

## Universal bounds on entropy production inferred from observed statistics

Eden Nitzan

*School of Physics & Astronomy, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel*Aishani Ghosal *Department of Biomedical Engineering, Tel Aviv University, Tel Aviv 6997801, Israel*Gili Bisker \**Department of Biomedical Engineering, Faculty of Engineering, The Center for Physics and Chemistry of Living Systems, The Center for Nanoscience and Nanotechnology, The Center for Light-Matter Interaction, Tel Aviv University, Tel Aviv 6997801, Israel*

(Received 29 November 2022; revised 27 April 2023; accepted 13 November 2023; published 15 December 2023)

Nonequilibrium processes break time-reversal symmetry and generate entropy. Living systems are driven out-of-equilibrium at the microscopic level of molecular motors that exploit chemical potential gradients to transduce free energy to mechanical work, while dissipating energy. The amount of energy dissipation, or the entropy production rate (EPR), sets thermodynamic constraints on cellular processes. Practically, calculating the total EPR in experimental systems is challenging due to the limited spatiotemporal resolution and the lack of complete information on every degree of freedom. Here, we propose an inference approach for a tight lower bound on the total EPR given partial information, based on an optimization scheme that uses the observed transitions and waiting times statistics. We introduce hierarchical bounds relying on the first- and second-order transitions, and the moments of the observed waiting time distributions, and apply our approach to two generic systems of a hidden network and a molecular motor, with lumped states. Finally, we show that a lower bound on the total EPR can be obtained even when assuming a simpler network topology of the full system.

DOI: [10.1103/PhysRevResearch.5.043251](https://doi.org/10.1103/PhysRevResearch.5.043251)

### I. INTRODUCTION

Advances in experimental techniques over the last few decades have opened new possibilities for studying systems at the single-molecule level [1–3]. In parallel, new theoretical approaches of stochastic thermodynamics for studying the physics of nonequilibrium, small fluctuating systems have emerged [4–6]. These include the mathematical relations describing symmetry properties of the stochastic quantities like work [7–10], heat [9], and entropy production [11,12], leading to fundamental limits on physical systems like heat engines [13–15], refrigerators [16], and biological processes [17,18].

Living systems operate far from equilibrium and constantly produce entropy. At the molecular level, the hydrolysis of fuel molecules, such as adenosine triphosphate (ATP), powers nonequilibrium cellular processes, utilizing part of the liberated free energy for physical work, while the rest is dissipated [5]. The dissipation, or entropy production, is a signature of irreversible processes and can be used as a direct measure of

the deviation from thermal equilibrium [19–22]. Therefore, the entropy production rate plays an important role in our understanding of the physics and underlying mechanism, governing biological and chemical processes [13–15,17,18,23].

Various studies have focused on estimating the mean entropy production rate using the thermodynamic uncertainty relations (TUR) using current fluctuations [24–29], fluctuations of first passage time [30,31], kinetic uncertainty relation in terms of the activity [32], or unified thermodynamic and kinetic uncertainty relations [33]. Other approaches utilize waiting-time distributions [34–36], machine learning [37–40], and single trajectory data [41–43]. Additional studies calculate higher moments of the full probability density function of the entropy production [44], use irreversible currents in stochastic dynamics described by a set of Langevin equations [45], or linear response theory [23].

Estimating the total EPR is only possible if we have knowledge regarding all of the degrees of freedom that are out of equilibrium [46,47]. However, due to practical limitations on the spatiotemporal resolution, not all of them can be experimentally accessible, and one can only obtain a lower bound on the total EPR for partially observed or coarse-grained systems [48–55].

The passive partial entropy production rate  $\sigma_{pp}$  is an estimator for the EPR calculated from the transitions between two observed states, which bounds the total EPR [48,56–58]. This estimator, however, fails to provide a nonzero bound in case of vanishing current over the observed link, i.e., at

\*bisker@tauex.tau.ac.il

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

stalling conditions [48]. Other EPR estimators for partially observed systems based on inequality relations like the TUR [24–26,32,33] also fail to provide a nontrivial bound on the total EPR in the absence of net flux in the system.

The Kullback-Leibler Divergence (KLD) estimator  $\sigma_{\text{KLD}}$  is based on the KLD, or the relative entropy, between the time-forward and the time-reversed path probabilities [21,46,59–64]. For semi-Markov processes, this estimator is a sum of two contributions. The first stems from transitions irreversibility or cycle affinities  $\sigma_{\text{aff}}$ , whereas the second stems from broken time-reversal symmetry reflected in irreversibility in waiting time distributions (WTD)  $\sigma_{\text{WTD}}$  [65]. Using the KLD estimator, one can obtain a nontrivial lower bound on the total EPR for second-order semi-Markov processes even in the absence of the net current [35,65–67]. Moreover, a lower bound on the total EPR can be obtained from the KLD between transition-based WTD [60,66,68].

Recently developed estimators solved an optimization problem to obtain a lower bound on the entropy production. For a discrete-time model, Ehrich proposed to search over the possible underlying systems that maintain the same observed statistics using knowledge on the number of hidden states [69]. For continuous-time models, Skinner and Dunkel minimized the EPR on a canonical form of the system that preserved the first- and second-order transition statistics to yield a lower bound on the total EPR  $\sigma_2$  [70]. The authors also formulated an optimization problem to infer the EPR in a system with two observed states using the waiting time statistics [34].

In this paper, we provide a tight bound on the total EPR by formulating an optimization problem based on the statistics of both transitions and waiting times. We use the first- and second-order statistics for the mass transition rates, and any chosen number of moments of the observed waiting time distributions. For a system with a known topology, we calculate the analytical expressions of the statistics as functions of the mass rates and the steady-state probabilities, which describe a possible underlying system and are used as variables in the optimization problem. These analytical expressions are then used to constrain the optimization variables to match the observed statistics. We show for a few continuous-time Markov chain systems that using the constraints of the mass rates and only the first moment of the WTD already provides close-to-total EPR value. Our approach outperforms other estimators, such as  $\sigma_{\text{pp}}$ ,  $\sigma_{\text{KLD}}$ ,  $\sigma_{\text{aff}}$ , and  $\sigma_2$ , in terms of the tightness of the lower bound. In the case of a complex model, where the formulation of the optimization problem might not be practical due to the number of constraints, or in case the full topology is not known, we show numerically that assuming a simpler underlying topology can provide a lower bound on the total EPR.

The paper is organized as follows. In Sec. II, we describe our model system and the coarse-graining approach. The results are presented in Sec. III: We discuss the estimator in Sec. III A, apply it to different systems in Sec. III B, demonstrate how the accuracy of the measured statistics affects the results of our estimator in Sec. III C, and finally, we show the results of the optimization problem assuming a simpler underlying model in Sec. III D. We conclude our findings in Sec. IV.

## II. MODEL

We assume a continuous time Markov chain over a finite and discrete set of states  $i = \{1, 2, \dots, N\}$ . A trajectory is described by a sequence of states and their corresponding residence times before a transition to the next state occurs. Being a Markovian process, the jump probabilities depend only on the current state.

The transition rates  $w_{ij}$  from state  $i$  to  $j$  determine the time evolution of the probabilities for the system to be in each state, according to the Master equation  $\frac{d}{dt}\mathbf{p}(t)^T = \mathbf{p}(t)^T\mathbf{W}$ , where  $T$  is the transpose operator, and  $\mathbf{W}$  is the rate matrix

$$[\mathbf{W}]_{ij} = \begin{cases} w_{ij} & j \neq i \\ -\lambda_i & j = i \end{cases} \quad (1)$$

$\mathbf{p}(t)$  is a column vector of the state probabilities at time  $t$ , with  $\sum_i p_i(t) = 1$ , and the diagonal entries are calculated according to  $\lambda_i = \sum_{j \neq i} w_{ij}$  for probability conservation.

At the long-time limit, the system eventually reaches a steady state  $\boldsymbol{\pi}$ , where  $\lim_{t \rightarrow \infty} p_i(t) = \pi_i$  such that  $0 = \frac{d}{dt}\boldsymbol{\pi}^T = \boldsymbol{\pi}^T\mathbf{W}$  [71].

The waiting time at each state  $i$  is an exponential random variable with mean waiting time of  $\tau_i = \lambda_i^{-1}$ .

The mass rates  $n_{ij}$  are defined as follows:

$$n_{ij} = \begin{cases} \pi_i w_{ij} & j \neq i \\ 0 & j = i \end{cases} \quad (2)$$

The probabilities of jumping from state  $i$  to state  $j$  can be written in terms of the mass rates,

$$p_{ij} = \frac{w_{ij}}{\lambda_i} = \frac{n_{ij}}{\sum_{j' \neq i} n_{ij'}} \quad (3)$$

The steady-state total EPR can be calculated by multiplying the net currents and the mass rate ratios (affinities), summing over all the links [5,6],

$$\begin{aligned} \sigma_{\text{tot}} &= \sum_{i,j} \pi_i w_{ij} \log \left( \frac{\pi_i w_{ij}}{\pi_j w_{ji}} \right) \\ &= \sum_{i,j} n_{ij} \log \left( \frac{n_{ij}}{n_{ji}} \right) \\ &= \sum_{i < j} (n_{ij} - n_{ji}) \log \left( \frac{n_{ij}}{n_{ji}} \right). \end{aligned} \quad (4)$$

Given a long trajectory of a total duration  $T$ , the steady-state probability  $\pi_i$  is the fraction of time spent in state  $i$ , and the mass rate  $n_{ij}$  is the number of transitions  $i \rightarrow j$  divided by  $T$ .

According to the definition of the mass rates in Eq. (2), at the steady state, mass conservation is satisfied at each state,

$$\forall_i : \sum_j n_{ij} = \sum_j n_{ji}. \quad (5)$$

In many practical scenarios, some of the microstates cannot be distinguished, and the transitions between them cannot be observed. In such a case, a set of states  $\{i_1, i_2, \dots, i_{N_i}\}$  is observed as a single coarse-grained state  $I$  [Fig. 1(a)]. The observed trajectory, therefore, includes only coarse-grained states and the combined residence time [Fig. 1(b)], and it is not necessarily a Markovian process [65]. Such a

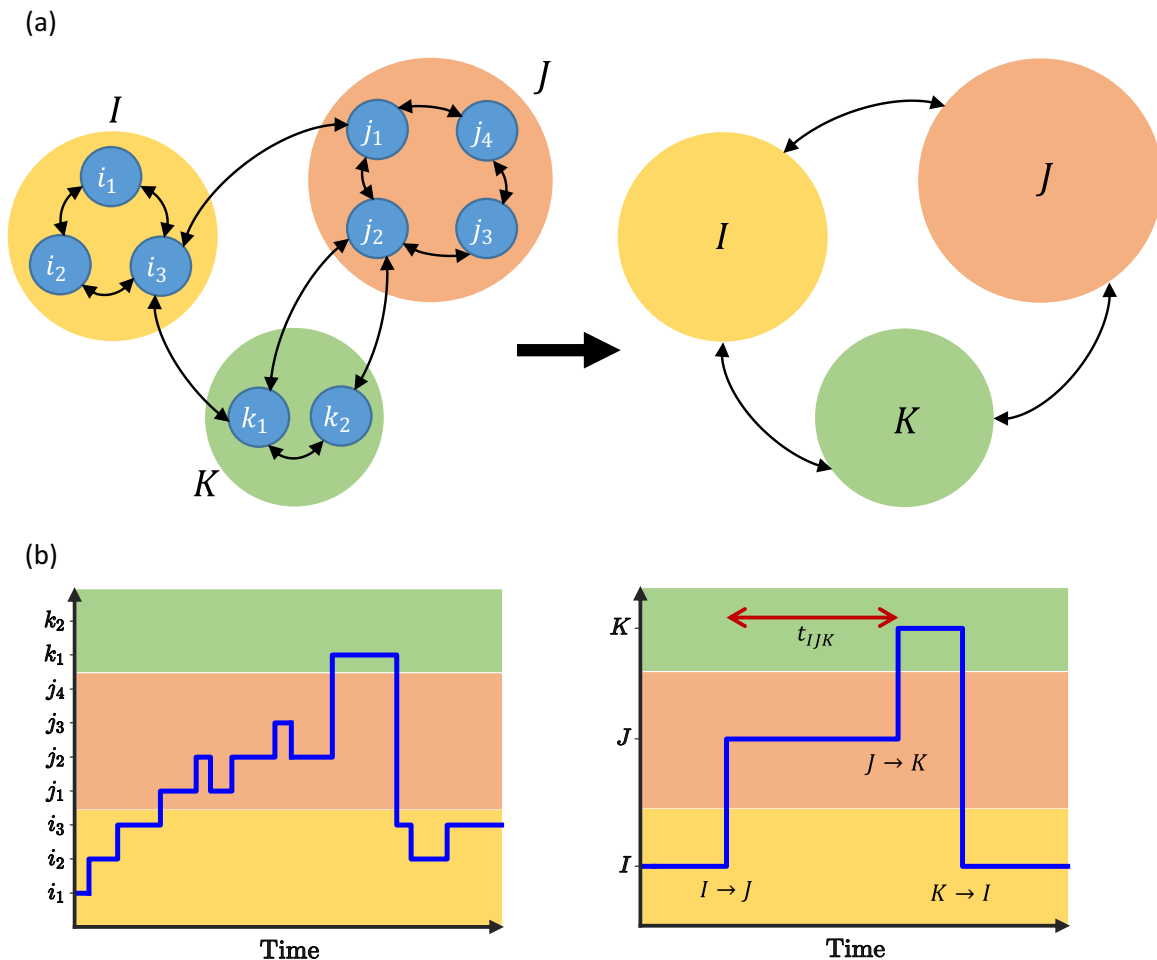


FIG. 1. Coarse graining. (a) The full Markovian system (left) and the coarse-grained system (right). (b) An example of a full trajectory (left) containing the actual states and the corresponding coarse-grained trajectory (right) containing only the observed states.

decimation procedure of lumping several states can give rise to semi-Markovian processes of any order depending on the topology of the network [70,72–74]. In this case, the observed statistics of two or more consecutive transitions may give us additional information on the process.

### III. RESULTS

#### A. Bounding the entropy production rate

Given a coarse-grained system with a model of the full underlying Markovian network topology, we can formulate an optimization problem for obtaining a tight bound on the total EPR. We consider a few observables: the coarse-grained steady-state probabilities  $\pi_I$ , which is the probability of observing the system in the coarse-grained state  $I$ ; the first-order mass rates  $n_{IJ}$ , which is the rate of observing the transition  $I \rightarrow J$ ; the second-order mass rates  $n_{IJK}$ , which is the rate of observing the transition  $I \rightarrow J$  followed by the transition  $J \rightarrow K$ ; and the conditional waiting time distributions  $\psi_{IJK}(t)$ , which is the distribution of waiting times in a coarse-grained state  $J$  before a transition to a coarse-grained state  $K$  occurs, conditioned on the previous transition being  $I \rightarrow J$ .

We search over the space of all possible underlying systems with the same topology as our hypothesized Markovian

model that give rise to the same observed statistics, while minimizing the EPR. Trivially, the EPR of the coarse-grained system at hand is bounded from below by the EPR of the underlying Markovian system with the same observed statistics after coarse graining, having the minimal value of entropy production.

#### 1. Analytical expressions of the observed statistics

The observed statistics of the coarse-grained system can be expressed analytically in terms of the mass rates and steady-state probabilities of the model underlying system. From probability and mass conservation  $\pi_I = \sum_{i \in I} \pi_i$ , and  $n_{IJ} = \sum_{i \in I, j \in J} n_{ij}$ , respectively. The mass conservation for the second-order transitions  $n_{IJK}$  must include all the paths starting at state  $i \in I$ , passing through a state in  $J$ , where any number of transitions might occur inside  $J$ , and jumping to state  $k \in K$ . To account for the transitions within  $J$ , we define the matrix  $P_{JJ}$  of the transition probabilities between states in  $J$ ,  $j_m, j_n \in J$ ,

$$[P_{JJ}]_{mn} = \begin{cases} p_{j_m j_n} & m \neq n \\ 0 & m = n \end{cases} \tag{6}$$

Summing over the possible transitions from  $I$ , transitions within  $J$ , and transitions to  $K$ , we have (see Appendix A)

$$n_{IJK} = \sum_{i \in I, k \in K} \mathbf{n}_{ij}^T [\mathbb{I} - \mathbf{P}_{JJ}]^{-1} \mathbf{p}_{jk}, \quad (7)$$

where  $\mathbb{I}$  is the identity matrix of the size of  $\mathbf{P}_{JJ}$ , and  $\mathbf{n}_{ij}$  and  $\mathbf{p}_{jk}$  are column vectors of the mass rates from state  $i \in I$  to any state  $j \in J$ , and jump probabilities from any state  $j \in J$  to a state  $k \in K$ , respectively,

$$\mathbf{n}_{ij}^T = [n_{ij_1}, n_{ij_2}, \dots, n_{ij_{N_j}}] \quad (8)$$

and

$$\mathbf{p}_{jk}^T = [p_{j_1k}, p_{j_2k}, \dots, p_{j_{N_j}k}]. \quad (9)$$

The conditional waiting time distribution  $\psi_{IJK}(t)$  can be calculated by the Laplace and inverse-Laplace transforms (full derivations can be found in Appendix B). We start from the Laplace transform of  $\psi_{ij}(t) = w_{ij}e^{-\lambda_i t}$ , the joint probability distribution of the transition  $i \rightarrow j$  and the waiting time in the Markovian state  $i$ ,

$$\tilde{\psi}_{ij}(s) = \mathcal{L}\{\psi_{ij}(t)\} = \int_0^\infty \psi_{ij}(t)e^{-ts} dt = \frac{w_{ij}}{s + \lambda_i}. \quad (10)$$

Note that for any function  $f(t)$ ,  $\tilde{f}(s \rightarrow 0) = \int_0^\infty f(t)e^{-ts} dt|_{s \rightarrow 0} = \int_0^\infty f(t) dt$  is the normalization of  $f(t)$ . Here,  $\psi_{ij}(t)$  is normalized to  $p_{ij}$ , i.e.,  $p_{ij} = \int_0^\infty \psi_{ij}(t) dt$  [Eq. (3)].

Now, we consider the simple case where the second-order transition through the coarse-grained state  $J$  starts and ends in specific Markovian states  $i \in I$  and  $k \in K$ , respectively. The Laplace transform of the distribution of waiting times in  $J$  before a transition to  $k$  occurs, given the previous transition was  $i \rightarrow J$  is

$$\tilde{\psi}_{ijk}(s) = \frac{\mathbf{p}_{ij}^T}{\sum_{j \in J} p_{ij}} [\mathbb{I} - \tilde{\Psi}_{JJ}(s)]^{-1} \tilde{\psi}_{jk}(s) \quad (11)$$

where

$$\tilde{\Psi}_{Jk}^T(s) = [\tilde{\psi}_{j_1k}(s), \tilde{\psi}_{j_2k}(s), \dots, \tilde{\psi}_{j_{N_j}k}(s)] \quad (12)$$

and  $\tilde{\Psi}_{JJ}(s)$  is a matrix of the Laplace transforms of every joint probability distribution of waiting times and transitions within  $J$ ,

$$\tilde{\Psi}_{JJ}(s) = \begin{cases} \tilde{\psi}_{j_m j_n}(s) & m \neq n \\ 0 & m = n \end{cases}. \quad (13)$$

We denote  $\tilde{\psi}_{iJK}(s) \equiv \sum_{k \in K} \tilde{\psi}_{ijk}(s)$ . Then, the Laplace transform of the conditional waiting time distribution is

$$\tilde{\psi}_{IJK}(s) = \sum_{i \in I} \frac{\pi_i}{\pi_I} \frac{\tilde{\psi}_{iJK}(s)}{\tilde{\psi}_{iJK}(s \rightarrow 0)}. \quad (14)$$

Finally, we apply an inverse Laplace transform to obtain the conditional probability density

$$\psi_{IJK}(t) = \mathcal{L}^{-1}\{\tilde{\psi}_{IJK}(s)\}. \quad (15)$$

We further impose mass conservation at each of the Markovian states according to Eq. (5), to make sure the solution represents a valid Markovian system.

## 2. Formalizing the optimization problem

Let  $\mathcal{S}$  be the real underlying Markovian system and let  $\mathcal{R}$  be a general underlying system with the same topology as  $\mathcal{S}$ , i.e., the same states and possible transitions as  $\mathcal{S}$ , but  $\mathcal{R}$  can have arbitrary mass rates and steady-state probabilities. Given the set of all systems  $\mathcal{R}$  with the same steady-state probabilities  $\pi_i^{\mathcal{R}} = \pi_i^{\mathcal{S}}$ , same first-order mass rates  $n_{ij}^{\mathcal{R}} = n_{ij}^{\mathcal{S}}$ , same second-order mass rates  $n_{iJK}^{\mathcal{R}} = n_{iJK}^{\mathcal{S}}$ , and the same conditional waiting time distributions  $\psi_{IJK}^{\mathcal{R}}(t) = \psi_{IJK}^{\mathcal{S}}(t)$ , as the system  $\mathcal{S}$ , the following inequality holds for the EPR of  $\mathcal{S}$  and  $\mathcal{R}$ ,  $\sigma(\mathcal{S})$  and  $\sigma(\mathcal{R})$ , respectively,

$$\sigma_{\text{tot}}(\mathcal{S}) \geq \min_{\mathcal{R}} \{ \sigma_{\text{tot}}(\mathcal{R}) | \forall_{I,J,K} : \pi_i^{\mathcal{R}} = \pi_i^{\mathcal{S}}, \quad n_{ij}^{\mathcal{R}} = n_{ij}^{\mathcal{S}}, \quad n_{iJK}^{\mathcal{R}} = n_{iJK}^{\mathcal{S}}, \quad \psi_{IJK}^{\mathcal{R}}(t) = \psi_{IJK}^{\mathcal{S}}(t) \} \equiv \sigma_{\text{opt}}^{(\infty)}, \quad (16)$$

where  $\sigma_{\text{opt}}^{(\infty)}$  is the minimal EPR value of all the possible underlying systems  $\mathcal{R}$ . The inequality holds since the real system  $\mathcal{S}$  belongs to the set of systems over which we minimize. The only variables of the optimization problem are  $n_{ij}$  and  $\pi_i$ , from which one can fully describe any of the possible underlying Markovian systems  $\mathcal{R}$ . All the constraints,  $\pi_i$ ,  $n_{ij}$ ,  $n_{iJK}$ , and  $\psi_{IJK}(t)$ , as well as the EPR objective function, depend on these variables. Note that these variables are bounded by  $0 \leq \pi_i \leq \pi_I$  and  $0 \leq n_{ij} \leq n_{ij}$ .

In contrast to the constraints on the steady-state probabilities and the first- and second-order mass rate values, the constraint on the waiting-time distributions requires an equality of continuous functions  $\psi_{IJK}(t)$ , which one cannot fully reconstruct from trajectory data of finite duration. Moreover, solving the optimization problem using a constraint on a function with nontrivial dependency on the optimization problem variables is extremely challenging. Thus, we modify the optimization, and instead, use the moments of the waiting time distributions,

$$\sigma_{\text{opt}}^{(n)}(\mathcal{S}) \equiv \min_{\mathcal{R}} \{ \sigma_{\text{tot}}(\mathcal{R}) | \forall_{I,J,K} : \pi_i^{\mathcal{R}} = \pi_i^{\mathcal{S}}, \quad n_{ij}^{\mathcal{R}} = n_{ij}^{\mathcal{S}}, \quad n_{iJK}^{\mathcal{R}} = n_{iJK}^{\mathcal{S}}, \quad \forall_{k \in \{1,2,\dots,n\}} : \langle t_{IJK}^k \rangle^{\mathcal{R}} = \langle t_{IJK}^k \rangle^{\mathcal{S}} \}, \quad (17)$$

where  $\langle t_{IJK}^k \rangle$  is the  $k$ th moment of the conditional waiting time distribution  $\psi_{IJK}(t)$ . Using an increasing number of moments, we can write the hierarchical bounds,

$$\forall_{n \in \mathbb{N}} : \sigma_{\text{tot}}(\mathcal{S}) \geq \sigma_{\text{opt}}^{(\infty)}(\mathcal{S}) \geq \sigma_{\text{opt}}^{(n)}(\mathcal{S}) \geq \dots \geq \sigma_{\text{opt}}^{(1)}(\mathcal{S}). \quad (18)$$

We can easily get the analytical expressions for the moments  $\langle t_{IJK}^k \rangle$  from the Laplace transform (see Appendix B),

$$\langle t_{IJK}^k \rangle = (-1)^k \frac{d^k \tilde{\psi}_{IJK}(s)}{ds^k} |_{s \rightarrow 0}. \quad (19)$$

Now, for each moment, we have an expression that depends on the optimization problem variables in a simpler way, which in turn, simplifies the calculations. After calculating the values of the observables for the optimization problem, we solve it using a global search nonlinear optimization algorithm [75].

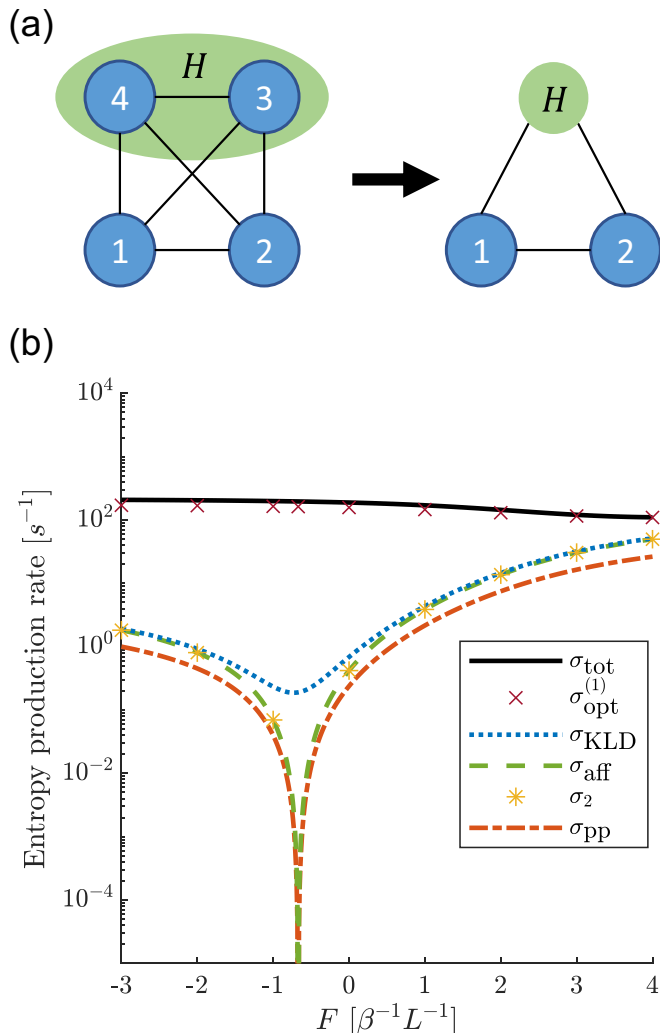


FIG. 2. Four-state system. (a) Illustration of the full four-state system topology, including the coarse graining of states 3 and 4 to state  $H$ . (b) Total EPR  $\sigma_{\text{tot}}$  (solid black line), our bound  $\sigma_{\text{opt}}^{(1)}$  (brown cross), KLD estimator  $\sigma_{\text{KLD}}$  (dotted blue line), affinity estimator  $\sigma_{\text{aff}}$  (dashed green line), two-step estimator  $\sigma_2$  (yellow Asterisk), and the passive partial entropy production  $\sigma_{\text{pp}}$  (dashed-dotted orange line). The rates used are  $w_{12} = 3 \text{ s}^{-1}$ ,  $w_{13} = 0 \text{ s}^{-1}$ ,  $w_{14} = 8 \text{ s}^{-1}$ ,  $w_{21} = 2 \text{ s}^{-1}$ ,  $w_{23} = 50 \text{ s}^{-1}$ ,  $w_{24} = 0.2 \text{ s}^{-1}$ ,  $w_{31} = 0 \text{ s}^{-1}$ ,  $w_{32} = 2 \text{ s}^{-1}$ ,  $w_{34} = 75 \text{ s}^{-1}$ ,  $w_{41} = 1 \text{ s}^{-1}$ ,  $w_{42} = 35 \text{ s}^{-1}$ ,  $w_{43} = 0.7 \text{ s}^{-1}$ .

## B. Examples

### 1. Four-state system

We consider a fully-connected network of four states, with two Markovian observed states  $\{1, 2\}$  and two hidden states  $\{3, 4\}$ , which are coarse grained to state  $H$  [Fig. 2(a)], resulting in second-order semi-Markov dynamics [65]. The observed statistics of interest are the steady-state probabilities  $\pi_1$ ,  $\pi_2$ , and  $\pi_H$ , the first-order mass rates  $n_{1H}$ ,  $n_{H1}$ ,  $n_{2H}$ , and  $n_{H2}$ , the second-order mass rates  $n_{1H2}$  and  $n_{2H1}$  and the  $k$ th moment of the conditional waiting time distributions  $\langle t_{1H1}^k \rangle$ ,  $\langle t_{1H2}^k \rangle$ ,  $\langle t_{2H1}^k \rangle$ , and  $\langle t_{2H2}^k \rangle$ . Notice we only used the second-order statistics through the coarse-grained state  $H$ , since states 1 and 2 are Markovian. Furthermore, we do not use  $n_{1H1}$  and  $n_{2H2}$  since they depend on the other mass rates:  $n_{1H1} =$

$n_{1H} - n_{1H2}$  and  $n_{2H2} = n_{2H} - n_{2H1}$ . The derivations of the analytical expressions of the second-order mass rates and the moments of the conditional waiting time moments, for this system, can be found in Appendix C.

We tune the transition rates over the observed link between states 1 and 2 according to  $w_{12}(F) = w_{12}e^{-\beta FL}$  and  $w_{21}(F) = w_{21}e^{\beta FL}$ , where  $\beta = T^{-1}$  is the inverse temperature (with  $k_B = 1$ ), and  $L$  is a characteristic length scale, to mimic external forcing. We compare the different EPR estimators on the system for several values for a driving force  $F$  over the observed link [Fig. 2(b)].

The passive partial EPR [48]

$$\begin{aligned} \sigma_{\text{pp}} &= (\pi_1 w_{12} - \pi_2 w_{21}) \log \left( \frac{\pi_1 w_{12}}{\pi_2 w_{21}} \right) \\ &= (n_{12} - n_{21}) \log \left( \frac{n_{12}}{n_{21}} \right). \end{aligned} \quad (20)$$

The KLD estimator is the sum of two contributions,

$$\begin{aligned} \sigma_{\text{KLD}} &= \sigma_{\text{aff}} + \sigma_{\text{WTD}} \\ &= \frac{1}{\mathcal{T}} \sum_{I,J,K} p_{IJK} \log \left( \frac{p([IJ] \rightarrow [JK])}{p([KJ] \rightarrow [JI])} \right) \\ &\quad + \frac{1}{\mathcal{T}} \sum_{I,J,K} p_{IJK} D[\psi_{IJK}(t) || \psi_{KJI}(t)], \end{aligned} \quad (21)$$

where  $p([IJ] \rightarrow [JK])$  is the probability to observe the transition  $J \rightarrow K$  given the previous transition was  $I \rightarrow J$ ,  $p_{IJK}$  is the probability to observe the second-order transition  $I \rightarrow J \rightarrow K$ , and  $D[p||q]$  is the KLD between the probability distributions  $p$  and  $q$ . As was previously shown, the hierarchy between the EPR estimators is  $\sigma_{\text{KLD}} \geq \sigma_{\text{aff}} \geq \sigma_{\text{pp}}$  [48,65].

The  $\sigma_2$  estimator is also formulated as an optimization problem searching over a canonical form of the system with the same observed statistics; however, it only considers the first- and second-order mass rates [70]. Its place in the hierarchy between the EPR estimators varies for different systems. While  $\sigma_2$  can be greater than  $\sigma_{\text{KLD}}$  in some cases [70], here, for the rate values we used,  $\sigma_2 < \sigma_{\text{KLD}}$ . In fact, although the values of  $\sigma_2$  and  $\sigma_{\text{aff}}$  appear to be similar [Fig. 2(b)], actually  $\sigma_2 < \sigma_{\text{aff}}$  for all of the values of  $F$  used.

At the stalling force, there is no current in the visible link, and we get  $\sigma_{\text{pp}} = \sigma_{\text{aff}} = \sigma_2 = 0$ , which is the trivial bound. In contrast,  $\sigma_{\text{KLD}}$  and our estimator  $\sigma_{\text{opt}}^{(1)}$  give a nontrivial bound. Moreover,  $\sigma_{\text{opt}}^{(1)}$  surpasses  $\sigma_{\text{KLD}}$  significantly and yields a tight bound. For this system, using higher moments in order to calculate  $\sigma_{\text{opt}}^{(2)}$  did not make any improvement compared to  $\sigma_{\text{opt}}^{(1)}$ .

While the example in this section is of a system with three observed states, two of which are Markovian, our approach can be generalized to any system. For example, see Appendix E for a system with four observed states, three of which are Markovian, where the  $\sigma_{\text{opt}}^{(1)}$  estimator still outperforms  $\sigma_{\text{KLD}}$  and  $\sigma_2$ .

### 2. Molecular motor

Here, we study a model of a molecular motor, illustrated in Fig. 3(a). The motor can physically move in space

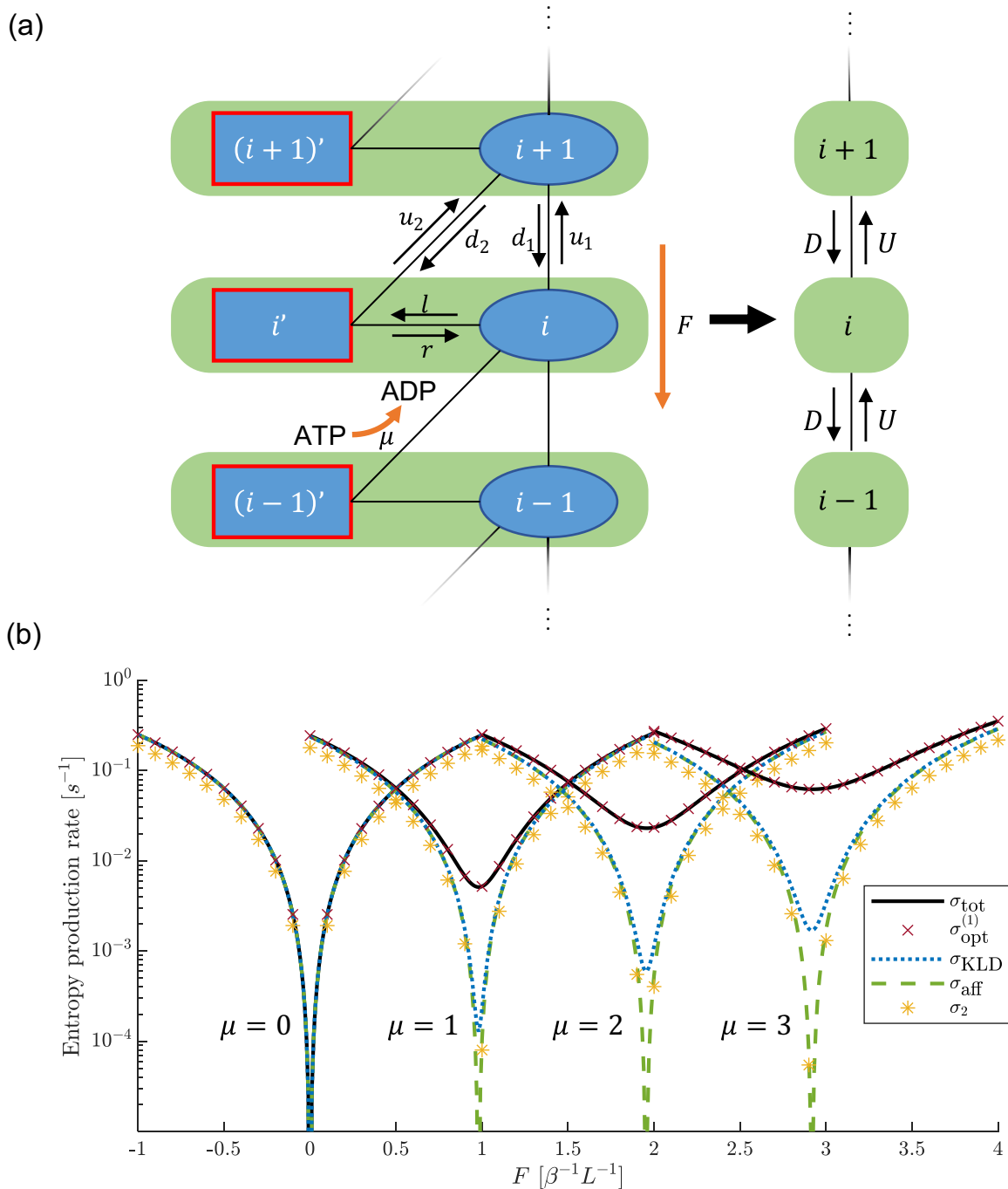


FIG. 3. Molecular motor. (a) Illustration of the full molecular motor system, including the coarse graining of the active (red boxed square) and passive (ellipse) states. (b) Total EPR  $\sigma_{\text{tot}}$  (solid black line), our bound  $\sigma_{\text{opt}}^{(1)}$  (brown cross), KLD estimator  $\sigma_{\text{KLD}}$  (dotted blue line), the affinity estimator  $\sigma_{\text{aff}}$  (dashed green line), and the two-step estimator  $\sigma_2$  (yellow asterisk). The rates used are  $w_r = w_l = w_{u2} = w_{d2} = 1 \text{ s}^{-1}$ ,  $w_{u1} = w_{d1} = 0.01 \text{ s}^{-1}$ .

(upward or downward),  $i \leftrightarrow i + 1$ , or change internal states (passive or active),  $i \leftrightarrow i'$ . An external source of chemical work  $\Delta\mu$  drives the upward spatial jumps from the active state, and a mechanical force  $F$  acts against it and drives the downward transitions. We assume that an external observer cannot distinguish between the internal states of the motor, but rather can only record its physical position. The observed statistics are thus of a second-order semi-Markov process [65].

Owing to the transnational symmetry in the model, we represent the molecular motor as a cyclic network of three coarse-grained states where each of them represents the physical location, lumping the active and passive internal states. We denote the steady-state probability of being in the passive and active states as  $\pi$  and  $\pi'$ , respectively. Notice that the probability of being in each physical location in the three-state cyclic system is the same, and that  $\pi$  and  $\pi'$  are the same for all of the physical locations, therefore,  $\pi + \pi' = 1/3$ .

We denote the upward and downward transitions from and to the passive state as  $u_1$  and  $d_1$ , respectively, the upward and downward transitions from and to the active state as  $u_2$  and  $d_2$ , respectively, and the transitions between the active and passive states at the same physical location as  $r$  (right) and  $l$  (left), respectively. The upward and downward coarse-grained transitions are labeled as  $U$  and  $D$ , respectively.

The observed statistics of interest are the first-order mass rates  $n_U$ ,  $n_D$ , the second-order mass rates  $n_{UU}$ ,  $n_{DD}$ , and the  $k$ th moment of the conditional waiting times  $\langle t_{UU}^k \rangle$ ,  $\langle t_{UD}^k \rangle$ ,  $\langle t_{DU}^k \rangle$ , and  $\langle t_{DD}^k \rangle$ . Note that we do not use  $n_{UD}$  and  $n_{DU}$ , since they depend on the other mass rates,  $n_{UD} = n_U - n_{UU}$  and  $n_{DU} = n_D - n_{DD}$ . Owing to the symmetry of the cycle representation of the coarse-grained system, in which the steady-state probabilities are equally distributed, we only need the constraints on the upward and downward transitions. The derivations of the analytical expressions of the second-order mass rates and the moments of the conditional waiting time distributions, for this system, can be found in Appendix D.

The chemical affinity  $\mu$ , arising from ATP hydrolysis for example, only affects the transitions  $u_2$  and  $d_2$ , whereas the external force  $F$  affects all of the spatial transitions  $u_1$ ,  $d_1$ ,  $u_2$ , and  $d_2$ . The transition rates then obey local detailed balance,  $w_{d1}/w_{u1} = e^{\beta FL}$  and  $w_{d2}/w_{u2} = e^{\beta(FL-\mu)}$ , where  $L$  is the length of a single spatial jump [65].

We compare the different EPR estimators for the molecular motor system for several values of  $\mu$ , and for each  $\mu$  value, we tune the external forcing parameter  $F$  [Fig. 3(b)]. Notice the passive partial EPR ( $\sigma_{pp}$ ) is not applicable for this system since all the original Markovian states are coarse grained.

The hierarchy of the different EPR estimators for the molecular motor, for the rate values we used, is  $\sigma_{\text{opt}}^{(1)} \geq \sigma_{\text{KLD}} \geq \sigma_{\text{aff}} \geq \sigma_2$ . At the stalling force for each value of  $\mu$ , where there is no visible current, we find  $\sigma_{\text{aff}} = \sigma_2 = 0$ , which is the trivial bound. In contrast, similar to the four-state system,  $\sigma_{\text{opt}}^{(1)}$  surpasses  $\sigma_{\text{KLD}}$  significantly and yields a tight bound.

### C. Importance of data accuracy

One of the hyperparameters defining the optimization problem is the constraint tolerance, which indicates the acceptable numerical error of the solution. If  $\epsilon$  is the absolute error of the trajectory statistics with respect to the true analytical ones, then the constraint tolerance must be equal to or greater than  $\epsilon$ . Otherwise, when the constraint tolerance is smaller than the absolute error of the statistics, the optimization problem might not converge or give an overestimation in the worst-case scenario.

In Fig. 4, we plot the absolute (and relative) error of a few statistics values calculated from several trajectories as a function of the trajectory length  $N$ , for both systems discussed in the previous sections. Moreover, using the analytical values of the statistics for maximum accuracy, we plot the results of our estimator  $\sigma_{\text{opt}}^{(1)}$  as a function of the constraint tolerance.

As expected, longer trajectory data result in a more accurate estimation of the observed statistics used for our optimization problem for both systems, as evident from the values of  $n_{1H}$ ,  $n_{1H2}$ , and  $\langle t_{1H2} \rangle$  for the four-state system [Fig. 4(a)], and from the values of  $n_U$ ,  $n_{UU}$ , and  $\langle t_{UU} \rangle$  for

the molecular motor [Fig. 4(b)]. For smaller errors, we can use a smaller constraint tolerance. The error bars in Figs. 4(a) and 4(b), which are the standard deviation of the values of the observables in different realizations of trajectories with the same size, can be used as a scale for the appropriate constraint tolerance.

For both systems, smaller constraint tolerance leads to a better estimator as the value of the lower bound on the EPR approaches the true analytical value [Figs. 4(c) and 4(d)], demonstrating the importance of an accurate estimation of the observables. We emphasize that data accuracy impacts all EPR estimators that rely on observed statistics, including  $\sigma_{\text{KLD}}$  and  $\sigma_2$ . The reliability of the estimation is inherently connected to the precision of the inferred observables, where longer trajectories naturally lead to smaller errors.

### D. Optimizing a simple model

Although our approach can be generalized to any number of hidden states, the analytical expressions for the observables become complicated, and the number of variables increases for a more complex coarse-grained topology. In turn, solving the optimization problem would require longer computation times. In order to test the performance of our estimator, we solved the optimization problem for a larger number of hidden states in a fully-connected network of four, five, and six states with only two Markovian observed states, assuming only two states are coarse grained [Fig. 5(a)]. Similarly, we tested the performance of our estimator for the case of the molecular motor with two, three, and four internal states at each physical position, assuming there are only two. While generally, the estimator gives a more accurate result for the case of the two hidden state, which matches the assumption, it still provides a lower bound on the total EPR with comparable accuracy for a larger number of hidden states in the two systems [Figs. 5(b) and 5(c)]. Moreover, our  $\sigma_{\text{opt}}^{(1)}$  estimator can still outperform other estimators,  $\sigma_{\text{KLD}}$  and  $\sigma_2$ , as demonstrated for the case of a five-state system with two Markovian observed states and one coarse-grained state of three internal states in Appendix F.

## IV. CONCLUSIONS

We present an estimator for the entropy production rate, which gives a tight bound by formulating an optimization problem using both transitions and waiting times statistics. Such data is readily available in different experimental settings, for example, tracking a bead attached to a rotating bacterial flagellum [70,76], or cortical granules embedded in the actomyosin cortex of an oocyte [77]. Our estimator can be applied to any system with a known topology, and it significantly surpasses previous estimators, as demonstrated for the two studied systems, the fully-connected hidden network, and the molecular motor. The variables for the optimization problem can be inferred from the observed statistics, where longer trajectories result in more accurate estimation and enable a smaller constraint tolerance value. Finally, for both systems, our approach can provide a lower bound on the total EPR for more complex systems, assuming a simpler underlying topology of the hidden states. Although we numerically showed that searching over all systems with a simpler topology of the

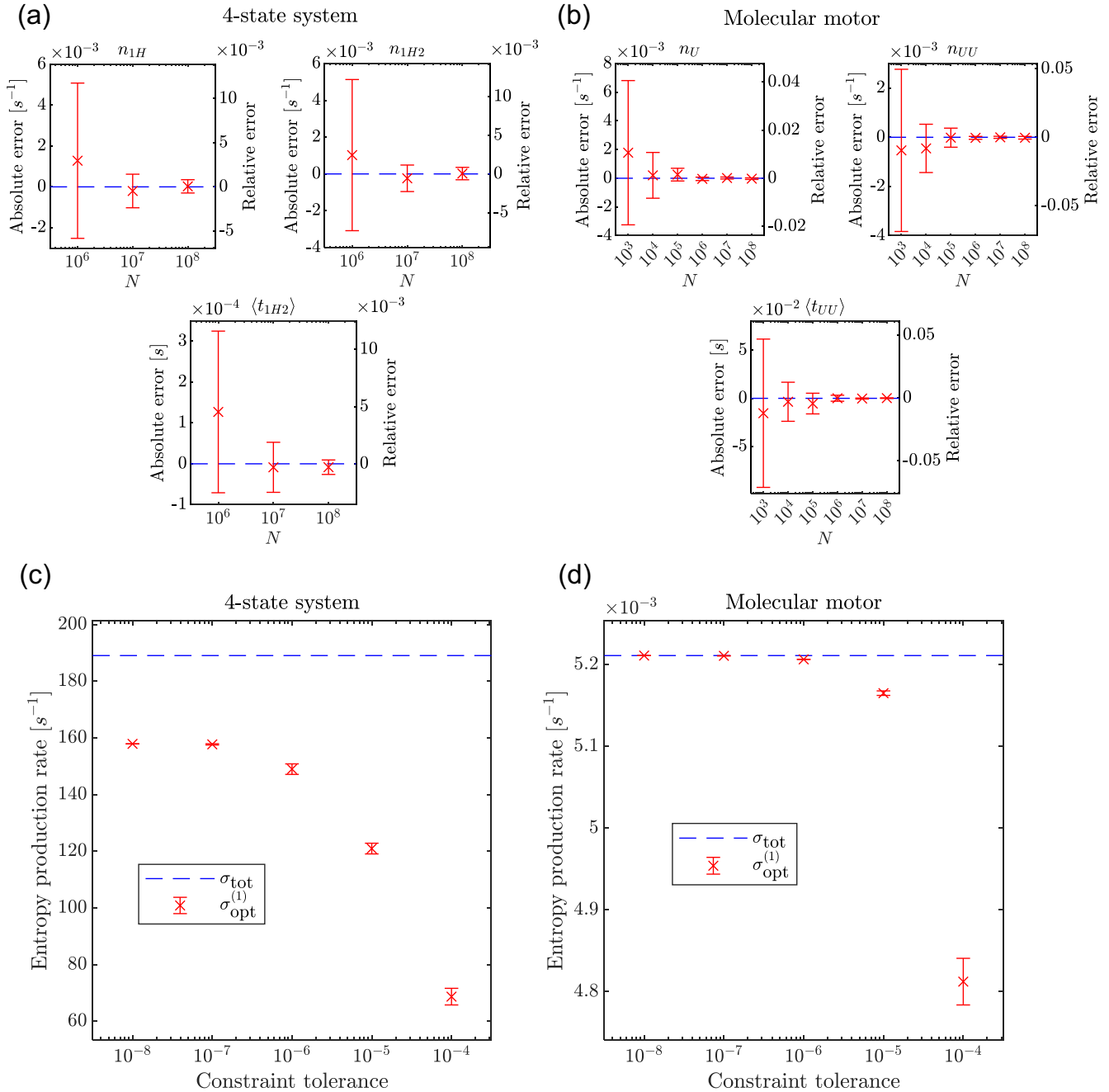


FIG. 4. Importance of data accuracy. (a) The error of some statistics of the four-state system for different values of the trajectory length  $N$ . The absolute and relative errors are on the left and right axes, respectively. (b) The error of some statistics of the molecular motor system for different values of the trajectory length  $N$ . The absolute and relative errors are on the left and right axes, respectively. (c) The error of  $\sigma_{\text{opt}}^{(1)}$  results for the four-state system for different constraint tolerance values, using the analytical statistics values. (d) The error of  $\sigma_{\text{opt}}^{(1)}$  results for the molecular motor system for different constraint tolerance values, using the analytical statistics values. Error bars stand for the standard deviation of 10 different realizations.

hidden part and the same observed statistics as the true system gave a lower bound on the total EPR for the two systems we studied, it remains an open problem to show this approach is universal. It would be interesting for future work to determine whether removing states from the hidden subnetwork can only decrease the entropy production, given the observed statistics are conserved.

In summary, our approach is based on an optimization problem formulated using the observed statistics of a partially accessible system, utilizing information on the underlying topology, in order to provide a tight lower bound on the total EPR. The estimator can be used as a benchmark for comparing the performance of other estimators that rely on coarse-grained or partial information about the system.



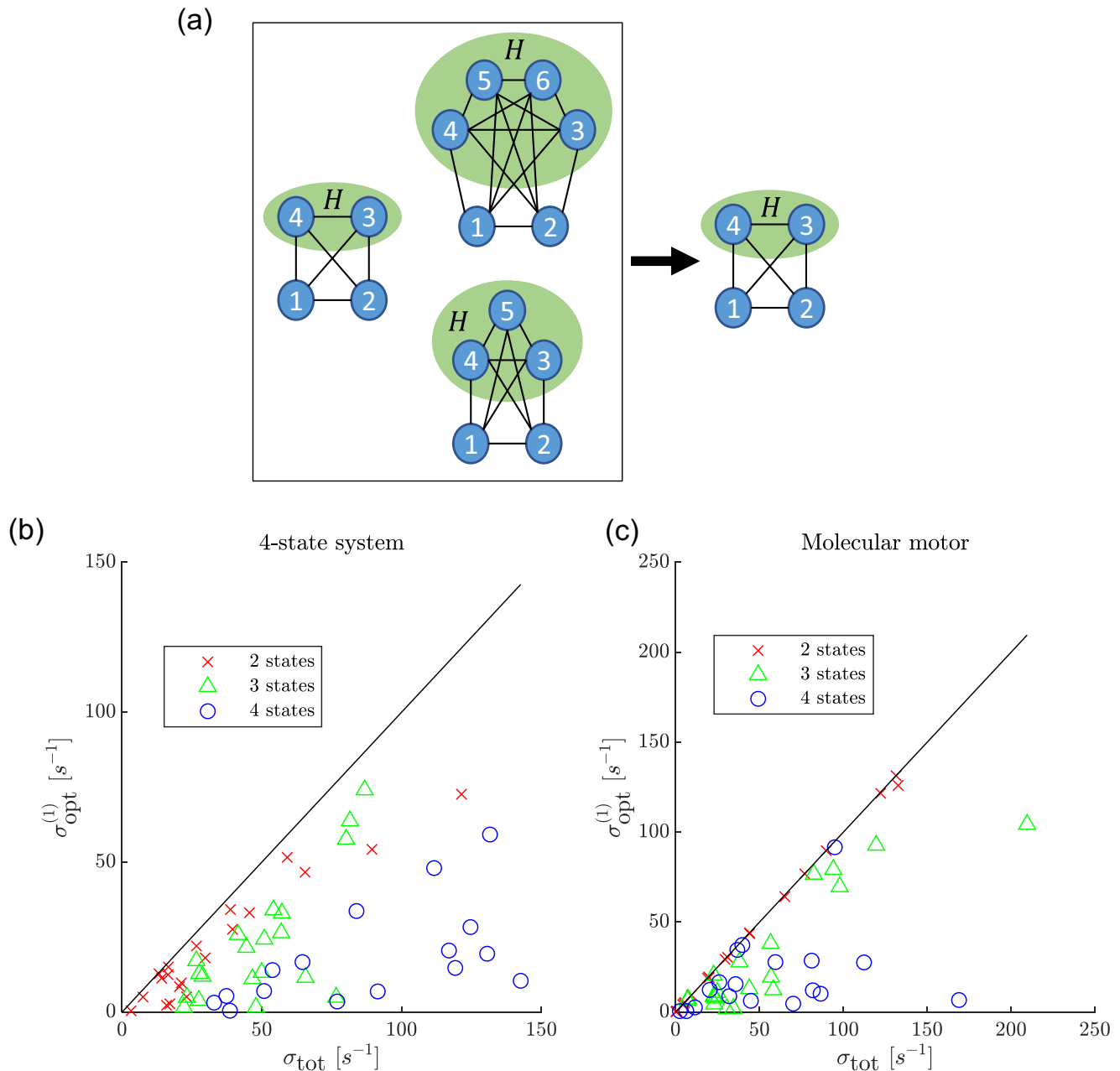


FIG. 5. Optimizing using a simple model. (a) Illustration of solving the optimization problem for a simple model with two hidden states (right), whereas the real system has more hidden states (left). (b) The results of  $\sigma_{\text{opt}}^{(1)}$  assuming the simple four-state model (two hidden states), when the real system has two (red cross), three (green triangle), or four (blue circle) hidden states. (c) The results of  $\sigma_{\text{opt}}^{(1)}$  assuming the simple molecular motor model (two hidden states), when the real system has two (red cross), three (green triangles), or four (blue circle) hidden states. For both systems, the results are presented for randomly generated transition rates (for each case) with statistics calculated from trajectories of length  $N = 10^8$  using a constraint tolerance of  $10^{-5}$ .

#### ACKNOWLEDGMENTS

G.B. acknowledges the Zuckerman STEM Leadership Program, and the Tel Aviv University Center for AI and Data Science (TAD). This work was supported by the ERC NanoNonEq 101039127, the Air Force Office of Scientific Research (AFOSR) under Award No. FA9550-20-1-0426, and by the Army Research Office (ARO) under Grant No. W911NF-21-1-0101. The views and conclusions contained in this document are those of the authors and should not be inter-

preted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government.

#### APPENDIX A: SECOND-ORDER MASS RATES

In order to find the second-order mass rates for two consecutive transitions between coarse-grained states  $n_{IJK}$ , we need to take into account every possible original state  $i \in I$ , every possible path within the coarse-grained state  $J$ , and every

possible transition from a state in  $J$  to every possible final state  $k \in K$ . Let us start by considering a specific initial Markovian state  $i \in I$  and a specific final Markovian state  $k \in K$  and calculate the mass rate  $n_{ijk}$ ,

$$\begin{aligned} n_{ijk} &= \sum_{N=0}^{\infty} \sum_{j_0, \dots, j_N \in J} n_{ij_0} P_{j_0 j_1} P_{j_1 j_2} \cdots P_{j_{N-1} j_N} P_{j_N k} \\ &= \sum_{N=0}^{\infty} \sum_{j', j'' \in J} n_{ij'} [P_{JJ}^N]_{j' j''} P_{j'' k} \\ &= \sum_{j', j'' \in J} n_{ij'} \left( \sum_{N=0}^{\infty} [P_{JJ}^N]_{j' j''} \right) P_{j'' k} \\ &= \sum_{j', j'' \in J} n_{ij'} [\mathbb{I} - P_{JJ}]_{j' j''}^{-1} P_{j'' k} \\ &= \mathbf{n}_{iJ}^T [\mathbb{I} - P_{JJ}]^{-1} \mathbf{p}_{Jk}. \end{aligned} \tag{A1}$$

The two summations are for all the possible lengths  $N$  of trajectories within  $J$ , and all the optional paths with the given length  $\{j_0, j_1, \dots, j_N\}$  in  $J$ . From mass conservation, we can now obtain the expression for  $n_{IJK}$  by summing over all the optional original  $i \in I$  and final  $k \in K$  states,

$$n_{IJK} = \sum_{i \in I} \sum_{k \in K} n_{ijk}. \tag{A2}$$

**APPENDIX B: CONDITIONAL WAITING TIME MOMENTS**

The waiting time at each Markovian state  $i$  is an exponentially distributed random variable  $\psi_i(t)$  with mean waiting time  $\tau_i = \lambda_i^{-1}$ ,

$$\psi_i(t) = \lambda_i e^{-\lambda_i t}. \tag{B1}$$

For the calculations, we used the joint distribution of the waiting time and the transition  $i \rightarrow j$ ,

$$\psi_{ij}(t) = w_{ij} e^{-\lambda_i t}. \tag{B2}$$

Notice that  $\psi_{ij}(t)$  is not normalized to 1 as  $\int_0^\infty \psi_{ij}(t) dt = p_{ij}$ .

The probability to observe a trajectory  $\gamma_N : i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_N$  with a total duration of  $T$  is

$$\begin{aligned} p(\gamma_N, T) &= \int_{\sum_{i=0}^{N-1} t_i = T} \psi_{i_0 i_1}(t_0) \psi_{i_1 i_2}(t_1) \cdots \psi_{i_{N-1} i_N}(t_{N-1}) \\ &\quad \times dt_0 dt_1 \cdots dt_{N-1}. \end{aligned} \tag{B3}$$

Since this is a convolution, we can perform a Laplace transform to get a simpler formula of multiplications of Laplace transforms of Markovian joint distributions of waiting times and transitions,

$$\tilde{p}(\gamma_N, s) = \tilde{\psi}_{i_0 i_1}(s) \tilde{\psi}_{i_1 i_2}(s) \cdots \tilde{\psi}_{i_{N-1} i_N}(s), \tag{B4}$$

where

$$\begin{aligned} \tilde{\psi}_{ij}(s) &= \int_0^\infty \psi_{ij}(t) e^{-st} dt = \int_0^\infty w_{ij} e^{-\lambda_i t} e^{-st} dt \\ &= w_{ij} \int_0^\infty e^{-(s+\lambda_i)t} dt = w_{ij} \left[ -\frac{e^{-(s+\lambda_i)t}}{s+\lambda_i} \right]_0^\infty = \frac{w_{ij}}{s+\lambda_i}. \end{aligned} \tag{B5}$$

In order to calculate the moments of the conditional waiting time distribution  $\psi_{IJK}(t)$  for the coarse-grained state  $J$  conditioned on an initial state in  $I$  and a final state in  $K$ , our strategy is to calculate its Laplace transform  $\tilde{\psi}_{IJK}(s)$ . We start by calculating  $\tilde{\psi}_{ijk}(s)$ , which is the Laplace transform of the waiting distribution in coarse-grained state  $J$ , before jumping to a specific Markovian state  $k \in K$ , given it came from a specific Markovian state  $i \in I$ . Since we want the waiting time in  $J$ , we sum over all of the paths with any length  $N$  inside  $J$  with a final transition to  $k \in K$ ,  $j_0 \rightarrow j_1 \rightarrow \dots \rightarrow j_N \rightarrow k$ , weighed by the probability to jump from  $i \in I$  to the first state  $j_0 \in J$ ,

$$\begin{aligned} \tilde{\psi}_{ijk}(s) &= \sum_{N=0}^{\infty} \sum_{j_0, \dots, j_N \in J} \frac{P_{ij_0}}{\sum_{j \in J} P_{ij}} \tilde{p}(j_0 \rightarrow j_1 \rightarrow \dots \rightarrow j_N \rightarrow k, s) \\ &= \sum_{N=0}^{\infty} \sum_{j_0, \dots, j_N \in J} \frac{P_{ij_0}}{\sum_{j \in J} P_{ij}} \tilde{\psi}_{j_0 j_1}(s) \cdots \tilde{\psi}_{j_{N-1} j_N}(s) \tilde{\psi}_{j_N k}(s) \\ &= \sum_{N=0}^{\infty} \sum_{j', j'' \in J} \frac{P_{ij'}}{\sum_{j \in J} P_{ij}} [\tilde{\Psi}_{JJ}(s)^N]_{j' j''} \tilde{\psi}_{j'' k}(s) \\ &= \sum_{j', j'' \in J} \frac{P_{ij'}}{\sum_{j \in J} P_{ij}} \sum_{N=0}^{\infty} [\tilde{\Psi}_{JJ}(s)^N]_{j' j''} \tilde{\psi}_{j'' k}(s) \\ &= \sum_{j', j'' \in J} \frac{P_{ij'}}{\sum_{j \in J} P_{ij}} [\mathbb{I} - \tilde{\Psi}_{JJ}(s)]_{j' j''}^{-1} \tilde{\psi}_{j'' k}(s) \\ &= \frac{\mathbf{p}_{iJ}^T}{\sum_{j \in J} P_{ij}} [\mathbb{I} - \tilde{\Psi}_{JJ}(s)]^{-1} \tilde{\mathbf{\psi}}_{Jk}(s), \end{aligned} \tag{B6}$$

where  $\tilde{\Psi}_{JJ}(s)$  is a matrix of size  $N_J \times N_J$ , and  $N_J$  is the number of Markovian states inside  $J$ ,

$$[\tilde{\Psi}_{JJ}(s)]_{j_1, j_2} = \begin{cases} \tilde{\psi}_{j_1 j_2}(s) & j_1 \neq j_2 \\ 0 & j_1 = j_2 \end{cases}. \tag{B7}$$

As mentioned in the main text we denote  $\tilde{\psi}_{iJK}(s) \equiv \sum_{k \in K} \tilde{\psi}_{ijk}(s)$ . Notice that  $\tilde{\psi}_{iJK}(s)$  is not normalized to 1 and it needs to be divided by  $\tilde{\psi}_{iJK}(s \rightarrow 0)$ , which is exactly the probability to jump from  $J$  to  $K$ , given the transition to  $J$  was from  $i$ ,

$$\tilde{\psi}_{iJK}^{\text{Normalized}}(s) = \frac{\tilde{\psi}_{iJK}(s)}{\tilde{\psi}_{iJK}(s \rightarrow 0)}. \tag{B8}$$

This results from the fact that we used  $\psi_{ij}(t)$ , which is normalized to  $p_{ij}$ .

In order to get  $\tilde{\psi}_{IJK}(s)$ , we sum  $\tilde{\psi}_{iJK}^{\text{Normalized}}(s)$  over all of the Markovian states  $i \in I$ , weighed by the corresponding probability  $\pi_i/\pi_I$  of being in state  $i$ , given the system is in the coarse-grained state  $I$ ,

$$\tilde{\psi}_{IJK}(s) = \sum_{i \in I} \frac{\pi_i}{\pi_I} \tilde{\psi}_{iJK}^{\text{Normalized}}(s). \tag{B9}$$

For a general probability density function  $f(t) : [0, \infty] \rightarrow [0, 1]$  the Laplace transform is

$$\tilde{f}(s) = \int_0^\infty f(t) e^{-st} dt \tag{B10}$$

and its  $k$ th derivative by  $s$  is

$$\frac{d^k \tilde{f}(s)}{ds^k} = (-1)^k \int_0^\infty t^k f(t) e^{-st} dt. \quad (\text{B11})$$

Taking the limit  $s \rightarrow 0$ ,

$$\frac{d^k \tilde{f}(s)}{ds^k} \Big|_{s \rightarrow 0} = (-1)^k \int_0^\infty t^k f(t) dt = (-1)^k \langle t^k \rangle, \quad (\text{B12})$$

we find the  $k$ th moment of the probability density function  $f(t)$ ,

$$\langle t^k \rangle = (-1)^k \frac{d^k \tilde{f}(s)}{ds^k} \Big|_{s \rightarrow 0}. \quad (\text{B13})$$

Therefore, the  $k$ th moment  $\langle t_{IJK}^k \rangle$  of the conditional waiting time distribution  $\psi_{IJK}(t)$  is

$$\langle t_{IJK}^k \rangle = (-1)^k \frac{d^k \tilde{\psi}_{IJK}(s)}{ds^k} \Big|_{s \rightarrow 0}. \quad (\text{B14})$$

### APPENDIX C: ANALYTICAL EXPRESSIONS FOR THE FOUR-STATE SYSTEM

The variables to consider for this system are the mass rates  $n_{ij}$  and the steady-state probabilities  $\pi_i$  for  $i, j \in \{1, 2, 3, 4\}$ , meaning a total of 16 variables. Note that  $\pi_1, \pi_2, n_{12},$  and  $n_{21}$  are fully observed. Therefore, we are left with 12 variables. With the following linear constraints, we can immediately reduce the problem to six variables.

#### 1. Linear constraints

We impose probability conservation, mass rate conservation in the hidden Markovian states, and mass rate conservation between an observed Markovian state and the hidden coarse-grained state.

##### a. Probabilities

From the conservation of the steady-state probability of the Markovian states within the coarse-grained hidden state,

$$\pi_H = \pi_3 + \pi_4. \quad (\text{C1})$$

##### b. Mass conservation at any Markovian state

We write the mass conservation for one of the hidden states (3 or 4), which for this system, is enough to guarantee the mass conservation for the other hidden state,

$$n_{13} + n_{23} + n_{43} = n_{31} + n_{32} + n_{34}. \quad (\text{C2})$$

##### c. First-order mass rates

Here, we require the mass rate conservation of transitions in and out of the hidden state, providing four constraint equations,

$$\begin{aligned} \forall_{i \in \{1,2\}} : n_{iH} &= n_{i3} + n_{i4}, \\ \forall_{i \in \{1,2\}} : n_{Hi} &= n_{3i} + n_{4i}. \end{aligned} \quad (\text{C3})$$

#### 2. Nonlinear constraints

The second-order mass rates and the conditional waiting times moments can be expressed only as a nonlinear function

of the optimization problem variables. Here, we show the full derivations of these relations.

##### a. Second-order mass rates

For this system, as mentioned in the text, we are interested in  $n_{1H2}$  and  $n_{2H1}$ , where the first and the last states are the observed Markovian states. From equation Eq. (A1)

$$n_{iHj} = \mathbf{n}_{iH}^T [\mathbb{I} - \mathbf{P}_{HH}]^{-1} \mathbf{P}_{Hj} \quad (\text{C4})$$

where

$$\mathbf{P}_{HH} = \begin{bmatrix} 0 & p_{34} \\ p_{43} & 0 \end{bmatrix} \quad (\text{C5})$$

and

$$[\mathbb{I} - \mathbf{P}_{HH}]^{-1} = \frac{1}{1 - p_{34}p_{43}} \begin{bmatrix} 1 & p_{34} \\ p_{43} & 1 \end{bmatrix}. \quad (\text{C6})$$

Plugging into Eq. (C4), we have

$$\begin{aligned} n_{iHj} &= \mathbf{n}_{iH}^T [\mathbb{I} - \mathbf{P}_{HH}]^{-1} \mathbf{P}_{Hj} \\ &= [n_{i3} \ n_{i4}] \left( \frac{1}{1 - p_{34}p_{43}} \begin{bmatrix} 1 & p_{34} \\ p_{43} & 1 \end{bmatrix} \right) \begin{bmatrix} p_{3j} \\ p_{4j} \end{bmatrix} \\ &= \frac{1}{1 - p_{34}p_{43}} [n_{i3} \ n_{i4}] \begin{bmatrix} p_{3j} + p_{34}p_{4j} \\ p_{4j} + p_{43}p_{3j} \end{bmatrix} \\ &= \frac{n_{i3}(p_{3j} + p_{34}p_{4j}) + n_{i4}(p_{4j} + p_{43}p_{3j})}{1 - p_{34}p_{43}}. \end{aligned} \quad (\text{C7})$$

Remember we can express  $p_{ij}$  in terms of the mass rates [Eq. (3)].

##### b. Conditional waiting time moments

We calculate the conditional waiting times moments  $\langle t_{iHj}^k \rangle$  for  $i, j \in \{1, 2\}$ , in terms of the problem variables. Based on Eq. (B14), we need to calculate  $\tilde{\psi}_{iHj}^{\text{Normalized}}(s)$ .

From Eq. (B6)

$$\tilde{\psi}_{iHj}(s) = \frac{\mathbf{P}_{iH}^T}{\sum_{h \in \{3,4\}} p_{ih}} [\mathbb{I} - \tilde{\Psi}_{HH}(s)]^{-1} \tilde{\Psi}_{Hj}(s). \quad (\text{C8})$$

Now, we can calculate  $\tilde{\Psi}_{Hj}(s)$  from Eqs. (12) and (B5),

$$\tilde{\Psi}_{Hj}(s) = \begin{bmatrix} \tilde{\psi}_{3j}(s) \\ \tilde{\psi}_{4j}(s) \end{bmatrix} = \begin{bmatrix} \frac{w_{3j}}{s + \lambda_3} \\ \frac{w_{4j}}{s + \lambda_4} \end{bmatrix}. \quad (\text{C9})$$

Given that [Eqs. (B7) and (B5)]

$$\begin{aligned} \tilde{\Psi}_{HH}(s) &= \begin{bmatrix} 0 & \tilde{\psi}_{34}(s) \\ \tilde{\psi}_{43}(s) & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \frac{w_{34}}{s + \lambda_3} \\ \frac{w_{43}}{s + \lambda_4} & 0 \end{bmatrix}. \end{aligned} \quad (\text{C10})$$

We can plug into Eq. (C8)

$$\begin{aligned}
 \tilde{\psi}_{iHj}(s) &= \frac{P_{iH}^T}{\sum_{h \in \{3,4\}} P_{ih}} [\mathbb{I} - \tilde{\Psi}_{HH}(s)]^{-1} \tilde{\psi}_{Hj}(s) \\
 &= \frac{1}{p_{i3} + p_{i4}} [p_{i3} \ p_{i4}] \left( \left( 1 - \frac{w_{34}w_{43}}{(s + \lambda_3)(s + \lambda_4)} \right)^{-1} \begin{bmatrix} 1 & \frac{w_{34}}{s + \lambda_3} \\ \frac{w_{43}}{s + \lambda_4} & 1 \end{bmatrix} \right) \begin{bmatrix} \frac{w_{3j}}{s + \lambda_3} \\ \frac{w_{4j}}{s + \lambda_4} \end{bmatrix} \\
 &= \frac{1}{p_{i3} + p_{i4}} \left( 1 - \frac{w_{34}w_{43}}{(s + \lambda_3)(s + \lambda_4)} \right)^{-1} [p_{i3} \ p_{i4}] \begin{bmatrix} \frac{w_{3j}}{s + \lambda_3} + \frac{w_{34}}{s + \lambda_3} \frac{w_{4j}}{s + \lambda_4} \\ \frac{w_{4j}}{s + \lambda_4} + \frac{w_{43}}{s + \lambda_4} \frac{w_{3j}}{s + \lambda_3} \end{bmatrix} \\
 &= \left( 1 - \frac{w_{34}w_{43}}{(s + \lambda_3)(s + \lambda_4)} \right)^{-1} \left[ \frac{p_{i3}}{p_{i3} + p_{i4}} \left( \frac{w_{3j}}{s + \lambda_3} + \frac{w_{34}}{s + \lambda_3} \frac{w_{4j}}{s + \lambda_4} \right) + \frac{p_{i4}}{p_{i3} + p_{i4}} \left( \frac{w_{4j}}{s + \lambda_4} + \frac{w_{43}}{s + \lambda_4} \frac{w_{3j}}{s + \lambda_3} \right) \right]. \tag{C11}
 \end{aligned}$$

Since the states  $i$  and  $j$  are Markovian, we just need to normalize this expression in order to get the desired result,

$$\begin{aligned}
 \tilde{\psi}_{iHj}(s \rightarrow 0) &= \left( 1 - \frac{w_{34}w_{43}}{\lambda_3\lambda_4} \right)^{-1} \left[ \frac{p_{i3}}{p_{i3} + p_{i4}} \left( \frac{w_{3j}}{\lambda_3} + \frac{w_{34}}{\lambda_3} \frac{w_{4j}}{\lambda_4} \right) + \frac{p_{i4}}{p_{i3} + p_{i4}} \left( \frac{w_{4j}}{\lambda_4} + \frac{w_{43}}{\lambda_4} \frac{w_{3j}}{\lambda_3} \right) \right] \\
 &= \frac{p_{i3}(p_{3j} + p_{34}p_{4j}) + p_{i4}(p_{4j} + p_{43}p_{3j})}{(p_{i3} + p_{i4})(1 - p_{34}p_{43})}. \tag{C12}
 \end{aligned}$$

Therefore

$$\tilde{\psi}_{iHj}^{\text{Normalized}}(s) = \frac{\tilde{\psi}_{iHj}(s)}{\tilde{\psi}_{iHj}(s \rightarrow 0)}. \tag{C13}$$

Finally, we get the moments from Eq. (B14).

In order to get the expressions of the derivatives, we used the package Sympy in Python.

#### APPENDIX D: ANALYTICAL EXPRESSIONS FOR THE MOLECULAR MOTOR SYSTEM

The variables to consider for the molecular motor system are the mass rates  $n_{u1}, n_{u2}, n_{d1}, n_{d2}, n_l, n_r$  and the steady-state probabilities  $\pi$  and  $\pi'$ , meaning a total of eight variables. With the following linear constraints, we can immediately reduce the problem to four variables.

##### 1. Linear constraints

As in the four-state system, we impose probability conservation, mass rate conservation in the Markovian states, and mass rate conservation for the observed transitions  $U$  and  $D$ .

###### a. Probabilities

From the conservation of the steady-state probability of the Markovian states within the coarse-grained states,

$$\pi + \pi' = \frac{1}{3}. \tag{D1}$$

###### b. Mass conservation at any Markovian state

We write the mass conservation for one of the hidden states (active or passive), which for this system, is enough to guarantee the mass conservation for the other hidden state,

$$n_r + n_{u2} = n_l + n_{d2}. \tag{D2}$$

##### c. First-order mass rates

Here, we require the mass rate conservation of transitions in and out of the coarse-grained state, providing two constraint equations,

$$n_U = n_{u1} + n_{u2}, \quad n_D = n_{d1} + n_{d2}. \tag{D3}$$

#### 2. Nonlinear constraints

Since we have two hidden states as in the four-state system, the results from Appendix C can be used here.

##### a. Second-order mass rates

We use the results for the four-state system in Eq. (C7), together with Eq. (A2). For  $n_{UU}$ , we need to sum over all the mass that goes up from the passive or active state, and then up again only to the passive state,

$$\begin{aligned}
 n_{UU} &= \frac{n_{u1}(p_{u1} + p_l p_{u2})}{1 - p_l p_r} + \frac{n_{u2}(p_{u1} + p_l p_{u2})}{1 - p_l p_r} \\
 &= \frac{(n_{u1} + n_{u2})(p_{u1} + p_l p_{u2})}{1 - p_l p_r}. \tag{D4}
 \end{aligned}$$

For  $n_{DD}$ , we need to sum over all the mass that goes down only from the passive state, and then down again to the passive or active state,

$$\begin{aligned}
 n_{DD} &= \frac{n_{d1}p_{d1} + n_{d2}p_r p_{d1}}{1 - p_l p_r} + \frac{n_{d1}p_{d2} + n_{d2}p_r p_{d2}}{1 - p_l p_r} \\
 &= \frac{(n_{d1} + n_{d2}p_r)(p_{d1} + p_{d2})}{1 - p_l p_r}. \tag{D5}
 \end{aligned}$$

##### b. Conditional waiting time moments

We account for all of the transitions through a coarse-grained state  $i$ , and specify in the following calculations the Markovian state before jumping to  $i$ , and the following Markovian state, after state  $i$ , where  $i'$  ( $i$ ) denoted an active

(passive) state. For example,  $(i - 1) \rightarrow (i + 1)$  represent two consecutive transitions,  $(i - 1) \rightarrow i \rightarrow (i + 1)$ .

Note that a transition upward is only to a passive state, so the previous state (being passive or active) in

the first transition does not affect the waiting time. Furthermore, a transition downward is only from a passive state.

From Eq. (B9),

$$\tilde{\psi}_{UU}(s) = \frac{\pi}{\pi + \pi'} \frac{\tilde{\psi}_{(i-1) \rightarrow (i+1)}(s)}{\tilde{\psi}_{(i-1) \rightarrow (i+1)}(s \rightarrow 0)} + \frac{\pi'}{\pi + \pi'} \frac{\tilde{\psi}_{(i-1)' \rightarrow (i+1)}(s)}{\tilde{\psi}_{(i-1)' \rightarrow (i+1)}(s \rightarrow 0)} = \frac{\tilde{\psi}_{(i-1) \rightarrow (i+1)}(s)}{\tilde{\psi}_{(i-1) \rightarrow (i+1)}(s \rightarrow 0)} \quad (\text{D6a})$$

and similarly

$$\begin{aligned} \tilde{\psi}_{UD}(s) &= \frac{\pi}{\pi + \pi'} \frac{(\tilde{\psi}_{(i-1) \rightarrow (i-1)} + \tilde{\psi}_{(i-1) \rightarrow (i-1)'}) (s)}{(\tilde{\psi}_{(i-1) \rightarrow (i-1)} + \tilde{\psi}_{(i-1) \rightarrow (i-1)'}) (s \rightarrow 0)} + \frac{\pi'}{\pi + \pi'} \frac{(\tilde{\psi}_{(i-1)' \rightarrow (i-1)} + \tilde{\psi}_{(i-1)' \rightarrow (i-1)'}) (s)}{(\tilde{\psi}_{(i-1)' \rightarrow (i-1)} + \tilde{\psi}_{(i-1)' \rightarrow (i-1)'}) (s \rightarrow 0)} \\ &= \frac{(\tilde{\psi}_{(i-1) \rightarrow (i-1)} + \tilde{\psi}_{(i-1) \rightarrow (i-1)'}) (s)}{(\tilde{\psi}_{(i-1) \rightarrow (i-1)} + \tilde{\psi}_{(i-1) \rightarrow (i-1)'}) (s \rightarrow 0)}. \end{aligned} \quad (\text{D6b})$$

Moreover

$$\tilde{\psi}_{DU}(s) = \frac{\tilde{\psi}_{(i+1) \rightarrow (i+1)}(s)}{\tilde{\psi}_{(i+1) \rightarrow (i+1)}(s \rightarrow 0)} \quad (\text{D6c})$$

and

$$\tilde{\psi}_{DD}(s) = \frac{(\tilde{\psi}_{(i+1) \rightarrow (i-1)} + \tilde{\psi}_{(i+1) \rightarrow (i-1)'}) (s)}{(\tilde{\psi}_{(i+1) \rightarrow (i-1)} + \tilde{\psi}_{(i+1) \rightarrow (i-1)'}) (s \rightarrow 0)}. \quad (\text{D6d})$$

Now we calculate all the terms in the numerators, using Eq. (C11) from the four-state system results,

$$\tilde{\psi}_{(i-1) \rightarrow (i+1)}(s) = \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \left(\frac{w_{u1}}{s + \lambda} + \frac{w_l}{s + \lambda} \frac{w_{u2}}{s + \lambda'}\right), \quad (\text{D7a})$$

$$\begin{aligned} (\tilde{\psi}_{(i-1) \rightarrow (i-1)} + \tilde{\psi}_{(i-1) \rightarrow (i-1)'}) (s) &= \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \left(\frac{w_{d1}}{s + \lambda} + \frac{w_{d2}}{s + \lambda}\right) \\ &= \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \frac{w_{d1} + w_{d2}}{s + \lambda}, \end{aligned} \quad (\text{D7b})$$

$$\begin{aligned} \tilde{\psi}_{(i+1) \rightarrow (i+1)}(s) &= \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \left[\frac{p_{d1}}{p_{d1} + p_{d2}} \frac{w_{u1}}{s + \lambda} + \frac{p_{d2}}{p_{d1} + p_{d2}} \frac{w_r}{s + \lambda'} \frac{w_{u1}}{s + \lambda}\right] \\ &= \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \frac{1}{p_{d1} + p_{d2}} \frac{w_{u1}}{s + \lambda} \left[p_{d1} + \frac{p_{d2} w_r}{s + \lambda'}\right], \end{aligned} \quad (\text{D7c})$$

$$\begin{aligned} (\tilde{\psi}_{(i+1) \rightarrow (i-1)} + \tilde{\psi}_{(i+1) \rightarrow (i-1)'}) (s) &= \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \left[\frac{p_{d1}}{p_{d1} + p_{d2}} \frac{w_{d1}}{s + \lambda} + \frac{p_{d2}}{p_{d1} + p_{d2}} \frac{w_r}{s + \lambda'} \frac{w_{d1}}{s + \lambda}\right. \\ &\quad \left.+ \frac{p_{d1}}{p_{d1} + p_{d2}} \frac{w_{d2}}{s + \lambda} + \frac{p_{d2}}{p_{d1} + p_{d2}} \frac{w_r}{s + \lambda'} \frac{w_{d2}}{s + \lambda}\right] \\ &= \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \left[\frac{p_{d1}}{p_{d1} + p_{d2}} \frac{w_{d1} + w_{d2}}{s + \lambda} + \frac{p_{d2}}{p_{d1} + p_{d2}} \frac{w_r}{s + \lambda'} \frac{w_{d1} + w_{d2}}{s + \lambda}\right] \\ &= \left(1 - \frac{w_l w_r}{(s + \lambda)(s + \lambda')}\right)^{-1} \frac{1}{p_{d1} + p_{d2}} \frac{w_{d1} + w_{d2}}{s + \lambda} \left[p_{d1} + \frac{p_{d2} w_r}{s + \lambda'}\right]. \end{aligned} \quad (\text{D7d})$$

All of the denominators from Eq. (D6) can be calculated by setting  $s \rightarrow 0$  in Eq. (D7). Finally, we get the moments from Eq. (B14).

In order to get the expressions of the derivatives, we used the package Sympy in Python.

## APPENDIX E: LARGER SYSTEMS

We apply our method to a system with five states, three of which are Markovian, and the other two are coarse grained to a single state  $H$  [Fig. 6(a)], for different values of an external

force  $F$  used to tune the transition rates over the observed link 1 - 2 according to  $w_{12}(F) = w_{12}e^{-\beta FL}$  and  $w_{21}(F) = w_{21}e^{\beta FL}$ . Comparing the results of our method to other bounds [Fig. 6(b)],  $\sigma_{\text{opt}}^{(1)}$  outperforms  $\sigma_{\text{KLD}}$  and  $\sigma_2$ , and trivially  $\sigma_{\text{aff}}$ .

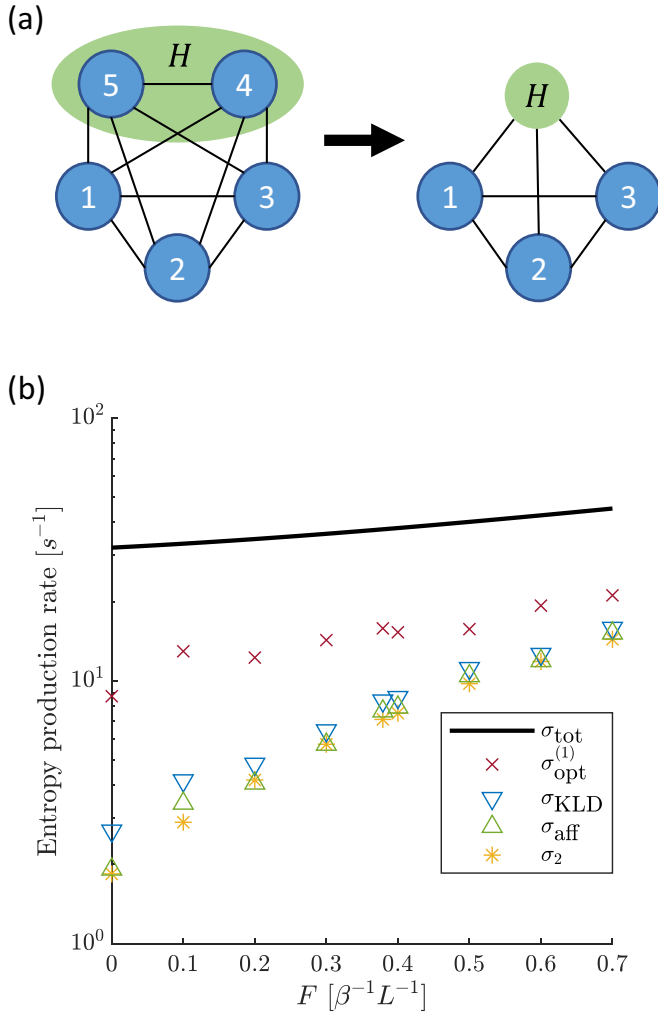


FIG. 6. Five-state system with four observed states. (a) Illustration of the full system topology, including the coarse graining of states 4 and 5 to state  $H$ . (b) Total EPR  $\sigma_{\text{tot}}$  (black line), our bound  $\sigma_{\text{opt}}^{(1)}$  (red cross), KLD estimator  $\sigma_{\text{KLD}}$  (blue downward-pointing triangle), affinity estimator  $\sigma_{\text{aff}}$  (green upward-pointing triangle), and two-step estimator  $\sigma_2$  (yellow asterisk). The rates used are  $w_{12} = 11 \text{ s}^{-1}$ ,  $w_{13} = 0 \text{ s}^{-1}$ ,  $w_{14} = 71 \text{ s}^{-1}$ ,  $w_{15} = 81 \text{ s}^{-1}$ ,  $w_{21} = 31 \text{ s}^{-1}$ ,  $w_{23} = 0 \text{ s}^{-1}$ ,  $w_{24} = 12 \text{ s}^{-1}$ ,  $w_{25} = 96 \text{ s}^{-1}$ ,  $w_{31} = 0 \text{ s}^{-1}$ ,  $w_{32} = 0 \text{ s}^{-1}$ ,  $w_{34} = 92 \text{ s}^{-1}$ ,  $w_{35} = 12 \text{ s}^{-1}$ ,  $w_{41} = 69 \text{ s}^{-1}$ ,  $w_{42} = 15 \text{ s}^{-1}$ ,  $w_{43} = 14 \text{ s}^{-1}$ ,  $w_{45} = 91 \text{ s}^{-1}$ ,  $w_{51} = 100 \text{ s}^{-1}$ ,  $w_{52} = 71 \text{ s}^{-1}$ ,  $w_{53} = 29 \text{ s}^{-1}$ ,  $w_{54} = 30 \text{ s}^{-1}$ .

#### APPENDIX F: COMPARING ESTIMATORS WHEN OPTIMIZING FOR A SIMPLE MODEL

We apply our method to a system with five states, two of which are Markovian, and the other three are coarse grained to a single state  $H$ . In order to solve for a case where one

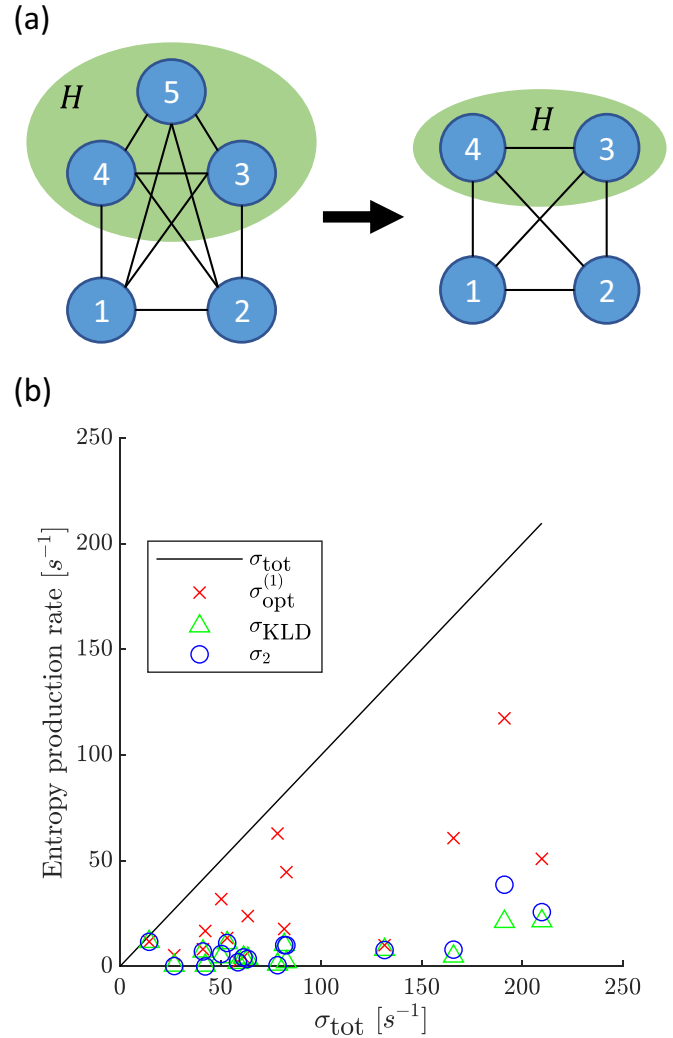


FIG. 7. Comparison to other bounds when assuming a simple topology. (a) Illustration of the full five-state system topology, including the coarse graining of states 3, 4, and 5 to state  $H$  (left) and the full-system topology we assume (right). (b) Total EPR  $\sigma_{\text{tot}}$  (solid black line), our bound  $\sigma_{\text{opt}}^{(1)}$  (red cross), KLD estimator  $\sigma_{\text{KLD}}$  (green upward-pointing triangle), and two-step estimator  $\sigma_2$  (blue circle). The results are presented for randomly generated transition rates with statistics calculated from trajectories of length  $N = 10^8$ . Values of the estimators with the same  $\sigma_{\text{tot}}$  correspond to the same system, showing that  $\sigma_{\text{opt}}^{(1)}$  outperforms both  $\sigma_{\text{KLD}}$  and  $\sigma_2$ .

does not have information about the underlying topology, we assume the simplest topology, with only two internal states [Fig. 7(a)], and write the optimization problem accordingly. Using our approach, not only do we get a lower bound on the total EPR, but also our estimator  $\sigma_{\text{opt}}^{(1)}$  outperforms the other estimators.

- [1] C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang, Optical tweezers in single-molecule biophysics, *Nat. Rev. Methods Primers* **1**, 25 (2021).
- [2] C. D. Kinz-Thompson, K. K. Ray, and R. L. Gonzalez Jr, Bayesian inference: The comprehensive approach to analyzing

single-molecule experiments, *Annu. Rev. Biophys.* **50**, 191 (2021).

- [3] C. Bustamante, L. Alexander, K. Maciuba, and C. M. Kaiser, Single-molecule studies of protein folding with optical tweezers, *Annu. Rev. Biochem.* **89**, 443 (2020).

- [4] C. Bustamante, Unfolding single RNA molecules: Bridging the gap between equilibrium and non-equilibrium statistical thermodynamics, *Q. Rev. Biophys.* **38**, 291 (2005).
- [5] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [6] C. Van den Broeck and M. Esposito, Ensemble and trajectory thermodynamics: A brief introduction, *Physica A* **418**, 6 (2015).
- [7] R. van Zon and E. G. D. Cohen, Stationary and transient work-fluctuation theorems for a dragged Brownian particle, *Phys. Rev. E* **67**, 046102 (2003).
- [8] F. Douarche, S. Joubaud, N. B. Garnier, A. Petrosyan, and S. Ciliberto, Work fluctuation theorems for harmonic oscillators, *Phys. Rev. Lett.* **97**, 140603 (2006).
- [9] S. Sabhapandit, Heat and work fluctuations for a harmonic oscillator, *Phys. Rev. E* **85**, 021108 (2012).
- [10] P. Visco, Work fluctuations for a Brownian particle between two thermostats, *J. Stat. Mech.* (2006) P06006.
- [11] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales, *Phys. Rev. Lett.* **89**, 050601 (2002).
- [12] S. Ciliberto, A. Imparato, A. Naert, and M. Tanase, Heat flux and entropy produced by thermal fluctuations, *Phys. Rev. Lett.* **110**, 180601 (2013).
- [13] I. A. Martínez, É. Roldán, L. Dinis, D. Petrov, J. M. Parrondo, and R. A. Rica, Brownian Carnot engine, *Nat. Phys.* **12**, 67 (2016).
- [14] C. Van den Broeck, N. Kumar, and K. Lindenberg, Efficiency of isothermal molecular machines at maximum power, *Phys. Rev. Lett.* **108**, 210602 (2012).
- [15] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck, The unlikely Carnot efficiency, *Nat. Commun.* **5**, 4721 (2014).
- [16] S. Mohanta, S. Saryal, and B. K. Agarwalla, Universal bounds on cooling power and cooling efficiency for autonomous absorption refrigerators, *Phys. Rev. E* **105**, 034127 (2022).
- [17] S. Bo, M. Del Giudice, and A. Celani, Thermodynamic limits to information harvesting by sensory systems, *J. Stat. Mech.* (2015) P01014.
- [18] N. P. Saadat, T. Nies, Y. Rousset, and O. Ebenhöf, Thermodynamic limits and optimality of microbial growth, *Entropy* **22**, 277 (2020).
- [19] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, Quantifying dissipation using fluctuating currents, *Nat. Commun.* **10**, 1666 (2019).
- [20] É. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, How far from equilibrium is active matter? *Phys. Rev. Lett.* **117**, 038103 (2016).
- [21] C. Maes and K. Netočný, Time-reversal and entropy, *J. Stat. Phys.* **110**, 269 (2003).
- [22] J. M. Parrondo, C. Van den Broeck, and R. Kawai, Entropy production and the arrow of time, *New J. Phys.* **11**, 073008 (2009).
- [23] P. Pietzonka, A. C. Barato, and U. Seifert, Universal bound on the efficiency of molecular motors, *J. Stat. Mech.* (2016) 124004.
- [24] N. Shiraishi, Optimal thermodynamic uncertainty relation in Markov jump processes, *J. Stat. Phys.* **185**, 19 (2021).
- [25] J. M. Horowitz and T. R. Gingrich, Thermodynamic uncertainty relations constrain non-equilibrium fluctuations, *Nat. Phys.* **16**, 15 (2020).
- [26] 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).
- [27] A. C. Barato and U. Seifert, Thermodynamic uncertainty relation for biomolecular processes, *Phys. Rev. Lett.* **114**, 158101 (2015).
- [28] S. K. Manikandan, S. Ghosh, A. Kundu, B. Das, V. Agrawal, D. Mitra, A. Banerjee, and S. Krishnamurthy, Quantitative analysis of non-equilibrium systems from short-time experimental data, *Commun. Phys.* **4**, 258 (2021).
- [29] S. K. Manikandan, D. Gupta, and S. Krishnamurthy, Inferring entropy production from short experiments, *Phys. Rev. Lett.* **124**, 120603 (2020).
- [30] T. R. Gingrich and J. M. Horowitz, Fundamental bounds on first passage time fluctuations for currents, *Phys. Rev. Lett.* **119**, 170601 (2017).
- [31] A. Pal, S. Reuveni, and S. Rahav, Thermodynamic uncertainty relation for first-passage times on Markov chains, *Phys. Rev. Res.* **3**, L032034 (2021).
- [32] I. Di Terlizzi and M. Baiesi, Kinetic uncertainty relation, *J. Phys. A: Math. Theor.* **52**, 02LT03 (2019).
- [33] Van Tuan Vo, T. Van Vu, and Y. Hasegawa, Unified thermodynamic–kinetic uncertainty relation, *J. Phys. A: Math. Theor.* **55**, 405004 (2022).
- [34] D. J. Skinner and J. Dunkel, Estimating entropy production from waiting time distributions, *Phys. Rev. Lett.* **127**, 198101 (2021).
- [35] A. Ghosal and G. Bisker, Inferring entropy production rate from partially observed Langevin dynamics under coarse-graining, *Phys. Chem. Chem. Phys.* **24**, 24021 (2022).
- [36] U. Kapustin, A. Ghosal, and G. Bisker, Utilizing time-series measurements for entropy production estimation in partially observed systems, [arXiv:2212.13487](https://arxiv.org/abs/2212.13487).
- [37] S. Otsubo, S. Ito, A. Dechant, and T. Sagawa, Estimating entropy production by machine learning of short-time fluctuating currents, *Phys. Rev. E* **101**, 062106 (2020).
- [38] D.-K. Kim, Y. Bae, S. Lee, and H. Jeong, Learning entropy production via neural networks, *Phys. Rev. Lett.* **125**, 140604 (2020).
- [39] Y. Bae, D.-K. Kim, and H. Jeong, Inferring dissipation maps from videos using convolutional neural networks, *Phys. Rev. Res.* **4**, 033094 (2022).
- [40] D.-K. Kim, S. Lee, and H. Jeong, Estimating entropy production with odd-parity state variables via machine learning, *Phys. Rev. Res.* **4**, 023051 (2022).
- [41] É. Roldán and J. M. R. Parrondo, Estimating dissipation from single stationary trajectories, *Phys. Rev. Lett.* **105**, 150607 (2010).
- [42] S. Otsubo, S. K. Manikandan, T. Sagawa, and S. Krishnamurthy, Estimating time-dependent entropy production from non-equilibrium trajectories, *Commun. Phys.* **5**, 11 (2022).
- [43] B. Lander, J. Mehl, V. Blickle, C. Bechinger, and U. Seifert, Noninvasive measurement of dissipation in colloidal systems, *Phys. Rev. E* **86**, 030401(R) (2012).
- [44] P. Padmanabha, D. M. Busiello, A. Maritan, and D. Gupta, Fluctuations of entropy production of a run-and-tumble particle, *Phys. Rev. E* **107**, 014129 (2023).
- [45] A. Dechant and S. I. Sasa, Entropic bounds on currents in Langevin systems, *Phys. Rev. E* **97**, 062101 (2018).

- [46] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Dissipation: The phase-space perspective, *Phys. Rev. Lett.* **98**, 080602 (2007).
- [47] G. Teza and A. L. Stella, Exact coarse graining preserves entropy production out of equilibrium, *Phys. Rev. Lett.* **125**, 110601 (2020).
- [48] G. Bisker, M. Polettini, T. R. Gingrich, and J. M. Horowitz, Hierarchical bounds on entropy production inferred from partial information, *J. Stat. Mech.* (2017) 093210.
- [49] M. Esposito, Stochastic thermodynamics under coarse graining, *Phys. Rev. E* **85**, 041125 (2012).
- [50] D. M. Busiello, J. Hidalgo, and A. Maritan, Entropy production for coarse-grained dynamics, *New J. Phys.* **21**, 073004 (2019).
- [51] A. Ghosal and G. Bisker, Entropy production rates for different notions of partial information, *J. Phys. D* **56**, 254001 (2023).
- [52] P. J. Foster, J. Bae, B. Lemma, J. Zheng, W. Ireland, P. Chandrakar, R. Boros, Z. Dogic, D. J. Needleman, and J. J. Vlassak, Dissipation and energy propagation across scales in an active cytoskeletal material, *Proc. Natl. Acad. Sci. USA* **120**, e2207662120 (2023).
- [53] S. Ro, B. Guo, A. Shih, T. V. Phan, R. H. Austin, D. Levine, P. M. Chaikin, and S. Martiniani, Model-free measurement of local entropy production and extractable work in active matter, *Phys. Rev. Lett.* **129**, 220601 (2022).
- [54] D. Lucente, A. Baldassarri, A. Puglisi, A. Vulpiani, and M. Viale, Inference of time irreversibility from incomplete information: Linear systems and its pitfalls, *Phys. Rev. Res.* **4**, 043103 (2022).
- [55] M. Baiesi and G. Falasco, Effective estimation of entropy production with lacking data, [arXiv:2305.04657](https://arxiv.org/abs/2305.04657).
- [56] N. Shiraishi, S. Ito, K. Kawaguchi, and T. Sagawa, Role of measurement-feedback separation in autonomous Maxwell's demons, *New J. Phys.* **17**, 045012 (2015).
- [57] N. Shiraishi and T. Sagawa, Fluctuation theorem for partially masked nonequilibrium dynamics, *Phys. Rev. E* **91**, 012130 (2015).
- [58] M. Polettini and M. Esposito, Effective thermodynamics for a marginal observer, *Phys. Rev. Lett.* **119**, 240601 (2017).
- [59] C. Maes, The fluctuation theorem as a Gibbs property, *J. Stat. Phys.* **95**, 367 (1999).
- [60] É. Roldán, J. Barral, P. Martin, J. M. Parrondo, and F. Jülicher, Quantifying entropy production in active fluctuations of the hair-cell bundle from time irreversibility and uncertainty relations, *New J. Phys.* **23**, 083013 (2021).
- [61] J. Horowitz and C. Jarzynski, Illustrative example of the relationship between dissipation and relative entropy, *Phys. Rev. E* **79**, 021106 (2009).
- [62] B. Gaveau, L. Granger, M. Moreau, and L. S. Schulman, Dissipation, interaction, and relative entropy, *Phys. Rev. E* **89**, 032107 (2014).
- [63] B. Gaveau, L. Granger, M. Moreau, and L. S. Schulman, Relative entropy, interaction energy and the nature of dissipation, *Entropy* **16**, 3173 (2014).
- [64] F. A. Cisneros, N. Fakhri, and J. M. Horowitz, Dissipative timescales from coarse-graining irreversibility, *J. Stat. Mech.* (2023) 073201.
- [65] I. A. Martínez, G. Bisker, J. M. Horowitz, and J. M. Parrondo, Inferring broken detailed balance in the absence of observable currents, *Nat. Commun.* **10**, 3542 (2019).
- [66] J. van der Meer, B. Ertel, and U. Seifert, Thermodynamic inference in partially accessible Markov networks: A unifying perspective from transition-based waiting time distributions, *Phys. Rev. X* **12**, 031025 (2022).
- [67] G. Bisker, I. A. Martinez, J. M. Horowitz, and J. M. Parrondo, Comment on "Inferring broken detailed balance in the absence of observable currents", [arXiv:2202.02064](https://arxiv.org/abs/2202.02064).
- [68] J. van der Meer, J. Degünther, and U. Seifert, Time-resolved statistics of snippets as general framework for model-free entropy estimators, *Phys. Rev. Lett.* **130**, 257101 (2023).
- [69] J. Ehrlich, Tightest bound on hidden entropy production from partially observed dynamics, *J. Stat. Mech.* (2021) 083214.
- [70] D. J. Skinner and J. Dunkel, Improved bounds on entropy production in living systems, *Proc. Natl. Acad. Sci. USA* **118**, e2024300118 (2021).
- [71] J. Schnakenberg, Network theory of microscopic and macroscopic behavior of master equation systems, *Rev. Mod. Phys.* **48**, 571 (1976).
- [72] C. Maes, K. Netočný, and B. Wynants, Dynamical fluctuations for semi-Markov processes, *J. Phys. A: Math. Theor.* **42**, 365002 (2009).
- [73] J. Zhang and T. Zhou, Markovian approaches to modeling intracellular reaction processes with molecular memory, *Proc. Natl. Acad. Sci. USA* **116**, 23542 (2019).
- [74] D. Hartich and A. Godec, Violation of local detailed balance despite a clear time-scale separation, *Phys. Rev. Res.* **5**, L032017 (2023).
- [75] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí, Scatter search and local NLP solvers: A multistart framework for global optimization, *INFORMS J. Comput.* **19**, 328 (2007).
- [76] J. A. Nirody, A. L. Nord, and R. M. Berry, Load-dependent adaptation near zero load in the bacterial flagellar motor, *J. R. Soc. Interface* **16**, 20190300 (2019).
- [77] T. H. Tan, G. A. Watson, Y.-C. Chao, J. Li, T. R. Gingrich, J. M. Horowitz, and N. Fakhri, Scale-dependent irreversibility in living matter, [arXiv:2107.05701](https://arxiv.org/abs/2107.05701).