Thermokinetic relations

Jean-Charles Delvenne[1] and Gianmaria Falasco[2]
[1]*Institute of Information and Communication Technologies, Electronics and Applied Mathematics, UCLouvain, 1348 Louvain-La-Neuve, Belgium*
[2]*Department of Physics and Astronomy, University of Padova, Via Marzolo 8, I-35131 Padova, Italy*

(Received 16 September 2022; revised 16 October 2023; accepted 30 November 2023; published 10 January 2024)

Thermokinetic relations bound thermodynamic quantities, such as entropy production of a physical system over a certain time interval, with statistics of kinetic (or dynamical) observables, such as mean total variation of the observable over the time interval. We introduce a thermokinetic relation to bound the entropy production or the nonadiabatic (or excess) entropy production for overdamped Markov jump processes, possibly with time-varying rates and nonstationary distributions. For stationary cases, this bound is akin to a thermodynamic uncertainty relation, only involving absolute fluctuations rather than the mean square, thereby offering a better lower bound far from equilibrium. For nonstationary cases, this bound generalizes (classical) speed limits, where the kinetic term is not necessarily the activity (number of jumps) but any trajectory observable of interest. As a consequence, in the task of driving a system from a given probability distribution to another, we find a tradeoff between nonadiabatic entropy production and housekeeping entropy production: the latter can be increased to decrease the former, although to a limited extent. We also find constraints specific to constant-rate Markov processes. We illustrate our thermokinetic relations on simple examples from biophysics and computing devices.

DOI: 10.1103/PhysRevE.109.014109

I. INTRODUCTION

Consider a stochastic overdamped physical system modeled by a continuous-time Markov process—a class of models able to capture many natural or engineered systems, especially relevant at the nanoscale: protein dynamics, chemical reactions, digital electronics among others [1].

Suppose that we want to drive the state of this system of interest from a probability distribution to another one via a well-chosen protocol, i.e. a time-varing sequence of transition rates. An example to which we will return at the end of this article is the task to drive an electronic memory from a "zero" to a "one."

In general, this change of state will come with a undesirable production of entropy in the Universe, through various irreversibilities. We would therefore like to design a protocol that minimizes the entropy production while reaching the final probability distribution of interest, or find useful lower bounds on the entropy produced by any protocol fulfilling the task. Strictly speaking, this question is ill-defined, in that even in overdamped systems one can always generally reach a state distribution from another at arbitrarily low entropy production cost, provided that the time-varying dynamics is sufficiently slow. Thus, we can only hope to obtain lower bounds on entropy production relatedly to kinetic variables, in the (informal) form of: "If the driving is this fast, then the entropy production must be that high." Such bounds are called (classical) *speed limits* [2–7].

Measuring how "fast" a system evolves can be done in a number of ways, the relevance of which are context- and application-dependent, thus one cannot hope to obtain a "universal" speed limit theorem of interest, but rather one speed limit for each situation of interest. A specific measure of "fastness" that has been used in most speed limits in the literature is the total expected *activity*, i.e., the expected number of jumps of a discrete-state Markov chain over the total time interval: a fast process is meant as a process that makes few jumps to reach the desired final distribution. The activity can be seen as an intrinsic measure of time of the process.

The speed limits results come in two categories: those which bound (from below) the total entropy production along the interval, and those which bound the excess (or nonadiabatic, or Hatano-Sasa) entropy production. The nonadiabatic entropy production is the part of the entropy production that is associated to the irreversible convergence of the probability distribution towards stationarity. In particular, it is zero for a stationary (possibly nonequilibrium) process. The bounds on nonadiabatic entropy production tend to be considered as the "better" kind of speed limits [2,3], because they offer nontrivial or significantly tighter results for protocols using "nonconservative forces" (i.e., generating more entropy production than just nonadiabatic entropy production), for example protocols working slowly near nonequilibrium stationary distributions for instance.

Recently, a version relating (total) entropy production with activity was proved tight [5], in that it is achievable by a specific protocol. This protocol happens to involve conservative forces, i.e., the entropy production is entirely nonadiabatic.

In this paper, we derive a speed limit for *nonadiabatic* entropy production that is close to tightness, in that it cannot be improved by more than 18 percent. We show in particular that nonconservative forces are useful to lower nonadiabatic entropy production to its smaller possible value, in contrast with minimum entropy production protocols. We also prove that if we limit ourselves to constant (or time-symmetric) protocols, in some cases the entropy production has an in-

compressible lower bound even regardless of the time, or the activity (number of jumps) of the transformation.

While intuitive and relevant in a number of circumstances, the activity (average number of jumps) as a measure of "time" or "slowness" has the disadvantage of being sensitive to the particular model chosen to capture the physical system. Indeed, the same physical system can often be modeled at different degrees of accuracy, with wildly different numbers of states. Models that are (exactly or approximately) equivalent from a kinetic and thermodynamic viewpoint may therefore exhibit wildly different activity, thus potentially lead to very loose speed limits. It seems more relevant in general to build bounds on physical observables, whose behavior is not too dependent of an arbitrary modeling choice.

A chief contribution of this article is precisely to build speed limits that relate (total) entropy production, not just to activity, but to any observable, symmetric or antisymmetric under time-reversal, of the trajectories. They also relate nonadiabatic entropy production, not just to activity, but to any trajectory observable with some flow preservation property. These observables, depending on the context, could be charge, displacement, work, heat, etc. We believe that this is in important step in making bounds derived from stochastic thermodynamics more widely applicable to real, complex systems.

In fact, our bounds are relevant, not only to situations where one seeks to drive the state of a system from a probability distribution to another—yielding a necessarily nonstationary process—but also when the task is merely to maintain a nonequilibrium probability distribution. This makes them akin to thermodynamic uncertainty relations. The latter were introduced as an inequality between entropy production of stationary nonequilibrium systems and mean-to-standard-deviation of a time-antisymmetric observable, and subsequently extended or refined to broader contexts (such as periodic processes) [8–19]. In contrast, the present bounds involve the absolute value of a fluctuating observable, rather than the mean square or variance. Numerical and theoretical comparisons indicate that the bound on entropy production is sometimes looser and sometimes tighter than the original thermodynamic uncertainty relation [8,9], and tends to be tighter far from equilibrium.

We call our bounds thermokinetic relations, because in each case, whether they compete with speed limits or with thermodynamic uncertainty relations, they offer a relation between a thermodynamic quantity (total or nonadiabatic entropy production) and kinetic (dynamic) quantities (related to statistical quantities of the process over a given time interval, such initial and final distributions, time, total activity, or other physical observables).

The article is organized as follows. In Sec. II we revise elementary mathematical and physical concepts related to entropy production for arbitrary time-varying or constant Markovian protocols. We prove in Sec. III our main result, a thermokinetic relation for entropy production, for any observable along the trajectories, first for constant-rate Markov chains, then, via integration over infinitesimal intervals, for time-varying Markov chains. A comparison is made with the thermodynamic uncertainty relation in Sec. IV. In Sec. V we show the connection with optimal transport theory, deducing a speed limit that involves Kantorovich costs—a more general concept than total variation distances or the Wasserstein distances. Thermokinetic relations involving nonadiabatic entropy production are derived in Sec. VI.

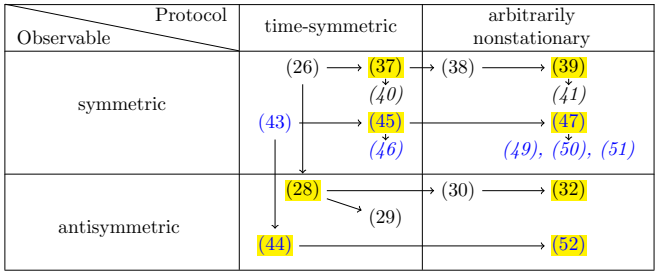

FIG. 1. Main equations derived in this paper and their logical interconnection. They are grouped on the basis of the observables and protocols required for their validity. Remember that the protocol is defined as the time sequence of transition rates of a Markov process. In particular, constant rates (but possibly nonstationary state probability distribution) is an example of time-symmetric protocol. Equation numbers in black refer to bounds on the total entropy production, equation numbers in blue refer to bounds on the nonadiabatic entropy production. For each category the main result is highlighted in yellow. For instance, Eq. (47) is the main bound on nonadiabatic entropy production for arbitrary protocols and time-symmetric observables. Results in italic are applications to the activity observable, which allows direct comparisons with the literature.

In Sec. VII, we discuss consequences and applications. One application is a biophysical molecular motor—kinesin moving along a microtubule—working in stationary conditions on which we compare our thermokinetic relations with thermodynamic uncertainty relations. Another application is the process of writing a bit in an electronic memory, modeled as rising the charge (or number of electrons) stored in a capacitance inserted in a larger, generic nonlinear electronic circuit.

In this paper a host of thermokinetic relations are derived. They relate total or nonadiabatic entropy production with time-antisymmetric observables (such as work or displacement) or time-symmetric observables (such as activity or Kantorovich costs, which include total variation distances or Wasserstein distances). They act as speed limits (nonstationary conditions) or thermodynamic uncertainty relations (stationary conditions). They apply to Markov chains with time-symmetric (e.g., constant) or arbitrary time-varying rates. All these variants are potentially useful depending on the system and task at hand. Figure 1 provides a concise summary of the main results of this paper and their relationships.

II. USEFUL CONCEPTS OF PROBABILITY THEORY AND STOCHASTIC THERMODYNAMICS

A. Kullback-Leibler divergence and total variation distance

Before any physics, we start with mathematical definitions and facts on divergences. Given an arbitrary space Ω endowed with the probability distribution p, the (ensemble) average, or expected value, of an observable $A : \Omega \to \mathbb{R}$ is denoted $\langle A \rangle_p$, or $\langle A \rangle$ if p is clear from the context. It is

defined as $\sum_{\omega\in\Omega} A(\omega)p(\omega)$. We use overall in this paper discrete sum notations for simplicity and familiarity, although a fully general notation would use integrals, such as $\int_\Omega A dp$ or $\int_\Omega A p(d\omega)$, to cover continuous situations as well.

Let us now assume that the space Ω is endowed with two probability distributions p and q, the Kullback-Leibler of p relative to q is a nonnegative, possibly infinite, real number defined as

$$D(p\|q) = \sum_{\omega\in\Omega} p(\omega) \ln \frac{p(\omega)}{q(\omega)}. \tag{1}$$

We adopt here again the sum notation for simplicity even though the definition makes sense for continuous and discrete distributions alike. In the former case, the sum is to be replaced with an integral and the ratio of probabilities with a probability density function (formally, the Radon-Nikodym derivative dp/dq).

The following well-known identity (chain rule for Kullback-Leibler divergence; see, e.g., Ref. [20]) will be repeatedly useful. Consider two probability distributions p and q over an arbitrary space Ω and an observable $X : \Omega \to \mathcal{X}$, taking values in an arbitrary set $\mathcal{X}$. Then we have the following decomposition into two nonnegative terms:

$$D(p\|q) = D(p_X\|q_X) + D(p\|q|X). \tag{2}$$

Here p_X is the probability distribution on the observable X, derived from probability distribution p on Ω as $p_X(x) = \sum_{\omega:X(\omega)=x} p(\omega)$. The conditional Kullback-Leibler divergence is defined as

$$\begin{aligned} D(p\|q|X) &= \sum_{x\in\mathcal{X}} p_X(x) D(p\|q|X=x) \\ &= \sum_{x\in\mathcal{X}} p_X(x) \sum_{\omega:X(\omega)=x} \frac{p(\omega)}{p_X(x)} \ln \frac{p(\omega)/p_X(x)}{q(\omega)/q_X(x)}. \end{aligned} \tag{3}$$

We introduce a simple bound for Kullback-Leibler divergences. Consider two arbitrary probability distributions p and q on the space Ω, and a nonnegative function $A : \Omega \to \mathbb{R}^+$ such that $p_A = q_A$ (in other words, the event $A = a$ has the same probability under p and under q, for all a), and $\langle A\rangle_p = \langle A\rangle_q > 0$. Then we can define $\varphi = A/\langle A\rangle$ as a probability density function on Ω with respect to both p and q, meaning that both φp and φq are valid probability distributions on Ω [assigning probability $\varphi(\omega)p(\omega)$ or $\varphi(\omega)q(\omega)$ to each ω, and summing to one: $\sum_{\omega\in\Omega} \varphi(\omega)p(\omega) = \sum_{\omega\in\Omega} \varphi(\omega)q(\omega) = 1$]. Then we can write

$$D(\varphi p\|\varphi q) \leqslant \varphi_{\max} D(p\|q), \tag{4}$$

where $\varphi_{\max} = A_{\max}/\langle A\rangle$ is the maximum (or supremum) of φ over Ω. In other words,

$$D(p\|q) \geqslant \frac{\langle A\rangle}{A_{\max}} D\left(\frac{Ap}{\langle A\rangle}\middle\|\frac{Aq}{\langle A\rangle}\right). \tag{5}$$

Equality holds when A takes only two values 0 and $A_{\max}$, with $p = q$ for all ω such that $A(\omega) = 0$. The proof of Eq. (4) goes as follows:

$$\begin{aligned} D(\varphi p\|\varphi q) &= D(\varphi p\|\varphi q|\varphi) \\ &= \sum_{z>0} z p(\varphi = z) D(\varphi p\|\varphi q|\varphi = z) \\ &\leqslant \varphi_{\max} \sum_{z>0} p(\varphi = z) D(\varphi p\|\varphi q|\varphi = z) \\ &= \varphi_{\max} \sum_{z>0} p(\varphi = z) \sum_{\omega:\varphi(\omega)=z} \frac{z p(\omega)}{z p(\varphi = z)} \ln \frac{z p(\omega)}{z q(\omega)} \\ &\leqslant \varphi_{\max} \sum_{z\geqslant 0} p(\varphi = z) \sum_{\omega:\varphi(\omega)=z} \frac{p(\omega)}{p(\varphi = z)} \ln \frac{p(\omega)}{q(\omega)} \\ &= \varphi_{\max} D(p\|q|\varphi) \\ &= \varphi_{\max} D(p\|q). \end{aligned} \tag{6}$$

Note that in case of an infinite space Ω, we may have φ taking arbitrarily high values with vanishingly small probabilities, in which case we can replace $\varphi_{\max}$ with an "effective" maximum; see Appendix A.

Another notion of divergence between two probability distributions is the total variation distance [20]:

$$d_{\mathrm{TV}}(p, q) = \frac{1}{2} \sum_{\omega\in\Omega} |p(\omega) - q(\omega)| \leqslant 1. \tag{7}$$

Equivalently:

$$d_{\mathrm{TV}}(p, q) = \langle \mathrm{sgn}(p - q)\rangle_p = \langle \mathrm{sgn}(q - p)\rangle_q. \tag{8}$$

Unlike the Kullback-Leibler divergence, this is a proper distance—in particular, it is symmetric and respects the triangle inequality. It is related to the Kullback-Leibler distance in the following way: We can find a convex increasing function h with $h(0) = 0$ such that for any distributions p, q on a space Ω we have

$$D(p\|q) \geqslant h[d_{\mathrm{TV}}(p, q)]. \tag{9}$$

Pinsker's inequality [20] states that $x \mapsto h(x) = 2x^2$ is such a function. Although well known, Pinsker's inequality is far from optimal. Vajda's bound [21] $h(x) = \ln \frac{1+x}{1-x} - 2\frac{x}{1+x}$ is much tighter for $d_{\mathrm{TV}}(p, q) \approx 1$ where Vajda's inequality correctly predicts an unbounded Kullback-Leibler divergence, unlike Pinsker's inequality. Even better is Gilardoni's bound [22], for

$$\begin{aligned} h(x) &= \ln \frac{1}{1-x} - (1-x)\ln(1+x) \\ &= \ln \frac{1+x}{1-x} - (2-x)\ln(1+x). \end{aligned} \tag{10}$$

Note that Pinsker's inequality remains very slightly better than Gilardoni's for small values $d_{\mathrm{TV}}(p, q) \approx 0$.

There is an optimal such convex function h^*, beating in particular Pinsker, Vajda, and Gilardoni's bounds. It is optimal in that for each $d \in [0, 1]$, there exist distributions p, q such that $d_{\mathrm{TV}}(p, q) = d$ and $D(p\|q) = h^*(d)$. Interestingly, these optimal distributions can always be chosen on a set Ω of two elements only. Despite this, h^* has no known explicit analytic expression, although it can easily be computed numerically through an implicit formula [23].

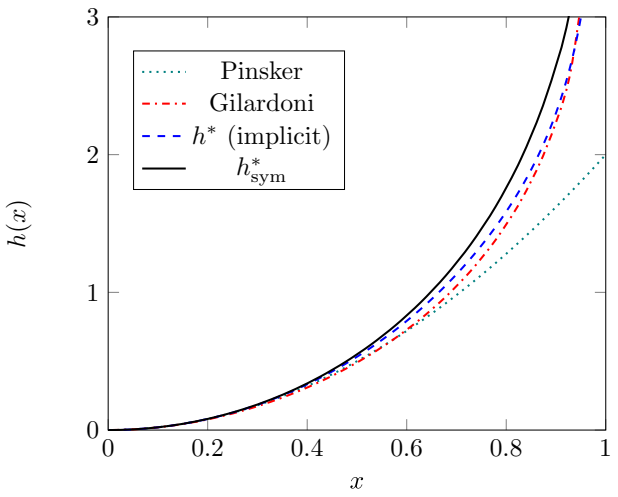

FIG. 2. The function $h(x)$ as given by Pinsker and Gilardoni, the optimal h^* given (in implicit form) by Fedotov *et al.* [23], and the tightest bound h^*_{sym} on the symmetric Kullback-Leibler divergence.

The situation is simpler for a symmetric version of Kullback-Leibler divergences, as we can write

$$\frac{D(p\|q)+D(q\|p)}{2} \geqslant h_{\mathrm{sym}}[d_{\mathrm{TV}}(p,q)]. \tag{11}$$

for some convex increasing function h_{sym}. Of course we can take any function h as above, for instance Pinsker's $h_{\mathrm{sym}}(x) = 2x^2$. Interestingly however, none of these are tight, and the tighest bound has an explicit formulation:

$$h^*_{\mathrm{sym}}(x) = x \ln \frac{1+x}{1-x} = 2x \operatorname{atanh} x, \tag{12}$$

as proved, e.g., in Refs. [7,22]. It is verified with equality for the two-element probability distributions $p = (\epsilon, 1-\epsilon)$ and $q = (1-\epsilon, \epsilon)$. In fact, it is more generally tight for any distributions p and q so that for some constant $c > 0$ and each $\omega \in \Omega$, either $|\ln p(\omega)/q(\omega)| = c$, or $q(\omega) = p(\omega) = 0$. We then have $d_{\mathrm{TV}}(p,q) = \tanh \frac{c}{2}$ and $\frac{D(p\|q)+D(q\|p)}{2} = c \tanh \frac{c}{2}$, as a direct calculation shows.

See Fig. 2 for a graphical comparison of these bounds.

B. Entropy production and its decompositions

In this section we recall some elementary concepts and well-known relations of the stochastic thermodynamics of Markov processes, written here in the language of Kullback-Leibler divergences. The goal of this section is to establish the equalities (23), (24), and (25) relating various concepts of entropy production, along with the necessary notations.

The entropy production of a time-varying Markov process over a discrete state space over a given time interval $[\tau_{\mathrm{ini}}, \tau_{\mathrm{fin}}]$ is defined as (in units such that the Boltzmann constant equals one)

$$\Delta\sigma = D(p\|\overline{p}) = \sum_{\omega\in\Omega} p(\omega) \ln \frac{p(\omega)}{\overline{p}(\omega)}, \tag{13}$$

where the sum runs over all trajectories ω of the Markov process on the time interval. Here the time-reversed probability distribution $\overline{p}(\omega)$ denotes the probability of the time-reversed trajectory $\overline{\omega}$ in the time-reversed process, whose initial state distribution is defined to be the final state distribution predicted by p, and the transition rates are defined as the transition rates of p, read backward in time (in case these rates are time-varying). We assume in this article an "overdamped" framework, where the time-reversed trajectory is simply the list of states and transitions read in reverse order.

Whenever Ω is the set of trajectories of a Markov process taking place in a state space $\mathcal{X}$, we define two observables X_{ini}, the initial state of the trajectory, and X_{fin} the final state of the trajectory. By construction of the time-reversed process $\overline{p}$ we have, in particular,

$$\begin{aligned}\overline{p}_{X_{\mathrm{fin}}}(x) &= \sum_{\omega: X_{\mathrm{fin}}(\omega)=x} \overline{p}(\omega) \\ &= \sum_{\omega: X_{\mathrm{ini}}(\overline{\omega})=x} \overline{p}(\omega).\end{aligned} \tag{14}$$

Since $\overline{p}(\omega)$ computes the probability of $\overline{\omega}$ starting with initial distribution $p_{X_{\mathrm{fin}}}$, we see by definition

$$p_{X_{\mathrm{fin}}} = \overline{p}_{X_{\mathrm{fin}}}. \tag{15}$$

When the transition rates are *constant* over time, or more generally *symmetric under time-reversal* , it is sometimes convenient to consider, as an intermediate quantity, the probability distribution $\tilde{p}(\omega) = p(\overline{\omega})$: the trajectory is reversed but not the process (i.e., we keep the same initial distribution and the same sequence of rates as in p). In particular,

$$\begin{aligned}\tilde{p}_{X_{\mathrm{fin}}}(x) &= \sum_{\omega: X_{\mathrm{fin}}(\omega)=x} \tilde{p}(\omega) \\ &= \sum_{\omega: X_{\mathrm{fin}}(\omega)=x} p(\overline{\omega}) \\ &= \sum_{\omega: X_{\mathrm{ini}}(\overline{\omega})=x} p(\overline{\omega}) \\ &= p_{X_{\mathrm{ini}}}.\end{aligned} \tag{16}$$

In summary we can write, in contrast to Eq. (15),

$$p_{X_{\mathrm{ini}}} = \tilde{p}_{X_{\mathrm{fin}}}. \tag{17}$$

Since the rates are assumed constant or time-symmetric, $\tilde{p}(\omega)$ and $\overline{p}(\omega)$ compute the probability of $\overline{\omega}$ with the same transition rates but with a different probability for the initial state of $\overline{\omega}$ (which is the final state of ω), which are $p_{X_{\mathrm{ini}}}$ and $p_{X_{\mathrm{fin}}}$ respectively.

Overall we can write, for constant-rate or time-symmetric-rate Markov chains:

$$\begin{aligned}D(p\|\overline{p}) &= \sum_{\omega\in\Omega} p(\omega) \ln \frac{p(\omega)}{\tilde{p}(\omega)} \frac{\tilde{p}(\omega)}{\overline{p}(\omega)} \\ &= D(p\|\tilde{p}) + \sum_{\omega\in\Omega} p(\omega) \ln \frac{p_{X_{\mathrm{ini}}}[X_{\mathrm{fin}}(\omega)]}{p_{X_{\mathrm{fin}}}[X_{\mathrm{fin}}(\omega)]} \\ &= D(p\|\tilde{p}) - D(p_{X_{\mathrm{fin}}}\|p_{X_{\mathrm{ini}}}) \\ &= D(p\|\tilde{p}|X_{\mathrm{fin}}) \\ &\leqslant D(p\|\tilde{p}).\end{aligned} \tag{18}$$

Equality holds for stationary Markov chains, or for infinitesimal time intervals, as over a very short interval $[\tau_{\mathrm{ini}}, \tau_{\mathrm{fin}}]$,

$D(p_{X_{\rm fin}}\|p_{X_{\rm ini}})$ is of order $(\tau_{\rm fin}-\tau_{\rm ini})^2$ (the square Fisher metric), thus is negligible before a nonzero $\Delta\sigma$ (which is typically in the order of $\tau_{\rm fin}-\tau_{\rm ini}$).

Considering now two distributions p and q on the trajectory space Ω of some Markov process (possibly with time-varying transition rates), which differ only in their initial distribution (obeying the same transition rates) one can write the Chain Rule for the initial state, finding

$$D(p\|q) = D(p_{X_{\rm ini}}\|q_{X_{\rm ini}}) + D(p\|q|X_{\rm ini}) = D(p_{X_{\rm ini}}\|q_{X_{\rm ini}}), \tag{19}$$

and for the final state,

$$D(p\|q) = D(p_{X_{\rm fin}}\|q_{X_{\rm fin}}) + D(p\|q|X_{\rm fin}). \tag{20}$$

Comparing the two equations we find

$$D(p_{X_{\rm ini}}\|q_{X_{\rm ini}}) - D(p_{X_{\rm fin}}\|q_{X_{\rm fin}}) = D(p\|q|X_{\rm fin}) \geqslant 0. \tag{21}$$

This is one standard way [20] to show that the Kullback-Leibler divergence between two state distributions driven by the same rates is nonincreasing. We can write further

$$\begin{aligned} D(p\|\overline{p}) &= D(p\|\overline{p}|X_{\rm fin}) + D(p_{X_{\rm fin}}\|\overline{p}_{X_{\rm fin}}) \\ &= D(p\|\overline{q}|X_{\rm fin}) \\ &= D(p\|q|X_{\rm fin}) + [D(p\|\overline{q}|X_{\rm fin}) - D(p\|q|X_{\rm fin})] \\ &= D(p\|q|X_{\rm fin}) + [D(p\|\overline{q}) - D(p\|q)]. \end{aligned} \tag{22}$$

We have used Eq. (15)—thus the facts that $p_{X_{\rm fin}} = \overline{p}_{X_{\rm fin}}$ and $q_{X_{\rm fin}} = \overline{q}_{X_{\rm fin}}$—and the fact that $\overline{q}(\omega)$ and $\overline{p}(\omega)$ are identical conditional to the final state of ω (as they obey the same time-varying rates).

Although this decomposition is true for any q obeying the same time-varying rates as p, it takes a particularly significant meaning when rates are constant in time, or time-symmetric, and q is a stationary process obeying those rates ($q_{X_{\rm fin}} = q_{X_{\rm ini}}$, thus $\overline{q} = \tilde{q}$). In this case $D(p\|q|X_{\rm fin})$ is called the *nonadiabatic entropy production* $\Delta\sigma_{\rm NA}$ [24] or excess entropy production [25] or Hatano-Sasa entropy production [26], and is zero if and only if p is also stationary.

Furthermore in this case $D(p\|\overline{q}) - D(p\|q) = D(p\|\overline{q}|X_{\rm ini}) \geqslant 0$ is called the *housekeeping entropy production* $\Delta\sigma_{\rm HK}$, or adiabatic entropy production [24], which is zero if and only if detailed balance is satisfied, i.e., if the q is an equilibrium, i.e., time-reversible ($q = \overline{q}$), i.e., driven by conservative forces.

In summary we have proved the following decomposition [24] for entropy production $\Delta\sigma = D(p\|\overline{p})$ on a constant-rate or time-symmetric-rate Markov process with q as the corresponding stationary process:

$$\begin{aligned} \Delta\sigma &= \Delta\sigma_{\rm NA} + \Delta\sigma_{\rm HK} \\ &= D(p\|q|X_{\rm fin}) + D(p\|\overline{q}|X_{\rm ini}). \end{aligned} \tag{23}$$

In addition, we have proved, in the same context,

$$\Delta\sigma = D(p\|\tilde{p}) - D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}). \tag{24}$$

Finally, another well-known observation is that for any set of trajectories Ω, we find

$$\begin{aligned} D(p\|\tilde{p}) &= \frac{1}{2}\sum_{\omega\in\Omega}(p(\omega) - p(\overline{\omega}))\ln\frac{p(\omega)}{p(\overline{\omega})} \\ &= \frac{D(p\|\tilde{p}) + D(\tilde{p}\|p)}{2}. \end{aligned} \tag{25}$$

To evaluate the entropy productions in case of a general continuous-time discrete-state Markov chain with time-varying rates, one can decompose the total time interval into successive infinitesimal time intervals $\tau_{\rm fin} - \tau_{\rm ini} = dt$, where the rates can be considered constant, and sum the entropy production computed over each interval. Note that in this process, the terms $D(p_{X_{\rm fin}}\|p_{X_{\rm ini}})$, of the negligible order $\mathcal{O}(dt^2)$, vanish. Recall as well that in an interval dt, the trajectory ω of a Markov jump process either is constant [no transition, with probability $1-\mathcal{O}(dt)$], or consists in a single transition between initial and final state [with probability $\mathcal{O}(dt)$]. We can safely neglect the possibility of two or more transitions in an interval dt, as it occurs with negligible probability $\mathcal{O}(dt^2)$.

III. THERMOKINETIC RELATIONS FOR ENTROPY PRODUCTION

A. Constant or time-symmetric rates

Let A be a time-symmetric nonnegative function on trajectories in Ω. By time-symmetric, we mean symmetric under time reversal, i.e., $A(\omega) = A(\overline{\omega})$ for all trajectories ω. Thus, for a constant-rate or time-symmetric-rate Markov process over arbitrary time intervals, we find, following Eqs. (24), (25), (5), (11), and (12),

$$\begin{aligned} \Delta\sigma + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) &\geqslant \frac{\langle A\rangle}{A_{\max}} h^*_{\rm sym}\left(d_{\rm TV}\left(\frac{Ap}{\langle A\rangle}, \frac{A\tilde{p}}{\langle A\rangle}\right)\right) \\ &= \frac{\langle A\rangle}{A_{\max}} h^*_{\rm sym}\left(\frac{\langle A\ \mathrm{sgn}(p-\tilde{p})\rangle}{\langle A\rangle}\right) \\ &= 2\frac{\langle A\ \mathrm{sgn}(p-\tilde{p})\rangle}{A_{\max}}\,\mathrm{atanh}\,\frac{\langle A\ \mathrm{sgn}(p-\tilde{p})\rangle}{\langle A\rangle}. \end{aligned} \tag{26}$$

This is our first thermokinetic relation for constant-rate or time-symmetric-rate Markov chains, relating entropy production to nonnegative symmetric observables. Recall that $h^*_{\rm sym}$ is the convex increasing function defined by Eq. (12).

Given *any* observable $f : \Omega \to \mathbb{R}$ on trajectories, we can build the nonnegative time-symmetric observable $A(\omega) = \frac{|f(\omega)|+|f(\overline{\omega})|}{2}$. Thus, we get, for any trajectory observable $f : \Omega \to \mathbb{R}$:

$$\Delta\sigma + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \geqslant \frac{\langle |f|\rangle}{|f|_{\max}} h^*_{\rm sym}\left(\frac{\left\langle \frac{|f|+|\overline{f}|}{2}\mathrm{sgn}(p-\tilde{p})\right\rangle}{\langle |f|\rangle}\right). \tag{27}$$

This implies a particularly simple form for time- *antisymmetric* [also called *current-like,* i.e., such that $f(\omega) = -f(\overline{\omega})$]

observable f:

$$\Delta\sigma + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \geqslant \frac{\langle|f|\rangle}{f_{\max}} h^*_{\rm sym}\left(\frac{|\langle f\rangle|}{\langle|f|\rangle}\right) = 2\frac{|\langle f\rangle|}{f_{\max}}\operatorname{atanh}\frac{|\langle f\rangle|}{\langle|f|\rangle}. \tag{28}$$

This is equivalent to the thermokinetic relation (26) above, expressed for time-antisymmetric trajectory observables instead. This is one of the main results of this article.

The right-hand side (r.h.s.) is increasing in the $|\langle f\rangle|$ argument, and is decreasing in the $\langle|f|\rangle$ argument and in the $f_{\max}$ argument. From these observations we may infer new valid inequalities, e.g., by replacing $\langle|f|\rangle$ with an upper bound in the r.h.s. For example, the trivial bound $\langle|f|\rangle \leqslant f_{\max}$ leads to the simpler, but weaker, thermokinetic relation:

$$\Delta\sigma + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \geqslant h^*_{\rm sym}\left(\frac{|\langle f\rangle|}{f_{\max}}\right). \tag{29}$$

B. General time-varying Markov processes

Let us now turn to the general case of time-varying rates over an arbitrary interval. We can decompose a time-varying Markov process into a succession of constant-rate Markov processes over infinitesimal intervals.

Let us first spell out what Eq. (28) becomes for an infinitesimal interval. Over each infinitesimal interval $[t, t+dt]$, we can consider that at most one transition occurs, as two transitions only occur with vanishing probability $\mathcal{O}(dt^2)$. Thus, the trajectories ω are either constant (no transition at all, with the system sitting at some state x) or with one transition. In the usual graph-theoretic representation of a discrete-state Markov chain [27], this transition is represented by one (directed) edge e from state (vertex) x to state (vertex) x'. In this context, an antisymmetric observable $f(\omega)$ over $[t, t+dt]$ takes value 0 if no transition happens, and some value $f(e)$ if ω is the transition e. Thus, f is described by a real number $f(e)$ attached to each edge e, so that $f(\bar{e}) = -f(e)$ (time-antisymmetry). The probability $p(e)$, i.e., the probability that $\omega = e$, also called the flow over e, is $p_{X_t}(x)r_t(x \to x')dt$, where $r_t(x \to x')$ is called the transition rate along e. The flow over e is an infinitesimal quantity, i.e., proportional to dt. Thus, although f takes real (noninfinitesimal) values on all transitions at all times, $\langle f\rangle = \sum_e f(e)p(e)$ is an infinitesimal quantity, proportional to dt. The same is true for $\langle|f|\rangle$ or $\langle A\rangle$. Thus, both sides of Eq. (28), for infinitesimal time interval dt are infinitesimal in the order of dt. Note finally that the term $D(p_{X_{\rm fin}}\|p_{X_{\rm ini}})$ of the order dt^2, thus negligible.

For a time-varying Markov chain over a noninfinitesimal interval $[\tau_{\rm ini}, \tau_{\rm fin}]$, with possibly time-varying real labels on edges $f_t(e)$ at each interval $[t, t+dt]$, a trajectory ω is a succession of (finitely many) transitions taking place over the interval, and we assign to ω the sum of all real numbers f labeling those transitions, which we can denote $\int f$ (recalling that this integral reduces to a sum, as on all but finitely many infinitesimal intervals we have $f = 0$). The expected value of this trajectory observable is $\langle\int f\rangle = \int\langle f\rangle$. Recall that $\langle f\rangle$ is an infinitesimal quantity, proportional to dt, thus $\int\langle f\rangle$ is indeed an integral in the usual sense.

We may now integrate Eq. (28) on the total interval $[\tau_{\rm ini}, \tau_{\rm fin}]$, with the following bound on the entropy production $\Delta\sigma$ over the total interval:

$$\Delta\sigma \geqslant \int \frac{\langle|f|\rangle}{f_{\max}} h^*_{\rm sym}\left(\frac{|\langle f\rangle|}{\langle|f|\rangle}\right) = 2\int \frac{|\langle f\rangle|}{f_{\max}}\operatorname{atanh}\frac{|\langle f\rangle|}{\langle|f|\rangle}. \tag{30}$$

The convexity of $h^*_{\rm sym}$ implies Jensen's inequality

$$\frac{\int w(t)h^*_{\rm sym}(g(t))dt}{\int w(t)dt} \geqslant h^*_{\rm sym}\left(\frac{\int w(t)g(t)dt}{\int w(t)dt}\right) \tag{31}$$

for any nonnegative weight function w and any function g (taking values between 0 and 1, where $h^*_{\rm sym}$ is defined). The triangle inequality also implies $\int|\langle f\rangle| \geqslant |\int\langle f\rangle|$. Applying these inequalities to Eq. (30), we find the thermokinetic relation for general time-varying (overdamped) Markov chains, the second central result of this article:

$$\Delta\sigma \geqslant \frac{\int|\langle f\rangle|}{f_{\max}}\, h^*_{\rm sym}\left(\frac{|\int\langle f\rangle|}{\int\langle|f|\rangle}\right) = 2\frac{|\int\langle f\rangle|}{f_{\max}}\,\operatorname{atanh}\frac{|\int\langle f\rangle|}{\int\langle|f|\rangle}. \tag{32}$$

Here $f_{\max}$ is the maximum (or supremum) value of $f(e)$ over all transitions e, and over all times t if the observable f is time-varying.

Note that although Eq. (32) is a weaker inequality than Eq. (30), it has the advantage of involving $\int\langle f\rangle$ and $\int\langle|f|\rangle$, which typically have a direct physical interpretation as we shall see in the examples. Moreover, Eq. (32) can be met with equality in some important cases as we detail hereafter.

IV. COMPARISON WITH THE THERMODYNAMIC UNCERTAINTY RELATION

In the stationary case (probability on state distribution constant over time, with in particular $p_{X_{\rm fin}} = p_{X_{\rm ini}}$), this relation (32) can be compared with the standard thermodynamic uncertainty relation:

$$\Delta\sigma \geqslant 2\frac{\langle f\rangle^2}{\mathrm{Var} f} = 2\frac{\langle f\rangle^2}{\langle f^2\rangle}, \tag{33}$$

here expressed over an infinitesimal time interval $\tau_{\rm fin} - \tau_{\rm ini} = dt$, where the relation is the tightest [19]. In this limit, $\langle f\rangle^2$ is of the order $\mathcal{O}(dt^2)$ thus is negligible compared to $\mathrm{Var} f = \langle f^2\rangle - \langle f\rangle^2$.

To compare Eqs. (32) and (33) fruitfully, we first consider time-antisymmetric observables f that can only take three values $f(e) = f_{\max}$, $f(e) = -f_{\max}$ or $f(e) = 0$ for any transition e. In that case we find $\langle f^2\rangle = f_{\max}\langle|f|\rangle$. Using the fact that $\operatorname{atanh} x > x$ for any $0 < x < 1$, we see that our thermokinetic relation (32) is always tighter than the thermodynamic uncertainty relation (33). They tend to be equally tight when the argument of atanh in Eq. (32) is small, e.g., close to equilibrium $[p(e) \approx p(\bar{e})]$.

Nevertheless for other observables, Eq. (33) may prove a better bound on the entropy production than Eq. (32). To

show this, we consider in Appendix B the best possible choice of f for Eq. (33), i.e., with the maximum mean-to-standard-deviation ratio, which makes (33) as tight as can be. This is the *hyperaccurate current* in the terminology of Ref. [28]. With this optimal choice, our thermokinetic relation (32) can be weaker (when close to equilibrium, $p \approx \tilde{p}$) or tighter (when far from equilibrium, $p/\tilde{p} \to 0$ or ∞) than the classic (33). It has the added advantage that it is valid for nonstationary cases as well.

In Sec. VII C we will offer a numerical illustration of our thermokinetic relation (28) and the thermodynamic uncertainty relation (33) on a biophysical example—kinesin moving along a microtubule.

V. SPEED LIMITS FOR GENERAL OPTIMAL TRANSPORT COSTS, INCLUDING THE TOTAL VARIATION DISTANCE

The relations above are dependent on the particular protocol driving the initial distribution towards the final distribution. It can be useful however to remove the dependence of the formula above on the specifics of the particular process, to get an absolute lower bound on the entropy production cost required to drive the system from a given state distribution p_{ini} to another state distribution p_{fin}. As explained in the Introduction, one must however constrain how "fast" the trajectory is, to obtain a useful tradeoff, and this can be done in various ways.

The chief contribution of this section is precisely to exploit the thermokinetic relations derived above to offer a flexible family of speed limits based on statistics of physical observables.

Consider that A, a time-symmetric nonnegative weight assigned to every trajectory ω in an arbitrary trajectory space Ω, represents the "cost" of ω. For each state $x \in X$ we assume we have a zero-cost trajectory ω_x going from x to x (e.g., the constant trajectory).

Given two probability distributions p_{ini} and p_{fin} on the state space X, we may now look for the probability distribution $p^*(\omega)$ on trajectories in Ω that drives the states from p_{ini} to p_{fin} with minimal cost $\langle A \rangle_{p^*}$. In other words, we want to pick p^* on Ω so that $p^*_{X_{\text{ini}}} = p_{\text{ini}}$ and $p^*_{X_{\text{fin}}} = p_{\text{fin}}$ (and no other constraint: for instance we do not require here that p^* derives from a Markovian process), while achieving the minimum possible value for $\langle A \rangle$. Finding such an optimal probability distribution p^* is essentially the standard *optimal transport* problem, as introduced by Monge [29] and by Kantorovich [30] and later extended by Rubinstein, Wasserstein, Dobrushin, etc. [31–33] We call the optimal average cost $\langle A \rangle_{p^*}$ the *Kantorovich cost* $K_A(p_{\text{ini}}, p_{\text{fin}})$ (also called earth mover's, or optimal transport cost in the literature, see Appendix C). Note that K_A is not necessarily a proper distance on the space on probability distributions on the state space X, although it is for many reasonable choices of Ω and A. More background and detail on optimal transport, comparing the general concept of Kantorovich cost with more specific concepts such as the 1-Wasserstein distance, and the total variation distance in particular, is provided in Appendix C.

Clearly, the average cost given by any probability distribution p on trajectories in Ω is an upper bound on the Kantorovich cost between the initial and final state distributions $p_{\text{ini}} = p_{X_{\text{ini}}}$ and $p_{\text{fin}} = p_{X_{\text{fin}}}$:

$$\langle A \rangle_p \geqslant K_A(p_{X_{\text{ini}}}, p_{X_{\text{fin}}}). \tag{34}$$

We can generally obtain a better bound on $K_A(p_{X_{\text{ini}}}, p_{X_{\text{fin}}})$ by "improving" the probability distribution p. From p we construct p_{red}, a *reduced* probability distribution on trajectories, in the following way. For each ω from x to $x' \neq x$ such that $p(\omega) > p(\overline{\omega})$, set $p_{\text{red}}(\omega) = p(\omega) - p(\overline{\omega})$, and $p_{\text{red}}(\overline{\omega}) = 0$. We also pick a zero-cost trajectory ω_x from x to x assigning $p_{\text{red}}(\omega_x) = p(\omega_x) + p(\overline{\omega})$. Similarly, we pick a zero-cost trajectory $\omega_{x'}$ from x' to x' and set $p_{\text{red}}(\omega_{x'}) = p(\omega_{x'}) + p(\overline{\omega})$. In other words, we "cancel" the flow along opposite directions and redirect the corresponding probability flow on self-loops. This keeps the probability distribution of initial and final states unperturbed, while decreasing the expected cost of trajectories. Thus, we can write

$$\langle A \rangle_p \geqslant \langle A \rangle_{p_{\text{red}}} \geqslant K_A(p_{X_{\text{ini}}}, p_{\text{fin}}). \tag{35}$$

The middle quantity can be rewritten as

$$\begin{aligned} \langle A \rangle_{p_{\text{red}}} &= \frac{1}{2} \sum_{\omega} |Ap(\omega) - Ap(\overline{\omega})| \\ &= \langle A \rangle_p d_{\text{TV}}\left(\frac{Ap}{\langle A \rangle_p}, \frac{A\tilde{p}}{\langle A \rangle_p} \right). \end{aligned} \tag{36}$$

Thus, we can weaken Eq. (26) and obtain the following speed limit for constant-rate or time-symmetric-rate Markov chains:

$$\Delta\sigma + D(p_{X_{\text{fin}}} \| p_{X_{\text{ini}}}) \geqslant \frac{\langle A \rangle_p}{A_{\max}} h^*_{\text{sym}}\left(\frac{K_A(p_{\text{ini}}, p_{\text{fin}})}{\langle A \rangle_p} \right). \tag{37}$$

We now consider arbitrary time-varying Markov chains. A trajectory can be decomposed into constant-rate Markov chains over infinitesimal time intervals, where at most one transition happens. The cost on each interval, written as A, is thus a cost assigned to each possible transition (as we request that constant paths have zero cost). This cost can possibly be time-varying. Integrating Eq. (37) over the trajectory on the whole time interval, and using convexity of h [proceeding as for the derivation of Eq. (32)]:

$$\Delta\sigma \geqslant \frac{\int \langle A \rangle}{A_{\max}} h^*_{\text{sym}}\left(\frac{\int dK_A}{\int \langle A \rangle} \right), \tag{38}$$

where $\int dK_A$ is the integral of costs $K_A(p_{X_t, X_{t+dt}})$ over the whole trajectory from p_{ini} to p_{fin}. Here $A_{\max}$ is the maximum value over all transitions and times, and $\int \langle A \rangle$ is the expected sum of costs of all transitions over the trajectory.

The quantity $\int dK_A$, which depends on all intermediate state probability distributions p_{X_t}, can itself be bounded from below by a Kantorovich cost (depending only on p_{ini} to p_{fin}) in the following way. Let us assign a cost denoted $\int A(\omega)$ to a trajectory ω over the whole interval $[\tau_{\text{ini}}, \tau_{\text{fin}}]$, defined as the integral of costs over infinitesimal transitions (i.e., the sum of cost of each transition along the trajectory, as we request that constant paths have zero cost). Then $\int \langle A \rangle$ is the mean value of this trajectory cost. As a result, the corresponding Kantorovich cost $K_{\int A}(p_{\text{ini}}, p_{\text{fin}})$ is no larger than the integral of Kantorovich costs over infinitesimal intervals. Note that if the cost function A is different for each time interval, then in

general the Kantorovich cost $K_{\int A}$ is not necessarily a *distance* —which does not hamper our results.

Thus, we may write

$$\Delta\sigma \geqslant \frac{\int\langle A\rangle}{A_{\max}}\; h^*_{\rm sym}\left(\frac{K_{\int A}(p_{\rm ini},p_{\rm fin})}{\int\langle A\rangle}\right). \tag{39}$$

This is one of the main results of this article. This speed limit has the advantage that $K_{\int A}(p_{\rm ini},p_{\rm fin})$ does not depend on the particular process driving $p_{\rm ini}$ to $p_{\rm fin}$. The quantity $\int\langle A\rangle$ is the expected total (integrated) cost of the whole trajectory.

An important example of the above is the *activity speed limit* [2–7]. Consider a connected Markov chain (with transitions possible between every pairs of states), and $A(\omega)=1$ if at least one transition occurs along the trajectory ω, and $A(\omega)=0$ otherwise (constant trajectory, no transitions). It is well known that the corresponding Kantorovich cost is precisely the total variation distance $d_{\rm TV}$ (see, e.g., Ref. [31]). Thus, for any constant-rate or time-symmetric-rate Markov chain over any time interval we find

$$\Delta\sigma + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \geqslant \langle A\rangle_p h^*_{\rm sym}\left(\frac{d_{\rm TV}(p_{\rm ini},p_{\rm fin})}{\langle A\rangle_p}\right), \tag{40}$$

where $\langle A\rangle_p$ is here the probability that at least one transition occurs. For an arbitrary time-varying Markov chain we find

$$\begin{aligned}\Delta\sigma &\geqslant \int\langle A\rangle_p\; h^*_{\rm sym}\left(\frac{d_{\rm TV}(p_{\rm ini},p_{\rm fin})}{\int\langle A\rangle_p}\right)\\ &= 2d_{\rm TV}(p_{\rm ini},p_{\rm fin})\,\operatorname{atanh}\frac{d_{\rm TV}(p_{\rm ini},p_{\rm fin})}{\int\langle A\rangle_p},\end{aligned} \tag{41}$$

where $\int\langle A\rangle_p$ is the total expected activity, i.e., the expected number of transitions (jumps) along the interval $[\tau_{\rm ini},\tau_{\rm fin}]$. This number of transitions is the kinetic parameter indicative of the slowness of the dynamics: Many steps needed to reach a given final distribution is thought of as a slow dynamics. Said otherwise, $\delta t=(\tau_{\rm fin}-\tau_{\rm ini})/\int\langle A\rangle_p$ can be seen as the typical time between two transitions, a natural timescale of the system, and the activity can be written as $(\tau_{\rm fin}-\tau_{\rm ini})/\delta t$, is a dimensionless measure of the duration of the process. Note that this formula was obtained in Ref. [5] (among with a specific family of Kantorovich costs, as explained in Appendices C and D).

In comparison, the general formula (38) allows arbitrary (symmetric, nonnegative, time-varying) costs for each transition of the symmetric weights. This allows to possibly model a physical cost of interest associated to the transition, and can mitigate the intrinsic dependence of activity on an arbitrary coarse graining level of the model at hand.

VI. BOUNDING NONADIABATIC ENTROPY PRODUCTION

The total entropy $\Delta\sigma$, used in the speed limits above, decomposes into nonadiabatic (NA) entropy and housekeeping entropy (22). Only the NA entropy is associated to the process of convergence to stationarity, while the housekeeping entropy may be non zero even at stationarity. In case of nonconservative forces, Eq. (32) is thus typically far from being tight, especially at stationarity [where the r.h.s. side is zero and the left-hand side (l.h.s.) is strictly positive] or around it.

It is therefore preferable to bound $\Delta\sigma_{\rm NA}$ (which cancels at stationarity) rather than $\Delta\sigma$, to obtain useful relations with kinetic quantities that also cancel at stationarity [2,3].

A. Constant or time-symmetric rates

Let us first focus on the case of constant-rate or time-symmetric-rate Markov processes, with corresponding stationary distribution q on trajectories. Assume we have a nonnegative symmetric observable A on the trajectories, i.e., a map $A:\Omega\to\mathbb{R}^+$ such that $A(\omega)=A(\overline{\omega})$, for any trajectory ω. We also assume $\langle A\rangle_q>0$ for non triviality. In particular, we have for instance $q_A=\tilde{q}_A=\overline{q}_A$. However, we know from stationarity that $q_{X_{\rm fin}}=\tilde{q}_{X_{\rm fin}}=\overline{q}_{X_{\rm fin}}=q_{X_{\rm ini}}$. Now assume that moreover A is such that the equality holds for the joint observable $(A,X_{\rm fin})$, i.e., $q_{AX_{\rm fin}}=\tilde{q}_{AX_{\rm fin}}=\overline{q}_{AX_{\rm fin}}$. In other words, even though the stationary process might not satisfy detailed balance ($q\neq\overline{q}$), we have "apparent" detailed balance from the joint knowledge of A and $X_{\rm fin}$. Let us restate this property in another way. For each possible value of $A=z$ of such an observable, there is a conservation law at each state x: the sets $\{\omega\in\Omega: A(\omega)=z \text{ and } X_{\rm ini}(\omega)=x\}$ and $\{\omega\in\Omega: A(\omega)=z \text{ and } X_{\rm fin}(\omega)=x\}$ have same probability. In other words, the global probability flow conservation at stationarity splits into several probability conservation laws on several sets of trajectories, each corresponding to a possible value of A.

This is the case for instance for the "activity" observable $A:\Omega\to\{0,1\}$ already used in Sec. V mapping a nonconstant (with at least one transition) trajectory ω to 1 and to 0 if ω is constant (no transition). Indeed the probability $q(A=1,X_{\rm fin}=x)$ records the probability flow entering x over the trajectory (from any initial state than x). However, $\tilde{q}(A=1,X_{\rm fin}=x)$ is the probability flow leaving the initial state x. From stationarity, these quantities must be equal.

Other practically relevant observables than activity may be eligible, depending on the specificities of the system at hand, as we see later in the case of electronic memories.

To prove a lower bound on NA entropy production, we write the following:

$$\begin{aligned}D(p\|q|X_{\rm fin}) &\geqslant D(p_{AX_{\rm fin}}\|q_{AX_{\rm fin}}|X_{\rm fin})\\ &= D(p_{AX_{\rm fin}}\|\overline{q}_{AX_{\rm fin}}|X_{\rm fin})\\ &= D(p_{AX_{\rm fin}}\|\overline{p}_{AX_{\rm fin}}|X_{\rm fin})\\ &= D(p_{AX_{\rm fin}}\|\tilde{p}_{AX_{\rm fin}}|X_{\rm fin}).\end{aligned} \tag{42}$$

We used here the fact that $q_{AX_{\rm fin}}=\overline{q}_{AX_{\rm fin}}$ and that $\overline{q}$, $\overline{p}$ and $\tilde{p}$ coincide when knowing the value of $X_{\rm fin}$ (initial state of the reversed path). We find by repeated use of the chain rule,

Eqs. (42), (5), and (9),

$$\begin{aligned}
&\Delta\sigma_{\rm NA} + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \\
&\quad\geqslant D(p_{AX_{\rm fin}}\|\tilde{p}_{AX_{\rm fin}}|X_{\rm fin}) + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \\
&\quad= D(p_{AX_{\rm fin}}\|\tilde{p}_{AX_{\rm fin}}|X_{\rm fin}) + D(p_{X_{\rm fin}}\|\tilde{p}_{X_{\rm fin}}) \\
&\quad= D(p_{AX_{\rm fin}}\|\tilde{p}_{AX_{\rm fin}}) \\
&\quad\geqslant \frac{\langle A\rangle_p}{A_{\rm max}} D\left(\frac{Ap_{AX_{\rm fin}}}{\langle A\rangle_p}\|\frac{A\tilde{p}_{AX_{\rm fin}}}{\langle A\rangle_p}\right) \\
&\quad= \frac{\langle A\rangle_p}{A_{\rm max}} D\left(\frac{Ap_{AX_{\rm fin}}}{\langle A\rangle_p}\|\frac{Ap_{AX_{\rm ini}}}{\langle A\rangle_p}\right) \\
&\quad= \frac{\langle A\rangle_p}{A_{\rm max}} h\left(d_{\rm TV}\left(\frac{Ap_{AX_{\rm fin}}}{\langle A\rangle_p}, \frac{Ap_{AX_{\rm ini}}}{\langle A\rangle_p}\right)\right) \\
&\quad= \frac{\langle A\rangle_p}{A_{\rm max}} h\left(\frac{\langle A\ \mathrm{sgn}(p_{AX_{\rm fin}} - p_{AX_{\rm ini}})\rangle}{\langle A\rangle_p}\right). \qquad (43)
\end{aligned}$$

This thermokinetic relation involves the nonadiabatic entropy production, for constant-rate or time-symmetric-rate Markov chains. Remember that h can taken as Pinsker's or Gilardoni's bound (10), or the optimal (nonexplicit) optimal bound h^*, as pictured in Fig. 2. We can reformulate it in terms of time-antisymmetric observables, similarly to Eq. (28). Let f be an antisymmetric observable on trajectories, such that at every state x and for every possible value $f = z$, the probability flow carried by trajectories of value z entering x equals, for the stationary flow q, the probability flow carried by trajectories of value $-z$ out of x. Then we find our main thermokinetic relation in this section:

$$\Delta\sigma_{\rm NA} + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \geqslant \frac{\langle|f|\rangle}{f_{\rm max}} h\left(\frac{|\langle f\rangle|}{\langle|f|\rangle}\right). \qquad (44)$$

We can also bring Eq. (43) to an optimal transport interpretation, still for constant or time-symmetric rates:

$$\begin{aligned}
&\Delta\sigma_{\rm NA} + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \\
&\qquad\geqslant \frac{\langle A\rangle_p}{A_{\rm max}} h\left(\frac{K_A(p_{X_{\rm ini}}, p_{X_{\rm fin}})}{\langle A\rangle_p}\right). \qquad (45)
\end{aligned}$$

In particular, for A the activity observable, i.e., when $\langle A\rangle_p$ is the probability that at least one transition occurs in the time interval, we find

$$\begin{aligned}
&\Delta\sigma_{\rm NA} + D(p_{X_{\rm fin}}\|p_{X_{\rm ini}}) \\
&\qquad\geqslant \frac{\langle A\rangle_p}{A_{\rm max}} h\left(\frac{d_{\rm TV}(p_{X_{\rm ini}}, p_{X_{\rm fin}})}{\langle A\rangle_p}\right). \qquad (46)
\end{aligned}$$

B. General time-varying Markov chains: The speed limit for nonadiabatic entropy production

Let us now turn to the general case of time-varying rates over an arbitrary interval. We can decompose a time-varying Markov process into a succession of constant-rate Markov processes over infinitesimal intervals. Over each infinitesimal interval dt, we can thus use Eq. (45), neglecting the subleading term $D(p_{X_{\rm fin}}\|p_{X_{\rm ini}})$ term. Integrating Eq. (45) on the total interval, and using convexity of h, we find

$$\begin{aligned}
\Delta\sigma_{\rm NA} &\geqslant \frac{\int\langle A\rangle}{A_{\rm max}}\ h\left(\frac{\int dK_A}{\int\langle A\rangle}\right) \\
&\geqslant \frac{\int\langle A\rangle}{A_{\rm max}}\ h\left(\frac{K_{\int A}(p_{\rm ini}, p_{\rm fin})}{\int\langle A\rangle}\right), \qquad (47)
\end{aligned}$$

where $\int dK_A$ is the integrated Kantorovich cost along the trajectory as in Eq. (32), and can be weakened to the trajectory-independent quantity $K_{\int A}(p_{\rm ini}, p_{\rm fin})$ as in Eq. (39). This thermokinetic relation is arguably the most central result of this article, and can also be called a speed limit.

It is also valid for time-varying observables A, as long as the conservation condition in Sec. VI A is valid at all times. In this case $A_{\rm max}$ is the maximum over all transitions over all times.

In the case of the activity observable, we obtain

$$\Delta\sigma_{\rm NA} \geqslant \int\langle A\rangle\ \ h\left(\frac{d_{\rm TV}(p_{\rm ini}, p_{\rm fin})}{\int\langle A\rangle}\right), \qquad (48)$$

where $\int\langle A\rangle$ is the total expected activity, i.e., the expected number of jumps in the whole time interval Δt. This is a novel bound on nonadiabatic (excess) entropy production. Recall that here h can be taken as $h(x) = 2x^2$ (Pinsker's bound), leading to

$$\Delta\sigma_{\rm NA} \geqslant 2\frac{d_{\rm TV}^2(p_{\rm ini}, p_{\rm fin})}{\int\langle A\rangle}. \qquad (49)$$

This is essentially the main result of Ref. [3], itself a refinement of Ref. [2]. The new bound (51) is substantially tighter, using e.g., Gilardoni's bound (10). Even slightly better is the bound obtained with $h = h^*$, although it has no simple explicit expression. All these bounds coincide in the limit of long times (high activity, slow driving). In the limit of low times or low activity (fast driving), Gilardoni's classical speed limit is unboundedly better than Pinsker's. Indeed it correctly predicts that the activity (expected number of jumps) to follow the trajectory must satisfy

$$\int\langle A\rangle > d_{\rm TV}(p_{\rm ini}, p_{\rm fin}), \qquad (50)$$

with an infinite NA entropy production at the limit case $\langle A\rangle = d_{\rm TV}$. In constrast, Pinsker's version only provides a finite lower bound.

Note that if the Markov chain transitions are known to take place on a certain graph, then we can improve (51) by replacing the total variation distance $d_{\rm TV}(p_{\rm ini}, p_{\rm fin})$ with the 1-Wasserstein distance on state distributions $W_1(p_{\rm ini}, p_{\rm fin})$ induced by the graph distance between states, see Appendix C for proof and explanations on 1-Wasserstein distances (which are special examples of Kantorovich costs):

$$\Delta\sigma_{\rm NA} \geqslant \int\langle A\rangle\ \ h\left(\frac{W_1(p_{\rm ini}, p_{\rm fin})}{\int\langle A\rangle}\right). \qquad (51)$$

This novel bound is the counterpart for nonadiabatic entropy production of the bounds developed in Ref. [5] for total entropy production. The total variation distance corresponds to the case of the complete graph, where all transitions are *a priori* allowed.

Finally, we also have a time-varying version of the thermokinetic relation bound (44) for antisymmetric observable f on transitions satisfying at all times the flow conservation condition:

$$\Delta\sigma_{\rm NA} \geqslant \frac{\int \langle |f| \rangle}{f_{\max}} h\left(\frac{|\int \langle f \rangle|}{\int \langle |f| \rangle} \right). \tag{52}$$

This form is perhaps the most practically convenient to use, as illustrated on the electronic memory below.

VII. DISCUSSION AND APPLICATIONS

A. Tightness of thermokinetic relations and the interest of nonconservative forces

It is known [5] that the speed limit (41) is tight for (total) entropy production and activity observable. This means that given $p_{\rm ini}$, $p_{\rm fin}$ and a given "budget" for the expected number of jumps one can find a time-varying Markov process that drives $p_{\rm ini}$ to $p_{\rm fin}$ within the prescribed number of jumps whose entropy production is the r.h.s. of Eq. (41). Remarkably, this optimal strategy only involves conservative forces, i.e., has zero housekeeping entropy production ($\Delta\sigma = \Delta\sigma_{\rm NA}$).

Comparing Eq. (51) to Eq. (41), we observe that the l.h.s. of the former is lower or equal to the l.h.s. of the latter ($\Delta\sigma_{\rm NA} \leqslant \Delta\sigma$), and the same for the r.h.s. ($h \leqslant h^*_{\rm sym}$). We can thus wonder how tight (51) is, as a bound on $\Delta\sigma_{\rm NA}$. In particular, can we find a Markov chain with $\Delta\sigma_{\rm NA}$ strictly smaller than the r.h.s. of Eq. (32)? It turns out that we can. For example, consider a simple three-state Markov chain with the master equation:

$$\begin{aligned} \dot{p}_1 &= p_2 - p_1, \\ \dot{p}_2 &= p_3 - p_2, \\ \dot{p}_3 &= p_1 - p_3 \end{aligned} \tag{53}$$

(the last equation being redundant with $\dot{p}_3 = -\dot{p}_1 - \dot{p}_2$ expressing probability conservation). Take $p_{X_{\rm ini}} = (1/31, 25/31, 5/31)$, and observe the Markov chain over an infinitesimal time interval dt. Thus, $p_{X_{\rm fin}} = p_{X_{\rm ini}} + (24/31, -20/31, -4/31)dt$, and the expected activity (probability of jump) is $\langle A \rangle = dt$. It turns out that $d_{\rm TV}(p_{X_{\rm ini}}, p_{X_{\rm fin}})/\langle A \rangle = 24/31 \approx 0.77$ and $\Delta\sigma_{\rm NA} \approx 45.1dt < \langle A \rangle h^*_{\rm sym}(0.77) \approx 49.5dt$. We also have $\Delta\sigma = +\infty$ since the flow is unidirectional, with a zero reverse flow. This means that, albeit minimum entropy production for a given trajectory is reached with a conservative force protocol, it is possible to reach an even lower NA entropy production with a nonconservative protocol (thus at the cost of higher total entropy production). Therefore, there is a nontrivial tradeoff between NA and housekeeping entropy production, as in some cases the latter can be increased to decrease the former. Said otherwise, nonequilibrium stationary states can be used to achieve faster transitions, for a given budget in $\Delta\sigma_{\rm NA}$.

Nevertheless the range of this tradeoff is quite limited. Indeed, $h^*_{\rm sym}$ and h^* always differ in value by less than 18%, in either direction. This gap goes to 0% both for small arguments (ratio $d_{\rm TV}/\int \langle A \rangle \ll 1$, slow kinetics), and large arguments (ratio $d_{\rm TV}/\int \langle A \rangle$ close to one, fast kinetics). Thus, the tradeoff can only exist in an intermediate range of speeds. We leave open the exact quantitative characterization of this tradeoff.

B. Constant or time-symmetric rates revisited

Let us now focus again on constant-rate (or time-symmetric-rate) Markov chains, in the case of activity observable. Weakening (40) with $\langle A \rangle \leqslant 1$ (as $\langle A \rangle$ is a probability), we find

$$\begin{aligned} \Delta\sigma + D(p_{X_{\rm fin}} \| p_{X_{\rm ini}}) &\geqslant \langle A \rangle h^*_{\rm sym}\left(\frac{d_{\rm TV}(p_{X_{\rm ini}}, p_{X_{\rm fin}})}{\langle A \rangle} \right) \\ &\geqslant h^*_{\rm sym}[d_{\rm TV}(p_{X_{\rm ini}}, p_{X_{\rm fin}})]. \end{aligned} \tag{54}$$

Interestingly, this implies that for those initial and final state distributions such that $h^*[d_{\rm TV}(p_{X_{\rm ini}}, p_{X_{\rm fin}})] \leqslant D(p_{X_{\rm fin}} \| p_{X_{\rm ini}}) < h^*_{\rm sym}[d_{\rm TV}(p_{X_{\rm ini}}, p_{X_{\rm fin}})]$, there is an incompressible lower bound for entropy production regardless of the activity or duration of the transformation, due to the constraints of constant rates. For instance, driving a system from $p_{X_{\rm ini}} = (0.15, 0.85)$ to $p_{X_{\rm fin}} = (0.95, 0.05)$ with constant rates, no matter how slowly, necessarily produces an entropy at least 0.14.

C. Kinesin molecular motor

We illustrate the bounds on a toy model describing the motion of the kinesin molecular motor along a microtubule powered by ATP hydrolysis and exerting a force F on a pulled cargo. This standard model is a two-state, eight-transition Markov chain in a nonequilibrium stationary state. The logarithm of transitions rates mediated by the phosphorylation of ATP are proportional to its chemical potential log[ATP], while the logarithm of transition rates in the direction (respectively, against) the pulling force F are proportional to the work $F\ell/(k_BT)$ [respectively, $-F\ell/(k_BT)$], where ℓ is the motor displacement and k_BT is the thermal energy of the environment. We refer to Refs. [19,34] for a detailed description of the model. In Fig. 3 we compare the thermokinetic bound Eq. (32) and the standard thermodynamic uncertainty relation for various values of ATP concentration and pulling force F, and compare the quality of the bounds on entropy production. The model is at equilibrium for [ATP] $= 0$ and $F = 0$. We see in particular that (32) tends to compare favorably to the thermodynamic uncertainty relation far from equilibrium (i.e., increasing ATP concentration). See caption of Fig. 3 for a more detailed analysis.

D. Switching a bit in an electronic memory

A logical bit (zero or one) can be encoded in simple or complex nonlinear circuit, depending on the functionality (NOT gate, memory, clock, etc.). In generic circumstances, such circuits can be modeled by an overdamped Markov chain, the states of which record the charge (number of electrons) present in each conductor of the circuit (for instance the plate of a capacitance) [35,36]. The circuit state is in probability distribution p_0 when encoding a logical zero, and in probability distribution p_1 when encoding a one. The switching from a zero to a one thus amounts to driving the system from $p_{\rm ini} = p_0$ to $p_{\rm fin} = p_1$. Assume, as is representative, that the value of the bit is retrieved by the reading of the charge (or

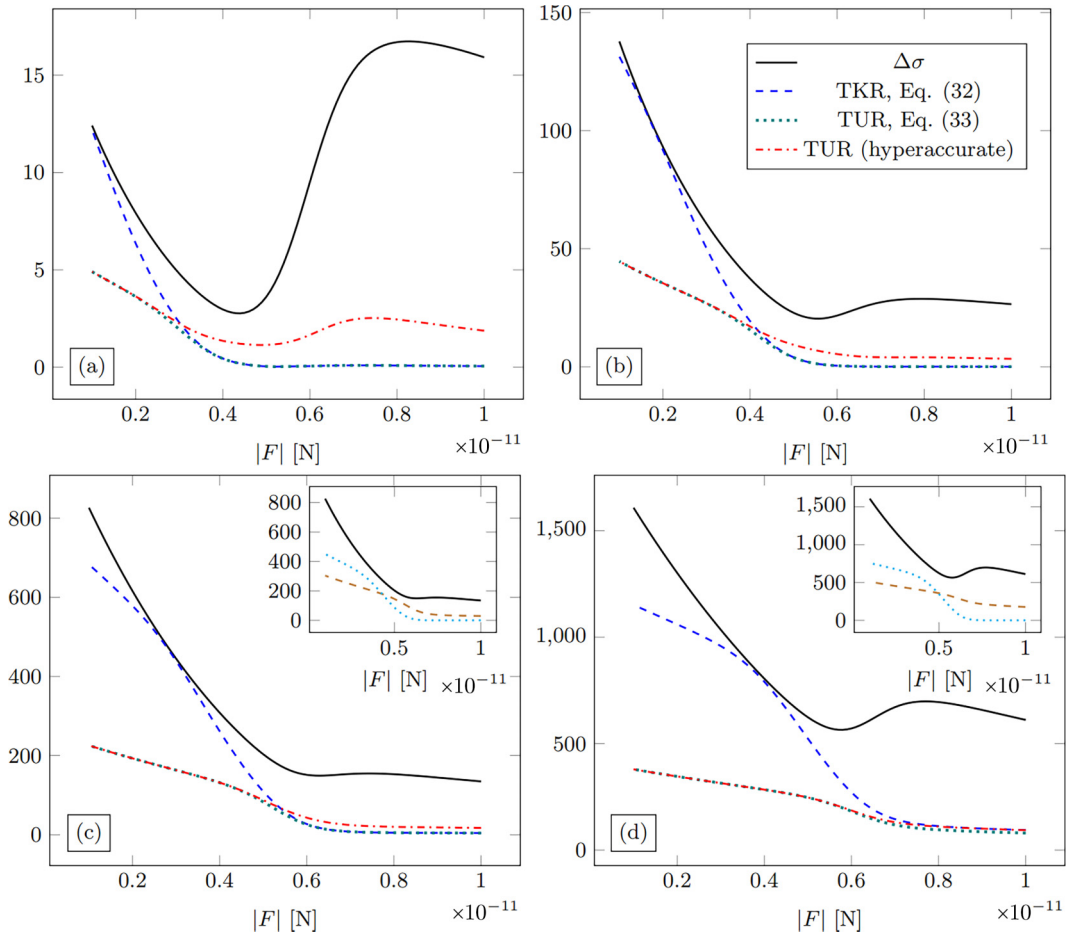

FIG. 3. Comparison of the bounds on the entropy production (solid) for the kinesin model in the stationary state. From panels (a–d) the ATP concentration increases as [ATP] = 1, 10, 10^2, 10^3 μM, respectively. For $f(e) = \mathrm{sgn}(p(e) - p(\bar{e}))$, the thermokinetic relation (32) (TKR, dashed) and the thermodynamic uncertainty relation (33) (TUR, dotted). We observe that the TKR (dashed) curve is always above the TUR (dotted) curve, thus offers a better bound on entropy production (solid), as predicted in Sec. IV. For the hyper-accurate current $f^*(e) = [p(e) - p(\bar{e})]/[p(e) + p(\bar{e})]$ (making Eq. (33) as tight as can be, see Appendix B), the thermodynamic uncertainty relation (33) (dashed-dotted). *Insets:* the new bound (32) for f representing the motor displacement (dashed) and the ATP current (dotted). We observe that the two TURs (dotted and dash-dotted) coincide (i.e., when $f^* \approx f$ on average), then the TKR (dashed) is a better bound than both TURs (dotted and dash-dotted), while when the two TUR curves diverge ($f^* \ll f$ on average), the hyper-accurate TUR (dash-dotted) performs better especially at low entropy production, in agreement with Appendix B.

equivalently, voltage) at the plate of a single "output" (linear or nonlinear) capacitance C_{out}, part of the full circuit. The distribution p_0 (respectively, p_1) results in a random charge Q_0 (respectively, Q_1) on C_{out}.

We now use thermokinetic relations to obtain a lower bound on the entropy produced while switching the system from a zero to a one, i.e., from $p_{\mathrm{ini}} = p_0$ to $p_{\mathrm{fin}} = p_1$. Note that Landauer's cost [37] here is zero, as we do not erase or reset a bit but simply switch it.

The antisymmetric observable $f = \Delta Q$ of interest on transitions is simply the "gradient" of charge, i.e., the increase of charge into the positive plate of C_{out} along the transition. Thus, $f = \pm Q_e$ or 0, where Q_e is the (positive) charge of the electron. It is possible that many transitions are zero valued, because they do not represent a change of charge on C_{out} (but on other parts of the complete circuit). In this way, the absolute value along the trajectory $\int \langle |f| \rangle$, counting the number of charge carriers flowing in or out the capacitance C_{out}, is less than the total activity of the full circuit along the trajectory. The net change of charge over the trajectory $\int \langle f \rangle$ is simply $Q_1 - Q_0$, the total mean charge separating a zero from a one. Since the maximum value for f is Q_e, we find

$$\Delta\sigma \geqslant 2\frac{|\langle Q_1\rangle - \langle Q_0\rangle|}{Q_e}\,\mathrm{atanh}\frac{|\langle Q_1\rangle - \langle Q_0\rangle|}{\int \langle |\Delta Q| \rangle}. \tag{55}$$

The quantity $\int \langle |\Delta Q| \rangle / Q_e$ can be considered as a dimensionless measure of the time taken by the driving, measured in the number of charges exchanged (in either direction) by the output capacitance. In particular, we recover Pinsker's bound, with a bound inversely proportional to time, consistently, e.g., with Refs. [38,39] in other contexts:

$$\Delta\sigma \geqslant 2\frac{|\langle Q_1\rangle - \langle Q_0\rangle|^2}{Q_e \int \langle |\Delta Q| \rangle}. \tag{56}$$

At stationarity, the number of charges into the positive plate of C_{out} match (on average) the number of charges out of that same plate. Thus, f verifies the "ocal conservation" conditions

stated in Sec. VI A to apply the nonadiabatic thermokinetic relation (52):

$$\Delta\sigma_{\rm NA} \geqslant 2\frac{\int\langle|\Delta Q|\rangle}{Q_e}h\left(\frac{|\langle Q_1\rangle-\langle Q_0\rangle|}{\int\langle|\Delta Q|\rangle}\right). \tag{57}$$

This relaxes in particular to the Pinsker's bound, stronger than (56) since it applies to the sole nonadiabatic entropy production:

$$\Delta\sigma_{\rm NA} \geqslant 2\frac{|\langle Q_1\rangle-\langle Q_0\rangle|^2}{Q_e\int\langle|\Delta Q|\rangle}. \tag{58}$$

Note that in this specific example, the absolute variation of charge $|\Delta q|$ over an infinitesimal interval is always 0 or Q_e. Thus, $Q_e|\Delta Q| = (\Delta Q)^2$, always (as already observed in a more general context in Sec. IV). Thus, $Q_e\langle|\Delta Q|\rangle = {\rm Var}(\Delta Q) = \mathcal{O}(dt)$ (as over an infinitesimal interval $\langle\Delta Q\rangle^2$ is negligible, in dt^2). We can thus write

$$\Delta\sigma_{\rm NA} \geqslant 2\frac{|\langle Q_1\rangle-\langle Q_0\rangle|^2}{\int{\rm Var}(\Delta Q)} = 2\frac{|\langle Q_1\rangle-\langle Q_0\rangle|^2}{(\tau_{\rm fin}-\tau_{\rm ini})\overline{{\rm Var}(\Delta Q)}}, \tag{59}$$

where $\overline{{\rm Var}(\Delta Q)}$ is the time-average of the variance rate and $\tau_{\rm fin}-\tau_{\rm ini}$ is the duration of the bit switch.

This formulation has the advantage that Q_e has disappeared from the equation, and remains true in the limit $Q_e \to 0$, i.e., when the charges are assumed to vary continuously, for instance in a Langevin model of the circuit, with Gaussian noise.

Note that an even better bound can be obtained in terms of the 1-Wasserstein distance. Indeed [see Appendix C, in particular Eq. (C6)] we can replace $|\langle Q_1\rangle-\langle Q_0\rangle|$ with $W_1(Q_0,Q_1)$, the Wasserstein distance between the distributions of random charge Q_0 and Q_1, which are two probability distributions on the real line. In particular, we have

$$\Delta\sigma_{\rm NA} \geqslant 2\frac{W_1(Q_0,Q_1)^2}{(\tau_{\rm fin}-\tau_{\rm ini})\overline{{\rm Var}(\Delta Q)}}. \tag{60}$$

This is a better bound than Eq. (59) because, from the general properties of Wasserstein distances [33], we have $W_1(Q_0,Q_1) \geqslant |\langle Q_1\rangle-\langle Q_0\rangle|$, with equality for example when Q_0 and Q_1 have the same distribution, only translated by $\langle Q_1\rangle-\langle Q_0\rangle$. A formally similar formula appears in Ref. [40] [Eq. (18) therein], that applies to Langevin equations with constant variance rate (white-noise intensity). On the l.h.s. it involves a different sort of entropy production (so-called MN excess entropy production), larger than our $\Delta\sigma_{\rm NA}$. On the r.h.s it involves the 2-Wasserstein distance, which is larger than the 1-Wasserstein distance [33]. Thus, it is not stronger or weaker than Eq. (60), and we leave the investigation of a possible relationship or common generalisation for future work.

This is a practical bound that can assess the efficiency of a wide range of electronic devices in storing and writing a bit. The methodology be adapted to nonelectronic overdamped devices as well (e.g., in chemical computing). We thus believe that our thermokinetic relations offer a flexible framework towards a useful characterization of the speed-dissipation tradeoff in complex real-world systems.

ACKNOWLEDGMENTS

J.-C.D. acknowledges funding from Project PDR/OL T.0060.21 "TheCirco" on Thermodynamics of Circuits for Computation, funded by the Fonds De La Recherche Scientifique - FNRS (Belgium) and FNR (Luxembourg). G.F. is funded by the European Union—NextGenerationEU—and by the program STARS@UNIPD with project "ThermoComplex." The anonymous referees are gratefully acknowledged for their suggestions of improvement.

APPENDIX A: EFFECTIVE MAXIMUM IN BOUNDS BETWEEN KULLBACK-LEILBER DIVERGENCES

We seek to improve Eqs. (5) and (4) by replacing $\varphi_{\max}$ and $A_{\max}$ with an "effective" maximum, whenever φ (thus A) is unbounded. We can safely assume that $D(p\|q)$ is finite, otherwise Eqs. (5) and (4) are trivially true. Thus, the sum (6), although gathering potentially infinitely many positive terms, converges, thus the sum of all terms for which $z > \varphi_0$, for a large enough threshold φ_0, is as small as we desire. Thus, for any $0 < \epsilon < 1$ there exists a φ_0 such that

$$p(\varphi > \varphi_0)D(\varphi p\|\varphi q|\varphi > \varphi_M) \tag{A1}$$

$$= \sum_{z>\varphi_0} zp(\varphi = z)D(\varphi p\|\varphi q|\varphi = z) \tag{A2}$$

$$< \epsilon D(\varphi p\|\varphi q). \tag{A3}$$

Then we find that

$$D(\varphi p\|\varphi q) = p(\varphi \leqslant \varphi_0)D(\varphi p\|\varphi q|\varphi \leqslant \varphi_0) + p(\varphi > \varphi_0)D(\varphi p\|\varphi q|\varphi > \varphi_0) \tag{A4}$$

$$\leqslant \varphi_0 p(\varphi \leqslant \varphi_0)D(p\|q|\varphi \leqslant \varphi_0) + \epsilon D(\varphi p\|\varphi q) \tag{A5}$$

$$\leqslant \varphi_0 D(p\|q) + \epsilon D(\varphi p\|\varphi q) \tag{A6}$$

$$\leqslant \frac{\varphi_0}{1-\epsilon}D(p\|q). \tag{A7}$$

Thus, we see that $\varphi_{\max}$ (the true maximum) can be replaced by $\varphi_{\rm effmax} = \frac{\varphi_0}{1-\epsilon}$ (an "effective" maximum), for each choice of ϵ and corresponding φ_0. Since ϵ determines the choice of φ_0, we can in principle optimize the choice of ϵ to get the best possible bound.

Note that this proof works also for the finite case, thus can be useful when A, albeit bounded, takes unconveniently large values with low probability.

APPENDIX B: THERMODYNAMIC UNCERTAINTY RELATIONS

Thermodynamic uncertainty relations are concerned with upper bounds on the ratio $|\langle f\rangle|/\sqrt{\langle f^2\rangle}$ (or, equivalently, $|\langle f\rangle|/{\rm StdDev}f$ where StdDev is the standard deviation), for a time-antisymmetric observable f.

For a constant-rate or time-symmetric-rate Markov chain, we find [19] that the time-antisymmetric observable with

highest such ratio is

$$f^*(\omega) = \frac{p(\omega) - p(\overline{\omega})}{p(\omega) + p(\overline{\omega})}. \tag{B1}$$

Indeed we find, for any time-antisymmetric observable f we have

$$\langle f \rangle = \langle f f^* \rangle, \tag{B2}$$

and, from Cauchy-Schwartz inequality:

$$\langle f \rangle^2 = \langle f f^* \rangle^2 \leqslant \langle f^2 \rangle^2 \langle (f^*)^2 \rangle, \tag{B3}$$

with equality for $f = f^*$. Moreover, we observe that

$$\langle f^* \rangle = \langle (f^*)^2 \rangle = \langle f^* \rangle^2 / \langle (f^*)^2 \rangle. \tag{B4}$$

Thus,

$$\langle f \rangle^2 / \langle f^2 \rangle \leqslant \langle f^* \rangle^2 / \langle (f^*)^2 \rangle, \tag{B5}$$

and, by simple algebra,

$$\langle f \rangle^2 / \mathrm{Var} f \leqslant \langle f^* \rangle^2 / \mathrm{Var} f^*. \tag{B6}$$

Moreover, we observe that $\langle |f^*| \rangle = d_{\mathrm{TV}}(p, \tilde{p})$ and $|f^*| \leqslant 1$.

From Eq. (29) with f^* we thus find

$$\Delta\sigma + D(p_{X_{\mathrm{fin}}} \| p_{X_{\mathrm{ini}}}) \geqslant 2 \langle f^* \rangle_p \mathrm{atanh}\left(\frac{\langle f^* \rangle_p}{d_{\mathrm{TV}}(p, \tilde{p})} \right). \tag{B7}$$

Although this is true for any constant-rate Markov chain, we consider it for a stationary Markov chain where the thermodynamic uncertainty relation applies and states:

$$\Delta\sigma \geqslant 2 \frac{\langle f^* \rangle^2}{\mathrm{Var} f^*} = 2 \frac{\langle f^* \rangle}{1 - \langle f^* \rangle}. \tag{B8}$$

This relation is proved [19] to be the tightest in the limit of short time intervals, where it simplifies to

$$\Delta\sigma \geqslant 2 \langle f^* \rangle. \tag{B9}$$

Comparing Eq. (B7) (derived from our thermokinetic relation) with Eq. (B9) (the best possible thermodynamic uncertainty relation) is now easy, as they only differ by the atanh factor, which extends from 0 to infinity. We thus see that no relation is always better than the other.

In particular, that when $p_e / p_{\overline{e}} \to \infty$ or 0 for every transition (infinitesimal-time trajectory) e, we find that $f^* \to \mathrm{sgn}(p_e - p_{\overline{e}})$ and $\langle f^* \rangle \to \langle |f^*| \rangle = d_{\mathrm{TV}}(p, \tilde{p})$, thus the thermokinetic relation (B7) correctly predicts an infinite entropy production, unlike the thermodynamic uncertainty relation (B9) which remains bounded.

Inversely when $p_e \approx p_{\overline{e}}$, then $f^* \ll \mathrm{sgn}(p_e - p_{\overline{e}})$, and $\langle f^* \rangle \to \langle |f^*| \rangle \ll d_{\mathrm{TV}}(p, \tilde{p})$, thus Eq. (B7) is looser than Eq. (34).

Note in passing that from Eqs. (B4), (B5), and (B7) we also obtain the following variant of thermodynamic uncertainty relation, valid for constant or time-symmetric rates, for stationary or nonstationary state distributions, and arbitrary antisymmmetric observables f:

$$\Delta\sigma + D(p_{X_{\mathrm{fin}}} \| p_{X_{\mathrm{ini}}}) \geqslant 2 \frac{\langle f \rangle^2}{\langle f^2 \rangle} \mathrm{atanh}\left(\frac{\langle f \rangle^2 / \langle f^2 \rangle}{d_{\mathrm{TV}}(p, \tilde{p})} \right). \tag{B10}$$

APPENDIX C: OPTIMAL TRANSPORT THEORY

In this Appendix we provide a brief summary of the problems solved by general optimal transport theory, in relation with the problem solved in this article and in Ref. [5].

The question of optimal transport was pioneered by Monge in 1781 and formalized in modern terms by Kantorovich in 1942.

The formal definition proposed by Kantorovich is the following. Consider $\Omega = X_{\mathrm{ini}} \times X_{\mathrm{fin}}$, for some probability spaces X_{ini} (with a given distribution p_{ini}) and X_{fin} (with a given distribution p_{fin}). Also consider a cost function $A : \Omega \to \mathbb{R}^+ \cup \{+\infty\}$.

We now consider all distributions p on Ω such that $p_{X_{\mathrm{ini}}} = p_{\mathrm{ini}}$ and $p_{X_{\mathrm{fin}}} = p_{\mathrm{fin}}$. We can see those p as "probabilistic maps" from X_{ini} to X_{fin}, transporting the distribution p_{ini} to p_{fin}. To such a distribution p we can associate the expected transport cost $\langle A \rangle_p = \sum_{\omega \in \Omega} A(\omega)$. As in the rest of this paper, this sum could be an integral depending on the context. Here ω is simply a pair $(x_{\mathrm{ini}}, x_{\mathrm{fin}})$.

The optimal transport problem is then to find a probability distribution p^* on Ω such that the expected cost $\langle A \rangle_{p^*}$ is minimum. This minimum cost $\langle A \rangle_{p^*}$, we denote it K_A, and call it *Kantorovich* cost (although many names are encountered in the literature, e.g., optimal transport cost, earth mover's cost, Monge-Kantorovich-Rubinstein cost, etc.). Formally:

$$p^* = \mathrm{argmin}_{p \text{ such that } p_{X_{\mathrm{ini}}} = p_{\mathrm{ini}} \text{ and } p_{X_{\mathrm{fin}}} = p_{\mathrm{fin}}} \langle A \rangle_p, \tag{C1}$$

$$K_A(p_{\mathrm{ini}}, p_{\mathrm{fin}}) = \langle A \rangle_{p^*}. \tag{C2}$$

In case where X_{ini} and X_{fin} are finite sets, this problem is also called the optimal assignment problem, or minimum cost matching problem.

In Monge's work, the motivation was to move earth to transform a pile of earth of a given shape into another pile of earth of another given shape. Kantorovich's main motivation was to allocate resources from production sites (e.g., coal mines) to places where they are useful (e.g., factories).

Another important situation is when $X_{\mathrm{ini}} = X_{\mathrm{fin}} = X$ is a metric space with distance $d(.,.)$. Then we can define, for any $k \geqslant 1$, the cost function $A(x_{\mathrm{ini}}, x_{\mathrm{fin}}) = d(x_{\mathrm{ini}}, x_{\mathrm{fin}})^k$. Then $K_A^{1/k}$ happens to define a distance between probability spaces on X, called the *k-Wasserstein* distance.

In this article, we consider the case where $X_{\mathrm{ini}} = X_{\mathrm{fin}} = X$ is a state space. We consider Ω as a set of abstract trajectories ω, each relating a initial state to a final state $x_{\mathrm{ini}}, x_{\mathrm{fin}}$. We consider a cost $A(\omega)$ on each trajectory ω. The optimal transport problem is to find a p^* on Ω minimizing the expected cost $\langle p^* \rangle$. Note that in this problem we do not impose p^* to be implementable as stochastic process of a certain class, e.g., a continuous-time Markov Chain.

The slight difference with the standard setting described above is that in our case Ω may be different from $X_{\mathrm{ini}} \times X_{\mathrm{fin}}$. Thus, possibly many trajectories $\omega \in \Omega$ go from x_{ini} to x_{fin}. This is an irrelevant difference however, as an optimal p^* may choose to concentrate probability from x_{ini} to x_{fin} on a single trajectory with smallest cost $A(\omega)$ relating these two states. If no trajectory in Ω relates x_{ini} to x_{fin}, this is interpreted as an infinite transport cost from x_{ini} to x_{fin}.

We request that A is symmetric under time reversal, so that the transport cost in our setting is symmetric:

$$K_A(p_{\rm ini}, p_{\rm fin}) = K_A(p_{\rm fin}, p_{\rm ini}). \tag{C3}$$

We also request that there is a zero-cost trajectory from any state to itself, so that the cost from a distribution to itself is zero:

$$K_A(p_{\rm ini}, p_{\rm ini}) = 0. \tag{C4}$$

Nevertheless, we do not impose anything else, so that K_A is not necessarily a distance (satisfying the triangle inequality).

If X is endowed with a distance $d(.,.)$, and for any states $x_{\rm ini}$, $x_{\rm fin}$ in X, the smallest cost $A(\omega)$ among all ω relating $x_{\rm ini}$ to $x_{\rm fin}$ is the distance $d(x_{\rm ini}, x_{\rm fin})$, then K_A is the corresponding 1-Wasserstein distance between $p_{\rm ini}$ and $p_{\rm fin}$.

Assume for example that A is the activity observable, as defined in Sec. V, i.e., the cost of transport of a state to any different state is one, thus is the discrete distance between states. Thus, in this case the Kantorovich cost $K_A(p_{\rm ini}, p_{\rm fin})$ is a 1-Wasserstein distance, which occurs to be the well-known total variation distance $d_{\rm TV}(p_{\rm ini}, p_{\rm fin})$.

A more general situation is the following. If we specify an undirected, unweighted connected graph whose nodes are the states of X, a set Ω of trajectories that can relate any state to any state, and a cost function $A : \Omega \to \mathbb{R}^+$ that assigns to any ω the distance between the extremities in the graph (number of edges of the shortest path), then we find the specific 1-Wasserstein distance on a graph used in Ref. [5].

If we now decompose trajectories into infinitesimal trajectories where at most one jump occurs, and we assume that these jumps necessarily occurs on an edge of the graph, then $\int \langle A \rangle_p$ is again the total expected activity (number of jumps) for a probability distribution p on trajectories Ω. In this case, $K_{\int A}(p_{\rm ini}, p_{\rm fin})$ is again the 1-Wasserstein distance $W_1(p_{\rm ini}, p_{\rm fin})$ for the graph distance. Thus, the bound Eq. (39) becomes

$$\begin{aligned}\Delta\sigma &\geqslant \int \langle A \rangle_p \; h^*_{\rm sym}\left(\frac{W_1(p_{\rm ini}, p_{\rm fin})}{\int \langle A \rangle_p}\right) \\ &= 2W_1(p_{\rm ini}, p_{\rm fin}) \,\mathrm{atanh}\, \frac{W_1(p_{\rm ini}, p_{\rm fin})}{\int \langle A \rangle_p}.\end{aligned} \tag{C5}$$

We see that Eq. (C5) is in general (for a given probability distribution p on the trajectories) a tighter bound than Eq. (42) (based on total variation distance) since $W_1 \geqslant d_{\rm TV}$. It also requires more knowledge on the distribution p, since it requires to know that some jumps between states are impossible under p. By contrast, the bound (42) assumes no knowledge on p other than being time-varying Markovian.

Because the condition of flow conservation mentioned in Sec. VI A is satisfied, the following bound is also verified for nonadiabatic (excess) entropy production:

$$\Delta\sigma_{\rm NA} \geqslant \int \langle A \rangle_p \; h\left(\frac{W_1(p_{\rm ini}, p_{\rm fin})}{\int \langle A \rangle_p}\right). \tag{C6}$$

As mentioned in the text, this is a new relation that is the counterpart for nonadiabatic entropy production of the bound (C5) derived in Ref. [5].

Suppose now that we are given a *weighted* undirected connected graph, with arbitrary nonnegative weights on the edge. Then the Kantorovich cost is indeed the 1-Wasserstein distance for the weighted distance between nodes of the graph (i.e., the shortest total weight of paths relating two nodes in the graph). In this case the cost of a trajectory (which is a time-stamped walk in the graph) is $\int A$ is the sum of costs $A(e)$ of all edges in the trajectory.

Even more general is when $A(e)$, the weight of edge e, is allowed to be time-dependent, in which case the cost $K_{\int A}$ is not even a distance in general (thus certainly not a 1-Wassertein distance). We then obtain Eqs. (39) and (48) in their full generality.

Similarly, if the time-varying weights on the edges satisfy the flow conservation condition of Sec. VI A at all times, we obtain Eq. (48) in full generality, provided that the flow conservation condition of Sec. VI A is satisfied at all times.

APPENDIX D: OPTIMAL PROTOCOL FOR THE ACTIVITY SPEED LIMIT

In this Appendix, we investigate conditions for which the speed limits (38) and (39) are tight, i.e., satisfied with equality for some Markov protocols. As a result, we find that the activity speed limit (42) can be satisfied with equality. In fact, we find that a slightly more general version than Eq. (42), involving activity and a specific class of Kantorovich costs, can be solved with equality. This reproduces the main result of Ref. [5]. The present derivation avoids the duality theory of optimal transport, and emphasizes the many degrees of freedom in the design of an optimal protocol, which is typically far from unique.

We first find at what condition the inequality (38), concerning constant-rate or symmetric-rate Markov chains, is in fact an equality. This inequality is obtained from chaining the inequalities (5) and (12). Looking at the conditions under which these two inequalities are tight [as spelled out under Eqs. (5) and (12)], we find that Eq. (38) is met with equality every time that, for all non-zero-probability trajectories ω in Ω, the cost function A takes either value 0 (and then $p = \tilde{p}$) or value $A_{\max}$ (and then $\ln \frac{p}{\tilde{p}} = c$), for some constants $A_{\max} > 0$ and $c > 0$. We can take $A_{\max} = 1$ without loss of generality. Note that A could take values that exceed $A_{\max}$ for zero-probability trajectories ω, as those trajectories are irrelevant to both sides of Eq. (38).

Let us spell out what this means for an infinitesimal time interval dt, where one can consider that either zero or one transition happens. This means that all the edges supporting a nonzero probability flow have unit cost $A = 1$ (and their reverse edge as well, out of symmetry of A). However, if no transitions happen (constant trajectory, i.e., waiting on a state), then $A = 0$ and $p = \tilde{p}$. Note that we can have other weights A on a zero-flow pair of edges, as they are irrelevant to both sides of Eq. (38). Then Eq. (38) is tight if on each edge with nonzero flow, the ratio between probability flows in each direction is e^c or e^{-c}, for some constant $c > 0$.

Suppose now, still along an infinitesimal interval dt, that we wish to design a flow p supported by unit-cost edges, with the constraint that we already know the reduced flow $p_{\rm red}(e)$ on each edge e [as defined above by Eq. (39)—in other words we know the net flow that p must achieve on each pair of edges] and the expected activity $\langle A \rangle = \sum_e p(e)$.

This problem has infinitely many solutions for p, provided that $\langle A\rangle > \sum_e p_{\rm red}(e)$. Among all those possible flows, we can now design one with the smallest entropy production, which satisfies Eq. (38) with equality. It suffices indeed to choose $c = 2\,\mathrm{atanh}\frac{\sum_e p_{\rm red}(e)}{\sum_e p(e)} = 2\,\mathrm{atanh}\frac{\sum A|p(e)-p(\overline{e})|}{2\langle A\rangle_p}$. Then for each e such that $p_{\rm red}(e) > 0$ we set $p(e) = \frac{p_{\rm red}(e)}{1-e^{-c}}$ and $p(\overline{e}) = \frac{p_{\rm red}(e)}{e^c-1} = \frac{p(e)}{e^c}$. In words, we "thicken" the reduced flow $p_{\rm red}$ to maintain the same net flow on every pair of edges as in $p_{\rm red}$, along with a constant forward/backward ratio on all pairs of edges.

Let us now assume that, in the above, $p_{\rm red}(e)$ has been chosen as the optimal transport flow p^*, minimizing $\langle A\rangle_{p^*}$ over flows with prescribed initial state distribution $p_{\rm ini} = p^*_{X_t}$ and final state distribution $p_{\rm fin} = p^*_{X_{t+dt}}$.

Let us provide an example of optimal transport. If we allow every possible transition (jumping directly from any state to any state), with unit weight $A(e) = 1$ on each edge, then the corresponding Kantorovich cost between $p_{\rm ini}$ and $p_{\rm fin}$ is just the total variation distance $d_{\rm TV}(p_{\rm ini}, p_{\rm fin}) = \frac{1}{2}\sum_x |\dot{p}_X|dt$, and the optimal transport flow is any flow transporting probability from "losing" states (those with a decreasing probability over the time interval) to "winning" states in amounts compatible with the prescribed $p_{\rm ini}$ and $p_{\rm fin}$.

Let us look at a more general case of optimal transport, as already explained briefly in Appendix C . We can restrict Ω (the set of trajectories allowed in the interval dt) to a subgraph of the complete graphs, to model the fact that only some transitions are physically possible in one direct jump, and assign cost $A = 1$ to each transition. Equivalently, we can keep Ω as the set of all possible transitions (from any state to any other state), assign cost $A = 1$ to the transitions that are physically possible in one jump, and a sufficiently large cost A to the other transitions that cannot be achieved directly. In this way the optimal transport (minimising the expected cost) will be concentrated on unit-cost edges. Thus, the Kantorovich cost is the minimum expected distance on the graph of allowed (unit-cost) edges. This is the case considered in Ref. [5].

Of course this optimal flow has no cycle of edges with nonzero flows, as one could then reduce the flow, if only by a small amount ϵ, thus reducing the expected activity by a multiple of ϵ with changing the state probability distributions.

A slightly more general property actually holds: assume that we find a cycle of edges with a nonzero flow on each edge *or their reverse* . This means that in this cycle $e_1 e_2 \dots e_n$ (say), some edges e_i support a nonzero flow, and some edges e_j have a zero flow, with a nonzero flow on $\overline{e}_i$. Then, we claim that the cycle has an even number of edges, with a nonzero flow on half of the edges, and a zero flow on the other half (with a nonzero flow on the reverse edge).

Indeed suppose that along a cycle, one finds more non-zero-flow edges than zero-flow edges. Then we can always remove a (possibly small, but nonzero) amount of flow ϵ from each edge e_i with nonzero flow, and add a flow ϵ to each $\overline{e}_j$ with a nonzero flow. Since each edge has a unit cost $A = 1$, this transfer of flow, which does not change at all the state probability distributions (as the total net flow into each state is unchanged), would have an overall smaller activity than $\langle A\rangle_{p^*}$, contradicting optimality of p^*. A symmetric argument holds for the case where one would find *less* non-zero-flow edges than zero-flow edges. The claim is proved.

This graph theoretic property on cycles has the interesting following thermodynamic consequence. Let us design a minimum entropy production flow p from such transport-optimal p^* and a given expected activity $\langle A\rangle_p$ by "thickening" p^* as shown above. Then on each edge, the ratio of forward over backward flow is either e^c or e^{-c}, for a given c. By the graph-theoretic property above, along any cycle with nonzero flows, the product of forward flows over backward flows is a product of equally many factors e^c and e^{-c}, thus is one. This shows that the householding entropy production is zero, i.e., the associated stationary flow satisfies detailed balance. In other words, this flow can in principle be achieved by conservative forces.

In summary, in all cases considered in this section, $\langle A\rangle$ is the expected activity, i.e., the average number of jumps over an infinitesimal interval dt. So far we have proved how to satisfy Eq. (38) with equality for any infinitesimal evolution of the state distribution, for a given expected activity, by an appropriate choice of flows. This solves the problem of designing minimum entropy production Markovian protocols given a certain expected activity over infinitesimal time intervals. The solution happens to involve only conservative forces.

We now consider the case of a noninfinitesimal time interval $[\tau_{\rm ini}, \tau_{\rm fin}]$. Assume that we wish to drive a system from a state distribution $p_{\rm ini}$ to a state distribution $p_{\rm fin}$ over an interval $[\tau_{\rm ini}, \tau_{\rm fin}]$. Assume we are allowed a total expected activity budget $\int A$.

First we solve the optimal transport problem, i.e., we find the flow $p^*_{\rm total}$ on each (allowed) edge that minimizes the total cost $\sum_e p^*(e)$. Note that the optimal transport is here a "one-step" protocol, not taking into account the need for a continuous-time Markov solution. So the solution provides for each edge the flow that must cross this edge over the whole time interval. To transform this one-step solution into a continuous-time solution, one can arbitrarily dispatch the total flow into infinitesimal flows for each infinitesimal intervals dt. For instance, a total flow $p^*_{\rm total}(e)$ over edge e can be achieved uniformly over time by assigning a constant infinitesimal flow $p^*_t(e) = p^*(e)dt/(\tau_{\rm fin} - \tau_{\rm ini})$ over each interval dt. This choice is by no means unique, as any nonuniform spreading over time is equally admissible. One can even spread the flow over time differently for each edge, provided that the probability of each state at all times never dips below zero. For each of those choices, $K_A(p_{\rm ini}, p_{\rm fin}) = \int dK_A = \int K_A(p_{X_t}, p_{X_{t+dt}})$.

We now wish to achieve Eq. (39) with equality. Equation (39) is obtained from summing Eq. (38) (over an infinitesimal interval) over the whole interval, exploiting the convexity of $h^*_{\rm sym}$. The equality in Eq. (38) for each infinitesimal interval remains an equality in Eq. (39) if the argument $\frac{K_A(p_{X_t}, p_{X_{t+dt}})}{\langle A\rangle_{p_{[t,t+dt]}}}$ is the same at all times t. Note that in this argument, both terms are infinitesimal quantities, proportional to dt. Hence, this ratio reads as an instantaneous speed, as the expected distance accomplished by transition. It is thus optimal to keep a constant speed at all times.

Thus, for each time interval $[t, t+dt]$, we assign an activity "budget" $\langle A\rangle_{[t,t+dt]}$ equal to $\frac{\int\langle A\rangle}{K_{\int A}(p_{\rm ini}, p_{\rm fin})} K_A(p_{X_t}, p_{X_{t+dt}})$. With this budget, we apply the methodology above over each infinitesimal time interval.

In summary: we have shown how to choose an optimal trajectory in the state probability space between p_{ini} and p_{fin}. We have shown that this can be done in many ways. Once this trajectory (and its time parametrization) is chosen, we choose the optimal flow for each time as shown above.

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

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

[3] V. T. Vo, T. Van Vu, and Y. Hasegawa, Unified approach to classical speed limit and thermodynamic uncertainty relation, Phys. Rev. E **102**, 062132 (2020).

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

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

[6] F. Tasnim and D. H. Wolpert, Thermodynamic speed limits for co-evolving systems, arXiv:2107.12471.

[7] G. Falasco, M. Esposito, and J.-C. Delvenne, Beyond thermodynamic uncertainty relations: Nonlinear response, error-dissipation tradeoffs, and speed limits, J. Phys. A: Math. Theor. **55**, 124002 (2022).

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

[9] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Dissipation bounds all steady-state current fluctuations, Phys. Rev. Lett. **116**, 120601 (2016).

[10] M. Polettini, A. Lazarescu, and M. Esposito, Tightening the uncertainty principle for stochastic currents, Phys. Rev. E **94**, 052104 (2016).

[11] J. M. Horowitz and T. R. Gingrich, Proof of the finite-time thermodynamic uncertainty relation for steady-state currents, Phys. Rev. E **96**, 020103(R) (2017).

[12] K. Proesmans and C. Van den Broeck, Discrete-time thermodynamic uncertainty relation, Europhys. Lett. **119**, 20001 (2017).

[13] T. R. Gingrich and J. M. Horowitz, Fundamental bounds on first passage time fluctuations for currents, Phys. Rev. Lett. **119**, 170601 (2017).

[14] A. Dechant and S. ichi Sasa, Current fluctuations and transport efficiency for general langevin systems, J. Stat. Mech. (2018) 063209.

[15] P. P. Potts and P. Samuelsson, Thermodynamic uncertainty relations including measurement and feedback, Phys. Rev. E **100**, 052137 (2019).

[16] Y. Hasegawa and T. Van Vu, Fluctuation theorem uncertainty relation, Phys. Rev. Lett. **123**, 110602 (2019).

[17] T. Van Vu and Y. Hasegawa, Thermodynamic uncertainty relations under arbitrary control protocols, Phys. Rev. Res. **2**, 013060 (2020).

[18] K. Liu, Z. Gong, and M. Ueda, Thermodynamic uncertainty relation for arbitrary initial states, Phys. Rev. Lett. **125**, 140602 (2020).

[19] G. Falasco, M. Esposito, and J.-C. Delvenne, Unifying thermodynamic uncertainty relations, New J. Phys. **22**, 053046 (2020).

[20] T. M. Cover, *Elements of Information Theory* (John Wiley & Sons, New York, NY, 1999).

[21] I. Vajda, Note on discrimination information and variation (corresp.), IEEE Trans. Inf. Theory **16**, 771 (1970).

[22] G. L. Gilardoni, An improvement on Vajda's inequality, in *In and Out of Equilibrium 2* (Springer, Berlin, 2008), pp. 299–304.

[23] A. A. Fedotov, P. Harremoës, and F. Topsoe, Refinements of Pinsker's inequality, IEEE Trans. Inf. Theory **49**, 1491 (2003).

[24] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, Phys. Rev. E **82**, 011143 (2010).

[25] Y. Oono and M. Paniconi, Steady state thermodynamics, Prog. Theor. Phys. Suppl. **130**, 29 (1998).

[26] T. Hatano and S. I. Sasa, Steady-state thermodynamics of Langevin systems, Phys. Rev. Lett. **86**, 3463 (2001).

[27] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, Vol. 107 (American Mathematical Society, Providence, RI, 2017).

[28] D. M. Busiello and S. Pigolotti, Hyperaccurate currents in stochastic thermodynamics, Phys. Rev. E **100**, 060102(R) (2019).

[29] G. Monge, Histoire de l'Académie royale des sciences, Année **1771**, 666 (1781).

[30] L. V. Kantorovich, On the translocation of masses, Dokl. Akad. Nauk SSSR **37**, 227 (1942) [Math. Sci. **133**, 1381 (2006)].

[31] Y. Ollivier, H. Pajot, and C. Villani, *Optimal Transport: Theory and Applications*, Vol. 413 (Cambridge University Press, Cambridge, UK, 2014).

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

[33] A. Figalli and F. Glaudo, *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows* (EMS Press, Helsinki, Finland, 2021).

[34] I. Di Terlizzi and M. Baiesi, Kinetic uncertainty relation, J. Phys. A Math. Theor. **52**, 02LT03 (2019).

[35] N. Freitas, J.-C. Delvenne, and M. Esposito, Stochastic and quantum thermodynamics of driven RLC networks, Phys. Rev. X **10**, 031005 (2020).

[36] N. Freitas, G. Falasco, and M. Esposito, Linear response in large deviations theory: A method to compute nonequilibrium distributions, New J. Phys. **23**, 093003 (2021).

[37] T. Sagawa, Thermodynamic and logical reversibilities revisited, J. Stat. Mech.: Theory Exp. (2014) P03025.

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

[39] J. S. Lee, S. Lee, H. Kwon, and H. Park, Speed limit for a highly irreversible process and tight finite-time Landauer's bound, Phys. Rev. Lett. **129**, 120603 (2022).

[40] A. Dechant, S.-I. Sasa, and S. Ito, Geometric decomposition of entropy production into excess, housekeeping, and coupling parts, Phys. Rev. E **106**, 024125 (2022).