))$, with $t \in [-f_f, -t_i]$.

(iii) *Adjoint process.* The system starts from NESS (6.13), and the dynamic protocol is $\lambda^{\rm Ad}(t) = \lambda(t), \varphi^{\rm Ad}(t) = -\varphi(\lambda(t))$, with $t \in [t_i, f_f]$.

(iv) *Adjoint backward process.* The system starts from NESS (6.14), and the dynamic protocol is $\lambda^{\rm AdB}(t) = \lambda(-t), \varphi^{\rm AdB}(t) = -\varphi(\lambda(-t))$, with $t \in [-f_f, -t_i]$. Here the superscript AdB means adjoint backward.

An infinitesimal step of each of these processes is illustrated in Fig. 5. The transition probabilities of these infinitesimal steps are calculated in Appendix, where we also show

$$\log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x}; dt)}{P_{\lambda,-\varphi}(\boldsymbol{x}|\boldsymbol{x}_1; dt)} = \log \frac{P_{\lambda,-\varphi}(\boldsymbol{x}_1|\boldsymbol{x}; dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1; dt)} = -d_{\boldsymbol{x}} U = -\beta d\mathcal{Q}^{\rm ex}, \tag{6.15a}$$

$$\log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x}; dt)}{P_{\lambda,-\varphi}(\boldsymbol{x}_1|\boldsymbol{x}; dt)} = \log \frac{P_{\lambda,-\varphi}(\boldsymbol{x}|\boldsymbol{x}_1; dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1; dt)} = \varphi_i \circ dx^i = -\beta d\mathcal{Q}^{\rm hk}. \tag{6.15b}$$

In the above, the subscripts (λ, φ) refer to the infinitesimal protocol of the forward and backward processes, whereas $(\lambda, -\varphi)$ refers to the infinitesimal protocol of the adjoint and backward adjoint processes, as illustrated in Fig. 5. The integrated house keeping and excess work and heat along the forward process are (with $f_i = T\varphi_i$)

$$\mathscr{W}^{\rm hk}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] = \int_\gamma f_i \circ dx^i, \tag{6.16a}$$

$$\mathscr{Q}^{\rm hk}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] = -\int_\gamma f_i \circ dx^i, \tag{6.16b}$$

$$\mathscr{W}^{\rm ex}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] = \int_\gamma d_\lambda H, \tag{6.16c}$$

$$\mathscr{Q}^{\rm ex}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] = \int_\gamma d_{\boldsymbol{x}} H. \tag{6.16d}$$

These functionals can be similarly defined for the backward, adjoint, and adjoint backward processes. We shall not list their concrete forms here, in order not to make the work excessively long. These functionals have the following symmetry properties:

$$\begin{aligned}\mathscr{W}^{\rm hk}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] &= -\mathscr{W}^{\rm hk}[\tilde{\gamma}, \lambda^{\rm B}, \varphi^{\rm B}] \\ &= -\mathscr{W}^{\rm hk}[\gamma, \lambda^{\rm Ad}, \varphi^{\rm Ad}] \\ &= \mathscr{W}^{\rm hk}[\tilde{\gamma}, \lambda^{\rm AdB}, \varphi^{\rm AdB}],\end{aligned} \tag{6.17a}$$

$$\begin{aligned}\mathscr{Q}^{\rm hk}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] &= -\mathscr{Q}^{\rm hk}[\tilde{\gamma}, \lambda^{\rm B}, \varphi^{\rm B}] \\ &= -\mathscr{Q}^{\rm hk}[\gamma, \lambda^{\rm Ad}, \varphi^{\rm Ad}] \\ &= \mathscr{Q}^{\rm hk}[\tilde{\gamma}, \lambda^{\rm AdB}, \varphi^{\rm AdB}],\end{aligned} \tag{6.17b}$$

$$\begin{aligned}\mathscr{W}^{\rm ex}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] &= -\mathscr{W}^{\rm ex}[\tilde{\gamma}, \lambda^{\rm B}, \varphi^{\rm B}] \\ &= \mathscr{W}^{\rm ex}[\gamma, \lambda^{\rm Ad}, \varphi^{\rm Ad}] \\ &= -\mathscr{W}^{\rm ex}[\tilde{\gamma}, \lambda^{\rm AdB}, \varphi^{\rm AdB}],\end{aligned} \tag{6.17c}$$

$$\begin{aligned}\mathscr{Q}^{\rm ex}[\gamma, \lambda^{\rm F}, \varphi^{\rm F}] &= -\mathscr{Q}^{\rm ex}[\tilde{\gamma}, \lambda^{\rm B}, \varphi^{\rm B}] \\ &= \mathscr{Q}^{\rm ex}[\gamma, \lambda^{\rm Ad}, \varphi^{\rm Ad}] \\ &= -\mathscr{Q}^{\rm ex}[\tilde{\gamma}, \lambda^{\rm AdB}, \varphi^{\rm AdB}].\end{aligned} \tag{6.17d}$$

Using Eqs. (6.15) for every small steps, as well as the initial pdfs Eqs. (6.13) and (6.14) for all processes, we can establish the following relations:

$$\frac{p_{\rm F}[\gamma]}{p_{\rm B}[\tilde{\gamma}]} = e^{\beta \mathscr{W}^{\rm hk}[\gamma] + \beta \mathscr{W}^{\rm ex}[\gamma] - \beta \Delta F(\lambda)}, \tag{6.18}$$

$$\frac{p_{\rm F}[\gamma]}{p_{\rm Ad}[\gamma]} = \frac{p_{\rm AdB}[\tilde{\gamma}]}{p_{\rm B}[\tilde{\gamma}]} = e^{\beta \mathscr{W}^{\rm hk}[\gamma]}, \tag{6.19}$$

$$\frac{p_{\rm F}[\gamma]}{p_{\rm AdB}[\tilde{\gamma}]} = \frac{p_{\rm Ad}[\gamma]}{p_{\rm B}[\tilde{\gamma}]} = e^{\beta \mathscr{W}^{\rm ex}[\gamma] - \beta \Delta F}. \tag{6.20}$$

We can now define the joint probability density for house-keeping and excess works for the forward process,

$$p_{\rm F}(W^{\rm hk}, W^{\rm ex}) \equiv \int \mathscr{D}[\gamma]\, p_{\rm F}[\gamma]\, \delta(W^{\rm hk} - \mathscr{W}^{\rm hk}[\gamma, \lambda, \boldsymbol{\varphi}]) \times \delta(W^{\rm ex} - \mathscr{W}^{\rm ex}[\gamma, \lambda, \boldsymbol{\varphi}]), \tag{6.21}$$

from which we can also obtain the marginal pdfs for the house-keeping work and excess work,

$$p_{\rm F}(W^{\rm hk}) \equiv \int_{W^{\rm ex}} p_{\rm F}(W^{\rm hk}, W^{\rm ex}), \tag{6.22}$$

$$p_{\rm F}(W^{\rm ex}) \equiv \int_{W^{\rm hk}} p_{\rm F}(W^{\rm hk}, W^{\rm ex}). \tag{6.23}$$

The notations we use for these pdfs are not precise from mathematical point of view, but should cause no difficulty in understanding. Similar probability density functions can also be defined for the backward, adjoint, and adjoint backward processes.

Following a proof similar to that in Sec. IV A, we can prove the following fluctuation theorems:

$$\begin{aligned} p_{\rm F}(W^{\rm hk}, W^{\rm ex}) &= e^{\beta(W^{\rm hk}+W^{\rm ex}-\Delta F)} p_{\rm B}(-W^{\rm hk}, -W^{\rm ex}) \\ &= e^{\beta W^{\rm hk}} p_{\rm Ad}(-W^{\rm hk}, W^{\rm ex}) \\ &= e^{\beta(W^{\rm ex}-\Delta F)} p_{\rm AdB}(W^{\rm hk}, -W^{\rm ex}). \end{aligned} \tag{6.24}$$

Note that the Crooks fluctuation theorem Eq. (6.12a) can be derived from the first equality of Eqs. (6.24). From these results we further obtain the following fluctuation theorems for the reduced pdf of house-keeping work and for excess work,

$$p_{\rm F}(W^{\rm hk}) e^{-\beta W^{\rm hk}} = p_{\rm Ad}(-W^{\rm hk}), \tag{6.25}$$

$$p_{\rm F}(W^{\rm ex}) e^{-\beta W^{\rm ex}+\beta\Delta F} = p_{\rm AdB}(-W^{\rm ex}), \tag{6.26}$$

as well as the corresponding work identities,

$$\langle e^{-\beta W^{\rm hk}} \rangle = 1, \tag{6.27}$$

$$\langle e^{-\beta W^{\rm ex}} \rangle = e^{-\beta\Delta F}. \tag{6.28}$$

Equations (6.26) and (6.28) were already derived in Sec. IV.

Because $\beta W^{\rm hk}$ and $\beta W^{\rm ex} - \beta\Delta F$ can be understood as the house-keeping EP and excess EP at the trajectory level, Eqs. (6.25)–(6.28) can be rewritten in terms of the total house-keeping EP and excess EP,

$$p_{\rm F}(\Delta S^{\rm hk}) e^{-\Delta S^{\rm hk}} = p_{\rm Ad}(-\Delta S^{\rm hk}), \tag{6.29}$$

$$p_{\rm F}(\Delta S^{\rm ex}) e^{-\Delta S^{\rm ex}} = p_{\rm AdB}(-\Delta S^{\rm ex}), \tag{6.30}$$

$$\langle e^{-\Delta S^{\rm hk}} \rangle = \langle e^{-\Delta S^{\rm ex}} \rangle = 1. \tag{6.31}$$

It is important to note that, unlike Eqs. (6.12), which are valid in arbitrary gauge, Eqs. (6.24)–(6.31) are valid only in the Gibbs gauge.

Relations (6.29) and (6.30) were derived in Ref. [28–30], where $\Delta S^{\rm hk}$ and $\Delta S^{\rm ex}$ are respectively called *adiabatic entropy production* and *nonadiabatic entropy production*. Equations (6.24), which are the most general versions of all these fluctuation theorems, have not been derived previously, to the best of our knowledge.

C. Implications of fluctuation theorems

Several significant implications can be inferred from the above fluctuation theorems.

(i) Even though the fluctuation theorems (6.24)–(6.28) are derived in Gibbs gauge, we must remember that the processes involved can be realized in infinitely many different gauges, each corresponding to a different experimental setup. This point becomes very important when applying the theory to concrete systems.

(ii) Let us for now assume that the control parameter λ is fixed, and that the system is in the steady regime. The excess work then vanishes identically at the trajectory level, according to Eq. (3.50b). The pdf of $W^{\rm ex}$ is then concentrated at zero. The steady-state free energy is also fixed, and hence $\Delta F = 0$. Hence the fluctuation theorem for the excess work becomes trivial. Also because λ are fixed, the forward process and backward process are then identical to each other, whereas the adjoint process and the adjoint backward process are also identical to each other. Note however the forward process and the adjoint process are generically different. Now Eq. (6.12a) and (6.25) respectively become

$$e^{-\beta W^{\rm hk}} p_{\rm F}(W^{\rm hk}) = p_{\rm F}(-W^{\rm hk}), \tag{6.32}$$

$$e^{-\beta W^{\rm hk}} p_{\rm F}(W^{\rm hk}) = p_{\rm Ad}(-W^{\rm hk}), \tag{6.33}$$

from which we deduce

$$p_{\rm F}(W^{\rm hk}) = p_{\rm Ad}(W^{\rm hk}). \tag{6.34}$$

Hence, even though the forward process and the adjoint process are generally different physical processes, they have the identical probability distribution of work in the steady regime. For the same reason, the backward process and the adjoint backward process have the identical probability distribution of work in the steady regime.

(iii) The fluctuation theorem (6.32) holds exactly for arbitrary time interval $\Delta t = t_f - t_i$. This is in strong contrast with the famous steady-state fluctuation theorem, which was only established as asymptotically exact in the long time limit. Hence the fluctuation theorem for house-keeping work (6.32) is a significant refinement of the steady-state fluctuation theorem. Note that all these derivations are applicable only for symmetric systems.

(iv) If the control parameter is not fixed, the fluctuation theorem of excess work becomes nontrivial. An interesting and very common scenario is that a small system is embedded in a large system in nonequilibrium dissipative state. For example, a polymer chain or a colloid is immersed in a gradient flow or in a temperature gradient, and at the same time manipulated by an optical tweezer. For these problems, the non-conservative force drives shear flow or the temperature gradient. The house-keeping work is an extensive quantity, whose fluctuations are difficult to measure experimentally. By strong contrast, the control parameter λ only couples to the small system. Hence the excess work and excess EP are typically small, whose fluctuations can be characterized by Eq. (6.26), and can be experimentally tested. We believe that this is the most interesting scenario.

VII. EXAMPLES OF SYMMETRIC SYSTEMS

In this section, we discuss three examples of symmetric models. We first describe the systems using the natural gauge defined by experimental setups, then find the NESS, and represent the systems in Gibbs gauge. We then calculate excess and house-keeping work, as well as the excess and house-keeping EP. Finally we discuss how backward process, adjoint process, and adjoint backward process can be realized.

FIG. 6. A polymer dragged by a constant force.

A. A dragged polymer

We consider a polymer dragged by a constant force. The polymer is modeled as two mass points connected by a harmonic spring. The force is acting only on one of the mass points. For simplicity, we treat the system as one dimensional, and assume that two mass points have the equal mass and friction constant. Both the spring constant and force may be controlled experimentally. The system is schematically shown in Fig. 6.

We start from the over-damped Langevin equations,

$$dx_1 + \frac{1}{\gamma}[k(x_1 - x_2) - f]dt = \sqrt{\frac{2T}{\gamma}}dW_1, \tag{7.1a}$$

$$dx_2 + \frac{1}{\gamma}k(x_2 - x_1)dt = \sqrt{\frac{2T}{\gamma}}dW_2, \tag{7.1b}$$

which can be written into the covariant form (3.1) with

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad U = \frac{1}{2}\beta k(x_1 - x_2)^2, \tag{7.2}$$

$$\boldsymbol{L} = \boldsymbol{B} = \frac{T}{\gamma}\boldsymbol{I}, \quad \boldsymbol{\varphi} = \begin{pmatrix} \beta f \\ 0 \end{pmatrix}, \tag{7.3}$$

where U and $\boldsymbol{\varphi}$ denote the generalized potential and nonconservative force in the natural gauge defined by experimental setup.

In the steady state, the polymer moves with average speed $f/2\gamma$, and the average extension of the spring should be such that the elastic force balances the friction, $\langle \Delta x \rangle = \langle x_1 - x_2 \rangle = f/2k$. Hence the steady state pdf, as well as the corresponding generalized potential, Hamiltonian, and nonconservative force in Gibbs gauge are

$$p_{\rm SS}(\boldsymbol{x}) = \frac{1}{L}\sqrt{\frac{\pi}{\beta k}}e^{-\beta k(x_1 - x_2 - f/2k)^2/2}, \tag{7.4}$$

$$U^{\rm G}(\boldsymbol{x}; k, f) = \frac{\beta k}{2}\left(x_1 - x_2 - \frac{f}{2k}\right)^2 + \frac{\log k}{2}, \tag{7.5}$$

$$H^{\rm G}(\boldsymbol{x}; k, f) = \frac{1}{2}k\left(x_1 - x_2 - \frac{f}{2k}\right)^2, \tag{7.6}$$

$$\boldsymbol{\varphi}^{\rm G} = \begin{pmatrix} \beta f/2 \\ \beta f/2 \end{pmatrix}, \tag{7.7}$$

The Hamiltonian $H^{\rm G}$ is determined by the generalized potential $U^{\rm G}$ up to an additive constant, which was chosen to maximally simplify $H^{\rm G}$. Using Eqs. (7.5) and (7.7), we easily verify Gibbs gauge condition Eq. (6.6).

The house-keeping EP and excess EP are respectively,

$$\Sigma^{\rm hk} = \int_{\boldsymbol{x}} \varphi_i^{\rm G} B^{ij} p \varphi_j^{\rm G} = \frac{f^2}{2T\gamma}, \tag{7.8}$$

$$\Sigma^{\rm ex} = \frac{T}{\gamma}\int_x p\left[\left(\partial_1 \log p + \beta k(x_1 - x_2) - \frac{1}{2}\beta f\right)^2 + \left(\partial_2 \log p - \beta k(x_1 - x_2) + \frac{1}{2}\beta f\right)^2\right], \tag{7.9}$$

which are indeed both positive. Notice that for this particular problem $\Sigma^{\rm hk}$ is independent of system state. Hence, the total EP is minimized by the NESS. Starting from any initial state, the EP decreases monotonically as a function of time. If we study the conformational behavior of the polymer in the moving frame, there is no way for us to deduce that the system is in a NESS instead of an equilibrium state!

At the trajectory level, the house-keeping work and excess work are given respectively by

$$d\mathscr{W}^{\rm hk} = T\varphi_i \circ dx^i = \frac{f}{2}(dx^1 + dx^2), \tag{7.10}$$

$$d\mathscr{W}^{\rm ex} = (\partial_k H)dk + (\partial_f H)df. \tag{7.11}$$

The Langevin equations (7.1) can be rewritten in terms of the center-of-mass coordinate $X = (x_1 + x_2)/2$ and relative coordinate $x = x_1 - x_2$,

$$dX - \frac{f}{2\gamma}dt = \sqrt{\frac{2T}{2\gamma}}dW_X, \tag{7.12a}$$

$$dx + \frac{1}{\gamma/2}(kx - f/2)dt = \sqrt{\frac{2T}{\gamma/2}}dW_x, \tag{7.12b}$$

where these two variables are completely decoupled. It becomes clear now that the house-keeping work Eq. (7.10) is precisely the work associated with the dragged motion of the center-of-mass coordinate X,

$$d\mathscr{W}^{\rm hk} = f \circ dX, \tag{7.13}$$

whereas the excess work is the work associated with the relative coordinate x,

$$d\mathscr{W}^{\rm ex} = d_\lambda H^{\rm G}, \tag{7.14}$$

where $H^{\rm G}$ is Eq. (7.6) expressed in terms of x. Since the Langevin equations Eq. (7.12) of x and X are completely decoupled, a theory of stochastic thermodynamics can be constructed for separately for x and for X. Each theory yields its own fluctuation theorem and work relation.

The EP associated with X and x can be obtain straightforwardly,

$$\Sigma_{\rm CM} = \int_X (\partial_X \log p(X) - \beta f)^2 \frac{T p(X)}{2\gamma} = \Sigma^{\rm hk} + \frac{T}{2\gamma}\int_X \frac{p'(X)^2}{p(X)}, \tag{7.15}$$

$$\Sigma_{\rm Rel} = \int_x (\partial_x \log p + \beta k(x - f/2k))^2 \frac{2T p(x)}{\gamma} = \Sigma^{\rm ex} - \frac{T}{2\gamma}\int_X \frac{p'(X)^2}{p(X)}. \tag{7.16}$$

They differ from $\Sigma^{\rm hk}$, $\Sigma^{\rm ex}$, given by Eqs. (7.8) and (7.9), only by a part, which depends on $p(X)$, but is independent

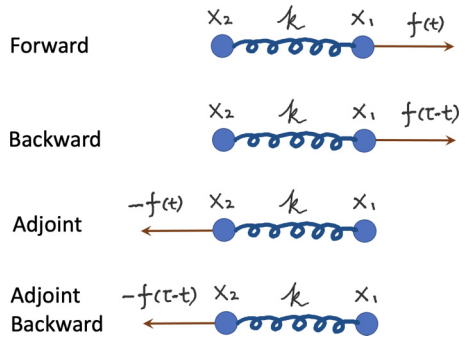

FIG. 7. How forward, backward, adjoint, and adjoint backward processes may be realized for a dragged polymer chain.

of both control parameters k, f. It therefore has no impact on fluctuation theorems.

Let us now discuss, for a given forward process, how backward process, adjoint process, and adjoint backward process can be realized. For concreteness, we assume that the spring constant k is fixed, and the dragging force $f(t)$ acting on x_1 follows a specified pattern in the forward process with $0 < t < \tau$. In the backward process, the force still acts on x_1 but varies as $f^{\rm B}(t) = f(\tau - t)$. The adjoint process is defined such that in the Gibbs gauge, the non-conservative force is the opposite of that of the forward process, whereas the generalized potential is the same. Such a process can be realized by apply a force $-f(t)$ on the second particle x_2. Finally in the adjoint process, the force is acting on x_2 and is given by $-f(\tau - t)$. These processes are illustrated in Fig. 7.

B. A particle in circular force field

This is a simplified version of the model was first discussed by Jarzynski *et al.* in Ref. [27]. As illustrated in Fig. 8, we consider a particle moving in a 2d isotropic potential $V(r, \lambda)$ and a circular non-conservative force field $g(r)\boldsymbol{e}_\theta$. The underdamped Langevin equations are

$$dp_x = \left(-\gamma\frac{p_x}{m} + f\frac{x}{r} - g\frac{y}{r}\right)dt + \sqrt{2\gamma T}dW_x, \quad (7.17)$$

$$dp_y = \left(-\gamma\frac{p_y}{m} + f\frac{y}{r} + g\frac{x}{r}\right)dt + \sqrt{2\gamma T}dW_y, \quad (7.18)$$

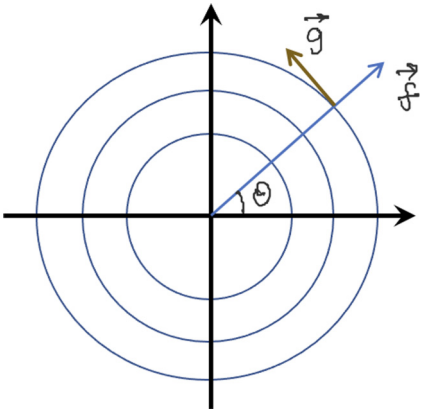

FIG. 8. A particle subjected to a conservative force $\boldsymbol{f}$ and non-conservative force $\boldsymbol{g}$.

where $f(r) = -V'(r)$. We consider the over-damped limit, where the Langevin equations become

$$dx = -\frac{fx - gy}{\gamma\, r}dt + \sqrt{\frac{2T}{\gamma}}dW_x, \quad (7.19)$$

$$dy = -\frac{fy + gx}{\gamma\, r}dt + \sqrt{\frac{2T}{\gamma}}dW_y, \quad (7.20)$$

which correspond to the covariant form (3.1) with

$$U = \beta V(r), \quad \boldsymbol{B} = \frac{T}{\gamma}\boldsymbol{I}, \quad \boldsymbol{\varphi} = \begin{pmatrix} -\beta g \sin\theta \\ \beta g \cos\theta \end{pmatrix}, \quad (7.21)$$

where θ is the polar angle. The FPE is

$$\partial_t p = \frac{T}{\gamma}\left[\partial_x\left(\partial_x - \frac{fx - gy}{T\, r}\right) + \partial_y\left(\partial_y - \frac{fy + gx}{T\, r}\right)\right]p. \quad (7.22)$$

It is easy to verify that the steady state is given by

$$p^{\rm SS}(r) = e^{-\beta V(r)}, \quad (7.23)$$

and the Gibbs gauge condition Eq. (6.6) is satisfied. Hence the parametrization Eq. (7.21) is already in the Gibbs gauge. The excess and house-keeping works are then easily obtained,

$$d\mathscr{W}^{\rm hk} = T\boldsymbol{\varphi} \circ d\boldsymbol{x} = \frac{g(r)}{r}(-y \circ dx + x \circ dy), \quad (7.24)$$

$$d\mathscr{W}^{\rm ex} = (\partial_\lambda V(r, \lambda))d\lambda. \quad (7.25)$$

The total EP can be written as

$$\Sigma = \frac{T}{\gamma}\int_{\boldsymbol{x}}\left[\left(\frac{\partial_x p}{p} - \frac{fx - gy}{T\, r}\right)^2 + \left(\frac{\partial_y p}{p} - \frac{fy + gx}{T\, r}\right)^2\right]. \quad (7.26)$$

Using integration by parts, as well as the fact that $p^{\rm SS}$ satisfies the steady state FPE, we can easily rewrite it into the sum of two positive parts,

$$\Sigma = \frac{T}{\gamma}\int_{\boldsymbol{x}} p(x, y)[(\partial_x \log p\, e^{\beta V})^2 + (\partial_y \log p\, e^{\beta V})^2] + \frac{1}{\gamma T}\int_{\boldsymbol{x}} p(x, y)\, g(r)^2, \quad (7.27)$$

where the first part is the excess EP and the second part is the house-keeping EP. Note that the house-keeping EP depends on $p(x, y)$. Hence the total EP is not minimized at the NESS.

Finally, the adjoint process and adjoint backward process can be obtained from the forward process and backward process by reversing the sign of $g(r)$, i.e., the direction of the non-conservative force.

C. Confined Brownian particle in shear flow

Our last example is a Brownian particle confined by a central force potential $V(\boldsymbol{x}) = k(\boldsymbol{x} - \boldsymbol{x}_0)^2/2$ in a shear flow, as illustrated in Fig. 9. Both k and $\boldsymbol{x}_0$ can be tuned externally. The velocity field of the fluid is

$$\vec{v}(\boldsymbol{x}, \boldsymbol{y}) = \frac{y}{\tau}\hat{x}. \quad (7.28)$$

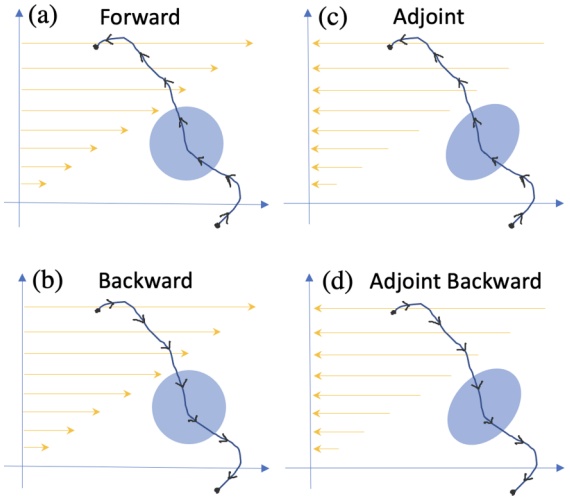

FIG. 9. A confined Brownian particle in shear flow. The disks represent the quadratic confining potentials. The yellow arrows indicate the velocity field of the fluid. The blue curves with arrows represent the loci of the center of the confining potential in different processes, which act as control parameters. Note that in the adjoint and adjoint backward processes, the confining potential is anisotropic, whereas the shear flow is reversed.

The over-damped Langevin equations that describe the motion of the Brownian particle are

$$\gamma\Big(\dot{x} - \frac{y}{\tau}\Big)dt + \partial_x V(\boldsymbol{x})dt = \sqrt{2\gamma T}\,dW_x, \tag{7.29}$$

$$\gamma\dot{y}\,dt + \partial_y V(\boldsymbol{x})dt = \sqrt{2\gamma T}\,dW_y. \tag{7.30}$$

Note that the friction force acting on particle is $\gamma(\boldsymbol{x} - y/\tau)$, which vanishes only if the particle moves with the fluid. The above Langevin equations can be written in the covariant form in the natural gauge with

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \boldsymbol{L} = \boldsymbol{B} = \frac{T}{\gamma}\boldsymbol{I}, \tag{7.31a}$$

$$U = \frac{\beta k}{2}(\delta x^2 + \delta y^2), \quad \boldsymbol{\varphi} = \begin{pmatrix} \beta\gamma y/\tau \\ 0 \end{pmatrix}. \tag{7.31b}$$

With the above parametrization, $(\gamma y/\tau)\hat{x}$ is interpreted as the non-conservative force acting on the particle due to the fluid in shear flow. If we treat γ/τ as a single parameter, it is even under time reversal, as one can see easily from unit analysis. (Both γ and τ are odd.) The non-conservative force as given in Eq. (7.31b) satisfies the detailed balance condition (3.6b), and hence our theory is applicable. Within this interpretation, then, the shear flow in the backward process is the same as that in the forward process, as shown in Fig. 9.

We may however treat τ as the control parameter with the friction coefficient γ being fixed. The combination γ/τ then changes sign under time reversal, i.e., it behaves as an odd parameter. The detailed balance condition (3.6b) is then no longer satisfied, and our theory is not applicable. It turns out that a different theory of stochastic thermodynamics may be developed for this interpretation, which is in the same spirit as the theory developed in Ref. [45]. The total EP calculated in these two theories are different from each other, yet all excess thermodynamic quantities are identical. Note, however, in both theories, the EP due to the shear flow itself, which is extensive in the size of the fluid, is not included.

A detailed comparison of these two theories will be supplied elsewhere. Here we adopt the first interpretation, by treating γ/τ as an even parameter. We can then follow the theory developed in Sec. III and write down the work and heat at the trajectory level (in the natural gauge):

$$đ\mathscr{W} = d_\lambda V + \frac{\gamma y}{\tau} \circ dx, \tag{7.32}$$

$$đ\mathscr{Q} = d_{\boldsymbol{x}} V - \frac{\gamma y}{\tau} \circ dx, \tag{7.33}$$

where $\circ$ is the product in Stratonovich's sense.

It is not easy to solve for the NESS for the general case. The situation is, however, greatly simplified if the shear flow is weak, so that the steady state remains approximately $e^{-\beta V(\boldsymbol{x})}$. In the Gibbs gauge, the generalized potential U^{G} and the non-conservative force φ_i^{G} are related to those in the natural gauge via a gauge transformation,

$$U^{\mathrm{G}} = U + \psi, \tag{7.34}$$

$$\varphi_i^{\mathrm{G}} = \varphi_i + \partial_i\psi, \tag{7.35}$$

where ψ is a quadratic function of $\boldsymbol{x}$, also small comparing with U. The Gibbs gauge condition (6.6) can be rewritten as

$$\partial_i\varphi_i^{\mathrm{G}} - \varphi_i^{\mathrm{G}}\partial_i U^{\mathrm{G}} = 0. \tag{7.36}$$

But U^{G} in Eq. (7.36) can be approximated by U, which leads to

$$\partial_i\varphi_i^{\mathrm{G}} - \varphi_i^{\mathrm{G}}\beta k\,\delta x^i = 0, \tag{7.37}$$

where $\delta x^i = x^i - x_0^i$. We expect that ψ is quadratic in $\delta\boldsymbol{x}$, and φ_i^{G} is linear in $\delta\boldsymbol{x}$. It is easy to verify that the solution to Eq. (7.37) is

$$\varphi_i^{\mathrm{G}} = \alpha\,\epsilon_{ij}\delta x_j, \tag{7.38}$$

where α is an arbitrary constant, whilst ϵ_{ij} is the antisymmetric tensor. The constant α is determined by substituting Eq. (7.38) into Eq. (7.35), and requiring that $\partial_i\partial_j\psi = \partial_i\partial_j\psi$. This also determines the gauge transformation ψ, as well as U^{G}, φ^{G} in the Gibbs gauge,

$$\psi = -\frac{\beta\gamma}{2\tau}(\delta x\delta y + 2\,\delta x\, y_0), \tag{7.39a}$$

$$U^{\mathrm{G}} = \frac{\beta k}{2}(\delta x^2 + \delta y^2) - \frac{\beta\gamma}{2\tau}(\delta x\delta y + 2\,\delta x\, y_0), \tag{7.39b}$$

$$\varphi_i^{\mathrm{G}} = \frac{\beta\gamma}{2\tau}\epsilon_{ij}\delta x^j, \tag{7.39c}$$

where the generalized potential U^{G} is determined only up to an additive constant, which depends on y_0 and all other parameters. It can be fixed of course by the normalization condition. Note that the non-conservative force field Eq. (7.39c) is precisely clockwise circular force field, opposite to that studied in Sec. VII B. The second term in the r.h.s. of Eq. (7.39b) arises due to the dragging effect of the shear flow. It has two effects: (1) the probability distribution of $\boldsymbol{x}$ anisotropic, and (2) its center is translated relative to the confining potential $V(\boldsymbol{x})$, along the $\boldsymbol{x}$ direction.

Now the adjoint process (again in Gibbs gauge) can be obtained from Eq. (7.39) by reversing the non-conservative force field, i.e., from clockwise to counterclockwise,

$$U^{\mathrm{Ad,G}} = \frac{\beta k}{2}(\delta x^2 + \delta y^2) - \frac{\beta\gamma}{2\tau}(\delta x\delta y + 2\,\delta x\, y_0), \quad (7.40a)$$

$$\varphi_i^{\mathrm{Ad,G}} = -\frac{\beta\gamma}{2\tau}\epsilon_{ij}\delta x^j. \quad (7.40b)$$

Now suppose we want to realize the adjoint process in the reversed shear flow, as shown in Fig. 9(c). The non-conservative force is then the opposite of that in Eq. (7.31b). The generalized potential then can be obtained from Eqs. (7.40) by making another gauge transform. The results are

$$U^{\mathrm{Ad}} = \frac{\beta k}{2}(\delta x^2 + \delta y^2) - \frac{\beta\gamma}{\tau}(\delta x\delta y + 2\,\delta x\, y_0), \quad (7.41a)$$

$$\varphi_i^{\mathrm{Ad}} = -\frac{\beta\gamma}{\tau}\delta_{i1}\delta y. \quad (7.41b)$$

The potential U^{Ad} corresponds to an asymmetric confining potential as shown in Fig. 9. The forward process and its corresponding adjoint backward process are also illustrated in Fig. 9.

It can be seen that in the present problem the forward process and the adjoint process are physically different. Yet Eq. (6.34) guarantees that for fixed control parameters, the distributions of house-keeping work in these two processes are identical to each other.

In typical experiments, the size of the fluid is microscopic, hence the total house-keeping entropy production and house-keeping work, which are due to both the sheared fluid and the confined particle, are extensive quantities, whose fluctuations are difficult to measure. Yet the excess entropy production and excess work are small quantities, whose fluctuations can be characterized by the fluctuation theorem of excess work. More detailed will be supplied in a separate paper.

VIII. CONCLUSIONS

With judicious choice of system variables, most classical nonequilibrium systems can be described by Markov processes, with either discrete or continuous state variables. With system parameters fixed, a Markov process converges either to an equilibrium state with detailed balance, or to a nonequilibrium steady state without detailed balance. The thermodynamics and stochastic thermodynamics for the former case has been well established, as the consequence of a very large body of papers due to many authors. By synthesizing and generalizing many important ideas and theories, in this paper we developed a general theory of thermodynamics and stochastic thermodynamics for latter case. The combination of the earlier papers in this sequel [48–50] and the present paper supply a consistent and relatively complete theoretical framework for thermodynamics and stochastic thermodynamics of all classical nonequilibrium systems with continuous state variables.

Our theory is based on splitting of thermodynamic forces into a conservative part and a non-conservative part, an operation that is gauge dependent. Each gauge defines an experimental setup, and the same physical process can be realized by many different experimental setups. In the special Gibbs gauge, there is a full correspondence between the thermodynamic theory of non-conservative systems and that of the conservative systems. Specializing to the stationary regime, our theory provides a rigorous justification of the Glansdorff-Prigogine stability criterion of NESS, as well as the steady-state thermodynamics, developed by Oono and Paniconi, as well as Sasa and Tasaki.

For asymmetric systems, entropy pumping constitutes a universal mechanism of entropy exchange between system and its environment, which does not involve dissipation. Using entropy pumping, a system may export entropy steadily and maintain a low entropy, far-from-equilibrium state. All living systems constantly probe their environments and adapt their behaviors accordingly. From statistical physics point of view, these may very well be understood in terms of entropy pumping. We feel that further exploration along this direction may help reveal the essential difference between living systems and nonliving systems. For small systems embedded in large nonequilibrium backgrounds, the fluctuation theorems of excess work and excess EP provide valuable characterization of fluctuations. All these issues will be explored in greater detail in future papers.

ACKNOWLEDGMENTS

The authors acknowledge support from NSFC via Grants No. 11674217(X.X.) and No. 12075016(F.L.), as well as Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01). M.D. thanks Chenxing Post Doctoral Incentive Project for financial support. X.X. is also thankful to Hong Qian for enlightening discussion about numerous relevant literatures.

APPENDIX: DERIVATION OF DETAILED FLUCTUATION THEOREMS

1. Derivation of Eq. (3.73)

We first sketch the results in the Appendix of Ref. [49], which are useful for our present purpose. For a full proof with all details, also see Ref. [76]. Here, we assume that the metric tensor is trivial, $g_{ij} = \delta_{ij}$. The most general form of Ito-Langevin equation is

$$dx^i - F^i(\boldsymbol{x}, \lambda)dt = b^{i\nu}(\boldsymbol{x}, \lambda)dW_\nu, \quad (A1)$$

where $F^i(\boldsymbol{x}, \lambda)$ is usually called the *systematic force*, and λ is the control parameter, which may vary with time. Consider a transition from $\boldsymbol{x}$ at time t to $\boldsymbol{x}_1$ at time $t + dt$, and let $\alpha \in (0, 1)$, so that

$$\boldsymbol{x}_\alpha = \boldsymbol{x} + \alpha(\boldsymbol{x}_1 - \boldsymbol{x}) = \boldsymbol{x} + \alpha d\boldsymbol{x} \quad (A2)$$

is an intermediate point between $\boldsymbol{x}$ and $\boldsymbol{x}_1$. The differential transition probability are given by Eqs. (A2) of Ref. [49] with the metric tensor set to δ_{ij},

$$d\mu(\boldsymbol{x}_1)P_\lambda(\boldsymbol{x}_1|\boldsymbol{x}; dt) = \frac{d\mu(\boldsymbol{x}_1)\, e^{-A^\alpha(\boldsymbol{x}_1, \boldsymbol{x}; dt, \lambda)}}{\sqrt{(4\pi\, dt)^n \det B^{ij}(\boldsymbol{x}_\alpha, \lambda)}}, \quad (A3a)$$

$$A^{\alpha}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda) = [dx^i - dt(F^i - 2\alpha\partial_k B^{ik})_\alpha]\frac{B^{-1}_{ij}(\boldsymbol{x}_\alpha,\lambda)}{4dt} \times [dx^j - dt(F^j - 2\alpha\partial_l B^{jl})_\alpha] + \alpha\,(\partial_i F^i)_\alpha dt - \alpha^2(\partial_i\partial_j B^{ij})_\alpha dt, \tag{A3b}$$

where $A^{\alpha}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda)$ is called the action, $d\mu(\boldsymbol{x}_1) = d^n\boldsymbol{x}_1$ is an infinitesimal volume element around $\boldsymbol{x}_1$, B^{-1}_{ij} is the inverse matrix of $B^{ij} = \sum_\nu b^{i\nu}b^{j\nu}/2$ [c.f. Eq. (2.2b)], and $(\cdots)_\alpha$ means that all functions inside the bracket are evaluated at $\boldsymbol{x}_\alpha$. The action only needs to be calculated up to the first order in dt, whereas dx^i, dt are all infinitesimal, and for typical paths, $dx \propto dt^{1/2}$. Consequently, it does not matter where we evaluate the functions in the last two terms of the action Eq. (A3b). The functions in the first term, however, need to be evaluated precisely at $\boldsymbol{x}_\alpha$. Here and below, we shall often hide $\boldsymbol{x}, \lambda$ dependence in F^i and B^{ij}, to simplify the notations.

For our purpose, it is most convenient to use $\alpha = 1/2$ version of the transition probability (A3a)

$$d\mu(\boldsymbol{x}_1)P_\lambda(\boldsymbol{x}_1|\boldsymbol{x};dt) = \frac{d\mu(\boldsymbol{x}_1)\,e^{-A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda)}}{\sqrt{(4\pi dt)^n \det B^{ij}(\boldsymbol{x}_{1/2})}}, \tag{A4a}$$

$$A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda) = [dx^i - dt(F^i - \partial_k B^{ik})_{1/2}]\frac{B^{-1}_{ij}(\boldsymbol{x}_{1/2})}{4dt} \times [dx^j - dt(F^j - \partial_l B^{jl})_{1/2}] + \frac{1}{2}(\partial_i F^i)dt - \frac{1}{4}(\partial_i\partial_j B^{ij})dt. \tag{A4b}$$

Using Eqs. (A4a) as well as the facts that $d\mu(\boldsymbol{x}) = d\mu(\boldsymbol{x}^*)$, and $\det B^{ij}(\boldsymbol{x}) = \det B^{ij}(\boldsymbol{x}^*)$, we easily find

$$\log\frac{P_\lambda(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}^*_1;dt)} = A^{1/2}(\boldsymbol{x}^*,\boldsymbol{x}^*_1;dt,\lambda^*) - A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda). \tag{A5}$$

The systematic force $F^i(\boldsymbol{x},\lambda)$ in Eq. (A1) can be decomposed into

$$F^i_{\rm R}(\boldsymbol{x},\lambda) \equiv \tfrac{1}{2}\big(F^i(\boldsymbol{x},\lambda) - \epsilon_i F^i(\boldsymbol{x}^*,\lambda^*)\big), \tag{A6}$$

$$F^i_{\rm IR}(\boldsymbol{x},\lambda) \equiv \tfrac{1}{2}\big(F^i(\boldsymbol{x},\lambda) + \epsilon_i F^i(\boldsymbol{x}^*,\lambda^*)\big), \tag{A7}$$

$$F^i(\boldsymbol{x},\lambda) = F^i_{\rm IR}(\boldsymbol{x},\lambda) + F^i_{\rm R}(\boldsymbol{x},\lambda), \tag{A8}$$

where $F^i_{\rm R}(\boldsymbol{x},\lambda)$ and $F^i_{\rm IR}(\boldsymbol{x},\lambda)$ are respectively the reversible part and the irreversible part of the forces. From these we derive (no summation over repeated indices below)

$$\epsilon_i F^i_{\rm R}(\boldsymbol{x}^*,\lambda^*) = -F^i_{\rm R}(\boldsymbol{x},\lambda), \tag{A9a}$$

$$\epsilon_i F^i_{\rm IR}(\boldsymbol{x}^*,\lambda^*) = F^i_{\rm IR}(\boldsymbol{x},\lambda), \tag{A9b}$$

$$\epsilon_i F^i(\boldsymbol{x}^*,\lambda^*) = F^i_{\rm IR}(\boldsymbol{x},\lambda) - F^i_{\rm R}(\boldsymbol{x},\lambda). \tag{A9c}$$

The following result can be proved using Eqs. (A4) and (A5) (c.f. Eq. (A22) in the Appendix of Ref. [49])

$$\log\frac{P_\lambda(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}^*_1;dt)} = \big(dx^i - dt\,F^i_{\rm R}(\boldsymbol{y})\big)B^{-1}_{ij}(\boldsymbol{y})\big(F^{IR}_j(\boldsymbol{y}) - \partial_l B_{jl}(\boldsymbol{y})\big) - dt\,\partial_i F^i_{\rm R}(\boldsymbol{y}). \tag{A10}$$

Now for the covariant Langevin equation (3.1) with non-conservative forces, we have

$$F^i = \partial_j L^{ij} - L^{ij}(\partial_j U - \varphi_j), \tag{A11a}$$

Using the detailed balance conditions (3.4), we can easily show

$$F^i_{\rm R} = \partial_j Q^{ij} - Q^{ij}(\partial_j U - \varphi_j), \tag{A11b}$$

$$F^i_{\rm IR} = \partial_j B^{ij} - B^{ij}(\partial_j U - \varphi_j). \tag{A11c}$$

Substituting Eqs. (A11) back into Eq. (A10), we find

$$\log\frac{P_\lambda(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda^*}(\boldsymbol{x}^*|\boldsymbol{x}^*_1;dt)} = (-\partial_i U(\boldsymbol{y}) + \varphi_i(\boldsymbol{y}))dx^i - Q^{ij}(\partial_i\varphi_j)dt, = -d_{\boldsymbol{x}}U + \varphi_i \circ dx^i - Q^{ij}(\partial_i\varphi_j)dt, \tag{A12}$$

where $\varphi_i \circ dx^i$ is a product in Stratonovich's sense,

$$\varphi_i \circ dx^i = \varphi_i(\boldsymbol{x} + d\boldsymbol{x}/2)dx^i, \tag{A13}$$

and $d_{\boldsymbol{x}}U$ is the differential of $U(\boldsymbol{x})$,

$$d_{\boldsymbol{x}}U = \partial_i U \circ dx^i = \partial_i U(\boldsymbol{y})dx^i. \tag{A14}$$

2. Derivation of Eqs. (6.15)

Here we consider models with all variables even under time reversal. Detailed balance then demands $Q^{ij} = 0$, and hence $L^{ij} = B^{ij} = B^{ji}$. Equation (A11) becomes

$$F^i = \partial_j B^{ij} - B^{ij}(\partial_j U - \varphi_i). \tag{A15}$$

The action in Eq. (A4b) then becomes

$$A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,\varphi) = [dx^i + dtB^{ik}(\partial_k U - \varphi_k)_{1/2}]\frac{B^{-1}_{ij}}{4dt} \times [dx^j + dtB^{jl}(\partial_l U - \varphi_l)_{1/2}] - \frac{1}{2}\,\partial_i(B^{ij}(\partial_j U - \varphi_i))dt + \frac{1}{4}(\partial_i\partial_j B^{ij})dt. \tag{A16}$$

Note that we explicitly demonstrate the dependence of the action on the non-conservative force φ.

The forward, backward, adjoint, and adjoint backward processes are defined in Sec. VI B. See also Fig. 5 for illustration of infinitesimal steps of all these processes. The transition probabilities of the infinitesimal forward and backward process are given by [c.f. Eq. (A4a)]

$$P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt) = \frac{e^{-A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,\varphi)}}{\sqrt{(4\pi dt)^n \det B^{ij}(\boldsymbol{x}_{1/2})}}, \tag{A17}$$

$$P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt) = \frac{e^{-A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,\varphi)}}{\sqrt{(4\pi dt)^n \det B^{ij}(\boldsymbol{x}_{1/2})}}. \tag{A18}$$

The action $A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,\varphi)$ can be obtained from Eq. (A16) by exchange of $\boldsymbol{x},\boldsymbol{x}_1$, or equivalently, by replacement $d\boldsymbol{x} \to -d\boldsymbol{x}$,

$$A^{1/2}(\boldsymbol{x},\boldsymbol{x}_1;dt,\lambda,\varphi) = [-dx^i + dtB^{ik}(\partial_k U - \varphi_k)_{1/2}]\frac{B^{-1}_{ij}}{4dt}[-dx^j + dtB^{jl}(\partial_l U - \varphi_l)_{1/2}] - \frac{1}{2}\,\partial_i(B^{ij}(\partial_j U - \varphi_i))dt + \frac{1}{4}(\partial_i\partial_j B^{ij})dt. \tag{A19}$$

Taking the ration between Eqs. (A17) and (A18), and using Eqs. (A16) and (A19), we find

$$\log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)} = -A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,\varphi) + A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,\varphi)$$
$$= -d_{\boldsymbol{x}}U + \varphi_i \circ dx^i. \tag{A20}$$

The transition probability of the infinitesimal *adjoint process* and *adjoint backward process* are

$$P_{\lambda,-\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt) = \frac{e^{-A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,-\varphi)}}{\sqrt{(4\pi dt)^n \det B^{ij}(\boldsymbol{x}_{1/2})}}, \tag{A21}$$

$$P_{\lambda,-\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt) = \frac{e^{-A^{1/2}(\boldsymbol{x},\boldsymbol{x}_1;dt,\lambda,-\varphi)}}{\sqrt{(4\pi dt)^n \det B^{ij}(\boldsymbol{x}_{1/2})}}, \tag{A22}$$

where the action $A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,-\varphi)$, $A^{1/2}(\boldsymbol{x},\boldsymbol{x}_1;dt,\lambda,-\varphi)$ can be obtained from Eq. (A16) and (A19) by changing the sign of φ,

$$A^{1/2}(\boldsymbol{x}_1,\boldsymbol{x};dt,\lambda,-\varphi) = [dx^i + dtB^{ik}(\partial_k U + \varphi_k)_{1/2}]\frac{B^{-1}_{ij}}{4dt}[dx^j + dtB^{jl}(\partial_l U + \varphi_l)_{1/2}]$$
$$-\frac{1}{2}\,\partial_i(B^{ij}(\partial_j U + \varphi_i))dt + \frac{1}{4}(\partial_i\partial_j B^{ij})dt. \tag{A23}$$

$$A^{1/2}(\boldsymbol{x},\boldsymbol{x}_1;dt,\lambda,-\varphi) = [-dx^i + dtB^{ik}(\partial_k U + \varphi_k)_{1/2}]\frac{B^{-1}_{ij}}{4dt}[-dx^j + dtB^{jl}(\partial_l U + \varphi_l)_{1/2}]$$
$$-\frac{1}{2}\,\partial_i(B^{ij}(\partial_j U + \varphi_i))dt + \frac{1}{4}(\partial_i\partial_j B^{ij})dt. \tag{A24}$$

Using the above results, we can prove

$$\log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,-\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)} = \log \frac{P_{\lambda,-\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)} = -d_{\boldsymbol{x}}U - e^U\,\partial_i(B^{ij}\varphi_j e^{-U})dt = -d_{\boldsymbol{x}}U = -\beta d\mathcal{Q}^{\rm ex}, \tag{A25}$$

where in the third equality we have used the Gibbs gauge condition Eq. (6.6). This is Eq. (6.15a). Similarly we can prove Eq. (6.15b)

$$\log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,-\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)} = \log \frac{P_{\lambda,-\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)} = \varphi_i \circ dx^i + e^U\,\partial_i(B^{ij}\varphi_j e^{-U})dt = \varphi_i \circ dx^i = -\beta d\mathcal{Q}^{\rm hk}. \tag{A26}$$

Using the preceding two equations, we obtain two equivalent decomposition of Eq. (A20)

$$d\mathcal{S}_{\rm EV} = \log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)} = -d_{\boldsymbol{x}}U + \varphi_i \circ dx^i$$
$$= \log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,-\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)} + \log \frac{P_{\lambda,-\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)}$$
$$= \log \frac{P_{\lambda,\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,-\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)} + \log \frac{P_{\lambda,-\varphi}(\boldsymbol{x}_1|\boldsymbol{x};dt)}{P_{\lambda,\varphi}(\boldsymbol{x}|\boldsymbol{x}_1;dt)}. \tag{A27}$$

[1] S.-i. Sasa and H. Tasaki, Steady state thermodynamics, J. Stat. Phys. **125**, 125 (2006).

[2] D. Groot, S. Ruurds, and P. Mazur, *Non-Equilibrium Thermodynamics* (Courier Corporation, North Chelmsford, 2013).

[3] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (John Wiley & Sons, New York, 1998).

[4] G. F. Mazenko, *Nonequilibrium Statistical Mechanics* (John Wiley & Sons, New York, 2008).

[5] D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (John Wiley & Sons, New York, 2014).

[6] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).

[7] P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (J. Willey & Sons, New York, 1971).

[8] L. Onsager, Reciprocal relations in irreversible processes. I, Phys. Rev. **37**, 405 (1931).

[9] L. Onsager, Reciprocal relations in irreversible processes. II, Phys. Rev. **38**, 2265 (1931).

[10] H. B. G. Casimir, On Onsager's principle of microscopic reversibility, Rev. Mod. Phys. **17**, 343 (1945).

[11] L. Onsager and S. Machlup, Fluctuations and irreversible processes, Phys. Rev. **91**, 1505 (1953).

[12] S. Machlup and L. Onsager, Fluctuations and irreversible process. II. Systems with kinetic energy, Phys. Rev. **91**, 1512 (1953).

[13] R. Kubo, The fluctuation-dissipation theorem, Rep. Prog. Phys. **29**, 255 (1966).

[14] J. Schnakenberg, Network theory of microscopic and macroscopic behavior of master equation systems, Rev. Mod. Phys. **48**, 571 (1976).

[15] P. Glansdorff, G. Nicolis, and I. Prigogine, The thermodynamic stability theory of non-equilibrium states, Proc. Natl. Acad. Sci. USA **71**, 197 (1974).

[16] J. Keizer and R. F. Fox, Qualms regarding the range of validity of the Glansdorff-Prigogine criterion for stability of non-equilibrium states, Proc. Natl. Acad. Sci. USA **71**, 192 (1974).

[17] Y. Oono and M. Paniconi, Steady state thermodynamics, Prog. Theor. Phys. Suppl. **130**, 29 (1998).

[18] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[19] L. Peliti and S. Pigolotti, *Stochastic Thermodynamics: An Introduction* (Princeton University Press, Princeton, 2021).

[20] C. Van den Broeck, Stochastic thermodynamics: A brief introduction, Phys. Complex Colloids **184**, 155 (2013).

[21] C. Jarzynski, Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale, Annu. Rev. Condens. Matter Phys. **2**, 329 (2011).

[22] G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Phys. Rev. E **60**, 2721 (1999).

[23] G. Gallavotti, and E. G. D. Cohen, Dynamical ensembles in stationary states, J. Stat. Phys. **80**, 931 (1995).

[24] J. L. Lebowitz and H. Spohn, A Gallavott-Cohen-type symmetry in the large deviation functional for stochastic dynamics, J. Stat. Phys. **95**, 333 (1999).

[25] T. Hatano and S.-i. Sasa, Steady-State Thermodynamics of Langevin Systems, Phys. Rev. Lett. **86**, 3463 (2001).

[26] T. Speck, and U. Seifert, Integral fluctuation theorem for the housekeeping heat, J. Phys. A: Math. Gen. **38**, L581 (2005).

[27] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, Path-integral analysis of fluctuation theorems for general Langevin processes, J. Stat. Mech. (2006) P08001.

[28] M. Esposito, and C. Van den Broeck, Three Detailed Fluctuation Theorems, Phys. Rev. Lett. **104**, 090601 (2010).

[29] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, Phys. Rev. E **82**, 011143 (2010).

[30] C. Van den Broeck and M. Esposito, Three faces of the second law. II. Fokker-Planck formulation, Phys. Rev. E **82**, 011144 (2010).

[31] H. Ge and H. Qian, Physical origins of entropy production, free energy dissipation, and their mathematical representations, Phys. Rev. E **81**, 051133 (2010).

[32] R. E. Spinney and I. J. Ford, Nonequilibrium Thermodynamics of Stochastic Systems with Odd and Even Variables, Phys. Rev. Lett. **108**, 170603 (2012).

[33] R. E. Spinney and I. J. Ford, EP in full phase space for continuous stochastic dynamics, Phys. Rev. E **85**, 051113 (2012).

[34] H. K. Lee, C. Kwon, and H. Park, Fluctuation Theorems and Entropy Production with Odd-Parity Variables, Phys. Rev. Lett. **110**, 050602 (2013).

[35] H. Qian, Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation, Phys. Rev. E **65**, 016102 (2001).

[36] H. Qian, Entropy production and excess entropy in a nonequilibrium steady-state of single macromolecules, Phys. Rev. E **65**, 021111 (2002).

[37] K. H. Kim and H. Qian, EP of Brownian Macromolecules with Inertia, Phys. Rev. Lett. **93**, 120602 (2004).

[38] K. H. Kim and H. Qian, Fluctuation theorems for a molecular refrigerator, Phys. Rev. E **75**, 022102 (2007).

[39] T. Munakata and M. L. Rosinberg, EP and fluctuation theorems under feedback control: The molecular refrigerator model revisited, J. Stat. Mech. (2012) P05010.

[40] M. L. Rosinberg, T. Munakata, and G. Tarjus, Stochastic thermodynamics of Langevin systems under time-delayed feedback control: Second-law-like inequalities, Phys. Rev. E **91**, 042114 (2015).

[41] F. Liu, H. Tong, R. Ma, and Z.-c. Ou-Yang, Linear response theory and transient fluctuation relations for diffusion processes: A backward point of view, J. Phys. A: Math. Theor. **43**, 495003 (2010).

[42] G. Li and Z. C. Tu, Stochastic thermodynamics with odd controlling parameters, Phys. Rev. E **100**, 012127 (2019).

[43] K. Yoshimura, A. Kolchinsky, A. Dechant, and S. Ito, Geometrical approach to excess/housekeeping EP in discrete systems, arXiv:2205.15227.

[44] C. Maes and K. Netočný, A nonequilibrium extension of the Clausius heat theorem, J. Stat. Phys. **154**, 188 (2014).

[45] T. Speck, J. Mehl, and U. Seifert, Role of External Flow and Frame Invariance in Stochastic Thermodynamics, Phys. Rev. Lett. **100**, 178302 (2008).

[46] C. Maes and K. Netočný, Revisiting the Glansdorff–Prigogine criterion for stability within irreversible thermodynamics, J. Stat. Phys. **159**, 1286 (2015).

[47] J. Wang, Landscape and flux theory of non-equilibrium dynamical systems with application to biology, Adv. Phys. **64**, 1 (2015).

[48] M. Ding, Z. Tu, and X. Xing, Covariant formulation of nonlinear Langevin theory with multiplicative Gaussian white noises, Phys. Rev. Res. **2**, 033381 (2020).

[49] M. Ding and X. Xing, Covariant nonequilibrium thermodynamics from Ito-Langevin dynamics, Phys. Rev. Res. **4**, 033247 (2022).

[50] M. Ding, Z. Tu, and X. Xing, Strong coupling thermodynamics and stochastic thermodynamics from the unifying perspective of time-scale separation, Phys. Rev. Res. **4**, 013015 (2022).

[51] H. Qian, Thermodynamics of the general diffusion process: Equilibrium supercurrent and nonequilibrium driven circulation with dissipation, Eur. Phys. J.: Spec. Top. **224**, 781 (2015).

[52] P. Ao, C. Kwon, and H. Qian, On the existence of potential landscape in the evolution of complex systems, Complexity **12**, 19 (2007).

[53] C. Kwon, P. Ao, and David J. Thouless, Structure of stochastic dynamics near fixed points, Proc. Natl. Acad. Sci. USA **102**, 13029 (2005).

[54] Strictly speaking, $-\partial_j U$ are thermodynamic forces in irreversible thermodynamics, which corresponds to the

deterministic limit of the covariant Langevin dynamics. For details, see Sec. V A of Ref. [49].

[55] U. Seifert, First and Second Law of Thermodynamics at Strong Coupling, Phys. Rev. Lett. **116**, 020601 (2016).

[56] P. Talkner and P. Hänggi, Open system trajectories specify fluctuating work but not heat, Phys. Rev. E **94**, 022143 (2016).

[57] C. Jarzynski, Stochastic and Macroscopic Thermodynamics of Strongly Coupled Systems, Phys. Rev. X **7**, 011008 (2017).

[58] P. Talkner and P. Hánggi, *Colloquium:* Statistical mechanics and thermodynamics at strong coupling: Quantum and classical, Rev. Mod. Phys. **92**, 041002 (2020).

[59] C. Jarzynski, NonEQ work theorem for a system strongly coupled to a thermal environment, J. Stat. Mech. P09005 (2004).

[60] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed., (Springer, Berlin, 2004).

[61] H. Risken, *Fokker-planck equation*, The Fokker-Planck Equation (Springer, Berlin, Heidelberg, 1996), p63-p95.

[62] T. Sagawa, Thermodynamics of information processing in small systems, Prog. Theor. Phys. **127**, 1 (2012).

[63] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[64] R. Filliger and P. Reimann, Brownian Gyrator: A Minimal Heat Engine on the Nanoscale, Phys. Rev. Lett. **99**, 230602 (2007).

[65] G. N. Bochkov and Yu. E. Kuzovlev, Fluctuation–dissipation relations. Achievements and misunderstandings, Phys. Usp. **56**, 590 (2013).

[66] C. Jarzynski, Comparison of far-from-equilibrium work relations, C. R. Phys. **8**, 495 (2007).

[67] C. Huang, M. Ding, and X. Xing, Information swimmer: Self-propulsion without energy dissipation, Phys. Rev. Res. **2**, 043222 (2020).

[68] J. C. Slonczewski, Current-driven excitation of magnetic multilayers, J. Magn. Magn. Mater. **159**, L1 (1996).

[69] C. Serpico, I. D. Mayergoyz, G. Bertotti, M. d'Aquino, and R. Bonin, Generalized Landau–Lifshitz–Gilbert equation for uniformly magnetized bodies, Phys. B: Condens. Matter **403**, 282 (2008).

[70] C. Aron, D. G. Barci, L. F. Cugliandolo, Z. G. Arenas, and G. S. Lozano, Magnetization dynamics: Path-integral formalism for the stochastic Landau–Lifshitz–Gilbert equation, J. Stat. Mech. (2014) P09008.

[71] M. Nishino and S. Miyashita, Realization of the thermal equilibrium in inhomogeneous magnetic systems by the Landau-Lifshitz-Gilbert equation with stochastic noise, and its dynamical aspects, Phys. Rev. B **91**, 134411 (2015).

[72] H. Tutu and T. Horita, Stochastic Landau-Lifshitz-Gilbert equation with delayed feedback field, Prog. Theor. Phys. **120**, 315 (2008).

[73] D. A. Garanin, Fokker-Planck and Landau-Lifshitz-Bloch equations for classical ferromagnets, Phys. Rev. B **55**, 3050 (1997).

[74] R. F. L. Evans, D. Hinzke, U. Atxitia, U. Nowak, R. W. Chantrell, and O. Chubykalo-Fesenko, Stochastic form of the Landau-Lifshitz-Bloch equation, Phys. Rev. B **85**, 014433 (2012).

[75] U. Atxitia, D. Hinzke, and U. Nowak, Fundamentals and applications of the Landau–Lifshitz–Bloch equation, J. Phys. D: Appl. Phys. **50**, 033003 (2017).

[76] M. Ding, and X. Xing, Time-Slicing Path-integral in Curved Space, Quantum **6**, 694 (2022).