PHYSICAL REVIEW E **106**, L032101 (2022)

Letter

Lower bound for entropy production rate in stochastic systems far from equilibrium

Domingos S. P. Salazar

Unidade de Educação a Distância e Tecnologia, Universidade Federal Rural de Pernambuco, 52171-900 Recife, Pernambuco, Brazil

(Received 1 April 2022; accepted 12 August 2022; published 1 September 2022)

We show that the Schnakenberg's entropy production rate in a master equation is lower bounded by a function of the weight of the Markov graph, here defined as the sum of the absolute values of probability currents over the edges. The result is valid for time-dependent nonequilibrium entropy production rates. Moreover, in a general framework, we prove a theorem showing that the Kullback-Leibler divergence between distributions $P(s)$ and $P'(s) := P(m(s))$, where m is an involution, $m(m(s)) = s$, is lower bounded by a function of the total variation of P and P', for any m. The bound is tight and it improves on Pinsker's inequality for this setup. This result illustrates a connection between nonequilibrium thermodynamics and graph theory with interesting applications.

DOI: 10.1103/PhysRevE.106.L032101

Introduction. Nonequilibrium physics has few general results and most of them can be traced back to the second law of thermodynamics. Among equivalent statements, the second law demands that the entropy production rate of a system is non-negative [1–5],

$$\frac{d}{dt}\langle\Sigma\rangle = \langle\pi\rangle \geqslant 0, \tag{1}$$

which means that the rate of entropy variation of the system plus environment is non-negative. The statement (1) is equivalent to all classic statements from Clausius [6,7], Carnot [8], and others [1], but we used a notation in (1) suggesting that the quantity $\langle\Sigma\rangle$ is an average. As a matter of fact, with the advent of stochastic thermodynamics [9–18], quantities such as heat, work, and even entropy production might be defined at the trajectory level. In order to account for fluctuations of thermodynamic quantities in small systems, the second law is now usually rephrased as an ensemble average of the entropy production.

However, for each system, one must identify what plays the role of the entropy production rate in terms of the system's dynamics. In several situations, either the entropy production $\langle\Sigma\rangle$ or the entropy production rate $\langle\pi\rangle$ might take the form

$$D(P|P') := \sum_s P(s)\ln\left(\frac{P(s)}{P(s')}\right) \geqslant 0, \tag{2}$$

where $s \in S$ is some event (for instance, a transition), s' is the time-reversed event, and $P(s)$ is a probability function. This expression appears in the context of Markov processes [19,20], heat exchange problems [21–27], and stochastic thermodynamics [2] in general for a suitable choice of S (see applications).

Formally, expression (2) is the Kullback-Leibler (KL) divergence of $P(s)$ and $P(s')$. From definition, the expression (2) satisfies (1). Moreover, it vanishes if and only if $P(s) = P(s')$ for all s, which is a form of detailed balance condition. It makes equilibrium a situation equivalent to microscopic reversibility. Actually, expression (2) appears as the entropy production rate in the analysis of Lyapunov stability of Markov processes to demonstrate that the system tends to equilibrium in the long time. A nonequilibrium steady state can be maintained by coupling the system to multiple reservoirs, resulting in a constant $\langle\pi\rangle > 0$. Thus, $D(P|P')$ quantifies a disagreement the system has with the detailed balance condition (equilibrium). Of course, there are other ways to quantify this disagreement.

Schnakenberg [19] proposed a network representation of the master equation, importing results from graph theory [28–30] to nonequilibrium thermodynamics. We explore this idea by considering the network representation of the vertices s for a given matching $\{(s, m(s))|s \in S\}$, where $m(m(s)) = s$. We define the weight of this matching as the following sum,

$$w := \frac{1}{2}\sum_s |P(s) - P(s')|, \tag{3}$$

which is the total variation distance of $P(s)$ and $P(s')$. As in the case of the entropy production rate (2), this weight w is non-negative and $w = 0$ if and only if $P(s) - P(s') = 0$ for all s, which is the detailed balance condition. From definition (3), we also have $w \geqslant 0$ finite, while $\langle\pi\rangle$ might diverge. Intuitively, the weight of an edge $|P(s) - P(s')|$ quantifies the disagreement with the detailed balance condition on that edge. Summing over all edges results in the amount of disagreement over the whole graph.

A natural question is if both measures of disagreement, $D(P|P')$ and w, are somehow connected. For now, we only checked that they agree in equilibrium, $D(P|P') = 0$ and $w = 0$. In this Letter, we show that

$$D(P|P') \geqslant 2w\tanh^{-1} w, \tag{4}$$

and the bound is tight, valid arbitrarily far from equilibrium, with saturation observed for a particular two-level system.

The Letter is organized as follows. We introduce the formalism and prove a general theorem about the relation between the KL divergence [31] and the total variation distance (TV) [32] when distributions are related by an involution. We apply this result to the heat exchange problem and comment on the exchange fluctuation theorem. Then, we apply this result to the master equation (22) and get a bound (4) in terms of the probability currents. We also illustrate the result with simulations and compare it with other known bounds for the KL divergence and the TV from statistics (Pinsker's and Bretagnolle-Hubber's). We apply the theorem to a general setup in stochastic thermodynamics and obtain a lower bound for the stochastic work. Finally, we apply the theorem to a time rotation problem, where a demon flips a coin and decide if time goes forward or backwards.

Formalism. First, we define some concepts. Let P and Q be probabilities in S, $\sum_s P(s) = \sum_s Q(s) = 1$, with $0 \leqslant P(s)$ and $Q(s) \leqslant 1$. The KL divergence is given by $D(P|Q) := \sum_s P(s)\ln[P(s)/Q(s)]$ [defined when $Q(s) = 0 \to P(s) = 0$] and the total variation distance, $\Delta(P, Q) := (1/2)\sum_s |P(s) - Q(s)|$. We prove the following result:

Theorem. Let $S \neq \emptyset$ be a countable set and let $P : S \to [0, 1]$ be any probability function. Let $m : S \to S$ be any involution, $m(m(s)) = s$. Define $P' := S \to [0, 1]$ as $P'(s) := P(m(s))$. Then, the bound

$$D(P|P') \geqslant 2\Delta(P, P')\tanh^{-1}(\Delta(P, P')) \tag{5}$$

is tight.

Proof. Let P and m be defined in S as in the theorem. If there is an $s \in S$ such that $P(s) = 0$ and $P'(s) = P(m(s)) \neq 0$, then $D(P|P')$ diverges while $\tanh^{-1}(\Delta(P, P'))$ might be finite. Therefore, we focus on the case with the condition $P(s) = 0 \to P(m(s)) = 0$ for all s (absolute continuity).

Define a function $\pi : S \to \mathbb{R}$ as

$$\pi(s) := \ln[P(s)/P(s')], \tag{6}$$

for $P(s) > 0$ [which makes $P(s') > 0$], where $s' := m(s)$. In the case $P(s) = 0$ [which makes $P(s') = 0$], define $\pi(s) := 0$. It is immediate that $\pi(s') = \ln[P(m(s))/P(m(s'))] = \ln[P(s')/P(s)] = -\pi(s)$. Note that the average of $\pi(s)$ is given by

$$\langle\pi\rangle := \sum_{s\in S}\pi(s)P(s) = \sum_{s\in S}P(s)\ln\left(\frac{P(s)}{P(s')}\right) = D(P|P'). \tag{7}$$

Moreover, note that the variable $\pi(s)$ satisfies a form of detailed fluctuation theorem from definition (6), $P(s)/P(s') = \exp[\pi(s)]$. Now we define another probability function over the pairs $E = \{e = \{s, m(s)\}|s \in S\}$, $p(e) : E \to [0, 1]$, given by

$$p(e) := P(s) + P(s'), \tag{8}$$

if $s \neq s'$ and $p(e) := P(s)$ if $s = s'$. Because m is an involution, each $s \in S$ belongs to a single pair $e \in E$.

We check the normalization of $p(e)$ as follows. Consider any ordering of the set $S = (s_1, s_2, \ldots)$ such that $r(s_i) = i$ is the rank of the element s in that ordering. Now define the subsets $E_0 = \{s|s = s'\}$, $E_< = \{s|s \neq s', r(s) < r(s')\}$, and $E_> = \{s|s \neq s', r(s) > r(s')\}$. Then,

$$\begin{aligned}\sum_{e\in E}p(e) &= \sum_{s\in E_<}[P(s) + P(m(s))] + \sum_{s\in E_0}P(s)\\ &= \sum_{s\in E_<}P(s) + \sum_{s\in E_>}P(s) + \sum_{s\in E_0}P(s) = \sum_{s\in S}P(s) = 1,\end{aligned} \tag{9}$$

where we use $m(s) \in E_< \to s \in E_>$, for $s \neq s$, and $p(e) = P(s)$, for $e = (s, s)$. In the last identity, we used that $E_< \cup E_> \cup E_0 = S$ forms a partition in S.

Now we define $\Pi(e) := p(e)^{-1}[P(s) - P(s')]\ln[P(s)/P(s')]$, for the pair $e = \{s, s'\}$, whenever $P(s) > 0$. When $P(s) = 0$, let $\Pi(e) := 0$. Then, we get the following from (7):

$$\langle\pi\rangle = \frac{1}{2}\sum_{s,s'\in S}[P(s) - P(s')]\ln\left(\frac{P(s)}{P(s')}\right) = \sum_{e\in E}\Pi(e)p(e). \tag{10}$$

We also compute the total variation of probabilities (P, P'):

$$\Delta(P, P') = \frac{1}{2}\sum_{s\in S}|P(s) - P(s')| = \sum_{e\in E}w(e)p(e), \tag{11}$$

where $w(e) := |P(s) - P(s')|/p(e)$, for $e = \{s, s'\}$, when $P(s) > 0$, and $w(e) := 0$, when $P(s) = 0$. Note that

$$|P(s) - P(s')| = \max[P(s), P(s')] - \min[P(s), P(s')], \tag{12}$$

which makes $w(e) = \rho - (1 - \rho)$ for $\rho = \max[P(s), P(s')]/p(e)$ for some $0.5 \leqslant \rho < 1$, using $\max[P(s), P(s')] + \min[P(s), P(s')] = p(e)$. Using ρ, we have

$$\Pi(e) = [\rho - (1 - \rho)]\ln\left(\frac{\rho}{1 - \rho}\right) = w(e)\ln\left(\frac{\rho}{1 - \rho}\right). \tag{13}$$

Finally, define the logistic function $\sigma(a) = \exp(a/2)/[\exp(a/2) + \exp(-a/2)]$ and find a such that $\rho = \sigma(a)$. Then, using $\ln[\rho/(1 - \rho)] = a$ and $w(e) = \rho - (1 - \rho) = \tanh(a/2) \to a = 2\tanh^{-1}(w(e))$ in (13), we get

$$\Pi(e) = aw(e) = 2w(e)\tanh^{-1}(w(e)) := B(w(e)), \tag{14}$$

where $B(w) = 2w\tanh^{-1}(w)$. The degenerate case, $P(s) = 0 \to \Pi(e) = 0$, $w(e) = 0$, also satisfies (14). Note that, for $0 < w < 1$, we have $B(w) > 0$, $B'(w) > 0$, and $B''(w) > 0$. Therefore, from Jensen's inequality,

$$\begin{aligned}\langle\pi\rangle &= \sum_{e\in E}\Pi(e)p(e)\\ &= \sum_{e\in E}B(w(e))p(e) \geqslant B\left(\sum_{e\in E}w(e)p(e)\right),\end{aligned} \tag{15}$$

and, after using $\langle\pi\rangle = D(P|P')$ from (7) and $\sum_e w(e)p(e) = \Delta(P, P')$ from (43), it results in

$$D(P|P') \geqslant 2\Delta(P, P')\tanh^{-1}(\Delta(P, P')), \tag{16}$$

which proves (5).

Saturation of the bound follows from (14). If S has a single element, then the bound is trivially saturated $D(P|P') =$

$B(\Delta(P, P')) = 0$, as $P = P'$. If S has more than one element, consider two distinct elements from $\{s_1, s_2\} \subseteq S$ and a $\tilde{P}$ such that $\tilde{P}(s_1) + \tilde{P}(s_2) = 1$, with involution $\tilde{m}(s_1) = s_2$, $\tilde{m}(s_2) = s_1$, and $\tilde{m}(s) = s$ otherwise. It makes $p(e) = 1$ for $e = \{s_1, s_2\}$, and $\Pi(e) = D(\tilde{P}|\tilde{P}')$ and $\Delta(\tilde{P}, \tilde{P}') = w(e)$, which results in the saturation $D(\tilde{P}|\tilde{P}') = B(\Delta(\tilde{P}, \tilde{P}'))$. For completeness, for a given total variation $\Delta = \Delta(P, P')$, if there is a lower bound $B^*(\Delta)$ for $D(P|P')$ such that $D(P|P') \geqslant B^*(\Delta) > B(\Delta)$, then the two-level system presented above violates it. Therefore, $B(\Delta)$ is the tightest bound for this setup. That completes the proof.

Remarks. Interestingly, time was not mentioned in the theorem and the fluctuation theorem was not assumed. This proof shows the important role played by the involution m in nonequilibrium physics: When you equip a probability space with an involution, you could think of the connected events $\{s, s'\}$ as forward and backward events in the arrow of time. Because they might have different probabilities, $P(s) > P(s')$, we pick s as the event forward in time. In this context, the fluctuation theorem is also a construct, because $\pi(s)$ behaves like a thermodynamic force [18], switching signs after an involution.

Additionally, when compared to off the shelf bounds between the KL divergence and the TV, we have $D(P|P') \geqslant B(\Delta(P, P')) \geqslant 2\Delta(P, P)^2$, which shows the bound improves on Pinsker's inequality [33] for this setup. It is also true for the Bretagnolle-Huber' [34] bound $D(P|P') \geqslant B(\Delta(P, P')) \geqslant -\ln[1 - \Delta(P, P')^2]$. See Ref. [35] for a discussion on the intuition behind the TV.

Also note that if (S_1, P, m) satisfies the theorem and $g : S_1 \to S_2$ is a bijection, then $(S_2, P \circ g^{-1}, g \circ m \circ g^{-1})$ satisfies the theorem. For that reason, different systems might be mapped into each other and the bound is universal. For completeness, although the theorem was proved for discrete sets S, it is possible to expand it to $S = \mathbb{R}^n$ with a differentiable involution m (see the Appendix for the case $S = \mathbb{R}$).

Application I: Heat-exchange problem. In a general heat-exchange problem, one has an energy variation between two consecutive measurements. The system starts at a density matrix ρ_0, and two consecutive energy measurements are performed, yielding values E_i (at $t = 0$), projecting the system at $|i\rangle\langle i|$, and E_j (at $t > 0$), both values in the spectrum $H = \{E_1, E_2, \ldots\}$. The energy variation is $\Delta E = E_j - E_i$. In this case, one has a set of energy gaps, $S = \{\Delta E = a - b | a, b \in H\}$, and $P(s) = P(\Delta E)$ given by

$$P(\Delta E) = \sum_{ij} \delta[\Delta E - (E_j - E_i)] \langle j|\Phi\big(|i\rangle\langle i|\big)|j\rangle\langle i|\rho_0|i\rangle, \tag{17}$$

where Φ is an operator from the quantum dynamical semigroup. Let the involution $m : S \to S$ be $m(\Delta E) = -\Delta E$. Check that $m(m(s)) = s$. Therefore, (S, P, m) satisfies the condition for the theorem. We have the following from (5),

$$D[P(\Delta E)|P(-\Delta E)] \geqslant B(w), \tag{18}$$

for $B(w) = 2w \tanh^{-1}(w)$, with w given by

$$w = \frac{1}{2} \sum_{\Delta E} |P(\Delta E) - P(-\Delta E)|. \tag{19}$$

Note that, as the theorem for (S, P, m) is general, we did not mention thermal distributions or detailed balance in this general setup.

Now for the specific case where the system satisfies the exchange fluctuation theorem (XFT), one has $\Sigma = \alpha \Delta E$ for some affinity α and the XFT [21–24,26] reads

$$\frac{P(\Delta E = \Sigma/\alpha)}{P(\Delta E = -\Sigma/\alpha)} = \exp(\Sigma). \tag{20}$$

In this case, Eq. (18) reads

$$\langle \Sigma \rangle \geqslant B[Pr(\Sigma > 0) - Pr(\Sigma < 0)], \tag{21}$$

where we used $\Sigma > 0 \leftrightarrow P(\Delta E = \Sigma/\alpha) > P(-\Delta E = \Sigma/\alpha)$ from (20) and $Pr(\Sigma > 0) := \sum P(\Delta E = \Sigma/\alpha)\theta(\alpha\Delta E)$ and $\theta(x) = 1(0)$ if $x > 0 (x < 0)$ is a step function.

Using that $B'(x) > 0$, inverting expression (18) results in a bound for the apparent violations of the second law, $Pr(\Sigma < 0)$, in terms of $\langle \Sigma \rangle$ recently proposed using similar methods [36]. The more general bound (18) is still valid without assuming the XFT or detailed balance.

Application II: Master equation. Another interesting application is to consider a continuous Markov process described by a master equation [37]:

$$\dot{p}_i = \sum_j W_{ij} p_j - W_{ji} p_i, \tag{22}$$

where $W_{ij} \geqslant 0$ is the transition rate from state j to state i and $p_i = p_i(t)$ is the probability of state i, $0 \leqslant p_i(t) \leqslant 1$, with condition $\sum_{i \neq j} W_{ij} p_j > 0$. Let $S = \{(i, j) | i, j \in \{1, \ldots, N\}, i \neq j\}$, where we deliberately removed the points (i, i).

The known expression for the entropy production rate from Schnakenberg [19,20,38,39] is

$$\overline{\pi} = \frac{1}{2} \sum_{ij} (W_{ji} p_i - W_{ij} p_j) \ln\left(\frac{W_{ji} p_i}{W_{ij} p_j}\right). \tag{23}$$

Now we find a natural tuple (S, P, m) for the application of the theorem. For any instant of time, we define a probability function $P : S \to [0, 1]$ as

$$P(s) = \frac{W_{ij} p_j}{\sum_{i \neq j} W_{ij} p_j} = \frac{W_{ij} p_j}{z} := P_{ij}, \tag{24}$$

for $s = (i, j)$, where $z := \sum_{i \neq j} W_{ij} p_j$. Note that P is normalized,

$$\sum_{s \in S} P(s) = z^{-1} \sum_{(i,j) \in S} W_{ij} p_j = z^{-1} \left(\sum_{i \neq j} W_{ij} p_j \right) = 1. \tag{25}$$

Let the involution $m : S \to S$ be $m(i, j) = (j, i)$. Check that $m(m(s)) = s$. Again, (S, P, m) meets the condition for the theorem (5).

Using $P(s) - P(s') = P_{ij} - P_{ji} = (W_{ij} p_j - W_{ji} p_i)/z$, we verify that the total variation $\Delta(P, P')$ is given as

$$\Delta(P, P') = \sum_{i > j} |J_{ij}|/z = w, \tag{26}$$

with $J_{ij} = W_{ij} p_j - W_{ji} p_i$ being the probability current. Also, we have $D(P|P') = \sum_s P(s) \ln[P(s)/P(s')] = \overline{\pi}/z$ as defined

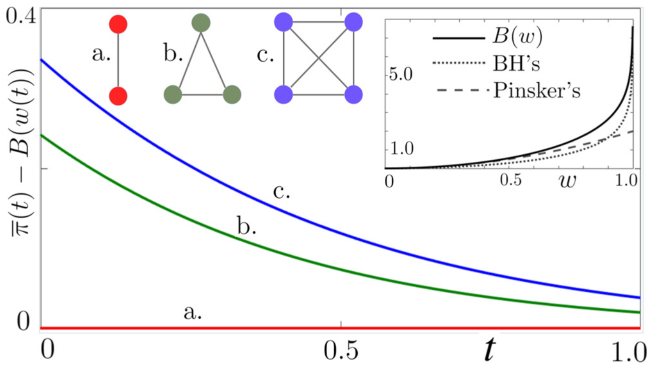

FIG. 1. The difference between the entropy production rate $\overline{\pi}$ and the bound $B(w)$ as a function of time for a two-level (a.), three-level (b.), and four-level (c.) systems described in the text. The gaps decay to zero with time as the systems approach equilibrium. The two-level system (a.) saturates the bound. Inset: The bound $B(w)$ as a function of w showing improvement over other known bounds, Pinsker's and Bretagnolle-Huber's (BH's).

in (23). In terms of $\overline{\pi}$, w, and z, theorem (5) now reads

$$\overline{\pi}/z \geqslant 2w \tanh^{-1}(w), \tag{27}$$

which is inequality (4), for $\overline{\pi} \to \overline{\pi}/z$. This bound was recently discovered using other methods [39,40]. In this context, the KL divergence and the TV are also analyzed in the context of classic speed limit and using Wasserstein distance [41,42]. For consistency of (27), check that $w \leqslant 1$,

$$w = \frac{1}{z}\sum_{i<j}|W_{ij}p_j - W_{ji}p_i| \leqslant \frac{1}{z}\sum_{i<j}(W_{ij}p_j + W_{ji}p_i) = 1. \tag{28}$$

In this case, the bound is saturated by a two-level system in contact with a thermal bath at infinite temperature. Let $(p_1, p_2) = (p, 1-p)$ be the initial state probabilities for some $0.5 < p < 1$ and the transition matrix given by $W_{12} = 1$ and $W_{21} = 1$, with $W_{11} = W_{22} = 0$, in the same spirit of (14).

In equilibrium, the system will have $w = 0$ (and $\overline{\pi} = 0$), satisfying the detailed balance condition $J_{ij} = 0$. Near equilibrium, $w \approx 0$, the bound expands to

$$\overline{\pi} \geqslant B(w) \approx 2w^2 + \tfrac{2}{3}w^4 + O(w^6), \tag{29}$$

and recognizing $\overline{\pi} = D(P|P')$, also $\Delta(P, P') = w$, Pinsker's inequality states that $\overline{\pi} = D(P|P') \geqslant 2\Delta(P, P')^2 = 2w^2$, so the bound $B(w)$ improves on it in order $O(w^4)$.

In Fig. 1, we simulate the master equation (22) for three different systems: (a.) a two-level system with $p(0) = [0.9; 0.1]$ and $W_{ij} = (1 - \delta_{ij})$; (b.) a three-level system with $p(0) = [0.8; 0.1; 0.1]$ and $W_{ij} = (1/2)(1 - \delta_{ij})$; and (c.) a four-level system with $p(0) = [0.7; 0.1; 0.1; 0.1]$ and $W_{ij} = (1/3)(1 - \delta_{ij})$. They correspond to a weak coupling approximation of a system and a thermal bath with infinite temperature. For each system, we compute $\overline{\pi}(t)$ and $w(t)$ and plot the gap, $\overline{\pi}(t) - B(w(t))$ vs t. Note that the two-level system saturates the bound, as expected. The three- and four-level systems display a gap that decays in time. In the inset of Fig. 1, we observe the plot $B(w)$ vs w, comparing the bound to other known bounds for the KL divergence as a function of TV. See that Pinsker's is tighter than Bretagnolle-Hubber's (BH's) for $w \approx 0$ (which corresponds to near equilibrium in the Markov dynamics). For $w \approx 1$, the behavior is inverted. In all domains, Fig. 1 shows $B(w)$ is a tighter bound than both.

Application III: Stochastic thermodynamics. Consider a trajectory $\Gamma = (x_0, \ldots, x_N)$ and its inverse $\Gamma^\dagger = (x_N, \ldots, x_0)$, where the indexes are time steps and $x_i \in D \subset \mathbb{R}$, representing some domain. The entropy production in stochastic thermodynamics is given as [2]

$$\langle\Sigma\rangle_F = \sum_\Gamma P_F(\Gamma) \ln\left(\frac{P_F(\Gamma)}{P_B(\Gamma^\dagger)}\right), \tag{30}$$

where P_F (P_B) is the forward (backward) probability over the set of trajectories. They are different because P_F depends on the protocol $\lambda(t) = (\lambda_0, \ldots, \lambda_N)$, while P_B depends on $\lambda(T-t) = (\lambda_N, \ldots, \lambda_0)$. Define the entropy production for the backward process, $\langle\Sigma\rangle_B$, by replacing $F \leftrightarrow B$ in (30).

The theorem is applied in this system as follows. Let $S = \{(\Gamma, \sigma)|\Gamma \in D^{N+1}, \sigma \in \{1, -1\}\}$. For any $s = (\Gamma, \sigma)$, define $P(s) = P(\Gamma, \sigma)$ as $P(\Gamma, 1) := (1/2)P_F(\Gamma)$ and $P(\Gamma, -1) := (1/2)P_B(\Gamma)$, and check $P(s)$ is normalized in S. Let $m(s) = m(\Gamma, \sigma) = (\Gamma^\dagger, -\sigma)$. Check that $m(m(s)) = s$. The tuple (S, P, m) satisfies the following condition for the theorem, $D(P|P') \geqslant B(w)$. In order to understand its meaning, let us compute the KLdivergence,

$$D(P|P') = \sum_{\Gamma,\sigma} P(\Gamma, \sigma) \ln\frac{P(\Gamma, \sigma)}{P(\Gamma^\dagger, -\sigma)} = \frac{\langle\Sigma\rangle_F + \langle\Sigma\rangle_B}{2}, \tag{31}$$

and the total variation,

$$\begin{aligned} w &= \frac{1}{2}\sum_{\Gamma,\sigma}|P(\Gamma, \sigma) - P(\Gamma^\dagger, -\sigma)| \\ &= \frac{1}{2}\sum_\Gamma |P_F(\Gamma) - P_B(\Gamma^\dagger)|, \end{aligned} \tag{32}$$

so that the theorem reads

$$\frac{\langle\Sigma\rangle_F + \langle\Sigma\rangle_B}{2} \geqslant B(w), \tag{33}$$

given in terms of $P_F(\Gamma)$ and $P_B(\Gamma^\dagger)$. For systems where $\langle\Sigma\rangle_{F,B} = \beta(\langle W\rangle_{F,B} - \Delta_{F,B}F) := \beta\langle W_{\rm irr}\rangle_{F,B}$ (irreversible work [2]), the theorem results in a lower bound for the sum of forward and backward irreversible works,

$$\langle W_{\rm irr}\rangle_F + \langle W_{\rm irr}\rangle_B \geqslant 2k_bTB(w), \tag{34}$$

for $\beta = (k_bT)^{-1}$ the inverse temperature and ΔF is the free energy variation. Combinations of forward and backward ensemble averages similar to those in (33) and (34) also appeared in previous results in the context of thermodynamic uncertainty relations [43,44].

Application IV: Time rotation and speed limit. Consider a system that outputs stochastic events $s \in S$ for every run with probability $P(s)$ and $m(s) = s'$ is the time-reversed event (or any involution), so that (S, P, m) satisfies the theorem. For every run, a demon draws a Bernoulli random variable ϵ and decides if it plays the dynamics ($\epsilon = 1$) or plays it backwards ($\epsilon = 0$) for that run. Take $\langle\epsilon\rangle = p$ and $0 \leqslant p \leqslant 1$. Over several runs, the observer sees the events s with probability $Q(s)$

given by

$$Q(s) = pP(s) + (1-p)P(s'), \tag{35}$$

Note that (S, Q, m) also satisfies the theorem. Over several runs, the observer will measure $D(Q|Q')$ as an attempt to assess the entropy production. In this case, the theorem reads

$$D(P|P') \geqslant D(Q|Q') \geqslant B[|2p-1|\Delta(P, P')], \tag{36}$$

where the first inequality follows from the convexity of the KL divergence and the second follows from the theorem, using $Q(s) - Q(s') = (2p-1)[P(s) - P(s')]$. From the theorem, the observer gets a result for $D(Q|Q')$ that is lower bounded by $\Delta(P, P')$ from the original process. This setup works as a time rotation because the demon mixes forward s and backward s' directions. If $p = 1/2$, note that $Q(s) = P(s) + P(s') = Q(s')$ and the observer will get $D(Q|Q') = 0$, which mimics equilibrium. For some $|2p-1| > 0$, $D(Q|Q')$ is lower bounded in terms of the original bound for $D(P|P')$, as $\Delta(P, P') = B^{-1} \circ B(\Delta(P, P'))$ and $B(\Delta(P, P'))$ is the original lower bound (5).

Alternatively, in another setup, for a set of probabilities $\{P_1, \ldots, P_n\}$, if one has $\sum_i q_i = \tau$, applying the theorem for each P_i results in

$$\sum_i q_i D(P_i|P_i') \geqslant 2\sum_i q_i w_i \tanh^{-1}\left(\sum_j (q_j/\tau) w_i\right), \tag{37}$$

using $B''(x) > 0$ again, which can be easily inverted to obtain a lower bound for τ akin to the classic speed limit, already observed for Markov systems [39,40].

Discussion and conclusions. We proved a theorem showing that the KL divergence is bounded by a function of the TV. In this case, we required that the distributions considered are related by an involution, $m(m(s)) = s$. This seemly arbitrary condition produced a tighter bound than off the shelf results between the KLdivergence and the TV (Pinsker's and Bretagnolle-Hubber's).

We argue that the involution condition has a physical meaning. It is a general way to introduce the arrow of time: flipping it acts as an involution. Therefore, the events s and s' could be seen as flipped versions of the arrow of time. Because of that, as each event has only one flipped version, the events are able to be collected in pairs $\{s, m(s)\}$ that form the edges of an undirected graph (as well as pairs of the type $\{s, s\}$). This physical constraint allows the discrepancies of P and P', measured in the KL divergence and the TV, to be tighter than the usual bounds from statistics. Moreover, the theorem translates to a bound between the entropy production rate and the weight of the matching of a graph. The bound is particularly useful, for instance, in situations where the estimation of the weight of the graph is more feasible than the actual entropy production rate.

In this context, the formalism presented in the theorem is general enough to be applied in the derivation of other bounds about the statistics of P and P', when related by an involution, instead of using the fluctuation theorem [21–27,36,45–53] as a starting point. As examples of bounds about the statistics of P and P', one has the thermodynamic uncertainty relation (TUR) [21,22,54,55] for the variance of $\pi(s) = \ln[P(s)/P'(s)]$, as well as a bound for negative values of $\pi(s)$, $\sum_{\pi(s)<0} P(s)$ [56], related to apparent violations of the second law as discussed in the applications. In other words, one could formulate some the bounds from nonequilibrium thermodynamics simply in terms of (S, P, m). The extent of this analogy is left for further research.

Appendix. For the case $S = \mathbb{R}$, we have $P(s) \to p(s)ds$ and $\sum_s \to \int ds$. The distribution p' is given by $p'(s) := p(m(s))|dm(s)/ds|$, the KL divergence is given as $D(p|p') = \int p(s)\ln(p(s)/p'(s))ds$ and the TV is given as $\Delta(p, p') = (1/2)\int |p(s) - p'(s)|ds$. Now we define another probability density function $f : \mathbb{R} \to [0, 1]$ given by

$$f(x) = \int_{-\infty}^{\infty} p(s)\delta[\pi(s) - x]ds, \tag{A1}$$

where $\delta(y)$ is the Dirac delta function and $\pi(s) = \ln[p(s)/p'(s)]$ for $p(s) \neq 0$ and $\pi(s) = 0$ for $p(s) = 0$. We check the normalization,

$$\int_x f(x)dx = \int_s p(s)\int_x \delta(\pi(s) - x)dxds = \int_s p(s)ds = 1, \tag{A2}$$

and compute the average $\langle x \rangle$,

$$\langle x \rangle := \int_x xf(x)dx = \int_s \pi(s)p(s)ds = D(p|p'). \tag{A3}$$

The function $f(x)$ satisfies a strong detailed fluctuation theorem, $f(x) = \exp(x)f(-x)$, where it follows from

$$f(x) = \int_s e^{\pi(s)}p'(s)\delta[\pi(s) - x]ds \tag{A4}$$

$$= e^x \int_{s'} p(s')\delta[\pi(s') + x]ds' = e^x f(-x), \tag{A5}$$

We also compute the total variation $\Delta(f, \hat{f})$, for $\hat{f} := f(-x)$, resulting in

$$\Delta(f, \hat{f}) = \frac{1}{2}\int_s |p(s) - p'(s)|ds = \Delta(p, p'). \tag{A6}$$

Finally, we follow the same steps as in the discrete case (14). Let $\langle x \rangle = \langle |x|\tanh(|x|/2)\rangle$ and $\Delta(f, \hat{f}) = \langle \tanh(|x|/2)\rangle$. Using $f_a(x) = [\delta(x-a)\exp(a/2) + \delta(x+a)\exp(-a/2)]/[2\cosh(a/2)]$, which results in $q(|x|) = \delta(|x| - a)$ as the distribution of $|x|$. In this case, note that

$$\langle x \rangle_{f_a} = a\tanh(a/2) = B\big(\Delta\big(f_a|\hat{f}_a\big)\big), \tag{A7}$$

for $B(w) = 2w\tanh^{-1}(w)$, which means that

$$D(p|p') = \langle x \rangle = \int_0^{\infty} B\big(\Delta\big(f_a|\hat{f}_a\big)\big)q(a)da \geqslant B(\Delta(p, p')), \tag{A8}$$

from Jensen's inequality. Saturation also follows from the two-level system (44).

[1] G. T. Landi and M. Paternostro, Rev. Mod. Phys. **93**, 035008 (2021).
[2] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[3] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).
[4] C. Bustamante, J. Liphardt, and F. Ritort, Phys. Today **58**, 43 (2005).
[5] M. Esposito, U. Harbola, and S. Mukamel, Rev. Mod. Phys. **81**, 1665 (2009).
[6] R. Clausius, Ann. Phys. (Berlin) **169**, 481 (1854).
[7] R. Clausius, Ann. Phys. **201**, 353 (1865).
[8] S. Carnot, *Réflexions sur la Puissance Motrice du feu et sur les Machines Propres à Développer cette Puissance*, Annales scientifiques de l'É.N.S. 2e série, tome 1 (Bachelier, Paris, 1872), pp. 393–457.
[9] C. Jarzynskia, Eur. Phys. J. B **64**, 331 (2008).
[10] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[11] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).
[12] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).
[13] G. Gallavotti and E. G. D. Cohen, J. Stat. Phys. **80**, 931 (1995).
[14] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).
[15] P. Hänggi and P. Talkner, Nat. Phys. **11**, 108 (2015); T. Baumgratz, M. Cramer, and M. B. Plenio, Phys. Rev. Lett. **113**, 140401 (2014).
[16] K. Saito and Y. Utsumi, Phys. Rev. B **78**, 115429 (2008).
[17] N. Freitas, J.-C. Delvenne, and M. Esposito, Phys. Rev. X **11**, 031064 (2021).
[18] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Phys. Rev. Lett. **116**, 120601 (2016).
[19] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
[20] T. Tomé and M. C. De Oliveira, Phys. Rev. Lett. **108**, 020601 (2012).
[21] Y. Hasegawa and T. Van Vu, Phys. Rev. Lett. **123**, 110602 (2019).
[22] A. M. Timpanaro, G. Guarnieri, J. Goold, and G. T. Landi, Phys. Rev. Lett. **123**, 090604 (2019).
[23] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).
[24] N. Merhav and Y. Kafri, J. Stat. Mech.: Theory Exp. **2010** (2010).
[25] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[26] C. Jarzynski and D. K. Wójcik, Phys. Rev. Lett. **92**, 230602 (2004).
[27] M. Campisi, J. Pekola, and R. Fazio, New J. Phys. **17**, 035012 (2015).
[28] D. Witthaut, F. Hellmann, J. Kurths, S. Kettemann, H. Meyer-Ortmanns, and M. Timme, Rev. Mod. Phys. **94**, 015005 (2022).
[29] J. W. Essam and M. E. Fisher, Rev. Mod. Phys. **42**, 272 (1970).
[30] B. Bollobás, *Modern Graph Theory, Graduate Texts in Mathematics*, Vol. 184 (Springer, New York, 1998).
[31] R. A. Kullback, S. and Leibler, Ann. Math. Stat. **22**, 79 (1951).
[32] L. Devroye and G. Lugosi, Total variation, in *Combinatorial Methods in Density Estimation* (Springer, New York, 2001), pp. 38–46.
[33] I. Csiszár and J. Körner, (Cambridge University Press, Cambridge, England, 2011).
[34] J. Bretagnolle and C. Huber, Z. Wahrscheinlichkeitstheorie Verw. Geb. **47**, 119 (1979).
[35] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.106.L032101 for a discussion on the intuition behind the TV.
[36] D. S. P. Salazar, Phys. Rev. E **103**, 022122 (2021).
[37] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2007).
[38] R. Ziener, A. Maritan, and H. Hinrichsen, J. Stat. Mech.: Theory Exp. **2015**, P08014 (2015).
[39] A. Dechant, J. Phys. A: Math. Theor. **55**, 094001 (2022).
[40] V. T. Vo, T. V. Vu, and Y. Hasegawa, arXiv:2203.11501.
[41] N. Shiraishi, K. Funo, and K. Saito, Phys. Rev. Lett. **121**, 070601 (2018).
[42] T. Van Vu and Y. Hasegawa, Phys. Rev. Lett. **126**, 010601 (2021).
[43] P. P. Potts and P. Samuelsson, Phys. Rev. E **100**, 052137 (2019).
[44] K. Proesmans and J. M. Horowitz, J. Stat. Mech.: Theory Exp. (2019) 054005.
[45] R. García-García, D. Domínguez, V. Lecomte, and A. B. Kolton, Phys. Rev. E **82**, 030104(R) (2010).
[46] B. Cleuren, C. Van den Broeck, and R. Kawai, Phys. Rev. E **74**, 021117 (2006).
[47] D. Andrieux, P. Gaspard, T. Monnai, and S. Tasaki, New J. Phys. **11**, 043014 (2009).
[48] D. Luposchainsky, A. C. Barato, and H. Hinrichsen, Phys. Rev. E **87**, 042108 (2013).
[49] I. Neri, É. Roldán, and F. Jülicher, Phys. Rev. X **7**, 011019 (2017).
[50] S. Pigolotti, I. Neri, É. Roldán, and F. Jülicher, Phys. Rev. Lett. **119**, 140604 (2017).
[51] N. Yunger Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, Phys. Rev. E **97**, 052135 (2018).
[52] M. Campisi and L. Buffoni, Phys. Rev. E **104**, L022102 (2021).
[53] M. V. S. Bonança and S. Deffner, Phys. Rev. E **105**, L012105 (2022).
[54] A. C. Barato and U. Seifert, Phys. Rev. Lett. **114**, 158101 (2015).
[55] P. Pietzonka and U. Seifert, Phys. Rev. Lett. **120**, 190602 (2018).
[56] D. S. P. Salazar, Phys. Rev. E **104**, L062101 (2021).