Letter

Compression algorithms reveal memory effects and static disorder in single-molecule trajectories

Kevin Song [1], Dmitrii E. Makarov [2,3,*] and Etienne Vouga [1]

[1] *Department of Computer Science, University of Texas at Austin, Austin, Texas 78712, USA*

[2] *Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA*

[3] *Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712, USA*

(Received 20 April 2022; revised 7 September 2022; accepted 17 January 2023; published 27 February 2023)

A key challenge in single-molecule studies is deducing underlying molecular kinetics from low-dimensional data, as distinct physical scenarios can exhibit similar observable behaviors such as anomalous diffusion. We show that information-theoretic analysis of single-molecule time series can reliably differentiate Markov (memoryless) from non-Markov dynamics and static from dynamic disorder. This analysis is based on the idea that non-Markov time series can be compressed, using lossless compression algorithms and transmitted within shorter messages than appropriately constructed Markov approximations. In practice, this method detects differences between Markov and non-Markov trajectories even when they are much smaller than the errors of the compression algorithm.

DOI: 10.1103/PhysRevResearch.5.L012026

I. INTRODUCTION

Single-molecule studies that track molecular conformations in real time have opened a window on biomolecular folding, function of molecular machines, and other cellular phenomena. A critical limitation of such experiments, however, is that they report on low-dimensional observables, which are projections of high-dimensional molecular motion. Such projected dynamics are known to be complex and often intractable; they are usually non-Markov processes exhibiting memory [1]. Yet, to describe the time evolution of experimental observables $x(t)$, phenomenological Markovian models, such as biased diffusion or random walk along x [2–4], are commonly invoked. Signatures of non-Markovian dynamics such as anomalous diffusion [5,6] have been reported (see, e.g., Refs. [7–13]), but the challenge then is to select the correct dynamical model out of the multitude of possibilities [5,14]. Data-driven Bayesian inference of models from single-molecule time series has enjoyed considerable success in recent years [15–20], but such studies published so far required physical insight to constrain the space of possible models, and they, too, often assume that the observed dynamics is a one-dimensional random walk even if the number of discrete states is not specified *a priori*. Moreover, Bayesian techniques, which sample full posteriors, come at added computational cost.

Is it possible to tell whether the observed experimental trajectory $x(t)$ can be explained by a Markov process or whether a non-Markov model or a model of a higher dimensionality is called for by the data? When the experimental observable x is a continuous variable, several Markovianity criteria have been found [21–24], but they provide only a necessary and not sufficient Markovianity condition, and for non-Markov processes, they do not quantify the memory length of the process. Other statistical Markovianity tests have recently been proposed for continuous-time jump processes [25]. For single-molecule measurements yielding discrete states, Markovianity of trajectories (or even of candidate hidden-state models) can be assessed by testing for exponentiality of dwell time distributions [26], but again, such exponentiality is only a necessary condition: it is easy to construct an example of a non-Markov random walk with exponential dwell time distributions. Here, we explore a different approach to the problem, which is based on Shannon's classic work [27], where he estimated the information content of printed English. We adapt Shannon's idea to the analysis of single-molecule trajectories (Fig. 1) and show that this method can readily detect memory and static disorder in single-molecule data.

We first describe how Shannon's method is applied to text. The applications to single-molecule data will follow. A text can be described as a sequence of letters, $\ldots i(t-1), i(t), i(t+1)\ldots$, where $i \in \{'A'..'Z'\}$. Let us assume for a moment that each letter occurs independently of the others, with some probability $P(i)$. While ASCII encoding of the letters requires 7 bits per symbol, we can use more bits for rare letters and fewer for common letters to compress the text, with the theoretical compression limit given by Shannon's entropy:

$$h^{(0)} = -\sum_{i=1}^{N} P(i)\log_2 P(i) \tag{1}$$

bits per character (where N is the alphabet size).

*makarov@cm.utexas.edu

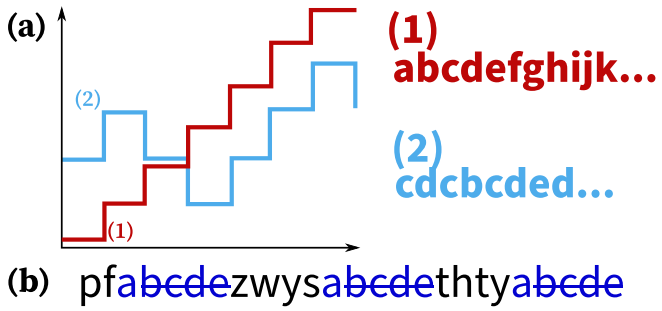

FIG. 1. (a) Compression applied to molecular trajectories and letter sequences. Two molecular motor trajectories. In the first (red), the motor keeps on stepping forward each time. In the second, the motor takes a random step at every direction. Intuitively, the first trajectory can be completely described by "take N steps forward," while a complete description of the second requires us to record every step. In information theory language, the first trajectory/string is characterized by low entropy/information and the second by a much higher one. (b) Memory makes character strings/trajectories compressible. A compression algorithm applied to the string will discover that each "a" is always followed by next four letters of alphabet. Thus, memory of "a" here persists over the next four characters. The algorithm may then take advantage of memory and shorten the string by simply not recording "bcde".

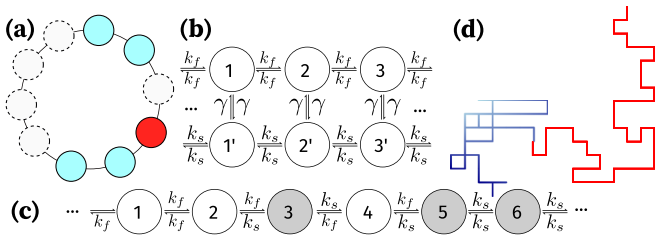

FIG. 2. A summary of models studied here. (a) Single-file diffusion on a ring lattice with $R = 10$ sites and $M = 5$ particles. Each particle (filled circle) can only move to an adjacent vacant lattice site (empty circle). The observer monitors the position of a single tagged particle (red). (b) A random walker with internal states. At each location i, the walker can be found in two experimentally indistinguishable states i and i', with jumps to the neighboring locations occurring with a higher rate k_f (lower rate k_s) for i (i'). Switches between the two internal states occur stochastically with a rate γ. (c) A static disorder model. Each site is randomly chosen to be fast or slow. Transitions leaving slow sites (gray) occur with rate k_s, while transitions leaving fast sites (white) occur with rate k_f. (d) Self-avoiding (SA) random walk (red) and non-SA walk (blue) on a square lattice.

Of course, the assumption of letter independence is unrealistic. A better model would account for the tendency of some letters to appear together: for example, a "t" is more likely to be followed by "h" than by "x". Such pairwise correlations can be included within a model that treats the text as a first-order Markov process, allowing its further compression. In this model, one measures the frequency $P(ij)$ with which a pair of letters ij occurs in the language and computes the conditional probability $T(i \to j) = P(ij)/P(i)$ of seeing j after i. Then the optimal compressed size (per letter) is given by the first-order entropy rate $h^{(1)}$:

$$h^{(1)} = -\sum_{i,j} P(i)T(i \to j)\log_2 T(i \to j). \tag{2}$$

Higher-order models of text are constructed similarly: For a sequence of k consecutive letters $S = i(m), i(m+1), \ldots, i(m+k-1)$ and any j, one can compute the conditional probability $T(S \to j) = P(Sj)/P(S)$ that the sequence S is followed by the letter j. The entropy rate of this kth-order Markov model is

$$h^{(k)} = -\sum_{j,S} P(S)T(S \to j)\log_2 T(S \to j). \tag{3}$$

A key observation is that $h^{(k+1)} \leqslant h^{(k)}$: Knowing more history helps us guess the next character, so the amount of new information revealed by the next character is lower. The true entropy rate of a non-Markov process is the limit $h = \lim_{k\to\infty} h^{(k)}$, which is bounded above by kth-order entropy $h^{(k)}$. This analysis (1) establishes that an English text is not a Markov process, i.e., that $h < h^{(1)}$, (2) quantifies the extent of the memory from observing how fast $h^{(k)}$ converges to h, (3) constructs a kth-order Markov model of the English language, and (4) provides a theoretical limit of how much the text can be compressed. Here, we examine whether similar considerations can be applied to molecular trajectories (Fig. 1), which are viewed as discrete time series $i(0), i(\Delta t), i(2\Delta t)\ldots$, where the molecular state i is sampled at time intervals Δt. Specifically, we consider several models that are commonly used to describe single-molecule phenomena (summarized in Fig. 2), with their dynamics sampled using kinetic Monte Carlo (see, e.g., Refs. [28–30]). We note in passing that earlier related work has explored construction of kth-order Markov models from single-molecule time series [31] and that the idea that a kth-order Markov process maximizes the entropy rate given the known transition probabilities $T(S \to j)$ can also be used to derive a kth-order master equation describing the process [32].

Direct application of Shannon's method to single-molecule data, however, is often not feasible: Evaluating Eq. (3) is computationally prohibitive unless k is small, and estimation of transition probabilities $T(S \to j)$ involving long sequences S from data becomes increasingly inaccurate given finite amounts of data [33], rendering such an entropy estimator unsuitable for long memory. Likewise, direct inspection of transition probabilities $T(S \to j)$, and particularly whether they depend only on the last symbol in the string S, can inform one about the validity of the first-order Markov assumption [34] but becomes prohibitive for high-order Markov models. Instead, here, we explore an approach that does not require construction of high-order Markov models of the process: We estimate the entropy rate h as the size of the output (per time step) of a lossless compression algorithm applied to the trajectory $i(t)$. Specifically, dictionary-type compression algorithms look for repeats of earlier (i.e., occurring at shorter time t) sequences in the data. If repeating data are found [Fig. 1(b)], the algorithm replaces later repeated sequences by references to the earlier occurrences, thereby reducing the output size of the algorithm relative to the original trajectory (in what follows, we report results using LZMA2 lossless encoding [35]).

In applying this idea to realistic trajectories, however, one should consider the errors introduced by the compression algorithm applied to trajectories of finite length. Although for an ergodic, stationary $i(t)$, the size of the output of the algorithm per time step is known to converge asymptotically to the theoretical entropy rate h of the process [36], molecular trajectories are usually not long enough to ensure such convergence in practice. Consequently, the results may depend on the trajectory length and on the specific compression algorithm used. To circumvent this, we estimate the errors introduced by the compression algorithm in the following way: We generate a synthetic first-order Markov process according to the trajectory-derived transition probabilities $T(i \to j)$, calculate its exact entropy rate $h^{(1)}$ according to Eq. (2), and estimate its entropy rate $\tilde{h}^{(1)}$ using the compression algorithm (throughout the rest of this Letter, tilde over an entropy rate indicates a raw compression-derived value). Assuming that the compression method introduces the same error to the original process as to its first-order Markov model that has the same transition probabilities $T(i \to j)$, we then correct the raw compression-derived entropy rate $\tilde{h}$ to estimate the entropy rate as

$$h \approx \tilde{h} - \tilde{h}^{(1)} + h^{(1)}. \tag{4}$$

While plausible, this correction is empirical; while remarkably effective in the examples studied below, we do not know how general it is. Note that, because of statistical errors in estimating the transition probabilities $T(i \to j)$, the estimated value $h^{(1)}$ in general differs from that of the exact first-order entropy rate of the process; for most of the cases studied here, however, this difference is negligible when compared with the errors introduced by the compression algorithm (Supplemental Material, Fig. S9 [37]). In other words, the statistical errors resulting from the last term in Eq. (4) are almost always immaterial (see, however, Supplemental Material, Fig. S8 [37], for an exception; see Supplemental Material, Fig. S9 [37], for a study of statistical errors as a function of the trajectory length). Further note that, for the purpose of detecting non-Markov behavior, the absolute value of h and the value of the last term in Eq. (4) are immaterial: The difference between the compressor-derived entropy rates $h \approx \tilde{h} - \tilde{h}^{(1)}$ already informs us about memory effects. Application of Eq. (4), however, provides a much more stringent test of the ability of the method not only to detect memory but also to estimate the actual entropy rate h. We find that, while the raw compressor-estimated vales show strong dependence on the trajectory length, on the specific compression algorithm used, or on change in the representation of the data [e.g., applying compression to the sequence of steps $l(t) = i(t) - i(t-1)$ instead of the original trajectory $i(t)$], the entropy rates corrected using Eq. (4) remain virtually the same (Supplemental Material, Figs. S2, S6, and S9 [37]).

Another issue that may affect the utility of the method in application to experimental rather than simulated trajectories is that the latter are usually partially degraded by noise. Interestingly, noise or loss of spatial resolution by itself may introduce additional memory not present in the noiseless dynamics, an effect that deserves a more extensive future study. In an example considered in Supplemental Material, Fig. S7 [37], noise effect on the estimated difference $\tilde{h} - \tilde{h}^{(1)}$ is less significant than on the absolute values, and thus, the compression algorithm still correctly detects non-Markovianity of the underlying process.

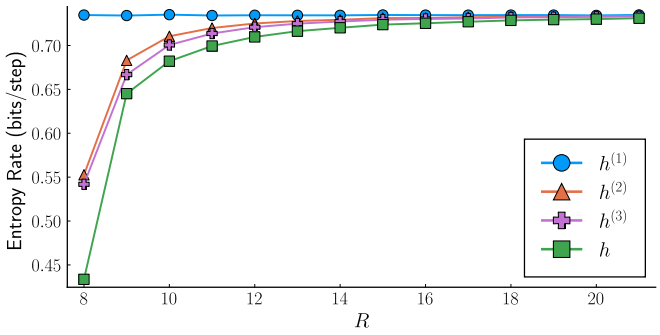

FIG. 3. Compression-derived [and corrected using Eq. (4)] entropy rates of a tracer particle undergoing single-file diffusion on a ring lattice shown as a function of the number R of lattice sites for a fixed number of random walkers, $M = 7$.

II. SINGLE-FILE DIFFUSION

A classic example of a random walk with long memory, single-file diffusion [38,39] [Fig. 2(a)], has applications as the prototype of diffusion in the crowded environment of a biological cell [5,40], passage of multiple solute particles across a biological channel [41], and non-Markovian barrier crossing [42]. Here, we use a discrete-time lattice formulation, in which M particles occupy discrete positions on a ring with R sites. A particle can move to an adjacent site if it is unoccupied, and each step of the single-file diffusion process consists of one such move chosen uniformly at random. An observer monitors the location $i(t)$ of a single tagged particle [red in Fig. 2(a)] as a function of the number t of successive steps.

It is instructive to consider the case with $R = 3$ and $M = 2$ because its true (infinite-order) entropy rate h and the first- and second-order entropy rates can be calculated analytically and are given by $h = 1$ bits/step, $h^{(1)} = 1.5$ bits/step, and $h^{(2)} = 1.25$ bits/step, respectively (Supplemental Material, Fig. S1 [37]). When using transition probabilities $T(i \to j)$, $T(ij \to m)$ estimated numerically from a sampled trajectory instead of their exact values, we obtain nearly identical estimates for $h^{(1)}$ and $h^{(2)}$ (Supplemental Material, Fig. S1 [37]). The true entropy rate, however, must be $h = 1$ bit/step, because the vacant site (dashed circle in Fig. 2(a) and Supplemental Material, Fig. S1 [37]) moves in a Markovian fashion with two equiprobable outcomes, and there is a bijective mapping between the positions of the vacancy and the tagged particle.

Corresponding compression-based estimates obtained using a simulated trajectory of $L = 10^9$ Monte Carlo steps are $\tilde{h}^{(1)} = 1.62$ bits/step and $\tilde{h} = 1.08$ bits/step. Using Eq. (4), the corrected compression-estimated values of the entropy rates are $h^{(1)} = 1.5$ bits/step and $h = 0.96$ bits/step, in better agreement with the theoretical values.

For larger numbers of walkers and sites, the exact entropy rate h is unknown. Figure 3 shows the compression-based estimate of h for $M = 7$ walkers as a function of the number of sites R; h is always lower than the first-order entropy rate $h^{(1)}$.

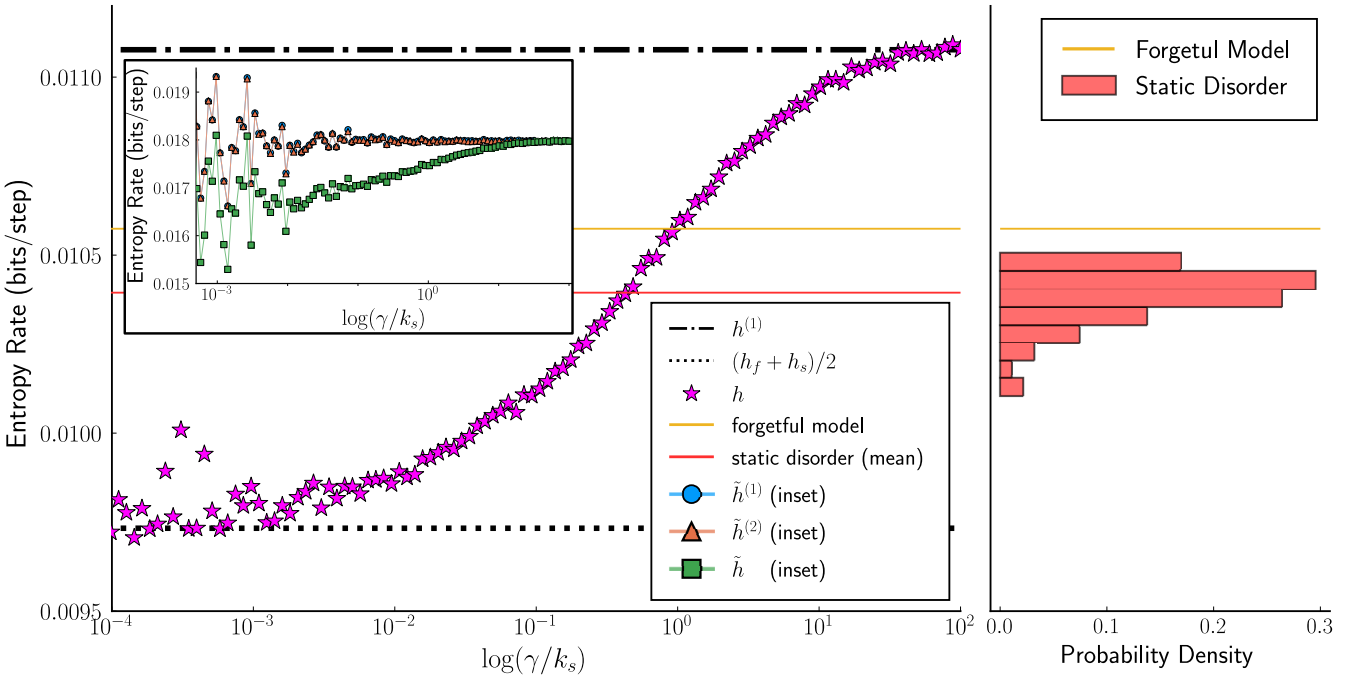

FIG. 4. Left: Estimated entropy rate for the random walk with internal states [Fig. 2(b)]. The ratio of the rates is $\frac{k_f}{k_s} = 10$. In the limit $\gamma \to \infty$, the process becomes Markov, with an entropy rate equal to that of a one-dimensional (1D) random walk with a jumping rate of $(k_s + k_f)/2$, and thus, $h \to h^{(1)}$ (dash-dotted line). At low switching rates, $\gamma \to 0$, the entropy rate is seen to approach the expected value equal to the mean of the entropy rates of two Markovian processes, the slow one (with the jump rate k_s) and the fast one (jump rate k_f). Red line indicates the entropy rate for a random walk with static disorder averaged over random arrangements of slow and fast sites placed on a ring of size $n = 110$. Yellow line indicates the entropy rate for a forgetful random walk. Right: Distribution of entropy rates for individual realizations of static disorder on the ring. Inset: Raw compression-derived entropy rates show significant noise, unlike their values corrected using Eq. (4).

In addition, we have also applied the compression method to higher-order Markov models of the same process ($k = 2,\ 3$), with the estimated values $h^{(2)}$ and $h^{(3)}$ also shown. When $M \approx R$, h is significantly lower than its finite-order estimates $h^{(k)}$, $k \leqslant 3$. This indicates strongly non-Markovian character of single-file diffusion not captured by including memory of past $k \leqslant 3$ states of the particle. As L increases, however, the clashes between walkers become increasingly unlikely, and each walker diffuses freely in the limit $R \gg M$, thus undergoing Markovian dynamics. Accordingly, the true entropy rate estimate h and its k-order Markovian estimates $h^{(k)}$ converge to the same value as R increases. Importantly, the compressor-estimated entropy rates follow the correct relative order $h^{(1)} > h^{(2)} > h^{(3)}$.

III. COARSE-GRAINED SYSTEMS

A fundamental source of dynamical memory is coarse graining [1]. An experiment cannot resolve the individual microscopic states of the molecule, so multiple microscopic states are effectively lumped into collective observable states. An example of coarse graining is given in Fig. 2(b) (additional examples are discussed in the Supplemental Material [37]). In Fig. 2(b), a random walker can be in one of two internal states; in one, the walk is fast (quantified by a jump rate k_f), and in the other, it is slow (jump rate k_s). The walker switches between the states stochastically [43], with a switching rate γ. Models of this type have been used to describe the dynamics of biomolecular motors [4,43–45]. The kinetic scheme [Fig. 2(b)] describing the system consists of fast (enumerated by i) and slow (enumerated by i') tracks, with the walker randomly switching between the two. The time evolution of the probabilities $P(i,t)$ and $P(i',t)$ to occupy sites i and i' obeys the continuous-time master equations:

$$\frac{dP(i,t)}{dt} = -2k_f P(i,t) + k_f P(i-1,t) + k_f P(i+1,t) - \gamma P(i,t) + \gamma P(i',t), \tag{5a}$$

$$\frac{dP(i',t)}{dt} = -2k_s P(i',t) + k_s P(i'-1,t) + k_s P(i'+1,t) - \gamma P(i',t) + \gamma P(i,t), \tag{5b}$$

which describe a Markov process. The two internal states i and i', however, correspond to the same observable position $i = 1, 2, \ldots$ of the walker. The states i and i' are then indistinguishable experimentally and thus lumped into a single coarse state characterized by position i. Unless $k_f = k_s$, the observed time evolution $i(t)$ of this position is non-Markovian.

The compression-derived entropy rates of this random walk are shown in Fig. 4 as a function of the switching rate γ and compared with the entropy rates of the first- and second-order Markov approximations. The true entropy rate is lower than that of the two approximations, indicating non-Markovianity of the dynamics. While significant statistical noise is observed in the raw compression-derived entropy rates (Fig. 4 inset), the corrected value of h [Eq. (4)] is

considerably less noisy and is lower than $h^{(1)}$; moreover, in the slow switching limit $\gamma \ll k_s$, it approaches the expected limit $h = (h_f + h_s)/2$, where h_s and h_f are the entropy rates of Markov random walks with jump rates k_s and k_f. Thus, the compression method detects the non-Markovianity of the random walk reliably even when the simulations have not fully converged.

IV. STATIC AND DYNAMIC DISORDER

The random walker with two internal states is an example of a model with dynamic disorder, where the (mean) lifetime of the random walker on a lattice site can be either long $[1/(2k_s)]$ or short $[1/(2k_f)]$ depending on a dynamical variable (the internal state of the walker). There are also static disorder models, where the lifetime of the walker is determined by a non-time-dependent variable, such as its spatial location. Given the same probabilities of being in the slow and fast states, one often uses these two types of models interchangeably, but the two models are not equivalent [46].

Can the compression method differentiate between the two kinds of disorder? Entropy rate is a measure of information gained about the random walker: Every time a new site is visited, the information gained consists of the direction of the step (1 bit of information for an unbiased walk) and of the time spent on this site. If the walker visits the same site again, less information will be gained in the case of static disorder, as information can already be inferred about the dwell time of this site from the time spent on this site in the previous visit.

To probe this effect, we introduce two additional models. In the static disorder model, each lattice site is randomly assigned to have either long or short average dwell time [Fig. 2(c)]. Transitions that leave the former occur with rate k_s, and transitions that leave the latter occur with rate k_f. The fractions of slow and fast sites are chosen such that, on average, the walker spends half of the time on slow sites and half of the time on fast ones. In the forgetful walker model, every time the walker transitions to a new state, the new state is randomly assigned to have rate k_f or k_s, regardless of its prior identity. Note that both models have the same first-order Markov model as that of the process described by Eqs. (5a) and (5b). If the above argument is correct, the forgetful random walker should have a higher entropy rate than the walker in the model where the site identity is frozen, and indeed, this is the case (Fig. 4). It is also instructive that the forgetful walker and static disorder entropy rates are greater than the entropy rate of the random walker described by Eq. (4) in the slow-switching limit, $k_s, k_f \gg \gamma$. This is because, in this limit, the consecutive steps are highly correlated, with a slow/fast step being likely followed by another slow/fast step, resulting in lower information gained and higher compressibility.

V. SELF-AVOIDING RANDOM WALKS

A random walk that is not allowed to cross its prior path [here, we consider a walk on a square lattice, Fig. 2(d)] offers an interesting example of a non-Markov process with infinite memory. A compression-based approach to the mathematically similar problem of computing the entropy of a polymer has been recently studied by Avinery *et al.* [47]. Since the frequencies with which left, right, up, and down steps are observed in a self-avoiding (SA) walk are the same, the conditional probabilities for making a step in any of these directions are equal to $\frac{1}{4}$, and thus, the first-order Markov model of the SA walk is simply the random walk with the SA condition removed (Supplemental Material, Fig. S4 [37]), with an entropy rate of $h^{(1)} = \log_2 4 = 2$ bits/step. Using transition probabilities estimated from walk trajectories, we find, numerically, a nearly identical first-order entropy rate $h^{(1)} \approx 2.00$ bits/step, and $h^{(2)} \approx 1.58$ for the second-order entropy rate. As with single-file diffusion, these values agree with known theory (Supplemental Material, Fig. S4 [37]). The true entropy rate can be estimated using the known asymptotic behavior of the total number of length L SA walks [48], $\Omega(L) \propto \mu^L$ as $L \to \infty$, with $2.625622 < \mu < 2.679193$ numerically estimated [49] for SA walks on the square lattice. This gives $h \approx \log_2 \mu \approx 1.4$. The corresponding compression-estimated entropy rates (Supplemental Material [37]) $h^{(2)} \approx 1.53$ and $h \approx 1.42$ bits/step are again close to the above values.

VI. CONCLUSIONS

Reconstruction of the underlying models of single-molecule dynamics from experimental observables has received much recent attention (see, e.g., Ref. [50] for a review) and remains a challenge in the field. Here, we showed that compression-derived entropy rate estimates could differentiate between Markov and non-Markov trajectories, as well as between models with dynamic and static disorder, even when the statistical errors or systematic errors introduced by the compression algorithm exceeded the difference between the entropy rate of the true trajectory and the candidate model. Moreover, this approach provides a measure of how long memory is: When the estimated entropy rate $h^{(k)}$ of the kth-order Markov model of the trajectory becomes close to the estimated true value h, the number k quantifies how many previous steps are remembered by the trajectory.

The method described here assumed ergodicity of the underlying dynamics; whether it could be applied to systems that, e.g., display aging phenomena [51] is an open question. Another limitation is that the observed variable is viewed as discrete. Applying the method to the continuous case would require digitizing the observed variable by measuring it with a finite resolution. The resulting entropy rate is known as the *epsilon entropy* $h(\epsilon)$ (the parameter ϵ quantifying the resolution), which can be viewed as an approximation to the Kolmogorov-Sinai entropy [52,53]. To our knowledge, the practical utility in using $h(\epsilon)$ to differentiate between stochastic processes with and without memory has not yet been explored, and it will be the subject of our future work.

ACKNOWLEDGMENTS

We are grateful to Alexander M. Berezhkovskii, Aljaz Godec, Gilad Haran, Hagen Hofmann, Anatoly Kolomeisky, and Benjamin Schuler for many discussions. Financial support from the Robert A. Welch Foundation (Grant No. F-1514 to D.E.M.), the National Science Foundation (Grants

No. CHE 1955552 to D.E.M. and No. IIS 1910274 to E.V.), and Adobe Inc. is gratefully acknowledged.

K.S. analyzed data; K.S., D.E.M., and E.V. designed and performed research and wrote this Letter.

[1] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).

[2] K. Neupane, A. P. Manuel, and M. Woodside, Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes, Nat. Phys. **12**, 700 (2016).

[3] H. A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions, Physica **7**, 284 (1940).

[4] A. B. Kolomeisky, *Motor Proteins and Molecular Motors* (CRC Press, Boca Raton, 2015).

[5] I. M. Sokolov, Models of anomalous diffusion in crowded environments, Soft Matter **8**, 9043 (2012).

[6] R. Metzler, J. H. Jeon, A. G. Cherstvy, and E. Barkai, Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking, Phys. Chem. Chem. Phys. **16**, 24128 (2014).

[7] P. Debnath, W. Min, X. S. Xie, and B. J. Cherayil, Multiple time scale dynamics of distance fluctuations in a semiflexible polymer: A one-dimensional generalized Langevin equation treatment, J. Chem. Phys. **123**, 204903 (2005).

[8] I. Grossman-Haham, G. Rosenblum, T. Namani, and H. Hofmann, Slow domain reconfiguration causes power-law kinetics in a two-state enzyme, Proc. Natl. Acad. Sci. USA **115**, 513 (2018).

[9] X. Hu, L. Hong, M. D. Smith, T. Neusius, X. Cheng, and J. C. Smith, The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time, Nat. Phys. **12**, 171 (2016).

[10] S. C. Kou and X. S. Xie, Generalized Langevin Equation with Fractional Gaussian Noise: Subdiffusion within a Single Protein Molecule, Phys. Rev. Lett. **93**, 180603 (2004).

[11] G. Luo, I. Andricioaei, X. S. Xie, and M. Karplus, Dynamic distance disorder in proteins is caused by trapping, J. Phys. Chem. B **110**, 9363 (2006).

[12] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Observation of a Power-Law Memory Kernel for Fluctuations within a Single Protein Molecule, Phys. Rev. Lett. **94**, 198302 (2005).

[13] H. Yang, G. Luo, P. Karnchanaphanurach, T. M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie, Protein conformational dynamics probed by single-molecule electron transfer, Science **302**, 262 (2003).

[14] G. Muñoz-Gil, G. Volpe, M. A. Garcia-March, E. Aghion, A. Argun, C. B. Hong, T. Bland, S. Bo, J. A. Conejero, N. Firbas *et al.*, Objective comparison of methods to decode anomalous diffusion, Nat. Commun. **12**, 6253 (2021).

[15] H. Y. Aviram, M. Pirchi, Y. Barak, I. Riven, and G. Haran, Two states or not two states: Single-molecule folding studies of protein L, J. Chem. Phys. **148**, 123303 (2018).

[16] J. N. Taylor, M. Pirchi, G. Haran, and T. Komatsuzaki, Deciphering hierarchical features in the energy landscape of adenylate kinase folding/unfolding, J. Chem. Phys. **148**, 123325 (2018).

[17] J. S. t. Bryan, I. Sgouralis, and S. Presse, Inferring effective forces for Langevin dynamics using Gaussian processes, J. Chem. Phys. **152**, 124106 (2020).

[18] Z. Kilic, I. Sgouralis, W. Heo, K. Ishii, T. Tahara, and S. Presse, Extraction of rapid kinetics from smFRET measurements using integrative detectors, Cell Rep. Phys. Sci. **2**, 100409 (2021).

[19] Z. Kilic, I. Sgouralis, and S. Presse, Generalizing HMMs to continuous time for fast kinetics: Hidden Markov jump processes, Biophys. J. **120**, 409 (2021).

[20] I. V. Gopich and A. Szabo, Decoding the pattern of photon colors in single-molecule FRET, J. Phys. Chem. B **113**, 10965 (2009).

[21] A. M. Berezhkovskii and D. E. Makarov, Single-molecule test for Markovianity of the dynamics along a reaction coordinate, J. Phys. Chem. Lett. **9**, 2190 (2018).

[22] R. Satija, A. M. Berezhkovskii, and D. E. Makarov, Broad distributions of transition-path times are fingerprints of multidimensionality of the underlying free energy landscapes, Proc. Natl. Acad. Sci. USA **117**, 27116 (2020).

[23] D. Hartich and A. Godec, Emergent Memory and Kinetic Hysteresis in Strongly Driven Networks, Phys. Rev. X **11**, 041047 (2021).

[24] A. Lapolla and A. Godec, Toolbox for quantifying memory in dynamics along reaction coordinates, Phys. Rev. Res. **3**, L022018 (2021).

[25] A. C. Titman and H. Putter, General tests of the Markov property in multi-state models, Biostatistics **23**, 380 (2022).

[26] J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt, and M. Rief, The complex folding network of single Calmodulin molecules, Science **334**, 512 (2011).

[27] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[28] R. Elber, D. E. Makarov, and H. Orland, *Molecular Kinetics in Condensed Phases: Theory, Simulation, and Analysis* (John Wiley & Sons, Hoboken, 2020).

[29] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. **22**, 403 (1976).

[30] D. E. Makarov, The master equation approach to problems in chemical and biological physics, in *Reviews in Computational Chemistry*, edited by A. L. Parrill and K. B. Lipkowitz, (John Wiley & Sons, Hoboken, 2017), Chap. 6.

[31] C. B. Li, H. Yang, and T. Komatsuzaki, Multiscale complex network of protein conformational fluctuations in single-molecule time series, Proc. Natl. Acad. Sci. USA **105**, 536 (2008).

[32] J. Lee and S. Presse, A derivation of the master equation from path entropy maximization, J. Chem. Phys. **137**, 074103 (2012).

[33] T. Schurmann and P. Grassberger, Entropy estimation of symbol sequences, Chaos **6**, 414 (1996).

[34] Y. Aït-Sahalia, J. Fan, and J. Jiang, Nonparametric tests of the Markov hypothesis in continuous-time models, Ann. Stat. **38**, 3129 (2010).

[35] https://tukaani.org/xz/.

[36] A. D. Wyner and J. Ziv, The sliding-window Lempel-Ziv algorithm is asymptotically optimal, Proc. IEEE **82**, 872 (1994).

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.L012026 for additional data referenced in the main text.

[38] T. E. Harris, Diffusion with "collisions" between particles, J. Appl. Probab. **2**, 323 (1965).

[39] J. L. Lebowitz and J. K. Percus, Kinetic equations and density expansions: Exactly solvable one-dimensional system, Phys. Rev. **155**, 122 (1967).

[40] J. Shin, A. M. Berezhkovskii, and A. B. Kolomeisky, Biased random walk in crowded environment: Breaking uphill/downhill symmetry of transition times, J. Phys. Chem. Lett. **11**, 4530 (2020).

[41] G. Hummer, J. C. Rasaiah, and J. P. Noworyta, Water conduction through the hydrophobic channel of a carbon nanotube, Nature (London) **414**, 188 (2001).

[42] A. Lapolla and 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).

[43] E. B. Stukalin and A. B. Kolomeisky, Transport of single molecules along the periodic parallel lattices with coupling, J. Chem. Phys. **124**, 204901 (2006).

[44] M. L. Mugnai, M. A. Caporizzo, Y. E. Goldman, and D. Thirumalai, Processivity and velocity for motors stepping on periodic tracks, Biophys. J. **118**, 1537 (2020).

[45] A. M. Berezhkovskii and D. E. Makarov, On the forward/backward symmetry of transition path time distributions in nonequilibrium systems, J. Chem. Phys. **151**, 065102 (2019).

[46] S. Burov and E. Barkai, Time Transformation for Random Walks in the Quenched Trap Model, Phys. Rev. Lett. **106**, 140602 (2011).

[47] R. Avinery, M. Kornreich, and R. Beck, Universal and Accessible Entropy Estimation Using a Compression Algorithm, Phys. Rev. Lett. **123**, 178102 (2019).

[48] J. Des Cloizeaux and G. Jannink, *Polymers in Solution: Their Modelling and Structure* (Clarendon Press, Oxford, 1990).

[49] R. Bauerschmidt, H. Duminil-Copin, J. Goodman, and G. Slade, Lectures on self-avoiding walks, in *Probability and Statistical Physics in Two and More Dimensions*, edited by D. D. Ellwood, C. Newman, V. Sidoravicius, and W. Werner (Clay Mathematics Institute, Cambridge, 2012), Vol. 15, pp. 395.

[50] D. E. Makarov, Barrier crossing dynamics from single-molecule measurements, J. Phys. Chem. B **125**, 2467 (2021).

[51] J. Klafter and I. M. Sokolov, *First Steps in Random Walks: From Tools to Applications* (Oxford University Press, Oxford, 2011).

[52] G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani, Predictability: A way to characterize complexity, Phys. Rep. **356**, 367 (2002).

[53] P. Gaspard and X.-J. Wang, Noise, chaos, and (ε,τ) entropy per unit time, Phys. Rep. **235**, 291 (1993).