PNAS RESEARCH ARTICLE | PHYSICS

Milestoning estimators of dissipation in systems observed at a coarse resolution

Kristian Blom[a], Kevin Song[b], Etienne Vouga[b], Aljaž Godec[a,1], and Dmitrii E. Makarov[c,1]

Edited by Anatoly B. Kolomeisky, Rice University, Houston, TX; received October 20, 2023; accepted March 14, 2024 by Editorial Board Member Peter J. Rossky

Many nonequilibrium, active processes are observed at a coarse-grained level, where different microscopic configurations are projected onto the same observable state. Such "lumped" observables display memory, and in many cases, the irreversible character of the underlying microscopic dynamics becomes blurred, e.g., when the projection hides dissipative cycles. As a result, the observations appear less irreversible, and it is very challenging to infer the degree of broken time-reversal symmetry. Here we show, contrary to intuition, that by ignoring parts of the already coarse-grained state space we may—via a process called milestoning—improve entropy-production estimates. We present diverse examples where milestoning systematically renders observations "closer to underlying microscopic dynamics" and thereby improves thermodynamic inference from lumped data assuming a given range of memory, and we hypothesize that this effect is quite general. Moreover, whereas the correct general physical definition of time reversal in the presence of memory remains unknown, we here show by means of physically relevant examples that at least for semi-Markov processes of first and second order, waiting-time contributions arising from adopting a naive Markovian definition of time reversal generally must be discarded.

entropy production | milestoning | semi-Markov processes | waiting-time distribution

Significance

Molecular machines operate far from equilibrium, dissipating energy and increasing the entropy of their surroundings. The irreversibility of their motion, even with the unprecedented spatiotemporal resolution of recent single-molecule studies, is blurred by thermal fluctuations and the fact that experiments typically observe only a coarse projection of microscopic dynamics. This makes the inference of broken time-reversal symmetry a daunting task. We describe, by means of diverse examples, how to achieve a drastic improvement in estimating dissipation from observed coarse trajectories. Our results are important, inter alia, for quantifying whether observables are affected by dissipative, ATP-hydrolysis-driven processes in the cell even if they do not display directional motion in an obvious way.

Although the nonequilibrium character of active systems may often be self-evident, e.g., the directed motion of a kinesin motor or persistent rotation of the F_1-ATPase (1), differentiating in general between equilibrium and nonequilibrium steady states at mesoscopic and microscopic scales is surprisingly difficult (2–6). A key reason is that broken time-reversal symmetry often emerges on length scales where detailed microscopic information is unavailable, e.g., below the diffraction limit. In particular, single-molecule experiments probe inherently coarse-grained representations of microscopic dynamics (7) and may thus underestimate (2) or even "be blind to" relevant slow dissipative degrees of freedom (8).

Coarse-graining may be a consequence of experimental limitations as well as of data analysis: For example, the location of a microscopic probe is, most optimistically, limited by the pixel size, which introduces spatial coarse-graining (9). Similarly, when protein folding is probed using optical tweezers (10) or fluorescence resonance energy transfer (2, 11, 12), many molecular conformations with, respectively, the same extension or with the same distance between fluorescent probes are lumped onto a single state. This projects the high-dimensional conformational dynamics onto a single lumped coordinate.

Consider the total steady-state entropy-production rate (i.e., dissipation rate), estimated as (13, 14)

$$\langle \dot{S}[x] \rangle \equiv \lim_{t\to\infty} \frac{1}{t} \left\langle \ln \frac{P[\{x(\tau)\}_{0\le\tau\le t}]}{P[\{\theta x(\tau)\}_{0\le\tau\le t}]} \right\rangle, \qquad [1]$$

where $x(\tau)$ denotes the system's state at time τ, $P[\{x(\tau)\}_{0\le\tau\le t}]$ is the probability of a forward path $\{x(\tau)\}_{0\le\tau\le t}$, and θ denotes the physically consistent time-reversal operation such that $P[\{\theta x(\tau)\}_{0\le\tau\le t}]$ corresponds to the probability of the time-reversed path. In Eq. **1**, Boltzmann's constant was set to 1, $\langle\cdot\rangle$ indicates averaging over $P[\{x(\tau)\}_{0\le\tau\le t}]$, and we tacitly assumed ergodic dynamics. When $x(\tau)$ is a Markov process, $\{\theta x(\tau)\}_{0\le\tau\le t} = \{\epsilon x(t-\tau)\}_{0\le\tau\le t}$ where ϵ denotes a sign change of all degrees of freedom that are odd under time reversal (such as momenta).

Note that in the presence of memory, the physically consistent form of time reversal θ is generally not known. In particular, naively setting $\{\theta x(\tau)\}_{0\le\tau\le t} \overset{!}{=} \{x(t-\tau)\}_{0\le\tau\le t}$ in Eq. **1** may yield an erroneous positive entropy production even for

Author contributions: K.B., A.G., and D.E.M. designed research; K.B., K.S., E.V., A.G., and D.E.M. performed research; K.B. contributed analytic tools; K.B. analyzed data; and K.B., A.G., and D.E.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.B.K. is a guest editor invited by the Editorial Board.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: agodec@mpinat.mpg.de or makarov@cm.utexas.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2318333121/-/DCSupplemental.

Published April 16, 2024.

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

coarse-grained representations of manifestly reversible microscopic dynamics, as shown in refs. 15–17 and elaborated below.

If the time-reversal operation and coarse-graining are carried out correctly, projected representations of microscopic dynamics are usually "less irreversible" than the true dynamics (16, 18–22). That is, if a coarse-grained representation $\hat{x}(t)$ of the full dynamics is considered, we typically expect to underestimate* the true entropy production (16, 18, 19, 23), i.e.,

$$0 \leq \langle \dot{S}[\hat{x}] \rangle \leq \langle \dot{S}[x] \rangle. \quad [2]$$

Recently, there has been a surge in interest in thermodynamic inference from coarse-grained, partially observed dynamics (16, 19, 20, 22, 24–26). The arguably most direct method to infer a lower bound on the entropy production is via the thermodynamic uncertainty relation (TUR) (9, 27–39). Moreover, approaches were developed that exploit the information encoded in the non-Markovian character to infer bounds on dissipation (21, 22, 40–42).

Milestoning at Coarse Resolution

Given the inequality in Eq. **2**, optimal estimation of entropy production from experimental observations remains an open question. Indeed, different coarse-grained representations yield different estimates; moreover, in any realistic experimental setting, neither the microscopic dynamics nor the precise coarse-graining are known. Recently, it was found that coarse-graining microscopic Markov dynamics via a method called "milestoning" (to be explained below) preserves the microscopic entropy production even in the absence of a timescale separation between hidden and observed degrees of freedom, if no dissipative cycles† become hidden by the coarse-graining, and the resulting milestone process is a 1st order semi-Markov process (16, 43). However, milestoning arbitrary Markov dynamics like other types of coarse-graining generally yields arbitrarily complex non-Markovian processes with long-range memory. Moreover, one can also milestone non-Markovian processes. The thermodynamics of such "long-range" non-Markovian processes remains only poorly understood.

The idea of milestoning, a method originally developed for efficient computation (44–46), is illustrated in Fig. 1, where we consider the example of a molecular motor whose dynamics is modeled as driven diffusion in a periodic potential on a ring (47). Owing to experimental limitations, the observation yields a binned version of the original microscopic trajectory $\varphi \to \hat{\varphi}$ (Fig. 1 *A* and *C*) with a finite-time resolution. This is an example of the common coarse-graining approach known as "lumping" (18). The binned trajectory may be highly non-Markovian even when the underlying detailed dynamics is Markov. As a result, the entropy production estimated with Eq. **1** under the assumption that the binned trajectory is a Markov random walk is far lower than the true value, see Fig. 1*D* as well as (43).

Milestoning involves further postprocessing of the binned trajectory, resulting in a trajectory hopping between a set of milestones that do not cover the entire configuration space‡

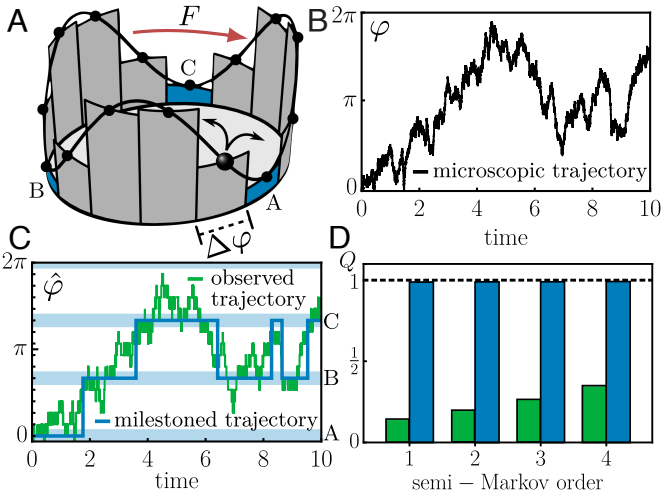

Fig. 1. Milestoning a coarse observation of a molecular motor. (*A*) Microscopic trajectory of a cyclical molecular motor modeled as diffusion along a periodic potential (the precise parameters of the model are given in *SI Appendix*). Owing to a finite spatiotemporal resolution, the motor state encoded in the angle φ is further binned into intervals of size $\Delta\varphi$ (here $\Delta\varphi = 2\pi/15$). Milestoning involves further postprocessing of the coarse trajectory: Some coarse states (here, the bins *A*, *B*, and *C*) are declared to be milestones (blue). Upon entering a milestone, the state of the milestone trajectory remains in said milestone until entering a different milestone for the first time. (*B*) Unobservable microscopic trajectory of the motor. (*C*) Observable binned trajectory of the motor (green), and the corresponding milestoned trajectory (blue). The entropy-production rate is estimated using Eq. **1** assuming different ranges of memory encoded in the so-called semi-Markov order; in the *k*th semi-Markov order, it is assumed that the motor's state at any given time *t* depends on the time it entered the state and the sequence of $k-1$ previously visited states. (*D*) Quality factor *Q*—the ratio of estimated entropy production and its exact value Eq. **6**)—for the observed (green) and milestoned (blue) trajectories, shown as a function of the assumed semi-Markov order. The strong dependence of the estimate on the assumed semi-Markov-order of the observed (i.e., binned) trajectory reflects memory in coarse dynamics (despite the fact that the microscopic trajectory is itself Markovian). In stark contrast, the milestoned trajectory, here approximately a 1st order semi-Markov process, yields the essentially exact entropy production.

(Fig. 1*C*). In the particular case of Fig. 1, milestoning of the highly non-Markov binned trajectory yields a 2nd order semi-Markov random walk, and the application of Eq. **1** recovers the true microscopic entropy production. Note that as a result of placing the milestones at equivalent sites in the potential, and taking a small bin size, the process is in fact approximately a 1st order semi-Markov random walk (recall that transition path-times of one-dimensional diffusion processes are reflection symmetric).

Here, we show—more generally and intriguingly—that milestoning, if used to postprocess lumped dynamics, can improve thermodynamic inference for a broad class of systems. In other words, milestoned trajectories obtained by discarding certain details of lumped trajectories (Fig. 1), can provide improved estimates of dissipation; as such, the milestoning analysis is directly applicable to experimentally observed dynamics, which are inherently lumped.

We explicitly address the scenario where lumping hides dissipative cycles and compare the milestoning estimate with the bound inferred from the TUR. We show how postlumped milestoning may be used systematically to improve thermodynamic inference, i.e., to enhance the precision of dissipation estimates from coarse observations. We stress the importance of considering a physically consistent time-reversal operation in the presence of memory and show that distinct end-state dependent waiting-

*In case of a perfect timescale separation between the observed and hidden degrees of freedom, the total entropy production is not necessarily underestimated by a coarse-grained representation.

†A dissipative cycle in a Markov process is a cycle for which the Kolmogorov cycle criterion is violated. The cycle is said to be hidden if the states of said cycle are coarse-grained into the same lump.

‡Note that in this example and throughout the rest of the paper milestones are (coarse) states, as in refs. 16, 43, 48, unlike in earlier versions (45, 46) where milestones were defined as interfaces between pairs of coarse states. In this example, both versions of milestoning give indistinguishable results. See ref. 43 for further discussion.

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

time contributions that emerge upon milestoning generally do not contribute to dissipation.

Milestoning Coarse Dynamics with Hidden Cycles

We begin with a toy model that contains hidden dissipative cycles (see refs. 49–51 for similar toy models). Models of this kind can emerge as descriptions of a system of magnetically coupled colloidal particles (8, 52), chemical reaction networks (53), circular hidden Markov models (54, 55), and conformational rearrangements in molecular machines (3). Here we consider a discrete-time Markov process as shown in Fig. 2*A*. The observed dynamics involves $N_{\rm lump}$ observable states arranged along a ring, and each such observable state contains $N_{\rm hid}$ microscopic states forming a smaller ring (Fig. 2*A*), such that the total number of states is given by $N_{\rm lump} \times N_{\rm hid}$. Within each lump there are two neighboring microstates which are connected to two microstates within the previous lump (e.g., microstate $1 \leftrightarrow 31$ and $2 \leftrightarrow 30$ in Fig. 2*A*), and two neighboring microstates which are connected to two microstates within the next lump (e.g., microstate $3 \leftrightarrow 5$ and $4 \leftrightarrow 8$ in Fig. 2*A*). The transition probabilities for internal microstates (i.e., microstates not connecting two lumps for $N_{\rm hid} > 4$) are $w_{\rm hid,+}$ in the clockwise direction, and $w_{\rm hid,-} = 1 - w_{\rm hid,+}$ in the counterclockwise direction. For microstates connecting two lumps, we set $w_{\rm lump,\pm}$ for transitions between the lumps, and $(1 - w_{\rm lump,\pm})w_{\rm hid,\pm}$ for internal transitions. Thus, for $w_{\rm hid,+} \neq w_{\rm hid,-}$, the lumped states hide dissipative cycles when $N_{\rm hid} > 2$. We now milestone the trajectory by declaring four equidistant lumps as milestones (see lumps I, III, V, and VII in Fig. 2*A*).

To evaluate $\langle \dot{S}[\hat{x}_i] \rangle$, where the subscript $i = \{\text{lump}, \text{mil}\}$ indicates whether the entropy production is estimated for the lumped trajectory or the milestone trajectory, we use Eq. **1** and measure lump/milestone-sequence frequencies in the trajectory (56–58). In the presence of long-range memory, this is a computationally demanding task requiring many independent steady-state trajectories or one extremely long trajectory. Even if this is given, a finite range of memory must be assumed in practical computations of path probabilities $P[\{x(\tau)\}_{0 \le \tau \le t}]$ and $P[\{\theta x(\tau)\}_{0 \le \tau \le t}]$. A numerical implementation of the inference is provided in ref. 59.

Note that the lumped trajectory $\hat{x}_i(t)$ is indeed a non-Markovian process. This is manifested in the fact that the estimate of $\langle \dot{S}[\hat{x}_i] \rangle$ strongly depends on the assumption of the underlying extent of memory. That is, considering only the preceding step—in the 1st order semi-Markov approximation—we assume that $\hat{x}_i(t)$ depends only on $\hat{x}_i(t-1)$ but not on $\hat{x}_i(t-k)$ for $k \ge 2$, leading to the one-step affinity estimate (assuming all coarse states are equivalent as in Fig. 2*A*)

$$\langle \dot{S}_1^{\rm aff}[\hat{x}_i] \rangle = \frac{p_+^i - p_-^i}{\langle \tau_i \rangle} \ln\left(\frac{p_+^i}{p_-^i}\right), \qquad [3]$$

where $\langle \tau_i \rangle$ is the average waiting time within a coarse-grained state (i.e., the time spent in a coarse-grained state before entering a new coarse-grained state), and $p_\pm^i$ are the forward/clockwise (+) and backward/counterclockwise (−) jump probabilities in the sequence of distinct coarse-grained states (i.e., upon removing repeated consecutive coarse-grained states).

In the kth order semi-Markov approximation, we assume that the probability of $\hat{x}_i(t)$ depends on the sequence $\{\hat{x}_i(t-1), \ldots, \hat{x}_i(t-k)\}$. To be concrete, for a 2nd order semi-Markov process (40) the jump probabilities and waiting times depend on the previous state. Naively, taking $\{\theta\hat{x}(\tau)\}_{0 \le \tau \le t} \overset{!}{=} \{\hat{x}(t-\tau)\}_{0 \le \tau \le t}$ in Eq. **1**, yields two contributions:

$$\lim_{t\to\infty} \frac{1}{t} \left\langle \ln \frac{P[\{\hat{x}_i(\tau)\}_{0 \le \tau \le t}]}{P[\{\hat{x}_i(t-\tau)\}_{0 \le \tau \le t}]} \right\rangle = \langle \dot{S}_2^{\rm aff}[\hat{x}_i] \rangle + \langle \dot{S}_2^{\rm wt}[\hat{x}_i] \rangle, \qquad [4]$$

where $\langle \dot{S}_2^{\rm aff}[\hat{x}_i] \rangle$ is the two-step affinity contribution, and $\langle \dot{S}_2^{\rm wt}[\hat{x}_i] \rangle$ a contribution of waiting times (40). In particular, for the model in Fig. 2*A* the two-step affinity contribution reads (40, 43) (assuming all coarse states are equivalent)

$$\langle \dot{S}_2^{\rm aff}[\hat{x}_i] \rangle = \frac{p_+^i - p_-^i}{\langle \tau_i \rangle} \ln\left(\frac{\phi_{++}^i}{\phi_{--}^i}\right), \qquad [5]$$

where $\phi_{\pm\pm}^i$ are conditional splitting probabilities of making a forward/backward jump, given that the previous jump occurred in the forward/backward direction. In anticipation that waiting times do not contribute to dissipation, we ignore them in the inference of $\langle \dot{S}_k[\hat{x}_i] \rangle$. Later, we also prove that they indeed result in a spurious contribution to the estimate for $\langle \dot{S}[x] \rangle$. For nonequivalent coarse states the correct expression for $\langle \dot{S}_1^{\rm aff} \rangle$ and $\langle \dot{S}_2^{\rm aff} \rangle$ are given in *SI Appendix*.

Taking a single ergodically long trajectory, we use Eq. **3** or Eq. **5** for the microscopic, lumped, and postlump milestoned trajectories to estimate the entropy production. To determine the accuracy of the estimates, we calculate the quality factor $0 \le Q \le 1$, equal to the ratio of the estimated entropy production to the true entropy production of the microscopic process:

$$Q = \langle \dot{S}_k^{\rm aff}[\hat{x}_i] \rangle / \langle \dot{S}[x] \rangle, \qquad [6]$$

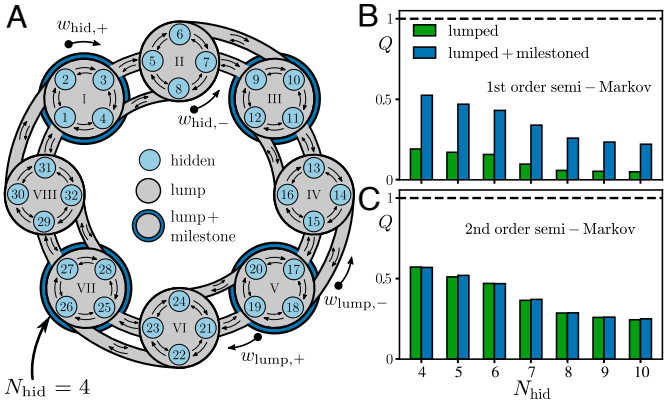

Fig. 2. Estimating entropy production from a coarse-grained discrete-time Markov chain with hidden cycles. (*A*) Schematic of a discrete-state Markov chain with 32 hidden microstates (indicated with Arabic numerals) lumped into eight domains (indicated with Roman numerals). In this example, each of the eight lumps contains $N_{\rm hid} = 4$ microstates. Such a partitioning arises naturally when the underlying process is observed with coarse resolution. Milestoning involves further postprocessing of the lumped process: Some lumped states (here, the domains I, III, V, and VII encircled in blue) are declared to be milestones. Once the trajectory enters a milestone, the coarse-grained state of the milestone trajectory remains in said milestone until it enters a different milestone for the first time. We choose transition rates $\{w_{\rm hid,+}, w_{\rm hid,-}\} = \{w_{\rm lump,+}, w_{\rm lump,-}\} = \{0.6, 0.4\}$, 32 lumped states, 2×10^8 discrete time steps, and we vary the number of hidden states from 4 to 10. (*B* and *C*) Quality factor Q given by Eq. **6** for the entropy production estimated from the lumped (green bars) and postlumped milestoned (blue bars) trajectory assuming (*B*) 1st order semi-Markov (Eq. **3**) and (*C*) 2nd order semi-Markov (Eq. **5**).

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

where $i = \{\text{lump, mil}\}$. We apply the 1st order (i.e., $k = 1$) semi-Markov and 2nd order (i.e., $k = 2$) semi-Markov approximation to estimate $\langle \dot{S}[\hat{x}_i] \rangle$, and the results are shown in Fig. 2 *B* and *C*. For $N_{\text{hid}} \geq 4$ the estimates from both coarse-grained trajectories (i.e., lumped and milestoned) are significantly lower than the true entropy-production rate. This is due to the presence of hidden dissipative cycles. Nevertheless, within the 1st order semi-Markov approximation, milestoning the lumped trajectory improves the entropy-production estimate (compare green and blue bars in Fig. 2*B*). On the 2nd order semi-Markov level milestoning neither improves nor worsens the estimate (Fig. 2*C*).

Interestingly, milestoning lumped dynamics apparently may improve the entropy-production estimate using Eq. **1** even in the presence of hidden dissipative cycles. Notably, in practice, the extent of memory is unlikely to be known, and computations at orders higher than 2 quickly become unfeasible.

Milestoning in Driven Single-File Diffusion

Next, we consider tracer-particle dynamics in discrete-time single-file diffusion on a ring (Fig. 3), a paradigmatic model of dynamics with long memory and anomalous diffusion (60–66) as well as a paradigm for diffusion in crowded media (67–71) and transport through biological channels (72). Let N_{sites} be the total number of sites on the ring, and N_{vac} the number of empty sites (vacancies), such that the total number of particles is $N = N_{\text{sites}} - N_{\text{vac}}$. At each discrete time step a particle is picked at random and shifted with a given probability to the left or right if the new site is unoccupied. Note that the latter condition introduces waiting times between jumps, i.e., it effectively gives rise to local Poissonian clocks in the limit of continuous time. We track the position of a tracer particle (see e.g., Fig. 3*A*, dark blue), which results in a lumped process, where multiple microscopic states correspond to the same location of the tracer particle but different arrangements of the unobserved "bath" particles.

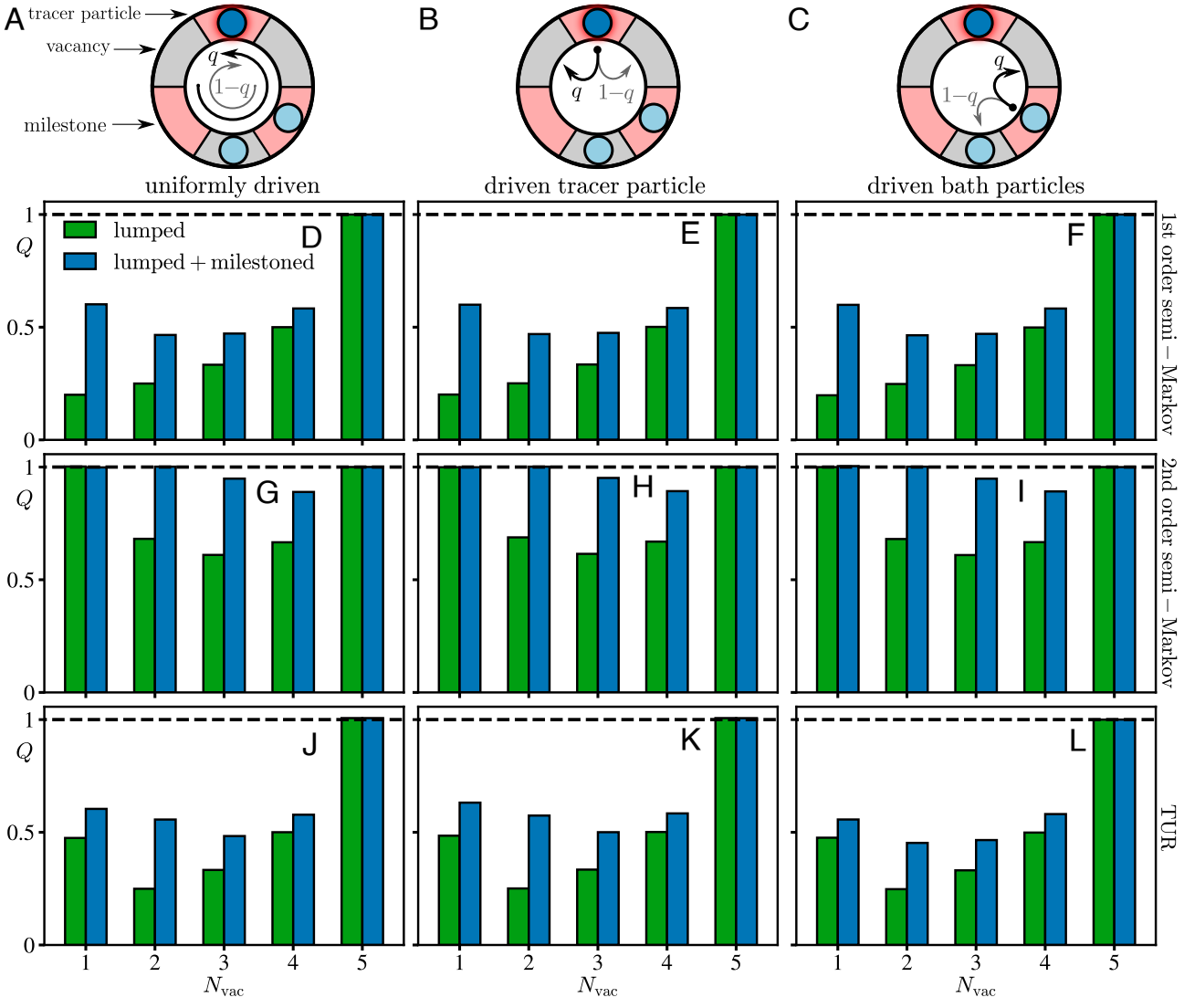

Fig. 3. Estimating the entropy production in single-file diffusion for different driving schemes. (*A–C*) Schematic of the different driving scenarios, where the tracer particle is indicated in dark blue and bath particles in light blue. Red sites indicate the milestone positions for the tracer particle. In each panel, we consider $N_{\text{sites}} = 6$ sites and vary the number of vacancies N_{vac} from 1 to 5. Asymmetric transition probabilities are $\{1-q, q\}=\{0.55, 0.45\}$. Entropy-production estimates are obtained from trajectories with 2×10^8 discrete steps. (*D–F*) Quality factor of the entropy production estimated with the 1st order semi-Markov approximation (Eq. **3**). (*G–I*) Quality factor of the entropy production estimated with the 2nd order semi-Markov approximation (Eq. **5**). (*J–L*) Quality factor of the entropy production estimated with the continuous-time TUR (Eq. **7**) using the tracer particle current.

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

We consider three different driving scenarios: i) a uniformly driven system with clockwise (w_+) and counterclockwise (w_-) transition probabilities $\{w_+, w_-\}=\{1-q, q\}$ (Fig. 3*A*), ii) a driven tracer particle with transition probabilities $\{w^{\mathrm{t}}_+, w^{\mathrm{t}}_-\}=\{1-q, q\}$ and bath particles with symmetric hopping probabilities $\{w_+, w_-\}=\{1/2, 1/2\}$ (Fig. 3*B*) (73–77), and iii) the opposite scenario where the bath particles are driven, and the tracer particle hops symmetrically (Fig. 3*C*) (78).

We further consider three maximally separated postlumped milestones on the ring (see e.g., Fig. 3*A*, red sites). This type of milestoning corresponds to that implemented in (43) and yields the exact entropy production in the case of a single particle (16, 43).

In each driving scenario, we evaluate, numerically, the exact (microscopic) entropy production $\langle \dot{S}[x] \rangle$ and compare it with $\langle \dot{S}^{\mathrm{aff}}_k[\hat{x}_i] \rangle$ assuming 1st order (i.e., $k = 1$) (Fig. 3 *D–F*) and 2nd order (i.e., $k = 2$) semi-Markov (Fig. 3 *G–I*) statistics for the lumped and postlumped milestoned trajectories, respectively. For all driving scenarios, and both semi-Markov approximations, we find that milestoning of the lumped trajectories significantly improves the entropy-production estimate.

For further comparison, we also estimate $\langle \dot{S}[\hat{x}_i] \rangle$ via the TUR (27, 79), which provides a universal trade-off between precision of any thermodynamic flux and the dissipation in the system (27, 38, 79). We choose the tracer-particle current $J^i_k(n) \equiv \hat{x}_i(n+k) - \hat{x}_i(n)$ in the lumped and postlumped milestoned trajectories, and determine the average $\langle J^i_k \rangle$ and variance $\mathrm{var}(J^i_k)$. The entropy production estimate obtained with the continuous-time TUR reads (27)

$$\langle \dot{S}_{\mathrm{TUR}}[\hat{x}_i] \rangle \equiv \sup_{k|1\le k\ll T} \left[\frac{2\langle J^i_k \rangle^2}{\mathrm{var}(J^i_k) k} \right], \qquad [7]$$

where the mean and variance of the current can be inferred as time averages from long trajectories (note that the particle position is a weakly ergodic process)

$$\langle J^i_k \rangle = \frac{1}{T-k} \sum_{n=1}^{T-k} J^i_k(n), \qquad [8]$$

$$\mathrm{var}(J^i_k) = \frac{1}{T-k} \sum_{n=1}^{T-k} (\langle J^i_k \rangle - J^i_k(n))^2, \qquad [9]$$

where T is the total trajectory length. The result for the quality factor is shown in Fig. 3 *J–L*. Here, by using the continuous-time TUR, we invoke the standard convergence of the discrete-time Markov chain to a continuous-time Markov process in the limit $k, T \to \infty$ under the condition $T \gg k$, justifying the above approach. Note, moreover, that for lumped dynamics it was proven that $0 \le \langle \dot{S}_{\mathrm{TUR}}[\hat{x}_{\mathrm{lump}}] \rangle \le \langle \dot{S}[x] \rangle$ (79), but is a priori not clear whether the TUR holds also for the postlumped milestoned dynamics. Here, we simply assume this to be true based on the numerical evidence and find that the milestoned estimates outperform the lumped estimates (see *Bottom* panel in Fig. 3).

Increasing Ignorance

We now show that, at first sight paradoxically, when milestoning is used, a larger distance between milestones and hence "more ignorance" may improve thermodynamic inference.

Let us focus on uniformly driven single-file diffusion with one vacancy, i.e., $N_{\mathrm{vac}} = 1$, as shown in Fig. 4. In this case,

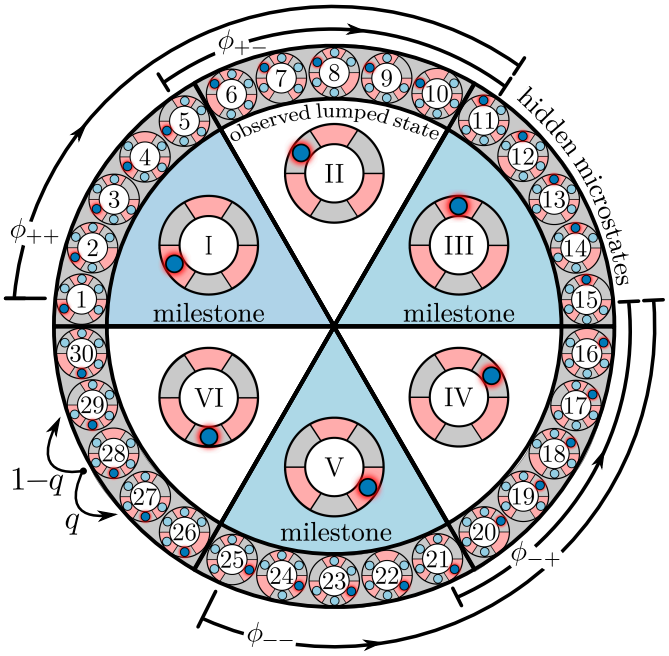

Fig. 4. Microscopic description of single-file diffusion with one vacancy. The outer circular layer depicts the hidden microstates: The underlying microscopic model is a random walk on this outer ring of states, with transition probabilities as indicated. The inner circle depicts the lumped states, where we only observe the location of the tracer particle (dark blue). Each lumped state contains $N_{\mathrm{sites}}-1$ hidden microstates. The random walk on milestones is a 2nd order semi-Markov process, with the sequence of milestones visited by the random walker determined by the corresponding splitting probabilities $\phi_{\pm\pm}$ indicated by the outer lines.

the microscopic dynamics of the system can be described by a particularly simple kinetic scheme, with the microscopic states arranged on a ring. Specifically, for a given number of sites N_{sites}, there are $N_{\mathrm{sites}}-1$ microstates with a fixed location of the tracer particle. There are in total $N_{\mathrm{sites}}(N_{\mathrm{sites}}-1)$ microscopic states, with groups of $N_{\mathrm{sites}}-1$ states forming the lumped states of the macroscopic system (i.e., the position of the tracer particle). We now designate some of those coarse states to be milestones. Specifically, suppose we take 3 milestones with N_C coarse states between adjacent milestones, such that $3 + 3N_C = N_{\mathrm{sites}}$ (see Fig. 4 where $N_{\mathrm{sites}} = 6$ and $N_C = 1$). For simplicity, we here assume that the milestones are equidistantly placed. In *SI Appendix*, we also consider nonequidistant milestones, for which the final result given by Eq. **10** remains identical.

We apply milestoning to obtain the trajectory $\hat{x}_{\mathrm{mil}}(t)$ as described before: The position of the tracer particle corresponds to the last visited milestone, resulting in a 2nd order semi-Markov process. Specifically, the milestone sequence can be described by the conditional splitting probabilities in the sequence of distinct states (i.e., upon removing repeated consecutive milestoned states) $\{\phi_{++}, \phi_{+-}, \phi_{-+}, \phi_{--}\}$, where ϕ_{+-}, for example, is the conditional probability of making a clockwise jump from one milestone to the next, given that the previous milestone jump was counterclockwise (i.e., III → I → III in Fig. 4).

To calculate the splitting probabilities ϕ_{++} and ϕ_{-+}, for example, we consider the random walker starting in the leftmost state of the segment of $N_{\mathrm{sites}}-1$ forming a milestone. If we denote this state by k, then ϕ_{-+} and ϕ_{++} are the probabilities to exit the interval $\{k-\Delta_-, \ldots, k+\Delta_+\}$ through its left and right boundaries with $\{\Delta_-, \Delta_+\}=\{N_C(N_{\mathrm{sites}}-1)+1, (N_C+1)(N_{\mathrm{sites}}-1)\}$. One may recognize that this problem is equivalent to the Gambler's

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

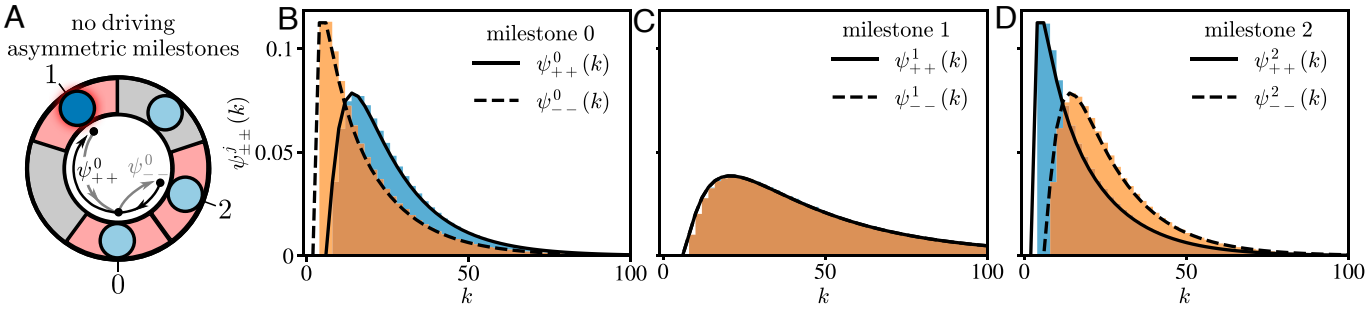

Fig. 5. Asymmetrically placed milestones result in distinct direction-dependent waiting-time distributions irrespective of the driving. (*A*) Symmetric single-file diffusion, that is, with no driving (i.e., $q = 1/2$), for $N_{\text{sites}} = 5$, and $N_{\text{vac}} = 1$. The three milestones correspond to the positions of the tracer particle (dark blue) indicated in red. (*B–D*) Waiting-time distributions $\psi^j_{\pm\pm}(k)$ in each of the milestones obtained by simulations (blue + orange) and by exact analytical results (full + dashed black lines, see *SI Appendix*).

ruin problem (80); its analytical solution is given explicitly in *SI Appendix*. Given these splitting probabilities, we can determine the entropy production at the level of the 1st order semi-Markov approximation using Eq. **3**, which gives

$$\langle \dot{S}_1^{\text{aff}}[\hat{x}_{\text{mil}}]\rangle = \frac{3/(N_{\text{sites}}-1)+3N_C}{3(1+N_C)}\langle \dot{S}[x]\rangle = \frac{3/(N_{\text{sites}}-1)+N_{\text{sites}}-3}{N_{\text{sites}}}\langle \dot{S}[x]\rangle, \qquad [10]$$

where $\langle \dot{S}[x]\rangle$ is the exact entropy production given by[§]

$$\langle \dot{S}[x]\rangle = \frac{2q-1}{N}\ln\left(\frac{q}{1-q}\right). \qquad [11]$$

We immediately see that Eq. **10** yields the exact entropy production in the limit $N_{\text{sites}} \to \infty$. This behavior can be understood, qualitatively, as follows: As the total intermilestone spacing increases, the relative difference between splitting probabilities such as, e.g., ϕ_{+-} and ϕ_{++} becomes negligible because the relative milestone size (in the microscopic view) is negligible compared to the milestone territory, which suppresses correlations between entry- and exit-directions to and from the milestone, respectively. In other words, the milestone trajectory approaches a 1st order semi-Markov process. But in the latter case, we expect milestoning to become exact (16, 43).

For the "bare" lumped trajectory, we simply set $N_C = 0$ in the first line of Eq. **10**, which gives

$$\langle \dot{S}_1^{\text{aff}}[\hat{x}_{\text{lump}}]\rangle = \frac{1}{N_{\text{sites}}-1}\langle \dot{S}[x]\rangle. \qquad [12]$$

Hence, in the limit $N_{\text{sites}} \to \infty$, the 1st order semi-Markov entropy production estimate in the lumped dynamics vanishes.

Spurious Waiting-Time Contribution and Kinetic Hysteresis

Notably, using Eq. **5** on the uniformly driven single-file system with one vacancy, we recover the exact entropy production, i.e.,

$$\langle \dot{S}_2^{\text{aff}}[\hat{x}_{\text{mil}}]\rangle = \langle \dot{S}_2^{\text{aff}}[\hat{x}_{\text{lump}}]\rangle = \langle \dot{S}[x]\rangle, \qquad [13]$$

which we also observe in Fig. 3 *G–I* at $N_{\text{vac}} = 1$. The two-step affinity also remains exact when the milestones are not placed equidistantly, which is shown explicitly in *SI Appendix*.

Importantly, in Eq. **13**, we have systematically discarded the waiting-time contribution $\langle \dot{S}_2^{\text{wt}}[\hat{x}_i]\rangle$ from Eq. **4** that follows from naively applying Eq. **1** to 2nd order semi-Markov trajectories (40). A direct computation shows (40)

$$\langle \dot{S}_2^{\text{wt}}[\hat{x}_{\text{mil}}]\rangle = \frac{1}{\langle \tau_{\text{mil}}\rangle}\sum_j\sum_{\pm} p_{j,\pm\pm} D[\psi^j_{\pm\pm}||\psi^j_{\mp\mp}] \geq 0, \qquad [14]$$

where $p_{j,\pm\pm} = \pi_{j\mp1}p_{j\mp1,\pm}\phi_{j,\pm\pm}$ are the jump probabilities centered around milestone j to make two consecutive jumps in the clockwise ($j-1 \to j \to j+1$) or counterclockwise ($j+1 \to j \to j-1$) direction, and where

$$D[\psi^j_{\pm\pm}||\psi^j_{\mp\mp}] \equiv \sum_{k\geq 1}\psi^j_{\pm\pm}(k)\ln[\psi^j_{\pm\pm}(k)/\psi^j_{\mp\mp}(k)], \qquad [15]$$

is the Kullback–Leibler divergence between the conditional waiting-time distributions for consecutive forward $\psi^j_{++}(k)$ and backward $\psi^j_{--}(k)$ transitions (analytical results for the waiting time distributions are shown in *SI Appendix*). Note that as soon as the conditional forward and backward waiting time distributions are distinct (i.e., asymmetric), i.e., $\psi^j_{++}(k) \neq \psi^j_{--}(k)$ for some k, Eq. **14** implies dissipation, irrespective of whether the microscopic dynamics is truly dissipative or not.

In Fig. 5, we show that as soon as the distance to the left/right neighboring milestone is not symmetric, we obtain $D[\psi^j_{\pm\pm}||\psi^j_{\mp\mp}] > 0$, and thus it follows from Eq. **13** that $\langle \dot{S}_2^{\text{aff}}[\hat{x}_i]\rangle + \langle \dot{S}_2^{\text{wt}}[\hat{x}_i]\rangle > \langle \dot{S}[x]\rangle$. Moreover, even under perfect (microscopic) detailed balance (i.e., q=1/2), we find that the waiting-time distributions are asymmetric (Fig. 5). Therefore, $D[\psi^j_{\pm\pm}||\psi^j_{\mp\mp}]>0$ and thus $\langle \dot{S}_2^{\text{wt}}[\hat{x}_{\text{mil}}]\rangle > \langle \dot{S}[x]\rangle = 0$. The above 2nd order semi-Markov example and the family of 1st order semi-Markov counterexamples in refs. 15–17 disprove that waiting times in general contribute to dissipation in Eq. **1**. The waiting-time contribution in Eq. **14** in our example quantifies the difference between the waiting-time distributions between forward versus backward transitions but does not contribute to dissipation.

Note that the postlumped milestoned process displays kinetic hysteresis (i.e., coarse-graining and time reversal do not commute; see *SI Appendix*) (16), whereas the original lumped dynamics does not. For the latter, the waiting-time distributions are

[§] The exact entropy production is conveniently calculated, in the case of a single vacancy, by using the fact that there is one-to-one correspondence between the trajectory of the tracer particle and that of the vacancy, and that the latter undergoes a biased Markovian random walk

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

symmetric irrespective of the lump sizes and presence or absence of driving, resulting in a vanishing waiting-time contribution $\langle \dot{S}_2^{\mathrm{wt}}[\hat{x}_{\mathrm{lump}}]\rangle = 0$. Therefore, in the presence of kinetic hysteresis in coarse-grained dynamics, the waiting-time contribution in Eq. **14** does not contribute to dissipation, and in its absence it typically (i.e., in our examples) seems to vanish (see, however, nonvanishing contributions in ref. 40). And since it does not seem to be possible in practice to test for the presence of kinetic hysteresis without the knowledge of the full dynamics, asymmetric waiting-time distributions generally cannot be used to infer dissipation via Eq. **14**, i.e., there is no implication between the two: $\langle \dot{S}[x]\rangle > 0 \implies \langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle > 0$ and in turn $\langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle > 0 \implies \langle \dot{S}[x]\rangle > 0$. Notably, the complete thermodynamically consistent definition of dissipation for 1st order semi-Markov processes has been established in ref. 16, while for 2nd order semi-Markov processes the affinity contribution $\langle \dot{S}_2^{\mathrm{aff}}[\hat{x}]\rangle$ seems to provide a correct lower bound, $\langle \dot{S}[x]\rangle \geq \langle \dot{S}_2^{\mathrm{aff}}[\hat{x}]\rangle$.

Discussion

Correct Notion of Time Reversal. Contrasting the setting where all hidden degrees of freedom relax instantaneously fast, that is solved essentially completely within the notion of local detailed balance (18), inferring dissipation from general coarse observations remains both conceptually and technically challenging and is still understood very poorly. Sufficiently far from equilibrium, challenges appear even in the presence of a timescale separation between the hidden and observed degrees of freedom (43). In the absence of such timescale separation, and for an arbitrary extent of memory, the problem remains unsolved. This is because a correct general mathematical definition of dissipation in the presence of memory is not known. In particular, the definition in Eq. **1** is incomplete, because the correct general form of the time-reversal operation θ remains elusive. Memory reflects (anti)persistence in dynamics, and it is therefore not surprising that it manifests some abstract analogy to momenta.

Whereas practical methods are available to infer the presence (81, 82) and range (58, 83, 84) of memory in coarse observations, the precise flavor of non-Markovianity is typically not known. In particular, if only coarse-grained trajectories are accessible, such as in experimental studies, it seems to be inherently impossible to check for the presence of kinetic hysteresis. Therefore, referring to the absence of kinetic hysteresis as a "mild assumption" (85) seems somewhat inappropriate. Either way, since $\langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle$ can be positive even under detailed balance, and one cannot check for kinetic hysteresis, a nonvanishing $\langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle$ generally cannot be used to infer dissipation.

Even if the order of memory is known, the correct notion of physical time-reversal generally remains elusive. For semi-Markov processes of 1st order, it has been proven that the transition-path periods (i.e., the duration of successful transitions between coarse states) are odd under time reversal (i.e., transition paths between any pair of states must be reverted in the time-reversed trajectory like momenta in inertial systems), and the thermodynamically consistent time-reversal operation was established (16). Importantly, by explicit counterexamples it was proven that waiting-time effects in 1st and 2nd order semi-Markov processes generally do not measure dissipation via Eq. **1** and Eq. **14**. Note that reverting all transition paths in the time-reversed trajectories ensures $\langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle = 0$, and thus has the same effects as simply ignoring $\langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle$. Moreover, we found that for the 2nd order semi-Markov process with no hidden dissipative cycles in Fig. 3 *G–I* (see also refs. 17 and 43), the two-step affinity exactly recovers the microscopic dissipation, which requires all waiting time contributions to be systematically discarded. To what extent and under which conditions this equality, i.e., $\langle \dot{S}_k^{\mathrm{aff}}[\hat{x}_i]\rangle = \langle \dot{S}[x]\rangle$ in the absence of hidden cycles, is true for general kth order semi-Markov processes ($k > 2$) still needs to be established. Whether a general form of time reversal θ exists that would always yield a consistent waiting-time contribution to dissipation remains unknown. One can construct 2nd (but not 1st) order semi-Markov processes without kinetic hysteresis that have "asymmetric" waiting time distributions whose contribution does not violate the inequality Eq. **2** (40). However, our results show that for 1st and 2nd order semi-Markov processes waiting-time effects as encoded in $\langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle$ are either nominally zero or else must be discarded, since one cannot know (nor can one test) for the presence of kinetic hysteresis without the knowledge of microscopic dynamics. We believe this to be a result of an incorrect or incomplete definition of time reversal θ when applying Eq. **1** to 1st and 2nd order semi-Markov processes. We expect this to also be true for processes with longer memory, where, however, evaluating the path probabilities becomes challenging.

Note, however, that we do not claim that waiting-time statistics generally do not contain information about dissipation (see e.g., refs. 19–21 for how to extract information from waiting time statistics). We only argue that the waiting-time contribution $\langle \dot{S}^{\mathrm{wt}}[\hat{x}]\rangle$ in Eq. **14** emerging from a naive application of Markovian time reversal $\{\theta\hat{x}_\tau\}_{0\leq\tau\leq t} \overset{!}{=} \{\hat{x}_{t-\tau}\}_{0\leq\tau\leq t}$ in Eq. **1** generally does not measure dissipation.

Practical Aspects of Milestoning. From a practical perspective, one therefore does not generally know how to correctly apply Eq. **1** to infer dissipation from observed non-Markovian trajectories. A convenient, practical, and consistent approach we push forward here is to first milestone trajectories and thereafter assume 1st or 2nd order semi-Markov correlations in the evaluation of Eq. **1** excluding any waiting-time contributions. Even if this assumption is not satisfied exactly (i.e., the memory is actually longer-ranged), a meaningful milestoning will systematically render the observations "closer" to the underlying microscopic dynamics and thus improve thermodynamic inference. Moreover, a sparser placement of milestones increases the accuracy of inference and is numerically more efficient.

One is therefore tempted to think that milestones should generally be placed as sparsely as possible for optimal thermodynamic inference. There is, however, a practical limitation. Consider the example of observing a tracer particle in a single-file. If we increase the spacing between milestones, the estimate for $\langle \dot{S}[\hat{x}_{\mathrm{mil}}]\rangle$ per milestone transition indeed gradually improves (i.e., increases). At large distances between milestones, however, the sequence of visited milestones will appear nearly deterministic, whereby transitions against the driving (i.e., time-reversed trajectory segments) will become increasingly unlikely. Thus, the evaluation of Eq. **1** using trajectory-derived path probabilities will require better and better sampling (see *SI Appendix* for a quantitative statement). Analytical estimates of transition probabilities (such as presented here) are of course not affected. In general, the placement of milestones should be optimized as a trade-off between accuracy and statistical feasibility.

Finally, we stress that the proposed postlumped milestoning based thermodynamic inference requires at least one dissipative cycle to be observable. If some dissipative cycles are hidden by the coarse-graining, milestoning-based estimation will yield a

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

true lower bound on dissipation. However, if truly all dissipative cycles become hidden, the method will fail, and one should resort to alternative transition-based approaches developed recently (19–22) to infer some lower bound on dissipation.

Materials and Methods

Stochastic simulations of the continuous-time model of a molecular motor were performed by integrating an overdamped Langevin equation using the Euler–Maruyama method (see, e.g., ref. 46. Stochastic trajectories of discrete-time models such as single-file random walk were generated using the standard kinetic Monte Carlo approach (46). Depending on the problem, the splitting probabilities required for entropy production estimates, as well as the waiting time distributions, were either estimated numerically from the trajectories or, in some cases, calculated analytically as further described in *SI Appendix*. The results of this work are derived in the main text, with some technical details relegated to *SI Appendix*, as indicated.

Supplementary Material. In *SI Appendix*, we provide details of all the calculations and simulation parameters for the results shown in Figs. 1-5.

Data, Materials, and Software Availability. The raw data and source codes to reproduce the results shown in this manuscript are publicly available at ref. 59.

ACKNOWLEDGMENTS. This project has received funding from the German Research Foundation (DFG) through the Heisenberg Program (grant GO 2762/4-1 to AG), from the Robert A. Welch Foundation (Grant No. F- 1514 to D.E.M.), the NSF (Grant Nos. CHE 1955552 to D.E.M. and IIS 2212048 to E.V.), from the Adobe Inc., and from the Alexander von Humboldt Foundation.

Author affiliations: [a]Mathematical biophysics Group, Max Planck Institute for Multidisciplinary Sciences, Göttingen 37077, Germany; [b]Department of Computer Science, University of Texas at Austin, Austin, TX 78712; and [c]Department of Chemistry and Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712

1. R. Yasuda, H. Noji, M. Yoshida, K. Kinosita Jr, H. Itoh, Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase. *Nature* **410**, 898–904 (2001).
2. A. Godec, D. E. Makarov, Challenges in inferring the directionality of active molecular processes from single-molecule fluorescence resonance energy transfer trajectories. *J. Phys. Chem. Lett.* **14**, 49–56 (2023).
3. C. Ratzke, B. Hellenkamp, T. Hugel, Four-colour FRET reveals directionality in the Hsp90 multicomponent machinery. *Nat. Commun.* **5**, 4192 (2014).
4. Y. Tu, The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by Maxwell's demons. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11737–11741 (2008).
5. F. S. Gnesotto, F. Mura, J. Gladrow, C. P. Broedersz, Broken detailed balance and non-equilibrium dynamics in living systems: A review. *Rep. Prog. Phys.* **81**, 066601 (2018).
6. J. Li, J. M. Horowitz, T. R. Gingrich, N. Fakhri, Quantifying dissipation using fluctuating currents. *Nat. Commun.* **10**, 1666 (2019).
7. D. E. Makarov, *Single Molecule Science: Physical Principles and Models* (CRC Press, 2015).
8. J. Mehl, B. Lander, C. Bechinger, V. Blickle, U. Seifert, Role of hidden slow degrees of freedom in the fluctuation theorem. *Phys. Rev. Lett.* **108**, 220601 (2012).
9. C. Dieball, A. Godec, Mathematical, thermodynamical, and experimental necessity for coarse graining empirical densities and currents in continuous space. *Phys. Rev. Lett.* **129**, 140601 (2022).
10. N. Q. Hoffer, M. T. Woodside, Probing microscopic conformational dynamics in folding reactions by measuring transition paths. *Curr. Opin. Chem. Biol.* **53**, 68–74 (2019).
11. B. Schuler, W. A. Eaton, Protein folding studied by single-molecule fret. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
12. H. S. Chung, W. A. Eaton, Protein folding transition path times from single molecule fret. *Curr. Opin. Struct. Biol.* **48**, 30–39 (2018).
13. L. Peliti, S. Pigolotti, *Stochastic Thermodynamics: An Introduction* (Princeton University Press, 2021).
14. U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* **75**, 126001 (2012).
15. H. Wang, H. Qian, On detailed balance and reversibility of semi-Markov processes and single-molecule enzyme kinetics. *J. Math. Phys.* **48**, 013303 (2007).
16. D. Hartich, A. Godec, Emergent memory and kinetic hysteresis in strongly driven networks. *Phys. Rev. X* **11**, 041047 (2021).
17. D. Hartich, A. Godec, Comment on "inferring broken detailed balance in the absence of observable currents" (2022).
18. M. Esposito, Stochastic thermodynamics under coarse graining. *Phys. Rev. E* **85**, 041125 (2012).
19. J. van der Meer, B. Ertel, U. Seifert, Thermodynamic inference in partially accessible Markov networks: A unifying perspective from transition-based waiting time distributions. *Phys. Rev. X* **12**, 031025 (2022).
20. P. E. Harunari, A. Dutta, M. Polettini, E. Roldán, What to learn from a few visible transitions' statistics? *Phys. Rev. X* **12**, 041026 (2022).
21. J. van der Meer, J. Degünther, U. Seifert, Time-resolved statistics of snippets as general framework for model-free entropy estimators. *Phys. Rev. Lett.* **130**, 257101 (2023).
22. J. Degunther, J. van der Meer, U. Seifert, Fluctuating entropy production on the coarse-grained level: Inference and localization of irreversibility (2023).
23. Q. Yu, D. Zhang, Y. Tu, Inverse power law scaling of energy dissipation rate in nonequilibrium reaction networks. *Phys. Rev. Lett.* **126**, 080601 (2021).
24. A. Puglisi, S. Pigolotti, L. Rondoni, A. Vulpiani, Entropy production and coarse graining in Markov processes. *J. Stat. Mech.* **2010**, P05015 (2010).
25. G. Teza, A. L. Stella, Exact coarse graining preserves entropy production out of equilibrium. *Phys. Rev. Lett.* **125**, 110601 (2020).
26. P. Talkner, P. Hänggi, Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical. *Rev. Mod. Phys.* **92**, 041002 (2020).
27. A. C. Barato, U. Seifert, Thermodynamic uncertainty relation for biomolecular processes. *Phys. Rev. Lett.* **114**, 158101 (2015).
28. J. M. Horowitz, T. R. Gingrich, Thermodynamic uncertainty relations constrain non-equilibrium fluctuations. *Nat. Phys.* **16**, 15–20 (2019).
29. T. R. Gingrich, G. M. Rotskoff, J. M. Horowitz, Inferring dissipation from current fluctuations. *J. Phys. A: Math. Theor.* **50**, 184004 (2017).
30. T. V. Vu, V. T. Vo, Y. Hasegawa, Entropy production estimation with optimal current. *Phys. Rev. E* **101**, 042138 (2020).
31. S. K. Manikandan, D. Gupta, S. Krishnamurthy, Inferring entropy production from short experiments. *Phys. Rev. Lett.* **124**, 120603 (2020).
32. S. Otsubo, S. Ito, A. Dechant, T. Sagawa, Estimating entropy production by machine learning of short-time fluctuating currents. *Phys. Rev. E* **101**, 062106 (2020).
33. J. Li, J. M. Horowitz, T. R. Gingrich, N. Fakhri, Quantifying dissipation using fluctuating currents. *Nat. Commun.* **10**, 1666 (2019).
34. T. Koyuk, U. Seifert, Quality of the thermodynamic uncertainty relation for fast and slow driving. *J. Phys. A: Math. Theor.* **54**, 414005 (2021).
35. C. Dieball, A. Godec, Coarse graining empirical densities and currents in continuous-space steady states. *Phys. Rev. Res.* **4**, 033243 (2022).
36. A. Dechant, S. Sasa, Continuous time reversal and equality in the thermodynamic uncertainty relation. *Phys. Rev. Res.* **3**, 042012 (2021).
37. A. Dechant, S. Sasa, Improving thermodynamic bounds using correlations. *Phys. Rev. X* **11**, 041061 (2021).
38. C. Dieball, A. Godec, Direct route to thermodynamic uncertainty relations and their saturation. *Phys. Rev. Lett.* **130**, 087101 (2023).
39. U. Seifert, Stochastic thermodynamics: From principles to the cost of precision. *Phys. A* **504**, 176–191 (2018).
40. I. A. Martínez, G. Bisker, J. M. Horowitz, J. M. Parrondo, Inferring broken detailed balance in the absence of observable currents. *Nat. Commun.* **10**, 3542 (2019).
41. J. Ehrich, Tightest bound on hidden entropy production from partially observed dynamics. *J. Stat. Mech. Theory Exp.* **2021**, 083214 (2021).
42. D. J. Skinner, J. Dunkel, Estimating entropy production from waiting time distributions. *Phys. Rev. Lett.* **127**, 198101 (2021).
43. D. Hartich, A. Godec, Violation of local detailed balance upon lumping despite a clear timescale separation. *Phys. Rev. Res.* **5**, L032017 (2023).
44. J. M. Bello-Rivas, R. Elber, Exact milestoning. *J. Chem. Phys.* **142**, 094102 (2015).
45. R. Elber, Perspective: Computer simulations of long time dynamics. *J. Chem. Phys.* **144**, 060901 (2016).
46. R. Elber, D. E. Makarov, H. Orland, *Molecular Kinetics in Condensed Phases: Theory, Simulation, and Analysis* (John Wiley & Sons, 2020).
47. A. B. Kolomeisky, M. E. Fisher, Molecular motors: A theorist's perspective. *Annu. Rev. Phys. Chem.* **58**, 675–695 (2007).
48. C. Schutte, F. Noé, J. Lu, M. Sarich, E. Vanden-Eijnden, Markov state models based on milestoning. *J. Chem. Phys.* **134**, 204105 (2011).
49. B. Ertel, J. van der Meer, U. Seifert, Operationally accessible uncertainty relations for thermodynamically consistent semi-Markov processes. *Phys. Rev. E* **105**, 044113 (2022).
50. G. Teza, A. L. Stella, Exact coarse graining preserves entropy production out of equilibrium. *Phys. Rev. Lett.* **125**, 110601 (2020).
51. E. Nitzan, A. Ghosal, G. Bisker, Universal bounds on entropy production inferred from observed statistics (2022).
52. M. Uhl, P. Pietzonka, U. Seifert, Fluctuations of apparent entropy production in networks with hidden slow degrees of freedom. *J. Stat. Mech. Theory Exp.* **2018**, 023203 (2018).
53. M. Polettini, A. Wachtel, M. Esposito, Dissipation in noisy chemical networks: The role of deficiency. *J. Chem. Phys.* **143**, 184103 (2015).
54. K. H. Choo, J. C. Tong, L. Zhang, Recent applications of hidden Markov models in computational biology. *Genom. Proteom. Bioinf.* **2**, 84–96 (2004).
55. M. Delorenzi, T. Speed, An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617–625 (2002).
56. E. Roldán, J. M. R. Parrondo, Estimating dissipation from single stationary trajectories. *Phys. Rev. Lett.* **105**, 150607 (2010).
57. E. Roldán, J. M. R. Parrondo, Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems. *Phys. Rev. E* **85**, 031129 (2012).
58. K. Song, R. Park, A. Das, D. E. Makarov, E. Vouga, Non-Markov models of single-molecule dynamics from information-theoretical analysis of trajectories. *J. Chem. Phys.* **159**, 064104 (2023).
59. K. Blom, K. Song, E. Vouga, A. Godec, D. E. Makarov, https://gitlab.gwdg.de/kblom/milestoning-estimators-for-entropy-production (2023).
60. T. E. Harris, Diffusion with "collisions" between particles. *J. Appl. Prob.* **2**, 323–338 (1965).
61. J. L. Lebowitz, J. K. Percus, Kinetic equations and density expansions: Exactly solvable one-dimensional system. *Phys. Rev.* **155**, 122–138 (1967).

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.

62. E. Barkai, R. Silbey, Theory of single file diffusion in a force field. *Phys. Rev. Lett.* **102**, 050602 (2009).
63. N. Leibovich, E. Barkai, Everlasting effect of initial conditions on single-file diffusion. *Phys. Rev. E* **88**, 032107 (2013).
64. A. Lapolla, A. Godec, Unfolding tagged particle histories in single-file diffusion: Exact single- and two-tag local times beyond large deviation theory. *New J. Phys.* **20**, 113021 (2018).
65. A. Lapolla, A. Godec, Single-file diffusion in a bi-stable potential: Signatures of memory in the barrier-crossing of a tagged-particle. *J. Chem. Phys.* **153**, 194104 (2020).
66. D. Hartich, A. Godec, Thermodynamic uncertainty relation bounds the extent of anomalous diffusion. *Phys. Rev. Lett.* **127**, 080601 (2021).
67. O. Bénichou, P. Illien, G. Oshanin, A. Sarracino, R. Voituriez, Tracer diffusion in crowded narrow channels. *J. Physics: Cond. Matt.* **30**, 443001 (2018).
68. P. Illien, O. Bénichou, C. Mejía-Monasterio, G. Oshanin, R. Voituriez, Active transport in dense diffusive single-file systems. *Phys. Rev. Lett.* **111**, 038102 (2013).
69. T. Bertrand, P. Illien, O. Bénichou, R. Voituriez, Dynamics of run-and-tumble particles in dense single-file systems. *New J. Phys.* **20**, 113045 (2018).
70. J. Shin, A. M. Berezhkovskii, A. B. Kolomeisky, Biased random walk in crowded environment: Breaking uphill/downhill symmetry of transition times. *J. Phys. Chem. Lett.* **11**, 4530–4535 (2020).
71. I. M. Sokolov, Models of anomalous diffusion in crowded environments. *Soft Matter* **8**, 9043–9052 (2012).
72. G. Hummer, J. C. Rasaiah, J. P. Noworyta, Water conduction through the hydrophobic channel of a carbon nanotube. *Nature* **414**, 188–190 (2001).
73. S. F. Burlatsky, G. Oshanin, M. Moreau, W. P. Reinhardt, Motion of a driven tracer particle in a one-dimensional symmetric lattice gas. *Phys. Rev. E* **54**, 3165–3172 (1996).
74. I. Lobaskin, M. R. Evans, K. Mallick, Matrix product solution for a partially asymmetric 1D lattice gas with a free defect. *J. Phys. A Math. Theor.* **55**, 205002 (2022).
75. A. Miron, D. Mukamel, Driven tracer dynamics in a one dimensional quiescent bath. *J. Phys. A Math. Theor.* **54**, 025001 (2020).
76. A. Miron, D. Mukamel, H. A. Posch, Phase transition in a 1D driven tracer model. *J. Phys. A Math. Theor.* **2020**, 063216 (2020).
77. A. Kundu, J. Cividini, Exact correlations in a single-file system with a driven tracer. *EPL* **115**, 54003 (2016).
78. T. Banerjee, R. L. Jack, M. E. Cates, Tracer dynamics in one dimensional gases of active or passive particles. *J. Stat. Mech. Theory Exp.* **2022**, 013209 (2022).
79. T. R. Gingrich, J. M. Horowitz, N. Perunov, J. L. England, Dissipation bounds all steady-state current fluctuations. *Phys. Rev. Lett.* **116**, 120601 (2016).
80. J. Vrbik, Gambler's ruin and first passage time. *Math. J.* **14**, 1–11 (2012).
81. A. M. Berezhkovskii, D. E. Makarov, Single-molecule test for Markovianity of the dynamics along a reaction coordinate. *J. Phys. Chem. Lett.* **9**, 2190–2195 (2018).
82. K. Engbring, D. Boriskovsky, Y. Roichman, B. Lindner, A nonlinear fluctuation-dissipation test for Markovian systems. *Phys. Rev. X* **13**, 021034 (2023).
83. A. Lapolla, A. Godec, Toolbox for quantifying memory in dynamics along reaction coordinates. *Phys. Rev. Res.* **3**, L022018 (2021).
84. K. Song, D. E. Makarov, E. Vouga, Compression algorithms reveal memory effects and static disorder in single-molecule trajectories. *Phys. Rev. Res.* **5**, L012026 (2023).
85. F. A. Cisneros, N. Fakhri, J. M. Horowitz, Dissipative timescales from coarse-graining irreversibility. *J. Stat. Mech.* **2023**, 073201 (2023).

Downloaded from https://www.pnas.org by Aljaz Godec on April 16, 2024 from IP address 84.59.54.79.