


## Housekeeping and excess entropy production for general nonlinear dynamics

Kohei Yoshimura <sup>1,\*</sup> Artemy Kolchinsky <sup>2,†</sup> Andreas Dechant <sup>3,†</sup> and Sosuke Ito <sup>1,2</sup>

<sup>1</sup>*Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

<sup>2</sup>*Universal Biology Institute, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

<sup>3</sup>*Department of Physics No. 1, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*



(Received 3 June 2022; revised 29 June 2022; accepted 21 December 2022; published 17 January 2023)

We propose a housekeeping/excess decomposition of entropy production for general nonlinear dynamics in a discrete space, including chemical reaction networks and discrete stochastic systems. We exploit the geometric structure of thermodynamic forces to define the decomposition; this does not rely on the notion of a steady state, and it even applies to systems that exhibit multistability, limit cycles, and chaos. In the decomposition, distinct aspects of the dynamics contribute separately to entropy production: the housekeeping part stems from a cyclic mode that arises from external driving, generalizing Schnakenberg’s cyclic decomposition to nonsteady states, while the excess part stems from an instantaneous relaxation mode that arises from conservative forces. Our decomposition refines previously known thermodynamic uncertainty relations and speed limits. In particular, it not only improves an optimal-transport-theoretic speed limit, but it also extends the optimal transport theory of discrete systems to nonlinear and nonconservative settings.

DOI: [10.1103/PhysRevResearch.5.013017](https://doi.org/10.1103/PhysRevResearch.5.013017)

### I. INTRODUCTION

The notion of “nonequilibrium” has two aspects: breaking of detailed balance, and nonstationarity. The housekeeping/excess decomposition of entropy production rate (EPR) is a way to deal with these two aspects of nonequilibrium systems [1–4]. Such a decomposition has previously been formulated based on steady states: the housekeeping EPR characterizes how detailed balance is broken in a steady state, while the excess part quantifies the additional dissipation that is needed to transition between steady states. The best known decomposition of this kind was proposed by Hatano and Sasa (HS) [2] (also known as the adiabatic/nonadiabatic decomposition [3]), while an important alternative decomposition was proposed by Maes and Netočný (MN) [4]. The practical utility of these decompositions is that they provide meaningful thermodynamic bounds for nonequilibrium systems where the integrated entropy production (EP) diverges by discounting the housekeeping EP, which also diverges, to give a finite excess EP [1].

However, previously known decompositions have a crucial limitation. Namely, they are not defined for general nonlinear dynamics, such as general chemical reaction networks. Such systems are not always globally stable and can exhibit various nontrivial phenomena such as limit cycles, bifurcations, and

multistability, in contrast to existing EPR decompositions, which assume global stability. More precisely, the HS decomposition has been generalized to chemical reaction networks only if there is a special steady state, called a complex balanced steady state [5,6], which is always globally stable [7]. However, even if a chemical reaction network has a stable steady state, it may not be complex balanced. In that case, it is possible for the HS decomposition to give a negative value of excess EPR, which makes it difficult to interpret the decomposition physically. Because a complex balanced steady state is “almost detailed balanced” in the sense that it always shows global stability and cannot exhibit genuine nonlinear phenomena such as multistability or chaos [8], this particular generalization does not help us to understand the various physically interesting situations that play a key role in biology and other complex chemical systems [9].

In this paper, we propose an EPR decomposition that applies to general nonlinear dynamics in a discrete space, which includes Markov jump processes and chemical reaction networks as special cases (Fig. 1). Because the decomposition is defined based on the geometric structure of thermodynamic forces, it does not require any information about the steady state or even its existence. Since it is not related to steady states, the decomposition is not only applicable to a wider range of dynamics, but it also provides crucial insights into the physical meaning of the individual terms: it allows us to interpret each term of the decomposition based on physical modes of dynamics, beyond the notion of stationarity. The housekeeping part reflects a cyclic mode on the graph of states that does not affect the apparent dynamics, and it quantifies how external driving breaks detailed balance at a given instant of time. As a result, we generalize Schnakenberg’s cyclic decomposition of steady-state dissipation into contributions stemming from cycles in the graph of states [10]. On the

\*kyoshimura@ubi.s.u-tokyo.ac.jp

†These authors contributed equally to this work.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

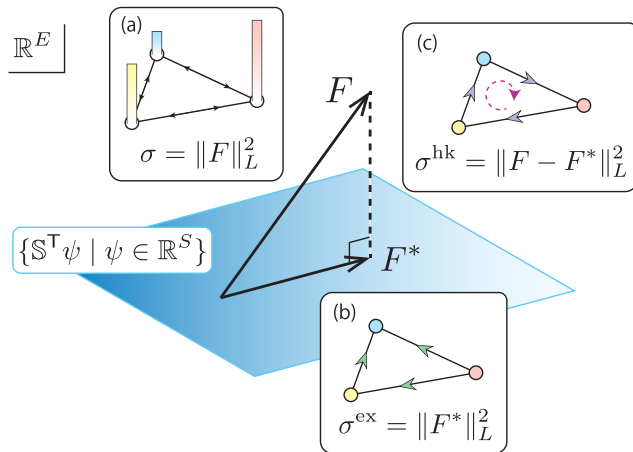


FIG. 1. We propose a decomposition of (a) the total entropy production rate (EPR) in general nonlinear dynamics in a discrete space (e.g., a chemical reaction network, and a Markov jump process) into (b) the excess part that reflects a relaxation mode, and (c) the housekeeping part that reflects a cyclic mode. The total EPR  $\sigma$  is expressed as the square norm of the thermodynamic force  $F \in \mathbb{R}^E$  with the metric  $L$ . The projection of the force onto the space of conservative forces  $\{S^T \psi \mid \psi \in \mathbb{R}^S\} \subset \mathbb{R}^E$  defined in Eq. (31) determines the conservative force  $F^*$  that provides the Onsager excess EPR  $\sigma^{\text{ex}}$ , which can be expressed as the sum of cyclic contributions as in Eq. (43). Because  $F^*$  and  $F - F^*$  are orthogonal with respect to the inner product given by  $L$ ,  $\sigma^{\text{ex}}$  and  $\sigma^{\text{hk}}$  decompose the total EPR,  $\sigma = \sigma^{\text{ex}} + \sigma^{\text{hk}}$ .

other hand, the excess term expresses a “pseudorelaxation” mode even if the system does not relax to a steady state by extracting a conservative component of the thermodynamic force. Because a conservative force is given by a potential difference of the internal states, it represents an internal free-energy-like contribution to dissipation. Using a corresponding instantaneous “pseudoequilibrium,” we obtain a gradient-flow form of an equation of motion, which was previously used to describe the relaxation to an equilibrium state [11,12]. In summary, our result allows us not only to decompose the EPR in general nonlinear dynamics in a discrete space, but also to separate the effects on dissipation stemming from the external cyclic mode and the internal relaxation mode.

While this physical interpretation of our decomposition is different from previous results, it is also consistent with existing theory of nonequilibrium thermodynamics and effectively refines several results. It gives much tighter thermodynamic uncertainty relations (TURs) and speed limits, which is concordant with the original philosophy of EPR decomposition [1] and strengthens its physical validity. Concretely, we derive refined versions of short-time and finite-time TURs for Markov jump processes [13–29], and a short-time TUR for chemical reaction networks [30]. We also generalize an optimal-transport-theoretic speed limit derived in [31] to general nonequilibrium systems.

More generally, we discuss connections between our decomposition and optimal transport theory. Optimal transport theory interprets the minimum transportation cost from one

probability distribution to another as a distance between them [32], providing a natural geometry for probability distributions. Recent studies have revealed that one can obtain thermodynamic bounds in continuous systems through this geometrical approach, associating thermodynamic costs with the optimal-transport-theoretic distance [28,29,33–40]. In particular, a distance called  $L^2$ -Wasserstein distance provides a very intuitive type of speed limit  $\tau \geq \mathcal{W}^2/\Sigma$ , with  $\tau$  being the time the system takes to change,  $\mathcal{W}$  is the  $L^2$ -Wasserstein distance between the initial and final state, and  $\Sigma$  is the integrated EP. However, such a simple bound has not been known for general discrete systems, except for a result restricted to detailed balanced stochastic systems [41], because the discretized  $L^2$ -Wasserstein distance has only been defined for such systems [42–44]. In this paper, we show that the local geometrical structure that defines our EPR decomposition reveals the physical meaning of the discrete  $L^2$ -Wasserstein distance as minimum excess EP, and it allows us to generalize it to general nonlinear dynamics in a discrete space. As a result, we obtain a simple speed limit for general nonlinear dynamics in a discrete space, in which the total integrated EP is replaced by the integrated excess EP, tightening the bound.

The paper is organized as follows. In Sec. II, we introduce basic concepts of nonlinear dynamics in a discrete space, which we characterize as a graph, and we propose a general equation of motion that contains master equations and chemical rate equations as special cases. We also define the edgewise Onsager coefficients and the induced inner product, which determine the aforementioned geometrical structure of thermodynamic forces. In addition, we review some relevant previous studies. In Sec. III, we derive the decomposition of the EPR, and we examine the properties of the excess and housekeeping EPRs, in particular the cyclic decomposition of the housekeeping EPR and a connection between gradient flow and the excess EPR. In Sec. IV, we apply this decomposition to derive refined versions of several TURs. In Sec. V, we introduce the generalized  $L^2$ -Wasserstein distance and examine its mathematical and physical features. We relate it to the excess EPR in Sec. VB, and we obtain the speed limit in Sec. VC. Section VI is devoted to the study of two examples that demonstrate how the decomposition of the EPR is obtained and characterize the distinct dynamical modes separately, both analytically and numerically. The first example is a two-level system attached to two heat baths at different temperatures, where the relaxation mode is described by a “mean” temperature of the two. The second example is the Brusselator model of chemical oscillation, which exhibits a limit cycle. The Appendixes contain mathematical details of the derivations and results.

## II. GENERAL FORMALISM OF DYNAMICS AND THERMODYNAMICS

### A. Dynamics

Let us consider general nonlinear Markovian dynamics on a graph whose nodes may include several “states” or “species.” It will be shown that the dynamics can lead to master equations and rate equations. Graph theoretical notions introduced below are summarized in Figs. 2 and 3 through

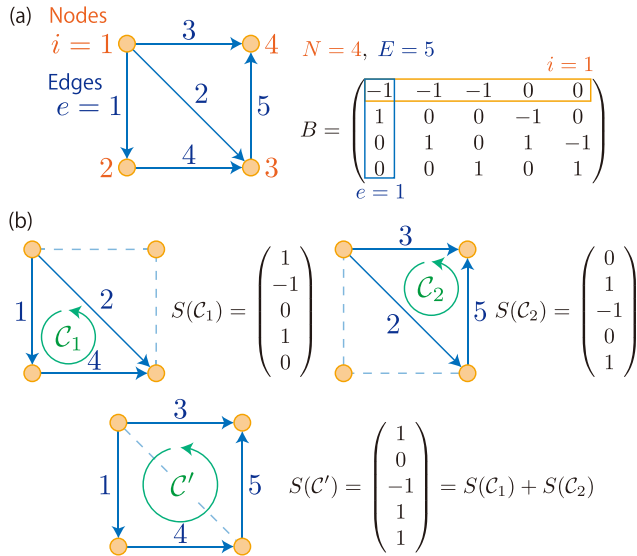


FIG. 2. Graph theoretical concepts are shown. (a) An example of a graph. There are  $N = 4$  nodes and  $E = 5$  directed edges. Each edge indicates a pair of reversible transitions or reactions. The incidence matrix  $B$  encodes the structural information of the graph. (b) In a Markov jump process, where each node is identified with a physical state, we define a cycle as a chain of transitions that leaves the system unchanged. Now there are two independent cycles,  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , whose corresponding vectors  $S(\mathcal{C}_1)$  and  $S(\mathcal{C}_2)$  span linear space  $\ker B$ . The other cycles, such as  $\mathcal{C}'$ , are given by superposing the two cycles. The notion of a cycle differs a bit in a chemical reaction network, as we explain in the next figure.

examples of Markov jump process and chemical reaction network respectively.

We consider a directed graph with nodes  $\{1, 2, \dots, N\}$  and edges  $\{1, 2, \dots, E\}$ . The stoichiometry of state or species  $\alpha \in \{1, 2, \dots, S\}$  at node  $i$  is indicated by  $v_{\alpha i} \in \mathbb{Z}_{\geq 0}$ . The distribution (concentration vector or statistical state) of the system at time  $t$  is indicated by a vector  $x(t) = [x_\alpha(t)] \in \mathbb{R}_{\geq 0}^S$ .

An edge represents a pair of reversible transitions or reactions and connects two nodes. We denote the starting node of edge  $e \in \{1, \dots, E\}$  as  $\iota(e)$  and its ending node by  $\iota'(e)$ . The reversed direction of edge  $e$  is indicated by  $-e$ . The structure of the graph is expressed by the incidence matrix  $B = (B_{ie})$  defined by

$$B_{ie} = \begin{cases} 1 & \text{if } \iota'(e) = i, \\ -1 & \text{if } \iota(e) = i, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

We write the possibly nonlinear and time-dependent flux of  $\iota(e) \rightarrow \iota'(e)$  as  $\mathcal{K}_e(x)$  and  $\iota'(e) \rightarrow \iota(e)$  as  $\mathcal{K}_{-e}(x)$ . We assume the fluxes are always positive,  $\mathcal{K}_{\pm e}(x) > 0$ . The current  $J_e(x)$  on edge  $e$  is given by  $J_e(x) = \mathcal{K}_e(x) - \mathcal{K}_{-e}(x)$ . Therefore, the net inflow to node  $i$  amounts to  $\sum_e B_{ie} J_e(x)$ . Defining  $\mathbb{S} = \nu B$ , where  $\nu = (\nu_{\alpha i})$ , we obtain a continuity equation

$$\frac{dx}{dt} = \mathbb{S}J(x(t)). \quad (2)$$

Throughout this paper, we consider general Markovian dynamics of a distribution  $x$  governed by this equation. It can

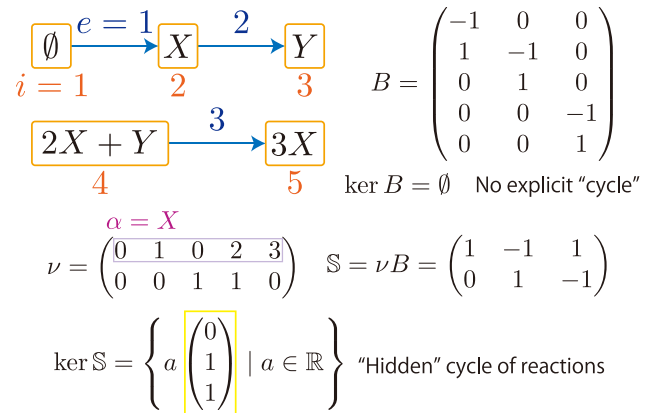


FIG. 3. An example of a graph theoretic description of a chemical reaction network. Note that although each edge has a specific direction, it represents a set of reversible reactions. The reaction network is the Brusselator, a paradigmatic model of chemical oscillation, which we consider as an example in Sec. VI B. Unlike Markov jump processes, a node in a chemical reaction network represents a so-called *complex* that contains some chemical species as represented by  $\nu$ . Complex  $\emptyset$  corresponds to external species whose concentrations are kept constant. Here we cannot find a “cycle” by just looking at the diagram, or equivalently, the incidence matrix  $B$ . However, the stoichiometric coefficient matrix  $\mathbb{S}$ , which governs the dynamics, has a nonempty kernel, which proves that there exists a cycle of reactions that does not change the system.

represent Markov jump processes and deterministic chemical reaction networks, as we show below. We sometimes assume that the flux is given by the law of mass action,

$$\mathcal{K}_e(x) = k_e \prod_{\alpha} x_{\alpha}^{\nu_{\alpha \iota(e)}}, \quad (3)$$

which, however, is not essential to our general results. Below, we explicitly note whenever we make the assumption of mass action kinetics. In addition, when mass action is used, the coefficients  $k_e(t) > 0$  may depend on time.

We sometimes abbreviate the dependence of some quantity  $f(x(t), t)$  on the state  $x(t)$  at time  $t$  and the explicit dependence on time  $t$  as  $f(t)$  if there is no ambiguity; we make the dependence implicit in the majority of the paper.

### 1. Example: Markov jump process

Consider the case in which each node  $i$  is identified with a state  $\alpha$ , so  $\nu = I$ , the identity. Then, by writing  $x$  as  $p$  and assuming the mass action law, Eq. (2) becomes the master equation

$$\frac{dp_i}{dt} = \sum_e B_{ie} (k_e p_{\iota(e)} - k_{-e} p_{\iota'(e)}), \quad (4)$$

where  $p_i$  can be interpreted as the occupation probability of state  $i$  if normalized  $\sum_i p_i = 1$ , and  $k_{\pm e}$  is the transition rate along edge  $\pm e$ . It can be written as

$$\frac{dp_i}{dt} = \sum_{j(\neq i)} (k_{j \rightarrow i} p_j - k_{i \rightarrow j} p_i) \quad (5)$$

because if  $e = i \rightarrow j$ ,  $B_{ie} = -1$  and if  $e = j \rightarrow i$ ,  $B_{ie} = 1$ , and only one of the two directions is contained in the summation in Eq. (4). We note that the linear form of flux is an example of mass action kinetics since

$$\mathcal{K}_e(p) = k_e \prod_{\alpha} p_{\alpha}^{\delta_{\alpha(e)}} = k_e p_{i(e)}, \quad (6)$$

where  $\delta$  is Kronecker's delta.

## 2. Example: Chemical reaction network

The general equation (2) can represent an open chemical reaction network by interpreting  $x$  as a concentration distribution  $c$  [5]. A node is interpreted as an aggregate of chemical species, technically called *complex*, which contains  $\nu_{\alpha i}$  molecules of chemical species  $\alpha$  and appears on the left or right-hand side of a chemical equation. Because the  $(\alpha, e)$ -element of  $\mathbb{S} = \nu B$  reads

$$\sum_i \nu_{\alpha i} B_{ie} = \nu_{\alpha i'(e)} - \nu_{\alpha i(e)}, \quad (7)$$

$\mathbb{S}$  is the conventional stoichiometric matrix.

When we consider chemical reaction networks, we do not assume the mass action kinetics in general so that we can treat a broad class of nonideal chemical reaction networks [45]. We put the assumption only when (i) we review previous studies that require it, e.g., gradient flow equations [11] and the HS decomposition [5,6], or (ii) we illustrate our results numerically where we need to provide a concrete form of  $\mathcal{K}$ .

When we discuss a result particular to a Markov jump process, we use the variable as  $p$  instead of  $x$  to indicate that the variable is a probability distribution; on the other hand, when we focus on a chemical reaction network, we will use  $c$  to indicate a concentration distribution.

## B. Thermodynamics

Let us define the thermodynamic force on edge  $e$  by

$$F_e(x) := \ln \frac{\mathcal{K}_e(x)}{\mathcal{K}_{-e}(x)}. \quad (8)$$

We assume the local detailed balance condition, namely that  $F_e$  gives the total change of the reduced entropy on the process that edge  $e$  represents [5,10,45], where “reduced” means that it is divided by a physical constant, the Boltzmann constant  $k_B$  for Markov jump processes, or the gas constant  $R$  for chemical reaction networks. We do not distinguish the entropy and the reduced one by setting the constant we consider to 1. Then, we obtain the following form of the entropy production rate (EPR):

$$\sigma(x) = \sum_e J_e(x) F_e(x). \quad (9)$$

For simplicity, hereafter we do not write the argument  $x$  explicitly.

In the seminal paper [10], Schnakenberg obtained an expression of the EPR by decomposing the network into cycles. He introduced the fundamental set of cycles  $\{\mathcal{C}_{\mu}\}_{\mu=1,\dots,M}$  of the graph defined by states and transitions in a Markov jump process. Each cycle corresponds to a vector  $S(\mathcal{C}_{\mu}) = [S_e(\mathcal{C}_{\mu})]_{e=1,\dots,E} \in \ker B$ , and the set  $\{S(\mathcal{C}_{\mu})\}_{\mu=1,\dots,M}$  forms a

basis of  $\ker B$ , where  $\ker$  indicates the kernel of the matrix. Such a set of vectors can also be introduced for a chemical reaction network, where it spans  $\ker \mathbb{S}$  [46]. In general, we refer to a basis of  $\ker \mathbb{S}$  rooted in the cycles of a graph as a cycle basis. A cycle basis is not unique in general; hereafter, we fix some choice of the basis. See also Figs. 2 and 3, where we provide concrete examples of cycles and clarify the slight difference between Markov jump processes and chemical reaction networks.

When the system is in a steady state, the current  $J^{\text{st}}$  is in the kernel of  $\mathbb{S}$ , so it can be expanded in the cycle basis as

$$J_e^{\text{st}} = \sum_{\mu} S_e(\mathcal{C}_{\mu}) \mathcal{J}_{\mu}. \quad (10)$$

Note that the superscript  $\text{st}$  is used to indicate the value of a quantity in the steady state. Each expansion coefficient  $\mathcal{J}_{\mu}$  quantifies the current flowing in the cycle  $\mathcal{C}_{\mu}$ . The conjugated cyclic forces are defined by

$$\mathcal{F}_{\mu} := \sum_e S_e(\mathcal{C}_{\mu}) F_e^{\text{st}}. \quad (11)$$

In terms of the currents and forces, we obtain the expression of the EPR as in Refs. [10,46],

$$\sigma^{\text{st}} = \sum_{\mu} \mathcal{J}_{\mu} \mathcal{F}_{\mu}, \quad (12)$$

which shows that the steady-state EPR can be divided into contributions from cycles, whose dissipation can be analyzed separately [47–49]. We will generalize this expression out of steady states in Sec. III A.

## C. Edgewise Onsager coefficient, gradient flow, and Wasserstein distance

We define the edgewise Onsager coefficient by

$$\ell_e(x) := \frac{\mathcal{K}_e(x) - \mathcal{K}_{-e}(x)}{\ln(\mathcal{K}_e(x)/\mathcal{K}_{-e}(x))} = \frac{J_e(x)}{F_e(x)} \quad (13)$$

and the diagonal matrix  $L(x) := \text{diag}(\ell_e(x))$ . When  $\mathcal{K}_e(x) = \mathcal{K}_{-e}(x)$ , by continuity we see  $\ell_e(x)$  is given by  $\mathcal{K}_e(x)$ . It is also easy to see that  $\ell_e(x) > 0$  and  $L(x)$  is positive-definite. For simplicity, we make  $L$  and  $\ell$ 's dependence on  $x$  implicit, but we stress that we always choose the current distribution  $x$  as the argument. We write  $L^{1/2} = \text{diag}(\sqrt{\ell_e})$ . For notational convenience, we define an inner product  $\langle u, v \rangle_L := u^T L v$  and the corresponding norm  $\|u\|_L := \sqrt{\langle u, u \rangle_L}$ . We can use this inner product to express the EPR as

$$\sigma = \|F\|_L^2, \quad (14)$$

which will be the starting point of our results [50]. Note that if we choose any other state like a steady state as the argument of  $L(\cdot)$ , then we can no longer give the EPR as a squared norm with  $L$  being the metric.

We remark that as a general property of the logarithmic mean,  $\ell_e(x)$  is sandwiched between arithmetic mean  $[\mathcal{K}_e(x) + \mathcal{K}_{-e}(x)]/2$  and geometric mean  $\sqrt{\mathcal{K}_e(x)\mathcal{K}_{-e}(x)}$  as

$$\sqrt{\mathcal{K}_e(x)\mathcal{K}_{-e}(x)} \leq \ell_e(x) \leq \frac{\mathcal{K}_e(x) + \mathcal{K}_{-e}(x)}{2}, \quad (15)$$

where the equalities hold if and only if  $\mathcal{K}_e(x) = \mathcal{K}_{-e}(x)$ .

In Ref. [10], Schnakenberg introduced a symmetric matrix  $\mathcal{L}$  and the reciprocal relation

$$\mathcal{J}_\mu = \sum_{\mu'} \mathcal{L}_{\mu\mu'} \mathcal{F}_{\mu'} \quad (16)$$

for the steady-state cycle current and force defined in Eqs. (10) and (11). By the definitions of  $\mathcal{J}$  and  $\mathcal{F}$ , we find that

$$\tilde{L}_{ee'} := \sum_{\mu, \mu'} S_e(C_\mu) \mathcal{L}_{\mu\mu'} S_{e'}(C_{\mu'}) \quad (17)$$

connects  $J^{\text{st}}$  and  $F^{\text{st}}$  as  $J^{\text{st}} = \tilde{L} F^{\text{st}}$ . However,  $\tilde{L}$  is not diagonal in general and does not coincide with  $L^{\text{st}}$ .

Nevertheless, the edgewise Onsager coefficient is useful to rewrite the equation of motion (2) into a gradient flow [11,12], which we review here. This is possible when there exists an equilibrium steady state  $x^{\text{eq}}$  that satisfies the detailed balance condition

$$\mathcal{K}_e(x^{\text{eq}}) = \mathcal{K}_{-e}(x^{\text{eq}}), \quad (18)$$

and the flux  $\mathcal{K}$  is given by the mass action law (3). Then the thermodynamic force is given by

$$F = -\mathbb{S}^\top \nabla D(x \| x^{\text{eq}}), \quad (19)$$

where  $\nabla = (\partial/\partial x_\alpha)$  is a column vector of the partial differential operators, and  $D(\cdot \| \cdot)$  is the generalized Kullback–Leibler (KL) divergence,

$$D(x \| x') := \sum_\alpha \left( x_\alpha \ln \frac{x_\alpha}{x'_\alpha} - x_\alpha + x'_\alpha \right), \quad (20)$$

which reduces to the usual KL divergence in the Markov jump case. Let us prove Eq. (19). The detailed balance condition leads to

$$\frac{k_e}{k_{-e}} = \frac{\prod_\alpha (x_\alpha^{\text{eq}})^{\nu_{\alpha'(e)}}}{\prod_\alpha (x_\alpha^{\text{eq}})^{\nu_{\alpha(e)}}}. \quad (21)$$

Substituting this into the definition of the thermodynamic force (8), we have

$$\begin{aligned} F_e &= \sum_\alpha (\nu_{\alpha(e)} - \nu_{\alpha'(e)}) \ln \frac{x_\alpha}{x_\alpha^{\text{eq}}} \\ &= -[\mathbb{S}^\top \Psi]_e \end{aligned} \quad (22)$$

with  $\Psi = \ln(x/x^{\text{eq}})$ , where we define the log of a vector as the vector of the logs of the elements. The proof ends by calculating the derivative of the KL divergence to show

$$\frac{\partial}{\partial x_\alpha} D(x \| x^{\text{eq}}) = \ln \frac{x_\alpha}{x_\alpha^{\text{eq}}} = \Psi_\alpha. \quad (23)$$

Because  $LF = J$ , by substituting Eq. (19) into Eq. (2), we obtain the gradient flow expression [11,42]

$$\frac{dx}{dt} = -\mathbb{S} L \mathbb{S}^\top \nabla D(x \| x^{\text{eq}}). \quad (24)$$

Note that the matrix  $\mathbb{S} L \mathbb{S}^\top$  is positive-semidefinite. When we consider a Markov jump process,  $\mathbb{S} L \mathbb{S}^\top$  reads  $BLB^\top$ , which is called the Laplacian matrix of a weighted graph [51].

A consequence of the above is that the KL divergence is a Lyapunov function when the system is autonomous, thus  $x^{\text{eq}}$

is time-invariant. The divergence decreases monotonically as

$$\begin{aligned} \frac{d}{dt} D(x \| x^{\text{eq}}) &= \left( \frac{dx}{dt} \right)^\top \nabla D(x \| x^{\text{eq}}) \\ &= -\|\mathbb{S}^\top \nabla D(x \| x^{\text{eq}})\|_L^2 \leq 0. \end{aligned} \quad (25)$$

Using Eqs. (14) and (19), we can confirm that the time derivative of the KL divergence is equal to the negative of the EPR.

For autonomous Markovian dynamics with a detailed balanced steady state, the gradient flow structure has also been utilized to measure the distance between two states. Maas proposed in Ref. [42] an  $L^2$ -Wasserstein distance measure

$$\mathcal{W}(p^{(0)}, p^{(1)}) := \inf_{p, \psi} \left( \int_0^1 dt \|B^\top \psi(t)\|_{L(p(t))}^2 \right)^{1/2}, \quad (26)$$

where the infimum is taken under the conditions

$$\frac{dp}{dt} = BL(p)B^\top \psi, \quad p(t=0) = p^{(0)}, \quad p(t=1) = p^{(1)}. \quad (27)$$

We can change the time interval from the unit one into an arbitrary one  $[0, \mathcal{T}]$  as

$$\mathcal{W}(p^{(0)}, p^{(1)}) = \inf_{p, \psi} \left( \mathcal{T} \int_0^\mathcal{T} dt \|B^\top \psi(t)\|_{L(p(t))}^2 \right)^{1/2}, \quad (28)$$

with the last condition of Eq. (27) replaced with  $p(\mathcal{T}) = p^{(1)}$  [52]. Note that the distance is a dynamics-wise measure because  $L$  depends not only on the current state but also on the given kinetics as in Eq. (13), or more explicitly as

$$\ell_e(p) = \frac{k_e p_{i(e)} - k_{-e} p_{i'(e)}}{\ln(k_e p_{i(e)}/k_{-e} p_{i'(e)})}.$$

Our definition looks unlike the original one [42], but the equivalence is easily proved (see Appendix A for the proof).

Vu and Hasegawa [31] found a speed limit by using the Wasserstein distance as

$$\tau \geq \frac{\mathcal{W}(p(0), p(\tau))^2}{\int_0^\tau dt \sigma}, \quad (29)$$

where  $\sigma$  is the EPR of the solution of a master equation  $\{p(t)\}_{t \in [0, \tau]}$ . The speed limit is obtained because the solution of a master equation satisfies the first condition in Eq. (27) as

$$\frac{dp}{dt} = BL(p)B^\top \Psi \quad (30)$$

with  $\Psi_i = -(\partial/\partial p_i)D(p \| p^{\text{eq}})$ . Since Eq. (28) involves minimization with respect to  $p$  and  $\psi$ , the Wasserstein distance is smaller than the corresponding expression evaluated for the original dynamics. By choosing the time interval of the solution  $\tau$  as the time interval in Eq. (28), we obtain the lower bound.

### III. ONSAGER-PROJECTIVE DECOMPOSITION OF EPR

We state our main result, a decomposition of the EPR in Markov jump processes and chemical reaction networks, which can be seen as the generalization of the Maes–Netočný decomposition to such systems [4]. Because we define the decomposition in a geometrical way by using a projection

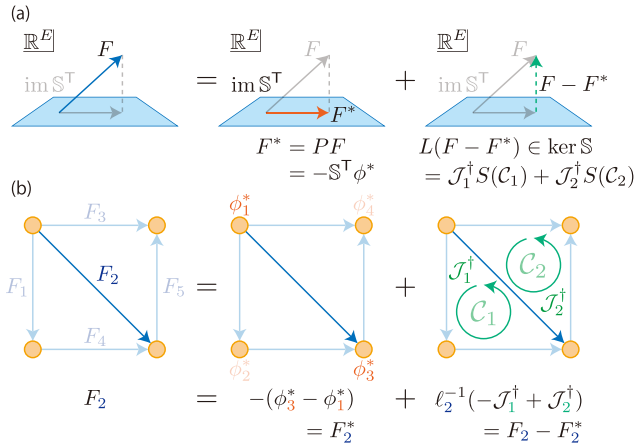


FIG. 4. Schematic diagrams of the decomposition of force, which is central to the decomposition of entropy production rate (34). (a) Geometrical illustration of decomposing force  $F$ . Projector  $P$  onto the space of conservative forces  $\text{im } \mathbb{S}$  is given in Eq. (31). The projected force  $F^*$  is orthogonal to the remainder  $F - F^*$  under the inner product  $\langle \cdot, \cdot \rangle_L$  introduced in Sec. II C. The remaining current  $J^{\text{cyc}} = L(F - F^*)$  belongs to  $\ker \mathbb{S}^\top$ , therefore it is cyclic and can be expanded in a cyclic basis as in Eq. (41). (b) The decomposition of force is exemplified on edge  $e = 2$  in the system considered in Fig. 2. The conservative force  $F^*$  is given by the difference of  $\phi^*$  which is assigned a value on each node. On the other hand, the remaining force  $F - F^*$  is associated with the coefficients of the linear combination of cycles.

operator and the edgewise Onsager coefficient (13), we term it *the Onsager-projective decomposition*. The basic idea of the technical details discussed in the following two paragraphs is summarized in Fig. 4.

Let us consider a matrix  $P$  that is the projection onto the image of  $\mathbb{S}^\top$  and orthogonal with respect to the matrix  $L$ . Formally, this means that for any  $f, g \in \mathbb{R}^E$ ,  $P$  satisfies  $\langle Pf, (I - P)g \rangle_L = 0$ , where  $I$  is the identity. This projection matrix is given by [53]

$$P = \mathbb{S}^\top (\mathbb{S}L\mathbb{S}^\top)^- \mathbb{S}L, \quad (31)$$

with  $A^-$  denoting the generalized inverse of a matrix  $A$ . Since  $\mathbb{S}$  does not have full rank in general, we need to consider a generalized inverse. Nevertheless, we may set it to a theoretically tractable one, like the Moore–Penrose inverse, because  $P$  does not depend on the choice of the generalized inverse [53]. To be more concrete, let the eigendecomposition of  $\mathbb{S}L\mathbb{S}^\top$  be  $U\Lambda U^{-1}$  with orthogonal matrix  $U$  and diagonal matrix  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$ , where  $\lambda_\alpha \neq 0$  and  $r = \text{rank } \mathbb{S}$ . Then, the Moore–Penrose inverse of  $\mathbb{S}L\mathbb{S}^\top$  can be computed as  $U\Lambda^+U^{-1}$  with  $\Lambda^+ = \text{diag}(\lambda_1^{-1}, \dots, \lambda_r^{-1}, 0, \dots, 0)$ , which is the Moore–Penrose inverse of  $\Lambda$ .

For any  $f \in \mathbb{R}^E$ ,  $Pf$  is also characterized by

$$Pf = \arg \min_{g \in \text{im } \mathbb{S}^\top} \|f - g\|_L^2. \quad (32)$$

Because  $Pf \in \text{im } \mathbb{S}^\top$ , there exists  $\phi(f) \in \mathbb{R}^S$ , which satisfies  $-\mathbb{S}^\top \phi(f) = Pf$ . The solution  $\phi(f)$  is not unique in general because the transformation  $\phi(f) \rightarrow \phi(f) + c$ , where  $c$  is a left null vector of  $\mathbb{S}$ , does not change  $-\mathbb{S}^\top \phi(f)$ . If  $\mathbb{S}$  is the

incidence matrix  $B$  of a connected graph,  $c$  is given by a vector of 1's times any scalar. When we consider chemical reaction networks, such a null vector is called a conservation law [5]. Nonetheless,  $Pf$  is unique, so we can consider  $\phi(f)$  as fixed.

Now, we can define the conservative force that plays a central role in the Onsager-projective decomposition. Define  $F^* := PF$  and  $\phi^* = \phi(F)$  such that  $F^* = -\mathbb{S}^\top \phi^*$ . The force  $F^*$  can be seen as a conservative force that reproduces the original (possibly nonconservative) dynamics because  $\mathbb{S}LF^* = \mathbb{S}LF = \mathbb{S}J = d_t x$ . This can be shown as follows. First, we have a decomposition

$$\begin{aligned} \mathbb{S}LF &= \mathbb{S}L(P + I - P)F \\ &= \mathbb{S}LF^* + \mathbb{S}L(I - P)F. \end{aligned} \quad (33)$$

The second term  $\mathbb{S}L(I - P)$  is zero because  $P$  is the orthogonal projection onto  $\text{im } \mathbb{S}^\top$  with respect to the metric  $L$ , i.e., for any  $\varphi \in \mathbb{R}^S$  and  $f \in \mathbb{R}^E$ ,  $P$  satisfies  $\langle \mathbb{S}^\top \varphi, (I - P)f \rangle_L = \varphi^\top \mathbb{S}L(I - P)f = 0$ . Because  $\varphi$  is now arbitrary, by setting  $f = F$ , we obtain  $\mathbb{S}L(I - P)F = 0$ . Thus  $d_t x = \mathbb{S}LF^*$  holds.

Using the conservative force  $F^*$ , we find a decomposition of the EPR that can be regarded as a generalization of the Maes–Netočný decomposition in Langevin systems to Markov jump processes and chemical reaction networks [4]. We define the *Onsager excess* and *housekeeping EPR* by

$$\sigma^{\text{ex}} := \|F^*\|_L^2, \quad \sigma^{\text{hk}} := \|F - F^*\|_L^2, \quad (34)$$

respectively. These terms give a non-negative decomposition of the EPR,

$$\sigma = \sigma^{\text{ex}} + \sigma^{\text{hk}}, \quad (35)$$

$$\sigma^{\text{ex}}, \sigma^{\text{hk}} \geq 0 \quad (36)$$

for both Markov jump processes and CRNs. They are always non-negative because  $L$  is positive-definite. From the orthogonality relation

$$\langle F^*, F - F^* \rangle_L = \langle PF, (I - P)F \rangle_L = 0, \quad (37)$$

the decomposition is proved as follows:

$$\begin{aligned} \sigma &= \langle F^* + (F - F^*), F^* + (F - F^*) \rangle_L \\ &= \langle F^*, F^* \rangle_L + \langle F - F^*, F - F^* \rangle_L. \end{aligned} \quad (38)$$

The two orthogonal components  $F^*$  and  $F - F^*$  of the thermodynamic force allow us to interpret the decomposition in a geometrical way; see Figs. 1 and 4. This is similar to the interpretation in the Langevin case [28,29].

In a steady state, where  $\mathbb{S}J^{\text{st}} = 0$  holds, the excess EPR vanishes and  $\sigma = \sigma^{\text{hk}}$  because

$$\sigma^{\text{ex}} = \langle F^*, F^* \rangle_{L^{\text{st}}} = -(\phi^*)^\top \mathbb{S}L^{\text{st}} F^* = -(\phi^*)^\top \mathbb{S}J^{\text{st}} = 0. \quad (39)$$

On the other hand, the housekeeping EPR is identically zero if and only if the steady state is detailed balanced [54]. Therefore, the decomposition (35) separates the EPR into a transient contribution  $\sigma^{\text{ex}}$  and a contribution  $\sigma^{\text{hk}}$  required to sustain a nonequilibrium steady state if a steady state exists.

Given the eigendecomposition  $\mathbb{S}L\mathbb{S}^\top = U\Lambda U^{-1}$  discussed earlier, we can write down the excess and housekeeping

TABLE I. Summary of consequences of the Onsager-projective decomposition discussed in Sec. III.

Generalization of Schnakenberg decomposition	Eq. (43)
Generalization of MN decomposition	Eq. (45)
Variational formulas	Eqs. (48)–(50)
Operational implications	Eq. (57)
Gradient flow	Eq. (59)

EPR as

$$\begin{aligned}\sigma^{\text{ex}} &= \sum_{\alpha=1}^r \lambda_{\alpha}^{-1} |[U^{-1}\mathbb{S}LF]_{\alpha}|^2, \\ \sigma^{\text{hk}} &= \sigma - \sum_{\alpha=1}^r \lambda_{\alpha}^{-1} |[U^{-1}\mathbb{S}LF]_{\alpha}|^2.\end{aligned}\quad (40)$$

In what follows in this section, we discuss several properties and consequences of the decomposition, which we summarize in Table I with corresponding equation numbers. Except for the operational implications discussed in Sec. III D using the variational formulas from Sec. III C, these results are basically independent from each other.

### A. Generalization of Schnakenberg's cyclic decomposition

Our decomposition leads to a generalization of Schnakenberg's cyclic decomposition of the steady-state EPR (12) to nonsteady states. Consider the current corresponding to  $F - F^*$ , that is,  $J^{\text{cyc}} := L(F - F^*)$ . This current is in the kernel of  $\mathbb{S}$  because  $\mathbb{S}LF = \mathbb{S}LF^*$ , therefore it is a cyclic current and can be expanded in the cycle basis as

$$J_e^{\text{cyc}} = \sum_{\mu} S_e(C_{\mu}) \mathcal{J}_{\mu}^{\dagger}. \quad (41)$$

This expansion is available whether the system is in a steady state or not. The conjugated force is also given by

$$\mathcal{F}_{\mu}^{\dagger} := \sum_e S_e(C_{\mu})(F_e - F_e^*) = \sum_e S_e(C_{\mu})F_e, \quad (42)$$

where we used  $\sum_e S_e(C_{\mu})F_e^* = -(\phi^*)^{\top} \mathbb{S} \mathbb{S}(C_{\mu}) = 0$ . Thus, we have generalized the decomposition in Eq. (11), which applies only in a steady state, to arbitrary states. Importantly, these cycle quantities allow us to write the Onsager housekeeping EPR as

$$\sigma^{\text{hk}} = \sum_{\mu} \mathcal{J}_{\mu}^{\dagger} \mathcal{F}_{\mu}^{\dagger} \quad (43)$$

because

$$\begin{aligned}\sum_{\mu} \mathcal{J}_{\mu}^{\dagger} \mathcal{F}_{\mu}^{\dagger} &= \sum_{\mu} \mathcal{J}_{\mu}^{\dagger} \sum_e S_e(C_{\mu})(F_e - F_e^*) \\ &= \sum_e J_e^{\text{cyc}}(F_e - F_e^*) = \|F - F^*\|_L^2.\end{aligned}\quad (44)$$

Therefore, our decomposition generalizes Schnakenberg's cyclic decomposing of the steady-state EPR [10] to a cyclic decomposition of the housekeeping EPR in non-steady-state systems.

### B. Generalization of Maes–Netočný decomposition

Next, we show that the decomposition can also be understood as a generalization of the MN decomposition [4]. In their paper, Maes and Netočný regarded the EPR as a functional of the potential landscape, and they minimized it by changing the landscape. The result is the following minimum EP principle: the potential landscape minimizing the EPR is the one for which the instantaneous distribution is a steady state. In our case, using the minimizing property of the projection operator (32), we have the expression

$$\sigma^{\text{hk}} = \inf_{\psi} \|F - (-\mathbb{S}^{\top}\psi)\|_L^2. \quad (45)$$

In a Markov jump process,  $F$  is written as  $F = -B^{\top}(u + \ln p) + F^{\text{nc}}$ , where  $u$  is some potential landscape and  $F^{\text{nc}}$  is a nonconservative driving force. We define  $J(v) := -LB^{\top}(v + \ln p) + LF^{\text{nc}}$  as a function of the potential landscape. Using these expressions, we find a formulation resembling Ref. [4],

$$\sigma^{\text{hk}} = \inf_v J(v)^{\top} L^{-1} J(v). \quad (46)$$

Minimization leads to the condition

$$BJ(v) = 0. \quad (47)$$

Thus, among all potential landscapes with Onsager matrix  $L$ , the entropy is minimized for the one in which the instantaneous distribution  $p$  is a steady state. This is the analog of Maes and Netočný's minimum EP principle [4]. We remark that fixing  $L$  is connected to fixing the diffusion matrix in the Langevin case (see Appendix B).

### C. Variational formulas

We can also extend other results derived in the Langevin case [28,29] to discrete dynamics. First, the following variational expressions hold:

$$\sigma^{\text{ex}} = \sup_{\psi} \frac{(\langle \mathbb{S}^{\top}\psi, F \rangle_L)^2}{\langle \mathbb{S}^{\top}\psi, \mathbb{S}^{\top}\psi \rangle_L}, \quad (48)$$

$$= \inf_{F' \in \ker \mathbb{S}L} \langle F - F', F - F' \rangle_L, \quad (49)$$

$$\sigma^{\text{hk}} = \sup_{F' \in \ker \mathbb{S}L} \frac{(\langle F', F \rangle_L)^2}{\langle F', F' \rangle_L}. \quad (50)$$

Note that  $\mathbb{S}L$  is the adjoint of  $\mathbb{S}^{\top}$  with respect to the inner product  $\langle \cdot, \cdot \rangle_L$ . As an important example,  $F - F^*$  is in its kernel because  $\mathbb{S}LF = \mathbb{S}LF^*$ . This space can be interpreted as the space of forces that do not contribute to the dynamics if we regard  $LF'$  as a probability current for any force  $F'$ .

The first expression (48) is obtained by combining the Cauchy–Schwarz inequality

$$(\langle \mathbb{S}^{\top}\psi, F^* \rangle_L)^2 \leq \langle \mathbb{S}^{\top}\psi, \mathbb{S}^{\top}\psi \rangle_L \langle F^*, F^* \rangle_L, \quad (51)$$

and the following equality:

$$\langle \mathbb{S}^{\top}\psi, F^* \rangle_L = \psi^{\top} \mathbb{S}LF^* = \psi^{\top} \mathbb{S}LF = \langle \mathbb{S}^{\top}\psi, F \rangle_L. \quad (52)$$

Equality is obtained in (51) when  $\mathbb{S}^{\top}\psi$  is parallel to  $F^*$ . The infimum representation (49) for the excess EPR is proven by

the following calculation: When  $F' \in \ker \mathbb{S}L$ ,

$$\begin{aligned} & \langle F - F', F - F' \rangle_L \\ &= \langle F^* + F - F^* - F', F^* + F - F^* - F' \rangle_L \\ &= \langle F^*, F^* \rangle_L + \langle F - F^* - F', F - F^* - F' \rangle_L, \end{aligned} \quad (53)$$

because

$$\langle F^*, F - F^* - F' \rangle_L = (\phi^*)^\top \mathbb{S}L(F - F^* - F') = 0. \quad (54)$$

The second term in the third line of Eq. (53) is non-negative and vanishes when  $F' = F - F^* \in \ker \mathbb{S}L$ ; thus the minimum value gives the excess EPR.

The supremum formula (50) of the housekeeping EPR can be shown in a parallel way to the excess case. For  $F' \in \ker \mathbb{S}L$ , we have  $(\langle F', F - F^* \rangle_L)^2 \leq \langle F', F' \rangle_L \langle F - F^*, F - F^* \rangle_L$  and  $\langle F', F - F^* \rangle_L = \langle F', F \rangle_L$ . The latter equality is because

$$\langle F', F^* \rangle_L = -(\phi^*)^\top \mathbb{S}L F' = 0. \quad (55)$$

Thus, the inequality

$$\sigma^{\text{hk}} \geq \frac{(\langle F', F \rangle_L)^2}{\langle F', F' \rangle_L} \quad (56)$$

is obtained for any  $F' \in \ker \mathbb{S}L$ . The equality is attained when  $F'$  is proportional to  $F - F^*$ , which also belongs to  $\ker \mathbb{S}L$ .

#### D. Minimum entropy production

The variational formula (49) allows us to interpret the excess EPR as a minimum EPR. In this section, we focus on Markov jump processes. In general, when we control rate constants to minimize the EPR with the dynamics  $d_t p$  unchanged, the EPR can be made arbitrarily small unless we make further assumptions [55]. To consider minimum EPR under additional constraints, recent studies have proposed to fix the geometrical mean of the transition rates  $\sqrt{k_e k_{-e}}$  on each edge [56], or to fix the backward rates  $k_{-e}$  [57]. They found that such constraints lead to nonconservative optimal forces. On the other hand, the excess EPR can be interpreted as the minimum EPR when we change transition rates while keeping the value of  $L(p)$  fixed, since Eq. (49) can be rewritten as

$$\begin{aligned} \sigma^{\text{ex}} &= \inf_{F' | d_t p = BL F'} \langle F', F' \rangle_L \\ &= \inf_{k' | d_t p = BL F_{k'}, L_{k'} = L} \langle F_{k'}, F_{k'} \rangle_L \\ &= \inf_{k' | d_t p = BL F_{k'}, L_{k'} = L} \sigma_{k'}, \end{aligned} \quad (57)$$

where  $F_{k'}$ ,  $L_{k'}$ , and  $\sigma_{k'}$  are the force, the Onsager matrix, and the EPR given by transition rates  $k'$  instead of the actual dynamics  $k$ , which provides  $L$ . The definition of the excess EPR (34) also shows that we can obtain a conservative optimal force in this minimization. Quite recently, a similar study [58] discussed minimization of integrated EP by fixing the time average of the sum of the edgewise coefficients, not the instantaneous coefficients. We summarize these minimizations in Table II.

Although the minimization is mathematically well defined, fixing the Onsager coefficients is not always interpretable

TABLE II. Comparison of dynamical constraints and optimal thermodynamic forces that minimize the EPR (or integrated EP). Here, NC and C stand for nonconservative and conservative. Vu [58] did not mention whether the optimal force is conservative or not.

	Ours	Remlein [56]	Ilker [57]	Vu [58]
Constraint	$\ell_e(p)$	$\sqrt{k_e k_{-e}}$	$k_{-e}$	$\tau^{-1} \int_0^\tau \ell_e(p) dt$
Force	C	NC	NC	?

from a physical point of view. Nonetheless, we can find a concrete meaning of this constraint in the continuum limit. In this limit, the edgewise Onsager coefficient becomes  $\ell_e(p) \approx D_e(t(e)) p_{i(e)} / (\Delta x)^2$  with  $\Delta x$  being the distance between nearest sites going to zero, and  $D_e(i)$  a kind of ‘‘diffusion coefficient’’.<sup>1</sup> Therefore, fixing  $L(p)$  is equivalent to assuming that the diffusion coefficients are constant, which is a natural assumption when we consider Langevin dynamics. For more detailed discussion on the asymptotic relation, see Appendix B.

#### E. Gradient flow

We now generalize Eqs. (24) and (25) using the Onsager-projective decomposition. Let  $\phi^*(t)$  be the potential that gives  $F^*(t) = -\mathbb{S}^\top \phi^*(t)$  at time  $t$ , and define a pseudocanonical distribution  $x^{\text{pcan}}(t)$  by

$$x_\alpha^{\text{pcan}}(t) = \frac{1}{Z_\alpha(t)} x_\alpha(t) e^{-\phi_\alpha^*(t)}. \quad (58)$$

In a Markov jump process,  $Z_\alpha(t) = \sum_i p_i(t) e^{-\phi_i^*(t)}$  is a normalization constant that does not depend on  $\alpha$ . In a chemical reaction network,  $Z(t)$  is any vector whose log becomes a conservation law, i.e.,  $\ln Z(t) = [\ln Z_\alpha(t)] \in \ker \mathbb{S}^\top$  holds. If there is no conservation law,  $Z(t)$  will be the vector of 1’s. Then, we have the relation

$$\sigma^{\text{ex}} = -\left. \frac{d}{dt} D(x(t) \| x^{\text{pcan}}(s)) \right|_{s=t}. \quad (59)$$

This result follows because

$$\begin{aligned} \left. \frac{d}{dt} D(x(t) \| x^{\text{pcan}}(s)) \right|_{s=t} &= \sum_\alpha \frac{dx_\alpha}{dt} (\phi_\alpha^* + \ln Z_\alpha) \\ &= J^\top \mathbb{S}^\top \phi^* = -\langle F, F^* \rangle_L = -\langle F^*, F^* \rangle_L, \end{aligned} \quad (60)$$

where the second equality follows from  $\sum_i B_{ie} = 0$  for Markov jump processes and from  $\mathbb{S}^\top \ln Z = 0$  for chemical reaction networks. In particular, in a Markov jump process, by the decomposition  $\phi^* = \epsilon^* + \ln p$ , the pseudocanonical distribution reads

$$p_i^{\text{pcan}} = \frac{1}{Z} e^{-\epsilon_i^*}. \quad (61)$$

If  $F^* = F$ , the pseudoenergy level  $\epsilon^*$  becomes the actual energy level of each state. The expression (61) is the reason why we refer to  $x^{\text{pcan}}$  as the pseudocanonical distribution.

<sup>1</sup>This relation was proved for one-dimensional systems in [59]. In Appendix B, we generalized it to multidimensional cases.



Although this pseudocanonical distribution depends on the instantaneous state, it indicates the momentary “goal” of the dynamics, even if the system is attracted to a limit cycle.

Along with the pseudocanonical distribution  $x^{\text{pcan}}$ , we can formally obtain a gradient flow representation of the continuity equation (2) even without detailed balance because

$$\begin{aligned} \mathbb{S}^\top \nabla D(x \| x^{\text{pcan}}) &= \mathbb{S}^\top \ln \frac{x}{x^{\text{pcan}}} \\ &= \mathbb{S}^\top (\phi^* + \ln Z) \\ &= -F^*, \end{aligned} \quad (62)$$

so that

$$-SLS^\top \nabla D(x \| x^{\text{pcan}}) = SLF^* = \frac{dx}{dt}. \quad (63)$$

This is a generalization of the gradient flow (24). We remark that, even if the rates entering the dynamics via (3) are time-independent,  $x^{\text{pcan}}$  generally depends on time; see Eq. (58).

#### F. Brief comparison with Hatano–Sasa decomposition

There are other ways to decompose the EPR into excess and housekeeping contributions. One well-known approach is the Hatano–Sasa (HS) decomposition [2], also called the adiabatic–nonadiabatic decomposition [3], which is given by

$$\sigma = \sigma_{\text{HS}}^{\text{ex}} + \sigma_{\text{HS}}^{\text{hk}}, \quad (64)$$

$$\sigma_{\text{HS}}^{\text{ex}} = J^\top (F - F^{\text{st}}), \quad (65)$$

$$\sigma_{\text{HS}}^{\text{hk}} = J^\top F^{\text{st}}. \quad (66)$$

If the flux is written in mass action form, the difference between the forces becomes

$$\begin{aligned} F_e - F_e^{\text{st}} &= \sum_{\alpha} (v_{\alpha l(e)} - v_{\alpha l'(e)}) \ln \frac{x_{\alpha}}{x_{\alpha}^{\text{st}}} \\ &= -[S^\top \Psi]_e \end{aligned} \quad (67)$$

with  $\Psi_{\alpha} = \ln(x_{\alpha}/x_{\alpha}^{\text{st}})$ . Then, we obtain another expression for the HS excess EPR,

$$\sigma_{\text{HS}}^{\text{ex}} = -J^\top S^\top \Psi, \quad (68)$$

$$\Psi = \ln \frac{x}{x^{\text{st}}}, \quad (69)$$

which is similar to the expression  $\sigma^{\text{ex}} = -J^\top S^\top \psi^*$  that follows from the definition (34) and  $-SLS^\top \phi^* = SJ$ . The non-negativity of the HS excess and housekeeping EPR can be proven only when the steady state satisfies  $BJ^{\text{st}} = 0$ . For a Markov-jump process, this condition is precisely the steady-state condition. For a chemical reaction network, however, this condition is called the *complex balance* condition and is stricter than the steady-state condition  $SJ^{\text{st}} = \nu BJ^{\text{st}} = 0$  [5,6]. On the other hand, the Onsager excess EPR is defined even in the absence of a stable steady state, which we will show in Sec. VIA through the Brusselator model that exhibits a limit cycle. For a more detailed discussion about the relation between our approach and the HS decomposition, see Appendix C.

TABLE III. Summary of TURs we derive in Sec. IV. Unless noted, the result is obtained only for Markov jump processes. Here, CRN stands for chemical reaction network.

	Short time	Finite time
Excess	Eq. (76) For CRN, Eq. (81)	Eq. (92)
Housekeeping	Eq. (88)	Eq. (87)

## IV. THERMODYNAMIC UNCERTAINTY RELATIONS

### A. Short-time TUR: Markov jump process

The variational expression (48) allows us to obtain a set of refined short-time TURs (Table III). In this section, we focus on Markov jump processes. We define the edgewise dynamical activity (traffic)  $\chi_e$  by [60]

$$\chi_e := k_e p_{l(e)} + k_{-e} p_{l'(e)}. \quad (70)$$

From the inequality in Eq. (15), we have

$$\ell_e = \frac{J_e}{F_e} \leq \frac{1}{2} \chi_e. \quad (71)$$

The equality holds if and only if  $p$  satisfies  $k_e p_{l(e)} = k_{-e} p_{l'(e)}$ , i.e., in the equilibrium state.

The numerator of the right-hand side in Eq. (48),  $\langle S^\top \psi, F \rangle_L$ , which reads  $\langle B^\top \psi, F \rangle_L$  now, is the time derivative of the expectation value  $\langle \psi \rangle = \sum_i p_i \psi_i$  because

$$\frac{d}{dt} \langle \psi \rangle = \psi^\top BJ = \psi^\top BLF = \langle B^\top \psi, F \rangle_L. \quad (72)$$

Note that here we assume that  $\psi$  does not have any time dependence. Thus, from the variational formula (48), we have a TUR

$$\frac{(d_t \langle \psi \rangle)^2}{\langle B^\top \psi, B^\top \psi \rangle_L} \leq \sigma^{\text{ex}} \quad (73)$$

for any observable  $\psi$ . Therefore, as long as we know the edgewise Onsager matrix  $L$ , we can estimate the excess EPR by measuring the rate of change of the average  $\langle \psi \rangle$ .

Because of the inequality in Eq. (71), the denominator is bounded from above by the short-time limit of the variance as

$$\langle B^\top \psi, B^\top \psi \rangle_L \leq \lim_{\tau \rightarrow 0} \frac{\text{Var}(\Delta_\tau \psi)}{2\tau} =: \frac{1}{2} D_\psi, \quad (74)$$

where  $\Delta_\tau \psi = \psi_{i(t+\tau)} - \psi_{i(t)}$  and the variance is asymptotically given by [23]

$$\text{Var}(\Delta_\tau \psi) = \tau \sum_e \chi_e ([B^\top \psi]_e)^2 + o(\tau). \quad (75)$$

This inequality shows a weaker bound,

$$\frac{2(d_t \langle \psi \rangle)^2}{D_\psi} \leq \sigma^{\text{ex}}. \quad (76)$$

Although equality can only be achieved in equilibrium, where we trivially have  $\sigma^{\text{ex}} = 0$ , the inequality can be tight near equilibrium, where the difference between the arithmetic and logarithmic mean in Eq. (71) is small. Moreover, Eq. (76) does not require information about the Onsager matrix  $L$ . Therefore, it is much more tractable than Eq. (73) and gives

an estimate of the excess EPR that depends only on the rate of change of the average and the short-time fluctuations of some observable. This inequality partly generalizes the short-time TUR derived in Ref. [23] for the case in which currents are expressed as changes of observables as above.

While the interpretation of the inequalities derived from Eq. (48) as TURs was straightforward, we need to develop a trajectory level description to understand the meaning of the inequality suggested by Eq. (50). This is done later when we discuss finite-time TURs.

### B. Short-time TUR: Chemical reaction network

Next, we consider chemical reaction networks. Because of the inequality between logarithmic mean and arithmetic mean (15), we have

$$\begin{aligned} \langle S^\top \psi, S^\top \psi \rangle_L &= \sum_e ([S^\top \psi]_e)^2 \frac{J_e}{F_e} \\ &\leq \frac{1}{2} \sum_e ([S^\top \psi]_e)^2 [\mathcal{K}_e(c) + \mathcal{K}_{-e}(c)]. \end{aligned} \quad (77)$$

The scaled diffusion coefficient of a chemical reaction network is defined by [6,30]

$$\tilde{D}_{\alpha\beta} := \frac{1}{2} \sum_e S_{\alpha e} S_{\beta e} [\mathcal{K}_e(c) + \mathcal{K}_{-e}(c)]. \quad (78)$$

Using this quantity, the right-hand side of Eq. (77) becomes

$$\frac{1}{2} \sum_e ([S^\top \psi]_e)^2 [\mathcal{K}_e(c) + \mathcal{K}_{-e}(c)] = \psi^\top \tilde{D} \psi. \quad (79)$$

Moreover, because

$$\langle S^\top \psi, F \rangle_L = \psi^\top S J = \frac{d}{dt} (\psi^\top c), \quad (80)$$

we obtain a TUR for a chemical reaction network

$$\frac{1}{\psi^\top \tilde{D} \psi} \left( \frac{d}{dt} (\psi^\top c) \right)^2 \leq \sigma^{\text{ex}}. \quad (81)$$

This inequality generalizes the previous result in Ref. [30]: because  $\sigma^{\text{ex}} \leq \sigma$ ,

$$\max_\alpha \frac{1}{\tilde{D}_{\alpha\alpha}} \left( \frac{dc_\alpha}{dt} \right)^2 \leq \sigma \quad (82)$$

follows by setting  $\psi_{\alpha'} = \delta_{\alpha'\alpha}$ .

### C. Finite-time TURs

The above results relate the excess EPR to the short-time fluctuations of observables. In the following, we obtain finite-time TURs for Markov jump process. Proofs of these relations are found in Appendix D. Note that these proofs rely on the expression for the path probability of the system, so their generalization to chemical reaction networks is challenging, and we leave it for future work.

In general, we consider a stochastic current observable  $\mathcal{J}_w$  with weight  $w_e(t)$ ,

$$\mathcal{J}_w = \sum_k w_{e_k}(t_k) \quad (83)$$

for a stochastic trajectory with jumps  $e_k$  at time  $t_k$  ( $k = 1, 2, \dots$ ). We assume the weight satisfies  $w_e = -w_{-e}$ . The average during a time interval  $[0, \tau]$  is given by

$$\langle \mathcal{J}_w \rangle_\tau = \int_0^\tau dt \sum_e w_e(t) J_e(t). \quad (84)$$

We introduce a decomposition of the current observable into the excess and the housekeeping contribution [29] as

$$\langle \mathcal{J}_w^{\text{ex}} \rangle_\tau := \int_0^\tau dt \sum_e w_e(t) J_e^*(t), \quad (85)$$

$$\langle \mathcal{J}_w^{\text{hk}} \rangle_\tau := \int_0^\tau dt \sum_e w_e(t) [J_e(t) - J_e^*(t)], \quad (86)$$

where we defined  $J_e^*(t) := \ell_e(t) F_e^*(t)$ . As shown above, these currents induce the same time evolution as  $J$ , since  $B J^* = B L F^* = B J$ .

By using the housekeeping component of a current observable, we can show the TUR for the Onsager housekeeping EP,

$$\frac{(\langle \mathcal{J}_w^{\text{hk}} \rangle_\tau)^2}{\text{Var}(\mathcal{J}_w)} \leq \frac{1}{2} \int_0^\tau dt \sigma^{\text{hk}}, \quad (87)$$

with variance  $\text{Var}(\mathcal{J}_w) := \langle (\mathcal{J}_w - \langle \mathcal{J}_w \rangle_\tau)^2 \rangle_\tau$ . Unlike the short-time TURs discussed above, this relation is valid for any finite time duration. It states that the conventional form of the steady-state TUR remains valid if we consider the respective housekeeping contributions for both the current and the entropy production.

Now, we can derive a short-time TUR for the housekeeping EPR. Since the time interval  $[0, \tau]$  can be replaced with  $[t, t + \Delta t]$ , by taking the limit  $\Delta t \rightarrow 0$ , we obtain the short-time TUR

$$\frac{2(d_t \langle \mathcal{J}_w^{\text{hk}} \rangle_t)^2}{\mathcal{D}_w} \leq \sigma^{\text{hk}} \quad (88)$$

with

$$\mathcal{D}_w = \sum_e \chi_e w_e^2 = \lim_{\Delta t \rightarrow 0} \frac{\text{Var}(\mathcal{J}_w)}{\Delta t} \quad (89)$$

because

$$\begin{aligned} d_t \langle \mathcal{J}_w^{\text{hk}} \rangle_t &= \sum_e w_e(t) [J_e(t) - J_e^*(t)] \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} dt' \sum_e w_e(t') [J_e(t') - J_e^*(t')]. \end{aligned} \quad (90)$$

When we assume the weight satisfies  $w = (w_e)_{e=1, \dots, E} \in \ker SL$ , we see that  $w^\top J^* = w^\top L F^* = 0$  because  $F^* \in \text{im } S^\top$ . Thus, we find another TUR,

$$\frac{2(d_t \langle \mathcal{J}_w \rangle_t)^2}{\mathcal{D}_w} \leq \sigma^{\text{hk}} \quad (91)$$

for  $w \in \ker SL$ . Because  $d_t \langle \mathcal{J}_w \rangle_t = w^\top J = \langle w, F \rangle_L$ , this is derived from Eq. (50) in the same way as we did in Sec. IV A. What is important here is that, by introducing a current observable, we can interpret the inequality we get as a TUR.

If we change the time interval as in Refs. [29,61], we can also derive a finite-time TUR for the Onsager excess

EP. We assume that the transition rates  $k_e(t)$  depend on  $t$  and  $\tau$  in the form of  $t/\tau$  only. For example, with  $k_0$  constant,  $k_e(t) = k_0 \exp(t/\tau)$  is acceptable but  $k_e(t) = k_0 \exp(t)$  and  $k_e(t) = \tau^{-1} \exp(t/\tau)$  are not. Physically, this assumption means that any external operation on the system is accelerated when  $\tau$  becomes smaller, and vice versa. We put the same assumption on the time dependence of the weight  $w$ . Under these assumptions, we can prove the finite-time TUR

$$\frac{(\tau \partial_\tau \langle \mathcal{J}_w \rangle_\tau - \langle \mathcal{J}_w^{\text{hk}} \rangle_\tau)^2}{\text{Var}(\mathcal{J}_w)} \leq \frac{1}{2} \int_0^\tau dt \sigma^{\text{ex}}, \quad (92)$$

where  $\partial_\tau$  is taken while holding fixed the normalized time  $t/\tau$  that defines  $k_e$  (and enters into the definition of  $\langle \mathcal{J}_w \rangle_\tau$ , Eq. (84), via  $J_e$ ).

When the weight is given by a scalar time-dependent observable  $\psi(t/\tau)$  as  $w = B^\top \psi$ ,  $\mathcal{J}_w$  becomes the change  $\Delta_\tau \psi = \psi(1) - \psi(0)$  and we have the relations

$$\begin{aligned} \langle \mathcal{J}_w \rangle_\tau &= \int_0^\tau dt \psi^\top B J = \int_0^\tau dt \frac{d}{dt} \langle \psi \rangle \\ &= \langle \psi(1) \rangle_{t=\tau} - \langle \psi(0) \rangle_{t=0} \end{aligned} \quad (93)$$

$$\langle \mathcal{J}_w^{\text{hk}} \rangle_\tau = \int_0^\tau dt \psi^\top B (J - J^*) = 0. \quad (94)$$

Then, Eq. (92) reads

$$\frac{[\tau d_\tau \langle \psi(1) \rangle_{t=\tau}]^2}{\text{Var}(\Delta_\tau \psi)} \leq \frac{1}{2} \int_0^\tau dt \sigma^{\text{ex}}. \quad (95)$$

This can be understood as a refinement of the TUR for time-dependent driving derived in Ref. [61]. In particular, when one considers the change of scalar observables, rather than currents, the resulting TUR can only estimate the excess part of the entropy production. By taking the limit  $\tau \rightarrow 0$ , we recover the short-time TUR (76).

Since the proofs of the finite-time TURs are lengthy and technically similar to existing results, we give them in Appendix D, only explaining their underlying idea briefly at this point. The structures of the proofs are essentially the same. The first step is to parametrize the rate constant and compare the two path probabilities for two close parameters. The difference between the two path probabilities is measured by a KL divergence that bounds from above the ratio of two quantities: (i) the difference between the averages of the current observable given by the different parameters, and (ii) the variance of the current observable. This inequality is obtained from the fluctuation-response inequality [26]. The difference between the observables can be shown to be equal to  $\langle \mathcal{J}_w^{\text{hk}} \rangle$  or  $\tau \partial_\tau \langle \mathcal{J}_w \rangle_\tau - \langle \mathcal{J}_w^{\text{hk}} \rangle$  depending on the way the rates are parametrized. At the same time, the KL divergence bounds from below the Onsager excess or housekeeping EP. Combining the two inequalities, we obtain the uncertainty relations.

## V. WASSERSTEIN GEOMETRY

### A. Generalization of the Wasserstein distance

In our final set of main theoretical results, we provide a generalization of the Wasserstein distance measure (26) and the speed limit (29) (see Table IV). In this section, we assume that the kinetics  $\mathcal{K}$  do not depend on time explicitly.

TABLE IV. Summary of the main results we obtain in Sec. V. Here ‘‘distance’’ means Wasserstein distance. The first three are mathematically novel, while the last two reveal physical relations between the Wasserstein geometry and the excess EP(R).

Generalization of distance	Eq. (100)
Constant speed property	Eq. (104)
Duality formula	Eq. (105)
Connection between distance and excess EPR	Eq. (110)
Speed limit with excess EP	Eq. (116)

As we reviewed in Sec. II C, the Wasserstein distance has been defined for Markov jump processes with detailed balanced steady states. Now, we find a general definition applicable to both Markov jump processes and chemical reaction networks:

$$\mathcal{W}(x^{(0)}, x^{(1)}) := \inf_{x,f} \left( \int_0^1 dt \|f(t)\|_{L(x(t))}^2 \right)^{1/2} \quad (96)$$

under the following conditions:

$$\frac{dx}{dt} = \mathbb{S}L(x)f, \quad x(t=0) = x^{(0)}, \quad x(t=1) = x^{(1)}. \quad (97)$$

As in Eq. (28), the unit time interval can be replaced with an arbitrary one  $[0, \mathcal{T}]$ , and we have

$$\mathcal{W}(x^{(0)}, x^{(1)}) = \inf_{x,f} \left( \mathcal{T} \int_0^\mathcal{T} dt \|f(t)\|_{L(x(t))}^2 \right)^{1/2} \quad (98)$$

for any  $\mathcal{T} > 0$ , with the last condition of Eq. (97) replaced with  $x(\mathcal{T}) = x^{(1)}$ . Because  $L(x) = \text{diag}(\ell_e(x))$  is defined by

$$\ell_e(x) = \frac{\mathcal{K}_e(x) - \mathcal{K}_{-e}(x)}{\ln(\mathcal{K}_e(x)/\mathcal{K}_{-e}(x))}, \quad (99)$$

$\mathcal{W}$  is defined in reference to some kinetics  $\mathcal{K}$ . Here,  $x(t)$  and  $f(t)$  are taken as independent variables in the minimization, which are related by  $d_t x = \mathbb{S}L(x)f$ ; in particular,  $f(t)$  is not necessarily equal to the original thermodynamic force  $F(x(t))$ . If there is no pair of  $x$  and  $f$  that satisfies the continuity equation in Eq. (97) for a given  $(x^{(0)}, x^{(1)})$ , we may let  $\mathcal{W} = \infty$ . However, we do not consider such a situation, because if  $(x^{(0)}, x^{(1)})$  are realized as the initial and final state of the dynamics (2), then  $x(t)$  and  $F(x(t))$  are a solution.

We can show that the optimal force that provides the value of  $\mathcal{W}$  is conservative. That is, we have the formula

$$\mathcal{W}(x^{(0)}, x^{(1)}) = \inf_{x,\psi} \left( \mathcal{T} \int_0^\mathcal{T} dt \|\mathbb{S}^\top \psi(t)\|_{L(x(t))}^2 \right)^{1/2} \quad (100)$$

with the conditions

$$\frac{dx}{dt} = \mathbb{S}L(x)\mathbb{S}^\top \psi, \quad x(t=0) = x^{(0)}, \quad x(t=\mathcal{T}) = x^{(1)}. \quad (101)$$

We give the proof of this formula in Appendix E.

The mathematical features of the generalized Wasserstein distance are expected to be similar to those of Maas’s distance [42]. In this paper, we demonstrate two important properties

of the Wasserstein distance, namely the constant speed property of a geodesic [43] and the Kantorovich duality [44].

We first explain that a geodesic with respect to the Wasserstein distance has constant speed. In Ref. [43], the authors proved that there exists a geodesic  $p^*$  accompanied by potential  $\psi^*$  with respect to their original definition of the Wasserstein distance (26). This geodesic has constant speed, meaning that  $\|B^\top \psi^*(t)\|_{L(p^*(t))} = \text{const.}$  As we show, the proof is also valid if we assume there is a geodesic  $x^*$  with potential  $\psi^*$  for the generalized Wasserstein distance (100). Because of the Cauchy–Schwarz inequality, for an arbitrary  $\{x, \psi\}$ , we have

$$\mathcal{T} \int_0^\mathcal{T} dt \|\mathbb{S}^\top \psi(t)\|_{L(x(t))}^2 \geq \left( \int_0^\mathcal{T} dt \|\mathbb{S}^\top \psi(t)\|_{L(x(t))} \right)^2. \tag{102}$$

Moreover, using a reparametrization discussed in Ref. [62], we can also prove the opposite inequality,

$$\mathcal{T} \int_0^\mathcal{T} dt \|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))}^2 \leq \left( \int_0^\mathcal{T} dt \|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))} \right)^2 \tag{103}$$

for the geodesic pair  $\{x^*, \psi^*\}$ . We provide the details of this technique in Appendix F. Because the equality of the Cauchy–Schwarz inequality (102) is attained only if  $\|\mathbb{S}^\top \psi(t)\|_{L(x(t))}$  is constant, the geodesic has the constant speed

$$\|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))} = \frac{\mathcal{W}(x^{(0)}, x^{(1)})}{\mathcal{T}}. \tag{104}$$

Next, for a Markov jump process, we prove the Kantorovich duality representation [32] of the Wasserstein distance as in Ref. [44]:

$$\frac{1}{2\mathcal{T}} \mathcal{W}(p^{(0)}, p^{(1)})^2 = \sup_{\psi} (\langle \psi(\mathcal{T}) \rangle_1 - \langle \psi(0) \rangle_0), \tag{105}$$

where  $\psi(t) \in \mathbb{R}^S$  ( $t \in [0, \mathcal{T}]$ ) has to satisfy

$$q^\top d_t \psi(t) + \frac{1}{2} \|B^\top \psi(t)\|_{L(q)}^2 \leq 0 \tag{106}$$

for any probability distribution  $q$ , and  $\langle \cdot \rangle_i$  ( $i = 0, 1$ ) is the expectation value under  $p^{(i)}$ . The maximizer  $\psi^{**}$  of Eq. (105) and the minimizer  $\{p^*, \psi^*\}$  of Eq. (100) provide the same conservative force,  $B^\top \psi^* = B^\top \psi^{**}$ . Furthermore, it can be shown that  $\psi^{**}$  and  $p^*$  attain the equality in Eq. (106) as

$$[p^*(t)]^\top d_t \psi^{**}(t) + \frac{1}{2} \|B^\top \psi^{**}(t)\|_{L(p^*(t))}^2 = 0, \tag{107}$$

which can be seen as a generalization of the Hamilton–Jacobi equation. While the duality representation characterizes the Wasserstein distance as the maximum potential difference without considering the time evolution of the distribution, we can use the Hamilton–Jacobi equation with the differential equation in Eq. (101) to obtain the optimal solution. We provide a proof of the duality formula (105) in Appendix G; there, we assume the existence of an optimal time evolution of  $p$  and  $\psi$ , that is, a geodesic. A proof that is more rigorous but restricted to systems with detailed balanced steady states can be found in Ref. [44].

### B. Wasserstein distance and Onsager excess EPR

Here we show that, analogously to the continuous case [28,29,34,36,39], the Wasserstein distance is connected with the excess entropy production. To this end, we employ the definition given in Eqs. (98) and (97).

Let  $\{\hat{x}(t)\}_{t \in [0, \tau]}$  be a solution of the continuity equation (2). Note that  $\tau$  indicates the physical time interval we focus on while  $\mathcal{T}$  is an arbitrary time parameter. If we set  $x^{(0)} = \hat{x}(t)$ ,  $x^{(1)} = \hat{x}(t + dt)$ , and  $\mathcal{T} = dt$  in Eq. (98), condition (97) reads

$$\begin{aligned} x(dt) - x(0) &= \mathbb{S}L(\hat{x}(t))f(0)dt + o(dt), \\ x(0) &= \hat{x}(t), \quad x(dt) = \hat{x}(t + dt). \end{aligned} \tag{108}$$

Hence for an infinitesimal time interval  $dt$ , the Wasserstein distance can be expanded as

$$\begin{aligned} \mathcal{W}(\hat{x}(t), \hat{x}(t + dt))^2 &= dt^2 \inf_{f, \hat{x}(t) = \mathbb{S}L(\hat{x}(t))f} \|f\|_{L(\hat{x}(t))}^2 + o(dt^2). \end{aligned} \tag{109}$$

The condition of minimization can be translated into  $\{f \mid F(\hat{x}(t)) - f \in \ker \mathbb{S}L\}$ , hence the leading term is identified with the excess EPR  $\sigma^{\text{ex}}$  from Eq. (49). We finally obtain the relation

$$\left( \lim_{dt \rightarrow 0} \frac{\mathcal{W}(\hat{x}(t), \hat{x}(t + dt))}{dt} \right)^2 = \sigma^{\text{ex}}. \tag{110}$$

Intuitively, in the definition of the Wasserstein distance Eq. (98), we optimize over the thermodynamic forces and the corresponding evolution of the state connecting the initial and final state. By contrast, in computing the excess EPR, we restrict the evolution of the state to that of the original dynamics and only optimize over the forces. In the short-time limit of a single time-step, there is no evolution of the state, so both problems are equivalent and we can identify the Wasserstein distance and the excess EPR with each other.

### C. Speed limit

We define  $\dot{l}$  and  $l$  by

$$\dot{l}(t) := \lim_{dt \rightarrow 0} \frac{\mathcal{W}(\hat{x}(t), \hat{x}(t + dt))}{dt}, \quad l := \int_0^\tau dt \dot{l}(t). \tag{111}$$

While  $\mathcal{W}(\hat{x}(0), \hat{x}(\tau))$  gives the shortest length between the initial state  $\hat{x}(0)$  and the final state  $\hat{x}(\tau)$ ,  $l$  quantifies the length of the path  $\{\hat{x}(t)\}_{t \in [0, \tau]}$  in terms of the Wasserstein metric. Using  $l$ , we have the speed limit, as in Ref. [39],

$$l^2 \leq \tau \int_0^\tau dt \sigma^{\text{ex}}. \tag{112}$$

This is because of the Cauchy–Schwarz inequality  $l^2 = (\int_0^\tau dt \dot{l})^2 \leq (\int_0^\tau dt) (\int_0^\tau dt \dot{l}^2)$  and the equality (110).

By the constant speed property, the following triangle inequality is proven for any  $t_1, t_2, t_3 \in [0, \tau]$ :

$$\mathcal{W}(\hat{x}(t_1), \hat{x}(t_3)) \leq \mathcal{W}(\hat{x}(t_1), \hat{x}(t_2)) + \mathcal{W}(\hat{x}(t_2), \hat{x}(t_3)). \tag{113}$$

This relation leads to an inequality between the two distance measures as

$$\mathcal{W}(\hat{x}(0), \hat{x}(\tau)) \leq \int_0^\tau dt \dot{l}(t) = l. \tag{114}$$

The equality holds only when the distribution changes at the constant speed

$$\dot{i}(t) = \frac{\mathcal{W}(\hat{x}(0), \hat{x}(\tau))}{\tau}. \quad (115)$$

By combining Eqs. (112) and (114), we obtain the speed limit

$$\mathcal{W}(\hat{x}(0), \hat{x}(\tau))^2 \leq \tau \int_0^\tau dt \sigma^{\text{ex}}. \quad (116)$$

This generalizes the Vu and Hasegawa speed limit (29), which was introduced in Ref. [31] for systems with detailed balanced steady states, to general Markovian dynamics on a graph that may have no stable steady state.

#### D. Relation to $L^1$ -Wasserstein distance

Another distance that has been considered in the theory of optimal transport is the  $L^1$ -Wasserstein distance [32,55]. This distance is defined by

$$\mathcal{W}_1(p^{(0)}, p^{(1)}) := \inf_{\Pi} \sum_{i,j} d_{ij} \Pi_{ij}, \quad (117)$$

where the infimum is taken over the couplings between the probability distributions  $p^{(0)}$  and  $p^{(1)}$ , that is,

$$\Pi_{ij} \geq 0, \quad \sum_j \Pi_{ij} = p_i^{(0)}, \quad \sum_i \Pi_{ij} = p_j^{(1)}, \quad (118)$$

and  $d_{ij}$  is a weight that satisfies the axioms of distance:  $d_{ij} = d_{ji} \geq 0$ ,  $d_{ij} = 0$  if and only if  $i = j$ , and  $d_{ij} \leq d_{ik} + d_{kj}$ . The  $(i, j)$ -element of a coupling represents how much of the mass on  $i$  is moved to  $j$ , and  $d$  evaluates the cost of the transportation. Here we set  $d_{ij} = 1$  if  $i$  and  $j$  are directly connected, and we assume that the graph of states is connected.

In Ref. [55], one of us derived the Kantorovich duality for this  $L^1$ -Wasserstein distance,

$$\mathcal{W}_1(p^{(0)}, p^{(1)}) = \sum_i \xi_i (p_i^{(1)} - p_i^{(0)}), \quad (119)$$

where  $\xi$  is a potential that satisfies  $\xi_j - \xi_i = d_{ij} = 1$  if and only if the optimal coupling  $\Pi^*$  is positive on  $(i, j)$ ,  $\Pi_{ij}^* > 0$ . Using this formula, we can prove the following relation between the  $L^2$ - and  $L^1$ -Wasserstein distance:

$$\frac{\bar{\mathcal{A}}[p^*]}{2} \mathcal{W}_2(p^{(0)}, p^{(1)})^2 \geq \mathcal{W}_1(p^{(0)}, p^{(1)})^2, \quad (120)$$

where we write the  $L^2$ -Wasserstein distance as  $\mathcal{W}_2$ , and  $\bar{\mathcal{A}}[p^*]$  is the time-averaged dynamical activity of the minimizer  $p^*$  of the  $L^2$ -Wasserstein distance,

$$\bar{\mathcal{A}}[p^*] = \frac{1}{\tau} \int_0^\tau dt \sum_e (k_e p_{i(e)}^* + k_{-e} p_{i'(e)}^*). \quad (121)$$

This is similar to the inequality  $\mathcal{W}_2(p_0, p_\tau) \geq \mathcal{W}_1(p_0, p_\tau)$  for the  $L^2$ - and  $L^1$ -Wasserstein distances of continuous probability distributions [32], apart from the prefactor that represents how frequently transitions occur in the optimal path on average. We provide the proof in Appendix H. We remark that in the discrete case, the  $L^2$ -distance defined above is a dynamics-wise measure and depends on the concrete values of the rates. By contrast, the  $L^1$ -distance depends only on the connectivity of the state network. This differs from the continuous case,

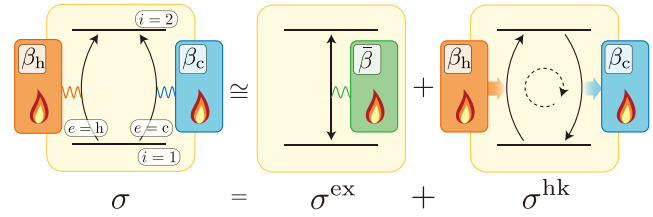


FIG. 5. Conceptual diagram of example VI A. A two-level system is attached to two heat reservoirs at different temperatures. The total dissipation (EPR)  $\sigma$  can be divided into excess EPR  $\sigma^{\text{ex}}$  and housekeeping EPR  $\sigma^{\text{hk}}$ . Both are clearly interpreted as follows:  $\sigma^{\text{ex}}$  corresponds to a relaxation caused by a single reservoir at the “mean” temperature  $\bar{\beta}$ , while  $\sigma^{\text{hk}}$  is attributed to heat transfer that does not change the state.

where both distances are defined in terms of the Euclidean distance.

## VI. EXAMPLES

In this section, we discuss two examples. One clearly shows how the Onsager-projective decomposition works and separates the two aspects of dynamics into excess and housekeeping terms. It illustrates that the short-time TUR (76) improves the bound given in Ref. [23]. The other demonstrates that the decomposition can also be obtained for a nonlinear chemical reaction network that exhibits a limit cycle. We numerically show that in such a system, the conventional HS decomposition fails. On the other hand, we do not discuss other results such as the generalization of finite-time TURs and the Wasserstein speed limit because they have been well studied in [25,61] and [58], respectively.

### A. Two-level system attached to two reservoirs

We illustrate our results in a simple two-level stochastic system. Consider a two-level system attached to two heat reservoirs at inverse temperatures  $\beta_h$  and  $\beta_c$ , respectively. We can calculate the analytical forms of the excess and housekeeping EPR. Let the energy of state  $i$  be  $\epsilon_i$  for  $i = 1, 2$  and  $\epsilon_2 > \epsilon_1$ . There are two kinds of transition associated with the distinct reservoirs. We label each transition with the reservoir at inverse temperature  $\beta_e$  by  $e = h$  or  $c$ , instead of 1 or 2, to avoid confusion. The incidence matrix is then given by

$$B = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}. \quad (122)$$

This setup is illustrated in Fig. 5.

By using the eigendecomposition of  $BLB^T$  and formula (31), we can find the projection matrix

$$P = \frac{1}{\ell_h + \ell_c} \begin{pmatrix} \ell_h & \ell_c \\ \ell_h & \ell_c \end{pmatrix}, \quad (123)$$

where  $\ell_e$  denotes the Onsager coefficient associated with a transition driven by heat bath  $e$ . It leads to the conservative force  $F^*$  as

$$\begin{aligned} F^* &= PF = \frac{\ell_h F_h + \ell_c F_c}{\ell_h + \ell_c} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= -B^T \phi^*, \end{aligned} \quad (124)$$

where, up to an additive constant,

$$\phi^* = -\frac{\ell_h F_h + \ell_c F_c}{\ell_h + \ell_c} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (125)$$

The choice of  $\phi^*$  is not unique, but  $\phi_2^* - \phi_1^* = -F_h^* = -F_c^*$  is uniquely determined, so we write  $\phi_2^* - \phi_1^*$  as  $\Delta\phi^*$ . If we do not consider any explicit external force, the forces read

$$F_e = -\beta_e \Delta\epsilon - \Delta \ln p \quad (126)$$

for  $e = h, c$ , where  $\Delta\epsilon = \epsilon_2 - \epsilon_1$  and  $\Delta \ln p = \ln p_2 - \ln p_1$ . Then, we obtain the expression

$$\Delta\phi^* = \bar{\beta} \Delta\epsilon + \Delta \ln p \quad (127)$$

with

$$\bar{\beta} = \frac{\ell_h \beta_h + \ell_c \beta_c}{\ell_h + \ell_c}. \quad (128)$$

Therefore,  $\phi^*$  can be thought of as a mean potential of the system attached with a single bath at inverse temperature  $\bar{\beta}$ . The excess EPR is then interpreted as the dissipation due to the relaxation to the pseudocanonical distribution  $p_i^{\text{pcan}} \propto e^{-\bar{\beta}\epsilon_i}$ , in view of Eq. (59).

On the other hand,  $F - F^*$  is written as

$$F - F^* = \frac{\Delta\epsilon \Delta\beta}{\ell_h + \ell_c} \begin{pmatrix} \ell_c \\ -\ell_h \end{pmatrix}, \quad (129)$$

where  $\Delta\beta := \beta_c - \beta_h$ . Then, the corresponding current

$$J^{\text{cyc}} := L(F - F^*) = \frac{\ell_h \ell_c \Delta\epsilon \Delta\beta}{\ell_h + \ell_c} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (130)$$

is cyclic ( $BJ^{\text{cyc}} = 0$ ). Moreover, the housekeeping EPR reads

$$\sigma^{\text{hk}} = \frac{\ell_h \ell_c}{\ell_h + \ell_c} (\Delta\epsilon)^2 (\Delta\beta)^2. \quad (131)$$

From this expression, we can see that the (inverse) temperature difference leads to a nonequilibrium steady state with a positive rate of dissipation.

We illustrate the TUR of Eq. (76) and compare our decomposition with the HS decomposition in Fig. 6. We plot the time evolution of the ratios

$$\eta = \frac{2(d_t \langle \psi \rangle)^2}{D_\psi \sigma^{\text{ex}}} \leq 1, \quad \eta' = \frac{2(d_t \langle \psi \rangle)^2}{D_\psi \sigma} \leq 1, \quad (132)$$

where we choose the observable  $\psi = (1, 0)^\top$ . Note, however, that in this case  $\eta$  and  $\eta'$  do not depend on the choice of  $\psi$ , as long as it is nonzero. We find that  $\eta'$  has a maximum value of  $t \simeq 0.3$  and cannot attain the upper bound 1, while  $\eta$  approaches 1 as  $t \rightarrow \infty$  (even though  $\sigma^{\text{ex}}$  vanishes at the same time). This calculation was done with parameters  $\beta_h = 1$ ,  $\beta_c = 2$ ,  $\epsilon_2 - \epsilon_1 = 1$ ,  $k_{-h} = k_{-c} = 1$ , and  $p_0(0)/p_1(0) = 10^3$ .

The inset shows the time evolution of the excess and housekeeping EPR for the two decompositions. Both decompositions give almost the same values in this system. We note that we can also express the HS housekeeping EPR with another mean temperature  $\bar{\beta}$ , which can be defined by

$$-\Delta\epsilon \bar{\beta} := \ln \frac{k_h^+ + k_c^+}{k_h^- + k_c^-}. \quad (133)$$

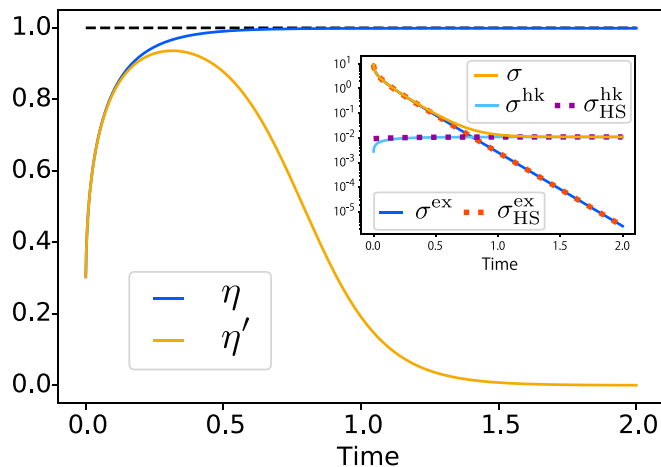


FIG. 6. Verification of the short-time TUR of Eq. (76). The ratios between the lower bound  $2(d_t \langle \psi \rangle)^2 / D_\psi$  vs EPR  $\sigma$  and excess EPR  $\sigma^{\text{ex}}$  are shown.  $\eta$  corresponds to  $\sigma^{\text{ex}}$  and  $\eta'$  to  $\sigma$ . We choose the observable  $\psi = (1, 0)^\top$  but, in this system, the lower bound does not depend on the choice of  $\psi \neq 0$ .  $\eta'$  decreases and tends to zero after its peak because  $\sigma$  does not vanish when  $t \rightarrow \infty$ . On the other hand,  $\eta$  approaches the upper bound of 1, although  $\sigma^{\text{ex}}$  diminishes at the same time. In the inset, we compare our decomposition with the HS decomposition. In this example, where a stable steady state exists, they behave in almost the same way.

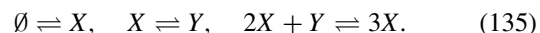
It satisfies  $\beta_h \leq \bar{\beta} \leq \beta_c$  if  $\beta_h \leq \beta_c$  and lets us rewrite the HS housekeeping EPR as

$$\sigma_{\text{HS}}^{\text{hk}} = \Delta\epsilon [(\bar{\beta} - \beta_h)J_h + (\bar{\beta} - \beta_c)J_c]. \quad (134)$$

However, the definition of  $\bar{\beta}$  is not straightforward, and this expression is less easy to interpret than Eq. (131).

## B. Brusselator model of chemical oscillation

Let us consider the Brusselator model of a chemical oscillation. This model consists of three reactions,



We assume the mass action law and set the rate constants  $k_1 = k_3 = k_{-1} = k_{-3} = 1$ ,  $k_2 = 10$ , and  $k_{-2} = 0.1$ . The concentrations of  $X$  and  $Y$  obey the following rate equation:

$$\begin{aligned} \frac{dc_X}{dt} &= 1 - c_X - 10c_X + 0.1c_Y + c_X^2 c_Y - c_X^3, \\ \frac{dc_Y}{dt} &= 10c_X - 0.1c_Y - c_X^2 c_Y + c_X^3. \end{aligned} \quad (136)$$

Here we denote the concentrations by  $c$  rather than  $x$ . This equation admits an unstable steady state  $(c_X, c_Y) = (1, 10)$  and has a limit cycle. The time evolution is shown in the uppermost panel of Fig. 7. In the smaller inset, we show the oscillation of the concentrations  $c_X$  and  $c_Y$ . The bigger graph shows its cyclic behavior. The blue star indicates the unstable fixed point.

We show the behavior of the EPRs along with  $\|\dot{c}\|^2 = \dot{c}_X^2 + \dot{c}_Y^2$  in the middle panel.  $\|\dot{c}\|^2$  is normalized so that it fits in the graph with the EPRs. This figure shows the correlation between  $\sigma^{\text{ex}}$  and  $\|\dot{c}\|^2$ . In this model, the faster the concen-

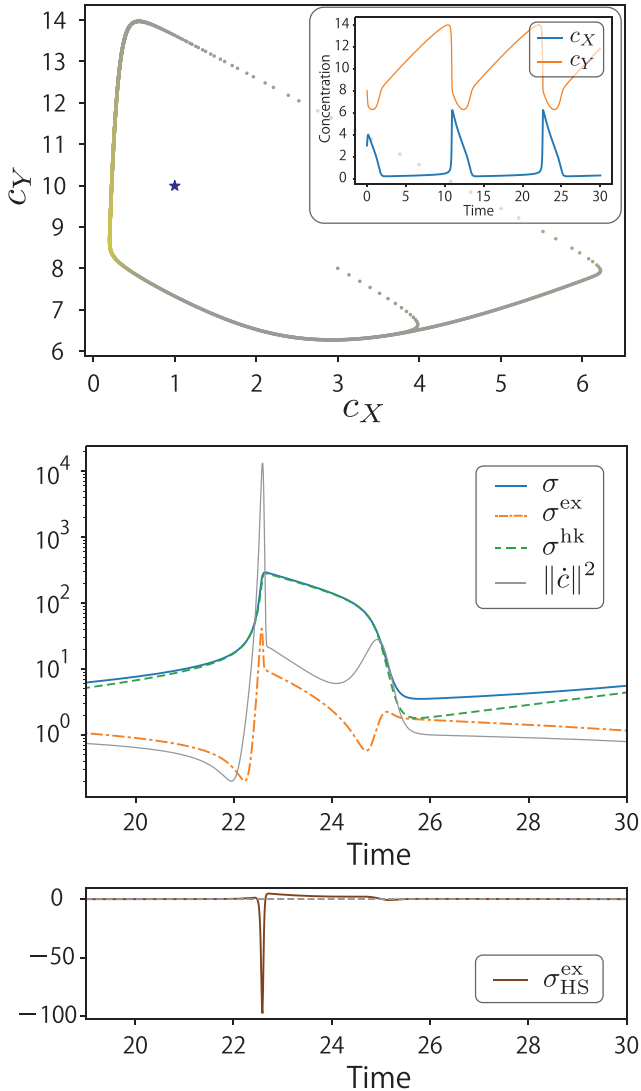


FIG. 7. Housekeeping/excess EPR decomposition in the Brusselator model of chemical oscillations. We show the time evolution of the concentration  $c = (c_X, c_Y)$  from Eq. (136) in the inset of the uppermost panel. The uppermost panel shows the trajectory of  $c$ , which starts from  $c(0) = (3, 8)$  and rapidly converges to the limit cycle. The blue star indicates the unstable fixed point  $c = (1, 10)$ . A point in the trajectory is yellower (brighter) when  $\sigma^{\text{ex}}$  is large relative to  $\sigma^{\text{hk}}$ . The middle figure shows EPRs and the squared norm  $\|\dot{c}\|^2 = \dot{c}_X^2 + \dot{c}_Y^2$  as a function of time. The norm is normalized so that it fits in the graph with EPRs, thus the value is in arbitrary units. Note the correlation between the Onsager excess EPR  $\sigma^{\text{ex}}$  and the speed of the concentration vector  $\|\dot{c}\|^2$  in the middle panel. As shown in the lowermost panel, in this system with an unstable steady state, the HS excess EPR can be negative.

tration changes, the bigger the Onsager excess EPR becomes. The color of the dots in the uppermost panel indicates the size of  $\sigma^{\text{ex}}$  relative to  $\sigma^{\text{hk}}$ . The yellow (bright) color corresponds to parts of the trajectory where  $\sigma^{\text{ex}}$  is large. This situation corresponds to around  $t \simeq 25.8$  in the middle panel.

Below the middle figure, we show that the HS excess EPR can be negative if one defines it using the unstable steady state  $x^{\text{st}}$ . Let us prove this fact in the general case.  $\sigma_{\text{HS}}^{\text{ex}}$  is given by

the time derivative of the KL divergence as

$$\frac{d}{dt} D(x \| x^{\text{st}}) = -\sigma_{\text{HS}}^{\text{ex}} \quad (137)$$

when the flux obeys the mass action law [3,5,6]. For a cyclic solution  $x$  with a period  $T > 0$ , we have

$$\int_0^T dt \sigma_{\text{HS}}^{\text{ex}} = D(x(0) \| x^{\text{st}}) - D(x(T) \| x^{\text{st}}) = 0. \quad (138)$$

This implies that, unless  $\sigma_{\text{HS}}^{\text{ex}}$  vanishes identically, it has to take negative values at some times during the cycle. We stress that this is in contrast to the Onsager projective decomposition, whose excess part is positive by construction.

## VII. DISCUSSION

In this paper, we developed a geometrical perspective on dissipation in general Markovian dynamics on networks. In particular, we found a geometrical excess/housekeeping decomposition of the entropy production rate that helps us to understand dynamics in terms of dissipation. Our decomposition leads to refinements of several thermodynamic uncertainty relations. We have also generalized the  $L^2$ -Wasserstein geometry of a Markovian dynamics on a network, which allows us to derive the thermodynamic speed limit. Because our results are valid for very general dynamics, we expect that they will offer novel insights in various concrete systems.

The geometrical decomposition given by the projection in (31) shows how dynamics can be divided into two modes, one being cyclic and one being a relaxation driven by a conservative force, which can be written as a gradient flow. Even if a stable steady state does not exist, we can identify an instantaneous target of the relaxation mode as in Eq. (58). The Onsager excess entropy production quantifies the dissipation due to this relaxation. On the other hand, the cyclic mode keeps the system out of equilibrium. The size of the cyclic mode defines the Onsager housekeeping entropy production, which can be further decomposed into contributions from individual cycles, generalizing Schnakenberg's decomposition as in Eq. (43). Because these interpretations apply to both Markov jump processes and chemical reaction networks, our decomposition provides a clear and general thermodynamic separation of motion.

We note that one may define a decomposition of the EPR using an orthogonal projection onto the image of  $B^T$  instead of  $\mathbb{S}^T$  as we have done. For a Markov jump process, it provides the same results, while for a chemical reaction network it gives a different decomposition. The excess part of this decomposition vanishes in a steady state only if the steady state is complex balanced, but both parts are always non-negative in contrast to the HS decomposition, thus this decomposition could provide another perspective to chemical reaction networks. However, it may be argued that the projection onto the image of  $\mathbb{S}^T$ , as investigated here, is more physically meaningful. This is because it corresponds to conservative forces defined in terms of the stoichiometric coefficients of the reactions, as is standard in chemical thermodynamics, rather than forces that depend explicitly on the reactant and product chemical complexes.

From the experimental point of view, the edgewise Onsager coefficient, and therefore the Onsager excess and housekeeping EPRs, require information about the kinetics. For this reason, it may be difficult to access them experimentally. However, recent studies have shown that TURs can be utilized to obtain estimates of the entropy production [23,24]. In a similar manner, we can use the refined TURs to estimate the excess and housekeeping EPR using measurable quantities. Although we cannot generally attain equality in the TUR for a discrete system out of equilibrium, since part of the estimation error stems from the log mean inequality between the edgewise Onsager coefficient and the edgewise activity [63], this error gets smaller as the system gets closer to a steady state. In a two-level system, we can approach equality, as shown in Fig. 6 of Sec. VIA, because the steady state is “detailed balanced” in the sense that there is no net current between the states, even though there is a steady flux between the reservoirs.

We have also extended Maas’s definition of  $L^2$ -Wasserstein distance to a general Markovian dynamics on a network. A mathematical advantage of Maas’s distance is that it can be related to master equation dynamics via the Wasserstein gradient flow [42], compared with other optimal transport theoretical distances between two discrete distributions, e.g., the  $L^1$ -Wasserstein distance [32,55] and another discretization of the  $L^2$ -Wasserstein distance [64]. We expect that a deterministic rate equation with a detailed-balanced steady state is also characterized as a Wasserstein gradient flow with the aid of our generalized Wasserstein distance, beyond the usual gradient flow expression as in Eq. (24) [11].

The relation between the geometrical decomposition and the generalized Wasserstein distance presented in Eq. (110) is parallel to recent results for Langevin systems [28,29]. Therefore, our generalization of Maas’s  $L^2$ -Wasserstein distance is not only important from a mathematical point of view, but it also has a physical interpretation as the minimum excess entropy of a process connecting the initial and final state. As discussed above, obtaining a meaningful “minimum entropy production” requires a constraint, that is, keeping some quantity fixed during the minimization [55–58,65]. In the present study, we fix the Onsager coefficients, whereas in previous studies [56,57,65] the symmetric parts or one of the two directions of the transition rates were fixed. While the latter may be the intuitive choice from an operational point of view, our approach establishes the connection between minimum entropy production, conservative forces, and optimal transport. Although we have shown how the excess EPR can be understood as a minimum EPR at each time point, further study is needed to investigate how the finite-time optimization of the Wasserstein distance works practically and how the different approaches are related to each other.

One drawback of the discrete  $L^2$ -Wasserstein distance is that it needs a kinetics to refer to. Since the reference kinetics has to be time-independent, we cannot obtain the Wasserstein speed limit in externally driven systems. Note that we cannot naively generalize the Wasserstein distance to the case in which the reference kinetics is time-dependent. While in the original definition the time interval is not definite, we should care about how to define the time integral for a time-dependent reference. Whichever way we choose, we cannot

use the conventional reparametrization technique, which is well described in Appendix F, because it requires the reference kinetics to be time-independent. As a result, it becomes difficult to prove several properties, such as the triangle inequality and constant speed property of a geodesic.

Let us mention a complementary study that we have recently done. In this paper, we considered an EPR decomposition by studying the space of thermodynamic forces using Euclidean geometry. However, it is also possible to define such a decomposition using the non-Euclidean information geometry. We develop this perspective in Ref. [66], where we derive several thermodynamic bounds based on an information-geometric excess/housekeeping decomposition.

Finally, we remark about a missing piece: nonlinear systems in continuous space. Two of the present authors, with another coauthor, presented a geometric decomposition of the EPR for linear stochastic systems in continuous space [28,29]. In this work, we developed it for Markov jump processes (linear stochastic system) and chemical master equations (nonlinear deterministic systems) in discrete space. The story remains to be written for nonlinear systems in continuous space, such as reaction-diffusion chemical systems and hydrodynamic systems. For such systems, geometric techniques for decomposing EPR may shed light on new kinds of thermodynamic constraints and tradeoffs.

## ACKNOWLEDGMENTS

K.Y. and S.I. thank Ryuna Nagayama and Kiyoshi Kanazawa for fruitful discussions. K.Y. is supported by a Grant-in-Aid for JSPS Fellows (Grant No. 22J21619). A.D. is supported by JSPS KAKENHI Grants No. 19H05795 and No. 22K13974. S.I. is supported by JSPS KAKENHI Grants No. 19H05796, No. 21H01560, and No. 22H01141, and UTEC-UTokyo FSI Research Grant Program.

## APPENDIX A: EQUIVALENCE OF DEFINITIONS OF WASSERSTEIN DISTANCE

Let us show that our definitions of the Wasserstein distance (26) and (100) are consistent with previous studies, e.g., Ref. [42]. That is, the approach from [42] is equivalent to ours if detailed balance holds. In Ref. [42], the author defines the relative distribution

$$\rho_i := \frac{p_i}{p_i^{\text{eq}}} \quad (\text{A1})$$

and defines the Wasserstein distance by

$$\begin{aligned} & \mathcal{W}'(p^{(0)}, p^{(1)}) \\ & := \inf_{p, \psi} \left( \int_0^1 dt \frac{1}{2} \sum_{i,j} (\psi_i - \psi_j)^2 k_{i \rightarrow j} p_i^{\text{eq}} \mathbb{L}(\rho_i, \rho_j) \right)^{1/2} \end{aligned} \quad (\text{A2})$$

with conditions

$$\frac{d}{dt} \rho_i + \sum_j (\psi_j - \psi_i) k_{i \rightarrow j} \mathbb{L}(\rho_i, \rho_j) = 0, \quad (\text{A3})$$

$$p(t=0) = p^{(0)}, \quad p(t=1) = p^{(1)}, \quad (\text{A4})$$



where  $\mathbb{L}$  is the logarithmic mean

$$\mathbb{L}(a, b) := \frac{a - b}{\ln a - \ln b}. \quad (\text{A5})$$

First, by symmetry, the integrand is rewritten as

$$\begin{aligned} & \frac{1}{2} \sum_{i,j} (\psi_i - \psi_j)^2 k_{i \rightarrow j} p_i^{\text{eq}} \mathbb{L}(\rho_i, \rho_j) \\ &= \sum_e ([B^\top \psi]_e)^2 k_e p_{i(e)}^{\text{eq}} \mathbb{L}(\rho_{i(e)}, \rho_{i'(e)}), \end{aligned} \quad (\text{A6})$$

where we used the detailed balance condition  $k_e p_{i(e)}^{\text{eq}} = k_{-e} p_{i'(e)}^{\text{eq}}$ . Moreover, the numerator of the logarithmic mean times  $k_e p_{i(e)}^{\text{eq}}$  gives  $J_e$  because

$$k_e p_{i(e)}^{\text{eq}} \rho_{i(e)} = k_e p_{i(e)}, \quad (\text{A7})$$

$$k_e p_{i(e)}^{\text{eq}} \rho_{i'(e)} = k_{-e} p_{i'(e)}^{\text{eq}} \rho_{i'(e)} = k_{-e} p_{i'(e)}. \quad (\text{A8})$$

Furthermore, the denominator reads

$$\ln \rho_{i(e)} - \ln \rho_{i'(e)} = \ln \frac{k_e p_{i(e)}}{k_{-e} p_{i'(e)}} = F_e \quad (\text{A9})$$

due to the detailed balance condition. Therefore, we have

$$k_e p_{i(e)}^{\text{eq}} \mathbb{L}(\rho_{i(e)}, \rho_{i'(e)}) = \frac{J_e}{F_e} = \ell_e, \quad (\text{A10})$$

and the integrand is equal to

$$\sum_e ([B^\top \psi]_e)^2 \ell_e = \|B^\top \psi\|_L^2. \quad (\text{A11})$$

We next consider the conditions for minimization. From Eq. (A10), we find that condition (A3) is equivalent to

$$\frac{d}{dt} p_i + \sum_j (\psi_j - \psi_i) \ell_{i \rightarrow j} = 0, \quad (\text{A12})$$

where  $\ell_{i \rightarrow j} = \ell_e$  if the edge  $e$  corresponding to the transition  $i \rightarrow j$  is in the set of edges  $\{1, \dots, E\}$ , and  $\ell_{i \rightarrow j} = 0$  if not (so  $k_{i \rightarrow j} = k_{j \rightarrow i} = 0$ ). Due to the symmetry  $\ell_{i \rightarrow j} = \ell_{j \rightarrow i}$ , we have

$$\begin{aligned} & \sum_j (\psi_j - \psi_i) \ell_{i \rightarrow j} \\ &= \sum_{e: i \rightarrow *} \ell_e \sum_j B_{je} \psi_j - \sum_{e: * \rightarrow i} \ell_e \sum_j B_{je} \psi_j \\ &= - \sum_e B_{ie} \ell_e \sum_j B_{je} \psi_j = -[BLB^\top \psi]_i, \end{aligned} \quad (\text{A13})$$

therefore the conditions are equivalent to ours in Eq. (27).

Finally, we complete proving the equivalence

$$\begin{aligned} & \inf_{p, \psi} \left( \frac{1}{2} \int_0^1 dt \sum_{i,j} (\psi_i - \psi_j)^2 k_{i \rightarrow j} p_i^{\text{eq}} \mathbb{L}(\rho_i, \rho_j) \right)^{1/2} \\ &= \inf_{p, \psi} \left( \int_0^1 dt \|B^\top \psi\|_L^2 \right)^{1/2} \end{aligned} \quad (\text{A14})$$

with equivalent conditions

$$\frac{d}{dt} \rho_i + \sum_j (\psi_j - \psi_i) k_{i \rightarrow j} \mathbb{L}(\rho_i, \rho_j) = 0, \quad (\text{A15})$$

$$p(t=0) = p^{(0)}, \quad p(t=1) = p^{(1)} \quad (\text{A16})$$

for the left-hand side and

$$\frac{d}{dt} p = BLB^\top \psi, \quad p(t=0) = p^{(0)}, \quad p(t=1) = p^{(1)} \quad (\text{A17})$$

for the right-hand side.

## APPENDIX B: ONSAGER COEFFICIENT AND DIFFUSION COEFFICIENT

We discuss how the Onsager coefficients are connected to the diffusion coefficient in Langevin systems when we consider a Markov jump process on a square lattice that approximates a continuous space. The result here was obtained in Ref. [58] for one-dimensional cases, and we generalize it to  $n$ -dimensional systems for an arbitrary dimension  $n$ .

Let  $i$  denote a point on an  $n$ -dimensional square lattice with the lattice constant  $\Delta x \ll 1$ . The discrete probability distribution  $p_i$  can be approximated by a smooth density function  $P(x)$  as  $p_i \simeq P(i)(\Delta x)^n$ . A transition can be designated by its starting point and the direction  $d = \pm 1, \dots, \pm n$ . With  $w_i^{(d)}$  being the transition rate of the jump from  $i$  in direction  $d$ , we have the master equation

$$\frac{d p_i}{dt} = \sum_{d=\pm 1}^{\pm n} (w_{i+d}^{(-d)} p_{i+d} - w_i^{(d)} p_i), \quad (\text{B1})$$

where we make the time dependence implicit and  $i+d$  is the point next to  $i$  in the direction of  $d$ . Note that now each edge is determined by the combination of the starting point and the direction of movement.

As in previous work [58,67], we assume the transition rates are expanded as the series

$$w_i^{(d)} = \frac{D_d(i)}{\Delta x^2} + \frac{f_d(i)}{\Delta x} + o(1). \quad (\text{B2})$$

Here, we assume  $D_d(x)$  to be a symmetric function [ $D_{-d}(x) = D_d(x)$ ] and  $f_d(x)$  antisymmetric [ $f_{-d}(x) = -f_d(x)$ ]. Then the master equation reads

$$\begin{aligned} \frac{d p_i}{dt} &= \frac{1}{\Delta x^2} \sum_{d=\pm 1}^{\pm n} [D_{-d}(i+d) p_{i+d} - D_d(i) p_i] \\ &+ \frac{1}{2\Delta x} \sum_{d=\pm 1}^{\pm n} [f_{-d}(i+d) p_{i+d} - f_d(i) p_i] + o(1) \\ &= \sum_{d=1}^n \frac{D_d(i+d) p_{i+d} + D_d(i-d) p_{i-d} - 2D_d(i) p_i}{\Delta x^2} \\ &- \sum_{d=1}^n \frac{f_d(i+d) p_{i+d} - f_d(i-d) p_{i-d}}{2\Delta x} + o(1). \end{aligned} \quad (\text{B3})$$

By taking the limit  $\Delta x \rightarrow 0$ , we have the Fokker–Planck equation

$$\frac{\partial}{\partial t} P(x) = - \sum_{d=1}^n \frac{\partial}{\partial x_d} (f_d(x) P(x)) + \sum_{d=1}^n \frac{\partial^2}{\partial x_d^2} (D_d(x) P(x)). \quad (\text{B4})$$

Now, we can show that to the leading order, the current and the force on edge  $(i, d)$  are given as

$$J_d(i) \simeq \frac{f_d(i) p_i}{\Delta x} - \frac{\partial_d (D_d(i) p_i)}{\Delta x}, \quad (\text{B5})$$

$$F_d(i) \simeq \frac{f_d(i)}{D_d(i)} \Delta x - \frac{\partial_d (D_d(i) p_i)}{D_d(i) p_i} \Delta x, \quad (\text{B6})$$

where  $\partial_d (D_d(i) p_i)$  stands for  $[D_d(i+d) p_{i+d} - D_d(i) p_i] / \Delta x$ . Thus we have

$$\ell_d(i) = \frac{D_d(i) p_i}{\Delta x^2} + o(\Delta x^{-2}), \quad (\text{B7})$$

which shows that fixing the functional form of the Onsager coefficients is similar to fixing the diffusion coefficient in a Langevin system.

### APPENDIX C: DETAILED COMPARISON WITH HS DECOMPOSITION

In addition to the discussion given in Sec. III F, we can further establish a connection between the HS decomposition and Onsager-projective decomposition. Let  $P_\Psi$  be the orthogonal projection operator onto the vector  $\mathbb{S}^\top \Psi$  with respect to the metric  $L$ , with  $\Psi$  given by Eq. (69),

$$P_\Psi := \frac{\mathbb{S}^\top \Psi \Psi^\top \mathbb{S}}{\|\mathbb{S}^\top \Psi\|_L^2}. \quad (\text{C1})$$

We define the pseudo-HS decomposition by

$$\sigma_{\text{pHS}}^{\text{ex}} := \|P_\Psi F\|_L^2, \quad (\text{C2})$$

$$\sigma_{\text{pHS}}^{\text{hk}} := \|(I - P_\Psi) F\|_L^2. \quad (\text{C3})$$

Because  $P_\Psi$  is an orthogonal projection operator, the sum of the two terms is equal to the total EPR:

$$\sigma = \sigma_{\text{pHS}}^{\text{ex}} + \sigma_{\text{pHS}}^{\text{hk}}. \quad (\text{C4})$$

Since  $PP_\Psi = P_\Psi$ , we have the inequality

$$\sigma_{\text{pHS}}^{\text{ex}} \leq \sigma^{\text{ex}}, \quad (\text{C5})$$

which allows us to define a coupling term as in Refs. [28,29],

$$\sigma^{\text{cpl}} := \sigma^{\text{ex}} - \sigma_{\text{pHS}}^{\text{ex}} \geq 0, \quad (\text{C6})$$

and decompose the EPR into three positive contributions

$$\sigma = \sigma_{\text{pHS}}^{\text{ex}} + \sigma^{\text{cpl}} + \sigma^{\text{hk}}. \quad (\text{C7})$$

However, the pseudo-HS decomposition does not generally coincide with the genuine HS decomposition. A straightforward calculation reveals their connection,

$$\sigma_{\text{pHS}}^{\text{ex}} = \frac{(\sigma_{\text{HS}}^{\text{ex}})^2}{\|\mathbb{S}^\top \Psi\|_L^2}. \quad (\text{C8})$$

If  $\sigma_{\text{HS}}^{\text{ex}} = \|\mathbb{S}^\top \Psi\|_L^2$  holds, the two give the same value, but it is not the case in general.

Moreover, the difference between  $\sigma_{\text{HS}}^{\text{ex}}$  and  $\|\mathbb{S}^\top \Psi\|_L^2$  is of the same order as each individual term even when the system is near a steady state; in other words, it is not small. Let  $\sup_\alpha |x_\alpha - x_\alpha^{\text{st}}| = \epsilon$  be very small, with  $x^{\text{st}}$  a steady state and assume  $d_t x = \mathbb{S}J$  is of the same order. Then,  $\sigma_{\text{HS}}^{\text{ex}}$  and  $\|\mathbb{S}^\top \Psi\|_L^2$  are of order of  $\epsilon^2$  because  $\Psi$  and  $\mathbb{S}J$  are both of order  $\epsilon$ . On the other hand, because  $\mathbb{S}L^{\text{st}} F^{\text{st}} = \mathbb{S}J^{\text{st}} = 0$ , the difference between  $\sigma_{\text{HS}}^{\text{ex}}$  and  $\|\mathbb{S}^\top \Psi\|_L^2$  becomes

$$\begin{aligned} \sigma_{\text{HS}}^{\text{ex}} - \|\mathbb{S}^\top \Psi\|_L^2 &= -\Psi^\top \mathbb{S} (J + L \mathbb{S}^\top \Psi) \\ &= -\Psi^\top \mathbb{S} L (F + \mathbb{S}^\top \Psi) \\ &= -\Psi^\top \mathbb{S} L F^{\text{st}} \\ &= -\Psi^\top \mathbb{S} (L - L^{\text{st}}) F^{\text{st}}, \end{aligned} \quad (\text{C9})$$

where the third line comes from Eq. (67). Here,  $\Psi$  and  $L - L^{\text{st}}$  are of the order of  $\epsilon$ , while  $F^{\text{st}}$  is generally of the order of 1, so that we see the difference is of the order of  $\epsilon^2$ , just as  $\sigma_{\text{HS}}^{\text{ex}}$  and  $\|\mathbb{S}^\top \Psi\|_L^2$ . The equality only holds when we have the detailed balance condition, that is, for conservative forces. In this case, we find  $\sigma = \sigma_{\text{HS}}^{\text{ex}} = \sigma_{\text{pHS}}^{\text{ex}}$ .

### APPENDIX D: DERIVATIONS OF FINITE-TIME TURs

#### 1. Housekeeping TUR

To prove the housekeeping TUR (87), let us consider a parametrization of transition rates by introducing the interpolated force  $F_e^\theta = F_e^* + \theta(F_e - F_e^*)$  as

$$k_e^\theta p_{l(e)} = \ell_e g(F_e^\theta), \quad (\text{D1})$$

$$\text{with } g(x) = \frac{x}{1 - e^{-x}}, \quad (\text{D2})$$

where  $\theta \in \mathbb{R}$  is an interpolation parameter. We also define  $J_e^\theta := k_e^\theta p_{l(e)} - k_{-e}^\theta p_{l'(-e)}$ . One can easily show that  $J_e^\theta = \ell_e F_e^\theta$  and  $J_e^\theta = J_e^* + \theta(J_e - J_e^*)$  hold. Because  $BJ^\theta = BJ$  for any  $\theta$  and any time, the solution of the original ( $\theta = 1$ ) master equation  $\{p(t)\}_{t \in [0, \tau]}$  solves  $d_t p(t) = BJ^\theta(t)$ .

In general, we denote the path probability for the master equation by  $\mathbb{P}$  and that for the master equation with modified rate constants  $k^\theta$ , such as those defined in Eq. (D1), by  $\mathbb{P}^\theta$ . The KL divergence between  $\mathbb{P}^\theta$  and  $\mathbb{P}^{\theta'}$  is given by

$$\begin{aligned} D(\mathbb{P}^\theta \| \mathbb{P}^{\theta'}) &= \int d\Gamma \mathbb{P}^\theta(\Gamma) \ln \frac{\mathbb{P}^\theta(\Gamma)}{\mathbb{P}^{\theta'}(\Gamma)} \\ &= \int_0^\tau dt \sum_{\pm e} \left( \ln \frac{k_e^\theta}{k_e^{\theta'}} + \frac{k_e^{\theta'}}{k_e^\theta} - 1 \right) k_e^\theta p_{l(e)}^\theta, \end{aligned} \quad (\text{D3})$$

where  $\Gamma$  indicates a path,  $\sum_{\pm e}$  denotes the summation over both directions of edges, and  $p^\theta$  is the solution of the master equation given by  $k^\theta$ . We always assume that the initial distribution does not depend on parameter  $\theta$ . The KL divergence can also be expressed as

$$D(\mathbb{P}^\theta \| \mathbb{P}^{\theta'}) = \int_0^\tau dt D(k^\theta p^\theta \| k^{\theta'} p^{\theta'}), \quad (\text{D4})$$

where  $D(k^\theta p^\theta \| k^{\theta'} p^{\theta'})$  is the generalized KL divergence between the two ‘‘distributions’’  $\{k_e^\theta p_{l(e)}^\theta\}_{e=\pm 1, \pm 2, \dots}$  and

$\{k_e^{\theta'} p_{i(e)}^{\theta'}\}_{e=\pm 1, \pm 2, \dots}$ , that is,

$$D(k^\theta p^\theta || k^{\theta'} p^{\theta'}) = \sum_{\pm e} \left[ k_e^\theta p_{i(e)}^\theta \ln \frac{k_e^\theta p_{i(e)}^\theta}{k_e^{\theta'} p_{i(e)}^{\theta'}} - k_e^\theta p_{i(e)}^\theta + k_e^{\theta'} p_{i(e)}^{\theta'} \right]. \quad (\text{D5})$$

When we consider  $\theta' = \theta + d\theta$  with a very small  $d\theta$ , the KL divergence becomes

$$\begin{aligned} D(\mathbb{P}^\theta || \mathbb{P}^{\theta+d\theta}) &= \int_0^\tau dt D(k^\theta p^\theta || k^{\theta+d\theta} p^{\theta+d\theta}) \\ &= \frac{d\theta^2}{2} \int_0^\tau dt \sum_{\pm e} k_e^\theta p_{i(e)}^\theta (\partial_\theta \ln k_e^\theta)^2 + o(d\theta^2). \end{aligned} \quad (\text{D6})$$

In the parametrization (D1), the relation

$$\partial_\theta \ln k_e^\theta = (F_e - F_e^*) \frac{d}{dx} \ln g(x) \Big|_{x=F_e^\theta} \quad (\text{D7})$$

holds. Thus, the KL divergence becomes

$$\begin{aligned} D(\mathbb{P}^\theta || \mathbb{P}^{\theta+d\theta}) &= \frac{d\theta^2}{2} \int_0^\tau dt \sum_{\pm e} \ell_e (F_e - F_e^*)^2 h(F_e^\theta) + o(d\theta^2), \end{aligned} \quad (\text{D8})$$

with

$$\begin{aligned} h(x) &= g(x) \left( \frac{d}{dx} \ln g(x) \right)^2 \\ &= \frac{x}{1 - e^{-x}} \left( \frac{e^x - x - 1}{x(e^x - 1)} \right)^2. \end{aligned} \quad (\text{D9})$$

One can easily show that  $h(x) + h(-x) \leq 1/2$ . Therefore, we have the following inequality regardless of the value of  $\theta$ :

$$D(\mathbb{P}^\theta || \mathbb{P}^{\theta+d\theta}) \leq \frac{d\theta^2}{4} \int_0^\tau dt \sigma^{\text{hk}}. \quad (\text{D10})$$

The fluctuation-response inequality (12) of Ref. [26] tells us that for any current observable  $\mathcal{J}_w$  defined in Sec. IV C,

$$\frac{(\langle \mathcal{J}_w \rangle_\tau^{\theta+d\theta} - \langle \mathcal{J}_w \rangle_\tau^\theta)^2}{2 \text{Var}^{\theta+d\theta}(\mathcal{J}_w)} \leq D(\mathbb{P}^{\theta+d\theta} || \mathbb{P}^\theta), \quad (\text{D11})$$

where  $\langle \cdot \rangle_\tau^\theta$  and  $\text{Var}^\theta(\cdot)$  are the average and variance regarding  $\mathbb{P}^\theta$ . Because

$$\langle \mathcal{J}_w \rangle_\tau^\theta = \langle \mathcal{J}_w^{\text{ex}} \rangle_\tau + \theta \langle \mathcal{J}_w^{\text{hk}} \rangle_\tau, \quad (\text{D12})$$

setting  $\theta = 1$  in Eq. (D11) and combining with Eq. (D10), we obtain the desired inequality in Eq. (87).

## 2. Excess TUR

Let us prove the relation (92). We separate the time dependence as  $t = \tau s$ , so that  $s \in [0, 1]$ , and write  $\tilde{f}(s; \tau) = f(\tau s)$  for any function of time  $f$ . The assumption on the functional form of the transition rate  $k$  leads to the conclusion that  $\bar{k}$  only

depends on  $s$ . The master equation reads

$$\partial_s \bar{p}(s; \tau) = \tau B \bar{J}(s; \tau), \quad (\text{D13})$$

$$\bar{J}_e(s; \tau) = \bar{k}_e(s) \bar{p}_{i(e)}(s; \tau) - \bar{k}_{-e}(s) \bar{p}_{i'(e)}(s; \tau). \quad (\text{D14})$$

It is a natural assumption that the initial distribution  $\bar{p}(0; \tau)$  does not depend on  $\tau$ .

We first consider the tilting of the transition rates with  $\tau$  fixed,

$$\bar{k}_e^\theta(s; \tau) = \bar{k}_e(s) \exp\left(\frac{\bar{J}_e^*(s; \tau)}{\bar{\chi}_e(s; \tau)}\theta\right), \quad (\text{D15})$$

where  $\theta$  is the parameter of the tilted transition rate. From the general formula (D6), the KL divergence between  $\mathbb{P}^0$  and  $\mathbb{P}^{d\theta}$  given by the tilting (D15) can be calculated as

$$D(\mathbb{P}^0 || \mathbb{P}^{d\theta}) = \frac{\tau}{2} d\theta^2 \int_0^1 ds \sum_e \frac{[\bar{J}_e^*(s; \tau)]^2}{\bar{\chi}_e(s; \tau)} + o(d\theta^2) \quad (\text{D16})$$

because  $\partial_\theta \ln \bar{k}_e^\theta(s) = \bar{J}_e^*(s; \tau) / \bar{\chi}_e(s; \tau)$ . From the log mean inequality, we also have

$$\begin{aligned} \frac{\tau}{2} \int_0^1 ds \sum_e \frac{[\bar{J}_e^*(s; \tau)]^2}{\bar{\chi}_e(s; \tau)} &\leq \frac{\tau}{4} \int_0^1 ds \sum_e \frac{[\bar{J}_e^*(s; \tau)]^2}{\bar{\ell}_e(s; \tau)} \\ &= \frac{1}{4} \int_0^\tau dt \sum_e \frac{[\bar{J}_e^*(t)]^2}{\bar{\ell}_e(t)} \\ &= \frac{1}{4} \int_0^\tau dt \sigma^{\text{ex}}. \end{aligned} \quad (\text{D17})$$

Thus, the fluctuation-response inequality provides

$$\frac{(\partial_\theta \langle \mathcal{J}_w \rangle_\tau^\theta |_{\theta=0})^2}{\text{Var}(\mathcal{J}_w)} \leq \frac{1}{2} \int_0^\tau dt \sigma^{\text{ex}}. \quad (\text{D18})$$

Next, we identify the quantity  $\partial_\theta \langle \mathcal{J}_w \rangle_\tau^\theta |_{\theta=0}$ . Let  $\{\bar{p}^\theta(s; \tau)\}_{s \in [0, 1]}$  be the solution of the master equation given by  $\bar{k}_e^\theta(s; \tau)$ ,

$$\partial_s \bar{p}^\theta(s; \tau) = \tau B \bar{J}^\theta(s; \tau), \quad (\text{D19})$$

$$\bar{J}_e^\theta(s; \tau) = \bar{k}_e^\theta(s; \tau) \bar{p}_{i(e)}^\theta(s; \tau) - \bar{k}_{-e}^\theta(s; \tau) \bar{p}_{i'(e)}^\theta(s; \tau), \quad (\text{D20})$$

where  $\theta = 0$  reproduces the original dynamics Eqs. (D13) and (D14). If we use the notation

$$\partial_\theta \bar{p}_i^\theta(s; \tau) |_{\theta=0} = r_i(s; \tau), \quad (\text{D21})$$

a straightforward calculation shows

$$\begin{aligned} \partial_\theta \bar{J}_e^\theta(s; \tau) |_{\theta=0} &= \bar{J}_e^*(s; \tau) + [\bar{k}_e(s) r_{i(e)}(s; \tau) - \bar{k}_{-e}(s) r_{i'(e)}(s; \tau)]. \end{aligned} \quad (\text{D22})$$

Because  $\tau B \bar{J}^* = \tau B \bar{L} \bar{F}^* = \tau B \bar{J} = \partial_s \bar{p}(s; \tau)$ , we find that  $r$  satisfies

$$\begin{aligned} \partial_s r_i(s; \tau) - \partial_s \bar{p}_i(s; \tau) &= \tau \sum_e B_{ie} [\bar{k}_e(s) r_{i(e)}(s; \tau) - \bar{k}_{-e}(s) r_{i'(e)}(s; \tau)]. \end{aligned} \quad (\text{D23})$$

We also consider another infinitesimal perturbation by changing  $\tau$  to  $\theta\tau$  where  $\theta = 1$  also reproduces the original dynamics

Eqs. (D13) and (D14). The master equation reads

$$\partial_s \bar{p}(s; \theta \tau) = \theta \tau B \bar{J}(s; \theta \tau). \quad (\text{D24})$$

By calculating

$$\partial_\theta \partial_s \bar{p}(s; \theta \tau)|_{\theta=1} = \partial_\theta [\theta \tau B \bar{J}(s; \theta \tau)]|_{\theta=1}, \quad (\text{D25})$$

we find

$$\begin{aligned} \partial_s [\tau \partial_\tau \bar{p}_i(s; \tau)] - \partial_s \bar{p}_i(s; \tau) &= \tau \sum_e B_{ie} \tau \partial_\tau \bar{J}_e(s; \tau) \\ &= \tau \sum_e B_{ie} [\bar{k}_e(s) [\tau \partial_\tau \bar{p}_{i(e)}(s; \tau)] - \bar{k}_{-e}(s) [\tau \partial_\tau \bar{p}'_{i(e)}(s; \tau)]]. \end{aligned} \quad (\text{D26})$$

It is the same equation as Eq. (D23). Because  $\tau \partial_\tau \bar{p}(s; \tau)$  and  $r(s; \tau)$  satisfy the same first-order differential equation and they share the initial condition  $\tau \partial_\tau \bar{p}(0; \tau) = r(0; \tau) = 0$ , we conclude that  $r(s; \tau) = \tau \partial_\tau \bar{p}(s; \tau)$  and

$$\bar{k}_e(s) r_{i(e)}(s; \tau) - \bar{k}_{-e}(s) r'_{i(e)}(s; \tau) = \tau \partial_\tau \bar{J}_e(s; \tau). \quad (\text{D27})$$

Thus, Eq. (D22) reads

$$\partial_\theta \bar{J}_e^\theta(s; \tau)|_{\theta=0} = \bar{J}_e^*(s; \tau) + \tau \partial_\tau \bar{J}_e(s; \tau), \quad (\text{D28})$$

and the current observable changes by tilting as

$$\begin{aligned} \partial_\theta \langle \mathcal{J}_w \rangle_\tau^\theta|_{\theta=0} &= \tau \int_0^1 ds \sum_e w_e(s) \partial_\theta \bar{J}_e^\theta(s; \tau)|_{\theta=0} \\ &= \langle \mathcal{J}_w^{\text{ex}} \rangle_\tau + \tau \int_0^1 ds \sum_e w_e(s) \tau \partial_\tau \bar{J}_e(s; \tau) \\ &= \langle \mathcal{J}_w^{\text{ex}} \rangle_\tau + \tau \partial_\tau \langle \mathcal{J}_w \rangle_\tau - \langle \mathcal{J}_w^{\text{hk}} \rangle_\tau \\ &= \tau \partial_\tau \langle \mathcal{J}_w \rangle_\tau - \langle \mathcal{J}_w^{\text{hk}} \rangle_\tau. \end{aligned} \quad (\text{D29})$$

Therefore, Eq. (D18) finally leads to the TUR,

$$\frac{(\tau \partial_\tau \langle \mathcal{J}_w \rangle_\tau - \langle \mathcal{J}_w^{\text{hk}} \rangle_\tau)^2}{\text{Var}(\mathcal{J}_w)} \leq \frac{1}{2} \int_0^\tau dt \sigma^{\text{ex}}. \quad (\text{D30})$$

#### APPENDIX E: DERIVATION OF EQ. (100)

We prove that  $\mathcal{W}$  provided in Eq. (98) is also given by Eq. (100). Let us introduce the functional

$$\begin{aligned} \mathcal{I}[x, f, \psi] &= \int_0^\tau dt \left( \frac{1}{2} \|f(t)\|_{L(x(t))}^2 \right. \\ &\quad \left. + \psi(t)^\top [d_t x(t) - \text{SL}(x(t))f(t)] \right). \end{aligned} \quad (\text{E1})$$

With  $\psi$  being a Lagrange multiplier, Eq. (98) is rewritten as

$$\frac{1}{2\mathcal{T}} \mathcal{W}(x^{(0)}, x^{(1)})^2 = \inf_{x, f} \sup_{\psi} \mathcal{I}[x, f, \psi]. \quad (\text{E2})$$

We assume that there exists a triple  $(x^*, f^*, \psi^*)$  that provides the optimal value of the right-hand side. Then, it satisfies

$$\frac{\delta \mathcal{I}}{\delta f_e(t)} = \ell_e(x^*(t)) f_e^*(t) - [\psi^*(t)^\top \mathbb{S}]_e \ell_e(x^*(t)) = 0, \quad (\text{E3})$$

so we have

$$f^*(t) = \mathbb{S}^\top \psi^*(t). \quad (\text{E4})$$

Therefore, the optimal force is conservative, and it is sufficient to consider conservative forces when we calculate the Wasserstein distance.

#### APPENDIX F: DERIVATION OF EQ. (103)

Here we prove the inequality in Eq. (103) by tracing the proof of Theorem 5.4 in Ref. [62]. Let  $\{x^*, \psi^*\}$  be a minimizer of the Wasserstein distance (100). We define a function with sufficiently small  $\epsilon > 0$  as

$$\tilde{s}_\epsilon(t) := \int_0^t dt' [\epsilon + \|\mathbb{S}^\top \psi^*(t')\|_{L(x^*(t'))}^2]^{1/2}. \quad (\text{F1})$$

Then,  $d_t \tilde{s}_\epsilon(t) > \epsilon$ , and its inverse  $\tilde{t}_\epsilon = \tilde{s}_\epsilon^{-1} : [0, \tilde{s}_\epsilon(\mathcal{T})] \rightarrow [0, \mathcal{T}]$  is well defined; thus,  $\tilde{s}_\epsilon$  works as a time coordinate. The functions satisfy

$$\frac{d\tilde{t}_\epsilon}{ds}(\tilde{s}_\epsilon(t)) = \left( \frac{d\tilde{s}_\epsilon}{dt}(t) \right)^{-1} = \frac{1}{[\epsilon + \|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))}^2]^{1/2}}. \quad (\text{F2})$$

If we change the time variable as  $\tilde{x}(s) := x^*(\tilde{t}_\epsilon(s))$  and  $\tilde{\psi}(s) := (d_s \tilde{t}_\epsilon(s)) \psi^*(\tilde{t}_\epsilon(s))$ , we find

$$\begin{aligned} d_s \tilde{x}(s) &= d_s \tilde{t}_\epsilon(s) d_t x^*(\tilde{t}_\epsilon(s)) = d_s \tilde{t}_\epsilon(s) \text{SL}(\tilde{x}(s)) \mathbb{S}^\top \psi^*(\tilde{t}_\epsilon(s)) \\ &= \text{SL}(\tilde{x}(s)) \mathbb{S}^\top \tilde{\psi}(s), \end{aligned} \quad (\text{F3})$$

so that  $(\tilde{x}, \tilde{\psi})$  falls in the feasible set of the minimization (100) with time duration  $\tilde{s}_\epsilon(\mathcal{T})$ . Therefore, we have

$$\begin{aligned} \mathcal{W}(x^{(0)}, x^{(1)})^2 &\leq \tilde{s}_\epsilon(\mathcal{T}) \int_0^{\tilde{s}_\epsilon(\mathcal{T})} ds \|\mathbb{S}^\top \tilde{\psi}(s)\|_{L(\tilde{x}(s))}^2 \\ &= \tilde{s}_\epsilon(\mathcal{T}) \int_0^\mathcal{T} dt \frac{d\tilde{s}_\epsilon}{dt}(t) \times \left( \frac{d\tilde{t}_\epsilon}{ds}(\tilde{s}_\epsilon(t)) \right)^2 \|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))}^2 \\ &= \tilde{s}_\epsilon(\mathcal{T}) \int_0^\mathcal{T} dt \frac{\|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))}^2}{\epsilon + \|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))}^2} \\ &\quad \times [\epsilon + \|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))}^2]^{1/2} \\ &\leq \tilde{s}_\epsilon(\mathcal{T})^2. \end{aligned} \quad (\text{F4})$$

By taking the limit  $\epsilon \rightarrow 0$ , we obtain the inequality

$$\mathcal{W}(x^{(0)}, x^{(1)}) \leq \tilde{s}_0(\mathcal{T}) = \int_0^\mathcal{T} dt \|\mathbb{S}^\top \psi^*(t)\|_{L(x^*(t))}. \quad (\text{F5})$$

#### APPENDIX G: DERIVATION OF EQ. (105)

We give a proof of the duality formula (105). We first prepare some relations regarding the functional  $\mathcal{I}$  defined in Appendix E. Here, we concentrate on probability distributions, so the argument of the functional is  $(p, f, \psi)$ . We again assume there exists an optimal solution  $(p^*, f^*, \psi^*)$  that provides the Wasserstein distance as

$$\begin{aligned} \frac{1}{2\mathcal{T}} \mathcal{W}(p^{(0)}, p^{(1)})^2 &= \mathcal{I}[p^*, f^*, \psi^*] \\ &= \frac{1}{2} \int_0^\mathcal{T} dt \|B^\top \psi^*(t)\|_{L(p^*(t))}^2, \end{aligned} \quad (\text{G1})$$

where we used the relation  $f^*(t) = B^\top \psi^*(t)$ , which was derived in Appendix E. Consider the functional derivative

$$\begin{aligned} \frac{\delta \mathcal{I}}{\delta p_i(t)} &= \frac{1}{2} \sum_e \frac{\partial \ell_e}{\partial p_i}(p(t)) |f_e(t)|^2 - d_t \psi_i(t) \\ &\quad - \frac{\partial}{\partial p_i} \psi^\top(t) BL(p(t)) f(t). \end{aligned} \quad (\text{G2})$$

It vanishes when  $(p, f, \psi) = (p^*, f^*, \psi^*)$ , so we have

$$d_t \psi_i^*(t) + \frac{1}{2} \frac{\partial}{\partial p_i} \|f^*(t)\|_{L(p^*(t))}^2 = 0. \quad (\text{G3})$$

Combining it with  $f^*(t) = B^\top \psi^*(t)$ , we find the equation

$$d_t \psi_i^*(t) + \frac{1}{2} \frac{\partial}{\partial p_i} \|B^\top \psi^*(t)\|_{L(p^*(t))}^2 = 0. \quad (\text{G4})$$

We next show that  $\psi^*$  gives the distance by the duality formula. Because of the equality for the log mean

$$a \frac{\partial}{\partial a} \mathbb{L}(a, b) + b \frac{\partial}{\partial b} \mathbb{L}(a, b) = \mathbb{L}(a, b) \quad (\text{G5})$$

shown in Ref. [43], we have the relation

$$\begin{aligned} &\sum_i p_i \frac{\partial}{\partial p_i} \ell_e(p) \\ &= W_e p_{i(e)} \frac{\partial}{\partial u_1} \mathbb{L}(W_e p_{i(e)}, W_{-e} p_{i'(e)}) \\ &\quad + W_{-e} p_{i'(e)} \frac{\partial}{\partial u_2} \mathbb{L}(W_e p_{i(e)}, W_{-e} p_{i'(e)}) = \ell_e(p), \end{aligned} \quad (\text{G6})$$

where  $u_1$  and  $u_2$  denote the first and second argument of  $\mathbb{L}$ . Combining it with Eq. (G4), we obtain the Hamilton–Jacobi equation

$$p^*(t)^\top d_t \psi^*(t) + \frac{1}{2} \|B^\top \psi^*(t)\|_{L(p^*(t))}^2 = 0. \quad (\text{G7})$$

Then, the Wasserstein distance becomes

$$\begin{aligned} &\frac{1}{\mathcal{T}} \mathcal{W}(p^{(0)}, p^{(1)})^2 \\ &= \int_0^\mathcal{T} dt \|B^\top \psi^*(t)\|_{L(p^*(t))}^2 = \int_0^\mathcal{T} dt \psi^*(t)^\top d_t p^*(t) \\ &= \langle \psi^*(\mathcal{T}) \rangle_1 - \langle \psi^*(0) \rangle_0 - \int_0^\mathcal{T} dt [d_t \psi^*(t)]^\top p^*(t) \\ &= \langle \psi^*(\mathcal{T}) \rangle_1 - \langle \psi^*(0) \rangle_0 + \frac{1}{2} \int_0^\mathcal{T} dt \|B^\top \psi^*(t)\|_{L(p^*(t))}^2, \end{aligned} \quad (\text{G8})$$

where we used the continuity equation  $d_t p^* = BL(p) B^\top \psi^*$  in the second equality of the second line, and we did the integration by parts in the second line. Recall that  $\langle \cdot \rangle_i$  indicates the expectation value under  $p^{(i)}$ . Therefore, we have the formula

$$\frac{1}{2\mathcal{T}} \mathcal{W}(p^{(0)}, p^{(1)})^2 = \langle \psi^*(\mathcal{T}) \rangle_1 - \langle \psi^*(0) \rangle_0. \quad (\text{G9})$$

Finally, we prove that  $\psi^*$  provides the maximum under the condition in Eq. (106) [also shown in Eq. (G13) below]. To show that  $\psi^*$  satisfies the condition, let  $g(p) := p^\top d_t \psi^*(t) + (1/2) \|B^\top \psi^*(t)\|_{L(p)}^2$ . Then,

$$\frac{\partial g}{\partial p_i}(p^*) = d_t \psi_i^* + \frac{1}{2} \frac{\partial}{\partial p_i} \|B^\top \psi^*\|_{L(p^*)}^2, \quad (\text{G10})$$

and it is zero because  $\psi^*$  satisfies Eq. (G4). As we will show later, since  $\ell_e(p)$  is a concave function,  $g(p)$  is also concave, so  $p^*$  gives the maximum value of  $g(p)$  [69]. Therefore, we conclude that  $g(p) \leq 0$ , and  $\psi^*$  satisfies the condition. Let  $\psi'$  also satisfy the condition. Then,

$$\begin{aligned} &\langle \psi'(\mathcal{T}) \rangle_1 - \langle \psi'(0) \rangle_0 \\ &= \int_0^\mathcal{T} dt d_t [(p^*)^\top \psi'] \\ &= \int_0^\mathcal{T} dt ((\psi')^\top BL(p^*) B^\top \psi^* + (p^*)^\top d_t \psi') \\ &\leq \int_0^\mathcal{T} dt \left( \langle B^\top \psi', B^\top \psi^* \rangle_{L(p^*)} - \frac{1}{2} \|B^\top \psi'\|_{L(p^*)}^2 \right) \\ &= \int_0^\mathcal{T} dt \left( \frac{1}{2} \|B^\top \psi^*\|_{L(p^*)}^2 - \frac{1}{2} \|B^\top (\psi^* - \psi')\|_{L(p^*)}^2 \right) \\ &\leq \frac{1}{2} \int_0^\mathcal{T} dt \|B^\top \psi^*\|_{L(p^*)}^2 = \langle \psi^*(\mathcal{T}) \rangle_1 - \langle \psi^*(0) \rangle_0, \end{aligned} \quad (\text{G11})$$

where the fourth line follows from the condition (106), and the fifth line is derived from a general property of an inner product. Therefore, we obtain the expression

$$\frac{1}{2\mathcal{T}} \mathcal{W}(p^{(0)}, p^{(1)})^2 = \sup_{\psi} (\langle \psi(\mathcal{T}) \rangle_1 - \langle \psi(0) \rangle_0) \quad (\text{G12})$$

$$\text{s.t. } q^\top d_t \psi(t) + \frac{1}{2} \|B^\top \psi(t)\|_{L(q)}^2 \leq 0$$

$$\text{for any probability distribution } q. \quad (\text{G13})$$

The concavity of  $\ell_e(p)$  follows from the concavity of the log mean because the map  $p \mapsto (W_e p_{i(e)}, W_{-e} p_{i'(e)})$  is linear. Let us prove that the log mean is concave. The log mean has an integral form

$$\mathbb{L}(x, y) = \int_0^1 dt x^t y^{1-t}. \quad (\text{G14})$$

Then, the Hessian  $H$  is given as

$$H_{xx} = \frac{\partial^2}{\partial x^2} \mathbb{L}(x, y) = \int_0^1 dt t(t-1) x^{t-2} y^{1-t}, \quad (\text{G15})$$

$$H_{yy} = \frac{\partial^2}{\partial y^2} \mathbb{L}(x, y) = \int_0^1 dt t(t-1) x^t y^{-1-t}, \quad (\text{G16})$$

$$H_{xy} = H_{yx} = \frac{\partial^2}{\partial x \partial y} \mathbb{L}(x, y) = \int_0^1 dt t(1-t) x^{t-1} y^{-t}. \quad (\text{G17})$$

The determinant of the Hessian is positive because

$$\begin{aligned} \det H &= H_{xx}H_{yy} - H_{xy}^2 \\ &= \int_0^1 dt t(t-1)x^{t-2}y^{1-t} \int_0^1 dt t(t-1)x^t y^{-1-t} \\ &\quad - \left( \int_0^1 dt t(t-1)x^{t-1}y^{-t} \right)^2 \\ &= \int_0^1 dt t(1-t)x^{t-2}y^{1-t} \int_0^1 dt t(1-t)x^t y^{-1-t} \\ &\quad - \left( \int_0^1 dt \sqrt{t(1-t)x^{t-2}y^{-1-t}} \sqrt{t(1-t)x^t y^{-1-t}} \right)^2, \\ &\geq 0, \end{aligned} \tag{G18}$$

where we used the Cauchy–Schwarz inequality in the last line. The trace is shown to be negative because  $t(t-1) \leq 0$  and  $x, y$  are positive by definition, so  $H_{xx}$  and  $H_{yy}$  are negative. Therefore, the Hessian  $H$  is negative-semidefinite and the log mean is concave.

#### APPENDIX H: DERIVATION OF EQ. (120)

In this Appendix, we provide the proof of the inequality in Eq. (120). First, we characterize the  $L^2$ -Wasserstein distance by a functional slightly different from what we used in the previous Appendixes. Next, we show relations between the  $L^2$ - and  $L^1$ -Wasserstein distance by using the derived expression.

Let  $\mathcal{F}$  be the functional,

$$\mathcal{F}[q, \psi] = \int_0^\tau dt \left( -\|B^\top \psi(t)\|_{L(q(t))}^2 + 2\psi(t)^\top d_t q(t) \right). \tag{H1}$$

Completing the square shows that this functional is maximized with respect to  $\psi$  when  $d_t q(t) = BL(q(t))B^\top \psi(t)$ , and then

$$\mathcal{F}'[q] := \sup_\psi \mathcal{F}[q, \psi] = \int_0^\tau dt \|B^\top \psi(t)\|_{L(q(t))}^2. \tag{H2}$$

If we choose a solution of the master equation  $\hat{p}$  as  $q$ , this amounts to the Onsager excess EP,  $\mathcal{F}'[\hat{p}] = \int_0^\tau dt \sigma^{\text{ex}}$ . The  $L^2$ -Wasserstein distance is given by

$$\mathcal{W}_2(\hat{p}(0), \hat{p}(\tau))^2 = \tau \inf_q \sup_\psi \mathcal{F}[q, \psi] \tag{H3}$$

with  $q$  satisfying the initial and final conditions,  $q(0) = \hat{p}(0)$  and  $q(\tau) = \hat{p}(\tau)$ . We write the minimizer as  $p^*$ , so that  $\mathcal{W}_2(\hat{p}(0), \hat{p}(\tau))^2 = \tau \mathcal{F}'[p^*]$ .

Let  $(q, \psi)$  satisfy the continuity equation  $d_t q = BL(q)B^\top \psi$ , so that  $\mathcal{F}'[q] = \mathcal{F}[q, \psi]$ . We now do not assume  $q$  is a solution of the master equation. The Cauchy–Schwarz inequality leads to

$$\begin{aligned} &\left( \int_0^\tau dt \varphi(t)^\top d_t q(t) \right)^2 \\ &= \left( \int_0^\tau dt \langle B^\top \varphi(t), B^\top \psi(t) \rangle_{L(q(t))} \right)^2 \\ &\leq \int_0^\tau dt \|B^\top \varphi(t)\|_{L(q(t))}^2 \int_0^\tau dt \|B^\top \psi(t)\|_{L(q(t))}^2 \end{aligned} \tag{H4}$$

for an arbitrary  $\varphi$ . Hence, by choosing  $\varphi(t) = \xi(t)$ , the potential of the duality in Eq. (119) between  $q(t)$  and  $q(t+dt)$  with  $dt$  an infinitesimal time interval, we have

$$\mathcal{F}'[q] \geq \frac{\left( \int_0^\tau dt \xi(t)^\top d_t q(t) \right)^2}{\int_0^\tau dt \|B^\top \xi(t)\|_{L(q(t))}^2}. \tag{H5}$$

According to the duality (119), if we assume  $q(0) = \hat{p}(0)$  and  $q(\tau) = \hat{p}(\tau)$ , the numerator reads

$$\begin{aligned} \int_0^\tau dt \xi(t)^\top d_t q(t) &= \int_0^\tau dt \lim_{\Delta t \rightarrow 0} \frac{\mathcal{W}_1(q(t), q(t+\Delta t))}{\Delta t} \\ &\geq \mathcal{W}_1(\hat{p}(0), \hat{p}(\tau)), \end{aligned} \tag{H6}$$

where we used the triangle inequality in the last inequality. The denominator is also bounded as

$$\begin{aligned} \|B^\top \xi(t)\|_{L(q(t))}^2 &= \sum_e \ell_e(q(t)) (\xi_{i(e)} - \xi_{i'(e)})^2 \leq \sum_e \ell_e(q(t)) \\ &\leq \frac{1}{2} \sum_e \chi_e(q(t)), \end{aligned} \tag{H7}$$

where we used  $|\xi_i - \xi_j| \leq 1$ . Recall that  $\chi_e(q(t)) = k_e q_{i(e)}(t) + k_{-e} q_{i'(e)}(t)$  is the edgewise dynamical activity and the log mean inequality  $\ell_e(q(t)) \leq (1/2)\chi_e(q(t))$ . Therefore, we obtain

$$\mathcal{F}'[q] \geq \frac{\mathcal{W}_1(\hat{p}(0), \hat{p}(\tau))^2}{\tau \bar{\mathcal{A}}[q]/2} \tag{H8}$$

for an arbitrary  $q$  with the appropriate initial and final conditions. If we choose  $q = p^*$ , the minimizer of  $\mathcal{F}'$  with the initial and final states  $(p(0), p(\tau))$ , we have the desired inequality

$$\frac{\bar{\mathcal{A}}[p^*]}{2} \mathcal{W}_2(\hat{p}(0), \hat{p}(\tau))^2 \geq \mathcal{W}_1(\hat{p}(0), \hat{p}(\tau))^2. \tag{H9}$$

We also find a lower bound for the Onsager excess EP as

$$\int_0^\tau dt \sigma^{\text{ex}} \geq \frac{\mathcal{W}_1(\hat{p}(0), \hat{p}(\tau))^2}{\tau \bar{\mathcal{A}}[\hat{p}]/2}. \tag{H10}$$

[1] Y. Oono and M. Paniconi, Steady state thermodynamics, *Prog. Theor. Phys. Suppl.* **130**, 29 (1998).  
 [2] T. Hatano and S.-i. Sasa, Steady-State Thermodynamics of Langevin Systems, *Phys. Rev. Lett.* **86**, 3463 (2001).

[3] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, *Phys. Rev. E* **82**, 011143 (2010).  
 [4] C. Maes and K. Netočný, A nonequilibrium extension of the Clausius heat theorem, *J. Stat. Phys.* **154**, 188 (2014).

- [5] R. Rao and M. Esposito, Nonequilibrium Thermodynamics of Chemical Reaction Networks: Wisdom from Stochastic Thermodynamics, *Phys. Rev. X* **6**, 041064 (2016).
- [6] H. Ge and H. Qian, Nonequilibrium thermodynamic formalism of nonlinear chemical reaction systems with Waage–Guldberg’s law of mass action, *Chem. Phys.* **472**, 241 (2016).
- [7] More precisely, systems with complex balanced steady states are globally stable within each stoichiometric compatibility class [70–72].
- [8] M. Feinberg, *Foundations of Chemical Reaction Network Theory* (Springer, Cham, 2019).
- [9] D. A. Beard and H. Qian, *Chemical Biophysics: Quantitative Analysis of Cellular Systems* (Cambridge University Press, Cambridge, 2008), Vol. 126.
- [10] J. Schnakenberg, Network theory of microscopic and macroscopic behavior of master equation systems, *Rev. Mod. Phys.* **48**, 571 (1976).
- [11] A. Mielke, A gradient structure for reaction–diffusion systems and for energy-drift-diffusion systems, *Nonlinearity* **24**, 1329 (2011).
- [12] A. Mielke, Geodesic convexity of the relative entropy in reversible Markov chains, *Calc. Var. Partial Diff. Eq.* **48**, 1 (2013).
- [13] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Dissipation Bounds All Steady-State Current Fluctuations, *Phys. Rev. Lett.* **116**, 120601 (2016).
- [14] P. Pietzonka, A. C. Barato, and U. Seifert, Universal bound on the efficiency of molecular motors, *J. Stat. Mech.* (2016) 124004.
- [15] J. M. Horowitz and T. R. Gingrich, Proof of the finite-time thermodynamic uncertainty relation for steady-state currents, *Phys. Rev. E* **96**, 020103(R) (2017).
- [16] K. Proesmans and C. Van den Broeck, Discrete-time thermodynamic uncertainty relation, *Europhys. Lett.* **119**, 20001 (2017).
- [17] P. Pietzonka, F. Ritort, and U. Seifert, Finite-time generalization of the thermodynamic uncertainty relation, *Phys. Rev. E* **96**, 012101 (2017).
- [18] A. Dechant and S.-i. Sasa, Current fluctuations and transport efficiency for general Langevin systems, *J. Stat. Mech.* (2018) 063209.
- [19] A. Dechant, Multidimensional thermodynamic uncertainty relations, *J. Phys. A* **52**, 035001 (2019).
- [20] Y. Hasegawa and T. Van Vu, Fluctuation Theorem Uncertainty Relation, *Phys. Rev. Lett.* **123**, 110602 (2019).
- [21] G. Falasco, M. Esposito, and J.-C. Delvenne, Unifying thermodynamic uncertainty relations, *New J. Phys.* **22**, 053046 (2020).
- [22] D. H. Wolpert, Uncertainty Relations and Fluctuation Theorems for Bayes Nets, *Phys. Rev. Lett.* **125**, 200602 (2020).
- [23] S. Otsubo, S. Ito, A. Dechant, and T. Sagawa, Estimating entropy production by machine learning of short-time fluctuating currents, *Phys. Rev. E* **101**, 062106 (2020).
- [24] S. K. Manikandan, D. Gupta, and S. Krishnamurthy, Inferring Entropy Production from Short Experiments, *Phys. Rev. Lett.* **124**, 120603 (2020).
- [25] K. Liu, Z. Gong, and M. Ueda, Thermodynamic Uncertainty Relation for Arbitrary Initial States, *Phys. Rev. Lett.* **125**, 140602 (2020).
- [26] A. Dechant and S.-i. Sasa, Fluctuation–response inequality out of equilibrium, *Proc. Natl. Acad. Sci. (USA)* **117**, 6430 (2020).
- [27] A. Dechant and S.-i. Sasa, Continuous time reversal and equality in the thermodynamic uncertainty relation, *Phys. Rev. Res.* **3**, L042012 (2021).
- [28] A. Dechant, S.-i. Sasa, and S. Ito, Geometric decomposition of entropy production in out-of-equilibrium systems, *Phys. Rev. Res.* **4**, L012034 (2022).
- [29] A. Dechant, S.-i. Sasa, and S. Ito, Geometric decomposition of entropy production into excess, housekeeping and coupling parts, *Phys. Rev. E* **106**, 024125 (2022).
- [30] K. Yoshimura and S. Ito, Thermodynamic Uncertainty Relation and Thermodynamic Speed Limit in Deterministic Chemical Reaction Networks, *Phys. Rev. Lett.* **127**, 160601 (2021).
- [31] T. Van Vu and Y. Hasegawa, Geometrical Bounds of the Irreversibility in Markovian Systems, *Phys. Rev. Lett.* **126**, 010601 (2021).
- [32] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, Heidelberg, 2009), Vol. 338.
- [33] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, Optimal Protocols and Optimal Transport in Stochastic Thermodynamics, *Phys. Rev. Lett.* **106**, 250601 (2011).
- [34] E. Aurell, K. Gawędzki, C. Mejía-Monasterio, R. Mohayae, and P. Muratore-Ginanneschi, Refined second law of thermodynamics for fast random processes, *J. Stat. Phys.* **147**, 487 (2012).
- [35] S. Bo, E. Aurell, R. Eichhorn, and A. Celani, Optimal stochastic transport in inhomogeneous thermal environments, *Europhys. Lett.* **103**, 10010 (2013).
- [36] A. Dechant and Y. Sakurai, Thermodynamic interpretation of Wasserstein distance, *arXiv:1912.08405* (2019).
- [37] Y. Chen, T. T. Georgiou, and A. Tannenbaum, Stochastic control and nonequilibrium thermodynamics: Fundamental limits, *IEEE Trans. Automat. Contr.* **65**, 2979 (2019).
- [38] R. Fu, A. Taghvaei, Y. Chen, and T. T. Georgiou, Maximal power output of a stochastic thermodynamic engine, *Automatica* **123**, 109366 (2021).
- [39] M. Nakazato and S. Ito, Geometrical aspects of entropy production in stochastic thermodynamics based on Wasserstein distance, *Phys. Rev. Res.* **3**, 043093 (2021).
- [40] O. M. Miangolarra, A. Taghvaei, Y. Chen, and T. T. Georgiou, Geometry of finite-time thermodynamic cycles with anisotropic thermal fluctuations, *arXiv:2203.12483* (2022).
- [41] V. T. Vo, T. Van Vu, and Y. Hasegawa, Unified approach to classical speed limit and thermodynamic uncertainty relation, *Phys. Rev. E* **102**, 062132 (2020).
- [42] J. Maas, Gradient flows of the entropy for finite Markov chains, *J. Funct. Anal.* **261**, 2250 (2011).
- [43] M. Erbar and J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy, *Arch. Ration. Mech. Anal.* **206**, 997 (2012).
- [44] M. Erbar, J. Maas, and M. Wirth, On the geometry of geodesics in discrete optimal transport, *Calc. Var. Partial Diff. Eq.* **58**, 19 (2019).
- [45] F. Avanzini, E. Penocchio, G. Falasco, and M. Esposito, Nonequilibrium thermodynamics of non-ideal chemical reaction networks, *J. Chem. Phys.* **154**, 094114 (2021).
- [46] M. Poletti and M. Esposito, Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws, *J. Chem. Phys.* **141**, 024117 (2014).
- [47] D. Andrieux and P. Gaspard, Fluctuation theorem for currents and Schnakenberg network theory, *J. Stat. Phys.* **127**, 107 (2007).

- [48] J. M. Horowitz and M. Esposito, Thermodynamics with Continuous Information Flow, *Phys. Rev. X* **4**, 031015 (2014).
- [49] J. van der Meer, B. Ertel, and U. Seifert, Thermodynamic Inference in Partially Accessible Markov Networks: A Unifying Perspective from Transition-Based Waiting Time Distributions, *Phys. Rev. X* **12**, 031025 (2022).
- [50] We note that in a withdrawn preprint [73], Ge and coauthors also found this characterization of the EPR (14) as the squared norm of the thermodynamic force with the Onsager coefficient being the metric. They tried to express the HS decomposition as an orthogonal decomposition, but they discovered an error in their proof and withdrew their preprint. In Sec. III F, we discuss that, in general, the HS decomposition is not given geometrically.
- [51] C. Godsil and G. F. Royle, *Algebraic Graph Theory* (Springer Science & Business Media, New York, 2001), Vol. 207.
- [52] Let  $\{p^*(t)\}_{t \in [0,1]}$  and  $\{\psi^*(t)\}_{t \in [0,1]}$  give the Wasserstein distance  $\mathcal{W}(p^{(0)}, p^{(1)})$ . Then for any  $\mathcal{T} > 0$ ,  $p'(t') := p^*(t'/\mathcal{T})$  and  $\psi'(t') := \psi^*(t'/\mathcal{T})/\mathcal{T}$  satisfy  $d_{t'} p'(t') = BL(p'(t'))B^T \psi'(t')$  and provide  $\mathcal{W}(p^{(0)}, p^{(1)})/\mathcal{T}$  by integrating  $\int_0^{\mathcal{T}} dt' \|B^T \psi'(t')\|_{L(p'(t'))}^2$ . This discussion is valid for more general settings like Eqs. (96) and (98).
- [53] S. Puntanen, G. P. Styan, and J. Isotalo, *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty* (Springer, Berlin, Heidelberg, 2011).
- [54] It is shown under the assumption that the dynamics obeys the law of mass action by seeing the equivalence between the following two conditions. One is that the force is conservative, which is equivalent to  $\sigma^{\text{hk}} = 0$ . The other is the so-called generalized Wegscheider condition  $\ln(k^+/k^-) \in \text{im } \mathbb{S}^T$ , which is a necessary and sufficient condition for the steady state to be detailed balanced [74].
- [55] A. Dechant, Minimum entropy production, detailed balance and Wasserstein distance for continuous-time Markov processes, *J. Phys. A* **55**, 094001 (2022).
- [56] B. Remlein and U. Seifert, Optimality of nonconservative driving for finite-time processes with discrete states, *Phys. Rev. E* **103**, L050105 (2021).
- [57] E. Ilker, Ö. Güngör, B. Kuznets-Speck, J. Chiel, S. Deffner, and M. Hinczewski, Shortcuts in Stochastic Systems and Control of Biophysical Processes, *Phys. Rev. X* **12**, 021048 (2022).
- [58] T. Van Vu and K. Saito, Thermodynamic unification of optimal transport: Thermodynamic uncertainty relation, minimum dissipation, and thermodynamic speed limits, [arXiv:2206.02684](https://arxiv.org/abs/2206.02684) (2022).
- [59] This relation was proved for one-dimensional systems in [58]. In Appendix B, we generalized it to multidimensional cases.
- [60] C. Maes and K. Netočný, Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states, *Europhys. Lett.* **82**, 30003 (2008).
- [61] T. Koyuk and U. Seifert, Thermodynamic Uncertainty Relation for Time-Dependent Driving, *Phys. Rev. Lett.* **125**, 260604 (2020).
- [62] J. Dolbeault, B. Nazaret, and G. Savaré, A new class of transport distances between measures, *Calc. Var. Partial Diff. Eq.* **34**, 193 (2009).
- [63] N. Shiraishi, Optimal thermodynamic uncertainty relation in Markov jump processes, *J. Stat. Phys.* **185**, 19 (2021).
- [64] S.-N. Chow, W. Huang, Y. Li, and H. Zhou, Fokker-Planck equations for a free energy functional or Markov process on a graph, *Arch. Ration. Mech. Anal.* **203**, 969 (2012).
- [65] P. Muratore-Ginanneschi, C. Mejía-Monasterio, and L. Peliti, Heat release by controlled continuous-time Markov jump processes, *J. Stat. Phys.* **150**, 181 (2013).
- [66] A. Kolchinsky, A. Dechant, K. Yoshimura, and S. Ito, Information geometry of excess and housekeeping entropy production, [arXiv:2206.14599](https://arxiv.org/abs/2206.14599) (2022).
- [67] C. Van den Broeck and M. Esposito, Three faces of the second law. II. Fokker-Planck formulation, *Phys. Rev. E* **82**, 011144 (2010).
- [69] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis* (Springer Science & Business Media, Berlin, Heidelberg, 2009), Vol. 317.
- [70] F. Horn and R. Jackson, General mass action kinetics, *Arch. Ration. Mech. Anal.* **47**, 81 (1972).
- [71] G. Craciun, Toric differential inclusions and a proof of the global attractor conjecture, [arXiv:1501.02860](https://arxiv.org/abs/1501.02860) (2015).
- [72] D. F. Anderson, A proof of the global attractor conjecture in the single linkage class case, *SIAM J. Appl. Math.* **71**, 1487 (2011).
- [73] H. Ge, W. H. Kim, and H. Qian, Thermodynamics and geometry of reversible and irreversible Markov processes, [arXiv:1108.4055](https://arxiv.org/abs/1108.4055) (2011).
- [74] S. Schuster and R. Schuster, A generalization of Wegscheider's condition. Implications for properties of steady states and for quasi-steady-state approximation, *J. Math. Chem.* **3**, 25 (1989).