Optimizing optimal transport: Role of final distributions in finite-time thermodynamics

Kaito Tojo [1,*,†] Rihito Nagase [1,*] Ken Funo [1] and Takahiro Sagawa [1,2,3]

[1] *Department of Applied Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[2] *Quantum-Phase Electronics Center (QPEC), The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[3] *Inamori Research Institute for Science (InaRIS), Kyoto-shi, Kyoto 600-8411, Japan*

(Received 26 September 2025; accepted 14 October 2026; published 16 November 2026)

Performing thermodynamic tasks within finite time while minimizing thermodynamic costs is a central challenge in stochastic thermodynamics. Here, we develop a unified framework for optimizing the thermodynamic cost of performing various tasks in finite time for overdamped Langevin systems. Conventional optimization of thermodynamic cost based on optimal transport theory leaves room for varying the final distributions according to the intended task, enabling further optimization. Taking advantage of this freedom, we use Lagrange multipliers to derive the optimal final distribution that minimizes the thermodynamic cost. Our framework applies to a wide range of thermodynamic tasks, including particle transport, thermal squeezing, and information processing such as information erasure, measurement, and feedback. Our results are expected to provide design principles for information-processing devices and thermodynamic machines that operate at high speed with low energetic costs.

DOI: 10.1103/9pdj-n1kj

I. INTRODUCTION

Optimal transport theory [1–3] is a powerful mathematical methodology that has developed from the optimization problem of how to transfer one probability distribution into another with the smallest possible transport cost. Solving such a transport problem yields an optimal transport plan and a distance between probability distributions known as the Wasserstein distance. In recent years, optimal transport theory has been applied across diverse fields, including deep learning [4], image processing [5], and molecular biology [6].

In the context of thermodynamics, the connections between optimal transport theory and stochastic thermodynamics [7–11] have been investigated. The lower bound on the entropy production set by the second law of thermodynamics is achievable only in the quasistatic limit, which makes it unattainable in finite time [12]. Therefore, tighter bounds than the second law have been pursued theoretically [12–26] and experimentally [27–34]. Specifically, optimal transport theory yields the fundamental achievable bound and the explicit optimal protocol in finite time. Such optimal speed limits have been formulated for overdamped Langevin systems [35–42], as well as for discrete systems [43–46], Markovian open quantum systems [43,46], and nonlinear systems [45,47]. More recently, optimal transport theory has been applied to information thermodynamics [48–51], revealing fundamental bounds on the thermodynamic cost of performing information erasure [39,40], measurement, and feedback [38,52–57] in finite time.

However, it is worth emphasizing that the straightforward application of optimal transport theory to thermodynamic tasks leaves room for further minimization of the thermodynamic cost. For example, when we consider the task of thermal squeezing, which aims to reduce the variance along a certain direction in the phase space, there are infinitely many distributions that share the same variance. Therefore, there remains room for further optimization of final distributions under the condition that the desired task is achieved [see Fig. 1(a)].

In this paper, we develop a unified framework for finite-time optimization of the thermodynamic cost for various tasks in overdamped Langevin systems. We take into account the freedom of the final distributions when performing thermodynamic tasks and identify the optimal final distribution. Our approach encodes thermodynamic tasks as constraints on the final distribution within the method of Lagrange multipliers, thereby deriving variational equations that identify the optimal final distribution [see Fig. 1(b)]. We can handle any kind of thermodynamic task as long as it can be expressed as equality constraints on functionals of the final distribution. This includes the expectation values and variances of physical quantities of the final distributions. We can also treat mutual information in a similar manner, which plays a crucial role in information processing.

Our framework applies broadly to thermodynamic tasks (see Fig. 1 and Table I), enabling us to obtain the truly optimal value of the thermodynamic cost and the protocol that achieves it. We demonstrate our framework by optimizing the thermodynamic cost of particle transport, thermal squeezing,

*These authors contributed equally to this work.

†Contact author: tojo@noneq.t.u-tokyo.ac.jp

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

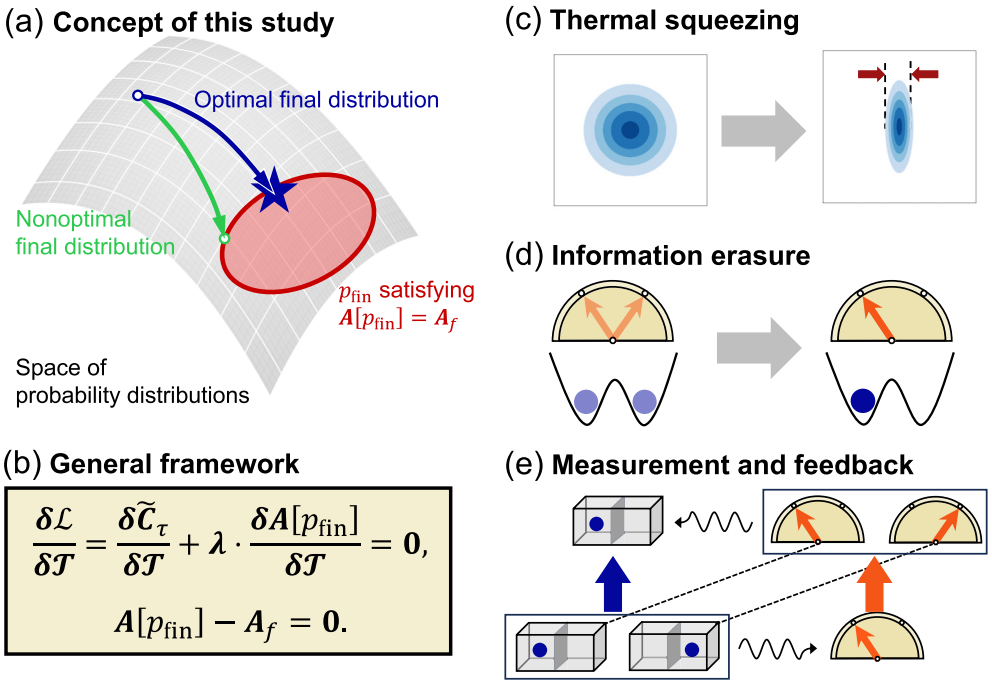


FIG. 1. Summary of the main results and their applications. (a) Schematic illustration indicating that further optimization is possible beyond conventional optimal transport. Even if optimal transport protocols are used (blue and green arrows), choosing a nonoptimal final distribution (green dot) yields a higher thermodynamic cost. Unlike the conventional setup that fixes the final distribution, we consider the freedom to choose a final distribution from the admissible set determined by the intended thermodynamic task. The boundary of the red region represents the set of admissible final distributions associated with each thermodynamic task, from which the optimal final distribution (blue star) is identified. (b) General framework that further optimizes the thermodynamic cost (Sec. III). We encode thermodynamic tasks as constraints on the final distributions, such that a quantity depending on the final distributions $\boldsymbol{A}[p_{\rm fin}]$ matches a desired value $\boldsymbol{A}_f$. The Lagrange multiplier method is employed to incorporate these constraints into the optimization problem. Solving the resulting variational equations yields the optimal transport map $\boldsymbol{\mathcal{T}}^{\rm opt}$ that transports the initial probability distribution to the optimal final distribution. Applications of our framework (Sec. IV): (c) thermal squeezing, (d) information erasure, and (e) measurement and feedback.

information erasure, work in quadratic potentials, control of free energy, and measurement and feedback. The framework also encompasses earlier studies, such as optimization of thermodynamic costs for information erasure [39,40] and optimization of work in systems with quadratic potentials [41]. The present framework extends the optimization of thermodynamic costs in systems far from equilibrium, beyond the slow-driving or linear response regime [17,18].

The structure of this paper is as follows. In Sec. II, we present notation, along with the thermodynamic speed limit based on optimal transport theory for overdamped Langevin systems. In Sec. III, we develop the general framework for optimizing the thermodynamic cost required to perform thermodynamic tasks in finite time. In Sec. IV, we apply our framework to the tasks summarized in Table I, showing that optimization of the thermodynamic cost required for such tasks can be treated. In Sec. V, we conclude the paper with a summary and outlook. In the Appendixes, we provide details of the derivations of our results.

II. NOTATION

We consider an overdamped Langevin system in d-dimensional Euclidean space in contact with a heat bath at temperature $T(=1/\beta)$, with mobility μ and diffusion coefficient $D = \mu T$. Throughout this paper, we set the Boltzmann

TABLE I. Applications of our framework (see Sec. IV). We show the thermodynamic cost C_τ and the constraints on the final distributions $\boldsymbol{A}[p_{\rm fin}]$ to perform given thermodynamic tasks.

Thermodynamic tasks	Thermodynamic cost C_τ	Constraints at $t = \tau$ associated with the task: $\boldsymbol{A}[p_{\rm fin}]$
Particle transport	Entropy production Σ_τ	Expectation value of position: $\langle \boldsymbol{r} \rangle_{\rm fin}$
Thermal squeezing	Σ_τ	Covariance matrix of position: $\langle (\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm fin})(\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm fin})^\top \rangle_{\rm fin}$
Information erasure [39,40]	Σ_τ, work W_τ	Cumulative distribution function at $r = 0$: $\int_{-\infty}^{0} {\rm d}r\, p_{\rm fin}(r)$
Work in systems with quadratic potentials [41]	W_τ	None
Control of free energy	Σ_τ	KL divergence: $D(p_{\rm fin}\|\|q)$
Measurement and feedback	Partial entropy production Σ_τ^X, Σ_τ^Y	Mutual information: $I_{\rm fin}$

constant to $k_B = 1$. The system is driven by a conservative force $-\nabla V_t(\boldsymbol{r})$ with $V_t(\boldsymbol{r})$ being a time-dependent potential, and a nonconservative force $\boldsymbol{F}_t^{\rm nc}(\boldsymbol{r})$. The time evolution of the probability distribution $p_t(\boldsymbol{r})$ is described by the Fokker-Planck equation [58,59]

$$\partial_t p_t(\boldsymbol{r}) = -\nabla \cdot (\boldsymbol{v}_t(\boldsymbol{r}) p_t(\boldsymbol{r})), \tag{1}$$

where $\boldsymbol{v}_t(\boldsymbol{r})$ denotes the mean local velocity

$$\boldsymbol{v}_t(\boldsymbol{r}) := \mu\big(-\nabla V_t(\boldsymbol{r}) + \boldsymbol{F}_t^{\rm nc}(\boldsymbol{r})\big) - D\nabla \ln p_t(\boldsymbol{r}). \tag{2}$$

We assume throughout the paper that the conservative force $-\nabla V_t(\boldsymbol{r})$ and the nonconservative force $\boldsymbol{F}_t^{\rm nc}(\boldsymbol{r})$ are fully controllable.

The entropy production Σ_τ is given by [7,8]

$$\Sigma_\tau := \frac{1}{D}\int_0^\tau dt \int d\boldsymbol{r}\, \|\boldsymbol{v}_t(\boldsymbol{r})\|^2 p_t(\boldsymbol{r}) \geqslant 0, \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm. The equality $\Sigma_\tau = 0$ holds if $\boldsymbol{v}_t(\boldsymbol{r}) = \boldsymbol{0}$ for all t, which is possible only in the quasistatic limit $\tau \to \infty$. The work done on the system W_τ is given by [7,8]

$$W_\tau := \int_0^\tau dt \int d\boldsymbol{r} \left(\frac{\partial V_t(\boldsymbol{r})}{\partial t} + \boldsymbol{F}_t^{\rm nc}(\boldsymbol{r}) \cdot \boldsymbol{v}_t(\boldsymbol{r})\right) p_t(\boldsymbol{r}). \tag{4}$$

By using the average energy of the system $E_t := \int d\boldsymbol{r}\, V_t(\boldsymbol{r}) p_t(\boldsymbol{r})$ and the Shannon entropy $S_t := -\int d\boldsymbol{r}\, p_t(\boldsymbol{r}) \ln p_t(\boldsymbol{r})$, we define the nonequilibrium free energy as $\mathcal{F}_t := E_t - TS_t$ [7,8]. We also define the equilibrium free energy $\mathcal{F}_t^{\rm eq} := -\beta^{-1} \ln \int d\boldsymbol{r}\, e^{-\beta V_t(\boldsymbol{r})}$. The nonequilibrium free energy is expressed as $\mathcal{F}_t = \beta^{-1} D(p_t || p_t^{\rm eq}) + \mathcal{F}_t^{\rm eq}$, where $p_t^{\rm eq}(\boldsymbol{r}) := e^{\beta(\mathcal{F}_t^{\rm eq} - V_t(\boldsymbol{r}))}$ and $D(p||q) := \int d\boldsymbol{r}\, p(\boldsymbol{r}) \ln(p(\boldsymbol{r})/q(\boldsymbol{r}))$ is the Kullback-Leibler (KL) divergence. We note that the work can be expressed as $W_\tau = \Delta\mathcal{F} + T\Sigma_\tau$, where $\Delta\mathcal{F} := \mathcal{F}_\tau - \mathcal{F}_0$ is the change in the nonequilibrium free energy.

We fix the initial distribution $p_{\rm ini}$, the final distribution $p_{\rm fin}$, and the operation time τ. The Benamou-Brenier formula [60] implies that the minimum entropy production is given by the L^2-Wasserstein distance from optimal transport theory [36,38]:

$$\min_{\substack{\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau} \\ \text{s.t. } p_0 = p_{\rm ini},\, p_\tau = p_{\rm fin}}} \Sigma_\tau = \frac{\mathcal{W}_2(p_{\rm ini}, p_{\rm fin})^2}{D\tau}. \tag{5}$$

Here, the minimum is taken over all protocols $\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau}$ that satisfy Eq. (1) with boundary conditions $p_{t=0} = p_{\rm ini}$ and $p_{t=\tau} = p_{\rm fin}$. The squared L^2-Wasserstein distance between $p_{\rm ini}$ and $p_{\rm fin}$ is defined as [1,2]

$$\mathcal{W}_2(p_{\rm ini}, p_{\rm fin})^2 := \min_{\mathcal{T}\ \text{s.t.}\ \mathcal{T}_\sharp p_{\rm ini} = p_{\rm fin}} \int d\boldsymbol{r}\, \|\mathcal{T}(\boldsymbol{r}) - \boldsymbol{r}\|^2 p_{\rm ini}(\boldsymbol{r}), \tag{6}$$

where, following the convention in mathematics, $\mathcal{T}_\sharp p_{\rm ini} = p_{\rm fin}$ means that the minimum in Eq. (6) is taken over transport maps $\mathcal{T}$ pushing $p_{\rm ini}$ forward to $p_{\rm fin}$. That is, $\mathcal{T}_\sharp p_{\rm ini} = p_{\rm fin}$ implies that $\int d\boldsymbol{r} f(\mathcal{T}(\boldsymbol{r})) p_{\rm ini}(\boldsymbol{r}) = \int d\boldsymbol{r} f(\boldsymbol{r}) p_{\rm fin}(\boldsymbol{r})$ for any continuous function f, which, when $\mathcal{T}$ is sufficiently smooth and has a nonsingular Jacobian $J_{\mathcal{T}}(\boldsymbol{r}) := \frac{\partial \mathcal{T}(\boldsymbol{r})}{\partial \boldsymbol{r}}$, can be written as

$$p_{\rm ini}(\boldsymbol{r}) = p_{\rm fin}(\mathcal{T}(\boldsymbol{r})) \det J_{\mathcal{T}}(\boldsymbol{r}). \tag{7}$$

III. GENERAL FRAMEWORK

In this section, we present our general framework. The problem we address is as follows: Given an initial distribution $p_{\rm ini}$, we optimize the thermodynamic cost C_τ required to perform various thermodynamic tasks while the system evolves according to Eq. (1) over the finite fixed time τ. In what follows, depending on the thermodynamic task, we take as the thermodynamic cost the entropy production Σ_τ, the work W_τ, or the partial entropy production Σ_τ^i $(i = X, Y)$ in bipartite systems, as listed in Table I.

A. Setup of optimization

First, we define a quantity $\boldsymbol{A}[p_{\rm fin}]$ dependent on the final distribution $p_{\rm fin}$, which includes various examples as listed in Table I. Then, we encode thermodynamic tasks as constraints on $\boldsymbol{A}[p_{\rm fin}]$ such that it matches a desired value $\boldsymbol{A}_f$. The crucial point is that there are (in general, infinitely) many distributions that satisfy this condition. Therefore, in addition to optimizing the protocol that achieves the minimum in Eq. (5), the final distribution $p_{\rm fin}$ can be further optimized.

We consider the thermodynamic cost C_τ required for achieving the above task. In general, we suppose that C_τ depends on the time evolution of the distribution and the mean local velocity, which we explicitly write as $C_\tau[\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau}]$. Then, the optimization problem can be formulated as

$$\min_{p_{\rm fin}\ \text{s.t.}\ \boldsymbol{A}[p_{\rm fin}] = \boldsymbol{A}_f} \min_{\substack{\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau} \\ \text{s.t. } p_0 = p_{\rm ini},\, p_\tau = p_{\rm fin}}} C_\tau[\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau}]. \tag{8}$$

We now show that the optimization problem Eq. (8) can be reduced to a more tractable expression. We first consider the special case where $C_\tau = \Sigma_\tau$. For a transport map $\mathcal{T}$, we define

$$\tilde{\Sigma}_\tau[\mathcal{T}] := \frac{1}{D\tau}\int d\boldsymbol{r}\, \|\mathcal{T}(\boldsymbol{r}) - \boldsymbol{r}\|^2 p_{\rm ini}(\boldsymbol{r}). \tag{9}$$

From Eqs. (5) and (6), we have

$$\min_{\substack{\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau} \\ \text{s.t. } p_0 = p_{\rm ini},\, p_\tau = p_{\rm fin}}} \Sigma_\tau[\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau}] = \min_{\mathcal{T}\ \text{s.t.}\ \mathcal{T}_\sharp p_{\rm ini} = p_{\rm fin}} \tilde{\Sigma}_\tau[\mathcal{T}]. \tag{10}$$

In addition, since $p_{\rm fin}$ is a functional of $\mathcal{T}$ via Eq. (7), $\boldsymbol{A}[p_{\rm fin}]$ can be written as $\boldsymbol{A}[\mathcal{T}]$. Therefore, our optimization problem Eq. (8) reduces to

$$\min_{p_{\rm fin}\ \text{s.t.}\ \boldsymbol{A}[p_{\rm fin}] = \boldsymbol{A}_f} \min_{\mathcal{T}\ \text{s.t.}\ \mathcal{T}_\sharp p_{\rm ini} = p_{\rm fin}} \tilde{\Sigma}_\tau[\mathcal{T}] = \min_{\mathcal{T}\ \text{s.t.}\ \boldsymbol{A}[\mathcal{T}] = \boldsymbol{A}_f} \tilde{\Sigma}_\tau[\mathcal{T}]. \tag{11}$$

A similar argument applies to other C_τ listed in Table I, by defining $\tilde{C}_\tau[\mathcal{T}]$ appropriately (see Secs. III C 2 and IV F for details).

In summary, we consider the optimization problem equivalent to Eq. (8), which takes the simple form

$$\min_{\mathcal{T}\ \text{s.t.}\ \boldsymbol{A}[\mathcal{T}] = \boldsymbol{A}_f} \tilde{C}_\tau[\mathcal{T}]. \tag{12}$$

B. Main formula

We adopt a Lagrange multiplier approach that takes the transport map as the optimization variable. The Lagrangian

$\mathcal{L}$ contains the transport map $\mathcal{T}$ and the Lagrange multipliers $\boldsymbol{\lambda}$ as variables, and is given by

$$\mathcal{L}[\mathcal{T}, \boldsymbol{\lambda}] = \tilde{C}_\tau[\mathcal{T}] + \boldsymbol{\lambda} \cdot (\boldsymbol{A}[\mathcal{T}] - \boldsymbol{A}_f). \tag{13}$$

The first-order optimality conditions read

$$\frac{\delta \tilde{C}_\tau[\mathcal{T}]}{\delta \mathcal{T}} + \boldsymbol{\lambda} \cdot \frac{\delta \boldsymbol{A}[\mathcal{T}]}{\delta \mathcal{T}} = 0, \tag{14}$$

$$\boldsymbol{A}[\mathcal{T}] - \boldsymbol{A}_f = 0. \tag{15}$$

Solving Eqs. (14) and (15) yields solutions $(\mathcal{T}^{\mathrm{opt}}, \boldsymbol{\lambda}^{\mathrm{opt}})$, from which the optimal value of the thermodynamic cost $C_\tau^{\mathrm{opt}} = \tilde{C}_\tau[\mathcal{T}^{\mathrm{opt}}]$, the optimal final distribution $p_{\mathrm{fin}}^{\mathrm{opt}}$, and the optimal protocol $\{p_t^{\mathrm{opt}}, \boldsymbol{v}_t^{\mathrm{opt}}\}_{0 \leqslant t \leqslant \tau}$ follow. The explicit form of the optimal mean local velocity is given by [36,60]

$$\boldsymbol{v}_t^{\mathrm{opt}}\left(\left(1 - \frac{t}{\tau}\right)\boldsymbol{r} + \frac{t}{\tau}\mathcal{T}^{\mathrm{opt}}(\boldsymbol{r})\right) = \frac{\mathcal{T}^{\mathrm{opt}}(\boldsymbol{r}) - \boldsymbol{r}}{\tau}. \tag{16}$$

These constitute the general framework shown in Fig. 1(b). We note that a similar approach has also been adopted previously in the setting of information erasure [39,40].

Throughout, we use the Einstein summation convention. We note that the details of the derivations are provided in Appendix B.

C. Variation of thermodynamic costs C_τ

As mentioned before, the thermodynamic cost C_τ we consider is the entropy production Σ_τ, the work W_τ, or the partial entropy production Σ_τ^i $(i = X, Y)$. In this subsection, we present the variational formulas for Σ_τ and W_τ, while Σ_τ^i $(i = X, Y)$ is discussed later in Sec. IV F. What we mainly treat below is the first term of Eq. (14).

1. C_τ: entropy production

We first consider the entropy production. From Eq. (9), the variation of $\tilde{\Sigma}_\tau[\mathcal{T}]$ with respect to the transport map $\mathcal{T}$ is given by

$$\frac{\delta \tilde{\Sigma}_\tau[\mathcal{T}]}{\delta \mathcal{T}} = \frac{2}{D\tau}(\mathcal{T}(\boldsymbol{r}) - \boldsymbol{r}) p_{\mathrm{ini}}(\boldsymbol{r}), \tag{17}$$

which is linear in $\mathcal{T}$.

2. C_τ: work

We next consider the work. When optimizing the work in finite time, we impose the boundary condition that the potentials V_t at $t = 0$ and $t = \tau$ are fixed to V_{ini} and V_{fin}, respectively [16]. Otherwise, even in the quasistatic limit, $W = \mathcal{F}_{\mathrm{fin}}^{\mathrm{eq}} - \mathcal{F}_{\mathrm{ini}}$ could be made arbitrarily small by varying V_0 or V_τ. Here, we fix V_0 and V_τ to eliminate this triviality.

Using $\tilde{\Sigma}_\tau[\mathcal{T}]$ in Eq. (9), we define $\tilde{W}_\tau[\mathcal{T}] := \Delta\mathcal{F}[\mathcal{T}] + \beta^{-1}\tilde{\Sigma}_\tau[\mathcal{T}]$. We note that since $\Delta\mathcal{F}[\mathcal{T}]$ depends only on p_{ini}, p_{fin}, V_{ini}, and V_{fin}, the optimization problem (8) with $C_\tau = W_\tau$ is equivalent to Eq. (12) with $\tilde{C}_\tau[\mathcal{T}] = \tilde{W}_\tau[\mathcal{T}]$. The variation of $\tilde{\Sigma}_\tau[\mathcal{T}]$ was obtained in Eq. (17), so it remains to vary $\mathcal{F}_{\mathrm{fin}}[\mathcal{T}]$, which is dependent on $\mathcal{T}$ via p_{fin}. Defining $p_{\mathrm{fin}}^{\mathrm{eq}}(\boldsymbol{r}) := e^{-\beta V_{\mathrm{fin}}(\boldsymbol{r})} / \int \mathrm{d}\boldsymbol{r}' e^{-\beta V_{\mathrm{fin}}(\boldsymbol{r}')}$, the dependence of $\mathcal{F}_{\mathrm{fin}}[\mathcal{T}]$ on the transport map is given by

$$\mathcal{F}_{\mathrm{fin}}[\mathcal{T}] = \beta^{-1} \int \mathrm{d}\boldsymbol{r}\, p_{\mathrm{ini}}(\boldsymbol{r}) \ln \frac{p_{\mathrm{ini}}(\boldsymbol{r})}{p_{\mathrm{fin}}^{\mathrm{eq}}(\mathcal{T}(\boldsymbol{r})) \det J_{\mathcal{T}}(\boldsymbol{r})} + \mathcal{F}_{\mathrm{fin}}^{\mathrm{eq}}. \tag{18}$$

Thus, we have

$$\frac{\delta \mathcal{F}_{\mathrm{fin}}[\mathcal{T}]}{\delta \mathcal{T}_i} = \beta^{-1}\Big[-p_{\mathrm{ini}}(\boldsymbol{r})\, \partial_{r_i'} \ln p_{\mathrm{fin}}^{\mathrm{eq}}(\boldsymbol{r}')\big|_{\boldsymbol{r}' = \mathcal{T}(\boldsymbol{r})} + \partial_{r_j}\{p_{\mathrm{ini}}(\boldsymbol{r}) \tilde{J}_{ji}(\boldsymbol{r})\}\Big]. \tag{19}$$

See Appendix B 1 for the details of the derivation. Here, $\tilde{J}_{ji}(\boldsymbol{r}) := (J_{\mathcal{T}}(\boldsymbol{r})^{-1})_{ji} = (J_{\mathcal{T}^{-1}}(\mathcal{T}(\boldsymbol{r})))_{ji} = \partial_{r_i'} \mathcal{T}_j^{-1}(\boldsymbol{r}')|_{\boldsymbol{r}' = \mathcal{T}(\boldsymbol{r})}$. Therefore, the variation of $\tilde{W}_\tau[\mathcal{T}]$ is given by

$$\frac{\delta \tilde{W}_\tau[\mathcal{T}]}{\delta \mathcal{T}_i} = \beta^{-1}\Big[-p_{\mathrm{ini}}(\boldsymbol{r})\, \partial_{r_i'} \ln p_\tau^{\mathrm{eq}}(\boldsymbol{r}')\big|_{\boldsymbol{r}' = \mathcal{T}(\boldsymbol{r})} + \partial_{r_j}\{p_{\mathrm{ini}}(\boldsymbol{r}) \tilde{J}_{ji}(\boldsymbol{r})\} + \frac{2}{D\tau}(\mathcal{T}_i(\boldsymbol{r}) - r_i) p_{\mathrm{ini}}(\boldsymbol{r})\Big]. \tag{20}$$

Although the potentials at $t = 0$ and $t = \tau$ are fixed, the potential shape may change discontinuously at those times. Indeed, it is known that quenches at $t = 0$ and $t = \tau$ are required in finite-time thermodynamic optimization [16].

D. Variation of $\boldsymbol{A}[p_{\mathrm{fin}}]$

We next derive the variational formula for $\boldsymbol{A}[p_{\mathrm{fin}}] = \boldsymbol{A}[\mathcal{T}]$ with respect to the transport map. Combining it with the variation of C_τ from the previous subsection yields the optimal final distributions $p_{\mathrm{fin}}^{\mathrm{opt}}$ via Eqs. (14) and (15).

In general, a quantity $\boldsymbol{A}[\mathcal{T}]$ can be written as a functional of the final distribution $p_{\mathrm{fin}}[\mathcal{T}]$ of the form $\int \mathrm{d}\boldsymbol{r}\, \boldsymbol{g}(p_{\mathrm{fin}}[\mathcal{T}(\boldsymbol{r})], \boldsymbol{r}) = \int \mathrm{d}\boldsymbol{r} \det J_{\mathcal{T}}(\boldsymbol{r}) \boldsymbol{g}(\frac{p_{\mathrm{ini}}(\boldsymbol{r})}{\det J_{\mathcal{T}}(\boldsymbol{r})}, \mathcal{T}(\boldsymbol{r}))$. For example, if the thermodynamic task is to make the expectation value of a physical observable $\boldsymbol{a}(\boldsymbol{r})$ at the final time $\langle \boldsymbol{a} \rangle_{\mathrm{fin}}$ equal to a desired value, then $\boldsymbol{g}(p_{\mathrm{fin}}[\mathcal{T}(\boldsymbol{r})], \boldsymbol{r}) = \boldsymbol{a}(\boldsymbol{r})\, p_{\mathrm{fin}}[\mathcal{T}(\boldsymbol{r})]$. Also, the Shannon entropy is described by taking $g(p_{\mathrm{fin}}[\mathcal{T}(\boldsymbol{r})], \boldsymbol{r}) = -p_{\mathrm{fin}}[\mathcal{T}(\boldsymbol{r})] \ln p_{\mathrm{fin}}[\mathcal{T}(\boldsymbol{r})]$. For measurement and feedback tasks, we fix the mutual information of the final distributions. Although mutual information does not take the above form because marginalization of the final probability distribution is involved, it can still be treated in a similar manner, as we discuss in Sec. IV.

For $\boldsymbol{A}[\mathcal{T}] = \int \mathrm{d}\boldsymbol{r}\, \boldsymbol{g}(p_{\mathrm{fin}}[\mathcal{T}(\boldsymbol{r})], \boldsymbol{r})$, the variation with respect to $\mathcal{T}_i$ is given by

$$\frac{\delta \boldsymbol{A}[\mathcal{T}]}{\delta \mathcal{T}_i} = \int \mathrm{d}\boldsymbol{r}' \left.\frac{\partial \boldsymbol{g}(X, \boldsymbol{r}')}{\partial X}\right|_{X = p_{\mathrm{fin}}(\boldsymbol{r}')} \frac{\delta p_{\mathrm{fin}}(\boldsymbol{r}')}{\delta \mathcal{T}_i(\boldsymbol{r})}. \tag{21}$$

By further using the expression for $\frac{\delta p_{\mathrm{fin}}[\mathcal{T}]}{\delta \mathcal{T}_i}$ obtained in Appendix B 2, we have

$$\frac{\delta \boldsymbol{A}[\mathcal{T}]}{\delta \mathcal{T}_i} = \partial_{r_j}\{\hat{\boldsymbol{g}}(\mathcal{T}(\boldsymbol{r}))\} \tilde{J}_{ji}(\boldsymbol{r}) p_{\mathrm{ini}}(\boldsymbol{r}), \tag{22}$$

where $\hat{\boldsymbol{g}}(\mathcal{T}(\boldsymbol{r})) = \frac{\partial \boldsymbol{g}(X, \mathcal{T}(\boldsymbol{r}))}{\partial X}|_{X = p_{\mathrm{fin}}(\mathcal{T}(\boldsymbol{r}))}$. Equation (22) is the general form for the variation of a quantity $\boldsymbol{A}[\mathcal{T}]$ associated with the thermodynamic task. In Sec. IV, we treat variations of specific quantities depending on the intended task.

An important choice of the quantity $\boldsymbol{A}[\mathcal{T}]$ is the expectation value of a physical observable at the final time $\boldsymbol{A}[\mathcal{T}] = \langle \boldsymbol{a} \rangle_{\rm fin}[\mathcal{T}] = \int \mathrm{d}\boldsymbol{r}\, \boldsymbol{a}(\boldsymbol{r}) p_{\rm fin}[\mathcal{T}(\boldsymbol{r})]$. In this case, we obtain simpler variational formulas. Since Eq. (7) implies

$$\langle \boldsymbol{a} \rangle_{\rm fin}[\mathcal{T}] = \int \mathrm{d}\boldsymbol{r}\, \boldsymbol{a}(\mathcal{T}(\boldsymbol{r})) p_{\rm ini}(\boldsymbol{r}), \tag{23}$$

we obtain by direct calculation

$$\frac{\delta \langle \boldsymbol{a} \rangle_{\rm fin}[\mathcal{T}]}{\delta \mathcal{T}} = \partial_{\boldsymbol{r}'} \boldsymbol{a}(\boldsymbol{r}')|_{\boldsymbol{r}'=\mathcal{T}(\boldsymbol{r})} p_{\rm ini}(\boldsymbol{r}). \tag{24}$$

We note that Eq. (24) is consistent with Eq. (22) when $\boldsymbol{g}(p_{\rm fin}[\mathcal{T}(\boldsymbol{r})], \boldsymbol{r}) = \boldsymbol{a}(\boldsymbol{r}) p_{\rm fin}[\mathcal{T}(\boldsymbol{r})]$ (see Appendix B 2 for details).

While our framework provides a necessary condition for global optimality in general, it guarantees global optimality if a feasible stationary solution $(\mathcal{T}^*, \boldsymbol{\lambda}^*)$ exists and the Lagrangian density $\mathscr{L}(\boldsymbol{r}, \mathcal{T}, \frac{\partial \mathcal{T}}{\partial \boldsymbol{r}}, \boldsymbol{\lambda})$, defined via $\mathcal{L} = \int \mathrm{d}\boldsymbol{r}\, \mathscr{L}$, is jointly convex in $(\mathcal{T}, \frac{\partial \mathcal{T}}{\partial \boldsymbol{r}})$ for all $\boldsymbol{r}$ with $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ [61]. One can verify that the density for the entropy production is convex in $\mathcal{T}$, as is the Lagrangian density for the partial entropy production introduced in Sec. IV F. However, the convexity of the density for the work and the density for $\boldsymbol{A}[\mathcal{T}]$ depends on the specific forms of the potential and constraints. In Appendix A, we further comment on imposing inequality constraints, the differentiability of the transport map, and the numerical cost of solving the optimization problem.

IV. APPLICATIONS

Using our framework developed in Sec. III, we demonstrate its applicability to the thermodynamic tasks listed in Table I. Throughout, the expectation value of a physical observable $\boldsymbol{a}(\boldsymbol{r})$ with respect to $p_{\rm ini}$ is denoted by $\langle \boldsymbol{a}(\boldsymbol{r}) \rangle_{\rm ini}$.

A. Particle transport

The simplest application of our framework is particle transport, that is, the task of transporting a particle from its initial position to a target position. We take the entropy production Σ_τ as the thermodynamic cost C_τ. We focus on transport problems in which only the expectation value of position is relevant. Accordingly, we impose on the final distribution the constraint $\boldsymbol{A}[p_{\rm fin}] = \langle \boldsymbol{r} \rangle_{\rm fin} = \boldsymbol{\mu}_f$, where $\boldsymbol{\mu}_f$ is the target position. From Eqs. (17) and (24), the equations corresponding to Eqs. (14) and (15) are given by

$$\frac{2 p_{\rm ini}(\boldsymbol{r})}{D\tau}(\mathcal{T}(\boldsymbol{r}) - \boldsymbol{r}) + \boldsymbol{\lambda} p_{\rm ini}(\boldsymbol{r}) = 0, \tag{25}$$

$$\int \mathrm{d}\boldsymbol{r}\, \mathcal{T}(\boldsymbol{r}) p_{\rm ini}(\boldsymbol{r}) - \boldsymbol{\mu}_f = 0, \tag{26}$$

which can be solved straightforwardly to yield

$$\mathcal{T}^{\rm opt}(\boldsymbol{r}) = \boldsymbol{r} + \boldsymbol{\mu}_f - \boldsymbol{\mu}_{\rm ini}, \tag{27}$$

$$p_{\rm fin}^{\rm opt}(\boldsymbol{r}) = p_{\rm ini}(\boldsymbol{r} - \boldsymbol{\mu}_f + \boldsymbol{\mu}_{\rm ini}), \tag{28}$$

where we define $\boldsymbol{\mu}_{\rm ini} := \langle \boldsymbol{r} \rangle_{\rm ini}$.

B. Thermal squeezing

Thermal squeezing is the reduction of fluctuations along a certain direction in phase space below that of a thermal state. Thermally squeezed states, when used as information carriers, enable information erasure beyond the Landauer bound [63], and when used as a heat bath, allow the design of heat engines beyond the Carnot bound [64,65]. Such thermal squeezing has been realized experimentally in mechanical oscillators [66] and in nanoparticles [67,68].

In Sec. IV A, we constrained only the expectation value of position $\boldsymbol{r}$ with respect to $p_{\rm fin}$. Here, by additionally constraining the covariance of final distributions, we apply our framework to thermal squeezing. We take the entropy production Σ_τ as the thermodynamic cost C_τ, and impose the constraints $\langle \boldsymbol{r} \rangle_{\rm fin} = \boldsymbol{\mu}_f$ and $\langle (\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm fin})(\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm fin})^\top \rangle_{\rm fin} = \Xi_f$ on the expectation and covariance of position, where $\cdot^\top$ represents the transpose.

Here, we derive the optimal transport map and the minimal entropy production for thermal squeezing. The arguments of the Lagrangian Eq. (13) are the transport map $\mathcal{T}(\boldsymbol{r})$ and the Lagrange multipliers $\boldsymbol{\nu}$ and Λ corresponding to constraining the expectation value and the covariance of position at the final time. Since the covariance matrix is a positive-definite symmetric matrix, Λ is taken to be symmetric. The term corresponding to constraining the covariance matrix is $\mathrm{Tr}[\Lambda(\langle (\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm fin})(\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm fin})^\top \rangle_{\rm fin} - \Xi_f)]$. Using Eqs. (17) and (24), the equations corresponding to Eqs. (14) and (15) are given by

$$\frac{2 p_{\rm ini}(\boldsymbol{r})}{D\tau}(\mathcal{T}(\boldsymbol{r}) - \boldsymbol{r}) + p_{\rm ini}(\boldsymbol{r})\boldsymbol{\nu} + 2 p_{\rm ini}(\boldsymbol{r}) \Lambda (\mathcal{T}(\boldsymbol{r}) - \boldsymbol{\mu}_f) = 0, \tag{29}$$

$$\int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \mathcal{T}(\boldsymbol{r}) - \boldsymbol{\mu}_f = 0, \tag{30}$$

$$\int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \mathcal{T}(\boldsymbol{r}) \mathcal{T}(\boldsymbol{r})^\top - \boldsymbol{\mu}_f \boldsymbol{\mu}_f^\top - \Xi_f = 0. \tag{31}$$

Solving these equations yields the optimal transport map

$$\mathcal{T}^{\rm opt}(\boldsymbol{r}) = \Xi_{\rm ini}^{-\frac{1}{2}} \left(\Xi_{\rm ini}^{\frac{1}{2}} \Xi_f \Xi_{\rm ini}^{\frac{1}{2}} \right)^{\frac{1}{2}} \Xi_{\rm ini}^{-\frac{1}{2}} (\boldsymbol{r} - \boldsymbol{\mu}_{\rm ini}) + \boldsymbol{\mu}_f, \tag{32}$$

where $\boldsymbol{\mu}_{\rm ini} := \langle \boldsymbol{r} \rangle_{\rm ini}$ and $\Xi_{\rm ini} := \langle (\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm ini})(\boldsymbol{r} - \langle \boldsymbol{r} \rangle_{\rm ini})^\top \rangle_{\rm ini}$. Thus, the optimal transport map is affine in $\boldsymbol{r}$. The optimal entropy production is given by

$$\Sigma_\tau^{\rm opt} = \frac{1}{D\tau} \left(\|\boldsymbol{\mu}_f - \boldsymbol{\mu}_{\rm ini}\|^2 + \mathrm{Tr}\left[\Xi_{\rm ini} + \Xi_f - 2 \left(\Xi_{\rm ini}^{\frac{1}{2}} \Xi_f \Xi_{\rm ini}^{\frac{1}{2}} \right)^{\frac{1}{2}} \right] \right). \tag{33}$$

As an example, we demonstrate thermodynamically optimal thermal squeezing for a non-Gaussian initial distribution in two dimensions (see Fig. 2). Figure 2(c) shows the trade-off relation that the smaller the variance in x of the final distributions, the larger the entropy production required for thermal squeezing.

We demonstrate that affine transport maps are optimal for particle transport and thermal squeezing even for non-Gaussian distributions, as illustrated in Fig. 2. We note that the optimal solutions in Eqs. (32) and (33) follow directly from the results of Ref. [69]. While it is sufficient to use quadratic potentials to implement optimal transport for Gaussian distributions [36], that is not the case for non-Gaussian

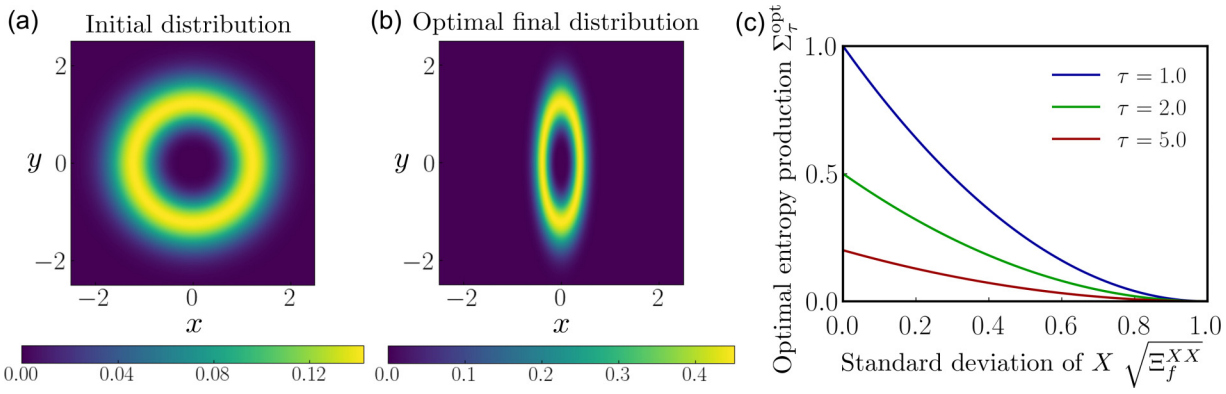


FIG. 2. Non-Gaussian example of thermodynamically optimal thermal squeezing in finite time. (a) Initial distribution before performing thermal squeezing. We take as the initial distribution the intensity profile of a normalized Laguerre-Gaussian mode with radial index 0 and azimuthal index 3 [62], which is a non-Gaussian distribution. (b) Final distribution after performing thermal squeezing with $\Xi_f^{XX} = 0.1$. (c) Dependence of the minimum entropy production $\Sigma_\tau^{\rm opt}$ on the standard deviation of X in the final distributions $\sqrt{\Xi_f^{XX}}$. This plot illustrates the trade-off relation that the smaller the variance of the final distributions, the larger the minimal entropy production required. The required minimum entropy production also decreases as the operation time τ increases. The diffusion coefficient is set to $D = 1$.

distributions, including optimal thermal squeezing for non-Gaussian distributions discussed here.

C. Work in systems with quadratic potentials

The probability distributions of overdamped Langevin systems trapped in quadratic potentials remain Gaussian if the initial distributions are Gaussian. Due to the analytical tractability of this setting, previous studies have derived the finite-time optimal control protocol [16] and the efficiency of a Brownian heat engine at maximum power [14].

Using our framework, we can analytically optimize the work in this setting and recover the results of earlier work [41]. This corresponds to taking $C_\tau = W_\tau$ and imposing no constraint on the final distributions. Given a Gaussian initial distribution, we minimize the work using a quadratic potential $V_t(\boldsymbol{r}) = \frac{1}{2}(\boldsymbol{r} - \boldsymbol{p}(t))^\top Q(t)(\boldsymbol{r} - \boldsymbol{p}(t))$, where $\boldsymbol{p}(t)$ is the center of the potential and $Q(t)$ is its strength, which is a positive-definite symmetric matrix. As in Sec. III C 2, we fix the potentials at $t = 0$ and $t = \tau$. Since $p_t(\boldsymbol{r})$ is Gaussian at all times, the optimal transport map is restricted to the form

$$\mathcal{T}(\boldsymbol{r}) = \Xi_{\rm ini}^{-\frac{1}{2}}\left(\Xi_{\rm ini}^{\frac{1}{2}}\Xi_{\rm fin}\Xi_{\rm ini}^{\frac{1}{2}}\right)^{\frac{1}{2}}\Xi_{\rm ini}^{-\frac{1}{2}}(\boldsymbol{r} - \boldsymbol{\mu}_{\rm ini}) + \boldsymbol{\mu}_{\rm fin}, \qquad (34)$$

where $\boldsymbol{\mu}_{\rm ini}$, $\boldsymbol{\mu}_{\rm fin}$, $\Xi_{\rm ini}$, and $\Xi_{\rm fin}$ are the means and covariance matrices of the initial and final distributions, respectively. Substituting this into the variational equation (20) and solving for $\boldsymbol{\mu}_{\rm fin}$ and $\Xi_{\rm fin}$ reproduces the results of Ref. [41].

D. Information erasure

We next consider information erasure as an application. Together with measurement and feedback discussed later, information erasure is an elementary operation of information processing. According to Landauer's principle [70], erasing one bit of information stored in a symmetric double-well potential necessarily generates at least $k_{\rm B}T \ln 2$ of heat, which has been verified experimentally [27–30,71,72]. The problem of optimizing the entropy production and the work required for finite-time information erasure has been addressed in earlier studies [39,40] by constraining the cumulative distribution function at $r = 0$, given as $\int_{-\infty}^{0} {\rm d}r\, p_{\rm fin}(r)$, which represents the probability that the particle is in the "0" state. In our framework, choosing the same thermodynamic cost C_τ and the same quantity $A[p_{\rm fin}]$ reproduces the results of Refs. [39,40].

On the other hand, by choosing other quantities as $\boldsymbol{A}[p_{\rm fin}]$, one can examine the erasure process from a different point of view. We explore this possibility in Appendix C by taking the exponentially weighted average of the position as $\boldsymbol{A}[p_{\rm fin}]$.

E. Control of free energy

We next take the nonequilibrium free energy as the quantity to be constrained. For a system in contact with a single heat bath, the fundamental bound on the extractable work is given by the change in nonequilibrium free energy, so preparing states with large free energy is important for designing work-storage devices. From this viewpoint, optimization of free energy has also been studied in the context of free-energy harvesting [73,74].

In this case, the variation of the free energy with respect to the transport map is given by Eq. (19). Using our framework, one can determine the optimal transport map and the optimal final distribution that yields the desired change in free energy. As a demonstration, we provide an example of optimal control of free energy in Appendix D.

F. Measurement and feedback

Besides information erasure, fundamental tasks in information processing include measurement and feedback. In the setting of information thermodynamics, the second law of thermodynamics has been generalized to include mutual information [50]. Recent studies have extended these results to finite-time processes using optimal transport theory [38,52–57]. Among these, studies on overdamped Langevin

systems [52,57] fix the final distributions, leaving room for further optimization.

1. Setup for measurement and feedback

We now introduce the setup of information thermodynamics. We consider a bipartite system XY, with X being the system of interest and Y the memory. The subsystems X and Y are each d dimensional, and the total system decomposes as $\boldsymbol{r} = (\boldsymbol{r}^X, \boldsymbol{r}^Y)^\top$, where $\boldsymbol{r}^X$ and $\boldsymbol{r}^Y$ are the positions of the subsystems X and Y, respectively. In the measurement process, Y evolves in time depending on X to generate correlations between X and Y. In the feedback process, in contrast, X evolves in time depending on Y to extract work or reduce fluctuations in X by consuming the correlation between X and Y. We write the mean local velocity as $\boldsymbol{v}_t^{XY}(\boldsymbol{r}) = (\boldsymbol{v}_t^X(\boldsymbol{r}), \boldsymbol{v}_t^Y(\boldsymbol{r}))^\top$ and the probability distribution of the total system as $p_t^{XY}(\boldsymbol{r})$. The thermodynamic costs of the subsystems under information processing are given by the partial entropy productions [48,50,51]:

$$\Sigma_\tau^i := \frac{1}{D}\int_0^\tau \mathrm{d}t \int \mathrm{d}\boldsymbol{r} \left\| \boldsymbol{v}_t^i(\boldsymbol{r}) \right\|^2 p_t^{XY}(\boldsymbol{r}), \quad (i = X, Y). \tag{35}$$

By definition, the total entropy production satisfies $\Sigma_\tau^{XY} = \Sigma_\tau^X + \Sigma_\tau^Y$.

We consider the optimization problem of thermodynamic costs of measurement and feedback under the constraint that the mutual information of the final distribution matches a desired value I_f. Although there is some freedom in choosing what to adopt as the thermodynamic cost, such as the partial entropy production or the total entropy production, we consider a setting in which we first optimize Σ_τ^X (resp. Σ_τ^Y) and subsequently optimize Σ_τ^Y (resp. Σ_τ^X) for the measurement (resp. feedback) process. For the measurement process, this optimization can be written as

$$\min_{\{p_t^{XY}, \boldsymbol{v}_t^{XY}\}_{0\leqslant t\leqslant\tau} \in S} \Sigma_\tau^Y\left[\left\{p_t^{XY}, \boldsymbol{v}_t^{XY}\right\}_{0\leqslant t\leqslant\tau}\right], \tag{36}$$

where S is the set defined as

$$S := \operatorname*{argmin}_{\substack{\{p_t^{XY}, \boldsymbol{v}_t^{XY}\}_{0\leqslant t\leqslant\tau} \text{ s.t.} \\ p_0^{XY} = p_{\mathrm{ini}}^{XY},\, I[p_\tau^{XY}] = I_f}} \Sigma_\tau^X\left[\left\{p_t^{XY}, \boldsymbol{v}_t^{XY}\right\}_{0\leqslant t\leqslant\tau}\right]. \tag{37}$$

The corresponding expression for the feedback process can be obtained by exchanging X and Y. In this case, as a result of the optimization, $\boldsymbol{v}_t^X = \boldsymbol{0}$ (resp. $\boldsymbol{v}_t^Y = \boldsymbol{0}$) holds throughout the process, and therefore $\Sigma_\tau^X = 0$ (resp. $\Sigma_\tau^Y = 0$) holds [57]. This situation corresponds to the usual case where the marginal probability distribution of the subsystem being measured (resp. performing feedback) does not change in time. Therefore, the optimization problem (36) can be rewritten as

$$\min_{\substack{p_{\mathrm{fin}} \text{ s.t.} \\ I_{\mathrm{fin}}[p_{\mathrm{fin}}] = I_f}} \min_{\substack{\{p_t^{XY}, \boldsymbol{v}_t^{XY}\}_{0\leqslant t\leqslant\tau} \text{ s.t.} \\ p_0^{XY} = p_{\mathrm{ini}}^{XY},\, p_\tau^{XY} = p_{\mathrm{fin}}^{XY}}} \Sigma_\tau^Y\left[\left\{p_t^{XY}, \boldsymbol{v}_t^{XY}\right\}_{0\leqslant t\leqslant\tau}\right]\Big|_{\Sigma_\tau^X = 0} \tag{38}$$

for the measurement process, where the counterpart for the feedback process follows by exchanging X and Y. In Appendix F, as a more general situation, we optimize a weighted sum of the partial entropy productions, which includes the total entropy production and the setting of this section as special cases.

Under this setting, we define the thermodynamic costs of measurement and feedback as $\Sigma_{\tau,\mathrm{meas}}^Y := \Sigma_\tau^Y|_{\Sigma_\tau^X = 0}$ and $\Sigma_{\tau,\mathrm{fb}}^X := \Sigma_\tau^X|_{\Sigma_\tau^Y = 0}$, respectively. The corresponding transport maps are given by $\boldsymbol{\mathcal{T}}_{\mathrm{meas}}(\boldsymbol{r}) = (\boldsymbol{r}^X, \boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}(\boldsymbol{r}^Y; \boldsymbol{r}^X))^\top$ for measurement and $\boldsymbol{\mathcal{T}}_{\mathrm{fb}}(\boldsymbol{r}) = (\boldsymbol{\mathcal{T}}_{\mathrm{fb}}^{X|Y}(\boldsymbol{r}^X; \boldsymbol{r}^Y), \boldsymbol{r}^Y)^\top$ for feedback, where $\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}$ and $\boldsymbol{\mathcal{T}}_{\mathrm{fb}}^{X|Y}$ denote the transport maps conditioned on subsystems X and Y, respectively.

In the remainder of this section, we consider only the measurement process. The equations presented below also hold for the feedback process by exchanging X and Y. In analogy with $\tilde{\Sigma}_\tau[\boldsymbol{\mathcal{T}}]$ introduced in Eq. (9), we define the corresponding quantity for the partial entropy productions as

$$\tilde{\Sigma}_{\tau,\mathrm{meas}}^Y\left[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}\right] := \frac{1}{D\tau}\int \mathrm{d}\boldsymbol{r} \left\| \boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}(\boldsymbol{r}^Y; \boldsymbol{r}^X) - \boldsymbol{r}^Y \right\|^2 p_{\mathrm{ini}}^{XY}(\boldsymbol{r}). \tag{39}$$

Meanwhile, a previous study [57] showed the following relation:

$$\begin{aligned} &\min_{\substack{\{p_t^{XY}, \boldsymbol{v}_t^{XY}\}_{0\leqslant t\leqslant\tau} \\ \text{s.t. } p_0^{Y|X} = p_{\mathrm{ini}}^{Y|X},\, p_\tau^{Y|X} = p_{\mathrm{fin}}^{Y|X}}} \Sigma_{\tau,\mathrm{meas}}^Y\left[\left\{p_t^{XY}, \boldsymbol{v}_t^{XY}\right\}_{0\leqslant t\leqslant\tau}\right] \\ &\quad = \min_{\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X} \text{ s.t. } \boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X} \sharp p_{\mathrm{ini}}^{Y|X} = p_{\mathrm{fin}}^{Y|X}} \tilde{\Sigma}_{\tau,\mathrm{meas}}^Y\left[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}\right], \end{aligned} \tag{40}$$

where $p^{Y|X}$ is the conditional probability distribution of Y given X. Therefore, by almost the same arguments as in Sec. III A, the optimization problem for the measurement process in Eq. (36) is reduced to

$$\Sigma_\tau^X = 0 \tag{41}$$

and

$$\min_{\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X} \text{ s.t. } I_{\mathrm{fin}}[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}] = I_f} \tilde{\Sigma}_{\tau,\mathrm{meas}}^Y\left[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}\right]. \tag{42}$$

The mutual information of the final distribution is given by $I_{\mathrm{fin}}[p_{\mathrm{fin}}^{XY}] = \int \mathrm{d}\boldsymbol{r}\, p_{\mathrm{fin}}^{XY}(\boldsymbol{r}) \iota_{\mathrm{fin}}(\boldsymbol{r})$, where $\iota_{\mathrm{fin}}(\boldsymbol{r}) := \ln[p_{\mathrm{fin}}^{XY}(\boldsymbol{r}) / (p_{\mathrm{fin}}^X(\boldsymbol{r}^X) p_{\mathrm{fin}}^Y(\boldsymbol{r}^Y))]$ is the stochastic mutual information at the final time. In this case, the Lagrangian is given by $\mathcal{L} = \tilde{\Sigma}_{\tau,\mathrm{meas}}^Y[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}] + \lambda(I_{\mathrm{fin}}[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}] - I_f)$. The variational formula for the partial entropy production with respect to the transport map is then given by

$$\frac{\delta \tilde{\Sigma}_{\tau,\mathrm{meas}}^Y\left[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}\right]}{\delta \boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}(\boldsymbol{r}^Y; \boldsymbol{r}^X)} = \frac{2}{D\tau}\left(\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}(\boldsymbol{r}^Y; \boldsymbol{r}^X) - \boldsymbol{r}^Y\right) p_{\mathrm{ini}}^{XY}(\boldsymbol{r}). \tag{43}$$

In general, the minima of the two partial entropy productions are incompatible and exhibit a trade-off. Previous studies [56,57] introduced the Pareto front of the partial entropy productions to represent such trade-offs between two incompatible quantities. The minimization problems considered here correspond to end points of the Pareto front in Ref. [57], and the optimal transport map at these points is the Knothe-Rosenblatt map [1,2,57].

2. Measurement

We consider a process in which the subsystem Y measures X within finite time τ to generate more mutual information than the initial distribution $I_f > I_{\mathrm{ini}}$. Here, we take the thermodynamic cost as $C_\tau = \Sigma_\tau^Y|_{\Sigma_\tau^X = 0}$ ($\tilde{C}_\tau = \tilde{\Sigma}_{\tau,\mathrm{meas}}^Y[\boldsymbol{\mathcal{T}}_{\mathrm{meas}}^{Y|X}]$) and

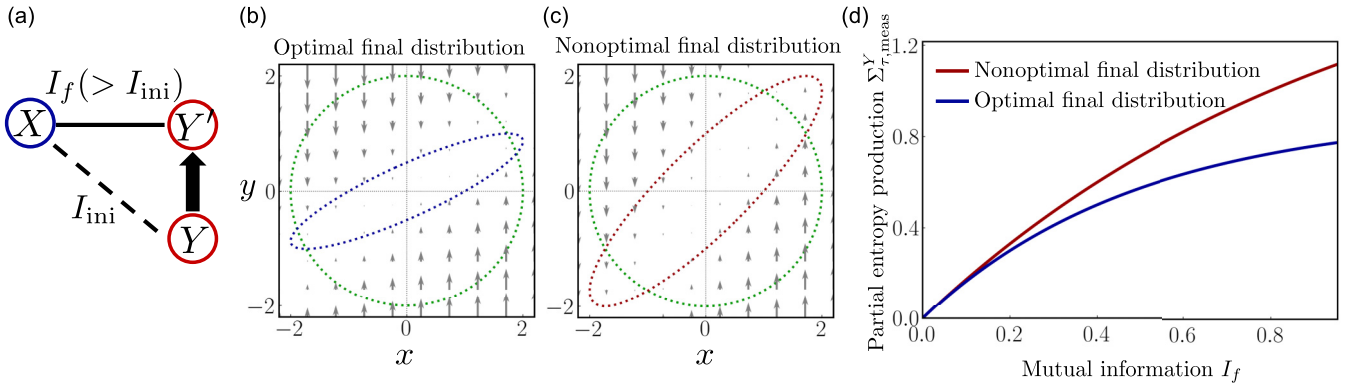


FIG. 3. An example of thermodynamically optimal finite-time measurement. (a) Schematic illustration of the measurement process. The probability current in the X direction vanishes, while the measuring system Y evolves in time to generate a correlation $I_f(>I_{\rm ini})$ between X and Y. (b), (c) Initial distribution (green dashed ellipse), optimal final distribution (blue dashed ellipse), and nonoptimal final distribution (red dashed ellipse). We represent each by an ellipse whose semiaxes are twice the standard deviations. The gray arrows show the optimal protocol $\boldsymbol{v}_t^{\rm opt}(\boldsymbol{r})$ at $t=\tau/2$ scaled by $1/8$. As the initial distribution, we take a state with zero mutual information, with covariance components $\Xi^{XX}=\Xi^{YY}_{\rm ini}=1.0$ and $\Xi^{XY}_{\rm ini}=\Xi^{YX}_{\rm ini}=0$. As the final distribution, we take a state with mutual information $I_f=\ln 2$. In this case, the optimal final distribution satisfies Eq. (46). As a nonoptimal distribution, we set $(\Xi^{YY}_{\rm fin})^{\rm nonopt}=\Xi^{YY}_{\rm ini}=1.0$. (d) Comparison of the partial entropy production required for the measurement process for the optimal and nonoptimal final distributions. The larger the desired mutual information I_f of the final distributions, the larger the thermodynamic cost required for measurement, which scales as $(\Sigma^Y_{\tau,\rm meas})^{\rm opt}\propto 1-e^{-2I_f}$. The parameters are set to $D=1.0$ and $\tau=1.1$.

the constraint on the final distributions to be $I_{\rm fin}[\mathcal{T}^{Y|X}_{\rm meas}]=I_f$. We have to obtain the variational formula for mutual information, which corresponds to the second term in Eq. (14). The variation of the mutual information is given by

$$\frac{\delta I_{\rm fin}\big[\mathcal{T}^{Y|X}_{\rm meas}\big]}{\delta \mathcal{T}^{Y|X}_{{\rm meas},i}}=\partial_{r_j}\{\iota_{\rm fin}(\mathcal{T}_{\rm meas}(\boldsymbol{r}))\}\tilde{J}_{ji}(\boldsymbol{r})p^{XY}_{\rm ini}(\boldsymbol{r}). \tag{44}$$

See Appendix B 3 for details. Combining this with Eq. (43), the variational equation to be solved is

$$\frac{2}{D\tau}\big(\mathcal{T}^{Y|X}_{{\rm meas},i}(\boldsymbol{r}^Y;\boldsymbol{r}^X)-r_i\big)p^{XY}_{\rm ini}(\boldsymbol{r}) + \lambda\partial_{r_j}\{\iota_{\rm fin}(\mathcal{T}(\boldsymbol{r}))\}\tilde{J}_{ji}(\boldsymbol{r})p^{XY}_{\rm ini}(\boldsymbol{r})=0. \tag{45}$$

Together with the constraint $I_{\rm fin}[\mathcal{T}^{Y|X}_{\rm meas}]=I_f$, we have the optimal transport map and the optimal final distribution that minimizes the partial entropy production in the measurement process.

As an illustration, we restrict ourselves to the case of Gaussian distributions throughout the process, which allows us to analytically determine the optimal final distribution and the partial entropy production for the optimal measurement process. In Fig. 3, we plot the initial distribution with zero mutual information, the optimal final distribution with mutual information I_f obtained from our framework, and a fixed nonoptimal final distribution with the same I_f. We also compute and plot the partial entropy production $\Sigma^Y_{\tau,\rm meas}$ for the optimal and nonoptimal final distributions. With the correlation coefficient $\rho:=\Xi^{XY}/\sqrt{\Xi^{XX}\Xi^{YY}}$, the mutual information is given by $I=-\frac{1}{2}\ln(1-\rho^2)$, and thus under the constraint $I_{\rm fin}=I_f$ we have $\rho_{\rm fin}=\rho_f=\sqrt{1-e^{-2I_f}}$. Under this condition, the optimal variance of Y, $(\Xi^{YY}_{\rm fin})^{\rm opt}$, and the optimal value of the partial entropy production required for measurement, $(\Sigma^Y_{\tau,\rm meas})^{\rm opt}$, are given by

$$\big(\Xi^{YY}_{\rm fin}\big)^{\rm opt}=\big(1-\rho_f^2\big)\Xi^{YY}_{\rm ini}, \tag{46}$$

$$\big(\Sigma^Y_{\tau,\rm meas}\big)^{\rm opt}=\frac{\Xi^{YY}_{\rm ini}\rho_f^2}{D\tau}. \tag{47}$$

See Appendix E for details.

3. Feedback

We consider a process in which the subsystem Y performs feedback on X in finite time τ to consume an amount $I_{\rm ini}-I_f>0$ of mutual information. Here, we take the thermodynamic cost as $C_\tau=\Sigma^X_\tau|_{\Sigma^Y_\tau=0}$ ($\tilde{C}_\tau=\tilde{\Sigma}^X_{\tau,\rm fb}[\mathcal{T}^{X|Y}_{\rm fb}]$). The variational equation to be solved is obtained from Eq. (45) by exchanging X and Y.

As in the measurement case, we restrict ourselves to Gaussian distributions throughout the process. Figure 4 shows an example of thermodynamically optimal feedback for Gaussian distributions. We compute and plot the partial entropy production $\Sigma^X_{\tau,\rm fb}$ for the optimal and nonoptimal final distributions. We plot the initial distribution with mutual information $I_{\rm ini}=\ln 3$. The optimal final distribution and a nonoptimal final distribution are also plotted in the case of $I_f=0$. In this setting, in analogy with the measurement process, we can analytically determine the optimal final distribution and the partial entropy production for the optimal feedback process. The optimal variance of X, $(\Xi^{XX}_{\rm fin})^{\rm opt}$, and the optimal value of the partial entropy production required for feedback, $(\Sigma^X_{\tau,\rm fb})^{\rm opt}$, are given by

$$\big(\Xi^{XX}_{\rm fin}\big)^{\rm opt}=\Big(\rho_{\rm ini}\rho_f+\sqrt{\big(1-\rho^2_{\rm ini}\big)\big(1-\rho_f^2\big)}\Big)^2\Xi^{XX}_{\rm ini}, \tag{48}$$

$$\big(\Sigma^X_{\tau,\rm fb}\big)^{\rm opt}=\frac{\Xi^{XX}_{\rm ini}}{D\tau}\Big[1-\Big\{\rho_{\rm ini}\rho_f+\sqrt{\big(1-\rho^2_{\rm ini}\big)\big(1-\rho_f^2\big)}\Big\}^2\Big]. \tag{49}$$

See Appendix E for details.

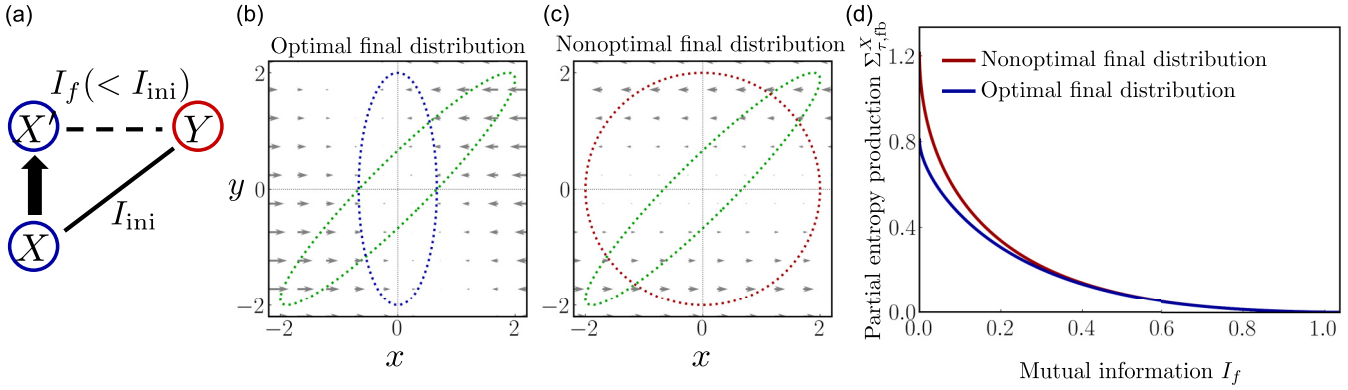


FIG. 4. An example of thermodynamically optimal finite-time feedback. (a) Schematic illustration of the feedback process. The probability current in the Y direction vanishes, while the measured system X evolves in time, consuming the correlation and transitioning to a state with $I_f(<I_{\rm ini})$. (b), (c) Initial distribution (green dashed ellipse), optimal final distribution (blue dashed ellipse), and nonoptimal final distribution (red dashed ellipse). We represent each by an ellipse whose semiaxes are twice the standard deviations. The gray arrows show the optimal protocol $\boldsymbol{v}_t^{\rm opt}(\boldsymbol{r})$ at $t=\tau/2$ scaled by $1/8$. As the initial distribution, we take a state with mutual information $I_{\rm ini}=\ln 3$. The covariance components are $\Xi^{XX}_{\rm ini}=\Xi^{YY}=1.0$ and $\Xi^{XY}_{\rm ini}=\Xi^{YX}_{\rm ini}=2\sqrt{2}/3$. As the final distribution, we take a state with mutual information $I_f=0$. In this case, the optimal final distribution satisfies Eq. (48). As a nonoptimal distribution, we set $(\Xi^{XX}_{\rm fin})^{\rm nonopt}=\Xi^{XX}_{\rm ini}=1.0$. (d) Comparison of the partial entropy production required for the feedback process for the optimal and nonoptimal final distributions. The smaller the mutual information I_f of the final distributions, the larger the thermodynamic cost required for feedback. The parameters are set to $D=1.0$ and $\tau=1.1$.

V. CONCLUSION AND DISCUSSION

We have constructed a general framework for optimizing the thermodynamic cost required to perform various tasks in overdamped Langevin systems. That is, we have developed a framework in which quantities depending on the final distributions attain desired values with minimal thermodynamic cost. To this end, we combined optimal transport theory with a variational method over the transport map. The standard setup of thermodynamic speed limits based on optimal transport theory fixes the initial and final distributions, whereas we relax that assumption and instead constrain quantities that depend on the final distributions specific to the thermodynamic task. Consequently, among the many possible final distributions that realize a given task, our general framework [Eqs. (14) and (15)] determines the optimal transport map leading to the optimal final distribution. In the broad setting of thermodynamic tasks where a functional of the final distribution is constrained to match a desired value, we derived the variational equation for the optimal transport map.

Our framework is applicable to far-from-equilibrium situations, and thus admits various applications (see Table I and Sec. IV). In contrast to the case where the final distribution is fixed, we demonstrated that further optimization of the thermodynamic cost is possible in analytically tractable examples and in numerical examples.

Thermodynamic speed limits based on optimal transport theory have been extended not only to the overdamped Langevin systems considered in this study [36–38], but also to Markov jump processes [43–46], open quantum systems [43,46], and nonlinear systems [45,47]. Extending our framework to these cases is left for future work. For Markov jump processes, previous studies [54,55] have tackled this problem for measurement and feedback setups, but a unified framework is still elusive.

From an experimental viewpoint, optimizing the partial entropy production generally requires complete control of nonconservative forces in addition to the potential [57]. Moreover, thermodynamically optimal information processing in multipartite systems demands flexible control of interactions. Even in a system with optical tweezers where optimal transport has been experimentally implemented [33], arbitrary control of interactions remains challenging.

ACKNOWLEDGMENTS

This work was supported by JST ERATO Grant No. JPMJER2302, Japan. K.T. and R.N. were supported by World-leading Innovative Graduate Study Program for Materials Research, Information, and Technology (MERIT-WINGS) of the University of Tokyo. K.F. was supported by JSPS KAKENHI Grants No. JP23K13036 and No. JP24H00831. T.S. was also supported by JST CREST Grant No. JPMJCR20C1 and by Institute of AI and Beyond of the University of Tokyo.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX A: SEVERAL REMARKS

In this Appendix, we provide several remarks on the general framework presented in Sec. III.

We first comment that the constraints on final distributions can be generalized to inequality constraints $\boldsymbol{A}[\mathcal{T}]\leqslant\boldsymbol{A}_f$ or $\boldsymbol{A}[\mathcal{T}]\geqslant\boldsymbol{A}_f$. In this case, the Lagrange multiplier $\boldsymbol{\lambda}$ must be non-negative or nonpositive, respectively, according to the Karush-Kuhn-Tucker conditions [61]. Conditions of

convexity of the Lagrangian density $\mathscr{L}(\boldsymbol{r}, \boldsymbol{\mathcal{T}}, \frac{\partial \mathcal{T}}{\partial \boldsymbol{r}}, \boldsymbol{\lambda})$ may differ from those for equality-constrained problems.

We also comment on the differentiability of the transport map. In deriving the Euler-Lagrange equation, we assume the differentiability of the transport map. However, in singular settings such as when the final distribution is concentrated at a single point [39,40], the solution of the Euler-Lagrange equation can fail to be differentiable. In such cases, it is known that singular solutions can be handled by a weak formulation of the Euler-Lagrange equation [61].

Finally, we discuss the numerical cost of our framework if analytic solutions are unavailable. Suppose we discretize each axis into k grid points and numerically solve the variational equation by employing Newton's method or one of its variants. In our case, for d-dimensional systems, the number of unknown variables scales as $\mathcal{O}(k^d)$ and the numerical cost of finding a stationary solution is $\mathcal{O}(k^d)$ per iteration. Therefore, one-dimensional systems can be solved efficiently, as $k \sim 10^4$ is feasible in practice and one can exploit the monotonicity $\mathcal{T}'(r) > 0$. In our numerical example of free energy control, we set $k = 601$. For two-dimensional systems, we expect that a discretization with $k \sim 10^2$–10^3 is still feasible in practice.

APPENDIX B: DERIVATION OF THE MAIN RESULTS

In this Appendix, we provide the details of the derivations of the results in Sec. III.

1. Variation of free energy

To obtain the variational formula for the work with respect to the transport map in Eq. (20), we need the variation of the nonequilibrium free energy in Eq. (19). In this Appendix, we give the details of that derivation.

For the state $p_{\rm fin}$ and the potential V_τ at time $t = \tau$, the nonequilibrium free energy is

$$\begin{aligned}\mathcal{F}_{\rm fin} &= \beta^{-1} D\big(p_{\rm fin} \big|\big| p_{\rm fin}^{\rm eq}\big) + \mathcal{F}_{\rm fin}^{\rm eq} \\ &= \beta^{-1} \int \mathrm{d}\boldsymbol{r}\, p_{\rm fin}(\boldsymbol{r}) \ln \frac{p_{\rm fin}(\boldsymbol{r})}{p_{\rm fin}^{\rm eq}(\boldsymbol{r})} + \mathcal{F}_{\rm fin}^{\rm eq} \\ &= \beta^{-1} \int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \ln \frac{p_{\rm ini}(\boldsymbol{r})}{p_{\rm fin}^{\rm eq}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{r})} + \mathcal{F}_{\rm fin}^{\rm eq},\end{aligned} \tag{B1}$$

where we used Eq. (7). Thus, we have

$$\begin{aligned}\frac{\delta \mathcal{F}_\tau}{\delta \mathcal{T}_i} = -\beta^{-1}\Bigg(&\frac{\delta}{\delta \mathcal{T}_i}\int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \ln p_{\rm fin}^{\rm eq}(\boldsymbol{\mathcal{T}}(\boldsymbol{r})) \\ &+ \frac{\delta}{\delta \mathcal{T}_i}\int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \ln \det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{r})\Bigg),\end{aligned} \tag{B2}$$

where the first term is given by

$$\frac{\delta}{\delta \mathcal{T}_i}\int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \ln p_{\rm fin}^{\rm eq}(\boldsymbol{\mathcal{T}}(\boldsymbol{r})) = p_{\rm ini}(\boldsymbol{r})\, \partial_{r_i'} \ln p_{\rm fin}^{\rm eq}(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}(\boldsymbol{r})}. \tag{B3}$$

On the other hand, the second term requires the variation of $\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{r})$. For this, we use Jacobi's formula [75]: For a matrix $A(t)$ differentiable with respect to a parameter t,

$$\frac{\mathrm{d} \det A(t)}{\mathrm{d}t} = \det A(t) \mathrm{tr}\left(A^{-1}(t)\frac{\mathrm{d}A(t)}{\mathrm{d}t}\right). \tag{B4}$$

Using Eq. (B4), we have

$$\begin{aligned}\delta \int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \ln \det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{r}) &= \int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \mathrm{tr}[J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{r})^{-1} \delta J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{r})] \\ &= \int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \tilde{J}_{ji}(\boldsymbol{r}) \delta\big(\partial_{r_j} \mathcal{T}_i(\boldsymbol{r})\big) \\ &= -\int \mathrm{d}\boldsymbol{r}\, \delta \mathcal{T}_i\, \partial_{r_j}(p_{\rm ini}(\boldsymbol{r}) \tilde{J}_{ji}(\boldsymbol{r})),\end{aligned} \tag{B5}$$

that is,

$$\frac{\delta}{\delta \mathcal{T}_i}\int \mathrm{d}\boldsymbol{r}\, p_{\rm ini}(\boldsymbol{r}) \ln \det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{r}) = -\partial_{r_j}(p_{\rm ini}(\boldsymbol{r}) \tilde{J}_{ji}(\boldsymbol{r})). \tag{B6}$$

Combining the above yields Eq. (19).

2. Variation of a functional of the final distribution

We give the details of the derivation of Eq. (22). We have

$$\delta \int \mathrm{d}\boldsymbol{r}\, \boldsymbol{g}(p_{\rm fin}(\boldsymbol{r}), \boldsymbol{r}) = \int \mathrm{d}\boldsymbol{r}\, \frac{\partial \boldsymbol{g}(X, \boldsymbol{r})}{\partial X}\bigg|_{X=p_{\rm fin}(\boldsymbol{r})} \delta p_{\rm fin}(\boldsymbol{r}), \tag{B7}$$

where, using Eq. (7),

$$\delta p_{\rm fin}(\boldsymbol{r}) = \delta \frac{p_{\rm ini}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}{\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}. \tag{B8}$$

We then need to express $\delta \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})$ as a variation with respect to $\boldsymbol{\mathcal{T}}(\boldsymbol{r})$. In Appendix B 2 a, we express $\delta \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})$ in terms of variations of $\boldsymbol{\mathcal{T}}(\boldsymbol{r})$, from which we obtain the variation of $p_{\rm fin}(\boldsymbol{r})$ in Eq. (B8), and then Eq. (22).

a. Variation of the inverse function of the transport map

Here, we derive the variation of $\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})$ needed for Eq. (B8). For an infinitesimal variation $\delta \boldsymbol{\mathcal{T}}$ of the transport map,

$$(\boldsymbol{\mathcal{T}} + \delta \boldsymbol{\mathcal{T}})((\boldsymbol{\mathcal{T}} + \delta \boldsymbol{\mathcal{T}})^{-1}(\boldsymbol{r})) = \boldsymbol{r} \tag{B9}$$

holds, while

$$\begin{aligned}\boldsymbol{\mathcal{T}}((\boldsymbol{\mathcal{T}} + \delta \boldsymbol{\mathcal{T}})^{-1}(\boldsymbol{r})) &= \boldsymbol{\mathcal{T}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) + J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) \delta \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}) \\ &= \boldsymbol{r} + J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) \delta \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}).\end{aligned} \tag{B10}$$

Subtracting the two sides of Eqs. (B9) and (B10) gives

$$\delta \boldsymbol{\mathcal{T}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) = -J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) \delta \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}). \tag{B11}$$

Multiplying on the left by $(J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})))^{-1} = J_{\boldsymbol{\mathcal{T}}^{-1}}(\boldsymbol{r})$ yields

$$\begin{aligned}\delta \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}) &= -(J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})))^{-1} \delta \boldsymbol{\mathcal{T}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) \\ &= -J_{\boldsymbol{\mathcal{T}}^{-1}}(\boldsymbol{r}) \delta \boldsymbol{\mathcal{T}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})),\end{aligned} \tag{B12}$$

that is,

$$\delta \mathcal{T}_i^{-1}(\boldsymbol{r}) = -\tilde{J}_{ij}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) \delta \mathcal{T}_j(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})). \tag{B13}$$

b. Variation of $p_{\rm fin}(\boldsymbol{r})$

Using the result obtained in Appendix B 2 a, the variation of $p_{\rm fin}(\boldsymbol{r})$ is given by

$$\begin{aligned}\delta p_{\rm fin}(\boldsymbol{r}) &= \delta\frac{p_{\rm ini}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}{\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}\\ &= \frac{\delta p_{\rm ini}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}{\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))} - \frac{p_{\rm ini}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\delta\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}{(\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})))^2}\\ &= \frac{\partial_{r_i'}p_{\rm ini}(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\delta\mathcal{T}_i^{-1}(\boldsymbol{r})}{\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}\\ &\quad - \frac{p_{\rm ini}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\mathrm{tr}[(J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})))^{-1}\delta J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))]}{\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}.\end{aligned} \tag{B14}$$

Here, $\delta J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))$ is given by

$$\begin{aligned}&\delta(J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})))_{ij}\\ &= \delta\big(\partial_{r_j'}\mathcal{T}_i(\boldsymbol{r}')|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\big)\\ &= \partial_{r_j'}(\mathcal{T}_i+\delta\mathcal{T}_i)(\boldsymbol{r}')|_{\boldsymbol{r}'=(\boldsymbol{\mathcal{T}}+\delta\boldsymbol{\mathcal{T}})^{-1}(\boldsymbol{r})} - \partial_{r_j'}\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\\ &= \partial_{r_j'}\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=(\boldsymbol{\mathcal{T}}+\delta\boldsymbol{\mathcal{T}})^{-1}(\boldsymbol{r})} - \partial_{r_j'}\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\\ &\quad + \partial_{r_j'}\delta\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\\ &= \partial_{r_k'}\partial_{r_j'}\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\delta\mathcal{T}_k^{-1}(\boldsymbol{r}) + \partial_{r_j'}\delta\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\\ &= H_{kji}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\delta\mathcal{T}_k^{-1}(\boldsymbol{r}) + \partial_{r_j'}\delta\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})},\end{aligned} \tag{B15}$$

where $H_{kji}(\boldsymbol{r}) = \partial_{r_k}\partial_{r_j}\mathcal{T}_i(\boldsymbol{r})$. Rewriting $((J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})))^{-1})_{ji}$ as $\tilde{J}_{ji}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))$ and using Eq. (B13), we have

$$\begin{aligned}\delta p_{\rm fin}(\boldsymbol{r}) = \frac{1}{\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}\big[&-\partial_{r_i'}p_{\rm ini}(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\tilde{J}_{ij}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\\ &\times\delta\mathcal{T}_j(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) - p_{\rm ini}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\tilde{J}_{ji}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\\ &\cdot\big\{-H_{kji}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\tilde{J}_{kl}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\delta\mathcal{T}_l(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\\ &+\partial_{r_j'}\delta\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\big\}\big].\end{aligned} \tag{B16}$$

Hence, Eq. (B7) is transformed into

$$\begin{aligned}\delta\int\mathrm{d}\boldsymbol{r}\,\boldsymbol{g}(p_{\rm fin}(\boldsymbol{r}),\boldsymbol{r}) &= \int\mathrm{d}\boldsymbol{r}\,\frac{\hat{\boldsymbol{g}}(\boldsymbol{r})}{\det J_{\boldsymbol{\mathcal{T}}}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))}\big[-\partial_{r_i'}p_{\rm ini}(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\tilde{J}_{ij}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\delta\mathcal{T}_j(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\\ &\quad -p_{\rm ini}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\tilde{J}_{ji}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\big\{-H_{kji}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\tilde{J}_{kl}(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r}))\delta\mathcal{T}_l(\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})) + \partial_{r_j'}\delta\mathcal{T}_i(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{r})}\big\}\big]\\ &= \int\mathrm{d}\boldsymbol{r}\,\hat{\boldsymbol{g}}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))[-\partial_{r_i}(p_{\rm ini}(\boldsymbol{r}))\tilde{J}_{ij}(\boldsymbol{r})\delta\mathcal{T}_j(\boldsymbol{r}) - p_{\rm ini}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r})\{-H_{kji}(\boldsymbol{r})\tilde{J}_{kl}(\boldsymbol{r})\delta\mathcal{T}_l(\boldsymbol{r}) + \partial_{r_j}\delta\mathcal{T}_i\}]\\ &= \int\mathrm{d}\boldsymbol{r}\,\hat{\boldsymbol{g}}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))[\{-\partial_{r_j}(p_{\rm ini}(\boldsymbol{r}))\tilde{J}_{ji}(\boldsymbol{r}) + p_{\rm ini}(\boldsymbol{r})\tilde{J}_{kl}(\boldsymbol{r})H_{jkl}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r})\}\delta\mathcal{T}_i - p_{\rm ini}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r})\partial_{r_j}\delta\mathcal{T}_i].\end{aligned} \tag{B17}$$

Using

$$\begin{aligned}\tilde{J}_{kl}(\boldsymbol{r})H_{jkl}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r}) &= \tilde{J}_{kl}(\boldsymbol{r})\big(\partial_{r_j}\partial_{r_k}\mathcal{T}_l(\boldsymbol{r})\big)\tilde{J}_{ji}(\boldsymbol{r})\\ &= \tilde{J}_{kl}(\boldsymbol{r})\big(\partial_{r_k}J_{lj}(\boldsymbol{r})\big)\tilde{J}_{ji}(\boldsymbol{r})\\ &= \partial_{r_k}(\tilde{J}_{kl}(\boldsymbol{r})J_{lj}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r})) - \big(\partial_{r_k}\tilde{J}_{kl}(\boldsymbol{r})\big)J_{lj}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r}) - \tilde{J}_{kl}(\boldsymbol{r})J_{lj}(\boldsymbol{r})\big(\partial_{r_k}\tilde{J}_{ji}(\boldsymbol{r})\big)\\ &= \partial_{r_k}(\tilde{J}_{kl}(\boldsymbol{r})\delta_{li}) - \big(\partial_{r_k}\tilde{J}_{kl}(\boldsymbol{r})\big)\delta_{li} - \delta_{kj}\big(\partial_{r_k}\tilde{J}_{ji}(\boldsymbol{r})\big)\\ &= -\partial_{r_j}\tilde{J}_{ji}(\boldsymbol{r}),\end{aligned} \tag{B18}$$

for $\boldsymbol{A}[p_{\rm fin}] = \int\mathrm{d}\boldsymbol{r}\,\boldsymbol{g}(p_{\rm fin}(\boldsymbol{r}),\boldsymbol{r})$, we have

$$\begin{aligned}\delta\int\mathrm{d}\boldsymbol{r}\,\boldsymbol{g}(p_{\rm fin}(\boldsymbol{r}),\boldsymbol{r}) &= \int\mathrm{d}\boldsymbol{r}\,\hat{\boldsymbol{g}}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))\big[\big\{-\partial_{r_j}(p_{\rm ini}(\boldsymbol{r}))\tilde{J}_{ji}(\boldsymbol{r}) - p_{\rm ini}(\boldsymbol{r})\partial_{r_j}(\tilde{J}_{ji}(\boldsymbol{r}))\big\}\delta\mathcal{T}_i - p_{\rm ini}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r})\partial_{r_j}\delta\mathcal{T}_i\big]\\ &= \int\mathrm{d}\boldsymbol{r}\,\partial_{r_j}\{\hat{\boldsymbol{g}}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))\}\tilde{J}_{ji}(\boldsymbol{r})p_{\rm ini}(\boldsymbol{r})\delta\mathcal{T}_i.\end{aligned} \tag{B19}$$

Therefore, we have Eq. (22).

Specifically, if $\boldsymbol{A}[p_{\rm fin}]$ is the expectation value of a physical observable at the final time $\boldsymbol{A}[p_{\rm fin}] = \langle\boldsymbol{a}\rangle_{\rm fin} = \int\mathrm{d}\boldsymbol{r}\,\boldsymbol{a}(\boldsymbol{r})p_{\rm fin}(\boldsymbol{r})$, then $\boldsymbol{g}(p_{\rm fin}(\boldsymbol{r}),\boldsymbol{r}) = \boldsymbol{a}(\boldsymbol{r})p_{\rm fin}(\boldsymbol{r})$. Substituting it into Eq. (22), we have

$$\begin{aligned}\partial_{r_j}\left\{\frac{\partial(\boldsymbol{a}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))X)}{\partial X}\bigg|_{X=p_{\rm fin}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))}\right\}\tilde{J}_{ji}(\boldsymbol{r})p_{\rm ini}(\boldsymbol{r}) &= \partial_{r_j}\{\boldsymbol{a}(\boldsymbol{\mathcal{T}}(\boldsymbol{r}))\}\tilde{J}_{ji}(\boldsymbol{r})p_{\rm ini}(\boldsymbol{r}) = \partial_{r_k'}\boldsymbol{a}(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}(\boldsymbol{r})}J_{kj}(\boldsymbol{r})\tilde{J}_{ji}(\boldsymbol{r})p_{\rm ini}(\boldsymbol{r})\\ &= \partial_{r_k'}\boldsymbol{a}(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}(\boldsymbol{r})}\delta_{ki}p_{\rm ini}(\boldsymbol{r}) = \partial_{r_i'}\boldsymbol{a}(\boldsymbol{r}')\big|_{\boldsymbol{r}'=\boldsymbol{\mathcal{T}}(\boldsymbol{r})}p_{\rm ini}(\boldsymbol{r}).\end{aligned} \tag{B20}$$

Thus, we have Eq. (24), which was derived directly in the main text.

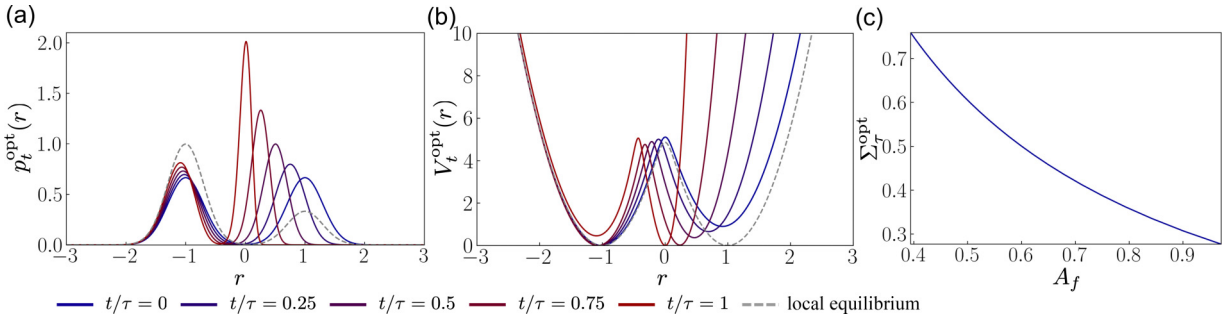


FIG. 5. Numerical example of optimal finite-time information erasure. (a) Intermediate probability distributions during the optimal erasure process. (b) Intermediate optimal potentials during the erasure process. The parameter values are $T = 1.0$, $D = 1.0$, $\tau = 1.0$, $m = 1.0$, $\sigma = 0.3$, and $\gamma = 2.0$. In panels (a) and (b), we set $A_f \simeq 0.5631$, for which the final erasure error, $\int_0^\infty {\rm d}r\, p_{\rm fin}(r)$, is 0.25. In panel (a), we plot the corresponding local equilibrium distributions $p^{\rm leq}(r) = 1.5 p_{\rm ini}(r)$ $(r \leqslant 0)$ and $p^{\rm leq}(r) = 0.5 p_{\rm ini}(r)$ $(r > 0)$ for comparison (dashed line). We fix the constant of the potential in Eq. (C6) so that the minimum value of the potential is zero. (c) Dependence of the minimal entropy production $\Sigma_\tau^{\rm opt}$ on the final value of $A[p_{\rm fin}]$. Panel (c) shows that decreasing $A[p_{\rm fin}]$ increases the minimal entropy production required for erasure.

3. Variation of mutual information

In this Appendix, we provide the details of the derivation of Eq. (44). We omit the subscript "meas" or "fb" for simplicity. The variation of the mutual information at the final time is given by

$$\begin{aligned}\delta I_{\rm fin} &= \delta \int {\rm d}\boldsymbol{r}\, p_{\rm fin}^{XY}(\boldsymbol{r}) \ln \frac{p_{\rm fin}^{XY}(\boldsymbol{r})}{p_{\rm fin}^{X}(\boldsymbol{r}^X) p_{\rm fin}^{Y}(\boldsymbol{r}^Y)} \\ &= \int {\rm d}\boldsymbol{r} \left[\left(\delta p_{\rm fin}^{XY}(\boldsymbol{r})\right) \iota_{\rm fin}(\boldsymbol{r}) + \delta p_{\rm fin}^{XY}(\boldsymbol{r}) \right] \\ &\quad - \int {\rm d}\boldsymbol{r}^X\, \delta p_{\rm fin}^{X}(\boldsymbol{r}^X) - \int {\rm d}\boldsymbol{r}^Y\, \delta p_{\rm fin}^{Y}(\boldsymbol{r}^Y) \\ &= \int {\rm d}\boldsymbol{r} \left(\delta p_{\rm fin}^{XY}(\boldsymbol{r})\right) \iota_{\rm fin}(\boldsymbol{r}), \end{aligned} \tag{B21}$$

where

$$\iota_{\rm fin}(\boldsymbol{r}) = \ln \frac{p_{\rm fin}^{XY}(\boldsymbol{r})}{p_{\rm fin}^{X}(\boldsymbol{r}^X) p_{\rm fin}^{Y}(\boldsymbol{r}^Y)} \tag{B22}$$

is the stochastic mutual information at time $t = \tau$. Using Eq. (22), we have

$$\frac{\delta I_{\rm fin}}{\delta \mathcal{T}_i} = \partial_{r_j} \{ \iota_{\rm fin}(\boldsymbol{\mathcal{T}}(\boldsymbol{r})) \} \tilde{J}_{ji}(\boldsymbol{r}) p_{\rm ini}^{XY}(\boldsymbol{r}). \tag{B23}$$

We note that Eq. (B23) is equivalent to Eq. (22) with $\hat{g}(\boldsymbol{r}) = \iota_{\rm fin}(\boldsymbol{r})$.

APPENDIX C: AN EXAMPLE OF INFORMATION ERASURE

In this Appendix, we give an example of finite-time optimal information erasure. We consider a one-dimensional system and take the exponential moment as the quantity to be constrained, which is given by

$$A[p_{\rm fin}] = \langle e^{\gamma r} \rangle_{\rm fin} = \int_{-\infty}^{\infty} {\rm d}r\, e^{\gamma r} p_{\rm fin}(r). \tag{C1}$$

We set $\gamma > 0$ and $0 < A_f < A_{\rm ini} = \int_{-\infty}^{\infty} {\rm d}r\, e^{\gamma r} p_{\rm ini}(r)$ so that erasure is successful in the sense that the final distribution is localized in the "0" state corresponding to the region $r \leqslant 0$. This quantity penalizes the final distribution for having a long positive tail at parameter γ in the positive r region, which corresponds to the "1" state. We note that we impose a different constraint from the one used in Refs. [39,40], where the cumulative distribution function at $r = 0$ was adopted. We choose the entropy production Σ_τ as the thermodynamic cost C_τ. We note that, if $\lambda > 0$, the Lagrangian density is convex in $\mathcal{T}(r)$, leading to the global optimality of the solution. The variational formulas corresponding to Eqs. (14) and (15) are given by

$$\frac{2}{D\tau}(\mathcal{T}(r) - r) p_{\rm ini}(r) + \lambda\gamma e^{\gamma \mathcal{T}(r)} p_{\rm ini}(r) = 0, \tag{C2}$$

$$\int_{-\infty}^{\infty} {\rm d}r\, e^{\gamma \mathcal{T}(r)} p_{\rm ini}(r) - A_f = 0. \tag{C3}$$

Equation (C2) gives the optimal transport map

$$\mathcal{T}^{\rm opt}(r) = r - \frac{1}{\gamma} W_0(\alpha e^{\gamma r}), \tag{C4}$$

where W_0 is the principal branch of the Lambert W function and we introduce the parameter $\alpha = \lambda D \tau \gamma^2/2 > 0$, which is uniquely determined by solving Eq. (C3).

In Fig. 5, we demonstrate the finite-time optimal information erasure for this choice of $A[p_{\rm fin}]$ by numerically determining α from Eq. (C3) and reconstructing the optimal transport map and protocol. We take the initial distribution to be an equally weighted mixture of two Gaussian distributions

$$p_{\rm ini}(r) = \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma^2}} \left(e^{-\frac{(r-m)^2}{2\sigma^2}} + e^{-\frac{(r+m)^2}{2\sigma^2}} \right), \tag{C5}$$

which is the Boltzmann distribution associated with the double-well potential given by

$$\begin{aligned} V_0(r) &= -\beta^{-1} \ln p_{\rm ini}(r) + {\rm const.} \\ &= \beta^{-1} \frac{r^2 + m^2}{2\sigma^2} - \beta^{-1} \ln \cosh\left(\frac{mr}{\sigma^2}\right) + {\rm const.} \end{aligned} \tag{C6}$$

We note that the potential $V_0(r)$ is a double-well potential if $m^2 > \sigma^2$, since the second derivative of $V_0(r)$ at $r = 0$ is given by $V_0''(0) = \beta^{-1}(\sigma^2 - m^2)/\sigma^4$. In contrast to Refs. [39,40], the optimal final distribution is not a singular distribution localized at $r = 0$ but a smooth distribution.

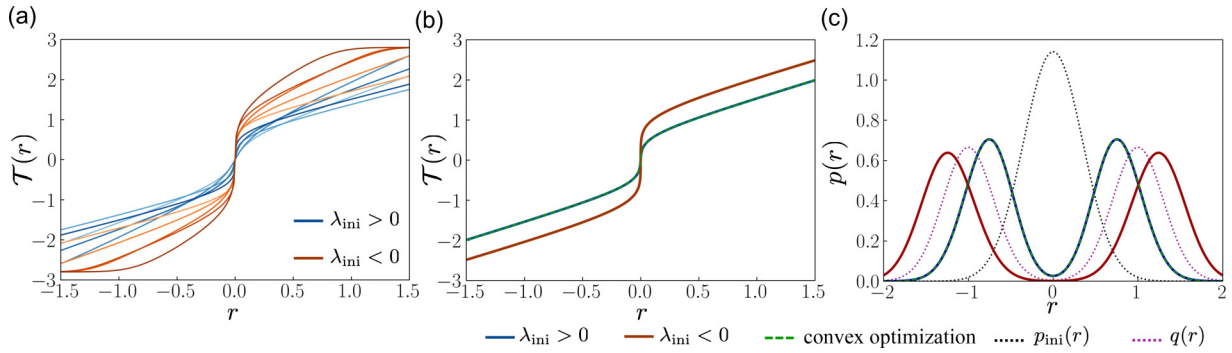


FIG. 6. Numerical example of optimal control of free energy in finite time. (a) Initial choices of transport map $\mathcal{T}(r)$, where blue lines correspond to initial choices of $\lambda_{\rm ini} > 0$ and red lines correspond to initial choices of $\lambda_{\rm ini} < 0$. (b) Stationary transport maps $\mathcal{T}(r)$ for different initial guesses shown in panel (a) obtained by numerically solving Eq. (D4) together with the constraint on the KL divergence. (c) Initial distribution $p_{\rm ini}(r)$ (black dotted line), equilibrium distribution with respect to the final potential $q(r)$ (purple dotted line), and optimized final distributions $p_{\rm fin}^{\rm opt}(r)$. Initial guesses with $\lambda_{\rm ini} > 0$ converge to the optimal transport map (green dashed line), which is obtained by solving the convex optimization problem (D5), while initial guesses with $\lambda_{\rm ini} < 0$ converge to another transport map. This indicates that the numerical iteration converges to a stationary solution, which is not globally optimal due to the nonconvexity for $\lambda_{\rm ini} < 0$. We take the initial distribution as a Gaussian distribution with mean 0 and standard deviation 0.35, and the final potential as a negative logarithm of a Gaussian mixture distribution with means ± 1 and standard deviation 0.3. We impose $D_{\rm KL}(p_{\rm fin}||q) = 0.3$, where $q(r)$ is the equilibrium distribution associated with the final potential. We set the diffusion coefficient, the inverse temperature and the operation time to $D = 1$, $\beta = 1$, and $\tau = 1$. As a result, the entropy production for the globally optimal solution ($\lambda_{\rm ini} > 0$) is $\Sigma_\tau^{\rm opt} = 0.249$ and that for the stationary solution ($\lambda_{\rm ini} < 0$) is $\Sigma_\tau = 0.923$.

Even for a single thermodynamic task such as information erasure, there is arbitrariness in the choice of the criterion that determines successful achievement of the task. Our framework can handle general constraints on the final distributions based on theoretical and experimental requirements when performing thermodynamic tasks.

APPENDIX D: AN EXAMPLE OF FREE ENERGY CONTROL

Here, we demonstrate the optimal control of free energy in finite time for a one-dimensional non-Gaussian distribution. To do so, we fix the potential at $t = \tau$ and impose the constraint that the final distribution $p_{\rm fin}$ has a fixed KL divergence from the equilibrium distribution with respect to the final potential $q(r) \propto e^{-\beta V_{\rm fin}(r)}$, which is equivalent to fixing the free energy change. Specifically, we fix the equilibrium distribution $q(r)$ as

$$q(r) = \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}}\left(e^{-\frac{(r-m)^2}{2\sigma^2}} + e^{-\frac{(r+m)^2}{2\sigma^2}}\right), \tag{D1}$$

which is a mixture of two Gaussian distributions with means $\pm m$ and standard deviation σ. Accordingly, the final potential is given by

$$\begin{aligned} V_{\rm fin}(r) &= -\beta^{-1}\ln q(r) + \text{const.} \\ &= -\beta^{-1}\ln\left(e^{-\frac{(r-m)^2}{2\sigma^2}} + e^{-\frac{(r+m)^2}{2\sigma^2}}\right) + \text{const.} \end{aligned} \tag{D2}$$

We consider the optimization problem of minimizing the entropy production Σ_τ under the constraint that the KL divergence is fixed at $D_{\rm KL}(p_{\rm fin}||q) = A_f < D_{\rm KL}(p_{\rm ini}||q)$. The Lagrangian for this problem is given by

$$\begin{aligned} \mathcal{L} = &\frac{1}{D\tau}\int {\rm d}r\,(\mathcal{T}(r) - r)^2 p_{\rm ini}(r) \\ &+ \lambda\left\{\int {\rm d}r\, p_{\rm ini}(r)\ln\frac{p_{\rm ini}(r)}{q(\mathcal{T}(r))\mathcal{T}'(r)} - A_f\right\}. \end{aligned} \tag{D3}$$

We comment on the convexity of the Lagrangian density in Eq. (D3) with respect to $\mathcal{T}(r)$ and $\mathcal{T}'(r)$. The Lagrangian density is not convex in $\mathcal{T}(r)$ in general since the convexity is governed by the curvature of $-\lambda\ln q$. For the present Gaussian mixture, $-\ln q$ is not convex when $m > \sigma$, since $-\frac{{\rm d}^2\ln q(r)}{{\rm d}r^2}|_{r=0} = (\sigma^2 - m^2)/\sigma^4 < 0$. In contrast, the Lagrangian density is convex in $\mathcal{T}'(r)$ for $\lambda > 0$ but not for $\lambda < 0$. As a result, the global optimality of the solution is not guaranteed in general.

Varying the Lagrangian (D3) with respect to $\mathcal{T}(r)$, we obtain the following equation:

$$\begin{aligned} \frac{2}{D\tau}(\mathcal{T}(r) - r)p_{\rm ini}(r) + \lambda\Bigg[&- p_{\rm ini}(r)\frac{\rm d}{{\rm d}r'}\ln q(r')\Big|_{r'=\mathcal{T}(r)} \\ &+ \frac{\rm d}{{\rm d}r}\left(\frac{p_{\rm ini}(r)}{\mathcal{T}'(r)}\right)\Bigg] = 0. \end{aligned} \tag{D4}$$

We solve this equation numerically and obtain the optimal transport map $\mathcal{T}^{\rm opt}(r)$ and the optimal final distribution $p_{\rm fin}$ (see Fig. 6). We then check whether the solution of Eq. (D4) matches the solution of the corresponding convex optimization problem in the Kantorovich formulation that guarantees global optimality:

$$\min_{p_{\rm fin}\ \text{s.t.}\ D_{\rm KL}(p_{\rm fin}||q)\leqslant A_f}\ \min_{\substack{\pi\ \text{s.t.} \\ \int {\rm d}r'\,\pi(r,r') = p_{\rm ini}(r), \\ \int {\rm d}r\,\pi(r,r') = p_{\rm fin}(r')}} \int {\rm d}r{\rm d}r'\,\frac{(r'-r)^2}{D\tau}\pi(r,r'). \tag{D5}$$

Here, the global optimality is guaranteed by the convexity of the sublevel set $\{p_{\rm fin} \mid D_{\rm KL}(p_{\rm fin}||q) \leqslant A_f\}$, the linearity of the cost function $\int {\rm d}r{\rm d}r'\,\frac{(r'-r)^2}{D\tau}\pi(r,r')$ with respect to π, and the linearity of the constraint $\int {\rm d}r\,\pi(r,r') = p_{\rm fin}(r')$. The optimal final distribution of Eq. (D5) satisfies

the equality in the constraint $D_{\mathrm{KL}}(p_{\mathrm{fin}}||q) \leqslant A_f$, so the solution of the problem in Eq. (D5) matches the globally optimal solution of the original problem. The proof of the equality is as follows. We assume that the optimal final distribution $p_{\mathrm{fin}}^{\mathrm{opt}}$ satisfies the inequality $D_{\mathrm{KL}}(p_{\mathrm{fin}}^{\mathrm{opt}}||q) < A_f$. Then, we can construct a new final distribution $p_{\mathrm{fin},\varepsilon} = (1-\varepsilon)p_{\mathrm{fin}}^{\mathrm{opt}} + \varepsilon p_{\mathrm{ini}}$ for sufficiently small $\varepsilon > 0$ that satisfies the inequality $D_{\mathrm{KL}}(p_{\mathrm{fin},\varepsilon}||q) < A_f$. Since $\mathcal{W}_2^2(p_{\mathrm{ini}}, p)$ is convex in p, we have $\mathcal{W}_2^2(p_{\mathrm{ini}}, p_{\mathrm{fin},\varepsilon}) \leqslant (1-\varepsilon)\mathcal{W}_2^2(p_{\mathrm{ini}}, p_{\mathrm{fin}}^{\mathrm{opt}}) + \varepsilon\mathcal{W}_2^2(p_{\mathrm{ini}}, p_{\mathrm{ini}}) = (1-\varepsilon)\mathcal{W}_2^2(p_{\mathrm{ini}}, p_{\mathrm{fin}}^{\mathrm{opt}}) < \mathcal{W}_2^2(p_{\mathrm{ini}}, p_{\mathrm{fin}}^{\mathrm{opt}})$. This contradicts the optimality of $p_{\mathrm{fin}}^{\mathrm{opt}}$, and thus the optimal final distribution satisfies the equality in the constraint $D_{\mathrm{KL}}(p_{\mathrm{fin}}^{\mathrm{opt}}||q) = A_f$.

APPENDIX E: DETAILS OF MEASUREMENT AND FEEDBACK

In this Appendix, we give the details of the examples treated in Sec. IV F. In both measurement and feedback, we consider two-dimensional systems with coordinates $\boldsymbol{r} = (x, y)^\top$, where x and y represent the degrees of freedom of the X and Y systems, respectively. Since what is actually transported is the conditional distribution, it suffices to consider one-dimensional Gaussian distributions. When the initial and final distributions are one-dimensional Gaussians with means $\mu_{\mathrm{ini}}, \mu_{\mathrm{fin}}$ and variances $\Xi_{\mathrm{ini}}, \Xi_{\mathrm{fin}}$, respectively, the transport map between them is given by [69]

$$\mathcal{T}(r) = \sqrt{\frac{\Xi_{\mathrm{fin}}}{\Xi_{\mathrm{ini}}}}(r - \mu_{\mathrm{ini}}) + \mu_{\mathrm{fin}}. \tag{E1}$$

1. Measurement

As the initial distribution, we take

$$p_{\mathrm{ini}}^{XY}(x, y) = \frac{1}{2\pi\sqrt{\Xi^{XX}\Xi_{\mathrm{ini}}^{YY}}} \exp\left(-\frac{x^2}{2\Xi^{XX}} - \frac{y^2}{2\Xi_{\mathrm{ini}}^{YY}}\right), \tag{E2}$$

which has zero mean and zero mutual information. Since the measurement does not change the marginal distribution of X, the mean and variance of x are fixed at $\mu_t^X = 0$ and $\Xi_t^{XX} = \Xi^{XX}$ for all $t \in [0, \tau]$. We restrict the final distributions to be Gaussian. We also adopt an ansatz that the solution to the variational equation (45) is given by

$$\mathcal{T}_{\mathrm{meas}}^{Y|X}(y; x) = \sqrt{\frac{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}{\Xi_{\mathrm{ini}}^{YY}}}y + \mu_{\mathrm{fin}}^Y + \frac{\Xi_{\mathrm{fin}}^{XY}}{\Xi^{XX}}x. \tag{E3}$$

Equation (E3) is the optimal transport map from the conditional distributions $p_{\mathrm{ini}}^{Y|X}(y|x)$ to $p_{\mathrm{fin}}^{Y|X}(y|x)$, which are Gaussian distributions with means $\mu_{\mathrm{ini}}^{Y|X}(x) = 0$ and $\mu_{\mathrm{fin}}^{Y|X}(x) = \mu_{\mathrm{fin}}^Y + \frac{\Xi_{\mathrm{fin}}^{XY}}{\Xi^{XX}}x$, and variances $\Xi_{\mathrm{ini}}^{Y|X} = \Xi_{\mathrm{ini}}^{YY}$ and $\Xi_{\mathrm{fin}}^{Y|X} = \Xi_{\mathrm{fin}}^{YY}(1-\rho_{\mathrm{fin}}^2)$, respectively [69].

In this case, $\tilde{J}_{11}(\boldsymbol{r}) = 1$, $\tilde{J}_{12}(\boldsymbol{r}) = 0$, $\tilde{J}_{21}(\boldsymbol{r}) = -\frac{\Xi_{\mathrm{fin}}^{XY}}{\Xi^{XX}}\sqrt{\frac{\Xi_{\mathrm{ini}}^{YY}}{\Xi_{\mathrm{fin}}^{YY}(1-\rho_{\mathrm{fin}}^2)}}$, and $\tilde{J}_{22} = \sqrt{\frac{\Xi_{\mathrm{ini}}^{YY}}{\Xi_{\mathrm{fin}}^{YY}(1-\rho_{\mathrm{fin}}^2)}}$. Since

$$\begin{aligned}
\frac{\partial}{\partial y}(\iota_{\mathrm{fin}}(\boldsymbol{\mathcal{T}}_{\mathrm{meas}}(\boldsymbol{r}))) &= \frac{\partial}{\partial y}\ln\frac{p_{\mathrm{fin}}^{XY}\left(x, \mathcal{T}_{\mathrm{meas}}^{Y|X}(y;x)\right)}{p_{\mathrm{fin}}^X(x)p_{\mathrm{fin}}^Y\left(\mathcal{T}_{\mathrm{meas}}^{Y|X}(y;x)\right)} \\
&= \frac{\partial}{\partial y}\ln\frac{p_{\mathrm{ini}}(x, y)}{p^X(x)\frac{\partial\mathcal{T}_{\mathrm{meas}}^{Y|X}(y;x)}{\partial y}p_{\mathrm{fin}}^Y\left(\mathcal{T}_{\mathrm{meas}}^{Y|X}(y;x)\right)} \\
&= -\frac{1}{\Xi_{\mathrm{ini}}^{YY}}y + \frac{1}{\Xi_{\mathrm{fin}}^{YY}}\sqrt{\frac{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}{\Xi_{\mathrm{ini}}^{YY}}}\left(\mathcal{T}_{\mathrm{meas}}^{Y|X}(y;x) - \mu_{\mathrm{fin}}^Y\right),
\end{aligned} \tag{E4}$$

we have

$$\begin{aligned}
&\frac{2}{D\tau}\left(\sqrt{\frac{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}{\Xi_{\mathrm{ini}}^{YY}}}y + \mu_{\mathrm{fin}}^Y + \frac{\Xi_{\mathrm{fin}}^{XY}}{\Xi^{XX}}x - y\right) + \lambda\sqrt{\frac{\Xi_{\mathrm{ini}}^{YY}}{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}} \\
&\times\left[-\frac{1}{\Xi_{\mathrm{ini}}^{YY}}y + \frac{1}{\Xi_{\mathrm{fin}}^{YY}}\sqrt{\frac{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}{\Xi_{\mathrm{ini}}^{YY}}}\left(\sqrt{\frac{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}{\Xi_{\mathrm{ini}}^{YY}}}y + \frac{\Xi_{\mathrm{fin}}^{XY}}{\Xi^{XX}}x\right)\right] = 0.
\end{aligned} \tag{E5}$$

By setting the coefficients of x, y, and the constant term to zero, we have

$$\frac{2}{D\tau}\frac{\Xi_{\mathrm{fin}}^{XY}}{\Xi^{XX}} + \lambda\frac{\Xi_{\mathrm{fin}}^{XY}}{\Xi^{XX}\Xi_{\mathrm{fin}}^{YY}} = 0, \tag{E6}$$

$$\frac{2}{D\tau}\left(\sqrt{\frac{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}{\Xi_{\mathrm{ini}}^{YY}}} - 1\right) + \lambda\left(-\frac{1}{\Xi_{\mathrm{ini}}^{YY}}\sqrt{\frac{\Xi_{\mathrm{ini}}^{YY}}{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}} + \frac{1}{\Xi_{\mathrm{fin}}^{YY}}\sqrt{\frac{\Xi_{\mathrm{fin}}^{YY}\left(1-\rho_{\mathrm{fin}}^2\right)}{\Xi_{\mathrm{ini}}^{YY}}}\right) = 0, \tag{E7}$$

$$\frac{2}{D\tau}\mu_{\mathrm{fin}}^Y = 0. \tag{E8}$$

Thus, we have

$$\mu^Y_{\rm fin} = 0, \tag{E9}$$

$$\Xi^{YY}_{\rm fin} = \left(1-\rho^2_{\rm fin}\right)\Xi^{YY}_{\rm ini}, \tag{E10}$$

$$\lambda = -\frac{2}{D\tau}\Xi^{YY}_{\rm fin}. \tag{E11}$$

Substituting $\rho_{\rm fin} = \rho_f$ obtained from the constraint yields the optimal variance of Y in the final distributions (46), and the partial entropy production required for measurement (47).

2. Feedback

As the initial distribution, we take

$$p^{XY}_{\rm ini}(x,y) = \frac{1}{2\pi\sqrt{\left(\Xi^{XX}_{\rm ini}\,\Xi^{YY} - \left(\Xi^{XY}_{\rm ini}\right)^2\right)}} \times \exp\left(-\frac{1}{2}\begin{pmatrix}x & y\end{pmatrix}\begin{pmatrix}\Xi^{XX}_{\rm ini} & \Xi^{XY}_{\rm ini}\\ \Xi^{XY}_{\rm ini} & \Xi^{YY}\end{pmatrix}^{-1}\begin{pmatrix}x\\ y\end{pmatrix}\right), \tag{E12}$$

which has zero mean and finite mutual information if $\Xi^{XY}_{\rm ini} > 0$. Since ideal feedback does not change the marginal distribution of Y, the mean and variance of y are fixed at $\mu^Y_t = 0$ and $\Xi^{YY}_t = \Xi^{YY}$ for all $t \in [0,\tau]$. As in the case of measurement, we restrict the final distributions to be Gaussian. We also assume that the solution to the variational equation is given by the ansatz:

$$\mathcal{T}^{X|Y}(x;y) = \sqrt{\frac{\Xi^{XX}_{\rm fin}\left(1-\rho^2_{\rm fin}\right)}{\Xi^{XX}_{\rm ini}\left(1-\rho^2_{\rm ini}\right)}}\left(x - \frac{\Xi^{XY}_{\rm ini}}{\Xi^{YY}}y\right) + \mu^X_{\rm fin} + \frac{\Xi^{XY}_{\rm fin}}{\Xi^{YY}}y, \tag{E13}$$

which is the optimal transport map from the conditional distributions $p^{X|Y}_{\rm ini}(x|y)$ to $p^{X|Y}_{\rm fin}(x|y)$, which are Gaussian distributions with means $\mu^{X|Y}_{\rm ini}(y) = \frac{\Xi^{XY}_{\rm ini}}{\Xi^{YY}}y$ and $\mu^{X|Y}_{\rm fin}(y) = \mu^X_{\rm fin} + \frac{\Xi^{XY}_{\rm fin}}{\Xi^{YY}}y$, and variances $\Xi^{X|Y}_{\rm ini} = \Xi^{XX}_{\rm ini}(1-\rho^2_{\rm ini})$ and $\Xi^{X|Y}_{\rm fin} = \Xi^{XX}_{\rm fin}(1-\rho^2_{\rm fin})$, respectively. Using these, and solving the equation obtained by exchanging X and Y in Eq. (45) in the same manner as in the measurement setting, we have

$$\mu^X_{\rm fin} = 0, \tag{E14}$$

$$\Xi^{XX}_{\rm fin} = \left(\rho_{\rm ini}\rho_{\rm fin} + \sqrt{\left(1-\rho^2_{\rm ini}\right)\left(1-\rho^2_{\rm fin}\right)}\right)^2\Xi^{XX}_{\rm ini}. \tag{E15}$$

Substituting $\rho_{\rm fin} = \rho_f$ from the constraint yields the optimal variance of X in the final distributions [Eq. (48)] and the partial entropy production required for feedback [Eq. (49)].

APPENDIX F: MEASUREMENT AND FEEDBACK MINIMIZING A WEIGHTED SUM OF PARTIAL ENTROPY PRODUCTIONS

In Sec. IV F, we considered measurement and feedback processes in which we first optimize one partial entropy production and subsequently optimize the other. In this Appendix, we optimize a weighted sum of the partial entropy productions, which includes the total entropy production and the setting of Sec. IV F as special cases.

For a weight $w \in (0,1)$, we define the weighted sum of the partial entropy productions, $\Sigma^w_\tau := (1-w)\Sigma^X_\tau + w\Sigma^Y_\tau (0 < w < 1)$ and use it as the thermodynamic cost. In particular, $\Sigma^{1/2}_\tau$ is equal to one half of the total entropy production. For a fixed final distribution, the limiting problems of minimizing Σ^w_τ as $w \to 0$ and $w \to 1$ are equivalent to the idealized measurement and feedback processes in the main text, respectively [57].

We consider the following optimization problem:

$$\min_{p_{\rm fin}\ \text{s.t.}\ I[p_{\rm fin}]=I_f}\ \min_{\substack{\{p_t,\boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau}\\ \text{s.t.}\ p_0=p_{\rm ini},\ p_\tau=p_{\rm fin}}} \Sigma^w_\tau[\{p_t,\boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau}]. \tag{F1}$$

That is, we seek the minimum value of Σ^w_τ required to prepare a final distribution with mutual information I_f, together with the optimal protocol that attains this minimum. Since the mean local velocities of both X and Y may be nonzero, the distinction between the system of interest and the memory can become ambiguous. The weighted sum of the partial entropy productions can be written as

$$\Sigma^w_\tau = \int_0^\tau {\rm d}t \int {\rm d}\boldsymbol{r}\, \boldsymbol{v}^{XY}_t(\boldsymbol{r})^\top D_w^{-1}\boldsymbol{v}^{XY}_t(\boldsymbol{r})p^{XY}_t(\boldsymbol{r}), \tag{F2}$$

where

$$D_w := \left(\frac{D}{1-w}I_d\right)\oplus\left(\frac{D}{w}I_d\right). \tag{F3}$$

Thus, Eq. (F2) can be interpreted as the total entropy production of a system with diffusion coefficients $D/(1-w)$ and D/w for X and Y, respectively.

To put the weighted sum in Eq. (F2) into a more tractable form, we introduce the rescaled coordinates

$$S_w := (\sqrt{1-w}I_d)\oplus(\sqrt{w}I_d), \tag{F4}$$

$$\tilde{\boldsymbol{r}} := S_w\boldsymbol{r}. \tag{F5}$$

Then, the probability distribution in the $\tilde{\boldsymbol{r}}$ coordinates is given by

$$\tilde{p}^{XY}_t(\tilde{\boldsymbol{r}}) = \frac{1}{\det S_w}p^{XY}_t\left(S_w^{-1}\tilde{\boldsymbol{r}}\right) = \frac{1}{[w(1-w)]^{d/2}}p^{XY}_t(\boldsymbol{r}). \tag{F6}$$

Under this transformation, the Fokker-Planck equation becomes

$$\frac{\partial\tilde{p}^{XY}_t(\tilde{\boldsymbol{r}})}{\partial t} = -\tilde{\boldsymbol{\nabla}}^\top\left(S_w\boldsymbol{v}^{XY}_t\left(S_w^{-1}\tilde{\boldsymbol{r}}\right)\tilde{p}^{XY}_t(\tilde{\boldsymbol{r}})\right). \tag{F7}$$

Here, $\tilde{\boldsymbol{\nabla}} := S_w^{-1}\boldsymbol{\nabla}$ denotes the gradient with respect to $\tilde{\boldsymbol{r}}$. Defining $\tilde{\boldsymbol{v}}^{XY}_t(\tilde{\boldsymbol{r}}) := S_w\boldsymbol{v}^{XY}_t(S_w^{-1}\tilde{\boldsymbol{r}})$, we obtain

$$\frac{\partial\tilde{p}^{XY}_t(\tilde{\boldsymbol{r}})}{\partial t} = -\tilde{\boldsymbol{\nabla}}^\top\left(\tilde{\boldsymbol{v}}^{XY}_t(\tilde{\boldsymbol{r}})\tilde{p}^{XY}_t(\tilde{\boldsymbol{r}})\right). \tag{F8}$$

Furthermore, Eq. (F2) becomes

$$\Sigma^w_\tau = \frac{1}{D}\int_0^\tau {\rm d}t\int {\rm d}\tilde{\boldsymbol{r}}\left\|\tilde{\boldsymbol{v}}^{XY}_t(\tilde{\boldsymbol{r}})\right\|^2\tilde{p}^{XY}_t(\tilde{\boldsymbol{r}}), \tag{F9}$$

which reduces the problem to the form of the total entropy production with diffusion coefficient D.

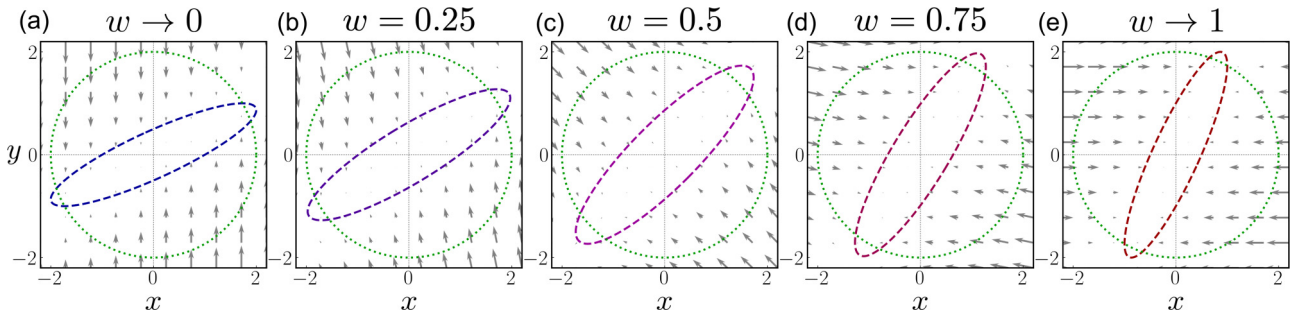


FIG. 7. An example of thermodynamically optimal finite-time measurement that minimizes a weighted sum of partial entropy productions. In panels (a)–(e), we plot the initial distribution (green dashed ellipse) and the optimal final distribution (dashed ellipse) for different values of the weight w. The cases of $w \to 0$ and $w \to 1$ correspond, respectively, to the idealized measurement and feedback discussed in the main text, since in these limits the optimal transport map for the weighted cost converges to the Knothe-Rosenblatt map [2,57]. Each distribution is represented by an ellipse whose semiaxes are twice the standard deviations. The gray arrows indicate the optimal protocol $\boldsymbol{v}_t^{\mathrm{opt}}(\boldsymbol{r})$ at $t = \tau/2$ scaled by $1/8$. As the initial distribution, we take a state with zero mutual information, with covariance components $\Xi_{\mathrm{ini}}^{XX} = \Xi_{\mathrm{ini}}^{YY} = 1.0$ and $\Xi_{\mathrm{ini}}^{XY} = \Xi_{\mathrm{ini}}^{YX} = 0$. As the final distribution, we take a state with mutual information $I_f = \ln 2$. The parameters are set to $D = 1.0$ and $\tau = 1.1$.

We next express the mutual information in the rescaled coordinates. The marginal probability distribution of X is given by

$$\tilde{p}_t^X(\tilde{\boldsymbol{r}}^X) = \int \mathrm{d}\tilde{\boldsymbol{r}}^Y \tilde{p}_t^{XY}(\tilde{\boldsymbol{r}}^X, \tilde{\boldsymbol{r}}^Y) = \frac{1}{(1-w)^{d/2}} p_t^X\big((\sqrt{1-w}^{-1}\tilde{\boldsymbol{r}}^X\big), \tag{F10}$$

and similarly,

$$\tilde{p}_t^Y(\tilde{\boldsymbol{r}}^Y) = \frac{1}{w^{d/2}} p_t^Y\big(\sqrt{w}^{-1}\tilde{\boldsymbol{r}}^Y\big). \tag{F11}$$

Thus, we obtain

$$\frac{\tilde{p}_t^{XY}(\tilde{\boldsymbol{r}}^X, \tilde{\boldsymbol{r}}^Y)}{\tilde{p}_t^X(\tilde{\boldsymbol{r}}^X)\tilde{p}_t^Y(\tilde{\boldsymbol{r}}^Y)} = \frac{p_t^{XY}(\boldsymbol{r}^X, \boldsymbol{r}^Y)}{p_t^X(\boldsymbol{r}^X)p_t^Y(\boldsymbol{r}^Y)}. \tag{F12}$$

Consequently, the mutual information is invariant under the coordinate transformation:

$$I = \int \mathrm{d}\boldsymbol{r}^X \mathrm{d}\boldsymbol{r}^Y p_t^{XY}(\boldsymbol{r}^X, \boldsymbol{r}^Y) \ln \frac{p_t^{XY}(\boldsymbol{r}^X, \boldsymbol{r}^Y)}{p_t^X(\boldsymbol{r}^X)p_t^Y(\boldsymbol{r}^Y)} = \int \mathrm{d}\tilde{\boldsymbol{r}}^X \mathrm{d}\tilde{\boldsymbol{r}}^Y \tilde{p}_t^{XY}(\tilde{\boldsymbol{r}}^X, \tilde{\boldsymbol{r}}^Y) \ln \frac{\tilde{p}_t^{XY}(\tilde{\boldsymbol{r}}^X, \tilde{\boldsymbol{r}}^Y)}{\tilde{p}_t^X(\tilde{\boldsymbol{r}}^X)\tilde{p}_t^Y(\tilde{\boldsymbol{r}}^Y)}. \tag{F13}$$

In analogy with Eq. (9), we define

$$\tilde{\Sigma}_\tau^w[\tilde{\boldsymbol{\mathcal{T}}}] := \frac{1}{D\tau} \int \mathrm{d}\tilde{\boldsymbol{r}} \, \|\tilde{\boldsymbol{\mathcal{T}}}(\tilde{\boldsymbol{r}}) - \tilde{\boldsymbol{r}}\|^2 \tilde{p}_{\mathrm{ini}}^{XY}(\tilde{\boldsymbol{r}}). \tag{F14}$$

Then, as in Eq. (10), we have

$$\min_{\substack{\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau} \\ \text{s.t. } p_0 = p_{\mathrm{ini}},\, p_\tau = p_{\mathrm{fin}}}} \Sigma_\tau^w[\{p_t, \boldsymbol{v}_t\}_{0\leqslant t\leqslant\tau}] = \min_{\tilde{\boldsymbol{\mathcal{T}}} \text{ s.t. } \tilde{\boldsymbol{\mathcal{T}}}\sharp\tilde{p}_{\mathrm{ini}} = \tilde{p}_{\mathrm{fin}}} \tilde{\Sigma}_\tau^w[\tilde{\boldsymbol{\mathcal{T}}}]. \tag{F15}$$

The variation of $\tilde{\Sigma}_\tau^w$ with respect to $\tilde{\boldsymbol{\mathcal{T}}}$ is given by

$$\frac{\delta\tilde{\Sigma}_\tau^w[\tilde{\boldsymbol{\mathcal{T}}}]}{\delta\tilde{\boldsymbol{\mathcal{T}}}(\tilde{\boldsymbol{r}})} = \frac{2}{D\tau}(\tilde{\boldsymbol{\mathcal{T}}}(\tilde{\boldsymbol{r}}) - \tilde{\boldsymbol{r}})\tilde{p}_{\mathrm{ini}}^{XY}(\tilde{\boldsymbol{r}}). \tag{F16}$$

The variation of the final mutual information is

$$\frac{\delta I_{\mathrm{fin}}[\tilde{\boldsymbol{\mathcal{T}}}]}{\delta\tilde{\mathcal{T}}_i(\tilde{\boldsymbol{r}})} = \partial_{\tilde{r}_j}\{\tilde{\imath}_{\mathrm{fin}}(\tilde{\boldsymbol{\mathcal{T}}}(\tilde{\boldsymbol{r}}))\}\tilde{J}_{ji}(\tilde{\boldsymbol{r}})\tilde{p}_{\mathrm{ini}}^{XY}(\tilde{\boldsymbol{r}}), \tag{F17}$$

where

$$\tilde{\imath}_{\mathrm{fin}}(\tilde{\boldsymbol{r}}) = \ln \frac{\tilde{p}_{\mathrm{fin}}^{XY}(\tilde{\boldsymbol{r}})}{\tilde{p}_{\mathrm{fin}}^X(\tilde{\boldsymbol{r}}^X)\tilde{p}_{\mathrm{fin}}^Y(\tilde{\boldsymbol{r}}^Y)}, \tag{F18}$$

and

$$\tilde{J}_{ji}(\tilde{\boldsymbol{r}}) = \partial_{\tilde{r}'_i}\tilde{\mathcal{T}}_j^{-1}(\tilde{\boldsymbol{r}}')\big|_{\tilde{\boldsymbol{r}}'=\tilde{\boldsymbol{\mathcal{T}}}(\tilde{\boldsymbol{r}})}. \tag{F19}$$

Together with the constraint $I_{\mathrm{fin}}[\tilde{\boldsymbol{\mathcal{T}}}] = I_f$, Eqs. (F16) and (F17) determine the optimal transport map $\tilde{\boldsymbol{\mathcal{T}}}_{\mathrm{opt}}$ and the corresponding value $\Sigma_\tau^{w,\mathrm{opt}}$. The protocol in the original coordinate system $\boldsymbol{r}$ is obtained by transforming the optimal quantities in the $\tilde{\boldsymbol{r}}$ coordinates back to the $\boldsymbol{r}$ coordinates.

As an example, we consider an optimal measurement process analogous to that in Sec. IV F 2 but minimizing Σ_τ^w with $w = 0,\ 0.25,\ 0.5,\ 0.75$, and 1. Restricting the distributions to be Gaussian throughout the process, we obtain the optimal final distributions as shown in Fig. 7. We see that the optimal final distributions, together with the optimal protocols, depend on the weight w.

The infimum of the optimization problem for measurement in Eqs. (36) and (F1) can become zero if arbitrary final distributions are allowed, because strong correlations can then be created on infinitesimally small length scales. Therefore, the class of admissible probability distributions needs to be restricted. A physically natural way to do so is to require that the probability distributions vary only on length scales where the overdamped description is valid. Another way, which we adopt in the examples, is to restrict the admissible probability distributions to Gaussian distributions, for which the infimum is nonzero.

[1] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, Heidelberg, 2009), Vol. 338.

[2] F. Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling* (Birkhäuser, Cham, 2015).

[3] G. Peyré and M. Cuturi, Computational optimal transport with applications to data sciences, Found. Trends Mach. Learn. **11**, 355 (2019).

[4] M. Arjovsky, S. Chintala, and L. Bottou, Wasserstein generative adversarial networks, in *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh (PMLR, Cambridge, MA, 2017), Vol. 70, pp. 214–223.

[5] Y. Rubner, C. Tomasi, and L. J. Guibas, The earth mover's distance as a metric for image retrieval, Int. J. Comput. Vision **40**, 99 (2000).

[6] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, and E. S. Lander, Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming, Cell **176**, 928 (2019).

[7] K. Sekimoto, *Stochastic Energetics* (Springer, Berlin, 2010), Vol. 799.

[8] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[9] L. Peliti and S. Pigolotti, *Stochastic Thermodynamics: An Introduction* (Princeton University Press, Princeton, NJ, 2021).

[10] N. Shiraishi, *An Introduction to Stochastic Thermodynamics* (Springer, Singapore, 2023).

[11] S. Ciliberto, Experiments in stochastic thermodynamics: Short history and perspectives, Phys. Rev. X **7**, 021051 (2017).

[12] N. Shiraishi, K. Saito, and H. Tasaki, Universal trade-off relation between power and efficiency for heat engines, Phys. Rev. Lett. **117**, 190601 (2016).

[13] C. Van den Broeck, Thermodynamic efficiency at maximum power, Phys. Rev. Lett. **95**, 190602 (2005).

[14] T. Schmiedl and U. Seifert, Efficiency at maximum power: An analytically solvable model for stochastic heat engines, Europhys. Lett. **81**, 20003 (2008).

[15] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, Efficiency at maximum power of low-dissipation Carnot engines, Phys. Rev. Lett. **105**, 150603 (2010).

[16] T. Schmiedl and U. Seifert, Optimal finite-time processes in stochastic thermodynamics, Phys. Rev. Lett. **98**, 108301 (2007).

[17] G. E. Crooks, Measuring thermodynamic length, Phys. Rev. Lett. **99**, 100602 (2007).

[18] D. A. Sivak and G. E. Crooks, Thermodynamic metrics and optimal paths, Phys. Rev. Lett. **108**, 190602 (2012).

[19] S. Blaber and D. A. Sivak, Optimal control in stochastic thermodynamics, J. Phys. Commun. **7**, 033001 (2023).

[20] N. Shiraishi, K. Funo, and K. Saito, Speed limit for classical stochastic processes, Phys. Rev. Lett. **121**, 070601 (2018).

[21] S. Ito, Stochastic thermodynamic interpretation of information geometry, Phys. Rev. Lett. **121**, 030605 (2018).

[22] S. Ito and A. Dechant, Stochastic time evolution, information geometry, and the Cramér-Rao bound, Phys. Rev. X **10**, 021056 (2020).

[23] A. C. Barato and U. Seifert, Thermodynamic uncertainty relation for biomolecular processes, Phys. Rev. Lett. **114**, 158101 (2015).

[24] 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).

[25] J. M. Horowitz and T. R. Gingrich, Thermodynamic uncertainty relations constrain non-equilibrium fluctuations, Nat. Phys. **16**, 15 (2020).

[26] D. H. Wolpert, J. Korbel, C. W. Lynn, F. Tasnim, J. A. Grochow, G. Kardeş, J. B. Aimone, V. Balasubramanian, E. De Giuli, D. Doty, N. Freitas, M. Marsili, T. E. Ouldridge, A. W. Richa, P. Riechers, É. Roldán, B. Rubenstein, Z. Toroczkai, and J. Paradiso, Is stochastic thermodynamics the key to understanding the energy costs of computation? Proc. Natl. Acad. Sci. USA **121**, e2321112121 (2024).

[27] 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).

[28] M. Scandi, D. Barker, S. Lehmann, K. A. Dick, V. F. Maisi, and M. Perarnau-Llobet, Minimally dissipative information erasure in a quantum dot via thermodynamic length, Phys. Rev. Lett. **129**, 270601 (2022).

[29] S. Dago, S. Ciliberto, and L. Bellon, Adiabatic computing for optimal thermodynamic efficiency of information processing, Proc. Natl. Acad. Sci. USA **120**, e2301742120 (2023).

[30] N. Barros, S. Whitelam, S. Ciliberto, and L. Bellon, Learning efficient erasure protocols for an underdamped memory, Phys. Rev. E **111**, 044114 (2025).

[31] Y.-H. Ma, R.-X. Zhai, J. Chen, C. P. Sun, and H. Dong, Experimental test of the $1/\tau$-scaling entropy generation in finite-time thermodynamics, Phys. Rev. Lett. **125**, 210601 (2020).

[32] S. A. M. Loos, S. Monter, F. Ginot, and C. Bechinger, Universal symmetry of optimal control at the microscale, Phys. Rev. X **14**, 021032 (2024).

[33] S. Oikawa, Y. Nakayama, S. Ito, T. Sagawa, and S. Toyabe, Experimentally achieving minimal dissipation via thermodynamically optimal transport, Nat. Commun. **16**, 10424 (2025).

[34] S. Pal, S. Saryal, D. Segal, T. S. Mahesh, and B. K. Agarwalla, Experimental study of the thermodynamic uncertainty relation, Phys. Rev. Res. **2**, 022044(R) (2020).

[35] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, Optimal protocols and optimal transport in stochastic thermodynamics, Phys. Rev. Lett. **106**, 250601 (2011).

[36] 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).

[37] A. Dechant and Y. Sakurai, Thermodynamic interpretation of Wasserstein distance, arXiv:1912.08405.

[38] M. Nakazato and S. Ito, Geometrical aspects of entropy production in stochastic thermodynamics based on Wasserstein distance, Phys. Rev. Res. **3**, 043093 (2021).

[39] K. Proesmans, J. Ehrich, and J. Bechhoefer, Finite-time Landauer principle, Phys. Rev. Lett. **125**, 100602 (2020).

[40] K. Proesmans, J. Ehrich, and J. Bechhoefer, Optimal finite-time bit erasure under full control, Phys. Rev. E **102**, 032105 (2020).

[41] Y. Chen, T. T. Georgiou, and A. Tannenbaum, Stochastic control and nonequilibrium thermodynamics: Fundamental limits, IEEE Trans. Autom. Control **65**, 2979 (2020).

[42] S. Chennakesavalu and G. M. Rotskoff, Unified, geometric framework for nonequilibrium protocol optimization, Phys. Rev. Lett. **130**, 107101 (2023).

[43] T. Van Vu and Y. Hasegawa, Geometrical bounds of the irreversibility in Markovian systems, Phys. Rev. Lett. **126**, 010601 (2021).

[44] A. Dechant, Minimum entropy production, detailed balance and Wasserstein distance for continuous-time Markov processes, J. Phys. A: Math. Theor. **55**, 094001 (2022).

[45] K. Yoshimura, A. Kolchinsky, A. Dechant, and S. Ito, Housekeeping and excess entropy production for general nonlinear dynamics, Phys. Rev. Res. **5**, 013017 (2023).

[46] T. Van Vu and K. Saito, Thermodynamic unification of optimal transport: Thermodynamic uncertainty relation, minimum dissipation, and thermodynamic speed limits, Phys. Rev. X **13**, 011013 (2023).

[47] R. Nagayama, K. Yoshimura, A. Kolchinsky, and S. Ito, Geometric thermodynamics of reaction-diffusion systems: Thermodynamic trade-off relations and optimal transport for pattern formation, Phys. Rev. Res. **7**, 033011 (2025).

[48] T. Sagawa and M. Ueda, Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics, Phys. Rev. Lett. **109**, 180602 (2012).

[49] J. M. Horowitz and M. Esposito, Thermodynamics with continuous information flow, Phys. Rev. X **4**, 031015 (2014).

[50] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[51] M. L. Rosinberg and J. M. Horowitz, Continuous information flow fluctuations, Europhys. Lett. **116**, 10007 (2016).

[52] A. Taghvaei, O. M. Miangolarra, R. Fu, Y. Chen, and T. T. Georgiou, On the relation between information and power in stochastic thermodynamic engines, IEEE Control Syst. Lett. **6**, 434 (2022).

[53] Y. Fujimoto and S. Ito, Game-theoretical approach to minimum entropy productions in information thermodynamics, Phys. Rev. Res. **6**, 013023 (2024).

[54] R. Nagase and T. Sagawa, Thermodynamically optimal information gain in finite-time measurement, Phys. Rev. Res. **6**, 033239 (2024).

[55] R. Nagase and T. Sagawa, Thermodynamic optimization of finite-time feedback protocols for Markov jump systems, Phys. Rev. E **112**, 024118 (2025).

[56] T. Kamijima, K. Funo, and T. Sagawa, Finite-time thermodynamic bounds and trade-off relations for information processing, Phys. Rev. Res. **7**, 013329 (2025).

[57] T. Kamijima, A. Takatsu, K. Funo, and T. Sagawa, Optimal finite-time Maxwell's demons in Langevin systems, Phys. Rev. Res. **7**, 023159 (2025).

[58] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed (Elsevier, Amsterdam, Netherlands, 2007).

[59] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2nd ed (Springer, Berlin, Heidelberg, 1996).

[60] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. **84**, 375 (2000).

[61] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control* (Springer, London, 2013).

[62] H. Kogelnik and T. Li, Laser beams and resonators, Appl. Opt. **5**, 1550 (1966).

[63] J. Klaers, Landauer's erasure principle in a squeezed thermal memory, Phys. Rev. Lett. **122**, 040602 (2019).

[64] J. Roßnagel, O. Abah, F. Schmidt-Kaler, K. Singer, and E. Lutz, Nanoscale heat engine beyond the Carnot limit, Phys. Rev. Lett. **112**, 030602 (2014).

[65] J. Klaers, S. Faelt, A. Imamoglu, and E. Togan, Squeezed thermal reservoirs as a resource for a nanomechanical engine beyond the Carnot limit, Phys. Rev. X **7**, 031044 (2017).

[66] D. Rugar and P. Grütter, Mechanical parametric amplification and thermomechanical noise squeezing, Phys. Rev. Lett. **67**, 699 (1991).

[67] M. Rashid, T. Tufarelli, J. Bateman, J. Vovrosh, D. Hempston, M. S. Kim, and H. Ulbricht, Experimental realization of a thermal squeezed state of levitated optomechanics, Phys. Rev. Lett. **117**, 273601 (2016).

[68] M. Kamba, N. Hara, and K. Aikawa, Quantum squeezing of a levitated nanomechanical oscillator, Science **389**, 1225 (2025).

[69] M. Gelbrich, On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces, Math. Nachr. **147**, 185 (1990).

[70] R. Landauer, Information is physical, Phys. Today **44**(5), 23 (1991).

[71] 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).

[72] Y. Jun, M. Gavrilov, and J. Bechhoefer, High-precision test of Landauer's principle in a feedback trap, Phys. Rev. Lett. **113**, 190601 (2014).

[73] J. Piñero, R. Solé, and A. Kolchinsky, Optimization of nonequilibrium free energy harvesting illustrated on bacteriorhodopsin, Phys. Rev. Res. **6**, 013275 (2024).

[74] A. Kolchinsky, I. Marvian, C. Gokler, Z.-W. Liu, P. Shor, O. Shtanko, K. Thompson, D. Wolpert, and S. Lloyd, Maximizing free energy gain, Entropy **27**, 91 (2025).

[75] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed (Wiley, Hoboken, NJ, 2019).