Thermodynamic Unification of Optimal Transport: Thermodynamic Uncertainty Relation, Minimum Dissipation, and Thermodynamic Speed Limits

Tan Van Vu[*] and Keiji Saito[†]

Department of Physics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

Thermodynamics serves as a universal means for studying physical systems from an energy perspective. In recent years, with the establishment of the field of stochastic and quantum thermodynamics, the ideas of thermodynamics have been generalized to small fluctuating systems. Independently developed in mathematics and statistics, the optimal transport theory concerns the means by which one can optimally transport a source distribution to a target distribution, deriving a useful metric between probability distributions, called the Wasserstein distance. Despite their seemingly unrelated nature, an intimate connection between these fields has been unveiled in the context of continuous-state Langevin dynamics, providing several important implications for nonequilibrium systems. In this study, we elucidate an analogous connection for discrete cases by developing a thermodynamic framework for discrete optimal transport. We first introduce a novel quantity called dynamical state mobility, which significantly improves the thermodynamic uncertainty relation and provides insights into the precision of currents in non-equilibrium Markov jump processes. We then derive variational formulas that connect the discrete Wasserstein distances to stochastic and quantum thermodynamics of discrete Markovian dynamics described by master equations. Specifically, we rigorously prove that the Wasserstein distance equals the minimum product of irreversible entropy production and dynamical state mobility over all admissible Markovian dynamics. These formulas not only unify the relationship between thermodynamics and the optimal transport theory for discrete and continuous cases, but also generalize it to the quantum case. In addition, we demonstrate that the obtained variational formulas lead to remarkable applications in stochastic and quantum thermodynamics, such as stringent thermodynamic speed limits and the finite-time Landauer principle. These bounds are tight and can be saturated for arbitrary temperatures, even in the zero-temperature limit. Notably, the finite-time Landauer principle can explain finite dissipation even at extremely low temperatures, which cannot be explained by the conventional Landauer principle.

I. INTRODUCTION

A. Background

Thermodynamics, which is built upon several axioms, is one of the most successful phenomenological theories for studying energy exchanges in macroscopic systems. Originally developed for the purpose of understanding the behavior of steam engines, thermodynamics has since been applied to various fields of science and engineering. The laws of thermodynamics show extraordinary universality and impose fundamental constraints on physical systems.

[*]tanvu@rk.phys.keio.ac.jp

[†]saitoh@rk.phys.keio.ac.jp

Beyond the macroscopic regime, the past two decades have witnessed substantial progress in extending the notions of conventional thermodynamics to microscopic systems, resulting in the frameworks of stochastic and quantum thermodynamics [1–5]. These comprehensive frameworks provide a means of investigating small nonequilibrium systems subject to significant fluctuations. Various universal relations have been discovered, including fluctuation theorems [6–11], thermodynamic uncertainty relations [12–15], thermodynamic speed limits [16–26], and refinements of the Landauer principle [27–32]. These equalities and inequalities characterize the fundamental limits of small systems and distinguish the possible from the impossible in terms of thermodynamics. They not only are theoretically important, but also lead to practical applications in estimating physically relevant quantities from experimental data, such as free energy [33] and dissipation [34–40]. In addition, information manipulations such as measurement, feedback, erasure, and copying have been incorporated into thermodynamics,

leading to significant developments in several subfields, such as the thermodynamics of information [41,42] and computation [43,44]. In parallel, concepts from other fields, such as the resource theory [45,46] and information geometry [47–49], have also been integrated into thermodynamics, generating new avenues of research and offering new tools for analyzing thermodynamic processes [50–58]. Accordingly, the integration of thermodynamics with other disciplines provides new insights into the understanding of nonequilibrium systems.

A key quantity in thermodynamics is entropy production, which quantifies the degree of irreversibility of thermodynamic processes. Entropy production plays a central role in the fundamental laws of thermodynamics and provides a quantitative characterization for investigating nonequilibrium processes; a comprehensive review regarding entropy production can be found in Ref. [59]. Recently, it has been shown that entropy production must be increased to achieve a high precision of currents [13] and fast state transformation [16]. However, minimizing entropy production is also a particularly relevant issue [60,61], because it is closely related to the energy lost to the environment. Owing to the critical role of entropy production, great efforts have been made to elucidate its properties and its relationship with other physical quantities [62–68].

Optimal transport [69], which is developed independently of thermodynamics, is a mature field in mathematics and statistics, and its theory concerns the optimal planning and optimal cost of transporting a distribution. Specifically, given the individual costs of transporting a unit weight of a resource from one location to another, the optimal transport problem is to determine the optimal means of redistributing the distribution of the resource into the desired distribution to yield the lowest total cost. Historically, the optimal transport problem was first defined by Monge in 1781 and has since been reformulated in a more general and well-defined form. Currently, this problem has several theoretical and practical applications in a variety of scientific fields, including statistics and machine learning [70], computer vision [71], linguistics [72], classical mechanics [73], and molecular biology [74]. It is noteworthy that the solution to this problem not only provides an optimal transport plan between distributions, but also defines a useful metric in the space of probability distributions. Although this metric is identified by several names in the literature, such as the Monge-Kantorovich distance or earth mover's distance, we refer to it as the Wasserstein distance throughout this paper.

Because the optimal transport theory is concerned with transformations of probability distributions that are commonly used to characterize the state of small systems, whether any connection exists between the two disciplines of optimal transport and stochastic thermodynamics is a natural question. Indeed, a deep connection between these fields has been elucidated in the context of overdamped Langevin dynamics, revealing that the problem of

minimizing entropy production can be mapped to the optimal transport problem [27,61,75–78]. More specifically, the minimum entropy production among all processes that transform the initial into the final distribution can be expressed in terms of the Wasserstein distance between the two distributions. In addition, the optimal transport plan provides a feasible solution for the optimal control protocol. The essence of this connection can be intuitively understood through the Benamou-Brenier formula [79], which is given by the following equality:

$$W_2(p^A, p^B) = \min \int_0^\tau \sqrt{D\sigma_t}\, dt = \min \sqrt{D\tau\Sigma_\tau}. \quad (1)$$

Here, $W_2(p^A, p^B)$ is the L^2-Wasserstein distance [cf. Eq. (10)], D is the diffusion coefficient, τ is the operational time, σ_t is the entropy production rate, Σ_τ is the total entropy production, and the minimum is taken over all the overdamped Langevin processes that transform distribution p^A into p^B. The variational relation (1) links two apparently unrelated quantities, namely, a mathematical metric and a thermodynamic cost. The metric $W_2(p^A, p^B)$ on the left-hand side of Eq. (1) is a static quantity that is determined only by two distributions, whereas the right-hand side represents a dynamical quantity that indicates the thermodynamic cost associated with overdamped Langevin dynamics. This formula leads to remarkable applications for overdamped Langevin dynamics, such as a thermodynamic speed limit [77] and a finite-time Landauer principle of information erasure for classical bits modeled by a continuous double-well potential [29,80], to name only two. The finite-time correction in the Landauer principle indicated by the speed limit expression is consistent with experimental observations [81]. Moreover, the bounds obtained from Eq. (1) are tight in the sense that, for any two given distributions, there always exists an overdamped Langevin dynamics that transforms the distributions and attains the equality of the bounds.

In contrast to continuous systems, a similar connection between optimal transport and thermodynamics is yet to be unveiled in discrete systems. We note that stochastic thermodynamics of discrete systems is highly relevant to experiments [82–85]. Even in continuous systems such as biological systems, the dynamics can be represented by effective discrete states [86–89]. In addition, the Landauer principle of information erasure problem is a statement for discrete bit operations [90]. Therefore, elucidating the thermodynamic interpretation of the discrete optimal transport problem is essential for an in-depth understanding of the nonequilibrium thermodynamic structure and, particularly, for its application to the state transformation speed.

To reveal this type of relationship for discrete systems, two nontrivial points are worth noting. First, the formula (1) cannot be extended directly to discrete cases, because no

exact correspondence to the diffusion coefficient exists in generic discrete systems. Even if a proper correspondence to the diffusion coefficient is defined for discrete cases, no guarantee can be given that the discrete Wasserstein distance can be expressed in the same manner as in Eq. (1). Second, previous studies show that, without any additional constraint, the distribution of Markov jump processes can always be transformed to the target distribution with arbitrarily small entropy production [91–93]. This implies that entropy production alone is insufficient to characterize the transport cost (i.e., the Wasserstein distance), or, equivalently, this implies that another quantity that plays the same role as the diffusion coefficient in continuous cases must be introduced along with entropy production. These technical remarks are an obstacle to elucidating the relationship between optimal transport and thermodynamics in discrete cases. Simultaneously, overcoming this obstacle is expected to reveal essential and common thermodynamic structures hidden in nonequilibrium processes.

With this background, we aim to elucidate the deep connection between thermodynamics and optimal transport in discrete cases (see Fig. 1 for illustration). Specifically, we develop discrete generalizations of the Benamou-Brenier formula in the context of Markovian open classical and quantum dynamics described by the master equations. Our formulas not only unify the relationship between optimal transport and stochastic thermodynamics for discrete and continuous cases, but also generalize to the quantum case. Moreover, by developing a thermodynamic framework for discrete optimal transport, we can derive fundamental bounds for nonequilibrium systems, including the thermodynamic uncertainty relation, thermodynamic speed limits, and finite-time Landauer principle for both classical and quantum systems. These bounds are tight and stronger

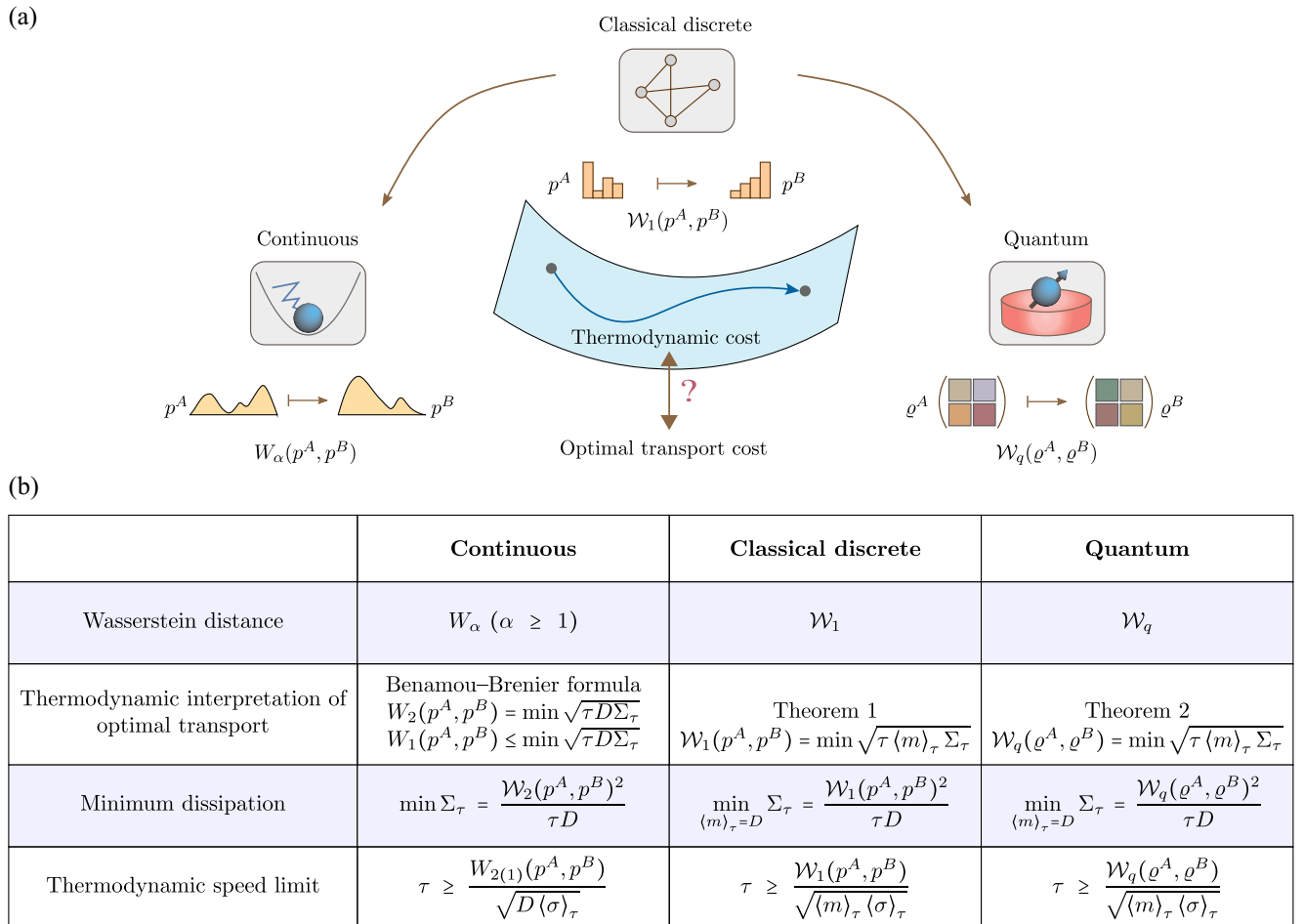|	Continuous	Classical discrete	Quantum
Wasserstein distance	W_α $(\alpha \geq 1)$	$\mathcal{W}_1$	$\mathcal{W}_q$
Thermodynamic interpretation of optimal transport	Benamou–Brenier formula $W_2(p^A, p^B) = \min \sqrt{\tau D \Sigma_\tau}$ $W_1(p^A, p^B) \leq \min \sqrt{\tau D \Sigma_\tau}$	Theorem 1 $\mathcal{W}_1(p^A, p^B) = \min \sqrt{\tau \langle m \rangle_\tau \Sigma_\tau}$	Theorem 2 $\mathcal{W}_q(\varrho^A, \varrho^B) = \min \sqrt{\tau \langle m \rangle_\tau \Sigma_\tau}$
Minimum dissipation	$\min \Sigma_\tau = \dfrac{W_2(p^A, p^B)^2}{\tau D}$	$\min\limits_{\langle m \rangle_\tau = D} \Sigma_\tau = \dfrac{\mathcal{W}_1(p^A, p^B)^2}{\tau D}$	$\min\limits_{\langle m \rangle_\tau = D} \Sigma_\tau = \dfrac{\mathcal{W}_q(\varrho^A, \varrho^B)^2}{\tau D}$
Thermodynamic speed limit	$\tau \geq \dfrac{W_{2(1)}(p^A, p^B)}{\sqrt{D \langle \sigma \rangle_\tau}}$	$\tau \geq \dfrac{\mathcal{W}_1(p^A, p^B)}{\sqrt{\langle m \rangle_\tau \langle \sigma \rangle_\tau}}$	$\tau \geq \dfrac{\mathcal{W}_q(\varrho^A, \varrho^B)}{\sqrt{\langle m \rangle_\tau \langle \sigma \rangle_\tau}}$

FIG. 1. (a) Schematic of stochastic and quantum thermodynamics of discrete Markovian dynamics and the optimal transport problem. This study aims to reveal the connection between thermodynamics and the optimal transport theory for discrete cases, including both classical and quantum systems. (b) Our thermodynamic unification of optimal transport and its consequences, including minimum dissipation and thermodynamic speed limits. By introducing dynamical state mobility m_t, we provide a unified, discrete generalization of the Benamou-Brenier formula for classical and optimal discrete systems. See Sec. I B for an explanation.

than previously reported results. Concerning the Landauer principle, finite-time information erasure should generate finite heat dissipation even at zero temperature. However, neither the original Landauer bound [90] nor finite-time corrections that have been proposed thus far for discrete systems [30,31] can explain heat dissipation at extremely low temperatures. By contrast, our expression for the first time can predict heat dissipation even at extremely low temperatures.

B. Summary of results

The main contributions of this study can be summarized as follows.

(i) *Dynamical state mobility and improved thermodynamic uncertainty relation.*—We define a novel kinetic quantity m_t [cf. Eq. (39)], which is essential to our results. The motivation for this definition is derived from Eq. (1), which suggests that a kinetic term is relevant for characterizing the Wasserstein distance. m_t is defined by the microscopic Onsager-like coefficients and reduces to the diffusion coefficient D in the continuous limit; thus, it is referred to as dynamical state mobility. Similar to dynamical activity [94], m_t should be measurable in experiments. Using this kinetic term, we derive an improved thermodynamic uncertainty relation for time-integrated currents in Markov jump processes, which can be expressed as [cf. Eq. (61)]

$$\frac{\langle J \rangle^2}{\operatorname{var}[J]} \leq \eta \frac{\Sigma_\tau}{2}, \tag{2}$$

where $\langle J \rangle$ and $\operatorname{var}[J]$ denote the mean and variance of an arbitrary current J, respectively, and $\eta := 2\mathcal{M}_\tau / \mathcal{A}_\tau \leq 1$ is an efficiency defined in terms of dynamical state mobility $\mathcal{M}_\tau := \int_0^\tau m_t dt$ and dynamical activity $\mathcal{A}_\tau$. The inequality (2) indicates that the precision of currents is constrained by the product of the thermodynamic and kinetic costs divided by the timescale. Moreover, it provides new insights into the relationship between precision and cost in Markov jump processes; that is, increasing only the thermodynamic cost does not guarantee high precision of currents. Instead, given the same timescale $\mathcal{A}_\tau$, the product of the thermodynamic and kinetic costs must be increased to achieve high precision. Notably, the relation (2) is tighter than the conventional thermodynamic uncertainty relation [13,14].

(ii) *Variational formulas that connect optimal transport to stochastic and quantum thermodynamics.*—Using the defined state mobility term, we derive variational formulas that relate the discrete Wasserstein distance to the thermodynamic and kinetic costs in Markovian dynamics. More specifically, we prove

the following equality for the classical case (cf. Theorem 1):

$$\mathcal{W}_1(p^A, p^B) = \min \int_0^\tau \sqrt{\sigma_t m_t} \, dt = \min \sqrt{\Sigma_\tau \mathcal{M}_\tau}, \tag{3}$$

where $\mathcal{W}_1(p^A, p^B)$ is the discrete L^1-Wasserstein distance between two distributions p^A and p^B [cf. Eq. (69)] and the minimum is over all admissible Markov jump processes that transform distribution p^A into p^B over a period τ with a given connectivity. The relation (3) provides a thermodynamic interpretation for the Wasserstein distance, implying that the Wasserstein distance is equal to the minimum product of the thermodynamic and kinetic costs. We also analogously generalize the formula (3) to the quantum case, in which the classical Wasserstein distance $\mathcal{W}_1(p^A, p^B)$ is replaced with a quantum Wasserstein distance $\mathcal{W}_q(\varrho^A, \varrho^B)$ between density matrices ϱ^A and ϱ^B. These formulas can be considered as discrete generalizations of the Benamou-Brenier formula known in continuous cases.

(iii) *Trade-off between irreversibility and state mobility.*—Through the developed variational formulas, we reveal a trade-off relation between the irreversibility and dynamical state mobility in discrete systems, which reads as follows:

$$\Sigma_\tau \mathcal{M}_\tau \geq \mathcal{W}_1(p^A, p^B)^2. \tag{4}$$

The inequality (4) implies that either the thermodynamic cost Σ_τ or kinetic cost $\mathcal{M}_\tau$ must be sacrificed (i.e., they cannot be simultaneously small) to evolve the system state.

(iv) *Minimum dissipation and optimal protocol.*—The problem of minimizing entropy production in discrete systems is trivial if no constraints exist on the transition rates. Our results shed new light on this issue. More specifically, the formula (3) implies that fixing the dynamical state mobility $\mathcal{M}_\tau$ is a reasonable constraint from which minimum dissipation can be immediately determined through the Wasserstein distance, and the optimal control protocol can be constructed from the optimal transport problem. When additional constraints exist on system dynamics, we show that a lower bound on minimum dissipation can be obtained (see Fig. 4 for illustration).

(v) *Thermodynamic speed limits.*—From the resulting variational formulas, we derive unified and stringent thermodynamic speed limits that place lower bounds on the time required for state transformation for both open classical and quantum systems. The classical bound reads [cf. Eq. (128)]

$$\tau \geq \frac{\mathcal{W}_1(p^A, p^B)}{\langle \sqrt{\sigma m} \rangle_\tau} \geq \frac{\mathcal{W}_1(p^A, p^B)}{\sqrt{\langle \sigma \rangle_\tau \langle m \rangle_\tau}}, \qquad (5)$$

where $\langle x \rangle_\tau$ denotes the time average of a time-dependent variable x_t. The quantum bound has the same form, where $\mathcal{W}_1(p^A, p^B)$ is replaced with $\mathcal{W}_q(\varrho^A, \varrho^B)$. The inequality (5) implies that the speed of state transformation is constrained by the time average of the product of the thermodynamic and kinetic costs. Because we start from the equality relations, these bounds are tight and can always be saturated for *arbitrary* temperatures. In other words, for an arbitrary pair of distributions or density matrices, there always exist Markovian dynamics that saturate the bounds. They are also stronger than previously known bounds [16,22].

(vi) *Finite-time Landauer principle.*—From the variational formulas, we derive finite-time lower bounds for heat dissipation Q incurred in classical and quantum information erasure [cf. Eqs. (142) and (150)]. The bounds characterize both finite-time and finite-error effects on heat dissipation. Several finite-time Landauer bounds are derived for discrete systems in previous studies [30,31]. However, these bounds encounter the same problem as the conventional Landauer bound; that is, they lose the predictive power in the low-temperature regime. By contrast, our new bounds are tight for *arbitrary* temperatures, even in the zero-temperature limit. A further simplified bound including a finite erasure error $\epsilon \geq 0$ reads [cf. Eq. (146)]

$$Q \geq T[\ln d - h(\epsilon)] + \frac{(1 - 1/d - \epsilon)^2}{\tau \beta \langle m \rangle_\tau}, \qquad (6)$$

where d is the system's dimension, T is the temperature of the heat bath, β is the inverse temperature, and $h(\epsilon) \geq 0$ is a function that vanishes as $\epsilon \to 0$. In the perfect-erasure ($\epsilon \to 0$) and quasistatic ($\tau \to \infty$) limits, the above bound reduces to the conventional Landauer bound $Q \geq T \ln d$.

C. Relevant literature

Here, we briefly discuss several relevant studies that attempt to link the (modified) Wasserstein distances to the thermodynamics of Markov jump processes.

The Benamou-Brenier formula has two facets. It not only provides a thermodynamic interpretation, but also reveals a geometric structure for the continuous L^2-Wasserstein distance. More specifically, W_2 can be interpreted as a Riemannian metric on the manifold of probability distribution functions. Although the discrete L^2-Wasserstein distance is well defined and widely used in the literature,

unfortunately, it does not possess a geometric interpretation, and its connection to thermodynamics also remains unclear. For this reason, many studies generalize the Wasserstein distance based on the geometric aspect of the Benamou-Brenier formula for discrete cases [95]. This modified Wasserstein distance places a lower bound on irreversible entropy production of Markov jump processes [22,96]. However, it is system dependent, because the transition rates are concretely used to define this distance.

In contrast to the previous direction, in this study, we consider the conventional discrete L^1-Wasserstein distance and focus on its thermodynamic interpretation. In this regard, Dechant obtains some interesting results by relating the discrete L^1-Wasserstein distance to the entropy production and dynamical activity of Markov jump processes [93]. Here, we consider a different approach by introducing the dynamical state mobility and obtain discrete generalizations of the Benamou-Brenier formula. This approach not only unifies the classical discrete and continuous cases, but also extends to the quantum case. Although we focus on the discrete L^1-Wasserstein distance, it is noteworthy that the obtained generalizations of the Benamou-Brenier formula are similar to that for the continuous L^2-Wasserstein distance. This suggests that the discrete L^1-Wasserstein distance may play the same role as the continuous L^2-Wasserstein distance in continuous cases.

The remainder of the paper is organized as follows. Section II presents a review of the optimal transport problem and the relevant existing results in the context of continuous-state overdamped Langevin dynamics. We particularly emphasize the Benamou-Brenier formula of the L^2-Wasserstein distances and their connections to stochastic thermodynamics. In Sec. III, we briefly introduce stochastic thermodynamics of classical Markovian dynamics. Next, we define the novel kinetic term m_t and discuss its relevant properties. We then derive the improved thermodynamic uncertainty relation for Markov jump processes and numerically demonstrate it. In Sec. IV, we describe the optimal transport problem in discrete cases and explain our first theorem that links the discrete Wasserstein distance to stochastic thermodynamics of Markov jump processes. The relationship between the obtained and existing results in continuous cases is also discussed. In Sec. V, we define a quantum Wasserstein distance and explain our second theorem that generalizes the variational formula to the quantum case. From the derived variational formulas, in Sec. VI, we describe two applications: the thermodynamic speed limits and the finite-time Landauer principle. In Sec. VII, we numerically demonstrate our findings. Finally, Sec. VIII presents a conclusion with a discussion and outlook. All detailed mathematical calculations and derivations can be found in the Appendixes.

II. REVIEW OF CONTINUOUS OPTIMAL TRANSPORT

In this section, we briefly review the optimal transport problem in continuous spaces and discuss the Benamou-Brenier formula, which provides a thermodynamic interpretation of the Wasserstein distances in the context of overdamped Langevin dynamics.

A. Optimal transport problem

First, we succinctly introduce the classical transport problem on the continuous space $\mathbb{R}^d$ with $d \geq 1$ (see Ref. [69] for details). The problem of optimal transport—that is, determining how a pile of earth can be optimally transported into another pile of the same volume but with a different shape—was originally introduced by Monge. The optimality here is interpreted to mean that the total transport cost is minimized with respect to a given cost metric. Suppose that the source and target piles of earth are characterized by probability distribution functions $p^A(x)$ and $p^B(x)$, respectively, on the space $\mathbb{R}^d$, and the cost metric is given by $c : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}_{\geq 0}$. Then, the Monge optimal transport problem is to identify a one-to-one map $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ that minimizes the objective function

$$\min_{\varphi} \int_{\mathbb{R}^d} c[x, \varphi(x)] p^A(x) dx, \qquad (7)$$

where the minimum is over all φ satisfying $p^A(x) = p^B[\varphi(x)] |\det[\nabla \varphi(x)]|$. However, this formulation presents an issue regarding the nonexistence of a valid transport map; that is, the map φ might not exist in discrete cases because no mass can be split. Fortunately, this issue was previously resolved by the relaxation of Kantorovich, which led to a more well-defined problem. Instead of a transport map $\varphi(x)$, Kantorovich considers a transport plan $\pi(x, y)$ that is a joint probability distribution function and represents a coupling of p^A and p^B. This transport plan allows us to split a single mass and transport it to multiple target locations. The Kantorovich problem can be formulated as an optimization of the following objective function:

$$\min_{\pi \in \Pi(p^A, p^B)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \pi(x, y) dx dy, \qquad (8)$$

where $\Pi(p^A, p^B)$ denotes the coupling set of joint probability distribution functions whose marginal distributions coincide with p^A and p^B:

$$\int_{\mathbb{R}^d} \pi(x, y) dy = p^A(x) \text{ and } \int_{\mathbb{R}^d} \pi(x, y) dx = p^B(y). \quad (9)$$

The concept of optimal transport provides a means for defining useful metrics on continuous spaces of probability distribution functions. By employing the cost metric of the Euclidean norm [i.e., $c(x, y) = \|x - y\|^\alpha$ for a positive

number $\alpha \geq 1$], the Kantorovich problem reduces exactly to the L^α-Wasserstein distance, which is defined as

$$W_\alpha(p^A, p^B)^\alpha := \min_{\pi \in \Pi(p^A, p^B)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^\alpha \pi(x, y) dx dy. \qquad (10)$$

The L^α-Wasserstein distance is a genuine metric and satisfies the triangle inequality. Applying Hölder's inequality, we can derive a hierarchical relationship, that is, $W_\alpha \leq W_{\alpha'}$ for $\alpha \leq \alpha'$. Of the several that exist, the L^1- and L^2-Wasserstein distances are particularly relevant from the thermodynamic and geometric perspectives. In the following, we discuss some remarkable properties of these two distances.

B. Benamou-Brenier formula

The Wasserstein distance can be expressed in a variational form in several ways. Interestingly, Benamou and Brenier developed a variational formula for the L^2-Wasserstein distance in terms of fluid mechanics [79]. The Benamou-Brenier formula casts the L^2-Wasserstein distance as a minimization problem of a time-integrated cost in terms of probability distribution functions and velocity fields:

$$W_2(p^A, p^B)^2 = \min_{v_t} \tau \int_0^\tau \int_{\mathbb{R}^d} \|v_t(x)\|^2 p_t(x) dx dt, \qquad (11)$$

where the minimum is over all smooth paths $\{v_t\}_{0 \leq t \leq \tau}$ subject to the continuity equation

$$\dot{p}_t(x) + \nabla \cdot [v_t(x) p_t(x)] = 0 \qquad (12)$$

with the initial and final conditions $p_0(x) = p^A(x)$ and $p_\tau(x) = p^B(x)$, respectively. Here, ∇ is the del operator, and $\cdot$ denotes the standard Euclidean inner product between vectors. Note that, given any absolutely continuous curve $\{p_t\}$, we can always find a velocity field $\{v_t\}$ that satisfies Eq. (12). The formulation (11) not only enables us to find a numerical scheme for computing W_2, but also provides the thermodynamic and geometric interpretations of the L^2-Wasserstein distance.

Next, we discuss a thermodynamic interpretation of the L^2-Wasserstein distance (see Appendix A for a geometric interpretation). Consider an overdamped system on the continuous space $\mathbb{R}^d$, which is constantly subject to a time-dependent force $F_t(x)$ and weakly coupled to a single heat bath. The system state at time t can be characterized by the probability distribution $p_t(x)$, the time evolution of which is described by the Fokker-Planck equation:

$$\dot{p}_t(x) = -\nabla \cdot [v_t(x) p_t(x)], \qquad (13)$$

$$v_t(x) = F_t(x) - D\nabla \ln p_t(x). \tag{14}$$

Note that the velocity field $v_t(x)$ of the system in Eq. (14) is an exact solution to the continuity equation (12), which drives the source distribution $p_0(x)$ to the target distribution $p_\tau(x)$. Based on the framework of stochastic thermodynamics, the irreversible entropy production during period τ can be calculated as [2]

$$\Sigma_\tau = \frac{1}{D}\int_0^\tau \int_{\mathbb{R}^d} \|v_t(x)\|^2 p_t(x)\,dx\,dt. \tag{15}$$

Irreversible entropy production clearly coincides with the time-integrated cost in the integration in Eq. (11), ignoring the scaling factor. Therefore, we can rewrite the Benamou-Brenier formula as

$$W_2(p^A, p^B) = \min_{v_t} \sqrt{D\tau\Sigma_\tau}. \tag{16}$$

From this, the following inequality can be immediately derived:

$$\Sigma_\tau \geq \frac{W_2(p_0, p_\tau)^2}{D\tau}. \tag{17}$$

Inequality (17) refines the second law of thermodynamics by providing a stronger bound on irreversible entropy production solely in terms of the initial and final distributions, given that the operational time and diffusion coefficient are fixed. The bound can be interpreted as a thermodynamic speed limit:

$$\tau \geq \frac{W_2(p^A, p^B)}{\sqrt{D\langle\sigma\rangle_\tau}}. \tag{18}$$

Moreover, it can also be applied to derive a finite-time Landauer principle [29,76]. It is noteworthy that the bound can be saturated for any pair of the initial and final distributions and is tight even in the zero-temperature limit.

Because $v_t(x)$ and $F_t(x)$ can be considered a one-to-one correspondence, we can obtain the following equality between irreversible entropy production and the Wasserstein distance:

$$\min_{F_t} \Sigma_\tau = \frac{W_2(p_0, p_\tau)^2}{D\tau}. \tag{19}$$

This relation implies that the minimum entropy production in all overdamped processes that transform one distribution into another can be determined exactly by the Wasserstein distance between the two distributions. In Refs. [76,79], it is demonstrated that the minimum in Eq. (11) can be achieved with a velocity field of the form $v_t(x) = -\nabla\phi_t(x)$, where $\phi_t(x)$ is a time-dependent potential. Thus, the minimum entropy production can always be achieved with

a conservative force $F_t(x) = -\nabla V_t(x)$, where $V_t(x) = \phi_t(x) - D\ln p_t(x)$ is a time-dependent potential.

III. STOCHASTIC THERMODYNAMICS OF DISCRETE SYSTEMS

In this section, we first briefly introduce the stochastic thermodynamics of classical discrete Markovian dynamics described by the master equation; for a comprehensive review, one can refer to Ref. [2]. We then define a novel physical quantity called dynamical state mobility, discuss its relevant properties, and derive an improved thermodynamic uncertainty relation.

A. Markov jump processes

We consider a discrete-state system with $N > 1$ states, which is weakly attached to single or multiple thermal reservoirs. Examples of these systems include diffusive processes on a lattice, biomolecular motors, chemical reaction networks, and quantum dots. The system can be described in terms of a time-dependent probability distribution $p_t := [p_1(t), \ldots, p_N(t)]^\top$, where $p_x(t)$ denotes the probability of finding the system in state x at time t. Assume that the system is modeled by a time-continuous Markov jump process and that the transitions from a state y to a state x occur at a non-negative rate $w_{xy}(t)$, which can be time dependent according to an external control protocol. The time evolution of the probability distribution is described by the master equation

$$\dot{p}_t = \mathsf{W}_t p_t, \tag{20}$$

where dot $\cdot$ denotes the time derivative and $\mathsf{W}_t = [w_{xy}(t)]$ denotes the matrix of the transition rates with $w_{xx}(t) = -\sum_{y(\neq x)} w_{yx}(t)$. We consider microscopically reversible dynamics, that is, $w_{xy}(t) > 0$ if and only if $w_{yx}(t) > 0$. Hereafter, we assume that the transition rates satisfy the *local* detailed-balance condition [2]:

$$\ln\frac{w_{xy}(t)}{w_{yx}(t)} = s_{xy}(t), \tag{21}$$

where $s_{xy}(t)$ denotes the entropy change in the environment due to the jump from state y to x at time t. If we fix the transition rates at any time, the system relaxes toward a stationary state, which may no longer be an equilibrium state.

In a case wherein the system is attached to a single reservoir at inverse temperature β and the transitions between states are induced by the energy difference, the entropy change reads

$$s_{xy}(t) = \beta[\varepsilon_y(t) - \varepsilon_x(t)], \tag{22}$$

where $\varepsilon_x(t)$ denotes the instantaneous energy level of state x at time t. Whenever this occurs, we say that the system

satisfies the *global* detailed-balance condition. Notably, the thermal state $p_x^{\text{eq}}(t) \propto e^{-\beta \varepsilon_x(t)}$ becomes the instantaneous stationary state of the system (i.e., $\mathsf{W}_t p_t^{\text{eq}} = 0$). Hereafter, we consider generic dynamics satisfying the local detailed-balance condition. However, dynamics satisfying Eq. (22) are used occasionally for physical interpretation of some quantities.

For convenience, we define the following quantities:

$$a_{xy}(t) := w_{xy}(t) p_y(t), \tag{23}$$

$$j_{xy}(t) := w_{xy}(t) p_y(t) - w_{yx}(t) p_x(t), \tag{24}$$

which quantify the frequency of jumps and the probability current from state y to x at time t, respectively.

B. Entropy production and dynamical activity

Given the previous setup, we now discuss some relevant thermodynamic quantities. One central quantity is irreversible entropy production, which quantifies the degree of irreversibility of the thermodynamic process. Within the framework of stochastic thermodynamics, total entropy production during a period τ can be defined as

$$\Sigma_\tau := \Delta S_{\text{sys}} + \Delta S_{\text{env}}, \tag{25}$$

where ΔS_{sys} and ΔS_{env} are the changes in the entropy of the system and the environment, respectively, expressed as

$$\Delta S_{\text{sys}} = S(p_\tau) - S(p_0), \tag{26}$$

$$\Delta S_{\text{env}} = \int_0^\tau \sum_{x \neq y} a_{xy}(t) s_{xy}(t) dt. \tag{27}$$

Here, system entropy production is quantified via the Shannon entropy $S(p) := -\sum_x p_x \ln p_x$, whereas the entropy change of the environment is defined as the sum of entropic contributions from each transition between states. Simple calculations show that the entropy production rate is always non-negative:

$$\sigma_t := \dot{\Sigma}_t = \sum_{x > y} [a_{xy}(t) - a_{yx}(t)] \ln \frac{a_{xy}(t)}{a_{yx}(t)} \geq 0. \tag{28}$$

The non-negativity of irreversible entropy production corresponds to the second law of thermodynamics. The equality of this zero bound is attained only when the system is in the instantaneous thermal state at all times.

Another essential quantity in nonequilibrium thermodynamics is dynamical activity, quantified by the amplitude of transitions between states as

$$a_t := \sum_{x \neq y} a_{xy}(t). \tag{29}$$

The average number of jumps during period τ can be calculated as

$$\mathcal{A}_\tau := \int_0^\tau a_t dt. \tag{30}$$

The time average of dynamical activity characterizes the timescale of thermodynamic processes. The higher the dynamical activity, the stronger the thermalization. Entropy production and dynamical activity are the time-antisymmetric and time-symmetric parts, respectively, of the path-integral action with respect to a time-reversed process [94]. Both quantities constrain the fluctuation of currents according to the thermodynamic and kinetic uncertainty relations [13,97].

C. Dynamical state mobility

1. Definition

Here, we introduce a new quantity called dynamical state mobility, which plays a crucial role in our results. Before getting into the details, let us briefly recall the linear response relations [98,99], which express the equalities between currents and forces in near-equilibrium systems. Consider an irreversible transport process driven by thermodynamic forces $F = [F_x]^\top$, such as affinities in temperatures or chemical potentials. Let $J = [J_x]^\top$ be the thermodynamic currents that characterize the response of the system to the applied forces. In a linear-response regime, the currents depend only on the thermodynamic forces and can be expressed by the following linear relations:

$$J_x = \sum_y \mu_{xy} F_y \quad \text{or} \quad J = \mathsf{L}F. \tag{31}$$

These relations (31) are referred to as linear response equations, where the coefficients μ_{xy} are known as Onsager kinetic coefficients and $\mathsf{L} = [\mu_{xy}]$ is called the Onsager matrix. Onsager reciprocal relations imply that, in the case of time-reversal symmetry, the Onsager matrix is symmetric (i.e., $\mu_{xy} = \mu_{yx}$). In addition, the entropy production rate can be expressed in a quadratic form of the forces as

$$\sigma = \sum_x J_x F_x = \sum_{x,y} \mu_{xy} F_x F_y \quad \text{or} \quad \sigma = F^\top \mathsf{L}F. \tag{32}$$

The non-negativity of the entropy production rate immediately derives that L is positive semidefinite.

The Onsager coefficients characterize the response of dynamics close to equilibrium at the *macroscopic* level. Nevertheless, they can be mimicked to dynamics far from equilibrium at the *microscopic* level. To show this, let us focus on local transitions between states. The generalized thermodynamic force associated with each transition from y to x is defined as [13]

$$f_{xy}(t) := \ln \frac{a_{xy}(t)}{a_{yx}(t)}, \tag{33}$$

which is the sum of the entropy changes in the system and environment derived from the jump. Since $j_{xy}(t)$ is the probability current associated with the transition from y to x, defining the following coefficient is logical:

$$m_{xy}(t) := \frac{a_{xy}(t) - a_{yx}(t)}{\ln a_{xy}(t) - \ln a_{yx}(t)} = \frac{j_{xy}(t)}{f_{xy}(t)}. \quad (34)$$

Intuitively, $\{m_{xy}(t)\}$ characterize the responses of the probability currents against the thermodynamic forces at the transition level. Notice that $\{m_{xy}(t)\}$ are always non-negative and symmetric [i.e., $m_{xy}(t) = m_{yx}(t) \geq 0$]. Equation (34) can also be rewritten in the form of

$$j_{xy}(t) = m_{xy}(t) f_{xy}(t), \quad (35)$$

which shows that the currents and thermodynamic forces can be linearly related in terms of the coefficients $\{m_{xy}(t)\}$. Moreover, the entropy production rate can be expressed in a quadratic form of the thermodynamic forces $\{f_{xy}(t)\}$ as

$$\sigma_t = \sum_{x>y} m_{xy}(t) f_{xy}(t)^2 = \sum_{x>y} \sigma_{xy}(t). \quad (36)$$

Here, we define the entropy production rate associated with each transition as $\sigma_{xy}(t) := m_{xy}(t) f_{xy}(t)^2$. Equations (35) and (36) have the same algebraic forms as Eqs. (31) and (32), respectively, which suggests that the coefficients $\{m_{xy}(t)\}$ play similar roles with the Onsager coefficients for far-from-equilibrium systems.

In a weak-thermodynamic-force limit [i.e., $|f_{xy}(t)| \rightarrow 0$], the coefficient $m_{xy}(t)$ reduces to the average dynamical activity between states x and y:

$$m_{xy}(t) \rightarrow \frac{a_{xy}(t) + a_{yx}(t)}{2}. \quad (37)$$

This is somewhat analogous to the Einstein relation on mobility in overdamped Langevin dynamics (see Table I). Note that the weak-thermodynamic-force limit is defined at the microscopic state level and can be achieved via two routes: the completely equilibrium limit in discrete systems and the continuous-state limit (e.g., the limit from the discrete hopping particle system to the overdamped Fokker-Planck equation). In the continuous-state limit,

TABLE I. Analogy between the dynamical state mobility and macroscopic mobility.

Microscopic level	Macroscopic level
$j_{xy} = m_{xy} f_{xy}$	$J = \mu F$
Einstein-like relation $\|f_{xy}\| \ll 1$	Einstein relation $\|F\| \ll 1$
$m_{xy} = (a_{xy} + a_{yx})/2$	$\mu = \beta D$

the difference between neighboring states x and y is infinitesimal, and, thus, the force $f_{xy}(t)$ is also infinitesimal because $a_{xy}(t)/a_{yx}(t) \simeq 1$. We discuss the continuous-state limit in the following subsection and show that the right-hand side in Eq. (37) is proportional to the diffusion coefficient. In general, the following relation holds for the coefficient $m_{xy}(t)$:

$$\sqrt{a_{xy}(t) a_{yx}(t)} \leq m_{xy}(t) \leq \frac{a_{xy}(t) + a_{yx}(t)}{2}. \quad (38)$$

It is, thus, natural to define the sum of $\{m_{xy}(t)\}$ over all transitions:

$$m_t := \sum_{x>y} m_{xy}(t). \quad (39)$$

For convenience, we refer to this term as dynamical state mobility throughout this paper. This nomenclature derives from the analogy between microscopic coefficients $\{m_{xy}(t)\}$ and the macroscopic mobility (see Table I).

To clarify further the identity of m_t, let us consider a case in which the system is attached to a single reservoir and satisfies the global detailed-balance condition. In this case, the master equation (20) can be written as [22]

$$\dot{p}_t = \mathsf{K}_t f_t, \quad (40)$$

where K_t is a symmetric, positive semidefinite matrix given by

$$\mathsf{K}_t := \frac{1}{2} \sum_{x \neq y} m_{xy}(t) \mathsf{E}_{xy}. \quad (41)$$

Here, $f_t = [f_1(t), \ldots, f_N(t)]^\top$ with $f_x(t) = -\ln p_x(t) + \ln p_x^{\text{eq}}(t)$ and $\mathsf{E}_{xy} = [e_{uv}] \in \mathbb{R}^{N \times N}$ is a matrix with $e_{yy} = e_{xx} = 1$, $e_{xy} = e_{yx} = -1$, and zeros in all other elements. The quantities $\{f_x(t)\}$ are identified as the entropic thermodynamic forces, which characterize how far the system is driven from the instantaneous equilibrium state. Equation (40) represents the linear relations between the rates $\dot{p}_t$ and forces f_t through the symmetric matrix K_t. Furthermore, the total entropy production rate can be written in a quadratic form as [22]

$$\sigma_t = f_t^\top \mathsf{K}_t f_t. \quad (42)$$

Therefore, Eqs. (40) and (42) can be viewed as far-from-equilibrium counterparts of Eqs. (31) and (32), respectively. The matrix K_t is, thus, identified as the Onsager-like matrix. Because $\{m_{xy}(t)\}$ are elements of K_t, they can be regarded as the Onsager-like kinetic coefficients for out-of-equilibrium systems. From the definition of K_t, we can

easily verify that m_t is exactly the sum of diagonal elements of the Onsager-like matrix:

$$m_t = \frac{1}{2} \operatorname{tr} \mathsf{K}_t. \tag{43}$$

Therefore, m_t can be identified as a kinetic term. The time integral of the kinetic term m_t can be considered as the kinetic cost of Markov jump processes, defined by

$$\mathcal{M}_\tau := \int_0^\tau m_t dt = \tau \langle m \rangle_\tau, \tag{44}$$

where we define the time-averaged quantity for arbitrary time-dependent quantity x_t as

$$\langle x \rangle_\tau := \tau^{-1} \int_0^\tau x_t dt. \tag{45}$$

By summing both sides of Eq. (38) for all $x > y$, we can prove that dynamical state mobility is upper bounded by dynamical activity as

$$m_t \le \frac{a_t}{2}. \tag{46}$$

The equality of Eq. (46) can be achieved in the equilibrium limit. It is, thus, evident that $\mathcal{M}_\tau \le \mathcal{A}_\tau / 2$.

2. Continuous limit

Next, we investigate m_t in the continuous limit. To this end, we consider an overdamped Brownian particle trapped in a one-dimensional potential $V_t(x)$. Let x_t denote the position of the particle at time t. Then, its dynamics is governed by the Langevin equation:

$$\dot{x}_t = F_t(x_t) + \sqrt{2D} \xi_t, \tag{47}$$

where $F_t(x) := -\partial_x V_t(x)$ is the total force applied on the particle, ξ_t is a zero-mean Gaussian white noise with variance $\langle \xi_t \xi_{t'} \rangle = \delta(t - t')$, and $D > 0$ is the diffusion coefficient. Let $p_t(x)$ be the probability distribution function of finding the particle in state x at time t. Then, its time evolution can be described by the Fokker-Planck equation:

$$\dot{p}_t(x) = -\partial_x [F_t(x) p_t(x) - D \partial_x p_t(x)], \tag{48}$$

where we set $\mu = 1$ for simplicity. We now consider the discretization of the Fokker-Planck equation (48) with space interval $\Delta x > 0$ and define $x_n := n\Delta x$. By defining the probability distribution and transition rates as

$$p_n(t) := p_t(x_n)\Delta x, \tag{49}$$

$$w_{n+1,n}(t) := \frac{D}{(\Delta x)^2} \exp\left[\frac{V_t(x_n) - V_t(x_{n+1})}{2D}\right]$$
$$\simeq \frac{F_t(x_n)}{2\Delta x} + \frac{D}{(\Delta x)^2}, \tag{50}$$

$$w_{n-1,n}(t) := \frac{D}{(\Delta x)^2} \exp\left[\frac{V_t(x_n) - V_t(x_{n-1})}{2D}\right]$$
$$\simeq \frac{-F_t(x_n)}{2\Delta x} + \frac{D}{(\Delta x)^2}, \tag{51}$$

we readily obtain the master equation:

$$\dot{p}_n(t) = w_{n,n-1}(t) p_{n-1}(t) + w_{n,n+1}(t) p_{n+1}(t)$$
$$- [w_{n+1,n}(t) + w_{n-1,n}(t)] p_n(t). \tag{52}$$

We note here from Eqs. (50) and (51) that

$$a_{n+1,n}(t) \simeq \left[\frac{F_t(x_n)}{2\Delta x} + \frac{D}{(\Delta x)^2}\right] p_n(t), \tag{53}$$

$$\frac{a_{n+1,n}(t) + a_{n,n+1}(t)}{2} = D p_t(x_n)(\Delta x)^{-1} + O(1). \tag{54}$$

The probability currents and thermodynamic forces can be calculated as

$$j_{n+1,n}(t) = F_t(x_n) p_t(x_n) - D \partial_x p_t(x_n) + O(\Delta x), \tag{55}$$

$$f_{n+1,n}(t) = \left[\frac{F_t(x_n)}{D} - \frac{\partial_x p_t(x_n)}{p_t(x_n)}\right] \Delta x + O(\Delta x^2). \tag{56}$$

From these expressions, $m_{n+1,n}(t)$ can be calculated via the definition (34), which gives

$$m_{n+1,n}(t) \to \frac{a_{n+1,n}(t) + a_{n,n+1}(t)}{2}$$
$$= D p_t(x_n)(\Delta x)^{-1} + O(1). \tag{57}$$

Equation (57) indicates that $m_{n+1,n}(t)$ converges to the value $[a_{n+1,n}(t) + a_{n,n+1}(t)]/2$, and these correspond to the product of the diffusion coefficient and probability distribution (see Table I). Summing both sides of Eq. (57) for all n, we obtain

$$m_t \to \frac{a_t}{2} \simeq \sum_n D p_t(x_n)(\Delta x)^{-1} = D(\Delta x)^{-2}. \tag{58}$$

This implies that m_t is proportional to D as the scaling factor is ignored. Thus, m_t should play the same role as that of the diffusion coefficient. It is noteworthy that the diffusion coefficient is exactly the diagonal Onsager coefficient of overdamped Langevin processes in the linear-response regime.

D. Thermodynamic uncertainty relation: State mobility is crucial in nonequilibrium

Here, we describe an improved thermodynamic uncertainty relation, showing that the kinetic cost of dynamical state mobility plays a critical role in constraining the precision of time-integrated currents. For simplicity, we consider a steady-state Markov jump process. The generalization to the case of an arbitrary initial state and arbitrary time-dependent driving is straightforward.

Let $\Gamma = \{x_0, (t_1, x_1), ..., (t_K, x_K)\}$ be a stochastic trajectory, in which the system is initially at state x_0 and subsequently jumps from state x_{k-1} to x_k at time t_k for each $k = 1, ..., K$. For each stochastic trajectory Γ, we consider a time-antisymmetric current J, defined as

$$J(\Gamma) := \sum_{k=1}^{K} \Upsilon_{x_k x_{k-1}}. \tag{59}$$

Here, $\{\Upsilon_{xy}\}$ are arbitrary real coefficients satisfying $\Upsilon_{xy} = -\Upsilon_{yx}$ for all x and y. Examples of currents include the entropy flux and heat flux to the environment by specific choices of $\{\Upsilon_{xy}\}$. The precision of current J can be quantified by the square of the current mean divided by its variance $\langle J \rangle^2 / \mathrm{var}[J]$. The conventional thermodynamic uncertainty relation [12,13] sets an upper bound on the precision in terms of the total entropy production, given by the following inequality:

$$\frac{\langle J \rangle^2}{\mathrm{var}[J]} \leq \frac{\Sigma_\tau}{2}. \tag{60}$$

Numerous studies generalize this relation to other dynamics, from classical to quantum [100–118].

We improve the thermodynamic uncertainty relation by proving that the precision of currents is upper bounded by the product of the thermodynamic and kinetic costs, as follows:

$$\frac{\langle J \rangle^2}{\mathrm{var}[J]} \leq \eta \frac{\Sigma_\tau}{2}, \tag{61}$$

where $\eta = 2\mathcal{M}_\tau / \mathcal{A}_\tau \leq 1$ can be regarded as an efficiency of dynamical activity (see Appendix C 3 for the proof). The new relation (61) is tighter than the conventional relation (60) and can be saturated in the case of a one-dimensional random walk. Although the conventional relation (60) implies that increasing dissipation is necessary to achieve high precision, it does not ensure the converse; that is, increasing dissipation is not sufficient for obtaining high precision of currents. This can be explained through our relation, where the kinetic contribution η appears in the bound in addition to the thermodynamic contribution. For systems far from equilibrium, it is tedious that $\eta \ll 1$,
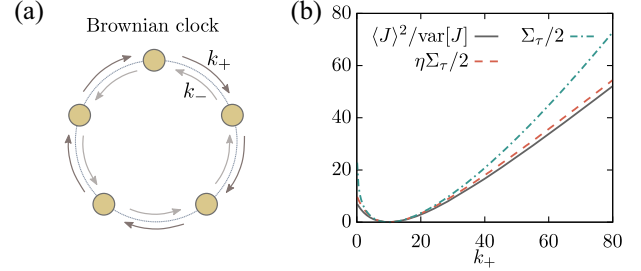

FIG. 2. Numerical illustration of the thermodynamic uncertainty relations. (a) Schematic of the five-state Brownian clock and (b) numerical verification. The current precision $\langle J \rangle^2 / \mathrm{var}[J]$, new bound $\eta \Sigma_\tau / 2$, and conventional bound $\Sigma_\tau / 2$ are indicated by the solid, dashed, and dash-dotted lines, respectively. The forward rate k_+ is varied, whereas the backward rate k_- is fixed at $k_- = 10$. The operational time is fixed at $\tau = 1$.

which equivalently indicates the unattainability of the conventional bound.

For a case in which the system is in an arbitrary initial state and is driven by a time-dependent protocol, the derived relation can be analogously generalized as

$$\frac{[(\tau \partial_\tau - v \partial_v) \langle J \rangle]^2}{\mathrm{var}[J]} \leq \eta \frac{\Sigma_\tau}{2}, \tag{62}$$

where v is a speed parameter of the control protocol [112].

In the following, we exemplify the derived thermodynamic uncertainty relation in a five-state Brownian clock [119]. The Brownian clock is modeled as an inhomogeneous biased random walk on a ring with five states [see Fig. 2(a)]. The clock's pointer transits from state x to $x + 1 (6 \equiv 1)$ at rate $k_+ > 0$, whereas the backward rate is $k_- > 0$. The net number of cycles completed by the pointer characterizes the clock's time. In other words, time can be counted by a stochastic current J that increases by 1 for each transition from state 5 to 1 and decreases by 1 for the reverse transition from state 1 to 5. The stochastic current J can be defined by setting $\Upsilon_{15} = 1 = -\Upsilon_{51}$ and $\Upsilon_{xy} = 0$ for others. Thus, the precision of the clock can be quantified by $\langle J \rangle^2 / \mathrm{var}[J]$.

We consider the clock operating in the stationary state. To investigate the quality of the bounds, we fix the backward rate $k_- = 10$ and vary the forward rate $k_+ \in (0, 80)$. For each parameter setting, we calculate the precision of the clock and the bounds of the conventional and new thermodynamic uncertainty relations using the full counting statistics. The numerical results are plotted in Fig. 2(b), which verify that the new bound is always tighter than the conventional bound. In particular, the new bound is tight even in the far-from-equilibrium regime.

IV. RESULTS ON DISCRETE OPTIMAL TRANSPORT

Thus far, the problem of optimal transport is discussed in terms of continuous spaces. In the following, we focus on

the case of discrete spaces and explain the discrete Wasserstein distance. We then state our first theorem, which connects the discrete Wasserstein distance to stochastic thermodynamics of Markov jump processes.

A. Optimal transport distance

The optimal transport problem in the discrete case is analogous to that in the continuous case except that we now deal with discrete N-dimensional distributions. Given two discrete distributions $p^A = [p_x^A]$ and $p^B = [p_x^B]$, the optimal means of transporting distribution p^A to p^B with respect to a cost matrix $C = [c_{xy}]$ becomes the focus. Here, $c_{xy} \geq 0$ denotes the cost of transporting a unit probability from p_y^A to p_x^B.

The transport problem can be formulated using a coupling $\pi = [\pi_{xy}]$ between the probability distributions p^A and p^B. Specifically, π is a joint probability distribution such that its marginal distributions coincide with p^A and p^B (i.e., the following conditions are satisfied for all x):

$$p_x^A = \sum_{y=1}^{N} \pi_{yx} \quad \text{and} \quad p_x^B = \sum_{y=1}^{N} \pi_{xy}. \qquad (63)$$

Each coupling, thus, defines a transport plan: For each x and y, we transport an amount π_{xy} from p_y^A to p_x^B. Thus, the discrete L^1-Wasserstein distance can be defined as the minimum transport cost over all admissible couplings:

$$W_1(p^A, p^B) := \min_{\pi \in \Pi(p^A, p^B)} \sum_{x,y} c_{xy} \pi_{xy}, \qquad (64)$$

where $\Pi(p^A, p^B)$ denotes a set of couplings between p^A and p^B. Once the cost matrix is provided, the discrete Wasserstein distance can be efficiently computed using a linear programming method. In addition, as long as the cost matrix satisfies

$$c_{xy} + c_{yz} \geq c_{xz} \qquad (65)$$

for any x, y, and z, the resulting distance fulfills the triangle inequality:

$$W_1(p^A, p^B) + W_1(p^B, p^C) \geq W_1(p^A, p^C). \qquad (66)$$

We observe that the definition of the Wasserstein distance depends on the cost matrix. In other words, each matrix of transport costs induces a quantitatively different measure of distance. Evidently, an infinite number of approaches can be used to choose the cost matrix. In the following, we consider the cost matrix and corresponding Wasserstein distance defined on the basis of a graph.

Let $\mathcal{G}(V, E)$ denote an undirected graph, where V and E are the sets of vertices and unordered edges, respectively. Then, any microscopically reversible Markov jump process can be associated with an undirected graph, in which $V = \{1, \ldots, N\}$ is the set of all states of the Markov jump process, and two vertices x and y are connected by an edge $(x, y) \in E$ if the transition between x and y is allowed. A jump process that has a unique steady state can be described by a connected graph; that is, for any unordered pair (x, y), a sequence of vertices $P = [v_1, \ldots, v_k]$ always exists such that $x = v_1$, $y = v_k$, and $(v_i, v_{i+1}) \in E$ for all $1 \leq i \leq k - 1$. A subgraph $\tilde{\mathcal{G}}$ of a graph $\mathcal{G}$ is one whose edge set is a subset of that of $\mathcal{G}$. In other words, $\tilde{\mathcal{G}}$ can be obtained from $\mathcal{G}$ by removing some edges. This is equivalent to setting some transition rates of the Markov jump process to zero. For convenience, hereafter, the underlying graph structure of a Markov jump process is referred to as its *topology*.

Given the topology of a jump process, we now can define the transport cost matrix. For each path P, let $\text{len}(P)$ denote its length, which is the number of edges contained in the path. The shortest-path distance from vertex x to vertex y can be defined as

$$d_{yx} := \min_{P} \{\text{len}(P)\}, \qquad (67)$$

where the minimum is over all paths that connect x to y. For undirected graphs, clearly $d_{xy} = d_{yx}$. From the definition of the distances $\{d_{yx}\}$, we can easily verify that the triangle inequality is fulfilled, that is,

$$d_{xy} + d_{yz} \geq d_{xz} \qquad (68)$$

for arbitrary vertices x, y, and z. Employing these shortest-path distances as the transport costs (i.e., $c_{xy} = d_{xy}$), we hereafter exclusively focus on the following discrete Wasserstein distance:

$$\mathcal{W}_1(p^A, p^B) := \min_{\pi \in \Pi(p^A, p^B)} \sum_{x,y} d_{xy} \pi_{xy}. \qquad (69)$$

It is noteworthy that only static information (i.e., graph connectivity) is required to define the Wasserstein distance at this time (see Fig. 3 for illustration).
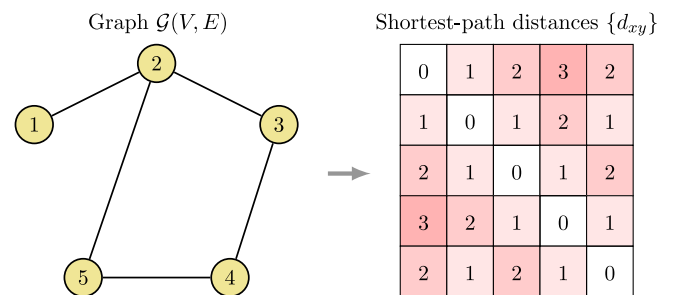

FIG. 3. Example of the Wasserstein distance defined based on a graph consisting of five vertices and five edges. Given the topology $\mathcal{G}(V, E)$, the shortest-path distances $\{d_{xy}\}$ can be calculated, from which the Wasserstein distance can be defined.

In a general case, since $d_{xy} \geq 1$ for all $x \neq y$, the Wasserstein distance is always lower bounded by the total variation distance:

$$\mathcal{W}_1(p^A, p^B) \geq \mathcal{T}(p^A, p^B) := \frac{1}{2} \sum_x |p_x^A - p_x^B|. \quad (70)$$

However, when the underlying graph is fully connected (i.e., the transition between any two states is admissible), the shortest-path distance becomes

$$d_{xy} = 1 - \delta_{xy}, \quad (71)$$

where δ_{xy} is the Kronecker delta of x and y. In this case, the Wasserstein distance coincides with the total variation distance (see Appendix D 1 for the proof):

$$\mathcal{W}_1(p^A, p^B) = \mathcal{T}(p^A, p^B). \quad (72)$$

B. Thermodynamic interpretation

With the above setup, we can now state the results. Conventionally, the discrete Wasserstein distance is defined mathematically using the transport cost matrix based on only the shortest-path distances. However, in the following theorem, we explicitly show an intimate relationship between the discrete Wasserstein distance defined in Eq. (69) and the stochastic thermodynamics of Markov jump processes.

Theorem 1: The Wasserstein distance based on a topology $\mathcal{G}(V, E)$ can be written in variational forms as

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \sqrt{\sigma_t m_t} dt \quad (73)$$

$$= \min_{\mathsf{W}_t} \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \quad (74)$$

Here, the minimum is taken over all transition rate matrices $\{\mathsf{W}_t\}_{0 \leq t \leq \tau}$ which satisfy the master equation (20) with the boundary conditions $p_0 = p^A$ and $p_\tau = p^B$ and induce subgraphs of $\mathcal{G}(V, E)$ for all times.

Theorem 1 is the first central result, and its sketch proof is given in the following. Note that the minimization is over all transition rate matrices which are microscopically reversible and induce subgraphs of $\mathcal{G}(V, E)$ for all times. This means that the transition rate between two states x and y must be fixed to zero for all times if no edge exists between vertices x and y in the graph $\mathcal{G}$. Otherwise, as long as an edge exists between the vertices, the transition rate can be arbitrarily controlled. Notably, the equality of Eq. (73) can always be ensured with dynamics that satisfy the global detailed-balance condition (see Appendix D 3 for the proof).

Proof.—Here, we provide an outline of the proof; see Appendix D 2 for a detailed derivation. The proof strategy can be mainly divided into the following two steps. We first prove that

$$\mathcal{W}_1(p^A, p^B) \leq \int_0^\tau \sqrt{\sigma_t m_t} dt \leq \sqrt{\Sigma_\tau \mathcal{M}_\tau} \quad (75)$$

holds for all admissible Markovian dynamics that transform p^A into p^B, and we then construct a specific process that attains the equality. Since the second inequality in Eq. (75) is simply a consequence of the Cauchy-Schwarz inequality, Eq. (75) can be proved after we verify the first inequality. This can be done by proving the following relation:

$$\mathcal{W}_1(p^A, p^B) \leq \int_0^\tau \sum_{x>y} |j_{xy}(t)| dt \leq \int_0^\tau \sqrt{\sigma_t m_t} dt. \quad (76)$$

The second inequality in Eq. (76) can be derived using the Cauchy-Schwarz inequality. Thus, we need only show the first inequality in Eq. (76). To this end, we map the Wasserstein distance to the minimum cost of the minimum cost flow problem in the field of graph theory. For this problem, we can show that the Wasserstein distance is exactly the optimal flow cost. Moreover, the Markov jump process also yields an admissible solution to the flow problem with the cost $\int_0^\tau \sum_{x>y} |j_{xy}(t)| dt$. Consequently, the first inequality in Eq. (76) is proved. Finally, we inversely translate the optimal solution of the minimum cost flow problem to construct a Markov jump process that attains the equality of Eq. (75). ∎

Some remarks regarding Theorem 1 are in order. First, Eqs. (73) and (74) provide a thermodynamic interpretation of the discrete Wasserstein distance; that is, $\mathcal{W}_1$ equals the minimum product of the thermodynamic and kinetic costs over all admissible Markovian dynamics that transform the source distribution into the target one. From a different perspective, it can be regarded as a trade-off between irreversible entropy production and dynamical state mobility; that is, to transform a probability distribution into another one, both Σ_τ and $\mathcal{M}_\tau$ cannot be simultaneously small:

$$\Sigma_\tau \mathcal{M}_\tau \geq \mathcal{W}_1(p^A, p^B)^2. \quad (77)$$

In other words, either the thermodynamic or kinetic cost must be sacrificed to achieve a feasible state transformation.

Second, we show that the discrete Wasserstein distance has analogous thermodynamic properties with the continuous L^2-Wasserstein distance. To this end, we rewrite Eq. (74) in the following form:

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \sqrt{\bar{D} \tau \Sigma_\tau}, \quad (78)$$

where we define time-averaged state mobility $\bar{D} := \langle m \rangle_\tau$. As previously shown, the kinetic term m_t reduces to the

diffusion coefficient in the continuous limit. Therefore, its time-averaged quantity $\bar{D}$ here plays the same role as the diffusion coefficient D does in the continuous case. Consequently, Eq. (78) can be regarded as the discrete analog of the Benamou-Brenier formula (16) known for the L^2-Wasserstein distance. Equation (78) immediately derives a lower bound on irreversible entropy production:

$$\Sigma_\tau \geq \frac{\mathcal{W}_1(p^A, p^B)^2}{\bar{D}\tau}. \tag{79}$$

This bound is tight and can always be attained for an arbitrary pair of distributions. In other words, the minimum entropy production among all feasible dynamics that have the same value of $\bar{D}$ is given by the Wasserstein distance:

$$\min_{\langle m \rangle_\tau = \bar{D}} \Sigma_\tau = \frac{\mathcal{W}_1(p^A, p^B)^2}{\bar{D}\tau}. \tag{80}$$

Equations (79) and (80) can be considered discrete analogs of Eqs. (17) and (19), respectively.

Third, Theorem 1 provides insights into the problem of minimizing entropy production in discrete Markovian dynamics. Previous studies show that the irreversible entropy production required to transform the initial into the final distribution can be arbitrarily small if no constraint is placed on the transition rates [91–93]. Equation (80) also confirms this, where the minimum entropy production depends on $\bar{D}$ and can be arbitrarily adjusted. Theorem 1 suggests that fixing $\bar{D}$ is a reasonable constraint under which the minimum entropy production is determined by the Wasserstein distance, as in the continuous case. Notably, as Appendix C 1 shows, $\bar{D}$ can be fixed to an arbitrary positive value, indicating the flexibility of the optimization problem. The optimal control protocol can also be constructed from optimal coupling, as shown in our proof of Theorem 1. Moreover, for arbitrary $\bar{D} > 0$, the minimum entropy production in Eq. (80) can always be attained using a system with conservative forces (see Appendix D 4 for the proof). It is noteworthy that the discussion thus far does not impose any other constraints on the transition rates, except for fixing $\langle m \rangle_\tau$. Therefore, if some additional constraints are placed on the transition rates, such as upper or lower bounds on the magnitude of transition rates, it may not be the case. Nevertheless, a lower bound can be derived for the minimum entropy production in this case. Let $\bar{D}_{\max}$ be the maximum of $\langle m \rangle_\tau$ among all processes that transform distribution p^A into p^B under these constraints. Then, the minimum entropy production is lower bounded by the Wasserstein distance and $\bar{D}_{\max}$ as

$$\Sigma_\tau \geq \frac{\mathcal{W}_1(p^A, p^B)^2}{\bar{D}_{\max}\tau}. \tag{81}$$
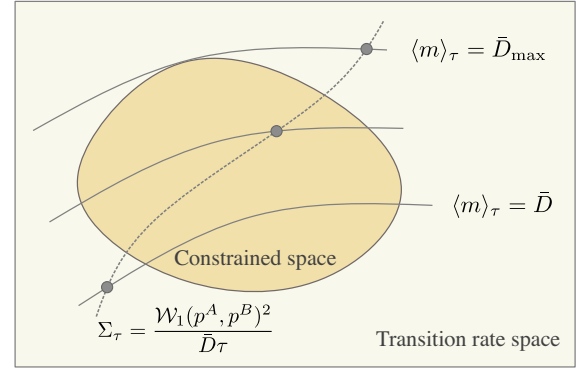

FIG. 4. Schematic of the thermodynamic structure of minimum dissipation in the space of transition rates. Solid lines represent dynamics that have the same value of the average state mobility $\langle m \rangle_\tau$. Black circles depict extreme points in which the minimum dissipation is attained, provided that the average state mobility is fixed. In the presence of other constraints on transition rates, the minimum dissipation is lower bounded by the Wasserstein distance and maximum value of $\langle m \rangle_\tau$.

The thermodynamic structure of minimum dissipation is illustrated in Fig. 4.

Finally, as shown in Sec. VI, the variational formulas (73) and (74) have crucial implications for thermodynamic speed limits and thermodynamic cost of information erasure at arbitrary temperatures.

Note that each topology induces a different Wasserstein metric. In the following, we consider a specific topology and discuss the relevance of the variational formulas (73) and (74). For other common topologies, see Appendix D 5.

One common topology is one-dimensional nearest-neighbor, in which a jump between states x and y is admitted if and only if $|x - y| = 1$. This topology is relevant to Brownian random walks and the discretization of a one-dimensional Langevin system. The shortest-path distances in this topology can be readily calculated as

$$d_{xy} = |x - y|. \tag{82}$$

Because this cost matrix is the discrete analog of the cost function $c(x, y) = |x - y|$ used in the definition of the continuous L^1-Wasserstein distance W_1, the discrete Wasserstein distance $\mathcal{W}_1$ should be reduced to W_1 in the continuous limit. Let Δx be the space interval. Then, $\mathcal{W}_1$ converges to W_1 as $\Delta x \to 0$ and $N \to \infty$:

$$\mathcal{W}_1(p^A, p^B)\Delta x \xrightarrow[\Delta x \to 0]{N \to \infty} W_1(p^A, p^B). \tag{83}$$

In addition, as shown in the proof of Theorem 1, $\mathcal{W}_1(p^A, p^B)$ can be expressed in terms of the probability currents as

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \sum_{x > y} |j_{xy}(t)| dt. \tag{84}$$

Equation (84) implies that the discrete Wasserstein distance is equal to the minimum sum of absolute probability currents. In the case considered here, the equality (84) reads

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \sum_{x=1}^{N-1} |j_{x+1,x}(t)| dt. \qquad (85)$$

Noticing that $\sum_{x=1}^{N-1} |j_{x+1,x}(t)|\Delta x \to \int_{\mathbb{R}} |j_t(x)| dx$ as $\Delta x \to 0$, we obtain the following limit:

$$\mathcal{W}_1(p^A, p^B)\Delta x \xrightarrow[\Delta x \to 0]{N \to \infty} \min_{j_t} \int_0^\tau \int_{\mathbb{R}} |j_t(x)| dx dt. \qquad (86)$$

Combining Eqs. (83) and (86) gives the following relation:

$$W_1(p^A, p^B) = \min_{j_t} \int_0^\tau \int_{\mathbb{R}} |j_t(x)| dx dt, \qquad (87)$$

where $j_t(x)$ is subject to the equation $\dot{p}_t(x) = -\partial_x j_t(x)$. Notably, Eq. (87) is exactly the Benamou-Brenier formula for the continuous L^1-Wasserstein distance in the one-dimensional case [120]. Therefore, we can conclude that Eq. (84) provides a unified generalization of the Benamou-Brenier formula for the L^1-Wasserstein distance.

Let us now consider the discretization of one-dimensional Langevin dynamics, that is, Markov jump processes with transition rates specified as in Eqs. (50) and (51). For these jump processes, the dynamical state mobility reduces to the diffusion coefficient in the continuous limit [i.e., $\mathcal{M}_\tau(\Delta x)^2 \to D\tau$ as $\Delta x \to 0$]. In the continuous case, the Benamou-Brenier formula (87) can be expressed as

$$
\begin{aligned}
W_1(p^A, p^B) &= \min_{j_t} \int_0^\tau \int_{\mathbb{R}} \sqrt{\sigma_t(x)m_t(x)} dx dt \\
&\leq \min_{F_t} \int_0^\tau \sqrt{D\sigma_t} dt = \min_{F_t} \sqrt{D\tau \Sigma_\tau}, \quad (88)
\end{aligned}
$$

where we define $m_t(x) := Dp_t(x)$ and the local entropy production rate $\sigma_t(x) := j_t(x)^2/[Dp_t(x)]$. From Theorem 1 and Eq. (88), we can conclude that, in the continuous limit, the equality in Theorem 1 might not be achieved with Markov jump processes whose transition rates are expressed as in Eqs. (50) and (51). This shows the difference between the discrete and continuous cases, where the discrete case has more degrees of freedom than the continuous case.

Theorem 1 characterizes the discrete Wasserstein distance $\mathcal{W}_1$ in terms of the thermodynamic and kinetic costs associated with Markovian dynamics. In Appendix D 6, we show that Theorem 1 has some useful corollaries that not only provide alternative expressions for $\mathcal{W}_1$, but also lead to stringent bounds for thermodynamic speed limits.

Using other combinations of irreversible entropy production, pseudo entropy production, and dynamical activity, the discrete Wasserstein distance $\mathcal{W}_1$ can be expressed in similar variational forms.

V. QUANTUM GENERALIZATION

We next generalize our framework to the quantum case. We first briefly introduce quantum thermodynamics of Markovian open quantum dynamics described by the Lindblad equations and define a quantum analog of dynamical state mobility. Then, we define a quantum Wasserstein distance and derive analogous variational formulas for the quantum Wasserstein distance in terms of thermodynamic cost.

A. Markovian open quantum dynamics

We consider a finite-dimensional open quantum system, which is attached to single or multiple thermal reservoirs. In the weak-coupling limit, the time evolution of the reduced density matrix can be described by the Lindblad master equation [121]:

$$\dot{\varrho}_t = \mathcal{L}_t(\varrho_t) := -i[H_t, \varrho_t] + \sum_k \mathcal{D}[L_k(t)]\varrho_t, \qquad (89)$$

where H_t is the time-dependent Hamiltonian, $\mathcal{D}$ is the dissipator given by $\mathcal{D}[L]\varrho := L\varrho L^\dagger - \{L^\dagger L, \varrho\}/2$, and $L_k(t)$ are jump operators. $[\circ, \star]$ and $\{\circ, \star\}$ denote the commutator and anticommutator of the two operators, respectively. Hereafter, we set the Planck constant to unity: $\hbar = 1$. To guarantee thermodynamically consistent dynamics, we assume that the jump operators satisfy the local detailed-balance condition [122,123]; that is, they come in pairs (k, k') such that

$$L_k(t) = e^{s_k(t)/2} L_{k'}(t)^\dagger, \qquad (90)$$

where $s_k(t) = -s_{k'}(t)$ denotes the entropy change in the environment due to the jump operator $L_k(t)$. In the case of a single reservoir at inverse temperature β, we can write $s_k(t) = \beta\omega_k(t)$, where $\omega_k(t)$ is the energy change associated with the kth jump.

B. Entropy production and dynamical activity

Given the previous setup, we can now introduce quantum entropy production and dynamical activity. Similar to the classical case, irreversible entropy production can be defined as the sum of entropy changes in the system and environment as

$$\Sigma_\tau := \Delta S_{\mathrm{sys}} + \Delta S_{\mathrm{env}}, \qquad (91)$$

where $\Delta S_{\mathrm{sys}} := S(\varrho_\tau) - S(\varrho_0)$ is the difference in the von Neumann entropy $S(\varrho) = -\mathrm{tr}\{\varrho \ln \varrho\}$ of the system

and ΔS_{env} denotes environmental entropy production, given by [122,123]

$$\Delta S_{\text{env}} := \int_0^\tau \sum_k \text{tr}\{L_k(t)\varrho_t L_k^\dagger(t)\} s_k(t) dt. \quad (92)$$

With this definition, we can prove that Σ_τ is always non-negative, which implies the second law of thermodynamics. For the case of a single reservoir and the jump operators that characterize transitions between energy eigenstates [i.e., $[L_k(t), H_t] = \omega_k(t) L_k(t)$], the entropy production of the environment reduces exactly to the conventional form [124]

$$\Delta S_{\text{env}} = -\beta \int_0^\tau \text{tr}\{H_t \dot{\varrho}_t\} dt. \quad (93)$$

Quantum dynamical activity can be analogously defined as in the classical case. The frequency of jumps at time t can be quantified as

$$a_t := \sum_k \text{tr}\{L_k(t)\varrho_t L_k^\dagger(t)\}, \quad (94)$$

and the average total number of jumps can be calculated as $\mathcal{A}_\tau := \int_0^\tau a_t dt$. Quantum dynamical activity characterizes the thermalization rate of thermodynamic processes. In addition, it is shown that quantum dynamical activity constrains the precision of generic counting observables and their first passage time in quantum jump processes [109,118].

It is convenient to alternatively express entropy production and dynamical activity defined in Eqs. (91) and (94), respectively. Let $\varrho_t = \sum_x p_x(t)|x_t\rangle\langle x_t|$ be the spectral decomposition of the density matrix ϱ_t. We then define transition rates between eigenbasis as $w_k^{xy}(t) := |\langle x_t|L_k(t)|y_t\rangle|^2 \geq 0$. Notice that $w_k^{xy}(t) = e^{s_k(t)} w_{k'}^{yx}(t)$. Taking the time derivative of $p_x(t) = \langle x_t|\varrho_t|x_t\rangle$, we obtain the following master equation for the distribution $\{p_x(t)\}$:

$$\dot{p}_x(t) = \sum_k \sum_{y(\neq x)} [w_k^{xy}(t) p_y(t) - w_k^{yx}(t) p_x(t)]. \quad (95)$$

Analogous to the classical case, we define

$$a_k^{xy}(t) := w_k^{xy}(t) p_y(t), \quad (96)$$

$$j_k^{xy}(t) := w_k^{xy}(t) p_y(t) - w_{k'}^{yx}(t) p_x(t). \quad (97)$$

Using these probability currents, we can write the master equation as

$$\dot{p}_x(t) = \sum_k \sum_{y(\neq x)} j_k^{xy}(t). \quad (98)$$

We emphasize that the classical-like master equation (95) is rigorously derived from Eq. (89). This equation is introduced only for the proof convenience of several properties that the dynamics (89) possesses. After some simple manipulations, we can prove that the entropy production rate $\sigma_t := \dot{\Sigma}_t$ can be analytically expressed as (see Appendix E 1 for the proof)

$$\sigma_t = \frac{1}{2} \sum_k \sum_{x,y} j_k^{xy}(t) \ln \frac{w_k^{xy}(t) p_y(t)}{w_{k'}^{yx}(t) p_x(t)}. \quad (99)$$

Besides, plugging the spectral decomposition of ϱ_t and inserting $\mathbb{1} = \sum_x |x_t\rangle\langle x_t|$ into Eq. (94), the dynamical activity rate can also be expressed as

$$a_t = \sum_k \sum_{x,y} w_k^{xy}(t) p_y(t) = \sum_k \sum_{x,y} a_k^{xy}(t). \quad (100)$$

Note that both σ_t and a_t, which can be written in terms of the transition rates of the master equation (95), are the entropy production and dynamical activity rates associated with the Lindblad dynamics (89), respectively.

C. Quantum dynamical state mobility

Analogous with the classical case, the quantum analog of dynamical state mobility can be defined as

$$m_t := \frac{1}{2} \sum_k e^{-s_k(t)/2} \langle L_k(t)^\dagger, [\![\varrho_t]\!]_{s_k(t)} \{\mathcal{P}_t[L_k(t)^\dagger]\}\rangle, \quad (101)$$

where $\langle X, Y \rangle := \text{tr}\{X^\dagger Y\}$ denotes the scalar inner product, $\mathcal{P}_t$ is a superoperator given by $\mathcal{P}_t[X] := X - \sum_x \langle x_t|X|x_t\rangle |x_t\rangle\langle x_t|$, and the tilted operator $[\![\phi]\!]_\theta(X)$ is defined for arbitrary density matrix ϕ, real number θ, and linear operator X as

$$[\![\phi]\!]_\theta(X) := e^{-\theta/2} \int_0^1 e^{\theta u} \phi^u X \phi^{1-u} du. \quad (102)$$

The quantum kinetic cost can be analogously defined as

$$\mathcal{M}_\tau := \int_0^\tau m_t dt. \quad (103)$$

From the mathematical definition in Eq. (101), interpreting the term m_t as a kinetic term may not be intuitive. In the following, we provide the physical interpretations of m_t from two perspectives.

First, by focusing on the master equation of the distribution $\{p_x(t)\}$, we can show that m_t is equal to the dynamical state mobility associated with Markovian jump dynamics (95):

$$m_t = \sum_k \sum_{x>y} \frac{a_k^{xy}(t) - a_{k'}^{yx}(t)}{\ln a_k^{xy}(t) - \ln a_{k'}^{yx}(t)}. \quad (104)$$

Note that, by applying the inequality (38) to Eq. (104), we can readily prove that m_t is upper bounded by the dynamical activity:

$$
\begin{aligned}
m_t &\leq \sum_k \sum_{x > y} \frac{a_k^{xy}(t) + a_{k'}^{yx}(t)}{2} \\
&= \frac{1}{2} \sum_k \sum_{x \neq y} a_k^{xy}(t) \\
&\leq \frac{a_t}{2}.
\end{aligned}
\tag{105}
$$

Second, let us consider the case of a single reservoir, in which the jump operators satisfy $[L_k(t), H_t] = \omega_k(t) L_k(t)$. In this case, the thermal state $\varrho_t^{\mathrm{eq}} := e^{-\beta H_t}/\mathrm{tr}\, e^{-\beta H_t}$ is always the instantaneous equilibrium state [i.e., $\mathcal{L}_t(\varrho_t^{\mathrm{eq}}) = 0$]. Note that the Lindblad master equation (89) can be rewritten as [22]

$$
\dot{\varrho}_t = \mathcal{U}_{\varrho_t}(t, f_t) + \mathcal{O}_{\varrho_t}(t, f_t),
\tag{106}
$$

where $f_t := -\ln \varrho_t + \ln \varrho_t^{\mathrm{eq}}$ is the quantum thermodynamic force and $\mathcal{U}_\phi(t, X)$ and $\mathcal{O}_\phi(t, X)$ are time-dependent super-operators defined, respectively, as

$$
\mathcal{U}_\phi(t, X) := i\beta^{-1}[X, \phi],
\tag{107}
$$

$$
\mathcal{O}_\phi(t, X) := \frac{1}{2} \sum_k e^{-s_k(t)/2}[L_k(t), [\![\phi]\!]_{s_k(t)}\{[L_k(t)^\dagger, X]\}].
\tag{108}
$$

The superoperators $\mathcal{U}$ and $\mathcal{O}$ characterize the unitary and dissipative parts of Lindblad dynamics, respectively. They linearly relate the rate of the density matrix to the thermodynamic force. In addition, the entropy production rate can be written in a quadratic form of the thermodynamic force as [22]

$$
\sigma_t = \langle f_t, \mathcal{O}_{\varrho_t}(t, f_t) \rangle.
\tag{109}
$$

Since Eqs. (106) and (109) are analogous to Eqs. (40) and (42) in the classical case, respectively, the super-operator $\mathcal{O}$ can be regarded as a quantum Onsager-like superoperator.

We now investigate the relationship between m_t and the Onsager-like superoperator $\mathcal{O}$. To this end, we employ the vectorization of a linear operator X as

$$
X = \sum_{i,j} x_{ij} |i\rangle\langle j| \to |X\rangle\!\rangle = \sum_{i,j} x_{ij} |i\rangle \otimes |j\rangle.
\tag{110}
$$

Using this representation, we can rewrite the Lindblad master equation (106) as

$$
|\dot{\varrho}_t\rangle\!\rangle = \mathsf{U}_t |f_t\rangle\!\rangle + \mathsf{O}_t |f_t\rangle\!\rangle,
\tag{111}
$$

where the linear matrices U_t and O_t are defined as

$$
\mathsf{U}_t := i\beta^{-1}(\mathbb{1} \otimes \varrho_t^\top - \varrho_t \otimes \mathbb{1}),
\tag{112}
$$

$$
\mathsf{O}_t := \frac{1}{2} \sum_k e^{-s_k(t)} \int_0^1 e^{s_k(t)u} \mathsf{O}_k(t, u)\, du.
\tag{113}
$$

Here, $\top$ denotes the matrix transpose, and $\mathsf{O}_k(t, x)$ is given by

$$
\begin{aligned}
&\mathsf{O}_k(t, u) \\
&:= L_k(t)\varrho_t^u L_k(t)^\dagger \otimes (\varrho_t^{1-u})^\top + \varrho_t^u \otimes [L_k(t)^\dagger \varrho_t^{1-u} L_k(t)]^\top \\
&\quad - L_k(t)\varrho_t^u \otimes [L_k(t)^\dagger \varrho_t^{1-u}]^\top - \varrho_t^u L_k(t)^\dagger \otimes [\varrho_t^{1-u} L_k(t)]^\top.
\end{aligned}
\tag{114}
$$

Note that U_t and O_t are the matrix representations of the superoperators $\mathcal{U}$ and $\mathcal{O}$, respectively. Simple algebraic calculations show that the term m_t can be related to the diagonal elements of the Onsager-like matrix O_t as (see Appendix E 2 for the proof)

$$
m_t = \frac{1}{2} \sum_x \langle x_t | \otimes |x_t\rangle^\top \mathsf{O}_t |x_t\rangle \otimes \langle x_t|^\top.
\tag{115}
$$

In this sense, m_t can be regarded as a quantum kinetic term.

D. Quantum optimal transport distance and thermodynamic interpretation

Although the classical Wasserstein distance is well formulated and studied, its quantum version remains under development. Several quantum generalizations of the Wasserstein distance have been proposed [22,120,125–129]. However, defining the quantum L^1-Wasserstein distance unambiguously by directly generalizing the classical distance has been shown to be impossible [130].

By a naive extension using quantum coupling, a quantum optimal transport distance can be defined as

$$
W_q(\varrho^A, \varrho^B) := \min_{\varrho^{AB} \in \Pi(\varrho^A, \varrho^B)} \mathrm{tr}\{C\varrho^{AB}\},
\tag{116}
$$

where the coupling $\Pi(\varrho^A, \varrho^B)$ denotes the set of density matrices ϱ^{AB} defined over the Hilbert space $\mathcal{H} \otimes \mathcal{H}$ and satisfy $\mathrm{tr}_B \varrho^{AB} = \varrho^A$ and $\mathrm{tr}_A \varrho^{AB} = \varrho^B$ and C is a cost matrix that must be properly chosen to guarantee that W_q is a distance. In the classical case, the total variation distance is a classical Wasserstein distance with an appropriate choice of the cost matrix C. It is, thus, natural to ask whether a cost matrix C exists such that the quantum version of the total variation distance (i.e., the trace distance) can be represented as a quantum Wasserstein distance defined in

Eq. (116). Unfortunately, Ref. [131] shows that the trace distance cannot be expressed in terms of this type of Wasserstein distance. In other words, for any choice of the cost matrix C, density matrices ϱ^A and ϱ^B always exist such that the distance W_q defined in Eq. (116) differs from the trace distance:

$$W_q(\varrho^A, \varrho^B) \neq \frac{1}{2}\|\varrho^A - \varrho^B\|_1 =: \mathcal{T}(\varrho^A, \varrho^B). \quad (117)$$

Our aim is to relate quantum optimal transport distances and dissipation in Lindblad dynamics. Note that Lindblad dynamics consist of a nondissipative unitary part and dissipative Lindblad part. Both parts jointly contribute to the time evolution of the system's density matrix. In the vanishing coupling limit, irreversible entropy production becomes zero, whereas the distance $W_q(\varrho_0, \varrho_\tau)$ may be positive since $\varrho_0 \neq \varrho_\tau$. Therefore, relating dissipation to the optimal transport distances defined in the current form (116) is impossible. Inspired by the dissipative structure of Lindblad dynamics, we define the following distance:

$$\mathcal{W}_q(\varrho^A, \varrho^B) := \frac{1}{2} \min_{V^\dagger V = \mathbb{1}} \|V\varrho^A V^\dagger - \varrho^B\|_1. \quad (118)$$

Here, the minimum is over all possible unitaries V. Intuitively, the distance $\mathcal{W}_q$ characterizes the state difference induced by the dissipative Lindblad part. Thus, it is expected to be relevant to dissipation. Note that in the zero-dissipation limit (i.e., the system is unitarily evolved), this distance also vanishes. Although the distance $\mathcal{W}_q$ is defined in a variational form, it can be analytically calculated using the eigenvalues of the density matrices. Interestingly, it becomes exactly the classical Wasserstein distance between the eigenvalue distributions:

$$\mathcal{W}_q(\varrho^A, \varrho^B) = \frac{1}{2}\sum_x |p_x^A - p_x^B| = \mathcal{T}(p^A, p^B), \quad (119)$$

where $\{p_x^A\}$ and $\{p_x^B\}$ are increasing eigenvalues of ϱ^A and ϱ^B, respectively (see Appendix E 3 for the proof). For this reason, hereafter, $\mathcal{W}_q$ is referred to as the quantum Wasserstein distance. Evidently, this distance satisfies the triangle inequality. However, it is a pseudo metric [i.e., $\mathcal{W}_q(\varrho^A, \varrho^B) = 0$ for $\varrho^A \neq \varrho^B$ is possible]. This originates from our goal of relating the defined distance to dissipation in Lindblad dynamics.

For the quantum Wasserstein distance previously defined, we provide the following thermodynamic interpretation.

Theorem 2: The quantum Wasserstein distance can be written in the following variational form:

$$\mathcal{W}_q(\varrho^A, \varrho^B) = \min_{\mathcal{L}_t} \int_0^\tau \sqrt{\sigma_t m_t}\, dt \quad (120)$$

$$= \min_{\mathcal{L}_t} \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \quad (121)$$

Here, the minimum is taken over all superoperators $\{\mathcal{L}_t\}_{0 \le t \le \tau}$ that satisfy the Lindblad master equation (89) with boundary conditions $\varrho_0 = \varrho^A$ and $\varrho_\tau = \varrho^B$.

Theorem 2 is the second central result, and its sketch proof is given in the following. Interestingly, Theorem 2 has the same structure as Theorem 1 in the classical case. This implies a universal relationship between the optimal transport distances and dissipation in classical and quantum discrete systems.

Proof.—We briefly describe the proof strategy; for a detailed derivation, see Appendix E 4. We first prove that the inequalities

$$\mathcal{W}_q(\varrho^A, \varrho^B) \le \int_0^\tau \sqrt{\sigma_t m_t}\, dt \le \sqrt{\Sigma_\tau \mathcal{M}_\tau} \quad (122)$$

hold for any Markovian open quantum dynamics and then construct a specific process that simultaneously attains all the equalities of Eq. (122). The inequalities in Eq. (122) can be proved similarly as in the classical case. To construct the dynamics that can achieve the equalities, we first construct a classical Markov jump process that transforms distribution p^A into p^B and satisfies

$$\mathcal{T}(p^A, p^B) = \int_0^\tau \sqrt{\sigma_t m_t}\, dt = \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \quad (123)$$

Here, $\{p_x^A\}$ and $\{p_x^B\}$ are increasing eigenvalues of ϱ^A and ϱ^B, respectively. Subsequently, we construct Lindblad dynamics based on this classical jump process such that the dynamics transforms density matrix ϱ^A into ϱ^B, and the quantities σ_t and m_t are identical to those in the classical jump process. We can verify that this quantum dynamics attains the equalities of Eq. (122). ∎

Similar to the classical case, the quantum Wasserstein distance can also be determined through the entropy production and dynamical activity associated with Markovian quantum dynamics, which is stated in Corollary 9 in Appendix E 5.

VI. APPLICATIONS FOR THERMODYNAMIC INTERPRETATION OF OPTIMAL TRANSPORT

In this section, we present applications for our central results, namely, Theorems 1 and 2. Specifically, we show that these variational formulas lead to stringent bounds for thermodynamic speed limits and information erasure at arbitrary temperatures.

A. Classical and quantum thermodynamic speed limits

The speed of state transformation in any system cannot be made arbitrarily fast because of physical constraints. This fact leads to a natural question: What is the ultimate limit for state transformation?

This question sparked a lot of research and gave rise to the concept of speed limits.

Precisely speaking, speed limits impose lower bounds on the operational time required for evolving a system from a given state to a target one. Originally, speed limits were derived for closed quantum systems, inspired by the Heisenberg time-energy uncertainty principle [132]. One of the celebrated results is the Mandelstam-Tamm bound, which applies to closed quantum systems and takes the following form:

$$\tau \geq \frac{\mathcal{B}(\varrho_0, \varrho_\tau)}{\langle \Delta H \rangle_\tau}, \tag{124}$$

where $\mathcal{B}(\varrho, \sigma) := \arccos \text{tr}\{|\sqrt{\varrho}\sqrt{\sigma}|\}$ is the Bures angle and $(\Delta H_t)^2 := \text{tr}\{H_t^2 \varrho_t\} - \text{tr}\{H_t \varrho_t\}^2$ is the energy fluctuation. Equation (124) implies that the speed of state transformation in closed quantum systems is constrained by the fluctuation of energy. Various types of speed limits are subsequently generalized for open quantum and classical systems [132–146] (see Ref. [147] for a comprehensive review).

Although several versions of classical and quantum speed limits exist for open systems, here we aim to develop *thermodynamic* bounds that satisfy two conditions: (i) They should be tight (i.e., for generic initial and final states, a configuration of the system always exists that transforms these states and saturates the bounds), and (ii) they should be physically interpretable (i.e., all quantities appearing in the bound are physically meaningful). In the following, we derive these thermodynamic speed limits from the variational formulas for both classical and quantum cases.

1. Classical case

We consider a discrete classical system modeled by a Markov jump process [Eq. (20)]. The system is driven by thermodynamic forces and evolves according to the laws of thermodynamics. Intuitively, to achieve fast transformation, we must pay some costs. In the following, we derive fundamental bounds on the operational time that is required to evolve the system's distribution to the target one.

Let $\mathcal{G}(V, E)$ be the underlying topology of the jump process (i.e., the graph connectivity that determines whether the transition between two states is allowed). Then, we can define the corresponding Wasserstein distance based on the graph $\mathcal{G}$. According to Theorem 1, we have

$$\mathcal{W}_1(p_0, p_\tau) = \min_{\mathsf{W}_t} \int_0^\tau \sqrt{\sigma_t m_t} dt = \min_{\mathsf{W}_t} \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \tag{125}$$

Since the system dynamics considered here is one of the admissible dynamics that transform p_0 into p_τ, the following inequalities follow immediately from the equality (125):

$$\mathcal{W}_1(p_0, p_\tau) \leq \int_0^\tau \sqrt{\sigma_t m_t} dt = \tau \langle \sqrt{\sigma m} \rangle_\tau \tag{126}$$

$$\leq \tau \sqrt{\langle \sigma \rangle_\tau \langle m \rangle_\tau}. \tag{127}$$

Consequently, we obtain lower bounds on the operational time in terms of the Wasserstein distance, thermodynamic cost, and kinetic cost as follows:

$$\tau \geq \frac{\mathcal{W}_1(p_0, p_\tau)}{\langle \sqrt{\sigma m} \rangle_\tau} \geq \frac{\mathcal{W}_1(p_0, p_\tau)}{\sqrt{\langle \sigma \rangle_\tau \langle m \rangle_\tau}}. \tag{128}$$

Equation (128) implies that both irreversible entropy production and state mobility jointly constrain the speed of state transformation. Using Corollary 7 in Appendix D 6 and following the same procedure, we also obtain similar but tighter bounds in terms of time-averaged entropy production and dynamical activity as

$$\tau \geq \frac{2\mathcal{W}_1(p_0, p_\tau)}{\langle \sigma \Phi(\sigma/2a)^{-1} \rangle_\tau} \geq \frac{2\mathcal{W}_1(p_0, p_\tau)}{\langle \sigma \rangle_\tau \Phi(\langle \sigma \rangle_\tau / 2\langle a \rangle_\tau)^{-1}}, \tag{129}$$

where $\Phi(x)$ is the inverse function of $x \tanh(x)$. Equations (128) and (129) are our new thermodynamic speed limits for classical Markov jump processes.

Some remarks are in order. First, the thermodynamic speed limits in Eqs. (128) and (129) are tight and saturable. More specifically, for generic initial and final distributions, we can always construct dynamics that satisfy the global detailed-balance condition and transform the initial distribution into the final one in a time duration equal to that of the lower bounds.

Second, our bounds are tight for arbitrary temperatures, even in the zero-temperature limit. Since $\langle \sigma \rangle_\tau = O(\beta)$, irreversible entropy production becomes infinite as $\beta \to +\infty$, whereas dynamical activity remains finite [i.e., $\langle a \rangle_\tau = O(1)$]. Nevertheless, we show in the following that our bounds remain useful in this low-temperature limit. Indeed, in the $\beta \to +\infty$ limit, bound (129) reduces to a nontrivial inequality $\tau \geq \mathcal{W}_1(p_0, p_\tau)/\langle a \rangle_\tau$. In addition, bound (128) also remains finite, because we can prove that $\beta \langle m \rangle_\tau$ does not diverge, in general. To this end, we assume that the energy levels are nondegenerate and the system is typically driven far from the instantaneous equilibrium. Since $m_{xy}(t)$ can be calculated as

$$\beta m_{xy}(t) = \frac{\beta[a_{xy}(t) - a_{yx}(t)]}{\ln a_{xy}(t) - \ln a_{yx}(t)}$$

$$= \frac{a_{xy}(t) - a_{yx}(t)}{\beta^{-1}[\ln p_y(t) - \ln p_x(t)] + \varepsilon_y(t) - \varepsilon_x(t)}, \tag{130}$$

we have

$$\beta m_{xy}(t) \xrightarrow{\beta \to +\infty} \frac{a_{xy}(t) - a_{yx}(t)}{\varepsilon_y(t) - \varepsilon_x(t)}, \tag{131}$$

which remains finite. Therefore, the term $\beta\langle m\rangle_\tau$ does not diverge in the zero-temperature limit $\beta \to +\infty$.

Third, we compare our results with existing bounds in the literature. In Ref. [16], a classical speed limit is obtained for Markov jump processes, which reads

$$\tau \geq \frac{\mathcal{T}(p_0, p_\tau)}{\sqrt{\langle\sigma\rangle_\tau \langle a\rangle_\tau/2}}. \tag{132}$$

Since $\mathcal{W}_1(p_0, p_\tau) \geq \mathcal{T}(p_0, p_\tau)$ and $m_t \leq a_t/2$ for all times, our speed limits in Eq. (128) are stronger than those in Eq. (132). Our bounds also suggest that the conventional bound (132) can be asymptomatically saturated only when $\mathcal{W}_1(p_0, p_\tau) = \mathcal{T}(p_0, p_\tau)$ (e.g., when the underlying graph is fully connected) and the system is always near the instantaneous equilibrium. In Refs. [24,26], another thermodynamic speed limit, which is tighter than the conventional bound (132), is derived as

$$\tau \geq \frac{2\mathcal{T}(p_0, p_\tau)}{\langle\sigma\rangle_\tau \Phi(\langle\sigma\rangle_\tau/2\langle a\rangle_\tau)^{-1}}. \tag{133}$$

Since $\mathcal{W}_1(p_0, p_\tau) \geq \mathcal{T}(p_0, p_\tau)$, our bound (129) is stronger than bound (133). The essential difference is that our bounds consider the topology of the jump process, whereas the conventional bounds do not.

2. Quantum case

Next, we consider an open quantum system described by the Markovian Lindblad master equation [Eq. (89)]. Following the same procedure as in the classical case, we derive stringent thermodynamic bounds on the operational time required to transform the initial density matrix into the final one.

From Theorem 2,

$$\mathcal{W}_q(\varrho_0, \varrho_\tau) = \min_{\mathcal{L}_t}\int_0^\tau \sqrt{\sigma_t m_t}\,dt = \min_{\mathcal{L}_t}\sqrt{\Sigma_\tau \mathcal{M}_\tau}, \tag{134}$$

we analogously obtain the following inequalities:

$$\mathcal{W}_q(\varrho_0, \varrho_\tau) \leq \int_0^\tau \sqrt{\sigma_t m_t}\,dt = \tau\langle\sqrt{\sigma m}\rangle_\tau \tag{135}$$

$$\leq \tau\sqrt{\langle\sigma\rangle_\tau \langle m\rangle_\tau}. \tag{136}$$

Consequently, we arrive at the following bounds on the operational time:

$$\tau \geq \frac{\mathcal{W}_q(\varrho_0, \varrho_\tau)}{\langle\sqrt{\sigma m}\rangle_\tau} \geq \frac{\mathcal{W}_q(\varrho_0, \varrho_\tau)}{\sqrt{\langle\sigma\rangle_\tau \langle m\rangle_\tau}}. \tag{137}$$

Equation (137) implies that the speed of state transformation in open quantum systems is constrained by irreversible entropy production and dynamical state mobility. Notably, it has the same form as the classical bound (128), indicating the unification of our results. Exploiting Corollary 9 in Appendix E 5 and repeating the same procedure yield other speed limits in terms of entropy production and dynamical activity, which read

$$\tau \geq \frac{2\mathcal{W}_q(\varrho_0, \varrho_\tau)}{\langle\sigma\Phi(\sigma/2a)^{-1}\rangle_\tau} \geq \frac{2\mathcal{W}_q(\varrho_0, \varrho_\tau)}{\langle\sigma\rangle_\tau \Phi(\langle\sigma\rangle_\tau/2\langle a\rangle_\tau)^{-1}}. \tag{138}$$

Equations (137) and (138) are the new quantum thermodynamic speed limits. Remarkably, these thermodynamic speed limits are tight and saturable. In other words, for any pair of generic initial and final states, a combination of Hamiltonian and jump operators always exists that attains the lower bound of the operational time. Moreover, they are useful even in the zero-temperature limit. The bounds imply that both the thermodynamic and kinetic costs play a crucial role in the change speed of open quantum systems.

We discuss the relevance of our results to previous studies. In Ref. [22], a thermodynamic speed limit is derived for Markovian open quantum dynamics and is given by

$$\tau \geq \frac{\mathcal{W}_q(\varrho_0, \varrho_\tau)}{\sqrt{\langle\sigma\rangle_\tau \langle a\rangle_\tau/2}}. \tag{139}$$

According to Eq. (105), we have $m_t \leq a_t/2$ for all t. Therefore, bound (139) is looser than the new bound (137). In addition, since $\Phi(x) \geq \sqrt{x}$ for all $x \geq 0$, the new bound (138) is also stronger than the conventional bound (139). In Ref. [18], another thermodynamic speed limit in terms of trace distance is derived for open quantum systems. Since the metrics used to measure the distance between quantum states in these bounds are different (i.e., the Wasserstein distance in our study and the trace distance in Ref. [18]), they cannot be directly compared. Nonetheless, by exploiting the dynamical state mobility introduced in this study, we can derive a similar but tighter speed limit in terms of the trace distance. The detailed form of this speed limit is presented in Appendix F 1. However, it is worth noting that the attainability of this bound is unclear.

B. Finite-time Landauer principle

The Landauer principle [90] implies that erasing information is always accompanied by a thermodynamic cost. More specifically, the thermodynamic cost required to erase a classical bit is at least $T\ln 2$, where T is the environment temperature. The lower bound $T\ln 2$ (referred to as the Landauer bound) not only plays a fundamental role in the thermodynamics of information and computation [4,41–44], it also resolves the paradox of Maxwell's demon [148].

Various classical and quantum platforms [81,149–153] have experimentally confirmed that the Landauer bound can be achieved in the slow quasistatic limit. However, practical computing requires fast memory erasure in a short time and, thus, in general, consumes a thermodynamic cost far beyond the Landauer bound. This background strongly motivates researchers to develop finite-time generalizations of the Landauer bound, which capture finite-time corrections and can better predict the erasure cost. Although several finite-time bounds have been developed for both classical and quantum discrete systems [30–32], the attainability of these bounds remains unclear. Moreover, these bounds have looser predictive power in the low-temperature regime. In the following, we attempt to derive finite-time bounds that are tight for arbitrary temperatures.

Before presenting the new bounds, we first describe the generic setup of information erasure for both classical and quantum cases. We consider a finite-dimensional discrete system attached to a thermal reservoir at temperature T. Information is encoded in the system state and subsequently erased by controlling the classical energy levels or the quantum Hamiltonian and driving the system toward its ground state. The erasure protocol should work for an arbitrary initial state; that is, any initial state should be reset close to the ground state in a finite time τ. This erasure process leads to a change in system entropy, which must be compensated for by the heat dissipated into the reservoir. Because we are interested in the average thermodynamic cost associated with the erasure protocol, considering the maximally mixed state as the initial state is convenient. Roughly speaking, the reasons for this are that the maximally mixed state is sufficient to understand the average dissipated heat of the erasure process for all initial states and that if a protocol can reliably reset the system from the maximally mixed state, then it does so also for an arbitrary state. A detailed discussion is given in the following.

1. Classical case

We consider an information erasure process using a d-state classical system, the dynamics of which is governed by the master equation. The transitions between states are mediated by a single thermal reservoir at temperature T. The system state is characterized by the probability distribution, which encodes information we want to erase. The energy levels are controlled according to a fixed protocol such that the system is always driven toward the ground state $p_* = [1, 0, \ldots, 0]^\top$, irrespective of the initial state.

Here, we explain why the initial state should be set to the uniform distribution $\bar{p} = [1/d, \ldots, 1/d]^\top$. First, let $\mathcal{Q}(p_0)$ be the heat dissipation of erasure for the initial distribution p_0. Then, due to the linearity of the master

equation and $\mathcal{Q}(\cdot)$, the average dissipation can be calculated as

$$\mathbb{E}[\mathcal{Q}(p_0)] = \mathcal{Q}(\mathbb{E}[p_0]) = \mathcal{Q}(\bar{p}), \qquad (140)$$

where $\mathbb{E}[\cdot]$ denotes the average over all possible initial distributions. Equation (140) implies that investigating the case with initial distribution $p_0 = \bar{p}$ is sufficient to understand the average dissipation. Second, let $\Lambda_\tau = \vec{T} \exp(\int_0^\tau \mathsf{W}_t dt)$ be the map that represents the erasure process, that is, $\Lambda_\tau p_0 = p_\tau$. We can then prove that if the uniform distribution can be erased within error $\delta > 0$ (i.e., $\|\Lambda_\tau \bar{p} - p_*\|_F \leq \delta$), the following inequality holds for arbitrary initial distribution p_0 (see Appendix F 2 for the proof):

$$\|\Lambda_\tau p_0 - p_*\|_F \leq \sqrt{2d\delta}. \qquad (141)$$

Equation (141) indicates that if a protocol can erase the uniform distribution, it can reliably do so also for arbitrary initial states.

We can now present the new bound. Let p_τ be the final distribution for the case $p_0 = \bar{p}$, and let $\epsilon := \mathcal{T}(p_\tau, p_*) = |1 - p_1(\tau)|$ be the erasure error, which should be sufficiently small. From Eq. (79), the heat dissipation is lower bounded by system entropy production and a finite-time correction term as

$$Q \geq -T\Delta S_{\mathrm{sys}} + \frac{\mathcal{W}_1(p_0, p_\tau)^2}{\tau \beta \langle m \rangle_\tau}. \qquad (142)$$

Equation (142) is regarded as the finite-time Landauer principle for classical systems. The bound is tight and can be saturated for arbitrary temperatures, even in the zero-temperature limit. As shown in the previous section, the term $\beta \langle m \rangle_\tau$ remains finite even when $\beta \to \infty$ (i.e., $T \to 0$). Therefore, bound (142) is useful for arbitrary temperatures. By contrast, the conventional Landauer bound becomes trivial in the low-temperature regime (i.e., $Q \geq 0$). We also note that bound (142) is tighter than the following bound:

$$Q \geq -T\Delta S_{\mathrm{sys}} + \frac{\mathcal{T}(p_0, p_\tau)^2}{\tau \beta \langle a \rangle_\tau / 2}, \qquad (143)$$

which is obtained from the conventional speed limit (132).

Bound (142) can be simplified by including the erasure error. To this end, we further bound the terms in Eq. (142) from below as

$$-\Delta S_{\mathrm{sys}} = \ln d - S(p_\tau) \geq \ln d - h(\epsilon), \qquad (144)$$

$$\mathcal{W}_1(p_0, p_\tau) \geq \mathcal{T}(p_0, p_\tau) \geq |1 - 1/d - \epsilon|, \qquad (145)$$

where $h(\epsilon) := -\epsilon \ln[\epsilon/(d-1)] - (1-\epsilon)\ln(1-\epsilon) \geq 0$ is a function of ϵ that vanishes as $\epsilon \to 0$. Consequently, we obtain the following bound on the average heat dissipation:

$$Q \geq T[\ln d - h(\epsilon)] + \frac{(1 - 1/d - \epsilon)^2}{\tau\beta\langle m\rangle_\tau}. \qquad (146)$$

Equation (146) imposes a lower bound on heat dissipation in terms of the operational time and erasure error. In the limit of perfect erasure (i.e., $\epsilon \to 0$), a simple bound can be derived:

$$Q \geq T \ln d + \frac{(1 - 1/d)^2}{\tau\beta\langle m\rangle_\tau}. \qquad (147)$$

For slow erasure (i.e., $\tau\beta\langle m\rangle_\tau \gg 1$), the second term in the lower bound vanishes. Thus, Eq. (147) recovers the conventional Landauer bound for the $d = 2$ case. By contrast, in the fast-erasure limit (i.e., $\tau\beta\langle m\rangle_\tau \ll 1$), this correction term becomes dominant, implying that fast erasure is accompanied by a thermodynamic cost far beyond the Landauer cost.

If we consider dynamical activity instead of dynamical state mobility, we can obtain another finite-time Landauer bound. By transforming the speed limit (129), we can show that heat dissipation is lower bounded by the Wasserstein distance and dynamical activity as

$$\frac{Q}{T} \geq -\Delta S_{\text{sys}} + 2\mathcal{W}_1(p_0, p_\tau)\tanh^{-1}\left(\frac{\mathcal{W}_1(p_0, p_\tau)}{\tau\langle a\rangle_\tau}\right). \quad (148)$$

This new bound is always tighter than the bound reported in Ref. [32], which uses the total variation distance to quantify the distance between probability distributions. In general, bound (148) can be either stronger or looser than bound (142).

2. Quantum case

Here, we consider a quantum process of erasing information. The erasure process is implemented using a controllable d-dimensional qudit system, which is attached to a thermal reservoir at temperature T. The density matrix of the qudit encodes the information we want to erase and then is driven toward the ground state $\varrho_* = |0\rangle\langle 0|$ by controlling the Hamiltonian.

Analogous to the classical case, the initial state is conveniently set to the maximally mixed state $\bar{\varrho} = \mathbb{1}/d$. This is because assigning the maximally mixed state to the initial state is sufficient to understand the average heat dissipation in the quantum case. In addition, we can show that if an erasure protocol can erase the maximally mixed state, it can reliably do so also for arbitrary initial states. More specifically, let $\Lambda_\tau(\cdot) = \vec{T}\exp(\int_0^\tau \mathcal{L}_t dt)(\cdot)$ be the quantum map that describes the erasure process [i.e., $\Lambda_\tau(\varrho_0) = \varrho_\tau$]. Then, if the maximally mixed state can be erased within error $\delta > 0$ [i.e., $\|\Lambda_\tau(\bar{\varrho}) - \varrho_*\|_F \leq \delta$], the erasure error for arbitrary initial state ϱ_0 can be upper bounded as follows [31]:

$$\|\Lambda_\tau(\varrho_0) - \varrho_*\|_F \leq \sqrt{2d\delta}. \qquad (149)$$

Equation (149) provides insight into the reliability of the erasure protocol by verifying the $\varrho_0 = \bar{\varrho}$ case.

Next, we present the finite-time bound for quantum information erasure. Because $\varrho_0 = \mathbb{1}/d$, the Wasserstein distance coincides with the trace distance [i.e., $\mathcal{W}_q(\varrho_0, \varrho_\tau) = \mathcal{T}(\varrho_0, \varrho_\tau)$]. Consequently, a finite-time bound on heat dissipation can be obtained as

$$Q \geq -T\Delta S_{\text{sys}} + \frac{\mathcal{T}(\varrho_0, \varrho_\tau)^2}{\tau\beta\langle m\rangle_\tau}. \qquad (150)$$

Equation (150) is the finite-time quantum Landauer principle, which is tight and can be saturated for arbitrary temperatures. In addition to the conventional Landauer term $-T\Delta S_{\text{sys}}$, a finite-time correction term exists in the lower bound, which does not vanish even in the zero-temperature limit. Therefore, the inequality (150) provides a stringent bound on heat dissipation for information erasure in both fast-driving and low-temperature regimes.

We compare bound (150) with an existing bound derived in Ref. [31], which reads

$$Q \geq -T\Delta S_{\text{sys}} + \frac{\mathcal{T}(\varrho_0, \varrho_\tau)^2}{\tau\beta\langle a\rangle_\tau/2}. \qquad (151)$$

Since $m_t \leq a_t/2$ for all times, it is immediately clear that bound (150) is always stronger than bound (151).

Next, we derive a simplified bound which includes the erasure error. Let $\epsilon := \mathcal{T}(\varrho_\tau, \varrho_*)$ be the erasure error, which should be small. We can then analogously bound the terms in Eq. (150) from below as

$$-\Delta S_{\text{sys}} = \ln d - S(\varrho_\tau) \geq \ln d - h(\epsilon), \qquad (152)$$

$$\mathcal{T}(\varrho_0, \varrho_\tau) \geq |\mathcal{T}(\varrho_0, \varrho_*) - \mathcal{T}(\varrho_\tau, \varrho_*)| = |1 - 1/d - \epsilon|, \qquad (153)$$

where we use an inequality relating the entropy difference between two quantum states to their trace distance in the first line [154]. Inserting Eqs. (152) and (153) into Eq. (150), we arrive at the following simple bound:

$$Q \geq T[\ln d - h(\epsilon)] + \frac{(1 - 1/d - \epsilon)^2}{\tau\beta\langle m\rangle_\tau}. \qquad (154)$$

Equation (154) is the simplified Landauer bound that includes finite-time and finite-error corrections. Remarkably, it has the same structure as the classical bound (142). In the limit of perfect and slow erasure, bound (154) reduces to the conventional Landauer bound as $d = 2$.

In analogy to the classical case, a finite-time Landauer bound in terms of the quantum Wasserstein distance and

dynamical activity can also be obtained. By rearranging the speed limit (138), we can prove that

$$\frac{Q}{T} \geq -\Delta S_{\text{sys}} + 2\mathcal{W}_q(\varrho_0, \varrho_\tau) \tanh^{-1}\left(\frac{\mathcal{W}_q(\varrho_0, \varrho_\tau)}{\tau \langle a \rangle_\tau}\right). \quad (155)$$

This finite-time bound can be considered a quantum analog of the classical bound (148).

VII. NUMERICAL DEMONSTRATIONS

Next, we numerically illustrate the applications of our results, the thermodynamic uncertainty relation, speed limits, and finite-time Landauer principles in several classical and quantum systems.

A. Illustration of classical thermodynamic speed limits and finite-time Landauer principle

First, we illustrate the classical speed limits and finite-time Landauer principle in a two-level system. The system dynamics can be described by a Markov jump process with transition rates given by

$$w_{xy}(t) = \gamma_{xy}(t) \frac{e^{\beta \varepsilon_y(t)}}{e^{\beta \varepsilon_y(t)} + e^{\beta \varepsilon_x(t)}}, \quad (156)$$

where $\gamma_{xy}(t) = \gamma_{yx}(t)$ are tunable parameters and $\varepsilon_x(t)$ denotes the instantaneous energy level of state x. For simplicity, we set $\varepsilon_t := \varepsilon_2(t) - \varepsilon_1(t)$ and $\gamma_{xy}(t) = 1$ for all transitions. The parameter $\{\varepsilon_t\}$ thus defines a time-dependent control protocol.

To illustrate the bounds, we consider an information erasure process in which an arbitrary initial distribution is always reset to the ground state within a finite error. We examine two control protocols, namely, optimal and nonoptimal. The optimality here refers to the dissipation of the least amount of heat while achieving the predetermined error.

The optimal protocol is numerically obtained by solving the minimization problem with the following objective functional:

$$\mathcal{F}_c[\{\varepsilon_t\}] := \lambda Q + (1 - \lambda)\mathcal{T}(p_\tau, p_*), \quad (157)$$

where $\lambda \in [0, 1]$ is a weighting factor. The functional $\mathcal{F}_c$ consists of two incompatible objectives, namely, heat dissipation and erasure error, which cannot be simultaneously small. To reduce the erasure error, we must pay the price of dissipation. Conversely, reducing dissipation could enhance the error between the final and ground states. As λ is fixed, the solution of the optimization problem corresponds to a Pareto-optimal protocol, in which heat dissipation cannot be further minimized without increasing the error. Imposing constraints on the control parameters is physically reasonable. Hereafter,

we consider the constraint $\varepsilon_t \in [10^{-2}, 10^1]$. To solve the problem (157) under both equality and inequality constraints, we discretize the control parameters into 1000 points and optimize the functional $\mathcal{F}$ with the aid of nonlinear programming solvers. We determine the weighting factor λ such that both the optimal and nonoptimal protocols reset the uniform distribution $\bar{p} = [1/2, 1/2]^\top$ to the ground state within the same error. The time variation of the protocol is plotted in Fig. 5(b). Notably, the increase in the energy gap between the two levels ε_1 and ε_2 is constant in the intermediate period but tends to slow down in the late period.

The nonoptimal protocol simply lifts the energy level $\varepsilon_2(t)$, forcing the system to descend to the ground state. The time-dependent control parameter ε_t is specified as

$$\varepsilon_t = 0.422 \times \exp\left(\frac{\tau + t}{2\tau - t}\right), \quad (158)$$

which is illustrated in Fig. 5(e). We can observe that, unlike the optimal protocol, the energy gap in the nonoptimal protocol rapidly increases in the late period. In the final time, this naive protocol should dissipate more heat than the optimal protocol.

The process of information erasure is performed within the period τ. At each time $t \leq \tau$, according to Eqs. (128) and (132), the operational time is lower bounded as

$$t \geq \frac{\mathcal{W}_1(p_0, p_t)}{\sqrt{\langle \sigma \rangle_t \langle m \rangle_t}} =: t_{c,1} \geq \frac{\mathcal{T}(p_0, p_t)}{\sqrt{\langle \sigma \rangle_t \langle a \rangle_t / 2}} =: t_{c,2}. \quad (159)$$

These bounds are numerically verified for the optimal and nonoptimal protocols in Figs. 5(c) and 5(f), respectively. As shown, the derived bound $t \geq t_{c,1}$ is tight and stronger than the existing bound $t \geq t_{c,2}$ for all times.

Likewise, the dissipated heat is lower bounded by the entropy change and finite-time correction term as

$$Q \geq -T\Delta S_{\text{sys}} + \frac{\mathcal{W}_1(p_0, p_t)^2}{t\beta \langle m \rangle_t} =: Q_{c,1}$$

$$\geq -T\Delta S_{\text{sys}} + \frac{\mathcal{T}(p_0, p_t)^2}{t\beta \langle a \rangle_t / 2} =: Q_{c,2}. \quad (160)$$

The numerical results are plotted in Figs. 5(d) and 5(g) for the optimal and nonoptimal protocols, respectively. As seen, the new lower bound $Q_{c,1}$ tightly bounds the dissipated heat Q in both protocols, whereas the existing lower bound $Q_{c,2}$ is loose and does not provide a good prediction for heat dissipation. Notice that the average heat dissipation at the final time $\tau = 10$ is approximately 0.33, which is far beyond the conventional Landauer bound $\beta^{-1} \ln 2 \approx 0.069$. This implies that the finite-time correction is dominant over the entropy change in this case.
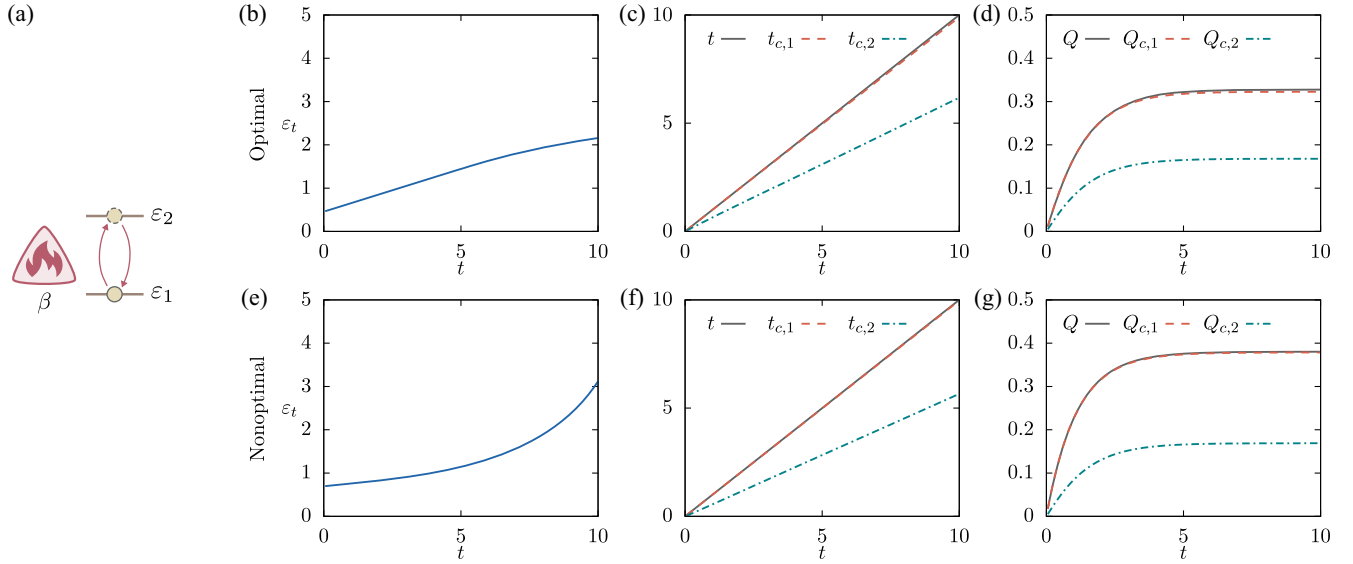
FIG. 5.   Numerical illustration of the classical thermodynamic speed limits and finite-time Landauer principle for both optimal and nonoptimal protocols. (a) Schematic of the two-level system. (b),(e) Time variations of the control parameters of the optimal and nonoptimal protocols. (c),(f) Numerical verification of the speed limits $t \geq t_{c,1} \geq t_{c,2}$ and (d),(g) finite-time Landauer principle $Q \geq Q_{c,1} \geq Q_{c,2}$. The other parameters are fixed as $\beta = 10$ and $\tau = 10$. The weighting factor λ is set such that $(1 - \lambda)/\lambda = 10^5$.

B. Illustration of quantum thermodynamic speed limits and finite-time Landauer principle

We next exemplify the quantum speed limits and finite-time Landauer principle with a simple model of information erasure using a spin-$1/2$ qubit. The qubit is weakly attached to a heat bath at inverse temperature β. The time evolution of the reduced density matrix can be described by the Lindblad equation with the Hamiltonian

$$H_t = \frac{\varepsilon_t}{2}[\cos(\theta_t)\sigma_z + \sin(\theta_t)\sigma_x] \qquad (161)$$

and jump operators

$$L_1(t) = \sqrt{\gamma \varepsilon_t(n_t + 1)}|0_t\rangle\langle 1_t|, \qquad (162)$$

$$L_2(t) = \sqrt{\gamma \varepsilon_t n_t}|1_t\rangle\langle 0_t|. \qquad (163)$$

Here, $\{|0_t\rangle, |1_t\rangle\}$ are the instantaneous energy eigenstates, $\sigma_{x,y,z}$ are the Pauli matrices, γ is the coupling strength, $n_t := 1/(e^{\beta\varepsilon_t} - 1)$, and ε_t and θ_t are time-dependent control parameters. More specifically, E_t characterizes the energy gap between the energy eigenstates, whereas θ_t quantifies the relative strength of coherent tunneling to energy bias [155]. The qubit is initially prepared in the state $\varrho_0 = \mathbb{1}/2$ and subsequently driven toward the ground state $\varrho_* = |0\rangle\langle 0|$ of σ_z. If θ_t is time invariant, quantum coherence in the energy eigenstates cannot be created, and the protocol is thus classical. Otherwise, it becomes a quantum protocol.

An infinite number of approaches can be used to reset the qubit with a probability close to 1. As in the classical case,

two protocols are considered, namely, Pareto-optimal and nonoptimal. Both protocols are designed to erase information with the same error.

The optimal protocol minimizes two incompatible objectives, the average dissipated heat and erasure error. Specifically, the protocol can be achieved by solving the minimization problem with the following multiobjective functional [31]:

$$\mathcal{F}_q[\{\varepsilon_t, \theta_t\}] := \lambda Q - (1 - \lambda)F(\varrho_\tau, \varrho_*), \qquad (164)$$

where $\lambda \in [0, 1)$ is a weighting factor and $F(\varrho, \sigma) := (\mathrm{tr}|\sqrt{\varrho}\sqrt{\sigma}|)^2$ is the fidelity of the two quantum states ϱ and σ [156]. Because of the physical limitations, placing constraints on the control parameters is natural. Hereafter, we impose the following lower and upper bounds on the parameters: $\varepsilon_t \in [10^{-1}, 10^2]$ and $\theta_t \in [-\pi, \pi]$. By numerically solving the nonlinear optimization problem (164), we can obtain the optimal protocol, as plotted in Fig. 6(b). As seen, the parameter θ_t is fixed to 0 for all times, implying that the optimal protocol is classical and does not generate any amount of energetic coherence. Furthermore, the energy gap increases gradually in the intermediate period and changes rapidly in the final stage.

The nonoptimal protocol is defined as

$$\varepsilon_t = 2.04 \times \exp\left(\frac{\tau + t}{2\tau - t}\right), \qquad \theta_t = \pi\left(\frac{t}{\tau} - 1\right), \quad (165)$$

which is plotted in Fig. 6(e). This protocol naively increases the energy gap while varying the coherent parameter.
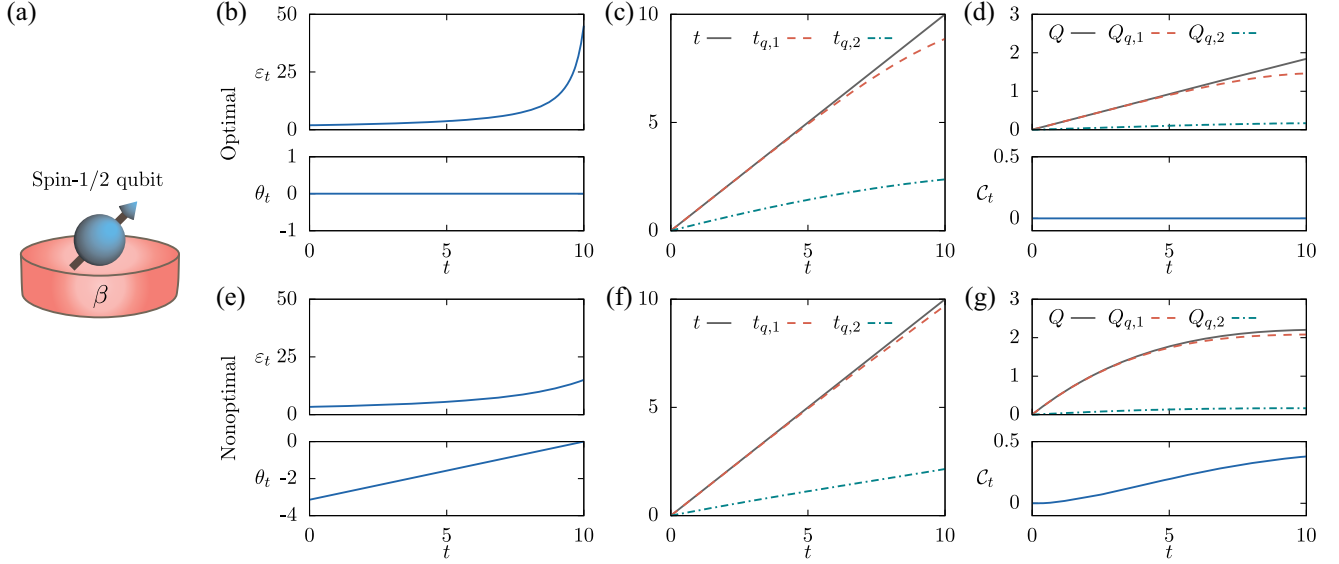
FIG. 6. Numerical illustration of the quantum thermodynamic speed limits and finite-time Landauer principle for both optimal and nonoptimal protocols. (a) Schematic of the spin-1/2 qubit, which is coupled to a heat bath at inverse temperature β. (b),(e) Time variations of the control parameters of the optimal and nonoptimal protocols. (c),(f) Numerical verification of the speed limits $t \geq t_{q,1} \geq t_{q,2}$ and (d),(g) finite-time Landauer principle $Q \geq Q_{q,1} \geq Q_{q,2}$ as well as the amount of quantum coherence $\mathcal{C}_t$. The other parameters are fixed as $\beta = 10$, $\gamma = 0.1$, and $\tau = 10$. The weighting factor λ is set such that $(1 - \lambda)/\lambda = 10^3$.

We also observe that the increase in the energy gap is different from that of the optimal case.

We first demonstrate the quantum thermodynamic speed limits. According to Eqs. (137) and (139), the operational time is lower bounded as follows:

$$t \geq \frac{\mathcal{W}_q(\varrho_0, \varrho_t)}{\sqrt{\langle \sigma \rangle_t \langle m \rangle_t}} =: t_{q,1} \geq \frac{\mathcal{W}_q(\varrho_0, \varrho_t)}{\sqrt{\langle \sigma \rangle_t \langle a \rangle_t / 2}} =: t_{q,2}. \quad (166)$$

We illustrate these bounds for both the optimal and nonoptimal protocols in Figs. 6(c) and 6(f), respectively. As shown, the derived bound $t \geq t_{q,1}$ is tight for all times and is stronger than the existing bound $t \geq t_{q,2}$.

Next, we verify the finite-time quantum Landauer principle. The lower bounds on the average heat dissipation are given by Eqs. (150) and (151) as

$$Q \geq -T\Delta S_{\text{sys}} + \frac{\mathcal{T}(\varrho_0, \varrho_t)^2}{t\beta\langle m \rangle_t} =: Q_{q,1}$$

$$\geq -T\Delta S_{\text{sys}} + \frac{\mathcal{T}(\varrho_0, \varrho_t)^2}{t\beta\langle a \rangle_t / 2} =: Q_{q,2}. \quad (167)$$

The numerical results are plotted in Figs. 6(d) and 6(g) for the optimal and nonoptimal protocols, respectively. As shown, the new bound $Q \geq Q_{q,1}$ is tight for all times, whereas the existing bound $Q \geq Q_{q,2}$ is loose. The optimal protocol also clearly dissipates less heat than does the nonoptimal protocol at the final time $\tau = 10$. In addition, note that the average heat dissipation in both protocols is

approximately 2, which is significantly greater than the conventional Landauer bound $\beta^{-1} \ln 2 \approx 0.069$.

We discuss the effect of quantum coherence in the finite-time erasure process. For convenience, we quantify the amount of energetic coherence using the ℓ_1-norm, which is one of the most general coherence monotones in the literature [157]:

$$\mathcal{C}_t := \int_0^t [|\langle 0_s | \varrho_s | 1_s \rangle| + |\langle 1_s | \varrho_s | 0_s \rangle|] ds. \quad (168)$$

That is, $\mathcal{C}_t$ is the time integral of the sum of absolute off-diagonal elements of quantum states in the basis of energy eigenstates. Since θ_t is invariant in the optimal protocol, the instantaneous energy eigenstates remain unchanged, and the density matrix is always diagonal in the eigenstates. Therefore, the total amount of quantum coherence generated is always zero (i.e., $\mathcal{C}_t = 0$), which is plotted in Fig. 6(d). On the other hand, for the nonoptimal protocol, θ_t varies over time and quantum coherence is generally generated. The positive finite value of $\mathcal{C}_t$ can be confirmed from Fig. 6(g). We can observe that the nonoptimal protocol that generates coherence is more dissipative than the optimal protocol that does not create coherence. This is consistent with the fact that the creation of quantum coherence leads to unavoidable dissipation [31]. This also suggests that a lower bound of heat dissipation that can capture the effect of coherence is desirable.

It is, therefore, worthwhile discussing how quantum coherence effects can be captured by the bounds. According to the definition (94) of quantum dynamical

activity, a_t can be explicitly expressed in terms of the diagonal elements of ϱ_t in the basis of energy eigenstates as

$$a_t = \gamma\varepsilon_t[n_t\langle 0_t|\varrho_t|0_t\rangle + (n_t + 1)\langle 1_t|\varrho_t|1_t\rangle]. \qquad (169)$$

As seen, a_t has no coherent contribution from the off-diagonal part of quantum state ϱ_t. On the other hand, quantum dynamical state mobility m_t implicitly includes a coherent contribution. This can be validated from Eq. (101) by noticing that the eigenbasis $\{|x_t\rangle\}$ of the spectral decomposition $\varrho_t = \sum_x p_x(t)|x_t\rangle\langle x_t|$ is generally different from the energy eigenstates $\{|0_t\rangle, |1_t\rangle\}$. Although the coherence present in the quantum state ϱ_t may contribute to the bounds through the boundary terms ΔS_{sys} and $\mathcal{T}(\varrho_0, \varrho_t)$, such contributions are negligibly small as compared to the time-extensive contribution $t\langle m\rangle_t$. Therefore, the derived bound $Q \geq Q_{q,1}$ can capture the effect of quantum coherence occurring during the process and precisely predict heat dissipation even in the presence of quantum coherence, whereas the existing bound $Q \geq Q_{q,2}$ cannot.

VIII. CONCLUSION AND OUTLOOK

In this study, we elucidated an intimate relationship between thermodynamics and discrete optimal transport for both classical and quantum cases. To this end, we introduced a novel physical term, namely, dynamical state mobility, which characterizes a complementary aspect of irreversible entropy production in the time evolution of a system. By deriving an improved thermodynamic uncertainty relation, we showed that dynamical state mobility plays a critical role in constraining the fluctuation of time-antisymmetric currents, thus providing insight into the precision of currents in Markov jump processes. Exploiting this term, we derived variational formulas that express the discrete Wasserstein distance in terms of the thermodynamic cost associated with Markovian dynamics. These formulas not only unify the relationship between thermodynamics and optimal transport for both discrete and continuous cases but also generalize to the Markovian quantum dynamics. From the variational formulas, we derived stringent thermodynamic speed limits and the finite-time Landauer principle. The obtained bounds are tight and can be saturated for an arbitrary pair of initial and final states and arbitrary temperatures.

Our theoretical frameworks also shed light on the minimization problem of entropy production in discrete Markov dynamics. Recent studies [91–93] have shown that entropy production can be optimized to be arbitrarily small if there are no constraints on the transition rates. Our results suggest that dynamical state mobility may be a reasonable constraint, because, once it is fixed, minimum entropy production is immediately determined by the discrete Wasserstein distance. In addition, the optimal protocol that attains the minimum entropy production can be constructed

from the optimal coupling between the initial and final distributions, which can be numerically computed in an efficient manner.

Although not explicitly stated in this study, our framework is also applicable to bipartite systems [158], in which two subsystems exchange information. In this case, the Wasserstein distance between the initial and final distributions of a subsystem can be expressed in terms of the entropy production of that subsystem and the information flow with another subsystem.

Our study opens several possible directions for future research, which are as follows.

(i) *Generalizing the formulations to include measurement and feedback control.*—Measurement and feedback control are ubiquitous in physics and biology. The thermodynamics of feedback control [159] has been intensively developed in recent years. In this study, we focused exclusively on discrete Markovian systems subjected to deterministic control protocols. Extending our framework to include the effects of measurement and feedback control would be significant, as it would provide a better understanding of the role of information in nonequilibrium systems. Because information obtained from measurements can enhance the precision of observables [160,161] and would violate the second law of thermodynamics [162], in addition to entropy production, information would be expected to play a crucial role in the speed of state transformation and heat dissipation of finite-time information erasure.

(ii) *Decomposition of entropy production.*—Decomposing entropy production is theoretically appealing, because it provides insight into the dissipative structure of thermodynamic processes. Previous studies have shown that irreversible entropy production of Markovian dynamics could be split into an adiabatic and nonadiabatic contribution in both discrete and continuous cases, which originates from the breaking of detailed balance [163,164]. For overdamped Langevin dynamics, recent studies [165,166] have introduced a new decomposition of the entropy production rate in terms of the continuous Wasserstein distance and a housekeeping entropy production rate as

$$\sigma_t = \frac{1}{D}\left(\lim_{dt \to 0} \frac{W_2(p_t, p_{t+dt})}{dt}\right)^2 + \sigma_t^{\mathrm{hk}}. \qquad (170)$$

The term σ_t^{hk} vanishes as the system is driven by a conservative force. Inspired by this decomposition, the entropy production rate of Markov jump processes can be split in a similar manner as

$$\sigma_t = \frac{1}{m_t}\left(\lim_{dt \to 0} \frac{\mathcal{W}_1(p_t, p_{t+dt})}{dt}\right)^2 + \tilde{\sigma}_t. \qquad (171)$$

The term $\tilde{\sigma}_t$ is non-negative and vanishes only when the system is driven by an optimal protocol, provided that m_t is fixed. Investigating the properties of the contribution $\tilde{\sigma}_t$ would be an interesting direction and may lead to a deep understanding of dissipation in Markov jump processes.

(iii) *Application to deterministic biochemical reaction networks.*—Although our framework deals with stochastic dynamics, generalizing the formulas to cases of deterministic dynamics such as biochemical reaction networks [167] would be interesting. This is feasible, because our results are derived from the master equation, which is similar to the deterministic rate equation characterizing the time evolution of biochemical reaction networks.

(iv) *Thermodynamic interpretation of the discrete L^2-Wasserstein distance.*—Thus far, we have investigated the connection between thermodynamics and optimal transport through the discrete L^1-Wasserstein distance. Although we showed that the discrete L^1-Wasserstein distance has aspects similar to the continuous L^2-Wasserstein distance, it remains an open question whether a thermodynamic interpretation exists for the discrete L^2-Wasserstein distance. Clarification of this interpretation is desirable and could lead to new fundamental thermodynamic bounds.

(v) *Formulation under constrained control protocols.*—In this study, we thermodynamically interpreted the discrete Wasserstein distance using Markov jump processes whose transition rates can be arbitrarily controlled without any constraint. However, in practice, some constraints may be imposed on the transition rates and protocols [92,93,168,169]. Developing analogous formulas for these settings would be highly relevant and broaden the range of applications. The specific form of transition rates given in Eqs. (50) and (51) also suggests that investigating this direction may reveal the thermodynamic role of the discrete L^2-Wasserstein distance.

ACKNOWLEDGMENTS

APPENDIX A: GEOMETRIC PROPERTY OF CONTINUOUS L^2-WASSERSTEIN DISTANCE

Here, we discuss a geometric interpretation of the Wasserstein distance. Specifically, we show that the L^2-Wasserstein distance can be interpreted as a Riemannian distance on the infinite-dimensional manifold M of probability distribution functions. For each distribution function $p(x)$, the tangent velocity space $\mathrm{Tan}_p M$ at point p can be defined as [170]

$$\mathrm{Tan}_p M := \{w(x)|\nabla \cdot [w(x)p(x)] = 0\}^\perp. \quad (A1)$$

Here, $\Omega^\perp$ denotes the orthogonal complement of a subspace Ω. In other words, $\mathrm{Tan}_p M$ contains all velocity fields $v(x)$ that satisfy

$$\int_{\mathbb{R}^d} [v(x) \cdot w(x)] p(x) dx = 0 \quad \forall w \quad \text{such that } \nabla \cdot (wp) = 0. \quad (A2)$$

The tangent space can be indirectly defined via the tangent velocity space as

$$T_p M := \{u| \exists v \in \mathrm{Tan}_p M \text{ such that } u + \nabla \cdot (vp) = 0\}. \quad (A3)$$

Consequently, a Riemannian metric $g_p : T_p M \times T_p M \to \mathbb{R}$ can be defined on the tangent space as

$$g_p(u_1, u_2) := \int_{\mathbb{R}^d} [v_1(x) \cdot v_2(x)] p(x) dx, \quad (A4)$$

where v_i is the velocity field corresponding to the tangent vector u_i ($i = 1, 2$). We can then show that $W_2(p^A, p^B)$ is exactly the geodesic distance between p^A and p^B induced by the defined metric:

$$W_2(p^A, p^B) = \min_{p_t} \left\{ \tau \int_0^\tau g_{p_t}(\dot{p}_t, \dot{p}_t) dt \right\}^{1/2}$$
$$= \min_{p_t} \int_0^\tau \sqrt{g_{p_t}(\dot{p}_t, \dot{p}_t)} dt. \quad (A5)$$

Here, we consider the fact that the geodesic distance between two points is equal to the minimum square root of the divergence taken over all possible paths connecting those points.

In general, obtaining a closed form for $W_2(p^A, p^B)$ is difficult except in the case in which p^A and p^B are normal distributions. Therefore, a lower bound on W_2 is often considered. It has been previously proved that $W_2(p^A, p^B)$ can be bounded from below by the means and covariances of distributions p^A and p^B as [171]

$$W_2(p^A, p^B)^2 \geq \|\mu_A - \mu_B\|^2$$
$$+ \mathrm{tr}\left\{ \Xi_A + \Xi_B - 2\sqrt{\sqrt{\Xi_A}\Xi_B\sqrt{\Xi_A}} \right\}, \quad (A6)$$

where μ_X and Ξ_X are the mean and covariance matrices, respectively, of the probability distribution p^X for $X \in \{A, B\}$.

APPENDIX B: USEFUL PROPOSITIONS

Proposition 3: Let $x = [x_1, \ldots, x_n]^\top$ and $y = [y_1, \ldots, y_n]^\top$ be vectors of real numbers and σ be a permutation of $\{1, \ldots, n\}$ such that if $x_i > x_j$, then $y_{\sigma(i)} \geq y_{\sigma(j)}$ for any i and j. Then, the following inequality holds:

$$\sum_{j=1}^{n} |x_j - y_j| \geq \sum_{j=1}^{n} |x_j - y_{\sigma(j)}|. \qquad (B1)$$

Proof.—Without loss of generality, we assume $x_1 \leq x_2 \leq \ldots \leq x_n$. Let ν be a permutation of $\{1, \ldots, n\}$ such that $\sum_{j=1}^{n} |x_j - y_{\nu(j)}|$ is minimum among all possible permutations and that the number of inversion pairs (i.e., $i > j$ and $y_{\nu(i)} < y_{\nu(j)}$) is minimum. Assume that two indices i and $j = i - 1$ exist such that $x_i \geq x_j$ and $y_{\nu(i)} < y_{\nu(j)}$. We then consider a new permutation ν' obtained from ν by swapping $\nu(i)$ and $\nu(j)$; that is, $\nu'(i) = \nu(j)$, $\nu'(j) = \nu(i)$, and $\nu'(k) = \nu(k)$ for all $k \neq i, j$. In this case, we can easily prove that

$$|x_i - y_{\nu(i)}| + |x_j - y_{\nu(j)}| \geq |x_i - y_{\nu'(i)}| + |x_j - y_{\nu'(j)}|. \quad (B2)$$

This means that

$$\sum_{j=1}^{n} |x_j - y_{\nu(j)}| \geq \sum_{j=1}^{n} |x_j - y_{\nu'(j)}|, \qquad (B3)$$

and the permutation ν' has fewer inversion pairs than the permutation ν, which contradicts the optimality of the permutation ν. Therefore, we have $y_{\nu(i)} \geq y_{\nu(j)}$ for any $x_i > x_j$. Consequently, the permutations σ and ν satisfy

$$\sum_{j=1}^{n} |x_j - y_{\sigma(j)}| = \sum_{j=1}^{n} |x_j - y_{\nu(j)}|, \qquad (B4)$$

from which Eq. (B1) is immediately proved because of the optimality of the permutation ν. ∎

Proposition 4: Let $x = [x_1, \ldots, x_n]^\top$ and $y = [y_1, \ldots, y_{n'}]^\top$ be vectors of non-negative numbers. If $\sum_{i=1}^{n} x_i = \sum_{j=1}^{n'} y_j$, then a matrix $Z = [z_{ij}] \in \mathbb{R}^{n \times n'}$ exists with non-negative elements such that

$$\sum_{j=1}^{n'} z_{ij} = x_i \quad \text{and} \quad \sum_{i=1}^{n} z_{ij} = y_j. \qquad (B5)$$

Proof.—We prove by induction on $k = n + n' \geq 2$. The $k = 2$ case is evident, since $n = n' = 1$ and $x_1 = y_1$; therefore, we can choose $z_{11} = x_1$. Supposing that it holds for all $k \leq \bar{k}$, we can consider an arbitrary case with $k = \bar{k} + 1$. Let $v = \min\{x_1, y_1\}$ and set $z_{11} = v$. Without loss of generality, we can assume that $v = x_1$. Then, $z_{1i} = 0$ for all $i \geq 2$. Consider two vectors $x' = [x_2, \ldots, x_n]^\top$ and $y' = [y_1 - x_1, \ldots, y_{n'}]^\top$ with $k' = n + n' - 1 = \bar{k}$. A matrix $Z' = [z'_{ij}] \in \mathbb{R}^{(n-1) \times n'}$ exists such that

$$\sum_{j=1}^{n'} z'_{ij} = x'_i \quad \text{and} \quad \sum_{i=1}^{n-1} z'_{ij} = y'_j. \qquad (B6)$$

Set $z_{ij} = z'_{(i-1)j}$ for all $i \geq 2$. Then, the matrix Z satisfies Eq. (B5). ∎

Proposition 5: For arbitrary real numbers $\{x_i\}$ and $\{y_i\}$ that satisfy $x_i y_i \geq 0$ for all i, the following inequality holds:

$$\sum_i \frac{x_i}{y_i} \sum_i x_i y_i \geq \sum_i \frac{x_i}{\coth(y_i/2)} \sum_i x_i \coth(y_i/2). \quad (B7)$$

Proof.—The inequality (B7) is equivalent to

$$\sum_{i > j} \left[\frac{x_i}{y_i} x_j y_j + \frac{x_j}{y_j} x_i y_i - \frac{x_i}{\coth(y_i/2)} x_j \coth(y_j/2) \right.$$
$$\left. - \frac{x_j}{\coth(y_j/2)} x_i \coth(y_i/2) \right] \geq 0. \qquad (B8)$$

It suffices to prove that each term in the above summation is non-negative, that is,

$$x_i x_j \left[\frac{y_j}{y_i} + \frac{y_i}{y_j} - \frac{\coth(y_j/2)}{\coth(y_i/2)} - \frac{\coth(y_i/2)}{\coth(y_j/2)} \right] \geq 0. \quad (B9)$$

Since $x_i x_j y_i y_j \geq 0$, Eq. (B9) is equivalent to

$$y_i^2 + y_j^2 - y_i y_j \left[\frac{\coth(y_j/2)}{\coth(y_i/2)} + \frac{\coth(y_i/2)}{\coth(y_j/2)} \right] \geq 0. \quad (B10)$$

Since $y \coth(y)$ and $y/\coth(y)$ are even functions, we can assume that $y_i \geq y_j \geq 0$ without loss of generality. The inequality (B10) can be rewritten as

$$\frac{y_i}{y_j} + \frac{y_j}{y_i} - \left[\frac{\coth(y_j/2)}{\coth(y_i/2)} + \frac{\coth(y_i/2)}{\coth(y_j/2)} \right]$$
$$= \left[\frac{y_i}{y_j} - \frac{\coth(y_j/2)}{\coth(y_i/2)} \right] \left[1 - \left\{ \frac{y_i}{y_j} \frac{\coth(y_j/2)}{\coth(y_i/2)} \right\}^{-1} \right] \geq 0. \qquad (B11)$$

Since $\coth(y)$ is a strictly decreasing function over $[0, +\infty)$, we have $y_i \coth(y_j/2) \geq y_j \coth(y_i/2)$. Therefore, Eq. (B11) is equivalent to

$$\frac{y_i}{y_j} \geq \frac{\coth(y_j/2)}{\coth(y_i/2)}. \qquad (B12)$$

The inequality (B12) is always valid, since $y \coth(y/2)$ is an increasing function over $[0, +\infty)$. Therefore, Eq. (B7) is proved. ∎

APPENDIX C: DERIVATION OF CALCULATIONS IN SEC. III

1. Property of m_t

Here, we show a relevant property of dynamical state mobility in terms of optimizing irreversible entropy production.

Lemma 6: For any Markov jump process $\{p_t; \mathsf{W}_t\}_{0 \le t \le \tau}$ and arbitrary positive constant $\bar{D}$, a Markov process $\{\tilde{p}_t; \tilde{\mathsf{W}}_t\}_{0 \le t \le \tau}$ exists that simultaneously satisfies the following conditions.

(i) The time evolution of probability distribution is the same (i.e., $\tilde{p}_t = p_t$ for all times).

(ii) The time-averaged state mobility is equal to $\bar{D}$ (i.e., $\langle \tilde{m} \rangle_\tau = \bar{D}$).

(iii) The associated product of entropy production and dynamical state mobility is smaller than that of the original process (i.e., $\tilde{\Sigma}_\tau \tilde{\mathcal{M}}_\tau \le \Sigma_\tau \mathcal{M}_\tau$).

Proof.—We consider another Markov jump process with the transition rate matrix $\tilde{\mathsf{W}}_t$ defined in the following manner. For each transition rate $w_{xy}(t) > 0$, we define

$$\tilde{w}_{xy}(t) = w_{xy}(t) + \frac{\alpha_{xy}(t)}{p_y(t)} \quad (x \ne y), \qquad (\text{C1})$$

where $\alpha_{xy}(t) = \alpha_{yx}(t)$ are real coefficients to be later determined. We can easily verify that $\tilde{j}_{xy}(t) = j_{xy}(t)$ for all $x \ne y$. Therefore, given that the initial distribution is the same (i.e., $\tilde{p}_0 = p_0$), we immediately obtain $\tilde{p}_t = p_t$ for all t, which fulfills condition (i). In addition, for any $\lambda > 0$, $\alpha_{xy}(t)$ always exists such that

$$\tilde{m}_{xy}(t) = \frac{\tilde{j}_{xy}(t)}{\ln[\tilde{w}_{xy}(t)\tilde{p}_y(t)] - \ln[\tilde{w}_{yx}(t)\tilde{p}_x(t)]} = \lambda |j_{xy}(t)|. \qquad (\text{C2})$$

This is because the following quantity can take an arbitrary positive value depending on the manner in which $\alpha_{xy}(t)$ is chosen:

$$\frac{w_{xy}(t)p_y(t) - w_{yx}(t)p_x(t)}{\ln \frac{w_{xy}(t)p_y(t) + \alpha_{xy}(t)}{w_{yx}(t)p_x(t) + \alpha_{xy}(t)}}. \qquad (\text{C3})$$

Choosing $\alpha_{xy}(t)$ such that Eq. (C2) is satisfied and setting $\lambda = \bar{D}\tau[\int_0^\tau \sum_{x>y} |j_{xy}(t)| dt]^{-1}$, we can calculate

$$
\begin{aligned}
\langle \tilde{m} \rangle_\tau &= \tau^{-1} \int_0^\tau \sum_{x>y} \tilde{m}_{xy}(t) dt \\
&= \tau^{-1} \lambda \int_0^\tau \sum_{x>y} |j_{xy}(t)| dt \\
&= \bar{D}, \qquad (\text{C4})
\end{aligned}
$$

which fulfills condition (ii). Finally, we prove that condition (iii) is also satisfied. To this end, we first note that

$$
\begin{aligned}
\tilde{\sigma}_t &= \sum_{x>y} \tilde{j}_{xy}(t) \ln \frac{\tilde{w}_{xy}(t)\tilde{p}_y(t)}{\tilde{w}_{yx}(t)\tilde{p}_x(t)} \\
&= \lambda^{-1} \sum_{x>y} |j_{xy}(t)|. \qquad (\text{C5})
\end{aligned}
$$

Consequently, condition (iii) can be verified as follows:

$$
\begin{aligned}
\tilde{\Sigma}_\tau \tilde{\mathcal{M}}_\tau &= \int_0^\tau \tilde{\sigma}_t dt \int_0^\tau \tilde{m}_t dt \\
&= \left(\int_0^\tau \sum_{x>y} |j_{xy}(t)| dt \right)^2 \\
&\le \left(\int_0^\tau \sum_{x>y} \sigma_{xy}(t) dt \right) \left(\int_0^\tau \sum_{x>y} m_{xy}(t) dt \right) \\
&= \Sigma_\tau \mathcal{M}_\tau. \qquad (\text{C6})
\end{aligned}
$$

It is noteworthy that if we choose $\bar{D} = \langle m \rangle_\tau$, condition (iii) implies $\tilde{\Sigma}_\tau \le \Sigma_\tau$. ∎

2. Lower bound of dynamical state mobility

Here, we provide a lower bound of dynamical state mobility in terms of entropy production and dynamical activity. By performing algebraic calculations, we can show that the kinetic coefficients $\{m_{xy}(t)\}$ can be expressed in terms of entropy production and dynamical activity rates at the transition level as

$$m_{xy}(t) = \frac{\sigma_{xy}(t)}{4} \Phi\left(\frac{\sigma_{xy}(t)}{2[a_{xy}(t) + a_{yx}(t)]} \right)^{-2}, \qquad (\text{C7})$$

where $\Phi(x)$ is the inverse function of $x \tanh(x)$. Since $x\Phi(x/y)^{-2}$ is a convex function over $(0, +\infty) \times (0, +\infty)$, we can derive a lower bound for m_t as

$$\frac{\sigma_t}{4} \Phi\left(\frac{\sigma_t}{2a_t} \right)^{-2} \le m_t. \qquad (\text{C8})$$

The inequality (C8) indicates that m_t can be lower bounded by both the entropy production and dynamical activity rates. Exploiting the convexity of $x\Phi(x/y)^{-2}$ also yields the following inequality:

$$\frac{\Sigma_\tau}{4} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau} \right)^{-2} \le \mathcal{M}_\tau. \qquad (\text{C9})$$

3. Proof of Eq. (61)

Through the Cramér-Rao inequality [102], the precision of time-antisymmetric currents can be upper bounded by pseudo entropy production as [172]

$$\frac{\langle J \rangle^2}{\text{var}[J]} \le \Sigma_\tau^{\text{ps}}, \qquad (\text{C10})$$

where Σ_τ^{ps} denotes pseudo entropy production given by

$$\Sigma_\tau^{\text{ps}} = \int_0^\tau \sum_{x>y} \frac{j_{xy}(t)^2}{a_{xy}(t) + a_{yx}(t)} dt. \qquad (\text{C11})$$

Σ_τ^{ps} is an empirical quantity that quantifies the degree of irreversibility. Unlike irreversible entropy production Σ_τ, which diverges in the presence of unidirectional transitions, pseudo entropy production always remains finite. However, it cannot be related directly to heat dissipation in thermodynamic processes. It is noteworthy that the magnitude relation $\Sigma_\tau^{ps} \leq \Sigma_\tau/2$ holds for all times.

Next, we prove the following inequality:

$$\Sigma_\tau^{ps} \leq \frac{\Sigma_\tau \mathcal{M}_\tau}{\mathcal{A}_\tau}. \tag{C12}$$

Noting that $j_{xy}(t) = a_{xy}(t) - a_{yx}(t)$ and $f_{xy}(t) = \ln[a_{xy}(t)/a_{yx}(t)]$, we can show that

$$a_{xy}(t) + a_{yx}(t) = j_{xy}(t)\frac{e^{f_{xy}(t)} + 1}{e^{f_{xy}(t)} - 1}$$
$$= j_{xy}(t)\coth[f_{xy}(t)/2]. \tag{C13}$$

Applying Proposition 5, we can prove the inequality (C12) as follows:

$$\begin{aligned}
\Sigma_\tau^{ps}\mathcal{A}_\tau &= \int_0^\tau \sum_{x>y} \frac{j_{xy}(t)^2}{a_{xy}(t) + a_{yx}(t)}dt \int_0^\tau \sum_{x>y}[a_{xy}(t) + a_{yx}(t)]dt \\
&= \int_0^\tau \sum_{x>y} \frac{j_{xy}(t)}{\coth[f_{xy}(t)/2]}dt \int_0^\tau \sum_{x>y}j_{xy}(t)\coth[f_{xy}(t)/2]dt \\
&\leq \int_0^\tau \sum_{x>y} \frac{j_{xy}(t)}{f_{xy}(t)}dt \int_0^\tau \sum_{x>y}j_{xy}(t)f_{xy}(t)dt \\
&= \Sigma_\tau \mathcal{M}_\tau.
\end{aligned} \tag{C14}$$

Combining Eqs. (C10) and (C12) gives the following thermodynamic uncertainty relation:

$$\frac{\langle J \rangle^2}{\text{var}[J]} \leq \frac{\Sigma_\tau \mathcal{M}_\tau}{\mathcal{A}_\tau} = \eta\frac{\Sigma_\tau}{2}. \tag{C15}$$

4. Additional illustration of Eq. (61)

Here, we numerically demonstrate the improved thermodynamic uncertainty relation in a thermoelectric device [173]. A thermoelectric device is an engine that transports electrons from a low- to a high-potential lead through a two-level quantum dot [see Fig. 7(a)]. Each energy level ε_i of the quantum dot is coupled to a lead with chemical potential $\mu_i(\mu_2 > \mu_1)$ and temperature T_c. Electrons enter and exit the quantum dot due to interactions with the leads. Because of the Coulomb repulsion between electrons, we can assume that at most one electron always exists in the quantum dot. The transitions between the two levels of the quantum dot are mediated by two heat baths, namely, cold and hot baths at temperatures T_c and $T_h(> T_c)$, respectively. From a thermodynamic perspective, the device can be considered a heat engine that converts some of the heat absorbed from the hot heat bath into work in the form of transporting electrons from a low to a high potential.

The thermoelectric device can be described by a Markov jump process with three states; that is, the quantum dot either (i) is an empty (state 0), (ii) contains one electron in energy level ε_1 (state 1), or (iii) contains one electron in

energy level $\varepsilon_2(> \varepsilon_1)$ (state 2). Electrons are exchanged with the leads at the following rates:

$$w_{i0} = \gamma_i/(1 + e^{x_i}), \qquad w_{0i} = \gamma_i e^{x_i}/(1 + e^{x_i}), \tag{C16}$$

where $\gamma_i > 0$ denotes the coupling strength to lead i and $x_i := (\varepsilon_i - \mu_i)/T_c$. The transition rates between the two energy levels of the quantum dot are given by

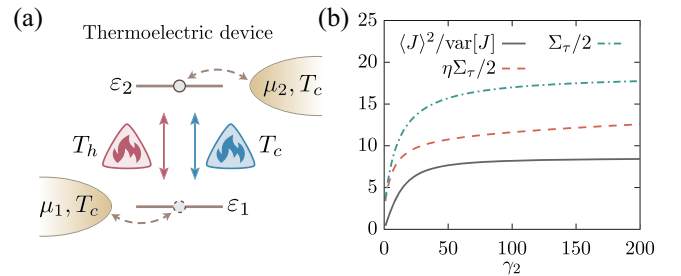$$w_{12} = w_{12}^c + w_{12}^h, \qquad w_{21} = w_{21}^c + w_{21}^h, \tag{C17}$$

FIG. 7. Numerical illustration of the thermodynamic uncertainty relations. (a) Schematic of the thermoelectric engine that transports electrons from the left to the right lead through the two-level quantum dot and (b) numerical verification. The current precision $\langle J \rangle^2/\text{var}[J]$, new bound $\eta\Sigma_\tau/2$, and conventional bound $\Sigma_\tau/2$ are depicted by the solid, dashed, and dash-dotted lines, respectively. γ_2 is varied, whereas other parameters are fixed as $\beta_c = 10$, $\beta_h = 0.1$, $\gamma_1 = 10$, $\gamma_c = \gamma_h = 1$, $\varepsilon_1 = 0$, $\varepsilon_2 = 1$, $\mu_1 = 0.4$, $\mu_2 = 0.6$, and $\tau = 1$.

where $w_{21}^a = \gamma_a/(e^{x_a} - 1)$, $w_{12}^a = \gamma_a e^{x_a}/(e^{x_a} - 1)$, $x_a :=$ $(\varepsilon_2 - \varepsilon_1)/T_a$ for $a \in \{c, h\}$, and γ_c and γ_h denote the coupling strengths to the heat baths. Here, the symbols c and h correspond to the cold and hot heat baths, respectively.

We consider the thermoelectric device operating in a stationary state. The current of interest is the net number of electrons transported between the leads. The stochastic current can be defined by setting $\Upsilon_{10} = 1 = -\Upsilon_{01}$ and $\Upsilon_{xy} = 0$ for others. The precision of the current over a finite period τ can be numerically calculated using full counting statistics.

We vary $\gamma_2 \in (0, 200]$ while fixing the remaining parameters. For each parameter setting, we calculate the precision of the electron current and the bounds of the conventional and new relations. As Fig. 7(b) shows, the new bound is always tighter than the conventional bound and more effectively predicts the current precision.

APPENDIX D: DERIVATION OF CALCULATIONS IN SEC. IV

1. Proof of Eq. (72)

Here, we prove that $\mathcal{W}_1(p, q) = \mathcal{T}(p, q)$ in the case of $d_{xy} = 1 - \delta_{xy}$. First, we prove that $\mathcal{W}_1(p, q) \geq \mathcal{T}(p, q)$. Let $S_+ = \{x|p_x \geq q_x\}$ and $S_- = \{x|p_x < q_x\}$. Evidently, $S_+ \cup S_- = \{1, 2, ..., N\}$. Moreover, since $\sum_x p_x = \sum_x q_x = 1$, we have

$$\sum_{x \in S_+}(p_x - q_x) = \sum_{x \in S_-}(q_x - p_x). \tag{D1}$$

Consequently, $\sum_x |p_x - q_x| = 2\sum_{x \in S_-}(q_x - p_x)$. Exploiting the positivity of d_{xy} and π_{xy}, we can bound $\mathcal{W}_1$ from below as follows:

$$\begin{aligned}
\mathcal{W}_1(p, q) &= \min_{\pi \in \Pi(p,q)} \sum_{x,y} d_{xy}\pi_{xy} \\
&\geq \min_{\pi \in \Pi(p,q)} \sum_{x \in S_-} \sum_y d_{xy}\pi_{xy} \\
&\geq \min_{\pi \in \Pi(p,q)} \sum_{x \in S_-} \sum_y d_{xy}(\pi_{xy} - \pi_{yx}) \\
&= \min_{\pi \in \Pi(p,q)} \sum_{x \in S_-} \sum_y (\pi_{xy} - \pi_{yx}) \\
&= \min_{\pi \in \Pi(p,q)} \sum_{x \in S_-} (q_x - p_x) \\
&= \frac{1}{2}\sum_x |p_x - q_x| \\
&= \mathcal{T}(p, q).
\end{aligned} \tag{D2}$$

Next, we show that this inequality can be attained with a specific coupling. Since $\sum_{x \in S_+}(p_x - q_x) = \sum_{x \in S_-}(q_x - p_x)$, according to Proposition 4, non-negative coefficients $\{z_{xy}\}$ defined over $S_- \times S_+$ always exist such that

$$\sum_{y \in S_+} z_{xy} = q_x - p_x \quad \forall x \in S_-, \tag{D3}$$

$$\sum_{y \in S_-} z_{yx} = p_x - q_x \quad \forall x \in S_+. \tag{D4}$$

We now construct a coupling $\pi = [\pi_{xy}]$ as follows:

$$\pi_{xx} = p_x \quad \forall x \in S_-, \tag{D5}$$

$$\pi_{xx} = q_x \quad \forall x \in S_+, \tag{D6}$$

$$\pi_{xy} = 0 \quad \forall x \in S_+ \quad \text{and} \quad y \neq x, \tag{D7}$$

$$\pi_{xy} = 0 \quad \forall y \in S_- \quad \text{and} \quad x \neq y, \tag{D8}$$

$$\pi_{xy} = z_{xy} \quad \text{otherwise}. \tag{D9}$$

We can verify that $\pi \in \Pi(p, q)$ and $\sum_{x,y} d_{xy}\pi_{xy} = \mathcal{T}(p, q)$. From the definition of the Wasserstein distance, we have

$$\mathcal{W}_1(p, q) \leq \sum_{x,y} d_{xy}\pi_{xy} = \mathcal{T}(p, q). \tag{D10}$$

Combining Eqs. (D2) and (D10) yields $\mathcal{W}_1(p, q) = \mathcal{T}(p, q)$.

2. Proof of Theorem 1

Here, we prove Theorem 1, which can be restated as

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \sqrt{\sigma_t m_t} dt = \min_{\mathsf{W}_t} \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \tag{D11}$$

To this end, we prove that rhs $\geq$ lhs and rhs $\leq$ lhs. First, we prove the former. According to the Cauchy-Schwarz inequality, we have

$$\sqrt{\Sigma_\tau \mathcal{M}_\tau} \geq \int_0^\tau \sqrt{\sigma_t m_t} dt, \tag{D12}$$

$$\begin{aligned}
\sqrt{\sigma_t m_t} &= \left(\sum_{x>y} m_{xy}(t) f_{xy}(t)^2 \sum_{x>y} m_{xy}(t)\right)^{1/2} \\
&\geq \sum_{x>y} \sqrt{m_{xy}(t) f_{xy}(t)^2} \sqrt{m_{xy}(t)} \\
&= \sum_{x>y} |j_{xy}(t)|.
\end{aligned} \tag{D13}$$

We then need only prove that

$$\int_0^\tau \sum_{x>y} |j_{xy}(t)| dt \geq \mathcal{W}_1(p^A, p^B). \quad (D14)$$

For this purpose, we map the optimal transport problem to a minimum cost flow problem. Let $\mathcal{G}(V, E)$ be the topology of Markov jump processes, from which the Wasserstein distance is defined. We consider a directed graph of $N + 2$ vertices: source vertex, target vertex, and N intermediate vertices $\{1, ..., N\}$ (see Fig. 8 for illustration). Each edge e of the graph is associated with a cost $c(e) \geq 0$ and capacity $a(e) > 0$ (i.e., the maximum flow that can be sent along this edge). The cost of sending a flow f along an edge e is, thus, $f \times c(e)$. The set of directed edges is as

$$\text{source} \rightarrow x: (c = 0, a = p_x^A), \quad (D15)$$

$$x \rightarrow \text{target}: (c = 0, a = p_x^B), \quad (D16)$$

$$x \leftrightarrow y: (c = 1, a = +\infty) \quad \text{if } (x, y) \in E. \quad (D17)$$

Consider a case in which an amount of flow 1 is sent from the source vertex to the target vertex. We can then prove that the minimum cost $\mathcal{C}$ of this flow problem is exactly the discrete Wasserstein distance. To this end, we first show that $\mathcal{C} \geq \mathcal{W}_1(p^A, p^B)$. Assume that $\mathcal{C}$ is attained by effectively sending a flow π_{xy} from source $\rightarrow y \rightarrow x \rightarrow$ target for each x and y. Since the shortest-path distance from y to x is d_{xy}, the total cost must be greater than or equal to $\sum_{x,y} d_{xy} \pi_{xy}$. Notice that $\{\pi_{xy}\}$ is a valid coupling. Therefore, we obtain $\mathcal{C} \geq \mathcal{W}_1(p^A, p^B)$ from the definition of the Wasserstein distance. We now need only prove
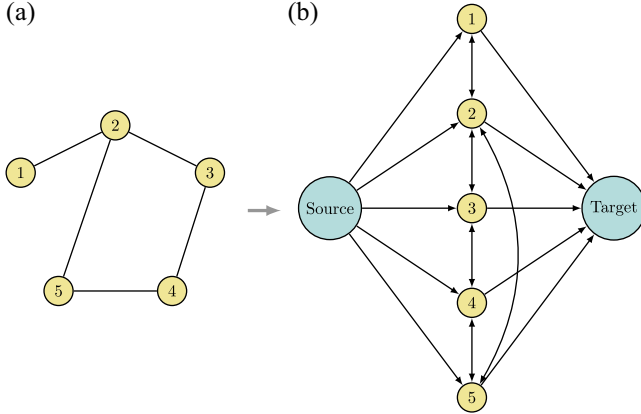
(a)   (b)

FIG. 8. Mapping the Wasserstein distance defined based on topology to a minimum cost flow problem. (a) Topology $\mathcal{G}(V, E)$ with five vertices and five edges, from which the Wasserstein distance is defined. (b) Directed graph of the minimum cost flow problem, which is constructed using the given topology $\mathcal{G}$. The minimum cost of the flow problem is equal to the Wasserstein distance.

the reverse statement $\mathcal{C} \leq \mathcal{W}_1(p^A, p^B)$. Assume that $\mathcal{W}_1(p^A, p^B)$ is achieved by an optimal transport plan $\pi^* = [\pi_{xy}^*]$ [i.e., for any pair (x, y), we move a probability π_{xy}^* from state y to state x with the cost of d_{xy} per unit probability]. For each x and y, let $P = [v_1, ..., v_k]$ be the shortest path of length d_{xy} that connects y to x; that is, $y = v_1$, $x = v_k$, $k - 1 = d_{xy}$, and $(v_i, v_{i+1}) \in E$ for all $1 \leq i < k$. We can then send an amount of flow π_{xy}^* along the path (source $\rightarrow v_1 \rightarrow \cdots \rightarrow v_k \rightarrow$ target). The total flow cost is exactly $\sum_{x,y} \pi_{xy}^* d_{xy} = \mathcal{W}_1(p^A, p^B)$; thus, $\mathcal{C} \leq \mathcal{W}_1(p^A, p^B)$. Consequently, we arrive at the equality $\mathcal{C} = \mathcal{W}_1(p^A, p^B)$.

We next show that the Markov jump process gives an admissible solution of the minimum cost flow problem with the cost $\int_0^\tau \sum_{x>y} |j_{xy}(t)| dt$. Consider discretization of the master equation with the time interval δt, where $\tau = K\delta t$. For each $k = 0, ..., K - 1$, we have

$$p_x[(k + 1)\delta t] = p_x(k\delta t) + \sum_{y(\neq x)} j_{xy}(k\delta t)\delta t. \quad (D18)$$

This means that we send an amount of flow $|j_{xy}(k\delta t)|\delta t$ from y to x if $j_{xy}(k\delta t) \geq 0$ and from x to y if $j_{xy}(k\delta t) < 0$. Since $j_{xy}(k\delta t) \neq 0$ only if x and y are directly connected by an edge, the cost of each transport is $|j_{xy}(k\delta t)|\delta t$. Therefore, the total flow cost associated with the Markov jump process is

$$\sum_{k=0}^{K-1} \sum_{x>y} |j_{xy}(k\delta t)|\delta t \xrightarrow{\delta t \rightarrow 0} \int_0^\tau \sum_{x>y} |j_{xy}(t)| dt. \quad (D19)$$

Since the Markov jump process realizes an admissible manner of sending flow from p^A to p^B, we obtain

$$\int_0^\tau \sum_{x>y} |j_{xy}(t)| dt \geq \mathcal{C} = \mathcal{W}_1(p^A, p^B), \quad (D20)$$

which verifies Eq. (D14). Consequently, combining Eqs. (D12)–(D14) yields

$$\text{rhs} \geq \min_{\mathsf{W}_t} \left\{ \int_0^\tau \sum_{x>y} |j_{xy}(t)| dt \right\}$$

$$\geq \mathcal{W}_1(p^A, p^B) = \text{lhs}. \quad (D21)$$

We next prove that lhs $\geq$ rhs by showing that the optimal cost $\mathcal{W}_1(p^A, p^B)$ can be achieved with a specific Markov jump process, the underlying graph of which is a subgraph of $\mathcal{G}(V, E)$ for all times. Note that the optimal transport plan can be represented as a sequence of transportation between neighboring states. Let $[(x_1, y_1, \chi_1), ..., (x_I, y_I, \chi_I)]$ denote the optimal transport plan; that is, at each step $1 \leq i \leq I$, we move a probability χ_i from state x_i to $y_i (\neq x_i)$. It is ensured

that at each step the probability of state x_i is always greater than or equal to χ_i. Since x_i and y_i are neighboring states, the total transport cost is $\sum_{i=1}^{I}\chi_i$. Thus,

$$\sum_{i=1}^{I}\chi_i = \mathcal{W}_1(p^A, p^B). \qquad (D22)$$

We now construct a Markov jump process of time period τ such that, for each $1 \leq i \leq I$, a probability χ_i is moved from x_i to y_i after time $t = i\Delta$, where $\Delta := \tau/I$. Specifically, we construct transition rates such that the probability distribution evolves as follows:

$$p_{x_i}[(i-1)\Delta + s] = p_{x_i}[(i-1)\Delta] - \frac{s}{\Delta}\chi_i, \qquad (D23)$$

$$p_{y_i}[(i-1)\Delta + s] = p_{y_i}[(i-1)\Delta] + \frac{s}{\Delta}\chi_i, \qquad (D24)$$

$$p_x[(i-1)\Delta + s] = p_x[(i-1)\Delta], \quad \forall x \neq x_i, y_i. \qquad (D25)$$

Here, $0 \leq s \leq \Delta$ is a time parameter. This time evolution of the probability distribution is effectively a two-level system, which can be realized using the following transition rates:

$$w_{y_i x_i}(t) = \frac{1}{\Delta(1 - e^{-\phi})}\frac{\chi_i}{p_{x_i}(t)}, \qquad (D26)$$

$$w_{x_i y_i}(t) = \frac{e^{-\phi}}{\Delta(1 - e^{-\phi})}\frac{\chi_i}{p_{y_i}(t)}, \qquad (D27)$$

$$w_{xy}(t) = 0 \quad \text{otherwise.} \qquad (D28)$$

Here, $\phi > 0$ is an arbitrary constant. During the time interval $[(i-1)\Delta, i\Delta]$, the underlying graph of this process has only one edge that connects vertices x_i and y_i. Thus, it is always a subgraph of $\mathcal{G}$. Using these transition rates, we can verify that

$$\dot{p}_x(t) = \sum_{y(\neq x)}j_{xy}(t) \quad \forall x. \qquad (D29)$$

Moreover,

$$\sigma_t = \sum_{x>y}j_{xy}(t)\ln\frac{a_{xy}(t)}{a_{yx}(t)} = \phi\sum_{x>y}|j_{xy}(t)|, \qquad (D30)$$

$$m_t = \sum_{x>y}\frac{j_{xy}(t)}{\ln[a_{xy}(t)/a_{yx}(t)]} = \frac{1}{\phi}\sum_{x>y}|j_{xy}(t)|. \qquad (D31)$$

In addition, note that

$$\sum_{x>y}|j_{xy}(t)| = \frac{\chi_i}{\Delta} \Rightarrow \int_{(i-1)\Delta}^{i\Delta}\sum_{x>y}|j_{xy}(t)|dt = \chi_i. \qquad (D32)$$

By summing both sides of Eq. (D32) for all $i = 1, \dots, I$, we obtain

$$\int_0^{\tau}\sum_{x>y}|j_{xy}(t)|dt = \sum_{i=1}^{I}\chi_i = \mathcal{W}_1(p^A, p^B). \qquad (D33)$$

Consequently, we have

$$\Sigma_\tau \mathcal{M}_\tau = \left(\int_0^{\tau}\sum_{x>y}|j_{xy}(t)|dt\right)^2$$
$$= \mathcal{W}_1(p^A, p^B)^2, \qquad (D34)$$

which completes the proof.

3. Equality in Theorem 1 can be achieved with global detailed-balance systems

Based on the previous construction of the dynamics that attains the equality in Theorem 1, we can further prove that the equality can be attained with global detailed-balance dynamics. Here, we prove this fact using a different approach.

Minimizing the integral term in Theorem 1 is equivalent to minimizing the cost function $\sigma_t m_t$ at each instance of time t. Consider the Lagrangian function

$$L(\mathsf{W}_t, \lambda) = \sigma_t m_t + \lambda^{\top}(\dot{p}_t - \mathsf{W}_t p_t), \qquad (D35)$$

where $\{p_t\}_{0 \leq t \leq \tau}$ is the probability distribution of a dynamics that attains the equality of Theorem 1. For simplicity, the time notation t is omitted hereafter. Taking the derivative of L with respect to w_{xy}, we have

$$\frac{\partial L}{\partial w_{xy}} = p_y(f_{xy} + 1 - e^{-f_{xy}})m + \frac{p_y(f_{xy} - 1 + e^{-f_{xy}})}{f_{xy}^2}\sigma$$
$$+ p_y(\lambda_y - \lambda_x) = 0. \qquad (D36)$$

Recall that $f_{xy} = \ln(a_{xy}/a_{yx})$. If $p_y = 0$, then w_{xy} can be arbitrarily determined. Therefore, we need only consider the nontrivial case $p_y \neq 0$. This leads to

$$(f_{xy} + 1 - e^{-f_{xy}})m + \frac{(f_{xy} - 1 + e^{-f_{xy}})}{f_{xy}^2}\sigma + \lambda_y - \lambda_x = 0. \qquad (D37)$$

Likewise, taking the derivative of L with respect to w_{yx} yields

$$(f_{yx} + 1 - e^{-f_{yx}})m + \frac{(f_{yx} - 1 + e^{-f_{yx}})}{f_{yx}^2}\sigma + \lambda_x - \lambda_y = 0.$$

(D38)

Notice that $f_{xy} = -f_{yx}$. Adding Eqs. (D37) and (D38) side by side, we obtain

$$(e^{f_{xy}} + e^{-f_{xy}} - 2)\left(\frac{\sigma}{f_{xy}^2} - m\right) = 0,$$

(D39)

which gives the solution $f_{xy} = 0$ or $f_{xy}^2 = \sigma/m$. Note that $f_{xy} = 0$ is equivalent to $j_{xy} = 0$, which implies that the transition between x and y does not contribute to the time evolution of the probability distribution. Therefore, such transitions need not be considered and can be eliminated by simply setting $w_{xy} = w_{yx} = 0$. Otherwise, if $f_{xy}^2 = \sigma/m$, then Eq. (D37) becomes

$$2f_{xy}m + \lambda_y - \lambda_x = 0,$$

(D40)

or, equivalently,

$$\ln\frac{w_{xy}}{w_{yx}} = \frac{\lambda_x - \lambda_y}{2m} + \ln p_x - \ln p_y.$$

(D41)

By defining an instantaneous energy $\beta\varepsilon_x := -\lambda_x/(2m) - \ln p_x$, we can verify that the transition rates satisfy the global detailed-balance condition:

$$\ln\frac{w_{xy}}{w_{yx}} = \beta(\varepsilon_y - \varepsilon_x).$$

(D42)

4. Minimum entropy production can be achieved with global detailed-balance systems

Here, we show that, given the time-averaged state mobility (i.e., $\langle m \rangle_\tau = \bar{D}$), there always exists a system that satisfies the global detailed balance and achieves the minimum entropy production:

$$\min_{\langle m \rangle_\tau = \bar{D}} \Sigma_\tau = \frac{\mathcal{W}_1(p_0, p_\tau)^2}{\bar{D}\tau}.$$

(D43)

According to the equality of Theorem 1 and Lemma 6, there exists a dynamics that satisfies Eq. (D43) with time-dependent probability distributions $\{p_t\}_{0 \le t \le \tau}$. We consider the following minimization problem:

$$\min \int_0^\tau \sigma_t dt,$$

(D44)

given that $\dot{p}_t = W_t p_t$ and $\langle m \rangle_\tau = \bar{D}$. Notice that the minimum value for this problem is exactly $\mathcal{W}_1(p_0, p_\tau)^2/(\bar{D}\tau)$. Consider the Lagrangian function

$$L(W_t, \lambda_t, \kappa) = \int_0^\tau \sigma_t dt + \int_0^\tau \lambda_t^\top(\dot{p}_t - W_t p_t)dt$$
$$+ \kappa\left(\int_0^\tau m_t dt - \bar{D}\tau\right).$$

(D45)

For simplicity, the time notation t is omitted hereafter. Taking the derivative of L with respect to w_{xy}, we have

$$\frac{\partial L}{\partial w_{xy}} = p_y(f_{xy} + 1 - e^{-f_{xy}}) + \frac{p_y(f_{xy} - 1 + e^{-f_{xy}})}{f_{xy}^2}\kappa$$
$$+ p_y(\lambda_y - \lambda_x) = 0.$$

(D46)

Following the same procedure as in Appendix D 3, we obtain the following relation:

$$2f_{xy} + \lambda_y - \lambda_x = 0,$$

(D47)

or, equivalently,

$$\ln\frac{w_{xy}}{w_{yx}} = \frac{\lambda_x - \lambda_y}{2} + \ln p_x - \ln p_y.$$

(D48)

By defining an instantaneous energy $\beta\varepsilon_x := -\lambda_x/2 - \ln p_x$, the transition rates satisfy the global detailed-balance condition:

$$\ln\frac{w_{xy}}{w_{yx}} = \beta(\varepsilon_y - \varepsilon_x).$$

(D49)

This means that the minimum entropy production (D44) can be achieved with conservative forces.

5. Particular topologies

a. Ring topology

Here, we consider a ring topology in which vertices x and $x + 1$ are connected for all x, where $N + 1 \equiv 1$. This topology can be seen in a one-dimensional asymmetric simple exclusion process on a ring of N sites and corresponds to a continuous-variable situation in which a single particle is driven in a periodic potential. For each integer number x, given $x = kN + r$, where $0 \le r \le N - 1$ is the remainder, we define $[x]_N := r$. Then, the shortest-path distance between states x and y can be calculated as

$$d_{xy} = \min\{[x - y]_N, N - [x - y]_N\}.$$

(D50)

In this case, the discrete Wasserstein distance can be written as

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \sum_{x=1}^N |j_{x+1,x}(t)| dt. \tag{D51}$$

We now consider the continuous case in which the particle is driven in a ring with a diameter $L = N\Delta x$. Taking the continuous limit of Eq. (D51), namely, $N \to \infty$ and $\Delta x \to 0$, we obtain the following relation:

$$\min_\pi \iint \min\{|x-y|, L - |x-y|\}\pi(x,y)dxdy$$
$$= \min_{j_t} \int_0^\tau \int_0^L |j_t(x)|dxdt, \tag{D52}$$

where $j_t(x)$ is subject to the continuity equation $\dot{p}_t(x) = -\partial_x j_t(x)$. The term on the left-hand side of Eq. (D52) is exactly the L^1-Wasserstein distance between probability distributions defined periodically over $[0, L]$ with the cost function

$$c(x,y) = \min\{|x-y|, L - |x-y|\}. \tag{D53}$$

Equation (D52), thus, provides a variational formula for the periodic L^1-Wasserstein distance.

b. Fully connected topology

Another topology is the fully connected topology; that is, for an arbitrary pair of two vertices, an edge always exists that connects them. In this case, the shortest-path distances become

$$d_{xy} = 1 - \delta_{xy}, \tag{D54}$$

and the discrete Wasserstein distance equals the total variation distance. Theorem 1 thus implies the following equality:

$$\mathcal{T}(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \sqrt{\sigma_t m_t} dt = \min_{\mathsf{W}_t} \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \tag{D55}$$

Here, the minimum is taken over all possible transition rate matrices; that is, the transition rate between any two states can be arbitrarily controlled. Although the total variation distance is widely used in previous studies, its connection with thermodynamics has thus far been veiled. Equation (D55) reveals a thermodynamic interpretation of this distance, showing that it equals the minimum product of the thermodynamic and kinetic costs given the full control of the transition rates.

6. Alternative variational expressions of the discrete Wasserstein distance

Corollary 7: The discrete Wasserstein distance can be expressed in terms of irreversible entropy production and dynamical activity as

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} dt \tag{D56}$$

$$= \min_{\mathsf{W}_t} \frac{\Sigma_\tau}{2} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1}. \tag{D57}$$

Proof.—We first prove that

$$\int_0^\tau \frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} dt \geq \mathcal{W}_1(p^A, p^B). \tag{D58}$$

Note that $x\Phi(x/y)^{-1}$ is a concave function over $(0, +\infty) \times (0, +\infty)$. Applying Jensen's inequality yields

$$\frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} \geq \sum_{x>y} \frac{\sigma_{xy}(t)}{2} \Phi\left(\frac{\sigma_{xy}(t)}{2[a_{xy}(t) + a_{yx}(t)]}\right)^{-1}$$
$$= \sum_{x>y} |j_{xy}(t)|. \tag{D59}$$

By taking the time integration of Eq. (D59) and using Eq. (84), we immediately prove Eq. (D58). Moreover, using the concavity of $x\Phi(x/y)^{-1}$ yields

$$\int_0^\tau \frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} dt \leq \frac{\Sigma_\tau}{2} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1}$$
$$\leq \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \tag{D60}$$

Thus, we have

$$\mathcal{W}_1(p^A, p^B) \leq \int_0^\tau \frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} dt \leq \frac{\Sigma_\tau}{2} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1}$$
$$\leq \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \tag{D61}$$

The proof is completed by taking the minimum of the terms on the right-hand side of Eq. (D61) over all admissible dynamics and applying Theorem 1. ∎

Equation (D57) implies that the discrete Wasserstein distance can be expressed in terms of irreversible entropy production and dynamical activity as

$$\mathcal{W}_1(p^A, p^B) = \min \frac{\Sigma_\tau}{2} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1}, \tag{D62}$$

which recovers the result obtained in Ref. [93].

Corollary 8: The discrete Wasserstein distance can be expressed in terms of pseudo entropy production and dynamical activity as

$$\mathcal{W}_1(p^A, p^B) = \min_{\mathsf{W}_t} \int_0^\tau \sqrt{\sigma_t^{\mathrm{ps}} a_t} \, dt \qquad \text{(D63)}$$

$$= \min_{\mathsf{W}_t} \sqrt{\Sigma_\tau^{\mathrm{ps}} \mathcal{A}_\tau}, \qquad \text{(D64)}$$

where $\sigma_t^{\mathrm{ps}} \coloneqq \dot{\Sigma}_t^{\mathrm{ps}}$ denotes the pseudo entropy production rate.

Proof.—The proof strategy is the same as in Corollary 7. We first prove that

$$\int_0^\tau \sqrt{\sigma_t^{\mathrm{ps}} a_t} \, dt \geq \mathcal{W}_1(p^A, p^B). \qquad \text{(D65)}$$

Applying the Cauchy-Schwarz inequality, we obtain

$$\sum_{x>y} |j_{xy}(t)| = \sum_{x>y} \frac{|j_{xy}(t)|}{\sqrt{a_{xy}(t) + a_{yx}(t)}} \sqrt{a_{xy}(t) + a_{yx}(t)}$$

$$\leq \sqrt{\sigma_t^{\mathrm{ps}} a_t}. \qquad \text{(D66)}$$

By taking the time integration of Eq. (D66) and using Eq. (84), we immediately prove Eq. (D65). Since $\Sigma_\tau^{\mathrm{ps}} \mathcal{A}_\tau \leq \Sigma_\tau \mathcal{M}_\tau$, the following relation holds:

$$\mathcal{W}_1(p^A, p^B) \leq \int_0^\tau \sqrt{\sigma_t^{\mathrm{ps}} a_t} \, dt \leq \sqrt{\Sigma_\tau^{\mathrm{ps}} \mathcal{A}_\tau} \leq \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \qquad \text{(D67)}$$

Taking the minimum of the terms on the right-hand side of Eq. (D67) over all admissible dynamics and using Theorem 1 complete the proof. ∎

Equation (D64) has the following implication. If dynamical activity $\mathcal{A}_\tau$ is fixed, then the minimum pseudo entropy production can be calculated using the Wasserstein distance as

$$\min \Sigma_\tau^{\mathrm{ps}} = \frac{\mathcal{W}_1(p^A, p^B)^2}{\mathcal{A}_\tau}, \qquad \text{(D68)}$$

which recovers the result reported in Ref. [93]. From Theorem 1 and Corollary 8, we can observe that, in the context of optimal transport, $(\Sigma_\tau, \mathcal{M}_\tau)$ and $(\Sigma_\tau^{\mathrm{ps}}, \mathcal{A}_\tau)$ are two thermodynamic-kinetic conjugate pairs.

APPENDIX E: DERIVATION OF CALCULATIONS IN SEC. V

1. Proof of Eq. (99)

Here, we derive an analytical expression of the entropy production rate σ_t. Taking the time derivative of irreversible entropy production, we can calculate the entropy production rate as

$$\sigma_t = -\mathrm{tr}\{\dot{\varrho}_t \ln \varrho_t\} + \sum_k \mathrm{tr}\{L_k(t)^\dagger L_k(t) \varrho_t\} s_k(t)$$

$$= \sum_k -\mathrm{tr}\{(\mathcal{D}[L_k(t)]\varrho_t) \ln \varrho_t\} + \mathrm{tr}\{L_k(t)^\dagger L_k(t) \varrho_t\} s_k(t)$$

$$= \sum_k \mathrm{tr}\{L_k(t)\varrho_t(s_k(t)L_k(t)^\dagger - [L_k(t)^\dagger, \ln \varrho_t])\}. \qquad \text{(E1)}$$

Notice that $w_k^{xy}(t) = e^{s_k(t)} w_{k'}^{yx}(t)$. Since $\mathrm{tr}\, A = \sum_x \langle x_t | A | x_t \rangle$ for any operator A, the entropy production rate can be calculated further as

$$\sigma_t = \sum_k \sum_x \langle x_t | L_k(t)\varrho_t(s_k(t)L_k(t)^\dagger - [L_k(t)^\dagger, \ln \varrho_t]) | x_t \rangle$$

$$= \sum_k \sum_{x,y} w_k^{xy}(t) p_y(t) \left[s_k(t) + \ln \frac{p_y(t)}{p_x(t)} \right]$$

$$= \frac{1}{2} \sum_k \sum_{x,y} j_k^{xy}(t) \ln \frac{w_k^{xy}(t) p_y(t)}{w_{k'}^{yx}(t) p_x(t)}. \qquad \text{(E2)}$$

Since $(a - b)\ln(a/b) \geq 0$ for all $a, b \geq 0$, the positivity of σ_t is immediately derived.

2. Proof of Eq. (115)

First, quantum dynamical state mobility can be expressed in terms of eigenvalues $\{p_x(t)\}$ and transition rates $\{w_k^{xy}(t)\}$ as

$$m_t = \frac{1}{2} \sum_k e^{-s_k(t)/2} \langle L_k(t)^\dagger, [\![\varrho_t]\!]_{s_k(t)} \{\mathcal{P}_t[L_k(t)^\dagger]\} \rangle$$

$$= \frac{1}{2} \sum_k e^{-s_k(t)} \int_0^1 e^{s_k(t)u} \langle L_k(t)^\dagger, \varrho_t^u \{\mathcal{P}_t[L_k(t)^\dagger]\} \varrho_t^{1-u} \rangle \, du$$

$$= \frac{1}{2} \sum_k e^{-s_k(t)} \sum_{x \neq y} \int_0^1 e^{s_k(t)u} w_k^{xy}(t) p_y(t)^u p_x(t)^{1-u} \, du. \qquad \text{(E3)}$$

Likewise, we can calculate

$$\sum_x \langle x_t| \otimes |x_t\rangle^\top O_k(t,u)|x_t\rangle \otimes \langle x_t|^\top$$

$$= \sum_x \langle x_t| \otimes |x_t\rangle^\top L_k(t)\varrho_t^u L_k(t)^\dagger \otimes (\varrho_t^{1-u})^\top |x_t\rangle \otimes \langle x_t|^\top + \sum_x \langle x_t| \otimes |x_t\rangle^\top \varrho_t^u \otimes [L_k(t)^\dagger \varrho_t^{1-u} L_k(t)]^\top |x_t\rangle \otimes \langle x_t|^\top$$

$$- \sum_x \langle x_t| \otimes |x_t\rangle^\top L_k(t)\varrho_t^u \otimes [L_k(t)^\dagger \varrho_t^{1-u}]^\top |x_t\rangle \otimes \langle x_t|^\top - \sum_x \langle x_t| \otimes |x_t\rangle^\top \varrho_t^u L_k(t)^\dagger \otimes [\varrho_t^{1-u} L_k(t)]^\top |x_t\rangle \otimes \langle x_t|^\top$$

$$= \sum_{x\neq y} [w_k^{xy}(t)p_x(t)^{1-u}p_y(t)^u + w_k^{yx}(t)p_x(t)^u p_y(t)^{1-u}]$$

$$= 2\sum_{x\neq y} w_k^{xy}(t)p_y(t)^u p_x(t)^{1-u}. \tag{E4}$$

Consequently, combining Eqs. (E3) and (E4) yields the desired relation:

$$\frac{1}{2}\sum_x \langle x_t| \otimes |x_t\rangle^\top O_t|x_t\rangle \otimes \langle x_t|^\top = \frac{1}{4}\sum_x \sum_k e^{-s_k(t)} \int_0^1 e^{s_k(t)u} \langle x_t| \otimes |x_t\rangle^\top O_k(t,u)|x_t\rangle \otimes \langle x_t|^\top du$$

$$= \frac{1}{2}\sum_k e^{-s_k(t)} \sum_{x\neq y} \int_0^1 e^{s_k(t)u} w_k^{xy}(t)p_y(t)^u p_x(t)^{1-u} du$$

$$= m_t. \tag{E5}$$

3. Proof of Eq. (119)

First, we prove that $\mathcal{W}_q(\varrho^A, \varrho^B) \leq \mathcal{T}(p^A, p^B)$. Let $\varrho^A = \sum_x p_x^A |x^A\rangle\langle x^A|$ and $\varrho^B = \sum_x p_x^B |x^B\rangle\langle x^B|$ be the spectral decompositions of the density matrices. Setting $V_* = \sum_x |x^B\rangle\langle x^B|x^A\rangle$, we can verify that V_* is a unitary operator and $V_*\varrho^A V_*^\dagger = \sum_x p_x^A |x^B\rangle\langle x^B|$. From the definition of $\mathcal{W}_q$, we have

$$\mathcal{W}_q(\varrho^A, \varrho^B) = \frac{1}{2}\min_{V^\dagger V=\mathbb{1}} \|V\varrho^A V^\dagger - \varrho^B\|_1$$

$$\leq \frac{1}{2}\|V_*\varrho^A V_*^\dagger - \varrho^B\|_1$$

$$= \frac{1}{2}\left\|\sum_x (p_x^A - p_x^B)|x^B\rangle\langle x^B|\right\|_1$$

$$= \frac{1}{2}\sum_x |p_x^A - p_x^B| = \mathcal{T}(p^A, p^B). \tag{E6}$$

Note that $\{p_x^A\}$ are increasing eigenvalues of $V\varrho^A V^\dagger$ for an arbitrary unitary operator V. Let $\varsigma_x(A)$ be the xth singular value of operator A in ascending order. Then,

$$\|A - B\|_1 = \sum_x \varsigma_x(A - B) \geq \sum_x |\varsigma_x(A) - \varsigma_x(B)| \tag{E7}$$

holds for arbitrary Hermitian operators A and B [174]. Applying the above inequality for $A = V\varrho^A V^\dagger$ and $B = \varrho^B$ yields $\|V\varrho^A V^\dagger - \varrho^B\|_1 \geq \sum_x |p_x^A - p_x^B|$, from which we immediately obtain $\mathcal{W}_q(\varrho^A, \varrho^B) \geq \mathcal{T}(p^A, p^B)$. Consequently, Eq. (119) is proved.

4. Proof of Theorem 2

Here, we prove Theorem 2, which can be restated as

$$\mathcal{W}_q(\varrho^A, \varrho^B) = \min_{\mathcal{L}_t} \int_0^\tau \sqrt{\sigma_t m_t}\, dt = \min_{\mathcal{L}_t} \sqrt{\Sigma_\tau \mathcal{M}_\tau}. \tag{E8}$$

First, we prove that rhs $\geq$ lhs. Note that

$$\sigma_t = \frac{1}{2}\sum_k \sum_{x,y} j_k^{xy}(t) \ln\frac{a_k^{xy}(t)}{a_{k'}^{yx}(t)}, \tag{E9}$$

$$m_t = \frac{1}{2}\sum_k \sum_{x\neq y} \frac{j_k^{xy}(t)}{\ln[a_k^{xy}(t)/a_{k'}^{yx}(t)]}. \tag{E10}$$

Applying the Cauchy-Schwarz inequality and triangle inequality, we obtain

$$\sqrt{\sigma_t m_t} \geq \frac{1}{2}\sum_k \sum_{x\neq y} |j_k^{xy}(t)|$$

$$\geq \frac{1}{2}\sum_x |\sum_k \sum_{y(\neq x)} j_k^{xy}(t)|$$

$$= \frac{1}{2}\sum_x |\dot{p}_x(t)|. \tag{E11}$$

Consequently, taking the time integration and applying Proposition 3 yield the following result:

$$
\begin{aligned}
\min_{\mathcal{L}_t} \sqrt{\Sigma_\tau \mathcal{M}_\tau} &\geq \min_{\mathcal{L}_t} \int_0^\tau \sqrt{\sigma_t m_t}\, dt \\
&\geq \frac{1}{2} \min_{\mathcal{L}_t} \sum_x \int_0^\tau |\dot{p}_x(t)|\, dt \\
&\geq \frac{1}{2} \min_{\mathcal{L}_t} \sum_x \Big| \int_0^\tau \dot{p}_x(t)\, dt \Big| \\
&= \frac{1}{2} \min_{\mathcal{L}_t} \sum_x |p_x(\tau) - p_x(0)| \\
&\geq \mathcal{W}_q(\varrho^A, \varrho^B).
\end{aligned}
\tag{E12}
$$

Next, we need only show that the equality in the inequality (E12) can be achieved with particular dynamics. First, we construct a Markov jump process with the transition rate matrix $\{W_t\}$ that transforms the initial distribution p^A into the final distribution p^B. Let us consider probability path $p_t = p^A + t(p^B - p^A)/\tau$. We then have $\dot{p}_x(t) = (p_x^B - p_x^A)/\tau$, which is invariant for all times t. We next define $S_+ := \{x | p_x^B \geq p_x^A\}$ and $S_- := \{x | p_x^B < p_x^A\}$. Then, $\sum_{x \in S_+} \dot{p}_x(t) = -\sum_{x \in S_-} \dot{p}_x(t)$. Let $\phi > 0$ be an arbitrary real positive number. According to Proposition 4, non-negative coefficients $\{z_{xy}\}$ exist such that

$$
\sum_{y \in S_-} z_{xy} = \frac{\dot{p}_x(t)}{1 - e^{-\phi}} \quad \forall\, x \in S_+,
\tag{E13}
$$

$$
\sum_{x \in S_+} z_{xy} = \frac{-\dot{p}_y(t)}{1 - e^{-\phi}} \quad \forall\, y \in S_-.
\tag{E14}
$$

Using these coefficients, we consider the following transition rates:

$$
w_{xy}(t) = \frac{z_{xy}}{p_y(t)} \quad \forall\, x \in S_+, y \in S_-,
\tag{E15}
$$

$$
w_{yx}(t) = e^{-\phi} \frac{z_{xy}}{p_x(t)} \quad \forall\, x \in S_+, y \in S_-,
\tag{E16}
$$

$$
w_{xy}(t) = 0 \quad \text{otherwise.}
\tag{E17}
$$

With these transition rates, we can verify that

$$
\dot{p}_x(t) = \sum_{y(\neq x)} j_{xy}(t) \quad \forall\, x.
\tag{E18}
$$

Moreover, the irreversible entropy production rate and dynamical state mobility associated with this Markov jump process can be calculated as

$$
\sigma_t = \sum_{x \in S_+, y \in S_-} j_{xy}(t) \ln \frac{a_{xy}(t)}{a_{yx}(t)} = \phi \sum_{x > y} |j_{xy}(t)|,
\tag{E19}
$$

$$
m_t = \sum_{x \in S_+, y \in S_-} \frac{j_{xy}(t)}{\ln[a_{xy}(t)/a_{yx}(t)]} = \phi^{-1} \sum_{x > y} |j_{xy}(t)|.
\tag{E20}
$$

In addition, note that

$$
\begin{aligned}
\sum_{x > y} |j_{xy}(t)| &= (1 - e^{-\phi}) \sum_{x \in S_+} \sum_{y \in S_-} z_{xy} \\
&= \sum_{x \in S_+} \dot{p}_x(t).
\end{aligned}
\tag{E21}
$$

Consequently, we have

$$
\begin{aligned}
\sqrt{\Sigma_\tau \mathcal{M}_\tau} &= \int_0^\tau \sum_{x > y} |j_{xy}(t)|\, dt \\
&= \sum_{x \in S_+} \int_0^\tau \dot{p}_x(t)\, dt \\
&= \sum_{x \in S_+} [p_x(\tau) - p_x(0)] \\
&= \sum_{x \in S_+} (p_x^B - p_x^A) \\
&= \mathcal{T}(p^A, p^B) \\
&= \mathcal{W}_q(\varrho^A, \varrho^B).
\end{aligned}
\tag{E22}
$$

We next construct Lindblad dynamics that transforms ϱ^A into ϱ^B and simultaneously satisfies the equality (E22). For each pair of positive transition rates $\{w_{xy}(t), w_{yx}(t)\}$, we define the corresponding jump operators $\tilde{L}_k(t) = \sqrt{w_{xy}(t)}|x^A\rangle\langle y^A|$ and $\tilde{L}_{k'}(t) = \sqrt{w_{yx}(t)}|y^A\rangle\langle x^A|$. We consider the following Lindblad equation:

$$
\dot{\tilde{\varrho}}_t = \sum_k D[\tilde{L}_k(t)]\tilde{\varrho}_t.
\tag{E23}
$$

As the initial state is diagonal in the eigenbasis $\{|x^A\rangle\}$, Eq. (E23) is equivalent to the classical Markov jump process previously constructed. Given the initial state $\tilde{\varrho}_0 = \varrho^A$, we can easily see that $\tilde{\varrho}_t$ is always diagonal in the eigenbasis $\{|x^A\rangle\}$ [i.e., $\tilde{\varrho}_t = \sum_x p_x(t)|x^A\rangle\langle x^A|$]. Moreover, from Eq. (E22), it is evident that

$$
\sqrt{\tilde{\Sigma}_\tau \tilde{\mathcal{M}}_\tau} = \mathcal{W}_q(\varrho^A, \varrho^B).
\tag{E24}
$$

Now, consider the unitary operator $U_\tau = \sum_x |x^B\rangle\langle x^A|$. A Hermitian Hamiltonian H exists such that $U_\tau = e^{-iH\tau}$.

Using this Hamiltonian, we consider the following Lindblad dynamics:

$$\dot{\varrho}_t = -i[H, \varrho_t] + \sum_k D[L_k(t)]\varrho_t, \qquad (E25)$$

where jump operators are given by $L_k(t) = U_t \tilde{L}_k(t) U_t^\dagger$ and $U_t := e^{-iHt}$. The density matrix ϱ_t is related to that in Eq. (E23) as $\varrho_t = U_t \tilde{\varrho}_t U_t^\dagger$. We can confirm that the dynamics (E25) transforms the density matrix $\varrho_0 = \varrho^A$ into $\varrho_\tau = \varrho^B$, and irreversible entropy production and dynamical state mobility remain unchanged:

$$\Sigma_\tau = \tilde{\Sigma}_\tau, \qquad (E26)$$

$$\mathcal{M}_\tau = \tilde{\mathcal{M}}_\tau. \qquad (E27)$$

Combining this with Eq. (E24), we can show that the inequality (E12) can be saturated as

$$\sqrt{\Sigma_\tau \mathcal{M}_\tau} = \mathcal{W}_q(\varrho^A, \varrho^B). \qquad (E28)$$

5. Quantum variational formula in terms of entropy production and dynamical activity

Corollary 9: The quantum Wasserstein distance can be expressed in terms of irreversible entropy production and dynamical activity as

$$\mathcal{W}_q(\varrho^A, \varrho^B) = \min_{\mathcal{L}_t} \int_0^\tau \frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} dt \qquad (E29)$$

$$= \min_{\mathcal{L}_t} \frac{\Sigma_\tau}{2} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1}. \qquad (E30)$$

Proof.—First, we prove that

$$\mathcal{W}_q(\varrho^A, \varrho^B) \le \int_0^\tau \frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} dt \le \frac{\Sigma_\tau}{2} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1}. \qquad (E31)$$

Noting that $x\Phi(x/y)^{-1}$ is a concave function and

$$\sigma_t = \frac{1}{2}\sum_{k,x,y} j_k^{xy}(t) \ln \frac{a_k^{xy}(t)}{a_{k'}^{yx}(t)} =: \frac{1}{2}\sum_{k,x,y} \sigma_k^{xy}(t), \quad (E32)$$

$$a_t = \frac{1}{2}\sum_{k,x,y} [a_k^{xy}(t) + a_{k'}^{yx}(t)], \qquad (E33)$$

we obtain the following result from Jensen's inequality:

$$\frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} \ge \sum_{k,x,y} \frac{\sigma_k^{xy}(t)}{4} \Phi\left(\frac{\sigma_k^{xy}(t)}{2[a_k^{xy}(t) + a_{k'}^{yx}(t)]}\right)^{-1}$$

$$= \frac{1}{2}\sum_{k,x,y} |j_k^{xy}(t)|$$

$$\ge \frac{1}{2}\sum_x |\dot{p}_x(t)|. \qquad (E34)$$

Taking the time integration, we can immediately prove Eq. (E31):

$$\mathcal{W}_q(\varrho^A, \varrho^B) \le \frac{1}{2}\int_0^\tau \sum_x |\dot{p}_x(t)| dt$$

$$\le \int_0^\tau \frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} dt$$

$$\le \frac{\Sigma_\tau}{2} \Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1}. \qquad (E35)$$

Next, we show that the equalities in Eq. (E31) can be attained with the dynamics constructed in the proof of Theorem 2. Notice that the density matrix $\tilde{\varrho}_t$ of Lindblad dynamics (E23) can be expressed as $\tilde{\varrho}_t = \sum_x p_x(t)|x^A\rangle\langle x^A|$. Therefore, the density matrix ϱ_t of Lindblad dynamics (E25) reads $\varrho_t = \sum_x p_x(t)U_t|x^A\rangle\langle x^A|U_t^\dagger$, the time-dependent eigenvectors of which are $|x_t\rangle = U_t|x^A\rangle$. The jump operators are given by $L_k(t) = \sqrt{w_{xy}(t)}U_t|x^A\rangle\langle y^A|U_t^\dagger$. Using these quantities, we can calculate

$$w_k^{x'y'}(t) = \delta_{xx'}\delta_{yy'}w_{xy}(t). \qquad (E36)$$

In addition, the entropy production and dynamical activity rates can be calculated as

$$\sigma_t = \phi\sum_{x>y} |j_{xy}(t)|, \qquad (E37)$$

$$a_t = \coth(\phi/2)\sum_{x>y} |j_{xy}(t)|, \qquad (E38)$$

where $\{j_{xy}(t)\}$ are probability currents in the classical Markov jump process. Consequently, we obtain the following relations:

$$\frac{\sigma_t}{2} \Phi\left(\frac{\sigma_t}{2a_t}\right)^{-1} = \frac{1}{2}\phi\sum_{x>y} |j_{xy}(t)|\Phi\left(\frac{\phi}{2\coth(\phi/2)}\right)^{-1}$$

$$= \sum_{x>y} |j_{xy}(t)|, \qquad (E39)$$

$$\frac{\Sigma_\tau}{2}\Phi\left(\frac{\Sigma_\tau}{2\mathcal{A}_\tau}\right)^{-1} = \int_0^\tau \sum_{x>y}|j_{xy}(t)|dt. \qquad \text{(E40)}$$

Combining Eqs. (E39), (E40), and (E22) verifies the equalities of Eq. (E31). ∎

APPENDIX F: DERIVATION OF CALCULATIONS IN SEC. VI

1. Thermodynamic speed limit in terms of the trace distance

The Wasserstein distance is used as a metric between quantum states in the speed limits in Eq. (137). Here, we show that another thermodynamic speed limit with a different metric can also be obtained. Specifically, we derive a speed limit using the trace distance in the following. Let $\varrho_t = \sum_x p_x(t)|x_t\rangle\langle x_t|$ be the spectral decomposition of the density matrix ϱ_t. Then, as previously shown in Ref. [18], we have

$$\|\dot{\varrho}_t\|_1 \le 2(\Delta H_t + \Delta H_t^D) + \sum_x |\dot{p}_x(t)|, \qquad \text{(F1)}$$

where

$$(\Delta H_t)^2 = \text{tr}\{H_t^2\varrho_t\} - (\text{tr}\{H_t\varrho_t\})^2, \qquad \text{(F2)}$$

$$(\Delta H_t^D)^2 = \text{tr}\{(H_t^D)^2\varrho_t\} - (\text{tr}\{H_t^D\varrho_t\})^2, \qquad \text{(F3)}$$

$$H_t^D := \sum_{x\neq y} \frac{i\langle x_t|\sum_k \mathcal{D}[L_k]\varrho_t|y_t\rangle}{p_y(t) - p_x(t)}|x_t\rangle\langle y_t|. \qquad \text{(F4)}$$

In addition, as shown in Eq. (E11), we can prove that

$$\sum_x |\dot{p}_x(t)| \le 2\sqrt{\sigma_t m_t}. \qquad \text{(F5)}$$

Taking the time integration and using the triangle inequality for the trace norm, we obtain

$$\begin{aligned}
\mathcal{T}(\varrho_0, \varrho_\tau) &\le \frac{1}{2}\int_0^\tau \|\dot{\varrho}_t\|_1 dt \\
&\le \tau\langle\Delta H + \Delta H^D + \sqrt{\sigma m}\rangle_\tau, \qquad \text{(F6)}
\end{aligned}$$

which yields the following speed limit:

$$\begin{aligned}
\tau &\ge \frac{\mathcal{T}(\varrho_0, \varrho_\tau)}{\langle\Delta H + \Delta H^D + \sqrt{\sigma m}\rangle_\tau} \\
&\ge \frac{\mathcal{T}(\varrho_0, \varrho_\tau)}{\langle\Delta H\rangle_\tau + \langle\Delta H^D\rangle_\tau + \sqrt{\langle\sigma\rangle_\tau\langle m\rangle_\tau}}. \qquad \text{(F7)}
\end{aligned}$$

Since $\langle m\rangle_\tau \le \langle a\rangle_\tau/2$, this new speed limit is stronger than the bound reported in Ref. [18], which reads

$$\tau \ge \frac{\mathcal{T}(\varrho_0, \varrho_\tau)}{\langle\Delta H\rangle_\tau + \langle\Delta H^D\rangle_\tau + \sqrt{\langle\sigma\rangle_\tau\langle a\rangle_\tau/2}}. \qquad \text{(F8)}$$

In the classical limit, the speed limit (F7) reduces to the following bound:

$$\tau \ge \frac{\mathcal{T}(p^A, p^B)}{\langle\sqrt{\sigma m}\rangle_\tau} \ge \frac{\mathcal{T}(p^A, p^B)}{\sqrt{\langle\sigma\rangle_\tau\langle m\rangle_\tau}}. \qquad \text{(F9)}$$

2. Proof of Eq. (141)

Equation (141) is the consequence of the following lemma.

Lemma 10: If an erasure protocol satisfies $\|\Lambda_\tau\bar{p} - p_*\|_F \le \delta$, where $\delta > 0$ is a sufficiently small number, then, for an arbitrary probability distribution p_0, the following inequality holds:

$$\|\Lambda_\tau p_0 - p_*\|_F \le \sqrt{2d\delta} \xrightarrow{\delta\to 0} 0. \qquad \text{(F10)}$$

Proof.—For any probability distribution p_0, a distribution p_0' always exists such that $p_0 + (d-1)p_0' = \mathbf{1} = dp_*$. Indeed, the distribution p_0' can be chosen as $p_0' = (\mathbf{1} - p_0)/(d-1)$. Here, $\mathbf{1} = [1, \ldots, 1]^\top$ is the all-one vector. We then define $p_\tau := \Lambda_\tau p_0$ and $p_\tau' := \Lambda_\tau p_0'$ and obtain the following relation:

$$\Lambda_\tau\bar{p} = \frac{1}{d}\Lambda_\tau[p_0 + (d-1)p_0'] = \frac{1}{d}p_\tau + \frac{d-1}{d}p_\tau'. \qquad \text{(F11)}$$

Therefore, the condition $\|\Lambda_\tau\bar{p} - p_*\|_F \le \delta$ is equivalent to

$$\|[p_\tau + (d-1)p_\tau']/d - p_*\|_F^2 \le \delta^2. \qquad \text{(F12)}$$

It suffices to prove that $\|p_\tau - p_*\|_F^2 \le 2d\delta$. From Eq. (F12), we have

$$\begin{aligned}
&|[p_{\tau,1} + (d-1)p_{\tau,1}']/d - p_{*,1}|^2 \\
&\le \|[p_\tau + (d-1)p_\tau']/d - p_*\|_F^2 \\
&\le \delta^2. \qquad \text{(F13)}
\end{aligned}$$

Consequently,

$$\frac{1 - p_{\tau,1}}{d} + (d-1)\frac{1 - p_{\tau,1}'}{d} \le \delta \Rightarrow 1 - d\delta \le p_{\tau,1} \le 1. \qquad \text{(F14)}$$

The last inequality in Eq. (F14) immediately derives $|1 - p_{\tau,1}| \le d\delta$ and $|p_{\tau,1}| \ge 1 - d\delta \ge 0$. From the inequality $\sum_{n=1}^d |p_{\tau,n}|^2 \le \sum_{n=1}^d p_{\tau,n} = 1$, the partial sum $\sum_{n=2}^d |p_{\tau,n}|^2$ can be upper bounded as

$$\sum_{n=2}^d |p_{\tau,n}|^2 \le 1 - |p_{\tau,1}|^2 \le 1 - (1 - d\delta)^2 = 2d\delta - d^2\delta^2. \qquad \text{(F15)}$$

Combining these inequalities, we obtain

$$\|p_\tau - p_*\|_F^2 = |1 - p_{\tau,1}|^2 + \sum_{n=2}^{d} |p_{\tau,n}|^2$$

$$\leq d^2\delta^2 + 2d\delta - d^2\delta^2 = 2d\delta, \quad \text{(F16)}$$

which completes the proof. ∎

---

[1] K. Sekimoto, *Stochastic Energetics* (Springer, Berlin, 2010), Vol. 799.

[2] U. Seifert, *Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines*, Rep. Prog. Phys. **75**, 126001 (2012).

[3] S. Vinjanampathy and J. Anders, *Quantum Thermodynamics*, Contemp. Phys. **57**, 545 (2016).

[4] J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, *The Role of Quantum Information in Thermodynamics—A Topical Review*, J. Phys. A **49**, 143001 (2016).

[5] S. Deffner and S. Campbell, *Quantum Thermodynamics* (Morgan & Claypool, San Rafael, 2019).

[6] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, *Probability of Second Law Violations in Shearing Steady States*, Phys. Rev. Lett. **71**, 2401 (1993).

[7] G. Gallavotti and E. G. D. Cohen, *Dynamical Ensembles in Nonequilibrium Statistical Mechanics*, Phys. Rev. Lett. **74**, 2694 (1995).

[8] G. E. Crooks, *Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences*, Phys. Rev. E **60**, 2721 (1999).

[9] C. Jarzynski, *Hamiltonian Derivation of a Detailed Fluctuation Theorem*, J. Stat. Phys. **98**, 77 (2000).

[10] M. Esposito, U. Harbola, and S. Mukamel, *Nonequilibrium Fluctuations, Fluctuation Theorems, and Counting Statistics in Quantum Systems*, Rev. Mod. Phys. **81**, 1665 (2009).

[11] M. Campisi, P. Hänggi, and P. Talkner, *Colloquium: Quantum Fluctuation Relations: Foundations and Applications*, Rev. Mod. Phys. **83**, 771 (2011).

[12] A. C. Barato and U. Seifert, *Thermodynamic Uncertainty Relation for Biomolecular Processes*, Phys. Rev. Lett. **114**, 158101 (2015).

[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] 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).

[15] J. M. Horowitz and T. R. Gingrich, *Thermodynamic Uncertainty Relations Constrain Non-equilibrium Fluctuations*, Nat. Phys. **16**, 15 (2020).

[16] N. Shiraishi, K. Funo, and K. Saito, *Speed Limit for Classical Stochastic Processes*, Phys. Rev. Lett. **121**, 070601 (2018).

[17] S. Ito, *Stochastic Thermodynamic Interpretation of Information Geometry*, Phys. Rev. Lett. **121**, 030605 (2018).

[18] K. Funo, N. Shiraishi, and K. Saito, *Speed Limit for Open Quantum Systems*, New J. Phys. **21**, 013006 (2019).

[19] S. Ito and A. Dechant, *Stochastic Time Evolution, Information Geometry, and the Cramér-Rao Bound*, Phys. Rev. X **10**, 021056 (2020).

[20] D. Gupta and D. M. Busiello, *Tighter Thermodynamic Bound on the Speed Limit in Systems with Unidirectional Transitions*, Phys. Rev. E **102**, 062121 (2020).

[21] 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).

[22] T. Van Vu and Y. Hasegawa, *Geometrical Bounds of the Irreversibility in Markovian Systems*, Phys. Rev. Lett. **126**, 010601 (2021).

[23] K. Yoshimura and S. Ito, *Thermodynamic Uncertainty Relation and Thermodynamic Speed Limit in Deterministic Chemical Reaction Networks*, Phys. Rev. Lett. **127**, 160601 (2021).

[24] J.-C. Delvenne and G. Falasco, *Tight Classical Speed Limits*, arXiv:2110.13050.

[25] D. S. P. Salazar, *Lower Bound for Entropy Production Rate in Stochastic Systems far from Equilibrium*, Phys. Rev. E **106**, L032101 (2022).

[26] V. T. Vo, T. Van Vu, and Y. Hasegawa, *Unified Thermodynamic-Kinetic Uncertainty Relation*, J. Phys. A **55**, 405004 (2022).

[27] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, *Optimal Protocols and Optimal Transport in Stochastic Thermodynamics*, Phys. Rev. Lett. **106**, 250601 (2011).

[28] J. Goold, M. Paternostro, and K. Modi, *Nonequilibrium Quantum Landauer Principle*, Phys. Rev. Lett. **114**, 060602 (2015).

[29] K. Proesmans, J. Ehrich, and J. Bechhoefer, *Finite-Time Landauer Principle*, Phys. Rev. Lett. **125**, 100602 (2020).

[30] Y.-Z. Zhen, D. Egloff, K. Modi, and O. Dahlsten, *Universal Bound on Energy Cost of Bit Reset in Finite Time*, Phys. Rev. Lett. **127**, 190602 (2021).

[31] T. Van Vu and K. Saito, *Finite-Time Quantum Landauer Principle and Quantum Coherence*, Phys. Rev. Lett. **128**, 010602 (2022).

[32] J. S. Lee, S. Lee, H. Kwon, and H. Park, *Speed Limit for a Highly Irreversible Process and Tight Finite-Time Landauer's Bound*, Phys. Rev. Lett. **129**, 120603 (2022).

[33] J. Gore, F. Ritort, and C. Bustamante, *Bias and Error in Estimates of Equilibrium Free-Energy Differences from Nonequilibrium Measurements*, Proc. Natl. Acad. Sci. U.S.A. **100**, 12564 (2003).

[34] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, *Quantifying Dissipation Using Fluctuating Currents*, Nat. Commun. **10**, 1666 (2019).

[35] S. K. Manikandan, D. Gupta, and S. Krishnamurthy, *Inferring Entropy Production from Short Experiments*, Phys. Rev. Lett. **124**, 120603 (2020).

[36] T. Van Vu, V. T. Vo, and Y. Hasegawa, *Entropy Production Estimation with Optimal Current*, Phys. Rev. E **101**, 042138 (2020).

[37] 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).

[38] D.-K. Kim, Y. Bae, S. Lee, and H. Jeong, *Learning Entropy Production via Neural Networks*, Phys. Rev. Lett. **125**, 140604 (2020).

[39] D. J. Skinner and J. Dunkel, *Estimating Entropy Production from Waiting Time Distributions*, Phys. Rev. Lett. **127**, 198101 (2021).

[40] A. Dechant and S.-i. Sasa, *Improving Thermodynamic Bounds Using Correlations*, Phys. Rev. X **11**, 041061 (2021).

[41] T. Sagawa, *Thermodynamics of Information Processing in Small Systems*, Prog. Theor. Phys. **127**, 1 (2012).

[42] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, *Thermodynamics of Information*, Nat. Phys. **11**, 131 (2015).

[43] C. H. Bennett, *The Thermodynamics of Computation—A Review*, Int. J. Theor. Phys. **21**, 905 (1982).

[44] D. H. Wolpert, *The Stochastic Thermodynamics of Computation*, J. Phys. A **52**, 193001 (2019).

[45] E. Chitambar and G. Gour, *Quantum Resource Theories*, Rev. Mod. Phys. **91**, 025001 (2019).

[46] M. Lostaglio, *An Introductory Review of the Resource Theory Approach to Thermodynamics*, Rep. Prog. Phys. **82**, 114001 (2019).

[47] P. Salamon and R. S. Berry, *Thermodynamic Length and Dissipated Availability*, Phys. Rev. Lett. **51**, 1127 (1983).

[48] G. Ruppeiner, *Riemannian Geometry in Thermodynamic Fluctuation Theory*, Rev. Mod. Phys. **67**, 605 (1995).

[49] S.-i. Amari and H. Nagaoka, *Methods of Information Geometry* (Oxford University, New York, 2000), Vol. 191.

[50] G. E. Crooks, *Measuring Thermodynamic Length*, Phys. Rev. Lett. **99**, 100602 (2007).

[51] E. H. Feng and G. E. Crooks, *Length of Time's Arrow*, Phys. Rev. Lett. **101**, 090602 (2008).

[52] B. B. Machta, *Dissipation Bound for Thermodynamic Control*, Phys. Rev. Lett. **115**, 260603 (2015).

[53] G. M. Rotskoff, G. E. Crooks, and E. Vanden-Eijnden, *Geometric Approach to Optimal Nonequilibrium Control: Minimizing Dissipation in Nanomagnetic Spin Systems*, Phys. Rev. E **95**, 012148 (2017).

[54] S. B. Nicholson, A. del Campo, and J. R. Green, *Nonequilibrium Uncertainty Principle from Information Geometry*, Phys. Rev. E **98**, 032106 (2018).

[55] M. Scandi and M. Perarnau-Llobet, *Thermodynamic Length in Open Quantum Systems*, Quantum **3**, 197 (2019).

[56] S. J. Bryant and B. B. Machta, *Energy Dissipation Bounds for Autonomous Thermodynamic Cycles*, Proc. Natl. Acad. Sci. U.S.A. **117**, 3478 (2020).

[57] P. Abiuso, H. J. D. Miller, M. Perarnau-Llobet, and M. Scandi, *Geometric Optimisation of Quantum Thermodynamic Processes*, Entropy **22**, 1076 (2020).

[58] K. Brandner and K. Saito, *Thermodynamic Geometry of Microscopic Heat Engines*, Phys. Rev. Lett. **124**, 040602 (2020).

[59] G. T. Landi and M. Paternostro, *Irreversible Entropy Production: From Classical to Quantum*, Rev. Mod. Phys. **93**, 035008 (2021).

[60] T. Schmiedl and U. Seifert, *Optimal Finite-Time Processes in Stochastic Thermodynamics*, Phys. Rev. Lett. **98**, 108301 (2007).

[61] S. Chennakesavalu and G. M. Rotskof, *Unifying Thermodynamic Geometries*, arXiv:2205.01205.

[62] K. Saito and A. Dhar, *Waiting for Rare Entropic Fluctuations*, Europhys. Lett. **114**, 50004 (2016).

[63] N. Shiraishi, K. Saito, and H. Tasaki, *Universal Trade-Off Relation between Power and Efficiency for Heat Engines*, Phys. Rev. Lett. **117**, 190601 (2016).

[64] I. Neri, E. Roldán, and F. Jülicher, *Statistics of Infima and Stopping Times of Entropy Production and Applications to Active Molecular Processes*, Phys. Rev. X **7**, 011019 (2017).

[65] S. Pigolotti, I. Neri, E. Roldán, and F. Jülicher, *Generic Properties of Stochastic Entropy Production*, Phys. Rev. Lett. **119**, 140604 (2017).

[66] P. Pietzonka and U. Seifert, *Universal Trade-Off between Power, Efficiency, and Constancy in Steady-State Heat Engines*, Phys. Rev. Lett. **120**, 190602 (2018).

[67] G. Manzano, R. Fazio, and E. Roldán, *Quantum Martingale Theory and Entropy Production*, Phys. Rev. Lett. **122**, 220602 (2019).

[68] G. Falasco and M. Esposito, *Dissipation-Time Uncertainty Relation*, Phys. Rev. Lett. **125**, 120604 (2020).

[69] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, 2008).

[70] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, *Optimal Mass Transport: Signal Processing and Machine-Learning Applications*, IEEE Signal Process. Mag. **34**, 43 (2017).

[71] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, *Optimal Mass Transport for Registration and Warping*, Int. J. Comput. Vis. **60**, 225 (2004).

[72] G. Huang, C. Guo, M. J. Kusner, Y. Sun, F. Sha, and K. Q. Weinberger, *Supervised Word Mover's Distance*, in *Advances in Neural Information Processing Systems*, Vol. 29 (Curran Associates, Inc., Spain, 2016).

[73] P. Koehl, M. Delarue, and H. Orland, *Statistical Physics Approach to the Optimal Transport Problem*, Phys. Rev. Lett. **123**, 040603 (2019).

[74] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube *et al.*, *Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming*, Cell **176**, 928 (2019).

[75] R. Jordan, D. Kinderlehrer, and F. Otto, *The Variational Formulation of the Fokker-Planck Equation*, SIAM J. Math. Anal. **29**, 1 (1998).

[76] E. Aurell, K. Gawędzki, C. Mejía-Monasterio, R. Mohayaee, and P. Muratore-Ginanneschi, *Refined Second Law of Thermodynamics for Fast Random Processes*, J. Stat. Phys. **147**, 487 (2012).

[77] A. Dechant and Y. Sakurai, *Thermodynamic Interpretation of Wasserstein Distance*, arXiv:1912.08405.

[78] M. Nakazato and S. Ito, *Geometrical Aspects of Entropy Production in Stochastic Thermodynamics Based on Wasserstein Distance*, Phys. Rev. Res. **3**, 043093 (2021).

[79] J.-D. Benamou and Y. Brenier, *A Computational Fluid Mechanics Solution to the Monge-Kantorovich Mass Transfer Problem*, Numer. Math. **84**, 375 (2000).

[80] K. Proesmans, J. Ehrich, and J. Bechhoefer, *Optimal Finite-Time Bit Erasure under Full Control*, Phys. Rev. E **102**, 032105 (2020).

[81] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, *Experimental Verification of Landauer's Principle Linking Information and Thermodynamics*, Nature (London) **483**, 187 (2012).

[82] S. Schuler, T. Speck, C. Tietz, J. Wrachtrup, and U. Seifert, *Experimental Test of the Fluctuation Theorem for a Driven Two-Level System with Time-Dependent Rates*, Phys. Rev. Lett. **94**, 180602 (2005).

[83] F. W. J. Hekking and J. P. Pekola, *Quantum Jump Approach for Work and Dissipation in a Two-Level System*, Phys. Rev. Lett. **111**, 093602 (2013).

[84] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, *Experimental Realization of a Szilard Engine with a Single Electron*, Proc. Natl. Acad. Sci. U.S.A. **111**, 13786 (2014).

[85] M. Josefsson, A. Svilans, A. M. Burke, E. A. Hoffmann, S. Fahlvik, C. Thelander, M. Leijnse, and H. Linke, *A Quantum-Dot Heat Engine Operating close to the Thermodynamic Efficiency Limits*, Nat. Nanotechnol. **13**, 920 (2018).

[86] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic Gene Expression in a Single Cell*, Science **297**, 1183 (2002).

[87] M. Schliwa and G. Woehlke, *Molecular Motors*, Nature (London) **422**, 759 (2003).

[88] J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt, and M. Rief, *The Complex Folding Network of Single Calmodulin Molecules*, Science **334**, 512 (2011).

[89] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, *Information Transduction Capacity of Noisy Biochemical Signaling Networks*, Science **334**, 354 (2011).

[90] R. Landauer, *Irreversibility and Heat Generation in the Computing Process*, IBM J. Res. Dev. **5**, 183 (1961).

[91] T. Van Vu and Y. Hasegawa, *Lower Bound on Irreversibility in Thermal Relaxation of Open Quantum Systems*, Phys. Rev. Lett. **127**, 190601 (2021).

[92] B. Remlein and U. Seifert, *Optimality of Nonconservative Driving for Finite-Time Processes with Discrete States*, Phys. Rev. E **103**, L050105 (2021).

[93] A. Dechant, *Minimum Entropy Production, Detailed Balance and Wasserstein Distance for Continuous-Time Markov Processes*, J. Phys. A **55**, 094001 (2022).

[94] C. Maes, *Frenesy: Time-Symmetric Dynamical Activity in Nonequilibria*, Phys. Rep. **850**, 1 (2020).

[95] J. Maas, *Gradient Flows of the Entropy for Finite Markov Chains*, J. Funct. Anal. **261**, 2250 (2011).

[96] K. Yoshimura, A. Kolchinsky, A. Dechant, and S. Ito, *Geometrical Approach to Excess/Housekeeping Entropy Production in Discrete Systems*, arXiv:2205.15227.

[97] I. D. Terlizzi and M. Baiesi, *Kinetic Uncertainty Relation*, J. Phys. A **52**, 02LT03 (2019).

[98] L. Onsager, *Reciprocal Relations in Irreversible Processes. I*, Phys. Rev. **37**, 405 (1931).

[99] L. Onsager, *Reciprocal Relations in Irreversible Processes. II*, Phys. Rev. **38**, 2265 (1931).

[100] K. Proesmans and C. V. den Broeck, *Discrete-Time Thermodynamic Uncertainty Relation*, Europhys. Lett. **119**, 20001 (2017).

[101] K. Brandner, T. Hanazato, and K. Saito, *Thermodynamic Bounds on Precision in Ballistic Multiterminal Transport*, Phys. Rev. Lett. **120**, 090601 (2018).

[102] Y. Hasegawa and T. Van Vu, *Uncertainty Relations in Stochastic Processes: An Information Inequality Approach*, Phys. Rev. E **99**, 062126 (2019).

[103] T. Van Vu and Y. Hasegawa, *Uncertainty Relations for Underdamped Langevin Dynamics*, Phys. Rev. E **100**, 032130 (2019).

[104] Y. Hasegawa and T. Van Vu, *Fluctuation Theorem Uncertainty Relation*, Phys. Rev. Lett. **123**, 110602 (2019).

[105] A. M. Timpanaro, G. Guarnieri, J. Goold, and G. T. Landi, *Thermodynamic Uncertainty Relations from Exchange Fluctuation Theorems*, Phys. Rev. Lett. **123**, 090604 (2019).

[106] G. Guarnieri, G. T. Landi, S. R. Clark, and J. Goold, *Thermodynamics of Precision in Quantum Nonequilibrium Steady States*, Phys. Rev. Res. **1**, 033021 (2019).

[107] F. Carollo, R. L. Jack, and J. P. Garrahan, *Unraveling the Large Deviation Statistics of Markovian Open Quantum Systems*, Phys. Rev. Lett. **122**, 130605 (2019).

[108] A. Dechant and S.-i. Sasa, *Fluctuation-Response Inequality out of Equilibrium*, Proc. Natl. Acad. Sci. U.S.A. **117**, 6430 (2020).

[109] Y. Hasegawa, *Quantum Thermodynamic Uncertainty Relation for Continuous Measurement*, Phys. Rev. Lett. **125**, 050601 (2020).

[110] T. Van Vu and Y. Hasegawa, *Thermodynamic Uncertainty Relations under Arbitrary Control Protocols*, Phys. Rev. Res. **2**, 013060 (2020).

[111] K. Liu, Z. Gong, and M. Ueda, *Thermodynamic Uncertainty Relation for Arbitrary Initial States*, Phys. Rev. Lett. **125**, 140602 (2020).

[112] T. Koyuk and U. Seifert, *Thermodynamic Uncertainty Relation for Time-Dependent Driving*, Phys. Rev. Lett. **125**, 260604 (2020).

[113] D. H. Wolpert, *Uncertainty Relations and Fluctuation Theorems for Bayes Nets*, Phys. Rev. Lett. **125**, 200602 (2020).

[114] H. J. D. Miller, M. H. Mohammady, M. Perarnau-Llobet, and G. Guarnieri, *Thermodynamic Uncertainty Relation in Slowly Driven Quantum Heat Engines*, Phys. Rev. Lett. **126**, 210603 (2021).

[115] Y. Hasegawa, *Thermodynamic Uncertainty Relation for General Open Quantum Systems*, Phys. Rev. Lett. **126**, 010602 (2021).

[116] J. S. Lee, J.-M. Park, and H. Park, *Universal Form of Thermodynamic Uncertainty Relation for Langevin Dynamics*, Phys. Rev. E **104**, L052102 (2021).

[117] A. Pal, S. Reuveni, and S. Rahav, *Thermodynamic Uncertainty Relation for Systems with Unidirectional Transitions*, Phys. Rev. Res. **3**, 013273 (2021).

[118] T. Van Vu and K. Saito, *Thermodynamics of Precision in Markovian Open Quantum Dynamics*, Phys. Rev. Lett. **128**, 140602 (2022).

[119] A. C. Barato and U. Seifert, *Cost and Precision of Brownian Clocks*, Phys. Rev. X **6**, 041053 (2016).

[120] Y. Chen, T. T. Georgiou, L. Ning, and A. Tannenbaum, *Matricial Wasserstein-1 Distance*, IEEE Control Syst. Lett. **1**, 14 (2017).

[121] G. Lindblad, *On the Generators of Quantum Dynamical Semigroups*, Commun. Math. Phys. **48**, 119 (1976).

[122] J. M. Horowitz and J. M. R. Parrondo, *Entropy Production along Nonequilibrium Quantum Jump Trajectories*, New J. Phys. **15**, 085028 (2013).

[123] G. Manzano, J. M. Horowitz, and J. M. R. Parrondo, *Quantum Fluctuation Theorems for Arbitrary Environments: Adiabatic and Nonadiabatic Entropy Production*, Phys. Rev. X **8**, 031037 (2018).

[124] R. Alicki, *The Quantum Open System as a Model of the Heat Engine*, J. Phys. A **12**, L103 (1979).

[125] E. A. Carlen and J. Maas, *An Analog of the 2-Wasserstein Metric in Non-Commutative Probability under which the Fermionic Fokker-Planck Equation Is Gradient Flow for the Entropy*, Commun. Math. Phys. **331**, 887 (2014).

[126] Y. Chen, T. T. Georgiou, and A. Tannenbaum, *Matrix Optimal Mass Transport: A Quantum Mechanical Approach*, IEEE Trans. Autom. Control **63**, 2612 (2018).

[127] R. Duvenhage, *Quadratic Wasserstein Metrics for von Neumann Algebras via Transport Plans*, J. Operator Theory **88**, 289 (2022).

[128] G. D. Palma, M. Marvian, D. Trevisan, and S. Lloyd, *The Quantum Wasserstein Distance of Order 1*, IEEE Trans. Inf. Theory **67**, 6627 (2021).

[129] S. Friedland, M. Eckstein, S. Cole, and K. Życzkowski, *Quantum Monge-Kantorovich Problem and Transport Distance between Density Matrices*, Phys. Rev. Lett. **129**, 110402 (2022).

[130] J. Agredo and F. Fagnola, *On Quantum Versions of the Classical Wasserstein Distance*, Stochastics **89**, 910 (2017).

[131] N. Yu, L. Zhou, S. Ying, and M. Ying, *Quantum Earth Mover's Distance, No-Go Quantum Kantorovich-Rubinstein Theorem, and Quantum Marginal Problem*, arXiv:1803.02673.

[132] L. Mandelstam and I. Tamm, *The Uncertainty Relation between Energy and Time in Non-relativistic Quantum Mechanics*, J. Phys. (Moscow) **9**, 249 (1945).

[133] N. Margolus and L. B. Levitin, *The Maximum Speed of Dynamical Evolution*, Physica (Amsterdam) **120D**, 188 (1998).

[134] A. del Campo, I. L. Egusquiza, M. B. Plenio, and S. F. Huelga, *Quantum Speed Limits in Open System Dynamics*, Phys. Rev. Lett. **110**, 050403 (2013).

[135] M. M. Taddei, B. M. Escher, L. Davidovich, and R. L. de Matos Filho, *Quantum Speed Limit for Physical Processes*, Phys. Rev. Lett. **110**, 050402 (2013).

[136] S. Deffner and E. Lutz, *Quantum Speed Limit for Non-Markovian Dynamics*, Phys. Rev. Lett. **111**, 010402 (2013).

[137] D. P. Pires, M. Cianciaruso, L. C. Céleri, G. Adesso, and D. O. Soares-Pinto, *Generalized Geometric Quantum Speed Limits*, Phys. Rev. X **6**, 021031 (2016).

[138] M. Okuyama and M. Ohzeki, *Quantum Speed Limit Is Not Quantum*, Phys. Rev. Lett. **120**, 070402 (2018).

[139] F. Campaioli, F. A. Pollock, F. C. Binder, and K. Modi, *Tightening Quantum Speed Limits for Almost All States*, Phys. Rev. Lett. **120**, 060409 (2018).

[140] B. Shanahan, A. Chenu, N. Margolus, and A. del Campo, *Quantum Speed Limits across the Quantum-to-Classical Transition*, Phys. Rev. Lett. **120**, 070401 (2018).

[141] S. Sun, Y. Peng, X. Hu, and Y. Zheng, *Quantum Speed Limit Quantified by the Changing Rate of Phase*, Phys. Rev. Lett. **127**, 100404 (2021).

[142] E. O'Connor, G. Guarnieri, and S. Campbell, *Action Quantum Speed Limits*, Phys. Rev. A **103**, 022210 (2021).

[143] R. Hamazaki, *Speed Limits for Macroscopic Transitions*, PRX Quantum **3**, 020319 (2022).

[144] S. Nakajima and Y. Utsumi, *Speed Limits of the Trace Distance in the Interaction Picture for Open Quantum System*, New J. Phys. **24**, 095004 (2022).

[145] L. P. García-Pintos, S. B. Nicholson, J. R. Green, A. del Campo, and A. V. Gorshkov, *Unifying Quantum and Classical Speed Limits on Observables*, Phys. Rev. X **12**, 011038 (2022).

[146] Y. Hasegawa, *Thermodynamic Bounds via Bulk-Boundary Correspondence: Speed Limit, Thermodynamic Uncertainty Relation, and Heisenberg Principle*, arXiv:2203.12421.

[147] S. Deffner and S. Campbell, *Quantum Speed Limits: From Heisenberg's Uncertainty Principle to Optimal Quantum Control*, J. Phys. A **50**, 453001 (2017).

[148] K. Maruyama, F. Nori, and V. Vedral, *Colloquium: The Physics of Maxwell's Demon and Information*, Rev. Mod. Phys. **81**, 1 (2009).

[149] Y. Jun, M. c. v. Gavrilov, and J. Bechhoefer, *High-Precision Test of Landauer's Principle in a Feedback Trap*, Phys. Rev. Lett. **113**, 190601 (2014).

[150] L. L. Yan, T. P. Xiong, K. Rehan, F. Zhou, D. F. Liang, L. Chen, J. Q. Zhang, W. L. Yang, Z. H. Ma, and M. Feng, *Single-Atom Demonstration of the Quantum Landauer Principle*, Phys. Rev. Lett. **120**, 210601 (2018).

[151] J. Hong, B. Lambson, S. Dhuey, and J. Bokor, *Experimental Test of Landauer's Principle in Single-Bit Operations on Nanomagnetic Memory Bits*, Sci. Adv. **2**, e1501492 (2016).

[152] O.-P. Saira, M. H. Matheny, R. Katti, W. Fon, G. Wimsatt, J. P. Crutchfield, S. Han, and M. L. Roukes, *Nonequilibrium Thermodynamics of Erasure with Superconducting Flux Logic*, Phys. Rev. Res. **2**, 013249 (2020).

[153] S. Dago, J. Pereda, N. Barros, S. Ciliberto, and L. Bellon, *Information and Thermodynamics: Fast and Precise Approach to Landauer's Bound in an Underdamped Micromechanical Oscillator*, Phys. Rev. Lett. **126**, 170601 (2021).

[154] K. M. R. Audenaert, *A Sharp Continuity Estimate for the von Neumann Entropy*, J. Phys. A **40**, 8127 (2007).

[155] A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger, *Dynamics of the Dissipative Two-State System*, Rev. Mod. Phys. **59**, 1 (1987).

[156] R. Jozsa, *Fidelity for Mixed Quantum States*, J. Mod. Opt. **41**, 2315 (1994).

[157] T. Baumgratz, M. Cramer, and M. B. Plenio, *Quantifying Coherence*, Phys. Rev. Lett. **113**, 140401 (2014).

[158] J. M. Horowitz and M. Esposito, *Thermodynamics with Continuous Information Flow*, Phys. Rev. X **4**, 031015 (2014).

[159] T. Sagawa and M. Ueda, *Nonequilibrium Thermodynamics of Feedback Control*, Phys. Rev. E **85**, 021104 (2012).

[160] T. Van Vu and Y. Hasegawa, *Uncertainty Relation under Information Measurement and Feedback Control*, J. Phys. A **53**, 075001 (2020).

[161] P. P. Potts and P. Samuelsson, *Thermodynamic Uncertainty Relations Including Measurement and Feedback*, Phys. Rev. E **100**, 052137 (2019).

[162] T. Sagawa and M. Ueda, *Second Law of Thermodynamics with Discrete Quantum Feedback Control*, Phys. Rev. Lett. **100**, 080403 (2008).

[163] T. Hatano and S.-i. Sasa, *Steady-State Thermodynamics of Langevin Systems*, Phys. Rev. Lett. **86**, 3463 (2001).

[164] M. Esposito and C. Van den Broeck, *Three Detailed Fluctuation Theorems*, Phys. Rev. Lett. **104**, 090601 (2010).

[165] C. Maes and K. Netočný, *A Nonequilibrium Extension of the Clausius Heat Theorem*, J. Stat. Phys. **154**, 188 (2014).

[166] A. Dechant, S.-i. Sasa, and S. Ito, *Geometric Decomposition of Entropy Production in Out-of-Equilibrium Systems*, Phys. Rev. Res. **4**, L012034 (2022).

[167] R. Rao and M. Esposito, *Nonequilibrium Thermodynamics of Chemical Reaction Networks: Wisdom from Stochastic Thermodynamics*, Phys. Rev. X **6**, 041064 (2016).

[168] A. Kolchinsky and D. H. Wolpert, *Work, Entropy Production, and Thermodynamics of Information under Protocol Constraints*, Phys. Rev. X **11**, 041024 (2021).

[169] P. Abiuso, V. Holubec, J. Anders, Z. Ye, F. Cerisola, and M. Perarnau-Llobet, *Thermodynamics and Optimal Protocols of Multidimensional Quadratic Brownian Systems*, J. Phys. Commun. **6**, 063001 (2022).

[170] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows: In Metric Spaces and in the Space of Probability Measures* (Springer, New York, 2008).

[171] M. Gelbrich, *On a Formula for the L^2-Wasserstein Metric between Measures on Euclidean and Hilbert Spaces*, Mathematische Nachrichten **147**, 185 (1990).

[172] N. Shiraishi, *Optimal Thermodynamic Uncertainty Relation in Markov Jump Processes*, J. Stat. Phys. **185**, 19 (2021).

[173] B. Rutten, M. Esposito, and B. Cleuren, *Reaching Optimal Efficiencies Using Nanosized Photoelectric Devices*, Phys. Rev. B **80**, 235122 (2009).

[174] R. Bhatia, *Matrix Analysis* (Springer, New York, 1996).