

Minimum entropy production, detailed balance and Wasserstein distance for continuous-time Markov processes

Andreas Dechant¹

¹*Department of Physics #1, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*
(Dated: October 5, 2021)

We investigate the problem of minimizing the entropy production for a physical process that can be described in terms of a Markov jump dynamics. We show that, without any further constraints, a given time-evolution may be realized at arbitrarily small entropy production, yet at the expense of diverging activity. For a fixed activity, we find that the dynamics that minimizes the entropy production is given in terms of conservative forces. The value of the minimum entropy production is expressed in terms of the graph-distance based Wasserstein distance between the initial and final configuration. This yields a new kind of speed limit relating dissipation, the average number of transitions and the Wasserstein distance. It also allows us to formulate the optimal transport problem on a graph in term of a continuous-time interpolating dynamics, in complete analogy to the continuous space setting. We demonstrate our findings for simple state networks, a time-dependent pump and for spin flips in the Ising model.

I. INTRODUCTION AND OVERVIEW

Entropy production quantifies how much a physical system is driven out of equilibrium. In practice, entropy production manifests itself in the form of dissipation, that is, energy that is irreversibly lost into the environment. Thus, minimizing the entropy production for a physical process is highly desirable from a practical point of view. But minimum entropy production is also interesting from a more fundamental point of view. There is a long-standing history of attempts to characterize non-equilibrium steady states via a minimum entropy production principle, see Ref. [1] for a historical and Ref. [2] for a more recent overview.

In this work, we consider the minimization of entropy production on a theoretical level, but motivated from the practical point of view. The general question we want to address is, given a physical process, what is the minimum amount of entropy production that is required to realize it? Obviously, in order to answer this question, we need to specify some details about the process and what are the parameters of the minimization. We will focus on Markovian dynamics in continuous time, for which entropy production can be defined, at least mathematically, in an unambiguous manner. For this class of dynamics, we can further differentiate between diffusion processes, where the state space is continuous, and jump processes, where the state space is discrete. In most cases, the continuous state space of diffusion processes is the d -dimensional real space \mathbb{R}^d and the dynamics can mathematically be described in terms of a Langevin equation [3] for the degrees of freedom or a Fokker-Planck equation [4] for their probability density. A physical process may then be characterized by the time-evolution of the probability density, or, if only the initial and final configuration are of interest, the initial and final probability density. In the latter case, it has been shown in Refs. [5–7], that the problem of minimizing the entropy production during the process is equivalent to calculating the Wasserstein distance between the two probability densities. The Wasserstein distance is a concept from optimal transport theory [8], which characterizes the optimal way of transforming an initial configuration into a final one. Since the Wasserstein distance characterizes the minimum entropy production, it also provides a lower bound on the entropy production [7]s. More recently, this approach has also been applied to the case where the detailed process instead of just the initial and final configurations is specified [9, 10].

For the case of a jump process, a similar formulation has only been developed very recently [11]. There, it was shown that a quantity resembling the Wasserstein distance, which is obtained by solving a modified optimal transport problem, provides a lower bound on the entropy production. However, this approach has two issues: First, the definition of the Wasserstein-distance-like quantity depends on the process itself and thus, unlike the Wasserstein distance in the continuous case, is not just a function of the initial and final configuration; it is further not directly related to the types of Wasserstein distance that are typically considered in the mathematical literature. Second, the argument of Ref. [11] only applies to systems whose transition rates satisfy detailed balance and thus does not allow for non-conservative forces. The latter point is important in light of another recent work [12], where it was shown that the entropy production is generally minimized by rates that do not satisfy detailed balance. This is in contrast to the continuous case, where the force that realizes minimum entropy production is always conservative and thus satisfies detailed balance [10, 13].

In this work, we address both of these issues. In Ref. [12] the entropy production was minimized under the condition that the symmetric part of the transition rates is fixed. We show that by adopting a slightly weaker condition, minimum entropy production does indeed correspond to rates that satisfy detailed balance. Specifically,

instead of the symmetric part of the rates, we fix the activity, which can be interpreted as the overall rate at which transitions occur in the system. Crucially, the minimum entropy production can be expressed in terms of the usual, graph-distance based Wasserstein distance, which only depends on the connectivity of the state space and the initial and final configuration. This generalizes the results of Refs. [5–7, 9, 10] to the case of a discrete state space.

Our paper is structured as follows: We consider a Markovian dynamics on a discrete state space, that is characterized by the time-dependent occupation probabilities $\mathbf{p}_t = (p_t(1), \dots, p_t(N))$ of the states labeled by $1, \dots, N$. This dynamics is introduced in detail in Section II. For a given time evolution of the occupation probabilities \mathbf{p}_t , we want to know is the minimum entropy production compatible with this time-evolution. As we discuss in section III, without any further constraints, the answer to this problem is trivial: we can make the entropy production arbitrarily small, so there is no finite lower bound. However, the price we have to pay for this is that the overall rate of transitions between states, quantified by the activity, diverges as the entropy production tends to zero. By contrast, as we show in Section IV, there exists a minimum value of the activity, that is, a given time-evolution requires a certain overall transition rate. The main result of this section is that the minimum activity can be expressed in terms of the usual Wasserstein distance on a graph between the initial and final occupation probability. While the minimum activity requires a diverging entropy production, we show in Section V that this formalism can also be used to construct the dynamics that minimize the entropy production for a given activity. This allows us to express the minimum entropy production in terms of the Wasserstein distance. These concepts are generalized to finite time-differences in Section VI, where, instead of the entire time-evolution, we only specify the initial and final occupation probabilities. This leads to a new type of speed limit for the transition between two states. In Section VII, we discuss the relation between a discrete and continuous state space and derive a generalization of the Benamou-Brenier formalism to the former case. Section VIII discusses the application of the results to a few simple state networks and to the Ising model.

II. CONTINUOUS-TIME MARKOV PROCESSES AND DETAILED BALANCE

We consider a continuous-time Markovian dynamics on a set of N states. Transitions between states j and i occur at a rate $W_t(i, j) \geq 0$ which may depend on time t through some external protocol. Then, the probability $p_t(i)$ of being in state i evolves according to the master equation [14, 15]

$$d_t p_t(i) = \sum_j (W_t(i, j)p_t(j) - W_t(j, i)p_t(i)), \quad (1)$$

with a prescribed set of initial occupation probabilities $\mathbf{p}_0 = (p_0(1), \dots, p_0(N))$. In the following, we assume that the rates satisfy the condition $W_t(i, j) = 0 \Leftrightarrow W_t(j, i) = 0$, i. e. that there are no unidirectional transitions. Then, we can parameterize the rates as [12]

$$W_t(i, j) = k(i, j)\omega_t(i, j) \exp\left(\frac{1}{2}A_t(i, j)\right), \quad (2)$$

with $k(i, j) = k(j, i) \in \{0, 1\}$, $\omega_t(i, j) = \omega_t(j, i) \geq 0$ and $A_t(i, j) = -A_t(j, i)$. Physically, $k(i, j)$ encodes the connectivity of the state network, i. e., between which states transitions are allowed ($k(i, j) = 1$) or forbidden ($k(i, j) = 0$). The parameters $\omega_t(i, j)$ constitute the symmetric part of the transition rates, which we can interpret as energetic barriers between the states. By contrast, the antisymmetric part $A_t(i, j)$ can be interpreted as the forces driving transitions between the states. Note that the parameterization Eq. (2) can represent an arbitrary set of transition rates, specifically, the symmetric and antisymmetric part are determined from the rates as

$$\omega_t(i, j) = \sqrt{W_t(i, j)W_t(j, i)}, \quad A_t(i, j) = \ln\left(\frac{W_t(i, j)}{W_t(j, i)}\right). \quad (3)$$

Note that, in the following sections, we will sometimes omit the subscript t in favor of a more compact notation when we consider the dynamics only at a specific time t . In general, the dynamics described by Eq. (1) are irreversible: Let us consider a trajectory $\Gamma = \{i(t)\}_{t \in [0, \tau]}$, i. e. we trace the instantaneous state of the system over a time-interval of length τ . Then, we can define a probability $P(\Gamma)$ of observing a specific trajectory. We can also imagine the time-reversed version of the dynamics, in which the system starts from the final state of Eq. (1), $\mathbf{p}_0^\dagger = \mathbf{p}_\tau$, and evolves according to the time-reversed protocol $W_t^\dagger(i, j) = W_{\tau-t}(i, j)$. The probability of observing the time-reversed trajectory $\Gamma^\dagger = \{i(\tau - t)\}_{t \in [0, \tau]}$ in the time-reversed dynamics is given by $P^\dagger(\Gamma^\dagger)$. The irreversibility of the system is measured by the entropy production [16]

$$\Delta S^{\text{irr}} = \sum_{\Gamma} P(\Gamma) \ln\left(\frac{P(\Gamma)}{P^\dagger(\Gamma^\dagger)}\right). \quad (4)$$

Mathematically, this is the Kullback-Leibler divergence between the path probabilities of the forward and reverse dynamics; this quantity is positive and vanishes only if, for every trajectory, the probability of observing its time-reversed version in the time-reversed dynamics is the same as for the forward trajectory in the original dynamics. In terms of the occupation probabilities and rates, the entropy production can be expressed as

$$\Delta S^{\text{irr}} = \int_0^\tau dt \sigma_t \quad \text{with} \quad \sigma_t = \frac{1}{2} \sum_{i,j} (W_t(i,j)p_t(j) - W_t(j,i)p_t(i)) \ln \left(\frac{W_t(i,j)p_t(j)}{W_t(j,i)p_t(i)} \right), \quad (5)$$

where $\sigma_t \geq 0$ is the rate of entropy production. The entropy production rate only vanishes, if for every pair of states (i, j) , the probability of a transition from j to i is precisely the same for as the reverse transition,

$$W_t(i,j)p_t(j) = W_t(j,i)p_t(i), \quad (6)$$

which is the detailed balance condition and equivalent to the system being in equilibrium [14]. From Eq. (1), this immediately implies $d_t \mathbf{p}_t = 0$ and thus, a vanishing entropy production rate can only occur if the occupation probabilities are independent of time, which is called a steady state. However, a steady state does not imply detailed balance; a steady state with a constant, non-vanishing rate of entropy production is referred to as a non-equilibrium steady state. We assume that, fixing $t = \hat{t}$, the dynamics described by Eq. (1) reaches a steady-state $\mathbf{p}_{\hat{t}}^{\text{st}}$ in the long-time limit. The detailed balance condition Eq. (6) in the steady state then is equivalent to

$$\exp(A_{\hat{t}}(i,j)) = \exp(\phi_{\hat{t}}^{\text{st}}(j) - \phi_{\hat{t}}^{\text{st}}(i)), \quad (7)$$

where we defined $\phi_{\hat{t}}^{\text{st}}(i) = -\ln(p_{\hat{t}}^{\text{st}}(i))$. This means that the forces $A_t(i,j)$ have to derive from a potential function $\phi_t(i)$ and are thus conservative. If $A_t(i,j)$ cannot be written in this way, that is, for non-conservative forces, detailed balance is broken and the steady state is out of equilibrium. We remark that both Eq. (6) and Eq. (7) are called detailed balance condition, the difference being that for Eq. (6) the system is already in the equilibrium state, whereas Eq. (7) only implies that an equilibrium state exists and will eventually be reached if the rates are kept constant. For convenience, we refer to Eq. (7) as the detailed balance condition in the following.

III. TIME EVOLUTION AT VANISHING ENTROPY PRODUCTION

From Eq. (5) it is obvious that the entropy production rate depends on both the transition rates and the occupation probabilities. These dependencies can be made explicit by writing [16]

$$\sigma_t = \sigma_t^{\text{m}} + \sigma_t^{\text{s}} \quad \text{with} \quad (8)$$

$$\sigma_t^{\text{m}} = \sum_{i,j} W_t(i,j)p_t(j) \ln \left(\frac{W_t(i,j)}{W_t(j,i)} \right) \quad \text{and} \quad \sigma_t^{\text{s}} = - \sum_i \ln(p_t(i)) d_t p_t(i).$$

The first term is generally referred to as the rate of entropy production in the medium; it quantifies the asymmetry in the transition rates. The second part is the rate of change of the Gibbs-Shannon entropy of the system; it depends explicitly on the time-evolution of the probabilities. The central question that we wish to answer in this work is, given a time-evolution of the occupation probabilities $\{\mathbf{p}_t\}_{t \in [0, \tau]}$, what is the minimum value of the entropy production compatible with this time-evolution. The motivation for this question is twofold: On the one hand, we may think of the evolution \mathbf{p}_t as the desired effect of driving the system, for example, when we want to pump the system into a higher-energy state. From this point of view, we want to minimize the dissipation associated with the driving. On the other hand, since any change in the probabilities \mathbf{p}_t implies a non-vanishing entropy production, we may ask how much we can infer about the dissipation in the system, which is often challenging to measure directly, from the observed evolution of \mathbf{p}_t . As it turns out, without any further qualifiers, the answer to the above question is trivial: The minimum entropy production associated with the time-evolution \mathbf{p}_t is zero. To see this, we define a new set of transition rates $W_t^{\mathcal{A}}(i,j)$, which leads to the same time-evolution at an arbitrary entropy production. First, we define currents $J_t(i,j)$ and traffic $D_t(i,j)$ as

$$J_t(i,j) = W_t(i,j)p_t(j) - W_t(j,i)p_t(i), \quad D_t(i,j) = W_t(i,j)p_t(j) + W_t(j,i)p_t(i). \quad (9)$$

$J_t(i,j)$ can be interpreted as the directed current across the transition from j to i , while $D_t(i,j)$ measures the rate of transitions along the transition in either direction. We then define the rates

$$W_t^{\mathcal{A}}(i,j) = \frac{1}{2p_t(j)} \left(\mathcal{A} + \text{sign}(J_t(i,j)) \right) |J_t(i,j)|, \quad (10)$$

with some real parameter $\mathcal{A} \geq 1$. A straightforward computation shows that the resulting currents are given by

$$J_t^{\mathcal{A}}(i, j) = J_t(i, j). \quad (11)$$

So, the rates Eq. (10) leave the currents invariant. Since, via Eq. (1), we have $d_t p_t(i) = \sum_j J_t(i, j)$, this also means that the time-derivative of the probabilities, and thus their time-evolution, remains unchanged. However, the entropy production rate associated with Eq. (10) is

$$\sigma_t^{\mathcal{A}} = \frac{1}{2} \sum_{i,j} |J_t(i, j)| \ln \left(\frac{\mathcal{A} + 1}{\mathcal{A} - 1} \right). \quad (12)$$

It is easily seen that, for given currents, this can take all values between zero (for $\mathcal{A} \rightarrow \infty$) to positive infinity (for $\mathcal{A} \rightarrow 1$). For large \mathcal{A} , in particular, we have

$$\sigma_t^{\mathcal{A}} \simeq \mathcal{A}^{-1} \sum_{i,j} |J_t(i, j)| + O(\mathcal{A}^{-3}). \quad (13)$$

Thus, we can make the entropy production rate arbitrarily small without altering the time-evolution of the occupation probabilities. However, there is a price we have to pay for this: If we consider the activity, which is defined as

$$\chi_t = \sum_{i,j \neq i} W_t(i, j) p_t(j) \quad (14)$$

and measures the overall rate of transitions in the system, we find

$$\chi_t^{\mathcal{A}} = \frac{\mathcal{A}}{2} \sum_{i,j} |J_t(i, j)|. \quad (15)$$

Thus, even as the entropy production vanishes, the overall rate of transitions in the system diverges, with the product of the two quantities approaching a constant value

$$\lim_{\mathcal{A} \rightarrow \infty} \sigma_t^{\mathcal{A}} \chi_t^{\mathcal{A}} = \frac{1}{2} \left(\sum_{i,j} |J_t(i, j)| \right)^2. \quad (16)$$

We remark that at the special value of the parameter \mathcal{A} at which the activity resulting from Eq. (10) is the same as the activity in the original system,

$$\mathcal{A}^* = \frac{[D_t]}{[|J_t|]} \quad \text{with} \quad [D_t] \equiv \sum_{i,j \neq i} D_t(i, j), \quad (17)$$

the corresponding entropy production is a guaranteed lower bound on the entropy production in the original dynamics,

$$\sigma_t^{\mathcal{A}^*} = \frac{1}{2} [|J_t|] \ln \left(\frac{[D_t] + [|J_t|]}{[D_t] - [|J_t|]} \right) \leq \sigma_t. \quad (18)$$

This inequality can be proven using elementary inequalities, see Appendix A. This means that, using Eq. (10), we have a dynamics with the same currents, time-evolution and activity as the original dynamics, yet at a reduced entropy production rate. The fact that we can reduce the entropy production at a given activity poses the question of whether Eq. (10) already constitutes the optimum or whether the entropy production may be reduced further.

IV. WASSERSTEIN DISTANCE AND MINIMAL ACTIVITY

Before we address the minimization of the entropy production, we are going to introduce the concept of Wasserstein distance on a graph. The intuitive idea of optimal transport [8, 17] is to find the optimal way of transforming a probability vector \mathbf{q} into \mathbf{p} , minimizing the cost of transporting probability in the process. If the cost of transporting probability is given by the distance over which the probability is transported, then the minimal cost is called Wasserstein distance. In the present setting, the connectivity matrix $k(i, j)$, which specifies which transitions are allowed or forbidden, is the adjacency matrix of a graph, whose vertices v_i are the states $i \in 1, \dots, N$ [18]. If $k(i, j) = 1$,

then the vertices v_i and v_j are connected by an edge $e(i, j)$, if $k(i, j) = 0$, then there is no edge between v_i and v_j . For now, we consider a simple undirected graph, that is $e(i, j)$ and $e(j, i)$ are equivalent and there is at most one edge between any pair of states. We call the collection of vertices and edges the graph $\mathcal{G}(v, e)$. We further define the graph distance $d_G(i, j)$ as the minimum number of edges in any path from v_i to v_j . Note that the shortest path between two vertices is generally not unique, so that there may be multiple shortest paths with length $d_G(i, j)$. This function satisfies all the axioms of a distance, $d_G(i, j) \geq 0$ with equality if and only if $v_i = v_j$, $d_G(i, j) = d_G(j, i)$ and $d_G(i, k) \leq d_G(i, j) + d_G(j, k)$ [18]. We assume that the graph is connected, that is In terms of this distance, the Wasserstein distance of order 1 or Kantorovich-Rubinstein distance [8] between two probability vectors \mathbf{p} and \mathbf{q} is defined as

$$\mathcal{W}(\mathbf{p}, \mathbf{q}) = \inf_{\Pi} \sum_{i, j} d_G(i, j) \Pi(i, j). \quad (19)$$

The infimum is taken over all couplings Π between \mathbf{p} and \mathbf{q} , which satisfy

$$\Pi(i, j) \geq 0, \quad \sum_i \Pi(i, j) = q(j), \quad \sum_j \Pi(i, j) = p(i). \quad (20)$$

This means that we are looking for the joint probability with marginals \mathbf{p} and \mathbf{q} that minimizes the expectation of the graph distance. As the name implies, the Wasserstein distance is a proper distance, satisfying the corresponding axioms [8]. An important property of Eq. (19) is that, in the above setting, the infimum always exists, however, the optimal coupling Π^* that realizes the minimal value is not unique. One reason is that the shortest path between any two vertices is not necessarily unique, so that we have several equivalent ways of transporting probability between the two vertices. Another reason is that, even if the shortest path is unique, we can easily construct a coupling that leads to the same value. To do so, assume we have two vertices with $\Pi(k, l) > 0$, which implies that we are moving some probability from l to k . Then, we consider one of the (possibly many) shortest paths from l to k which we denote by $\pi(k, l)$. By definition $\pi(k, l)$ consists of $d_G(k, l) \equiv m$ edges. We denote the directed edge from j to i by $\vec{e}(i, j)$ and write $\pi(k, l) = (\vec{e}(k, j_{m-1}), \vec{e}(j_{m-1}, j_{m-2}), \dots, \vec{e}(j_1, l))$. Now, we define a new coupling by setting $\tilde{\Pi}(k, l) = 0$ and $\tilde{\Pi}(i, j) = \Pi(i, j) + \Pi(k, l)$ for all $\vec{e}(i, j) \in \pi(k, l)$. Intuitively, this means that, instead of moving probability directly from v_l to v_k , we first move it from v_l to v_{j_1} , then from v_{j_1} to v_{j_2} and so on, finally arriving at v_k . Since, along the shortest path, the graph distance is additive, we are moving the same amount of probability by the same distance and thus the value of Eq. (19) is the same for Π and $\tilde{\Pi}$. This means that for any coupling Π (which does not have to be the optimal one), we can always construct an equivalent coupling that is non-zero only the edges of the graph. In the following, we can thus assume that $\Pi(i, j) = 0$ whenever the edge $\vec{e}(i, j)$ is not part of the graph.

Now we make the connection to the Markov jump dynamics Eq. (1). We specialize Eq. (19) to the case $\mathbf{p} = \mathbf{p}_{t+dt}$ and $\mathbf{q} = \mathbf{p}_t$, that is, we consider the solution of Eq. (1) at two infinitesimally different times. From Eq. (20), we then obtain the condition

$$d_t p(i) = \frac{p_{t+dt}(i) - p_t(i)}{dt} = \frac{1}{dt} \sum_j (\Pi(i, j) - \Pi(j, i)). \quad (21)$$

Since, by the above discussion, we can assume that $\Pi(i, j) > 0$ only if $k(i, j) = 1$ ($W_t(i, j) > 0$) we may thus identify

$$\Pi(i, j) = W_t(i, j) p_t(j) dt. \quad (22)$$

This provides a one-to-one correspondence between the off-diagonal elements of the coupling and the transition rates. Eq. (20) further fixes the diagonal elements as

$$\Pi(i, i) = \left(1 - \sum_{k \neq i} W_t(k, i) dt \right) p_t(i), \quad (23)$$

which is positive for sufficiently small dt . We can thus compute the activity associated with a given coupling,

$$\chi_t = \sum_{i, j \neq i} W_t(i, j) p_t(j) = \frac{1}{dt} \sum_{i, j \neq i} \Pi(i, j) = \frac{1}{dt} \sum_{i, j} d_G(i, j) \Pi(i, j). \quad (24)$$

In the last step, we used that all pairs of vertices for which $\Pi(i, j) > 0$ we have either $d_G(i, j) = 1$ (if $i \neq j$) or $d_G(i, j) = 0$ (if $i = j$). Apart from the factor dt , this is precisely the same as the functional in Eq. (19). Thus,

minimizing the activity for a given connectivity and time-evolution is precisely the same as computing the Wasserstein distance between \mathbf{p}_t and \mathbf{p}_{t+dt} and

$$\chi_t^* = \frac{1}{dt} \mathcal{W}(\mathbf{p}_{t+dt}, \mathbf{p}_t). \quad (25)$$

Next, we want to exploit the known properties of the optimal coupling Π^* . A central result of optimal transport theory is the Kantorovich duality [8]. For the present case, this result states that there exists a function $\psi(i)$ defined on the vertices that satisfies

$$\psi(j) - \psi(i) = d_G(i, j) \quad \text{for all } (i, j) \text{ with } \Pi^*(i, j) > 0. \quad (26)$$

In other words, for all edges on which the optimal coupling is non-zero, the graph distance can be written in terms of a potential function $\psi(i)$. Since we may assume that the optimal coupling is non-zero only on the edges of the graph, this implies

$$\psi(j) - \psi(i) = 1 \quad \text{for all } (i, j) \text{ with } \Pi^*(i, j) > 0, \quad (27)$$

that is, we may characterize the optimal coupling by a potential landscape with energy difference ± 1 between neighboring sites. More precisely, we have $\psi(j) - \psi(i) = 1$ if probability is transported from j to i ($\Pi^*(i, j) > 0$) and $\psi(j) - \psi(i) = -1$ if probability is transported from i to j ($\Pi^*(j, i) > 0$), so that flow of probability described by the optimal coupling is always downhill in the potential. This also implies that for given i and j , probability flows between them in only one direction and the transition rates Eq. (22) defined by Π^* are unidirectional $W_t^*(i, j) > 0 \Rightarrow W_t^*(j, i) = 0$. From the point of view of minimizing the activity, this is reasonable, as any reverse transitions also contribute to the activity. However, this also means that at the minimal activity, the entropy production rate diverges. Since our original goal was minimizing the entropy production, it seems we have only achieved the opposite. However, as we will see in the next section, the information about dynamics realizing minimal activity will actually be useful for minimizing the entropy production. We remark that, if the optimal coupling Π^* is known, the energy landscape $\psi(i)$ is determined uniquely up to a constant shift [8]. Conversely, the knowing the energy landscape also determines the Wasserstein distance, which can be written as

$$\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0) = \sum_i \psi(i) (p_{\tau+dt}(i) - p_t(i)). \quad (28)$$

V. MINIMUM ENTROPY PRODUCTION AT FIXED ACTIVITY

Minimizing the entropy production rate requires minimizing Eq. (5) with respect to the transition rates under the constraints that (i) the rates lead to the correct time-evolution Eq. (1) and that (ii) the resulting activity Eq. (14) has the same value as in the original dynamics. To simplify the notation we introduce the parameters $C(i, j) = -C(j, i)$ via

$$A(i, j) = 2C(i, j) + \ln \left(\frac{p(i)}{p(j)} \right), \quad (29)$$

and omit the subscript t in the following. In terms of these parameters, the master equation, entropy production rate and activity can be written as

$$d_t p(i) = 2 \sum_j \sqrt{p(i)p(j)} k(i, j) \omega(i, j) \sinh(C(i, j)), \quad (30a)$$

$$\sigma = 2 \sum_{i, j} \sqrt{p(i)p(j)} k(i, j) \omega(i, j) C(i, j) \sinh(C(i, j)), \quad (30b)$$

$$\chi = \sum_{i, j \neq i} \sqrt{p(i)p(j)} k(i, j) \omega(i, j) \cosh(C(i, j)). \quad (30c)$$

Then, we minimize Eq. (30b) with respect to $\omega(i, j)$ and $C(i, j)$ under the constraints Eq. (30a) and Eq. (30c). Taking the derivative of Eq. (30) with respect to $\omega(i, j)$ and $C(i, j)$ yields the conditions for a stationary point,

$$k(i, j) (C(i, j) \sinh(C(i, j)) + \gamma \cosh(C(i, j))) = k(i, j) (\lambda(j) - \lambda(i)) \sinh(C(i, j)), \quad (31a)$$

$$k(i, j) \omega(i, j) (C(i, j) \cosh(C(i, j)) + (\gamma + 1) \sinh(C(i, j))) = k(i, j) \omega(i, j) (\lambda(j) - \lambda(i)) \cosh(C(i, j)), \quad (31b)$$

where γ and $\lambda(i)$ are Lagrange multipliers. For all (i, j) such that $k(i, j)\omega(i, j) > 0$, this results in the condition

$$\left(\tanh(C(i, j))\right)^2 = \frac{\gamma}{1 + \gamma} \equiv q^2 \quad \text{with} \quad 0 \leq q \leq 1. \quad (32)$$

Thus for all pairs of states with a non-zero transition rate between them, $C(i, j)$ has to be constant, specifically

$$C(i, j) = s(i, j)\text{artanh}(q) \quad (33)$$

with $s(i, j) = -s(j, i) \in \{-1, 1\}$. Plugging this into Eq. (31a) yields

$$\lambda(j) - \lambda(i) = s(i, j)\mathcal{E}_q \quad \text{with} \quad \mathcal{E}_q = \left(\text{artanh}(q) + \frac{q}{1 - q^2}\right). \quad (34)$$

Thus the Lagrange multipliers $\lambda(i)$ define an energy landscape, in which any two states with a non-vanishing probability flow between them are separated by an energy \mathcal{E}_q . This is precisely the condition that is satisfied by the optimal transport coupling Π^* ! We choose

$$\lambda(i) = \mathcal{E}_q\psi(i), \quad \omega(i, j) = \frac{\sqrt{1 - q^2}}{2qdt} \frac{\Pi^*(i, j) + \Pi^*(j, i)}{\sqrt{p(i)p(j)}}. \quad (35)$$

For this choice, we obtain from Eq. (30)

$$d_t p(i) = \frac{1}{dt} \sum_j (\Pi^*(i, j) - \Pi^*(j, i)), \quad (36a)$$

$$\sigma = \frac{2\text{artanh}(q)}{dt} \sum_{i, j \neq i} \Pi^*(i, j), \quad (36b)$$

$$\chi = \frac{1}{qdt} \sum_{i, j \neq i} \Pi^*(i, j). \quad (36c)$$

The first equation is satisfied because of Eq. (21) and the third equation can be solved for q ,

$$q = \frac{\chi^*}{\chi}, \quad (37)$$

where we used Eq. (24) and Eq. (25). This yields the minimal entropy production rate

$$\sigma^* = 2\chi^* \text{artanh}\left(\frac{\chi^*}{\chi}\right). \quad (38)$$

In summary, for a given connectivity of the state network $k(i, j)$ and time-evolution \mathbf{p}_t , there exists a minimal activity χ^* , which is equal to the Wasserstein distance between \mathbf{p}_t and \mathbf{p}_{t+dt} . The minimal entropy production rate at a given activity χ is then determined by the ratio of χ^* and χ and diverges logarithmically as $\chi \rightarrow \chi^*$. We remark that the above is different from the approach pursued in Ref. [12], where the symmetric part $\omega(i, j)$ of the rates was kept fixed. Intuitively, fixing the activity is a weaker constraint compared to fixing the symmetric parts of all the rates. Consequently, as we discuss in Appendix B, we expect that the minimum entropy production Eq. (38) is smaller than the minimum entropy production of Ref. [12] in most cases, even though this relation is not strict.

The somewhat convoluted notation of the above derivation does not lend itself to physical intuition. However, we can use Eq. (10) to gain a more intuitive understanding. Suppose that, in Eq. (10), instead of the original rates $W(i, j)$, we use the rates $W^*(i, j)$ corresponding to the optimal coupling Π^* via Eq. (22). Since the rates $W^*(i, j)$ are unidirectional, we then find

$$W^{\mathcal{A}}(i, j) = \begin{cases} \frac{\mathcal{A}+1}{2} W^*(i, j) & \text{if } W^*(i, j) > 0 \\ \frac{\mathcal{A}-1}{2} W^*(j, i) & \text{if } W^*(j, i) > 0. \end{cases} \quad (39)$$

The corresponding entropy production rate and activity are given by

$$\sigma^{\mathcal{A}} = 2\chi^* \text{artanh}\left(\frac{1}{\mathcal{A}}\right) \quad (40a)$$

$$\chi^{\mathcal{A}} = \mathcal{A}\chi^*. \quad (40b)$$

For $\mathcal{A} = 1$, we re-obtain the unidirectional rates corresponding to minimal activity and diverging entropy production. For any $\mathcal{A} > 1$, however, the reverse transition rate becomes non-zero and thus the entropy production rate is finite. Thus Eq. (39) corresponds to a symmetrized version of the unidirectional rates $W^*(i, j)$. For the choice $\mathcal{A} = \chi/\chi^*$, the activity is equal to the activity of the original dynamics and the entropy production yields the minimal value at this activity. In that sense, the minimal entropy production is obtained by first solving the optimal transport problem and then symmetrizing the resulting transition rates by just the right amount. In terms of Eq. (2), the symmetric and antisymmetric part of the transition rates realizing the minimal entropy production rate are given by

$$\omega^*(i, j) = \sqrt{\left(\frac{\chi}{\chi^*}\right)^2 - 1} \frac{\Pi^*(i, j) + \Pi^*(j, i)}{2\sqrt{p(i)p(j)}dt} \quad (41a)$$

$$A^*(i, j) = \phi(j) - \phi(i) + 2\text{artanh}\left(\frac{\chi^*}{\chi}\right)(\psi(j) - \psi(i)), \quad (41b)$$

with $\phi(i) = -\ln p(i)$. This form explicitly shows that the forces $A^*(i, j)$ are derived from a potential and thus satisfy the detailed balance condition Eq. (6). This potential includes two contributions: The term involving $\phi(i)$ derives from differences between the instantaneous occupation probabilities; it favors transitions from states with high probability to states with low probability. The second term involving $\psi(i)$ derives from the structure of the optimal coupling Eq. (27). Since $\psi(j) - \psi(i) = \pm 1$, we may interpret the prefactor as the energy scale defining the energy landscape. As we approach the minimal activity, $\chi \rightarrow \chi^*$, the energy scale diverges, while the symmetric part $\omega(i, j)$ tends to zero, which reproduces the finite unidirectional transition rates corresponding to the solution of the optimal transport problem.

We remark that we can also calculate the pseudo-entropy production rate [19], which is defined by

$$\rho = \sum_{i,j} \frac{(W(i, j)p(j) - W(j, i)p(i))^2}{W(i, j)p(j) + W(j, i)p(i)}. \quad (42)$$

Like the entropy production, the pseudo-entropy production quantifies the breaking of detailed balance in the system. While it cannot be expressed in terms of the time-forward and time-reversed path probabilities like Eq. (4), it has the advantage that it remains finite in the presence of unidirectional transitions. It is further a lower bound on both the entropy production rate and the activity: $\rho_t \leq \sigma_t$ and $\rho_t \leq 2\chi_t$, where the former bound turns into an equality close to equilibrium. For Eq. (39), the pseudo-entropy production rate is given by

$$\rho^{\mathcal{A}} = \frac{2\chi^*}{\mathcal{A}}. \quad (43)$$

Solving Eq. (30) by minimizing ρ instead of σ yields precisely the same conditions, so that the entropy production rate and pseudo-entropy production rate are simultaneously minimized for $\mathcal{A} = \chi/\chi^*$,

$$\rho^* = \frac{2(\chi^*)^2}{\chi}. \quad (44)$$

VI. FINITE TIMES

In the derivation of the minimum entropy production rate Eq. (38), we focused on the situation where the entire evolution of the probabilities is specified. However, in many situations, only the initial and final configuration \mathbf{p}_0 and \mathbf{p}_τ are of interest. In this case, instead of the instantaneous activity χ_t , it is more natural to fix its time-integral

$$\int_0^\tau dt \chi_t = \langle M \rangle, \quad (45)$$

which is equal to the average number of transitions in the time interval $[0, \tau]$. Then, we want to minimize the total entropy production Eq. (5) during the time τ , keeping the average number of transitions fixed. As before, we first establish some properties of the optimal transport problem. Suppose we know an optimal coupling Γ^* between \mathbf{p}_0 and \mathbf{p}_τ . Then, from Eq. (20), we have

$$p_\tau(i) - p_0(i) = \sum_j (\Gamma^*(i, j) - \Gamma^*(j, i)). \quad (46)$$

Then, we define

$$\Pi^*(i, j) = \begin{cases} \Gamma^*(i, j) \frac{dt}{\tau} & \text{for } i \neq j \\ 1 - \sum_k \Gamma^*(k, i) \frac{dt}{\tau} & \text{for } i = j. \end{cases} \quad (47)$$

This satisfies

$$\frac{p_\tau(i) - p_0(i)}{\tau} = \frac{1}{dt} (\Pi^*(i, j) - \Pi^*(j, i)). \quad (48)$$

and, for sufficiently small dt/τ , $\Gamma^*(i, j) > 0$. If we can find a dynamics that satisfies

$$\mathbf{p}_t = \mathbf{p}_0 + \frac{t}{\tau} (\mathbf{p}_\tau - \mathbf{p}_0) \quad \Rightarrow \quad d_t \mathbf{p}_t = \frac{\mathbf{p}_\tau - \mathbf{p}_0}{\tau}, \quad (49)$$

that is, whose solution interpolates between \mathbf{p}_0 and \mathbf{p}_τ at a constant rate of change of the probability, then, by Eq. (48), Π^* is a coupling between \mathbf{p}_t and \mathbf{p}_{t+dt} . We further have

$$\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0) = \sum_{i,j} d_G(i, j) \Gamma^*(i, j) = \frac{\tau}{dt} \sum_{i,j} d_G(i, j) \Pi^*(i, j). \quad (50)$$

Suppose we divide the time interval $[0, \tau]$ into $K \gg 1$ intervals $[kdt, (k+1)dt]$. We can write the above as

$$\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0) = K \sum_{i,j} d_G(i, j) \Pi^*(i, j). \quad (51)$$

Because the Wasserstein distance satisfies the triangle inequality, we also have

$$\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0) \leq K \mathcal{W}(\mathbf{p}_{t+dt}, \mathbf{p}_t), \quad (52)$$

where we used that the distance between \mathbf{p}_t and \mathbf{p}_{t+dt} only depends on the (constant) rate of change $d_t \mathbf{p}_t$ because of Eq. (21). Comparing Eq. (51) and Eq. (52), we then have

$$\sum_{i,j} d_G(i, j) \Pi^*(i, j) \leq \mathcal{W}(\mathbf{p}_{t+dt}, \mathbf{p}_t). \quad (53)$$

Since Π^* is a coupling between \mathbf{p}_t and \mathbf{p}_{t+dt} , it is already is an optimal one, since, by definition, the Wasserstein distance is equal to the infimum of the quantity on the left. This means that for a constant rate of change of the probability, Eq. (49), we have

$$\frac{\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0)}{\tau} = \frac{\mathcal{W}(\mathbf{p}_{t+dt}, \mathbf{p}_t)}{dt}. \quad (54)$$

In other words, a constant rate of probability change between the initial and final state yields the geodesic between the two states with respect to the metric defined by the Wasserstein distance. The coupling Π^* also defines the rates that lead to the dynamics Eq. (49) via Eq. (22).

The above discussion shows that the optimal process that minimizes the total number of transitions is one whose rate of change in the probability is constant. For this process, we have

$$\langle M \rangle^* = \tau \chi_t^* = \tau \frac{\mathcal{W}(\mathbf{p}_{t+dt}, \mathbf{p}_t)}{dt} = \mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0). \quad (55)$$

That is, the minimal number of transitions required to transform the initial state \mathbf{p}_0 into the final state \mathbf{p}_τ is given by the Wasserstein distance between the two states. This provides a direct interpretation of the Wasserstein distance between two states in terms of the process connecting the two states. Further, since χ_t^* and χ_t are independent of time, we obtain from Eq. (38)

$$\Delta S^{\text{irr},*} = 2\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0) \text{artanh} \left(\frac{\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0)}{\langle M \rangle} \right). \quad (56)$$

Thus, the minimal amount of entropy production that is required to transform \mathbf{p}_0 into \mathbf{p}_τ is given in terms of the Wasserstein distance and the average number of transitions. The proof that a constant rate of change in the

probability indeed minimizes the entropy production is given in Appendix C. For an arbitrary dynamics, the right-hand side Eq. (56) constitutes a lower bound on the entropy production during the process. Since the number of transitions is in often proportional to time $\langle M \rangle = \tau \bar{\chi}$, with the time-averaged activity $\bar{\chi}$, Eq. (56) also translates into a speed limit for the transition from \mathbf{p}_0 to \mathbf{p}_τ ,

$$\tau \geq \frac{\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0)}{\bar{\chi} \tanh\left(\frac{\Delta S^{\text{irr}}}{2\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0)}\right)}. \quad (57)$$

That is, the minimal time for a transition between two states is determined by the Wasserstein distance, the activity and the amount of dissipation. Using the inequalities $\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0) \geq \delta(\mathbf{p}_\tau, \mathbf{p}_0)$ with the total variation distance $\delta(\mathbf{p}, \mathbf{q}) = \sum_i |p(i) - q(i)|/2$ and $\tanh(x) \leq x$ we re-obtain the speed limit derived in Ref. [20],

$$\tau \geq \frac{2\delta(\mathbf{p}_\tau, \mathbf{p}_0)^2}{\bar{\chi} \Delta S^{\text{irr}}}. \quad (58)$$

However, Eq. (57) is tighter than this bound. In particular, since the hyperbolic tangent approaches unity for large arguments, we see that in the limit of large dissipation, the activity becomes the limiting factor for the transition time,

$$\tau \geq \frac{\mathcal{W}(\mathbf{p}_\tau, \mathbf{p}_0)}{\bar{\chi}}. \quad (59)$$

This means that, in contrast to what is suggested by Eq. (58), strong driving and large dissipation do not allow us to realize arbitrarily fast transition times. Another appealing feature of Eq. (57) is that its derivation shows that the speed limit is tight, that is, there always exists a process that realizes the minimal transition time. Instead of a speed limit, Eq. (56) may also be interpreted as a tradeoff relation between dissipation and precision. Since the Wasserstein distance is equal to the minimum number of transitions, we may write Eq. (56) as

$$\Delta S^{\text{irr}} \geq 2\langle M \rangle^* \text{artanh}\left(\frac{\langle M \rangle^*}{\langle M \rangle}\right). \quad (60)$$

If the dynamics transforms the initial into the final state with the minimum number of transitions, it has perfect precision: Transitions only occur in the “correct” direction and never in reverse. Eq. (60) implies that this requires infinite dissipation. If we sacrifice precision and allow transitions to also occur in the reverse direction, then the dissipation becomes finite, however, we also require more transitions to realize the same change in the state. We remark that similar tradeoff relations have been obtained for Brownian clocks [21] and stochastic currents [22]. The relation Eq. (60) establishes a tradeoff relation on the basis of the system state itself, rather than derived quantities. In addition to the lower bound corresponding to Eq. (58), we can also give an upper bound on the minimum entropy production. The maximal graph distance between any two vertices is defined as the diameter of the graph $\max_{(i,j) \in \mathcal{G}} d_{\mathcal{G}}(i, j) = d(\mathcal{G})$ [18], which also is an upper bound on the Wasserstein distance [8]. We thus obtain lower and upper bounds on the minimum entropy production,

$$2\delta(\mathbf{p}_\tau, \mathbf{p}_0) \text{artanh}\left(\frac{\delta(\mathbf{p}_\tau, \mathbf{p}_0)}{\langle M \rangle}\right) \leq \Delta S^{\text{irr},*} \leq 2d(\mathcal{G}) \text{artanh}\left(\frac{d(\mathcal{G})}{\langle M \rangle}\right). \quad (61)$$

The lower bound depends only on the initial and final state via their total variation distance and is independent of the connectivity of the state space. $\delta(\mathbf{p}_\tau, \mathbf{p}_0)$ is the Wasserstein distance on a complete graph. By contrast, the upper bound depends only on the connectivity of the state space and is independent of the initial and final state. $d(\mathcal{G})$ equals the Wasserstein distance between an initial and final state each concentrated on one of two maximally distant vertices.

VII. RELATION TO LANGEVIN DYNAMICS

The results of the preceding section imply three alternative but equivalent ways of formulating the optimal transport problem of computing the Wasserstein distance Eq. (19). All of them involve constructing a Markov jump process of the form Eq. (1) interpolating between the two probability vectors $\mathbf{q} = \mathbf{p}_0$ and $\mathbf{p} = \mathbf{p}_\tau$.

1. Find the Markov jump process with initial state $\mathbf{p}_0 = \mathbf{q}$ and final state $\mathbf{p}_\tau = \mathbf{p}$ that minimizes the average number of transitions $\langle M \rangle$. Then, the Wasserstein distance is given by $\mathcal{W}(\mathbf{p}, \mathbf{q}) = \langle M \rangle^*$.

2. Find the Markov jump process with initial state $\mathbf{p}_0 = \mathbf{q}$ and final state $\mathbf{p}_\tau = \mathbf{p}$ that minimizes the entropy production ΔS^{irr} , Eq. (5), at a given average number of transitions $\langle M \rangle$. Then, the Wasserstein distance is related to the minimum entropy production via Eq. (56).
3. Find the Markov jump process with initial state $\mathbf{p}_0 = \mathbf{q}$ and final state $\mathbf{p}_\tau = \mathbf{p}$ that minimizes the pseudo entropy production $\Delta R^{\text{irr}} = \int_0^\tau dt \rho_t$, Eq. (42), at a given average number of transitions $\langle M \rangle$. Then, the Wasserstein distance is related to the minimum pseudo entropy production via

$$\Delta R^{\text{irr},*} = \frac{2\mathcal{W}(\mathbf{p}, \mathbf{q})^2}{\langle M \rangle} = \frac{2\mathcal{W}(\mathbf{p}, \mathbf{q})^2}{\tau \bar{\chi}}. \quad (62)$$

The last formulation, in particular, is appealingly similar to the well-known Benamou-Brenier formula [23] for the L^2 -Wasserstein distance in \mathbb{R}^n . We recall Eq. (9) defining the probability current $J_t(i, j)$ and traffic $D_t(i, j)$,

$$J_t(i, j) = W_t(i, j)p_t(j) - W_t(j, i)p_t(i), \quad D_t(i, j) = W_t(i, j)p_t(j) + W_t(j, i)p_t(i). \quad (63)$$

We define the discrete gradient and divergence operator

$$\begin{aligned} (\text{grad}(a))(i, j) &= \frac{1}{2}(a(i) - a(j)), \\ (\text{div}(A))(i) &= \frac{1}{2} \sum_j (A(j, i) - A(i, j)). \end{aligned} \quad (64)$$

Note that the gradient operator defines an antisymmetric $N \times N$ -matrix in terms of the N -vector \mathbf{a} , while the divergence operator defines an N -vector in terms of the antisymmetric part of the $N \times N$ -matrix A . With this notation, we can write Eq. (1) as a continuity equation

$$d_t \mathbf{p}_t = -\text{div}(J_t). \quad (65)$$

Let us further define the inner products for vectors and matrices, respectively

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_i a(i)b(i) \quad \text{and} \quad \langle A, B \rangle = \sum_{i,j} A(i, j)B(i, j). \quad (66)$$

With this definition, it is straightforward to check that we have the discrete integration by parts formula

$$\langle \mathbf{a}, \text{div}(A) \rangle = -\langle \text{grad}(\mathbf{a}), A \rangle. \quad (67)$$

The pseudo-entropy production and average number of transitions can be written as

$$\Delta R^{\text{irr}} = \int_0^\tau dt \left\langle J_t, \frac{J_t}{D_t} \right\rangle \quad (68a)$$

$$\langle M \rangle = \frac{1}{2} \int_0^\tau dt \langle D_t, 1_{N \times N} \rangle, \quad (68b)$$

where the division is taken as component-wise division and $1_{N \times N}$ is a matrix with all entries equal to 1. In summary, the statement of the optimal transport problem in terms of the pseudo-entropy is the following: Find

$$\Delta R^{\text{irr},*} = \inf_{J_t, D_t} \int_0^\tau dt \left\langle J_t, \frac{J_t}{D_t} \right\rangle \quad (69)$$

such that the continuity equation Eq. (65) with boundary conditions $\mathbf{p}_0 = \mathbf{q}$ and $\mathbf{p}_\tau = \mathbf{p}$ the constraint Eq. (68b) are satisfied. Then, the Wasserstein distance is related to the minimum pseudo-entropy production by Eq. (62). By comparison, for a Langevin dynamics ($\mathbf{x} \in \mathbb{R}^n$) with diffusion coefficient \mathcal{D} , the probability density $p_t(\mathbf{x})$ satisfies the Fokker-Planck equation

$$\partial_t p_t(\mathbf{x}) = -\text{div}(\mathbf{j}_t(\mathbf{x})) \quad \text{with} \quad \mathbf{j}_t(\mathbf{x}) = \mathbf{f}_t(\mathbf{x})p_t(\mathbf{x}) - \mathcal{D} \text{grad}(p_t(\mathbf{x})) \quad (70)$$

and the entropy production is given by

$$\Delta S^{\text{irr}} = \int_0^\tau dt \int d\mathbf{x} \frac{|\mathbf{j}_t(\mathbf{x})|^2}{\mathcal{D}p_t(\mathbf{x})} = \int_0^\tau dt \left\langle \mathbf{j}_t, \frac{\mathbf{j}_t}{\mathcal{D}p_t} \right\rangle, \quad (71)$$

where the inner product is defined as

$$\langle \mathbf{a}, \mathbf{b} \rangle = \int d\mathbf{x} \mathbf{a}(\mathbf{x}) \cdot \mathbf{b}(\mathbf{x}), \quad (72)$$

where \cdot is the standard scalar product in \mathbb{R}^n . As has been discussed in Refs. [5–7], we may consider the optimization problem

$$\Delta S^{\text{irr},*} = \inf_{\mathbf{j}_t, p_t} \int_0^\tau dt \left\langle \mathbf{j}_t, \frac{\mathbf{j}_t}{\mathcal{D}p_t} \right\rangle \quad (73)$$

under the constraint that the continuity equation Eq. (70) with boundary conditions $p(\mathbf{x})_0 = q(\mathbf{x})$ and $p(\mathbf{x})_\tau = p(\mathbf{x})$ is satisfied. The solution to this problem is related to the Wasserstein distance of order 2 between $p_0(\mathbf{x})$ and $p_\tau(\mathbf{x})$ via

$$\Delta S^{\text{irr},*} = \frac{\mathcal{W}_2(p_\tau, p_0)^2}{\tau \mathcal{D}}. \quad (74)$$

The latter is defined as

$$\mathcal{W}_2(p_\tau, p_0) = \left(\inf_{\Pi} \int d\mathbf{x} \int d\mathbf{y} \|\mathbf{x} - \mathbf{y}\|^2 \Pi(\mathbf{x}, \mathbf{y}) \right)^{\frac{1}{2}}, \quad (75)$$

where $\Pi(\mathbf{x}, \mathbf{y})$ is a joint probability density with marginals $p_\tau(\mathbf{x})$ and $p_0(\mathbf{y})$. The resemblance between Eq. (69) and Eq. (73) is obvious. There are, however, two crucial differences between the two results: First, the Wasserstein distance in Eq. (74) is defined in terms of the order 2 distance on euclidean space, whereas Eq. (62) involves the order 1 distance on a graph. And second, the quantity in the denominator is different, specifically, the diffusion coefficient \mathcal{D} appears in the Langevin case and the time-averaged activity $\bar{\chi}$ in the Markov jump case. However, given their striking similarity, we conjecture that Eq. (74) indeed emerges from Eq. (62), provided that the jump dynamics Eq. (1) tends to a Langevin dynamics in a suitable continuum limit. In this limit, we expect that the constraint on the number of transitions is equivalent to fixing the diffusion coefficient \mathcal{D} . We further remark that, in the continuum limit, the quantities ΔS^{irr} and ΔR^{irr} become equivalent, so both problems involve a minimization of the entropy production under constraints.

There is, however, another, more fundamental difference between the continuous and discrete case: Let us return to the formulation of Section V, that is, we fix the time evolution of the occupation probabilities \mathbf{p}_t and the activity χ_t and minimize the entropy production rate σ_t . For the Langevin case, a similar setup—fixing the diffusion coefficient instead of the activity—was discussed in Refs. [10, 13]. In that case, it was found that the force that realizes the minimum entropy production dynamics is the gradient of a potential, which is the equivalent of the detailed balance condition Eq. (7). In the Langevin case, there is further a one-to-one correspondence between the potential and the time-dependent probability density, that is, there is only one potential (up to an additive constant) that gives rise to a given time-evolution. As a consequence, if the dynamics is described by a potential force, this force already minimizes the entropy production rate. As we found in Section V, the antisymmetric part of the transition rates also satisfies the detailed balance condition and thus is derived from a potential. However, in contrast to the Langevin case, the one-to-one relation between the potential and the time-evolution of \mathbf{p}_t is lost. The reason is that, in principle, we may vary the symmetric and antisymmetric part of the transition rates separately, so that different potentials may give rise to the same time-evolution. This is clear from the potential landscape of the minimum entropy production dynamics, which has a fixed energy difference between any two connected states, while this condition is not satisfied by a general potential. This means that, even if the rates of the original dynamics satisfy the detailed balance condition Eq. (7), we can generally reduce the entropy production rate further while keeping the time-evolution and the activity invariant. The consequence of this difference is that, while in the Langevin case, the minimum entropy production for a given time-evolution can be identified as the excess entropy production [13] (in the sense of non-equilibrium thermodynamics, see also Refs. [24–26]) and the remaining entropy production as the housekeeping part, this identification is not straightforward in the Markov jump case.

VIII. ILLUSTRATIVE EXAMPLES

A. Simple networks

While the matrix $k(i, j)$ specifies the connectivity of the state network of the original dynamics, for the minimum entropy production dynamics, it is the maximal connectivity. The difference between the two is that, even if a

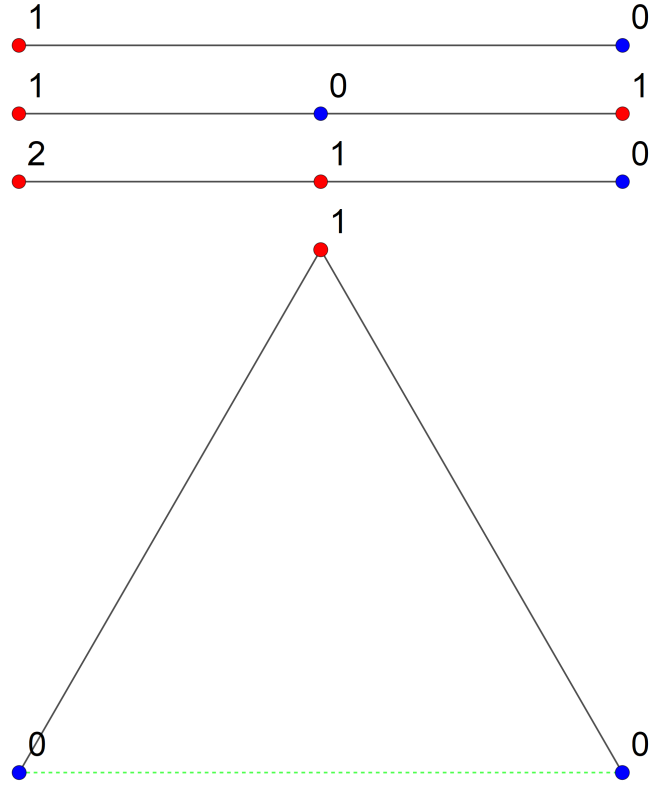


FIG. 1. Possible configurations for $N = 2$ and $N = 3$ states. Red dots represent states with $d_t p_t(i) < 0$ (sources of probability), while blue dots represent states with $d_t p_t(i) > 0$ (sinks of probability). The numbers are the energies $\psi(i)$ of each state, chosen such that the lowest energy is zero. Solid lines are connect states with a non-vanishing probability flow between them in the minimum entropy production dynamics, while the green dashed lines connect states between which a connection is permitted in principle ($k(i, j) = 1$), but between which there is no probability flow in the minimum entropy production dynamics. From top to bottom: For $N = 2$ states, there is only one possible non-trivial configuration and the Wasserstein distance is equal to the total variation distance. For $N = 3$ states on a line, the Wasserstein distance may be equal (upper configuration) or larger (lower configuration) than the total variation distance, depending on the change in the occupation probabilities. Note that the second configuration has a maximal energy difference of 3. For a loop with $N = 3$ states, two states are always effectively disconnected in the minimum entropy production dynamics and the Wasserstein distance is equal to the total variation distance.

connection between two states is allowed in principle ($k(i, j) = 1$), the transition rates Eq. (39) between the two states may be zero. Physically, this corresponds to inserting a large energy barrier between the two states, effectively preventing transitions between them. One particular example for this are loops consisting of an odd number of states. Along such a loop, it is not possible to define an energy landscape with constant step size in a consistent manner. Thus, the minimum entropy production dynamics never contains odd loops. While loops with an even number of states are allowed in principle, we can always find a choice for the optimal coupling Π^* that does not contain any loops, such that the corresponding graph is a tree. Precisely speaking, we may also have a collection of trees without any connection between them, provided that the total change in probability in any one tree is zero. The ambiguity in the choice of the optimal coupling reflects the non-uniqueness of the graph distance in the presence of even loops: States on opposite sides of the loop are always connected by (at least) two paths of the same length, so we can move probability between them along either path without changing the value of the Wasserstein distance.

Let us consider a few simple state networks. For $N = 2$ states, we always have one state with $d_t p_t(1) > 0$ and one state with $d_t p_t(2) < 0$. Then, we assign the energies $\psi(1) = 0$ and $\psi(2) = 1$, which ensures that probability flows from state 2 to state 1. In this case, the Wasserstein distance is equal to the total variation distance. As remarked before, this is always the case for a complete graph, i. e. $k(i, j) = 1$ for all pairs of states. More generally, the Wasserstein distance is also equal to the total variation distance, if any state i with $d_t p_t(i) > 0$ is connected to all states j with $d_t p_t(j) < 0$ and vice versa. In this case, the energies $\psi(i)$ can be chosen such that $\psi(i) = 0$ for $d_t p_t(i) > 0$ and $\psi(i) = 1$ for $d_t p_t(i) < 0$, that is, the largest energy difference between any two states is 1. For $N = 3$, there are two possible configurations that result in a connected graph: Either the three states can be arranged on a line, e. g. $k(2, 1) = k(3, 2) = 1$ and $k(1, 3) = 0$. In this case, the connectivity of the minimum entropy production dynamics

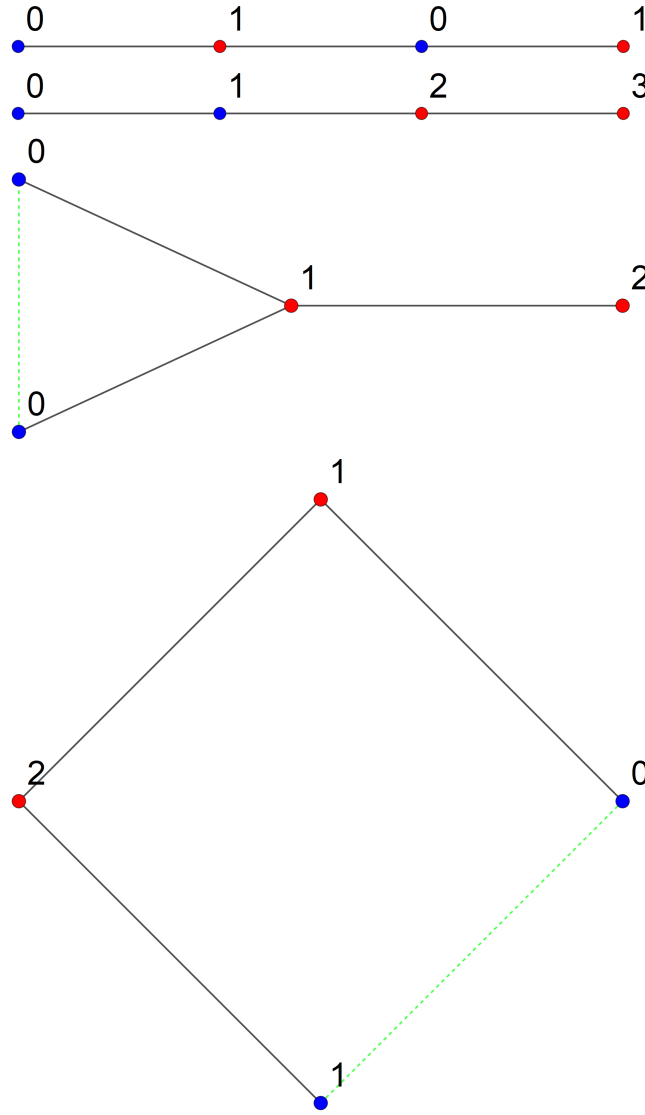


FIG. 2. Some configurations for $N = 4$ states. The symbols are the same as in Fig. 1. As for three states, a linear configuration may lead to a Wasserstein distance that is equal to (upper configuration) or larger than (lower configuration) the total variation distance. If the original graph contains any odd-numbered loop, those loops are cut in the minimum entropy production dynamics. Finally, for an even-numbered loop, there always exists a tree-like configuration. However, in this case, we may also transport (some or all of the) probability from the left to the right state via the lower path without changing the Wasserstein distance, so we can also find an equivalent configuration which contains the loop.

is generally the same as the original graph. However, even for this very simple configuration, the Wasserstein distance is generally larger than the total variation distance: If $d_t p_t(1) > 0$, $d_t p_t(2) > 0$ and $d_t p_t(3) < 0$, then we cannot avoid moving some probability from state 1 to state 3, and thus over a graph distance of $d(1, 3) = 2$. The other possibility is that the three states form a loop, e. g. $k(2, 1) = k(3, 2) = k(1, 3) = 1$. In this case, since the loop contains an odd number of states, the minimum entropy production dynamics will always have zero probability flow across one of the edges and we obtain a configuration that can be arranged on a line. However, since the original graph is complete, the Wasserstein distance is always equal to the total variation distance in this case. The possible configurations for 2 and 3 states are shown in Fig. 1. For $N = 4$ there are already several qualitatively different possibilities, taking into account the connectivity and distribution of probability changes. Some examples are shown in Fig. 2. The number of possible configurations increases rapidly with the number of states, two random examples for $N = 10$ states are shown in Fig. 3.

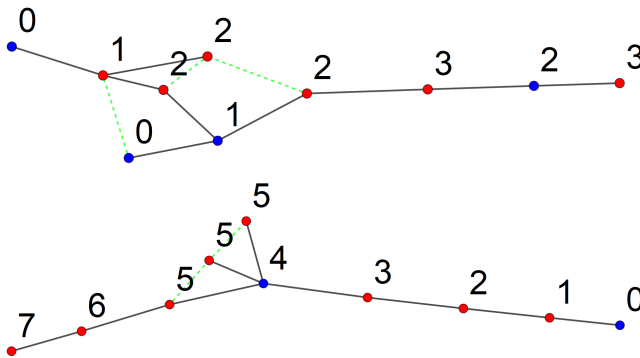


FIG. 3. Two sample configurations for $N = 10$ states. The symbols are the same as in Fig. 1. For the upper configuration, the Wasserstein distance is about 1.43 times, for the lower configuration about 3.52 times the total variation distance.

B. Pumping current on a ring

In the previous section, we only discussed the static structure of the minimum entropy production process for a given initial and final state. Now we want to extend this to a dynamic situation, where a system is actually driven by an external protocol. We consider a ring of N states, with transitions between nearest neighbors occurring according to the time-dependent rates

$$W_t(i, j) = \omega_0 \exp\left(-\frac{1}{2}(E_t(i) - E_t(j))\right) \quad (76)$$

$$\text{with } E_t(i) = \epsilon_0 \left(\frac{1}{2} - \frac{1}{2} \cos\left(2\pi\left(t + \frac{j-1}{N}\right)\right)\right).$$

The energy of each state varies periodically between 0 and ϵ_0 , with a phase-shift of $2\pi/N$ between two neighboring states. As a consequence, the location of the minimum energy state moves around the ring and thus the rates Eq. (76) correspond to pumping a probability current around the ring. Note that the rates satisfy the detailed balance condition Eq. (7) for any time t . The parameter ω_0 quantifies the rate at which transitions occur between the different states and thus determines how well the system can follow the time-dependence of the rates. For $\omega_0 \gg 1$, the transitions between the states are fast compared to the change in the rates and thus the occupation probabilities are close to the instantaneous equilibrium probabilities, $p_t(i) \simeq p_t^{\text{eq}}(i)$ with

$$p_t^{\text{eq}}(i) = \frac{e^{-E_t(i)}}{\sum_j e^{-E_t(j)}}. \quad (77)$$

For $\omega_0 \ll 1$, on the other hand, the rates change faster than transitions between the individual states occur; in this case, the transitions are approximately determined by the rates averaged over one period and we have $p_t(i) \approx 1/N$. In the long-time limit, the dynamics Eq. (76) lead to time-periodic occupation probabilities. While these do not have an analytic expression, they are easily obtained numerically, which allows us to calculate the entropy production rate σ_t Eq. (5) and activity χ_t Eq. (14) at any given time. With the time-evolution of the occupation probabilities and the activity, we can then solve the optimization problem in Eq. (30) for any instant of time and thus compute the minimum activity χ_t^* , and the minimum entropy production rate σ_t^* at the given activity χ_t . We can also compute the energy landscape for the minimum entropy production dynamics, which is defined using Eq. (41) as

$$E_t^*(i) = -\ln(p_t(i)) + 2\text{artanh}\left(\frac{\chi_t^*}{\chi_t}\right)\psi_t(i), \quad (78)$$

where $\psi_t(i)$ is determined by the optimal coupling via Eq. (27). The results are shown in Fig. 4 for fast transitions ($\omega_0 = 10$) in Fig. 4 and for slow transitions ($\omega_0 = 0.1$) in Fig. 5. First, we note that the activity is approximately constant as a function of time. This means that in this case, we would obtain the same results by fixing the time-averaged activity instead of its instantaneous value. Next, even though the rates Eq. (76) already satisfy detailed balance, we can further reduce the entropy production rate while keeping the time-evolution of the occupation probabilities and the activity invariant. The relative reduction $\sigma_t^*/\sigma_t \approx 0.7$ has the same order of magnitude for both fast and

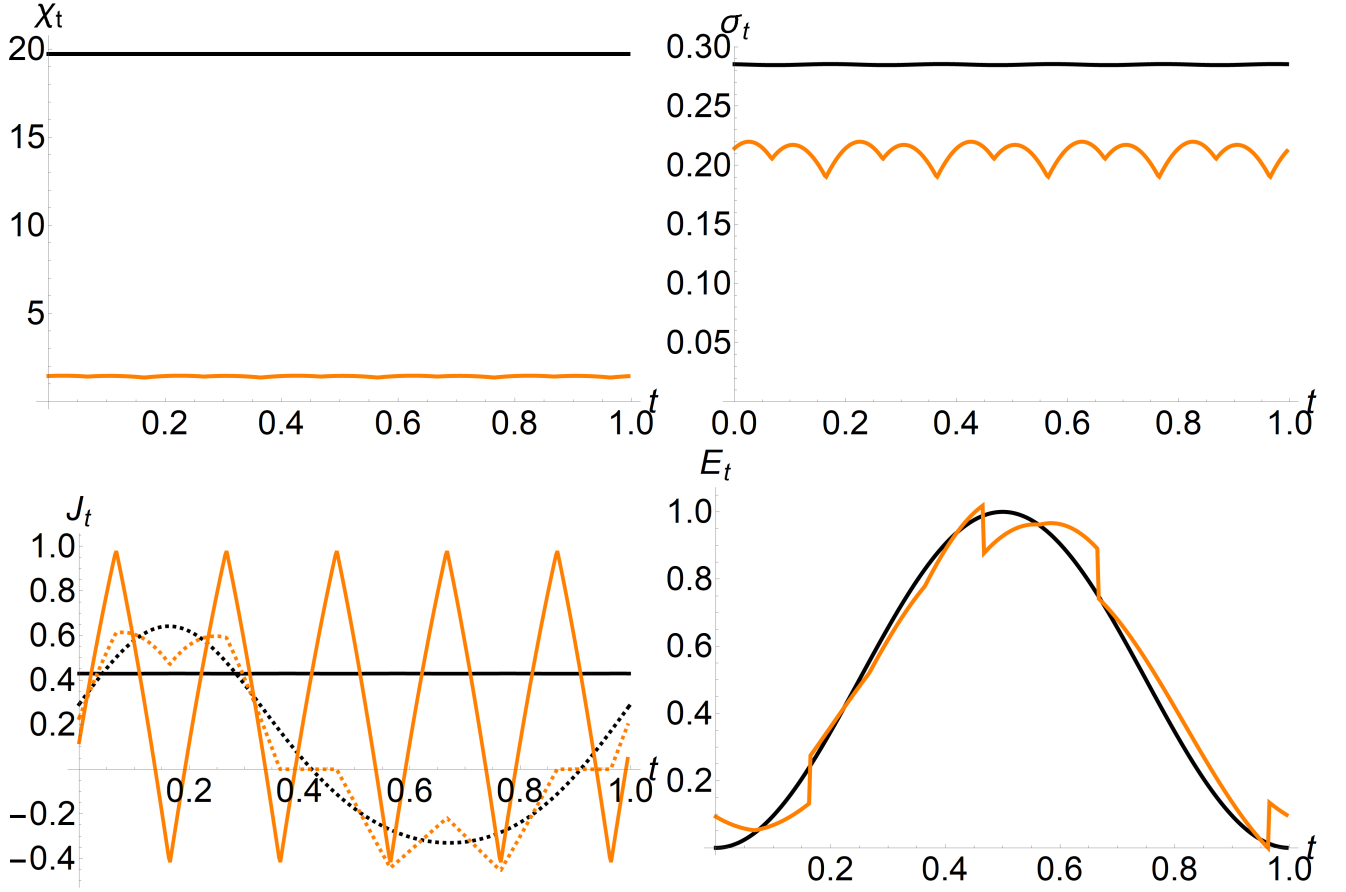


FIG. 4. Results for the pumping model Eq. (76) with $N = 5$, $\epsilon_0 = 1$ and $\omega_0 = 10$ as a function of time for one period of the driving. (Top left) The activity χ_t (black) and the minimal activity χ_t^* (orange). (Top right) The entropy production rate σ_t (black) and the minimum entropy production rate σ_t^* (orange). (Bottom left) The probability current $J_t(2, 1)$ (dashed) and the total probability current around the ring (solid) for the original dynamics (black) and the minimum entropy production dynamics (orange). (Bottom right) The energy $E_t(1)$ (black, see Eq. (76)) and the energy $E_t^*(1)$ in the minimum entropy production dynamics (orange, see Eq. (78)).

slow transitions. If the transitions between states are fast compared to the driving, then we observe a finite current and approximately constant around the ring, while for slow transitions, where the system cannot follow the driving, the overall current vanishes. While this behavior is the same for the current in the minimum entropy production dynamics when averaged over one period, the instantaneous current shows marked oscillations, both for fast and slow transitions. At first glance, this is counter-intuitive, as it indicates that driving a current back and forth across the ring can reduce the entropy production rate. However, the latter is actually determined by the local currents between individual states, rather than the global current in the entire ring. Still, it is not trivial that optimizing the local currents would lead to a more erratic behavior in the overall current. Finally, with respect to the energy landscape, for fast transitions, we observe that the energy landscape of the minimum entropy production dynamics, Eq. (76), closely follows the energies in the original dynamics. The reason is that, in this limit, the second term in Eq. (78) is small, since the activity is much larger than the minimum value, so the energy E_t^* is mostly determined by the equilibrium probabilities Eq. (77) and thus the original energy landscape. By contrast, for slow transitions, the occupation probabilities are approximately constant and equal and thus enter Eq. (78) only as a trivial constant shift. In this limit, we observe a marked step-like behavior of the minimum entropy production energy landscape, which reflects the fact that the direction of the probability flow and thus the optimal coupling change in a non-smooth manner during the time-evolution.

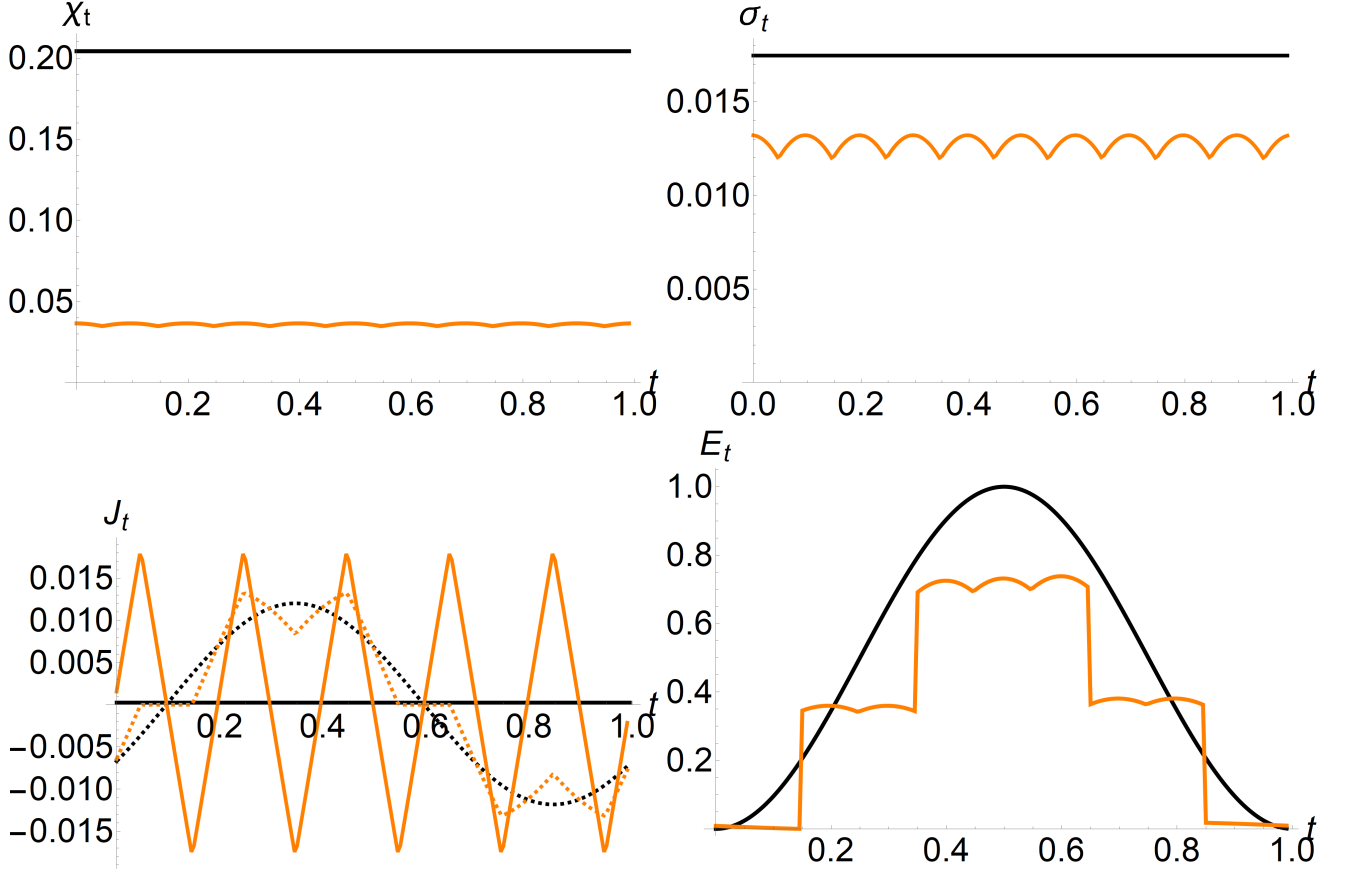


FIG. 5. Results for the pumping model Eq. (76) with $N = 5$, $\epsilon_0 = 1$ and $\omega_0 = 0.1$ as a function of time for one period of the driving. (Top left) The activity χ_t (black) and the minimal activity χ_t^* (orange). (Top right) The entropy production rate σ_t (black) and the minimum entropy production rate σ_t^* (orange). (Bottom left) The probability current $J_t(2, 1)$ (dashed) and the total probability current around the ring (solid) for the original dynamics (black) and the minimum entropy production dynamics (orange). (Bottom right) The energy $E_t(1)$ (black, see Eq. (76)) and the energy $E_t^*(1)$ in the minimum entropy production dynamics (orange, see Eq. (78)).

C. Spin flips in the Ising model

The Ising model is a paradigmatic example for a physical system on a discrete state space. We consider a collection of K classical spins arranged on a d -dimensional square lattice. Each spin can have one of two directions, up (denoted by $\sigma(k) = +1$) and down (denoted by $\sigma(k) = -1$). The spins interact via a pairwise ferromagnetic interaction between nearest neighbors, characterized by the Hamiltonian

$$H_t(\boldsymbol{\sigma}) = -J_t \sum_{\langle kl \rangle} \sigma(k)\sigma(l) - h_t \sum_k \sigma(k), \quad (79)$$

where J_t quantifies the strength of the interaction, h_t is an external field and $\langle kl \rangle$ denotes the sum over nearest-neighbor pairs. Thus, the first term is minimized when all spins are parallel, whereas the second term is minimized when all spins are aligned with the external field. For simplicity, we assume that the parameters J_t and h_t are the same for all spins, but may change as a function of time. The state i of the system is determined by the value of each spin, that is, the vector $\boldsymbol{\sigma} = (\sigma(1), \dots, \sigma(K))$, so that there are $N = 2^K$ different states. We assume that transitions between different states can occur only via single spin flips, so that $k(i, j) = 1$ if and only if i and j differ by the orientation of exactly one spin, and that the transition rates have the form

$$W_t(i, j) = k(i, j)\omega \exp\left(-\frac{\beta_t}{2}(H_t(i) - H_t(j))\right), \quad (80)$$

where $\beta_t = 1/T_t$ is the inverse temperature and the constant ω quantifies the overall speed of the dynamics. Since the transition rates satisfy detailed balance, the steady state occupation probability (at fixed parameters) is the

Boltzmann-Gibbs equilibrium

$$p_t^{\text{eq}}(i) = \frac{e^{-\beta_t H_t(i)}}{\mathcal{Z}_t}, \quad (81)$$

where $\mathcal{Z}_t = \sum_i e^{-\beta_t H_t(i)}$ is the partition function. The system can be driven out of equilibrium by varying the parameters, typically the external field h_t and temperature β_t , as a function of time according to a given protocol. This leads to a time-dependent occupation probability $p_t(i)$ which is generally different from the instantaneous Boltzmann-Gibbs state. A typical example is a quench, where the system starts out in the equilibrium state corresponding to initial parameters values h_i and β_i . Then, at $t = 0$, the value of the parameters is instantaneously changed to h_f and β_f and we can observe the random evolution of the system's state and, by repeating this procedure sufficiently often, the occupation probability $p_t(i)$, as it relaxes to the new equilibrium state. Using Eq. (8), we can write the total entropy production during the process as

$$\Delta S^{\text{irr}} = \Delta S^{\text{m}} + \Delta S^{\text{s}} \quad \text{with} \quad (82a)$$

$$\Delta S^{\text{m}} = - \int_0^\tau dt \beta_t \sum_i H_t(i) \partial_t p_t(i) \quad (82b)$$

$$\Delta S^{\text{s}} = - \sum_i \left(\ln(p_\tau(i)) p_\tau(i) - \ln(p_0(i)) p_0(i) \right). \quad (82c)$$

For the case of a quench the value of the temperature and the functional form of the Hamiltonian are independent of time during the process and we can further write

$$\Delta S^{\text{m}} = -\beta_f \sum_i H_f(i) (p_\tau(i) - p_0(i)). \quad (83)$$

The advantage of this formulation is that it only depends on the initial and final occupation probabilities and the Hamiltonian after the quench. Let us suppose that we wait for a sufficiently long time after the quench so that the occupation probabilities have relaxed to their equilibrium values Eq. (81). Then, the total entropy production is a function of only the initial and final Hamiltonian

$$\Delta S^{\text{irr}} = -\beta_f \langle H_f \rangle_i + \beta_i \langle H_i \rangle_i + \ln \left(\frac{\mathcal{Z}_f}{\mathcal{Z}_i} \right), \quad (84)$$

where $\langle \dots \rangle_i$ denotes an average with respect to the Boltzmann-Gibbs equilibrium corresponding to the initial Hamiltonian. For the graph defined by the transition rates Eq. (80), the graph distance between any two states is the number of spins whose direction differs between the two states. Further, the number of transitions $\langle M \rangle$ is the average number of spin flips during the process and the Wasserstein distance $\mathcal{W}(\mathbf{p}_f^{\text{eq}}, \mathbf{p}_i^{\text{eq}})$ corresponds to the minimum average number of spin-flips that is required to transform the initial occupation probability into the final one. Since the Wasserstein distance likewise only depends on the initial and final occupation probability, we can thus derive a criterion for the number of spin flips during the quench,

$$\langle M \rangle \geq \frac{\mathcal{W}(\mathbf{p}_f^{\text{eq}}, \mathbf{p}_i^{\text{eq}})}{\tanh \left(\frac{\Delta S^{\text{irr}}}{2\mathcal{W}(\mathbf{p}_f^{\text{eq}}, \mathbf{p}_i^{\text{eq}})} \right)}. \quad (85)$$

Note that, unlike the entropy production and the Wasserstein distance, $\langle M \rangle$ depends not only on the initial and final state but on the dynamics of the system after the quench. Thus, this inequality provides a constraint on the dynamics of the system after a quench. Note that, so far, we have not used the specific form of the Ising-Hamiltonian Eq. (79), so the above result is valid for any quench dynamics between two equilibrium configurations, even though the interpretation of $\langle M \rangle$ of course depends on the physical system. As a concrete example, we consider a one-dimensional chain consisting of $K = 4$ spins. We fix the energy scale by setting $J = 1$ and consider a quench from $h_i = -h_0$ to $h_f = h_0$, i. e. a reversal of the external field. The lower bound on $\langle M \rangle$ is shown as a function of the temperature for various values of h_0 in Fig. 6. We see that the number of spin flips is always bounded from below by the number of spins. For low temperature, this result is expected, since the statistics are dominated by states in which most spins are aligned with the external field, so, on average, we have to flip every spin to reach the new equilibrium configuration. However, for high temperature, this bound is not so easy to understand, since the contribution from all states is of the same order and the initial and final equilibrium configuration are only slightly different. Moreover, we find that the bound on the number of spin flips is larger for a weaker external field. For a strong external field, the spins will align

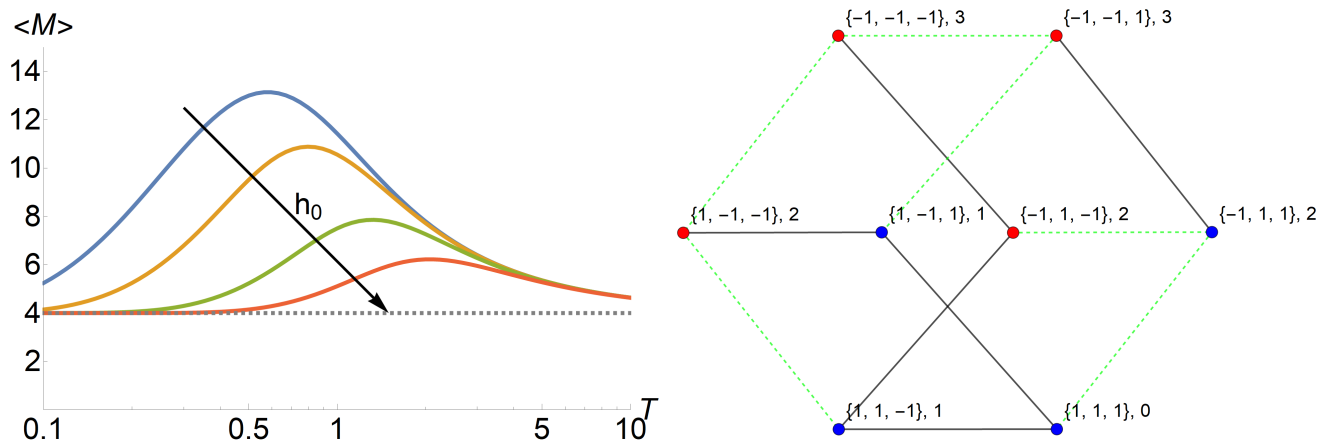


FIG. 6. (Left panel) The lower bound Eq. (85) on the number of spin flips for a field reversal in a four-state Ising chain as a function of temperature. The lines correspond to $h_0 = 0.1, 0.2, 0.5, 1.0$ from top to bottom. (Right panel) The graph corresponding to the minimum entropy production dynamics for a three-state Ising chain under field reversal for $h_0 = 0.5$ and $T = 1.0$. The numbers in braces represent the spin configuration, the second number is the energy $\psi(i)$ of the state; the symbols are the same as in Fig. 1.

with the external field in an almost deterministic manner; for a weak field, by contrast, the dynamics of the spins is dominated by the interactions between the spins and thus events where a spin randomly aligns with the external field just to flip back because of the orientation of its neighbors increase the number of spin flips during the quench.

Since the number of states grows exponentially with the number of spins, it quickly becomes unfeasible to explicitly compute the Wasserstein distance. However, as noted above, for low temperature, the statistics is dominated by the states with the minimum energy, which corresponds to fully ordered states in the presence of an external field. Then, reversing the direction of the field requires us to flip every spin, which means that the Wasserstein distance is approximately equal to K . We thus obtain the simpler approximate bound in the low temperature limit

$$\langle M \rangle \geq \frac{K}{\tanh\left(\frac{\Delta S^{\text{irr}}}{2K}\right)}. \quad (86)$$

As noted in Eq. (60), this can be interpreted as a tradeoff between precision and dissipation: Reducing the number of spin flips requires increasing the dissipation. While such a relation is intuitively expected, Eq. (86) puts it on a quantitative footing.

IX. DISCUSSION

In this work, we have explored the connection between minimum entropy production, detailed balance and Wasserstein distance. Our results suggest several possible new research directions. First, as remarked in Section VII, there is a striking resemblance between the discrete jump and continuous diffusion cases. While we conjectured that the two results are indeed equivalent in the continuum limit, this should be shown explicitly. The most obvious candidate are hypercubic lattice models that correspond to discrete representations of \mathbb{R}^d . However, it would be particularly interesting to explore other cases as well, for example triangulations of smooth surfaces.

Second, as remarked in the introduction, another approach to relate entropy production in jump processes to a modified Wasserstein distance has been developed recently [11]. While technically different from our approach, it would be interesting to see whether the two approaches can be related. Further, in Ref. [11], the same approach was generalized to open quantum systems, so a natural question is whether the current approach also can be applied to this setting. The main challenge here is to develop a reasonable graph-based representation of the density matrix as opposed to the probability vector.

Finally, in the Langevin case, the concept of minimum entropy production was used in Ref. [13] to define a decomposition of the entropy production into excess and housekeeping parts. The excess entropy production rate, which is equal to the minimal entropy production rate, vanishes only in a steady state, while the housekeeping part vanishes for a conservative force. This allows to separate the dissipation in the system into contributions due to time-dependent driving and non-conservative forces. Further, the excess entropy can be expressed in terms of the Wasserstein distance [10]. Formally, we may also define the excess part as the minimum production rate in the Markov jump case

and express it in terms of the Wasserstein distance. It is tempting to identify the remainder as the housekeeping part. However, as we saw in Section VIII, the entropy production rate is generally not minimized for an arbitrary conservative force, which implies that the housekeeping part can be non-zero even for conservative forces. Thus, it needs to be clarified what the physical meaning of this “housekeeping entropy” is.

ACKNOWLEDGMENTS

The author gratefully acknowledges inspiring and helpful discussions with S. Ito and S.-i. Sasa.

Appendix A: Proof of Eq. (18)

Eq. (18) provides a lower bound on the entropy production rate for a specific choice of the parameter \mathcal{A} in Eq. (10). Adopting the notation of Eq. (30), the inequality Eq. (18) reads

$$2 \sum_{(i,j)} m(i,j) C(i,j) \sinh(C(i,j)) \geq 2 \sum_{(i,j)} m(i,j) |\sinh(C(i,j))| \operatorname{artanh} \left(\frac{\sum_{(i,j)} m(i,j) |\sinh(C(i,j))|}{\sum_{(i,j)} m(i,j) \cosh(C(i,j))} \right), \quad (\text{A1})$$

where we defined the symmetric matrix $m(i,j) = \sqrt{p(i)p(j)} k(i,j) \omega(i,j)$, whose entries are positive, and the sum runs over all unequal index pairs (i,j) with $i \neq j$. This is equivalent to

$$\tanh \left(\frac{\sum_{(i,j)} m(i,j) |C(i,j)| |\sinh(C(i,j))|}{\sum_{(i,j)} m(i,j) |\sinh(C(i,j))|} \right) \geq \frac{\sum_{(i,j)} m(i,j) |\sinh(C(i,j))|}{\sum_{(i,j)} m(i,j) \cosh(C(i,j))}, \quad (\text{A2})$$

where we used that the hyperbolic sine is an odd function so that $x \sinh(x) = |x| |\sinh(x)|$. We define

$$P(i,j) = \frac{m(i,j) |\sinh(C(i,j))|}{\sum_{(i,j)} m(i,j) |\sinh(C(i,j))|}, \quad (\text{A3})$$

which is normalized probability with respect to (i,j) ,

$$P(i,j) \geq 0 \quad \text{and} \quad \sum_{i,j} P(i,j) = 1. \quad (\text{A4})$$

Denoting the average with respect to this probability by $\langle \dots \rangle$, we can write the left-hand side of Eq. (A2) as

$$\begin{aligned} \tanh \left(\frac{\sum_{(i,j)} m(i,j) |C(i,j)| |\sinh(C(i,j))|}{\sum_{(i,j)} m(i,j) |\sinh(C(i,j))|} \right) &= \tanh(\langle |C| \rangle) \\ &\geq \langle \tanh(|C|) \rangle = \frac{\sum_{(i,j)} m(i,j) |\sinh(C(i,j))| |\tanh(C(i,j))|}{\sum_{(i,j)} m(i,j) |\sinh(C(i,j))|}, \end{aligned} \quad (\text{A5})$$

where we used that the hyperbolic tangent is a concave function for positive arguments and employed Jensen's inequality. Plugging this into Eq. (A2), we have to show that

$$\sum_{(i,j)} m(i,j) \frac{|\sinh(C(i,j))|^2}{\cosh(C(i,j))} \geq \frac{\left(\sum_{(i,j)} m(i,j) |\sinh(C(i,j))| \right)^2}{\sum_{(i,j)} m(i,j) \cosh(C(i,j))}. \quad (\text{A6})$$

However, this is nothing but the Cauchy-Schwarz inequality, for we can write,

$$\begin{aligned} \left(\sum_{(i,j)} m(i,j) |\sinh(C(i,j))| \right)^2 &= \left(\sum_{(i,j)} \frac{\sqrt{m(i,j)} |\sinh(C(i,j))|}{\sqrt{\cosh(C(i,j))}} \sqrt{m(i,j) \cosh(C(i,j))} \right)^2 \\ &\leq \sum_{(i,j)} \frac{m(i,j) |\sinh(C(i,j))|^2}{\cosh(C(i,j))} \sum_{(i,j)} m(i,j) \cosh(C(i,j)). \end{aligned} \quad (\text{A7})$$

Thus, we have proven Eq. (18) by means of elementary inequalities.

Appendix B: Relation to the minimum entropy production rate of Ref. [12]

In Ref. [12], the entropy production rate was minimized under the constraint that the symmetric part $\omega(i, j)$ of the rates Eq. (2) remains fixed, i. e. the entropy production rate was minimized only with respect to the antisymmetric part $A(i, j)$. While, intuitively, this is a stronger constraint than fixing the activity, the relation between the resulting minimum entropy production rates has to be determined explicitly. A central result of Ref. [12] is that, when fixing the symmetric part of the rates, the optimal forces $A(i, j)$ that minimize the entropy production rate do generally not satisfy the detailed balance condition Eq. (7). Let us consider an alternative minimization problem. Instead of the entropy production rate, we minimize the functional

$$g = \sigma - 2\chi \quad (\text{B1})$$

under the constraint that the time evolution of the occupation probabilities and the symmetric part $\omega(i, j)$ are kept fixed. This is similar to the problem considered in Ref. [12], the difference being that we subtract twice the activity from the entropy production rate. We recall the notation introduced in Eq. (30)

$$d_t p(i) = 2 \sum_j \sqrt{p(i)p(j)} k(i, j) \omega(i, j) \sinh(C(i, j)), \quad (\text{B2a})$$

$$\sigma = 2 \sum_{i,j} \sqrt{p(i)p(j)} k(i, j) \omega(i, j) C(i, j) \sinh(C(i, j)), \quad (\text{B2b})$$

$$\chi = \sum_{i,j \neq i} \sqrt{p(i)p(j)} k(i, j) \omega(i, j) \cosh(C(i, j)). \quad (\text{B2c})$$

Minimizing g with respect to $A(i, j)$ is equivalent to minimizing with respect to $C(i, j)$ since the two are related via the linear transformation Eq. (29). We obtain the condition for a stationary point

$$k(i, j) \omega(i, j) C(i, j) \cosh(C(i, j)) = k(i, j) \omega(i, j) (\lambda(j) - \lambda(i)) \cosh(C(i, j)), \quad (\text{B3})$$

which implies that, whenever $k(i, j) \omega(i, j) > 0$, we should have

$$C(i, j) = \lambda(j) - \lambda(i). \quad (\text{B4})$$

This means that the transition rates that minimize g have to satisfy the detailed balance condition. Mathematically, subtracting the term 2χ precisely cancels the term that leads to a breaking of detailed balance in Ref. [12]. Now suppose that we have found the rates that minimize g . We denote the corresponding entropy production $\sigma_{g|\omega}$ and activity $\chi_{g|\omega}$ with a subscript $g|\omega$ indicating that they were obtained by minimizing g while fixing ω . Since the rates satisfy detailed balance, the result of Ref. [12] implies that there exists another set of rates, possessing the same symmetric part and the same time evolution, which do not satisfy detailed balance and lead to a smaller entropy production rate $\sigma_{\sigma|\omega} < \sigma_{g|\omega}$. We denote the corresponding activity by $\chi_{\sigma|\omega}$. However, since we minimized g , we have the inequality

$$g_{g|\omega} = \sigma_{g|\omega} - 2\chi_{g|\omega} \leq \sigma_{\sigma|\omega} - 2\chi_{\sigma|\omega} = g_{\sigma|\omega}. \quad (\text{B5})$$

Since $\sigma_{\sigma|\omega} < \sigma_{g|\omega}$, this implies

$$\chi_{\sigma|\omega} < \chi_{g|\omega}. \quad (\text{B6})$$

So the minimum entropy production dynamics of Ref. [12] necessarily also has a lower activity than the dynamics minimizing g . Now we fix the activity $\chi_{\sigma|\omega}$ and again minimize the entropy production rate according to Section V. We thus obtain a yet lower minimum entropy production

$$\sigma_{\sigma|\omega}^* = 2\chi_{\sigma|\omega}^* \operatorname{artanh}\left(\frac{\chi_{\sigma|\omega}^*}{\chi_{\sigma|\omega}}\right). \quad (\text{B7})$$

Note that since χ^* only depends on the time evolution of the occupation probabilities and the connectivity of the state network, it is the absolute minimum value of the activity and independent of whether we start from the original dynamics or the minimum entropy production dynamics at fixed ω . Since we have $\chi_{\sigma|\omega} < \chi_{g|\omega}$, we have the inequalities

$$\sigma_{\sigma|\omega} \geq \sigma_{\sigma|\omega}^* \geq \sigma_{g|\omega}^*. \quad (\text{B8})$$

Thus, starting from the detailed balanced dynamics minimizing g , the procedure of Section V always yields an entropy production rate that is smaller than the one of the non-detailed balanced dynamics minimizing the entropy production rate for fixed ω . However, when compared to the original dynamics, we have

$$\sigma_{\sigma|\omega} \geq \sigma_{\sigma|\omega}^* = 2\chi^* \operatorname{artanh}\left(\frac{\chi^*}{\chi_{\sigma|\omega}}\right) \leq 2\chi^* \operatorname{artanh}\left(\frac{\chi^*}{\chi}\right) = \sigma^*, \quad (\text{B9})$$

since the minimum entropy production dynamics of Ref. [12] does not always lead to a smaller activity compared to the original dynamics. As a consequence, either one of the minimum entropy production values σ^* and $\sigma_{\sigma|\omega}$ may be the smaller one. However, we expect the relation $\sigma_{\sigma|\omega} \geq \sigma^*$ to hold in most cases. The reason for this becomes apparent when looking at Eq. (B2): Both the entropy production rate and the activity are convex functionals of the parameters $C(i, j)$. Thus, minimizing the entropy production rate corresponds to choosing the parameters $C(i, j)$ in Eq. (B2b) as small as possible, while still satisfying Eq. (B2a). Except for special cases, this choice should also reduce the value of the activity, Eq. (B2c). Thus, we may expect $\chi_{\sigma|\omega} \leq \chi$ and thus $\sigma_{\sigma|\omega} \geq \sigma^*$ via Eq. (B9).

Appendix C: Minimum entropy production dynamics for finite time

In Eq. (56), we argued that the minimum entropy production for the process connecting an initial probability \mathbf{p}_i to a final probability \mathbf{p}_f is realized by a process with a constant rate of change of the probability. Here, we want to show this fact explicitly. We can divide the problem of minimizing the entropy production into two steps. First, for a given time evolution of the probability \mathbf{p}_t and activity χ_t , we minimize the entropy production rate at each instant of time. This leads to the value of σ_t^* given by Eq. (38). Second, we minimize the resulting entropy production

$$\Delta S^{\text{irr}} = \int_0^\tau dt \sigma_t^* = 2 \int_0^\tau dt \chi_t^* \operatorname{artanh}\left(\frac{\chi_t^*}{\chi_t}\right). \quad (\text{C1})$$

under the constraints that $\mathbf{p}_0 = \mathbf{p}_i$ and $\mathbf{p}_\tau = \mathbf{p}_f$ and that the total number of transitions is given by

$$\langle M \rangle = \int_0^\tau dt \chi_t. \quad (\text{C2})$$

We note that, as discussed in Section IV the minimum activity χ_t^* depends only on the time-derivative of the probability vector $d_t \mathbf{p}_t$. Next, we use Eq. (40), which allows us to construct a dynamics with $\chi_t = \mathcal{A}_t \chi_t^*$, where the choice of $\mathcal{A}_t \geq 1$ is arbitrary. This means that we can treat χ_t and χ_t^* as independent. Then, we have from Eq. (C1),

$$\Delta S^{\text{irr}} = \int_0^\tau dt \sigma_t^* = 2 \int_0^\tau dt \chi_t^* \operatorname{artanh}\left(\frac{1}{\mathcal{A}_t}\right) \quad (\text{C3})$$

with the constraint on the activity now being

$$\langle M \rangle = \int_0^\tau dt \mathcal{A}_t \chi_t^*. \quad (\text{C4})$$

First, we fix χ_t^* , and, from the Euler-Lagrange equation for the minimization with respect to \mathcal{A}_t we find

$$2 \operatorname{artanh}\left(\frac{1}{\mathcal{A}_t}\right) + \lambda \mathcal{A}_t = 0, \quad (\text{C5})$$

where λ is a Lagrange multiplier. This condition implies that $\mathcal{A}_t \equiv \mathcal{A}$ has to be independent of time. Plugging this into the above equations, we obtain

$$\Delta S^{\text{irr}} = 2 \operatorname{artanh}\left(\frac{1}{\mathcal{A}}\right) \int_0^\tau dt \chi_t^* \quad \text{and} \quad \langle M \rangle = \mathcal{A} \int_0^\tau dt \chi_t^*, \quad (\text{C6})$$

and, solving the second equation for \mathcal{A} ,

$$\Delta S^{\text{irr}} = 2 \int_0^\tau dt \chi_t^* \operatorname{artanh}\left(\frac{\int_0^\tau dt \chi_t^*}{\langle M \rangle}\right). \quad (\text{C7})$$

This is a monotonously increasing function of $\int_0^\tau dt \chi_t^*$, so minimizing it requires minimizing $\int_0^\tau dt \chi_t^* = \int_0^\tau \mathcal{W}(\mathbf{p}_{t+dt}, \mathbf{p}_t)$ under the constraint of the initial and final probability. However, we already saw in Section VI that a constant rate of change of the probability defines the geodesic in Wasserstein space, so the former indeed minimizes the entropy production and we obtain Eq. (56).

-
- [1] E. T. Jaynes, The minimum entropy production principle, *Ann. Rev. Phys. Chem.* **1980**, 579 (1980).
 - [2] C. Maes and K. Netocny, Minimum entropy production principle, *Scholarpedia* **8**, 9664 (2013), revision #134780.
 - [3] W. M. Coffey and Y. P. Kalmykov, *The Langevin equation: with applications to stochastic problems in physics, chemistry, and electrical engineering* (World Scientific, 2017).
 - [4] H. Risken, *The Fokker-Planck Equation* (Springer Berlin, 1986).
 - [5] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, Optimal protocols and optimal transport in stochastic thermodynamics, *Phys. Rev. Lett.* **106**, 250601 (2011).
 - [6] E. Aurell, K. Gawedzki, C. Mejía-Monasterio, R. Mohayaee, and P. Muratore-Ginanneschi, Refined second law of thermodynamics for fast random processes, *J. Stat. Phys.* **147**, 487 (2012).
 - [7] A. Dechant and Y. Sakurai, Thermodynamic interpretation of wasserstein distance, arXiv preprint arXiv:1912.08405 (2019).
 - [8] C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften (Springer Berlin Heidelberg, 2008).
 - [9] M. Nakazato and S. Ito, Geometrical aspects of entropy production in stochastic thermodynamics based on Wasserstein distance, arXiv preprint arXiv:2103.00503 (2021).
 - [10] A. Dechant, S. i. Sasa, and S. Ito, Geometric decomposition of entropy production in out-of-equilibrium systems (2021), arXiv:2109.12817 [cond-mat.stat-mech].
 - [11] T. Van Vu and Y. Hasegawa, Geometrical bounds of the irreversibility in markovian systems, *Phys. Rev. Lett.* **126**, 010601 (2021).
 - [12] B. Remlein and U. Seifert, Optimality of nonconservative driving for finite-time processes with discrete states, *Phys. Rev. E* **103**, L050105 (2021).
 - [13] C. Maes and K. Netočný, A nonequilibrium extension of the Clausius heat theorem, *J. Stat. Phys.* **154**, 188 (2014).
 - [14] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, Vol. 1 (Elsevier, 1992).
 - [15] C. W. Gardiner, *Handbook of stochastic methods: for physics, chemistry and the natural sciences* (Springer, 2002).
 - [16] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
 - [17] C. Villani, *Topics in optimal transportation (No. 58)* (American Mathematical Soc., 2003).
 - [18] D. B. West, *Introduction to Graph Theory*, Introduction to Graph Theory (Prentice Hall, 1996).
 - [19] N. Shiraishi, Optimal thermodynamic uncertainty relation in markov jump processes, arXiv preprint arXiv:2106.11634 (2021).
 - [20] N. Shiraishi, K. Funo, and K. Saito, Speed limit for classical stochastic processes, *Phys. Rev. Lett.* **121**, 070601 (2018).
 - [21] A. C. Barato and U. Seifert, Cost and precision of brownian clocks, *Phys. Rev. X* **6**, 041053 (2016).
 - [22] P. Pietzonka and U. Seifert, Universal trade-off between power, efficiency, and constancy in steady-state heat engines, *Phys. Rev. Lett.* **120**, 190602 (2018).
 - [23] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.* **84**, 375 (2000).
 - [24] Y. Oono and M. Paniconi, Steady state thermodynamics, *Prog. Theor. Phys. Suppl.* **130**, 29 (1998).
 - [25] T. Hatano and S.-i. Sasa, Steady-state thermodynamics of Langevin systems, *Phys. Rev. Lett.* **86**, 3463 (2001).
 - [26] T. S. Komatsu, N. Nakagawa, S.-i. Sasa, and H. Tasaki, Steady-state thermodynamics for heat conduction: microscopic derivation, *Phys. Rev. Lett.* **100**, 230602 (2008).