This is the accepted manuscript made available via CHORUS. The article has been published as:

Nonequilibrium uncertainty principle from information geometry

Schuyler B. Nicholson, Adolfo del Campo, and Jason R. Green

Phys. Rev. E **98**, 032106 — Published 5 September 2018

DOI: 10.1103/PhysRevE.98.032106

Nonequilibrium uncertainty principle from information geometry

Schuyler B. Nicholson,[1] Adolfo del Campo,[2,3] and Jason R. Green[1,2,3,*]

[1] *Department of Chemistry, University of Massachusetts Boston, Boston, MA 02125*
[2] *Department of Physics, University of Massachusetts Boston, Boston, MA 02125*
[3] *Center for Quantum and Nonequilibrium Systems, University of Massachusetts Boston, Boston, MA 02125*

(Dated: August 14, 2018)

With a statistical measure of distance, we derive a classical uncertainty relation for processes traversing nonequilibrium states both transiently and irreversibly. The geometric uncertainty associated with dynamical histories that we define is an upper bound for the entropy production and flow rates, but it does not necessarily correlate with the shortest distance to equilibrium. For slowly driven systems, we show that our uncertainty lower bounds the rate of energy fluctuations. For a model one-bit memory device, we find that expediting the erasure protocol increases the maximum dissipated heat and geometric uncertainty. A driven version of Onsager's three-state model shows that a set of dissipative, high-uncertainty initial conditions, some of which are near equilibrium, scar the state space.

I. INTRODUCTION

Myriad phenomena generate structures and patterns that are unique outside of thermodynamic equilibrium. Efforts to understand these processes stretch back to the very beginnings of thermodynamics – a pinnacle of physics that encapsulates the quantitative understanding of energy transfer and transformations [1]. A powerful approach to studying thermodynamic processes focuses on uncertainty principles [2–4]. Thermal uncertainty relations have strong resemblances to their quantum counterparts and rest on the foundations of equilibrium statistical mechanics. The recent introduction of nonequilibrium uncertainty relations [5, 6] has generated a flurry of activity [7–13], but these results are largely restricted to nonequilibrium steady-states. They leave open the question of whether there are uncertainty relations for processes that are transient and nonstationary. We address this question here.

There are growing links between thermodynamics and information [14–17], some of which place bounds [18–20] on entropy changes [21]. One important example in this context is the erasure of physically-stored information, which dissipates heat and limits the computational power of physical devices [22]. There is still much to be done to disentangle physical and logical irreversibility in order to clarify the processing of information and thermodynamic function [23]. Of particular interest are extending predictions into practically important regimes where erasure is fast and devices are small – when dynamics and statistical fluctuations rule. Progress in this direction requires a firm grasp on the information in the distributions [24] sampled by processes driven transiently away from equilibrium.

For nonstationary processes, it is natural to treat the distributions evolving under certain control parameters through ideas formalized in information geometry [25–29]. There, the focus is on the structure of the manifold of probability distributions, along with the distance and velocity paths traversed by the dynamics of the system. Though often presented in a general setting [25], information geometry has connections to thermodynamics [30–34]. For nonstationary irreversible processes, results are scarce, however, and our understanding of connections between thermodynamics and information geometry remains incomplete. A significant challenge to the development of a statistical-mechanical theory for nonstationary processes is that there are few restrictions on the possible nonequilibrium distributions over paths or states. Here, we establish a fundamental connection between the acceleration of the Shannon entropy and the Fisher information that enables us to bring the mathematical machinery of information geometry to bear on the problem.

II. THEORY

A. Notation and setting

At the ensemble level, a path is the set of probability distributions a system samples as it evolves over a finite time interval. We define the set of probability distributions $\mathbb{P}(\Omega) = \{p : \Omega \to \mathbb{R} \,|\, p_x(t) > 0, \forall x \in \Omega, \sum_x p_x(t) = 1\}$. A subset of these distributions belong to the manifold $\Theta = \{p(x|\theta(t)) : \theta(t) = \{\theta^1(t), \theta^2(t), \ldots, \theta^N(t)\}\}$, where $\theta(t)$ represents the time-dependent control parameters [25] determining the path across the manifold. Empirically, one could sample trajectories through the system state space and construct a distribution at each moment in time from the ensemble of realizations (each distribution being a point on Θ in the large sample limit). Together, these distributions are the "path" in probability space between the initial distribution, $p(t_0)$, and the final distribution, $p(t_f)$, over the time interval $\tau = t_f - t_0$. Here, our main results are independent of the form of the distributions. We only assume the dynamics is governed

* jason.green@umb.edu

by the master equation

$$\dot{p}_x(t) = \sum_y W_{xy}(\theta(t))p_y(t), \tag{1}$$

where $\dot{p}_x(t) = dp_x(t)/dt$ and $W_{xy}(t)$ is the transition rate from state $y \to x$. The occupation probability $p_x(t) = p(x|\theta(t))$ for state x is conditional on the control parameters $\theta(t)$. The rate matrix $\mathbb{W}(t)$ also depends on $\theta(t)$ and follows the usual conventions: for $W_{xy}(t) \in \mathbb{W}(t)$, $W_{xy}(t) > 0$ when $x \neq y$ and $W_{yy}(t) = -\sum_{x\neq y} W_{xy}(t)$ so that $\sum_x W_{xy}(t) = 0$.

A system satisfies detailed balance if the currents or thermodynamic fluxes, $C_{xy}(t) = W_{xy}(t)p_y(t) - W_{yx}(t)p_x(t)$, are zero for all x, y. Otherwise, the existence of current implies the system is undergoing an irreversible process [35]. The current is related to the dynamics through the master equations, $\dot{p}_x(t) = \sum_y C_{xy}(t)$. But, it does not satisfy the requirements of a metric and so cannot be used to quantify the distance from equilibrium. However, it is well known that the Fisher information is a metric [36], providing a notion of distinguishability between neighboring distributions related by the time-evolution of the dynamics. Here, we arrive at the Fisher information and the "geometric uncertainty" accumulated along a path across Θ through the matrix,

$$E_{xy}(t) = W_{xy}(t) - \frac{C_{xy}(t)}{2p_y(t)}. \tag{2}$$

The results that follow are built on the foundation set by the properties of this matrix (Appendix). Even when the current is nonzero, this matrix satisfies a detailed balance condition, $E_{xy}(t)p_y(t) = E_{yx}(t)p_x(t)$. It is similar to a symmetric matrix and, thus, has a complete set of eigenvectors and real eigenvalues [37]. Matrices with a similar form and function are known for discrete-time, discrete-state Markov chains [38, 39] but not for continuous-time Markovian dynamics.

B. Fisher information and thermodynamics

The matrix E allows us to connect the Fisher information (from information geometry) to the entropic acceleration (from thermodynamics). A more common approach uses the Fisher matrix [25],

$$g_{ij} = \sum_x p_x(t) \frac{\partial \ln p_x(t)}{\partial \theta_i} \frac{\partial \ln p_x(t)}{\partial \theta_j}, \tag{3}$$

which is a metric tensor that gives a statistical measure of distance over a manifold of probability distributions, $ds^2 = \sum_{i,j} g_{ij} d\theta_i d\theta_j$. The Fisher information, $I_F(t)$, reflects a change in a probability distribution with respect to a set of control parameters [40]. When parametrized by time it is

$$I_F(t) = \sum_{i,j} \frac{d\theta_i}{dt} g_{ij} \frac{d\theta_j}{dt} = \sum_x p_x(t) \left[\frac{d \ln p_x(t)}{dt}\right]^2. \tag{4}$$

Thus far, the Fisher information is purely a mathematical construction. However, we can relate it to the acceleration of the entropy through the entropy production/flow and make a connection between information and thermodynamics.

Shannon [41] showed that $I_y(t) = -\ln p_y(t)$ is the information associated with state y. The difference $I_{xy}(t) = -\ln p_y(t)/p_x(t)$ is then the local difference in information or relative "surprise" about states y and x. With this context, consider the Shannon entropy rate,

$$\dot{S}(t) \equiv -\sum_x \dot{p}_x(t) \ln p_x(t) = -\langle\langle I(t) \rangle\rangle, \tag{5}$$

which we express as an average, $\langle\langle \cdot \rangle\rangle \equiv \sum_{x,y} W_{xy}(t)p_y(t)[\cdot]$, that is equivalent to an average over the current (up to a factor of 1/2) [42]. The connection to nonequilibrium thermodynamics comes from decomposing the entropy rate at any instant in time, $\dot{S} = \dot{S}^i + \dot{S}^e$, into the entropy production rate from sources in the system, $\dot{S}^i$, and the rate of entropy exchange with the environment, $\dot{S}^e = -\sum_{x,y} W_{xy}(t)p_y(t) \ln W_{xy}(t)/W_{yx}(t)$ [43]. The entropy production $\dot{S}^i = \langle\langle F \rangle\rangle$ is an average of the generalized forces, $F_{xy} = \ln W_{xy}(t)p_y(t) - \ln W_{yx}(t)p_x(t)$ [42], which multiplied by Boltzmann's constant, k_B, are the thermodynamic affinities [35]. Here, we set $k_B = 1$.

The second derivative of the Shannon entropy is the "entropic acceleration",

$$\ddot{S}(t) = -\frac{d}{dt}\langle\langle I(t) \rangle\rangle = -\sum_x \ddot{p}_x(t) \ln p_x(t) - \langle\langle \dot{I}(t) \rangle\rangle. \tag{6}$$

Our first main result is that this acceleration relates to the Fisher information, which for nonstationary irreversible Markovian processes is an average over the rate of information change in the system,

$$I_F(t) = -\sum_{x,y} W_{xy}(t)p_y(t) \frac{d}{dt} \ln \left[\frac{E_{yx}(t)}{E_{xy}(t)}\right] = \langle\langle \dot{I}(t) \rangle\rangle. \tag{7}$$

Combining Eq. (6) and (7), shows that the Fisher information and the entropic acceleration are related:

$$\ddot{S}(t) = -\sum_x \ddot{p}_x(t) \ln p_x(t) - I_F(t) = \ddot{S}^i + \ddot{S}^e. \tag{8}$$

This result can be cast in matrix form with $E(t)$ (Appendix), which can also be expressed in terms of the thermodynamic forces. The entropic acceleration measures the rate at which the bulk information changes in time, the Fisher information is the local rate of information change on average, and the remainder is their sum, $\mathcal{C} = -\sum_x \ddot{p}_x(t) \ln p_x(t) = d\langle\langle I \rangle\rangle/dt + \langle\langle dI/dt \rangle\rangle$.

C. Uncertainty and deviations from the geodesic

Now, by introducing a measure of uncertainty over a path across Θ, these results enable us to show that for any

initial and final distribution, the entropy rate is bounded from above by the contributions from the local and bulk information rates and the geometric uncertainty about the path. Rao showed that the Fisher information matrix satisfies the requirements of a metric [36] and, so, the Fisher information relates to the line element between two distributions infinitesimally displaced from one another, $ds^2 = I_F(t)\,dt^2$. The length, $\mathcal{L}$, of a path on the manifold Θ can then be measured with the statistical distance [44], $\mathcal{L} = \int_{t_0}^{t_f} dt\sqrt{I_F(t)}$. The Cauchy-Schwarz inequality yields the statistical divergence,

$$\mathcal{J} \equiv \tau \int_{t_0}^{t_f} dt\, I_F(t) \geq \mathcal{L}^2. \tag{9}$$

Previous work has shown that $\mathcal{J} - \mathcal{L}^2 \geq 0$ is a temporal variance [27, 28], and that, in one representation, can measure cumulative fluctuations in the rate coefficients for irreversible decay processes [45].

In the current context, the difference between the two terms of the Cauchy-Schwarz inequality equals the variance or "geometric uncertainty" of the path connecting $p(t_0)$ and $p(t_f)$ that measures the cumulative deviations from the geodesic. To illustrate this interpretation, we define the time average for a function, $A(t)$, as $\mathbb{E}[A(t)] = \tau^{-1}\int_{t_0}^{t_f} dt\, A(t)$. The difference between the time average of I_F and the squared time-average of $\sqrt{I_F}$ over the path is the time-averaged variance,

$$\sigma^2 = \frac{\mathcal{J} - \mathcal{L}^2}{\tau^2} = \mathbb{E}\left[I_F\right] - \mathbb{E}\left[\sqrt{I_F}\right]^2 \geq 0. \tag{10}$$

This geometric uncertainty is the cumulative deviation from the geodesic connecting the initial and final distributions. It depends on the path and the initial and final distributions. We expect it to be nonzero for most irreversible processes. One notable exception are paths following the geodesic connecting two distributions. These paths correspond to the condition $\mathcal{J} = \mathcal{L}^2$ [44] and a variance of zero. These "certain" paths are irreversible, nonstationary paths with zero geometric uncertainty.

It has previously been shown that measuring cumulative deviations from the geodesic amounts to measuring the cumulative fluctuations in nonequilibrium observables [33, 45]. Past work has also used statistical distances (though with other metrics) to measure the dissipation associated with quasistatic transformations [46]. These results, however, do not connect thermodynamic quantities such as the entropic acceleration to the Fisher information for general nonstationary irreversible processes as we do here.

Our second main result is a bound on the entropy rate by the geometric uncertainty. It follows from recognizing that the variance satisfies the inequality:

$$\sigma^2 \leq \frac{1}{\tau}\int_{t_0}^{t_f} dt\left(I_F + \mathbb{E}\left[\sqrt{I_F}\right]^2\right) \leq \frac{2\mathcal{J}}{\tau^2}. \tag{11}$$

Where we use, $\mathcal{J} - \mathcal{L}^2 \geq 0$ and the nonnegativity of the variance. Defining $\mathcal{I} \equiv \tau^{-1}\int_{t_0}^{t_f} dt\,\langle\langle \dot{I}\rangle\rangle$, this relation becomes

$$\mathcal{I}\,\sigma^{-2} \geq \frac{1}{2}. \tag{12}$$

The intuition behind this uncertainty relation is that different paths across the manifold of probability distributions Θ can lower the time-averaged rate of information change $\mathcal{I}$, but only at the expense of a corresponding decrease in uncertainty (a smaller excursion from the geodesic). Simply put, the uncertainty places a bound on the cumulative rate of information change. It is worth noting that this information-uncertainty ratio is valid for nonstationary, irreversible paths over any finite time interval between arbitrary probability distributions. To test this inequality, one only needs the basic ingredients of a Markov state model, models that have proven useful for discovering collective variables and analyzing rare events in diverse areas, including protein (un)folding [47, 48].

Another way to write the uncertainty relation is in terms of the entropic acceleration, Eq. (8). Upon integrating, it becomes a bound on the entropy rate

$$\Delta\dot{S} \leq \mathcal{C} - \frac{\sigma^2\tau}{2}, \tag{13}$$

where, again, $\mathcal{C} = -\int_{t_0}^{t_f} dt\sum_x \ddot{p}_x(t)\ln p_x(t)$ and $\Delta\dot{S} = \dot{S}(t_f) - \dot{S}(t_0)$. The quantities $\mathcal{C}$ and $\sigma^2\tau/2$ can both be zero only in a stationary state – that is, this uncertainty relation applies specifically to the nonstationary regime – and are measurable from occupation/transition probabilities. A direct connection to previous stationary uncertainty relations appears to be a subtle question. However, the present result has a clear physical meaning: the entropy rate (from thermodynamics) is bounded by contributions from the local and bulk information dynamics and cumulative deviations from the geodesic (from information geometry). When no heat is exchanged, these information dynamics bound the entropy production. And, when there are no internal sources of entropy production, they bound the entropy (heat) flow.

The geometric uncertainty measures deviations from the geodesic. Along the geodesic, the average local rate of information change is constant and the path is certain:

$$\mathcal{I} = \langle\langle \dot{I}\rangle\rangle = \langle\langle \ddot{S}^i\rangle\rangle + \langle\langle \ddot{S}^e\rangle\rangle \qquad \text{(geodesic)}. \tag{14}$$

In other words, the average local entropic acceleration is constant, independent of time and distance from the final distribution.

D. Special cases

The results so far avoid any assumptions about the probability distributions, rate of driving, or "closeness" to equilibrium. If the system is undergoing a process at (nonequilibrium) steady-state, $\dot{p}_x = 0$, the Fisher

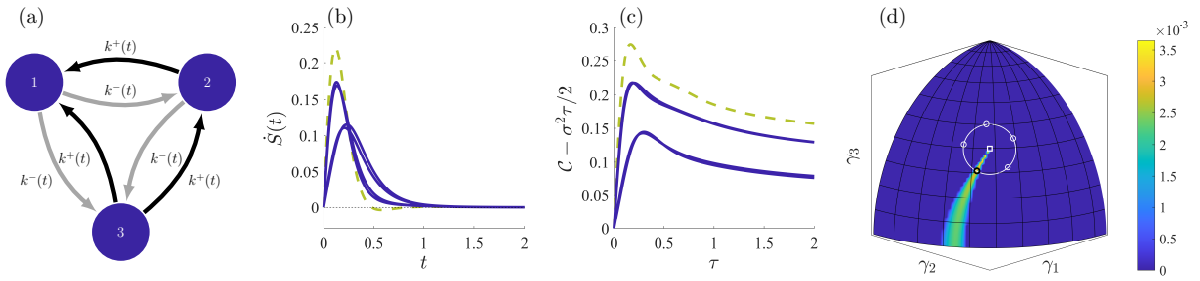

FIG. 1. (a) Kinetic scheme for an asymmetric three-state model. Time-dependent transitions rates are $k^+(t) = 4\,\mathrm{atan}(\omega_1 t)$ and $k(t)^- = 4\,\mathrm{atan}(\omega_2 t)$ with $\omega_1 = 4$ and $\omega_2 = 6$. Using $\mathrm{atan}(\omega_i t)$ ensures all initial conditions evolve to the stationary distribution $p_x^\infty = (1/3, 1/3, 1/3)$. Each initial condition in (b) and (c) is marked by a circle in (d), which shows the positive octant of a sphere colored by $\dot{S}(t = t_f)$. Here, t_f is the initial time for each path to reach the stationary distribution (white square). At $t = 0$, $\dot{S}(t_0) = 0$. The scar on the state space in (d) emerges over time from initial conditions that ultimately dissipate and have higher path uncertainties. One initial condition is highlighted with a circle and corresponds to the dashed lines in (b-c). (c) The upper bound for $\dot{S}(t)$, $\mathcal{C} - \sigma^2\tau/2$, for the five select initial conditions as a function of the time interval $\tau = t - t_0$.

information is $I_F = 0$. As a result, there is zero path uncertainty $\sigma^2 = \mathcal{I} = 0$ and the local entropy production rate and entropy flow rate are exactly in balance: $\langle\langle \ddot{S}^i \rangle\rangle = \langle\langle \dot{F} \rangle\rangle = -\langle\langle \ddot{S}^e \rangle\rangle$.

While the results in the previous section are independent of the form of the probability distributions, the special case of an exponential energy distribution gives further physical insight. Of particular interest is a single-component, homogeneous, closed system at thermal equilibrium with a reservoir at inverse temperature $\beta = 1/k_B T$:

$$p_x(t) = Z^{-1} e^{-\beta \epsilon_x(t)}. \tag{15}$$

in which the energy of each state of the system is driven slowly. Under similar conditions, Crooks [49] showed that for the canonical ensemble, the Fisher information is equal to the infinitesimal change in energy with respect to a change in the control parameter,

$$I_F(\theta) = \beta^2 \left\langle \left(\frac{\partial \epsilon}{\partial \theta} - \left\langle \frac{\partial \epsilon}{\partial \theta} \right\rangle \right)^2 \right\rangle. \tag{16}$$

If $\theta = \beta$ is the control parameter, this expression becomes the variance measuring energy fluctuations around equilibrium. Here, we derive a similar result. The key differences is that we fix β and explicitly incorporate time. In this case, I_F measures the fluctuations in the energy rates.

Given the time-dependent probability distribution above, the rate of change is:

$$\dot{p}_x(t) = -\beta \dot{\epsilon}_x(t) p_x(t) + \beta p_x(t) \sum_x \dot{\epsilon}_x(t) p_x(t). \tag{17}$$

Writing the ensemble average as $\sum_x p_x(t)[\cdot] = \langle \cdot \rangle$, gives

$$\frac{d \ln p_x(t)}{dt} = \beta \langle \dot{\epsilon}(t) \rangle - \beta \dot{\epsilon}_x(t) \tag{18}$$

and

$$I_F(t) = \sum_x p_x(t) \left(\frac{d \ln p_x(t)}{dt} \right)^2 = \beta^2 \mathrm{Var}[\dot{\epsilon}(t)]. \tag{19}$$

Here, the energy is playing the role of our control parameter, and β is held fixed, so the Fisher information measures the fluctuations in the energy rate. In this special case, the uncertainty relation, Eq. (12), becomes

$$\frac{\beta^2}{\tau} \int_{t_0}^{t_f} dt\, \mathrm{Var}[\dot{\epsilon}(t)] \geq \frac{\sigma^2}{2}. \tag{20}$$

The uncertainty over the path lower bounds time-averaged fluctuations of the energy rates. The geodesic is a path traversed when systems have no fluctuations in the energy rate. All other paths will have a positive variance. During a nonstationary process operating near this bound, lowering the time-average fluctuations in energy flux will mean smaller excursions from the geodesic where these fluctuations and $I_F(t)$ are constant: $I_F(t) = \beta^2 \mathrm{Var}\,[\dot{\epsilon}(t)] \geq 0$.

III. APPLICATIONS

A. Uncertainty scarring in a single-cycle chemical reaction

To illustrate these results, we adapt the kinetic scheme Fig. (1a), used by Onsager to demonstrate the reciprocal relations of irreversible thermodynamics [50]. The model consists of three states and a kinetics driven by the time-dependent rate coefficients, $k^+(t) = 4\,\mathrm{atan}(\omega_1 t)$ and $k^-(t) = 4\,\mathrm{atan}(\omega_2 t)$, with $\omega_1 = 4$ and $\omega_2 = 6$. The inverse tangent function ensures that for large t, every path reaches the same stationary distribution, $p_x^\infty = (1/3, 1/3, 1/3)$. Our criterion for a path to reach the stationary distribution is that each initial condition must evolve to be within $\|p_x(t) - p_x^\infty\|_2 \leq 5 \times 10^{-3}$ of the stationary distribution. Under the transformation $\gamma_x(t) \equiv \sqrt{p_x(t)}$, the system travels across the positive octant of a sphere, Fig. (1d).

This system and driving protocol localize the effects of the initial condition on the geometric uncertainty about

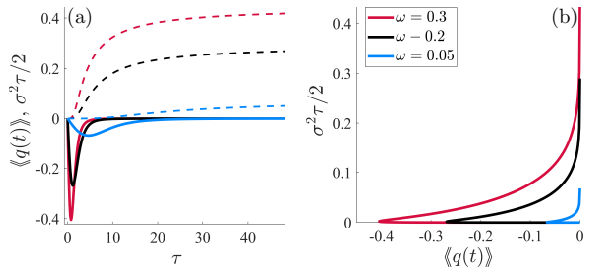

FIG. 2. Accelerating the erasure of information (destroying state $x = 0$), by increasing ω, generates more (a) uncertainty ($\sigma^2\tau/2$, dashed lines) about the erasure path and maximum heat dissipated ($\langle\langle q(t)\rangle\rangle$, solid lines). (b) The heat dissipated, $\langle\langle q(t)\rangle\rangle$, follows the geodesic ($\sigma^2 \approx 0$) up until the maximum amount of heat is being released. Beyond that point, the system undergoes entropic deceleration. Time series are from $p(t_0) = [0.5, 0.5]$ to $p(t_f) = [1, 0]$ for $\omega = [0.05, 0.2, 0.3]$ shown in blue, black, and red lines, respectively.

the nonequilibrium path. What we find is that the distance from the stationary state says little about the uncertainty or the entropy rate (dissipation rate). Figure (1d) shows $\Delta\dot{S} = \dot{S}(t)$ for all physically-relevant initial conditions (color indicates final $\dot{S}$). Five initial conditions are marked (open circles), each point equidistant from the equilibrium state, p_x^∞ (white square). While these initial conditions are all equally "far" from equilibrium, their entropy rates exhibit different behavior over time. Color in Figure (1b) corresponds to the initial conditions highlighted in (d). The dashed line represents initial condition originating from the scar. This initial condition, like all others originating from the scar, has a larger maximum value of $\dot{S}(t)$ than those equidistant from p_x^∞, but launched from off it. These scar-initial conditions are also unique in that they all dissipate for a period of time. Initial conditions off the scar do not dissipate. Regardless of the dissipative nature of these paths, the uncertainty relation holds. The time dependence of the upper bound is shown in Fig. (1c) for the same five initial conditions. Again, those initial conditions originating in the scar have a larger uncertainty and upper bound the entropy rate. Overall, these results are evidence that the distance from the stationary state can be a poor predictor of transient nonequilibrium behavior.

B. Landauer's principle and information erasure

Perhaps nowhere is the connection between information and thermodynamics more apparent than in Landauer's principle [51]. According to this principle, erasing one bit of information requires the dissipation of at least $k_B T \ln 2$ thermal energy as heat. Since the bound on $\Delta\dot{S}(t)$ in Eq. (13), holds for any Markovian evolution between any two distributions, we can explore the connection between the entropy (heat) dissipation and the uncertainty about the erasure path. To examine correlation between erasure paths and the heat release, we consider the erasure of one-bit of information in a model memory device.

The model 1-bit memory device initially consists of two states $x = \{1, 0\}$ that are equally probable, $p(t_0) = [0.5, 0.5]$. To measure the heat release, we choose $W(t)$ such that every distribution is of the form $p_x(t) = Z(t)^{-1} e^{-\beta\epsilon_x(t)}$, where $Z(t)$ is the partition function, β is the inverse temperature set to one, and $\epsilon_x(t)$ is the energy at state x.

To ensure that our dynamics generates the prescribed sequence of distributions, we work backwards. Given $p_x(t)$ for $t_0 \leq t \leq t_f$, we estimate $\dot{p}_x(t) = [p_x(t + \delta t) - p_x(t)]/\delta t$, where $\delta t = 1 \times 10^{-4}$ was used for all simulations. To find the $W(t)$ that satisfies $W(t)\,|p(t)\rangle = |\dot{p}(t)\rangle$, we solve the system of linear equations using conservation of probability, $W_{11}(t) = -W_{21}(t)$, $W_{22}(t) = -W_{12}(t)$. There is an infinite family of $W(t)$ that satisfies these conditions given by

$$W(t) = \begin{bmatrix} -A & \frac{\dot{p}_1(t) + A p_1(t)}{p_2(t)} \\ A & -\frac{\dot{p}_1(t) + A p_1(t)}{p_2(t)} \end{bmatrix}. \tag{21}$$

We set A, which is arbitrary, to one. To calculate the second inequality in the main text, Eq. (13), we need to evaluate $-\sum_x \ddot{p}_x(t) \ln p_x(t)$. Given $p_x(t) = Z(t)^{-1} e^{-\beta\epsilon_x(t)}$, the first derivative is $\dot{p}_x(t) = \beta p_x(t)\,(\dot{\epsilon}_2(t) p_2(t) - \dot{\epsilon}_x(t))$. For the energies, we hold $\epsilon_1(t) = \epsilon_2(t_0)$ constant, making our initial distribution $p_x(t_0) = [0.5, 0.5]$, and vary $\epsilon_2(t) = c_1 + c_2 \pi^{-1} \operatorname{atan}(\omega t)$, where we use $c_1 = 0.2$, and $c_2 = 20$. The first derivative of $\epsilon_2(t)$ is $\dot{\epsilon}_2(t) = 20\omega\pi^{-1}(1 + \omega^2 t^2)^{-1}$.

The parameter ω controls the rate at $k_B T \ln 2$ of energy that is dissipated. This restriction on $p_x(t)$ means that the entropy rate, $\dot{S}(t) = -\sum_{x,y} W_{xy} p_y (\epsilon_x(t) - \epsilon_y(t)) = \langle\langle q(t)\rangle\rangle$, is the average heat exchange between the system and surroundings at an instant in time. From Eq. (13), we know that $\langle\langle q(t)\rangle\rangle \leq \mathcal{C} - \sigma^2\tau/2$. For this system, both $\mathcal{C}$ and $\sigma^2\tau/2$ are positive and $\langle\langle q(t)\rangle\rangle \leq 0$, as expected. The higher the rate at which heat is dissipated, the larger the erasure path uncertainty, Fig. (2a). The uncertainty does not increase at a constant rate during the erasure protocol. As energy is initially dissipated, up to the maximum value, the system approximately follows the geodesic across Θ and $\sigma^2 \approx 0$. Figure (2b) shows $\sigma^2\tau/2$ is near zero until the system reaches a state of maximal dissipation, after which the path moves off the geodesic. The faster physically-stored information is erased, the faster energy is dissipated and the greater the resulting uncertainty about the path to equilibrium.

IV. CONCLUSIONS

For processes arbitrarily far from equilibrium, we have established a bound on the entropy production and flow rates via the uncertainty in the path connecting any two arbitrary distributions. This uncertainty relation holds

when the system evolves under a time-inhomogeneous Markovian dynamics, making it applicable to a broad class of nonequilibrium processes. It is clear that even for the classical single-cycle system, the proximity to the stationary state is a poor indicator of uncertainty: initial conditions that are statistically equidistant from the stationary state can have dramatically different geometric uncertainties, uncertainties that we showed are linked to the entropy rate. When erasing information in a model one-bit memory device, we find that increasing the speed of erasure comes at the expense of increasing the rate of energy dissipation and the geometric uncertainty about the path to equilibrium. For the special case of exponential distributions, the fluctuations in the energy rate are bounded by the geometric uncertainty of the path. This result suggests that information geometry and stochastic thermodynamics are intimately linked. We expect these results to be usefully applied to other kinetic phenomena, such as (bio)chemical reactions [52, 53] and to further expand the understanding of processes away from equilibrium, both near and far.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant number W911NF-14-1-0359 and the John Templeton Foundation. S. B. N. acknowledges financial support from the Office of Global Programs, University of Massachusetts Boston. We thank Sosuke Ito for pointing out the relation between the generalized forces and the matrix E and Tamiki Komatsuzaki for comments on the manuscript.

Appendix: Properties of the matrix, E

In this section, we will derive important properties of the matrix,

$$E_{xy}(t) = W_{xy}(t) - \frac{C_{xy}(t)}{2p_y(t)}, \tag{A.1}$$

that underlie our results. We will use bra and ket notation $\langle\cdot|\cdot\rangle$, so that given the $N \times N$ matrix, E, we have $[E\,|p\rangle]_x = \sum_y E_{xy}p_y$ and $[\langle p|\,E]_y = \sum_x p_x E_{xy}$.

1. Master equation

The master equation can be recast in terms of the matrix $E(t)$. Multiplying $E(t)$ by the probability $p_y(t)$ and summing gives,

$$\begin{aligned}
\sum_y E_{xy}(t)p_y(t) &= \sum_y \left(W_{xy}(t) - \frac{C_{xy}(t)}{2p_y(t)}\right) p_y(t), \\
\sum_y E_{xy}(t)p_y(t) &= \dot{p}_x(t) - \frac{\dot{p}_x(t)}{2}, \\
2\sum_y E_{xy}(t)p_y(t) &= \dot{p}_x(t) = \sum_y W_{xy}(t)p_y(t).
\end{aligned} \tag{A.2}$$

The final line shows that $E(t)$ is an alternative representation of the dynamics governed by the master equation.

We should note that $E(t)$ is not a trivial transformation of $W(t)$. For example, $2E(t) \neq W(t)$. Instead, multiplying $E(t)$ by two, shows that if $C_{xy}(t) = 0\ \forall\, x, y$ then $E(t) = W(t)$. In general, however, $2E_{xy}(t) = W_{xy}(t) + \hat{W}_{xy}(t)$, where $\hat{W}_{xy}(t)$ is the time-reversed rates defined [54] by

$$\begin{aligned}
\hat{W}_{xy}(t)p_y(t) &= W_{yx}(t)p_x(t), \\
\hat{W}_{xy}(t) &= \frac{1}{p_y(t)} W_{yx}(t)p_x(t).
\end{aligned} \tag{A.3}$$

The matrix $\hat{W}(t)$ defines the microscopically reversible dynamics of $W_{xy}(t)$, which do not satisfy detailed balance in general. In fact, only when $W(t)$ satisfies detailed balance, does $W_{xy}(t)p_y(t) = \hat{W}_{xy}(t)p_y(t)$.

2. Detailed-balance condition

The rate matrix has the properties that $W_{xy}(t) \in \mathbb{W}(t)$, $W_{xy}(t) > 0$ for $x \neq y$ and $W_{yy}(t) = -\sum_{x\neq y} W_{xy}(t)$ so that $\sum_x W_{xy}(t) = 0$. Using the last property, the master equation becomes

$$\begin{aligned}
\dot{p}_x(t) &= \sum_{y\neq x} \left[W_{xy}(t)p_y(t) - W_{yx}(t)p_x(t)\right] \\
&= \sum_y C_{xy}(t).
\end{aligned} \tag{A.4}$$

The master equation exhibits detailed balance if each of the currents vanish; that is, when the currents or thermodynamic fluxes, $C_{xy}(t) = W_{xy}(t)p_y(t) - W_{yx}(t)p_x(t)$, are zero for all x, y. Otherwise, the existence of current implies the system is undergoing an irreversible process [35].

Even for processes that are driven or transiently away from equilibrium, $C_{xy}(t) \neq 0$, the matrix $E(t)$ satisfies a similar detailed balance condition. Since the master equation can be recast in terms of $E(t)$, it also has an analogous form with source and sink terms. Applying the definition of $E(t)$ to the master equation gives

$$\begin{aligned}
\dot{p}_x(t) &= \sum_{y\neq x} \left[E_{xy}(t)p_y(t) - E_{yx}(t)p_x(t) + C_{xy}(t)\right] \\
&= \sum_y \left[C^E_{xy}(t) + C_{xy}(t)\right].
\end{aligned} \tag{A.5}$$

In the final line, we define the current (for $E(t)$) between states x and y, $C^E_{xy}(t)$. Comparing this result to Eq. (A.4) suggests $C^E_{xy}(t) = 0$, akin to detailed balance, but valid when $C_{xy}(t) \neq 0$. To prove this detailed balance condition, we can expand the current for $E(t)$:

$$\begin{aligned} C^E_{xy}(t) &\equiv E_{xy}(t)p_y(t) - E_{yx}(t)p_x(t) \\ &= W_{xy}(t)p_y(t) - \frac{C_{xy}(t)}{2} - W_{yx}(t)p_x(t) + \frac{C_{yx}(t)}{2} \\ &= \frac{C_{xy}(t)}{2} + \frac{C_{yx}(t)}{2} \\ &= 0. \end{aligned} \tag{A.6}$$

The last equality follows from the anti-symmetry of the current, $C_{xy}(t) = -C_{yx}(t)$. When detailed balance is satisfied for $C_{xy}(t)$ it is also satisfied for $C^E_{xy}(t)$: the condition for both is $W_{xy}(t)p_y(t) = W_{yx}(t)p_x(t)$.

3. Symmetrization

As is done at equilibrium with $W(t)$, we can show that $E(t)$ is similar to a symmetric matrix, $S(t)$: $E(t) \backsim S(t)$. First, we define

$$S_{xy}(t) = \frac{1}{\sqrt{p_x(t)}} E_{xy}(t) \sqrt{p_y(t)}. \tag{A.7}$$

To show $S(t)$ is symmetric, we use $S(t)$ and the detailed balance of $E(t)$:

$$\begin{aligned} E_{xy}(t)p_y(t) &= E_{yx}(t)p_x(t), \\ \sqrt{p_x(t)}S_{xy}(t)\sqrt{p_y(t)} &= \sqrt{p_y(t)}S_{yx}(t)\sqrt{p_x(t)}, \\ S_{yx}(t) &= S_{xy}(t). \end{aligned} \tag{A.8}$$

Since $S(t)$ is a real, symmetric (Hermitian) matrix, it has a complete set of eigenvectors and real eigenvalues [37]. Since $E(t)$ is similar to $S(t)$, it also has a complete set of eigenvectors and real eigenvalues. We note that the eigenvectors and eigenvalues of both matrices are time dependent.

The E-representation of the master equation, Eq. (A.2), does not imply that $W(t)$ is similar to $S(t)$. From the definitions of $E(t)$ and $S(t)$, Eq. (A.1) and Eq. (A.7), we know

$$E_{xy}(t) = \sqrt{p_x(t)} S_{xy}(t) \frac{1}{\sqrt{p_y(t)}}.$$

Re-writing in terms of $S(t)$ gives:

$$S_{xy}(t) = \frac{1}{\sqrt{p_x(t)}} W_{xy}(t) \sqrt{p_y(t)} - \frac{C_{xy}(t)}{2\sqrt{p_x(t)p_y(t)}}.$$

By inspection, $W(t)$ is only similar to $S(t)$ when $C_{xy}(t) = 0 \; \forall \; x, y$ and $E(t) = W(t)$. It can be further shown that just because $E(t) \backsim S(t)$ through $p_x(t)$, this does not imply $\exists \; S'(t) \backsim W(t)$, where $S'_{xy}(t) = S'_{yx}(t)$. Defining $S'(t)$ as

$$S'_{xy}(t) = \frac{1}{\sqrt{p_x(t)}} W_{xy}(t) \sqrt{p_y(t)}.$$

and using the expression for the current,

$$\begin{aligned} C_{xy}(t) &= W_{xy}(t)p_y(t) - W_{yx}(t)p_x(t) \\ &= \sqrt{p_x(t)}S'_{xy}(t)\sqrt{p_x(t)} - \sqrt{p_y(t)}S'_{yx}(t)\sqrt{p_x(t)}, \end{aligned}$$

yields

$$S'_{xy}(t) = \frac{C_{xy}(t)}{\sqrt{p_x(t)p_y(t)}} + S'_{yx}(t).$$

Thus, $S'(t)$ cannot be symmetric unless detailed balance is satisfied, $C(t) = 0$, or S' is the zero matrix. It is the nonstationary, irreversibility of the system that prevents $W(t)$ from satisfying a similarity transform – irreversibility is built into $E(t)$.

4. Surprisal rate

Shannon [41] identified the information gained or surprise in observing the state y as $-\ln p_y(t)$ [24]. Using $\sum_x E_{xy}(t) = [\langle 1| E(t)]_y$, where $\langle 1|$ is a row vector of ones, the surprisal rate is related to E by

$$\begin{aligned} [\langle 1| E(t)]_y &= \sum_x \left[W_{xy}(t) - \frac{C_{xy}(t)}{2p_y(t)} \right] \\ &= \frac{\dot{p}_y(t)}{2p_y(t)} = \frac{1}{2}\frac{d \ln p_y(t)}{dt}. \end{aligned} \tag{A.9}$$

This relationship also implies, through conservation of probability, that $\langle 2| E(t) |p(t)\rangle = \sum_y \dot{p}_y(t) = 0$.

5. Fisher information

Underlying the principal results of the main text is that $E(t)$ is related to the Fisher information, $I_F(t)$, through $I_F(t) = -\langle 2| \dot{E}(t) |p(t)\rangle = \langle 2| E(t) |\dot{p}(t)\rangle$. Differentiating $\langle 2| E(t) |p(t)\rangle$ with respect to time gives

$$\frac{d}{dt} \langle 2| E(t) |p(t)\rangle = \langle 2| E(t) |\dot{p}(t)\rangle + \langle 2| \dot{E}(t) |p(t)\rangle . \tag{A.10}$$

The first term on the right-hand side is the Fisher information,

$$\begin{aligned} \langle 2| E(t) |\dot{p}(t)\rangle &= 2\sum_{x,y} \left(W_{xy}(t)\dot{p}_y(t) - \frac{C_{xy}(t)\dot{p}_y(t)}{2p_y(t)} \right) \\ &= \sum_{x,y} \frac{1}{p_y(t)} C_{yx}(t)\dot{p}_y(t) \\ &= \sum_y \frac{\dot{p}_y(t)^2}{p_y(t)} \\ &= I_F(t). \end{aligned} \tag{A.11}$$

The second term is the negative of the Fisher information:

$$\begin{aligned}\langle 2|\,\dot{E}(t)\,|p(t)\rangle &= 2\sum_{xy}\left(\dot{W}_{xy}(t)p_y(t)\right.\\ &\qquad\left.+\frac{C_{xy}(t)\dot{p}_y(t)}{2p_y(t)}-\frac{\dot{C}_{xy}(t)}{2}\right)\\ &=\sum_{x,y}\frac{C_{xy}(t)\dot{p}_y(t)}{p_y(t)}\\ &=-\sum_{x,y}\frac{C_{yx}(t)\dot{p}_y(t)}{p_y(t)}\\ &=-I_F(t). \end{aligned}\tag{A.12}$$

From the first to the second line, we use conservation of probability, $d/dt\sum_x W_{xy}(t)=0$, and $d/dt\sum_{xy}C_{xy}(t)=0$. Plugging Eq. (A.11) and Eq. (A.12) into Eq. (A.10), we see that the conservation of probability leads to $\frac{d}{dt}\langle 2|\,E(t)\,|p(t)\rangle = I_F(t)-I_F(t)=0$.

6. Fisher information as entropic acceleration

With the properties discussed so far, we can arrive at the first main result: for systems with dynamics that are governed by continuous-time master equations, the Fisher information is part of the entropic acceleration. To show this, we recognize that the matrix $E(t)$ in Property IV can be expressed in terms of the generalized thermodynamic forces,

$$\begin{aligned}[\langle 2|\,E(t)\,]_y &= \frac{1}{p_y(t)}\sum_x C_{yx}(t)\\ &=\frac{1}{p_y(t)}\sum_x W_{yx}(t)p_x(t)-W_{xy}(t)p_y(t)\\ &=\frac{1}{p_y(t)}\sum_x W_{xy}(t)p_y(t)\left(\frac{W_{yx}(t)p_x(t)}{W_{xy}(t)p_y(t)}-1\right)\\ &=\sum_x W_{xy}(t)e^{-F_{xy}(t)}, \end{aligned}\tag{A.13}$$

where $F_{xy}(t)=\ln W_{xy}(t)p_y(t)-\ln W_{yx}(t)p_x(t)$. In Property V, we proved that $I_F=-\langle 2|\,\dot{E}(t)\,|p(t)\rangle$. Using this relation, together with the definition of the thermodynamic forces, shows the Fisher information is given by

$$\begin{aligned}I_F(t) &= -\langle 2|\,\dot{E}(t)\,|p(t)\rangle\\ &=-\sum_{x,y}\left[\frac{d}{dt}\left(W_{xy}(t)e^{-F_{xy}(t)}\right)\right]p_y(t)\\ &=-\sum_{x,y}\left(\dot{W}_{xy}(t)e^{-F_{xy}(t)}p_y(t)\right.\\ &\qquad\left.-W_{xy}(t)\dot{F}_{xy}(t)e^{-F_{xy}(t)}p_y(t)\right)\\ &=-\sum_{x,y}W_{xy}(t)p_y(t)\frac{d}{dt}\ln\left(\frac{p_y(t)}{p_x(t)}\right)\\ &=\langle\!\langle\dot{I}(t)\rangle\!\rangle. \end{aligned}\tag{A.14}$$

Since $-\ln p_y(t)$ is the surprisal of state y at time t, the quantity $\langle\!\langle\dot{I}(t)\rangle\!\rangle$ is the time rate of change in the surprisal difference (between y and x).

To connect $I_F(t)$ to the entropic acceleration, we differentiate the entropy rate,

$$\dot{S}(t)=\sum_{x,y}W_{xy}(t)p_y(t)\ln\left[\frac{p_y(t)}{p_x(t)}\right],\tag{A.15}$$

with respect to time:

$$\begin{aligned}\ddot{S}(t) &= -\sum_{x,y}\frac{d}{dt}\left[W_{xy}(t)p_y(t)\right]I_{xy}-\sum_{x,y}W_{xy}(t)p_y(t)\dot{I}_{xy}(t)\\ &=\sum_{x,y}\frac{d}{dt}\left[W_{xy}(t)p_y(t)\right]I_{xy}(t)-I_F(t)\\ &=-\sum_x\ddot{p}_x(t)\ln p_x(t)-I_F(t), \end{aligned}\tag{A.16}$$

where $I_{xy}=-\ln p_y(t)+\ln p_x(t)$. From Eq. (A.14), the second term in the entropic acceleration is minus the Fisher information. The entropy rate $\dot{S}(t)$ is an average over the change in information $I(t)$, so $\ddot{S}(t)$ can be thought of as the rate of change of the bulk, or average information, $\ddot{S}(t)=d\langle\!\langle I(t)\rangle\!\rangle/dt$. The Fisher information, though, is an average over the rate of change in information between each set of states x and y, $I_F=\langle\!\langle\dot{I}(t)\rangle\!\rangle$. Therefore, the first term on the right hand side of Eq. (A.16) is the sum of the bulk information rate and the average local information rate,

$$\begin{aligned}-\sum_x\ddot{p}_x(t)\ln p_x(t) &= \ddot{S}+I_F\\ &=\frac{d\langle\!\langle I(t)\rangle\!\rangle}{dt}+\left\langle\!\!\left\langle\frac{dI(t)}{dt}\right\rangle\!\!\right\rangle. \end{aligned}\tag{A.17}$$

[1] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed. (John Wiley & Sons, Inc., 1985).

[2] B. Mandelbrot, IRE Transactions on Information Theory **2**, 190 (1956).

[3] F. Schlögl, J. Phys. Chem. Solids **49**, 679 (1988).

[4] J. Uffink and J. van Lith, Found. Phys. **29**, 655 (1999).
[5] A. C. Barato and U. Seifert, Phys. Rev. Lett. **114**, 158101 (2015).
[6] T. R. Gingrich, J. Horowitz, N. Perunov, and J. England, Phys. Rev. Lett. **116**, 120601 (2016).
[7] C. Maes, Physical review letters **119**, 160601 (2017).
[8] J. M. Horowitz and T. R. Gingrich, Phys. Rev. E **96**, 020103 (2017).
[9] P. Pietzonka, F. Ritort, and U. Seifert, arXiv:1702.07699 (2017).
[10] N. Shiraishi, K. Saito, and H. Tasaki, Phys. Rev. Lett. **117**, 190601 (2016).
[11] N. Shiraishi, arXiv:1706.00892 (2017).
[12] K. Proesmans and C. Van den Broeck, EPL **119**, 20001 (2017).
[13] A. Dechant and S. Sasa, arXiv:1708.08653 (2017).
[14] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[15] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[16] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[17] D. J. A. E. Allahverdyan and G. Mahler, J. Stat. Mech. **2009**, P09011 (2009).
[18] D. Hartich, A. Barato, and U. Seifert, J. Stat. Mech. **2014**, P02016 (2014).
[19] J. M. Horowitz and M. Esposito, Phys. Rev. X **4**, 031015 (2014).
[20] S. Yamamoto, S. Ito, N. Shiraishi, and T. Sagawa, Phys. Rev. E **94**, 052121 (2016).
[21] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Nat. Phys. **11**, 131 (2015).
[22] S. Lloyd, Nature **406**, 1047 (2000).
[23] A. B. Boyd and D. M. J. P. Crutchfield, New J. Phys. **18**, 023049 (2016).
[24] T. M. Cover and J. A. Thomas, *Elements of Information theory*, 2nd ed. (Wiley, 2006).
[25] S. Amari and H. Nagaoka, *Methods of information geometry*, Vol. 191 (American Mathematical Soc., 2007).
[26] D. Brody and N. Rivier, Phys. Rev. E **51**, 1006 (1995).
[27] J. Heseltine and E. Kim, J. Phys. A **49**, 175002 (2016).
[28] S. B. Nicholson and E. Kim, Entropy **18**, 258 (2016).
[29] M. Oizumi, N. Tsuchiya, and S. Amari, Proc. Natl. Acad. Sci. U.S.A. **113**, 14817 (2016).
[30] F. Weinhold, J. Chem. Phys. **63**, 2479 (1975).
[31] G. E. Crooks, Phys. Rev. Lett. **99**, 100602 (2007).
[32] G. Ruppeiner, Phys. Rev. A. **20**, 1608 (1979).
[33] D. A. Sivak and G. E. Crooks, Phys. Rev. Lett. **108**, 190602 (2012).
[34] S. Lahiri, J. Sohl-Dickstein, and S. Ganguli, arXiv:1603.07758 (2016).
[35] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
[36] R. C. Rao, J. Royal Statistics Society **B10**, 159 (1945).
[37] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge university press, 2012).
[38] S. B. Nicholson, L. S. Schulman, and E. Kim, Phys. Lett. A **377**, 1810 (2013).
[39] S. B. Nicholson and E. Kim, Physica Scripta **91**, 044006 (2016).
[40] B. R. Frieden, *Science from Fisher Information*, Vol. 2 (Cambridge University Press, 2004).
[41] C. E. Shannon, The Bell System Technical Journal **27**, 623 (1948).
[42] M. Esposito and C. Van den Broeck, Phys. Rev. E **82**, 011143 (2010).
[43] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[44] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
[45] S. W. Flynn, H. C. Zhao, and J. R. Green, J. Chem. Phys. **141**, 104107 (2014); J. W. Nichols, S. W. Flynn, and J. R. Green, *ibid.* **142**, 064113 (2015).
[46] P. Salamon and R. S. Berry, Phys. Rev. Lett. **51** (1983).
[47] C.-B. Li and T. Komatsuzaki, Phys. Rev. Lett. **111**, 058301 (2013).
[48] B. E. Husic and V. S. Pande, J. Am. Chem. Soc. **140**, 2386 (2018).
[49] G. E. Crooks, *Fisher information and statistical mechanics*, Tech. Rep. (Citeseer, 2011).
[50] L. Onsager, Phys. Rev. **37**, 405 (1931).
[51] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[52] A. C. Barato, D. Hartich, and U. Seifert, New J. Phys. **16**, 103024 (2014).
[53] T. McGrath, N. S. Jones, P. R. ten Wolde, and T. E. Ouldridge, Phys. Rev. Lett. **118**, 028101 (2017).
[54] J. Norris, *Markov Chains*, Vol. 1st (Cambridge University Press, 1997).