Memory effects and static disorder reduce information in single-molecule signals

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

[2]Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712

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

Abstract

A key theoretical challenge posed by single-molecule studies is the inverse problem of deducing the underlying molecular dynamics from the time evolution of low-dimensional experimental observables. Toward this goal, a variety of low-dimensional models have been proposed as descriptions of single-molecule signals, including random walks with or without conformational memory and/or with static or dynamics disorder. Differentiating among different models presents a challenge, as many distinct physical scenarios lead to similar experimentally observable behaviors such as anomalous diffusion and nonexponential relaxation. Here we show that information-theory-based analysis of single-molecule time series, inspired by Shannon's work studying the information content of printed English, can differentiate between Markov (memoryless) and non-Markov single-molecule signals and between static and dynamic disorder. In particular, non-Markov time series are more predictable and thus can be compressed and transmitted within shorter messages (i.e. have a lower entropy rate) than appropriately constructed Markov approximations, and we demonstrate that in practice the LZMA compression algorithm reliably differentiates between these entropy rates across several simulated dynamical models.

*makarov@cm.utexas.edu

1. Introduction.

Single-molecule studies that track molecular conformations in real time have opened a new window on biomolecular folding, function of molecular machines, and other cellular phenomena. A critical limitation of such experiments, however, is that they usually track only a few quantities of interest over time, even though the observed behavior is driven by underlying physical processes described by orders of magnitude more degrees of freedom. In other words, the experimental data describes a low-dimensional projection of high-dimensional molecular dynamics. It is known that such projected dynamics are highly complex and often intractable, their common property being that they exhibit memory effects[1]; that is, they are non-Markov processes. Yet to interpret single-molecule data, phenomenological Markovian models are commonly invoked for observed quantities of interest $x(t)$, such as diffusion[2, 3] or evolution of x along a biased random walk on a lattice when x is discrete[4]. Signatures of non-Markovian dynamics such as anomalous diffusion, where the mean-square displacement of x grows nonlinearly with time[5, 6], or nonexponential relaxation have been previously reported (see, e.g., refs.[7-13]), but the challenge, then, is to choose the right dynamical model out of the multitude of possibilities[5]. Data-driven Bayesian inference models of single-molecule time series have enjoyed considerable success in recent years[14-19], but they usually require physical insight in order 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.

Is it possible to tell whether the observed signal 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? For the case where the experimental observable x is a continuous variable, one such Markovianity criterion has been established recently[20] based on analyzing spatial intervals $[a, b]$ and comparing to theoretical inequalities the frequency with which the trajectory $x(t)$ transitions through or loops back when entering the interval, but this approach is inapplicable to processes where the observed experimental quantity takes on discrete values. Other non-Markovianity signatures have been established[21, 22], and an information-theoretic method based on computing the mutual information of the true dynamics and its Markovian model has been proposed[23]. For discrete-state processes, a solution was already described by Shannon in his classic work[24] where he estimated the information content (i.e. the Shannon entropy) of printed English. Specifically, a printed text may be viewed as a discrete-time signal

… $i(t-1), i(t), i(t+1)$ …, where the discrete variable $i \in \{1, …, N\}$ encodes the text characters (N being the alphabet size). In the simplest – and clearly unrealistic – language model, the text is a sequence of statistically independent characters, and its statistical properties are completely specified by the frequencies $P(i), 1 \le i \le N$, with which each character appears in the text. A binary representation of the alphabet requires $\log_2 N$ bits per character, but by encoding frequently occurring characters with fewer bits and rare ones with more bits, one can compress the text, with the theoretical compression limit being given by Shannon's entropy H (in bits per character):

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

A better language model, however, recognizes that certain adjacent pairs of letters are more frequent than others. For example, "t" is more likely to be followed by "h" but not by "w". This observation can be used to compress the text further, by encoding the text as a sequence of character pairs rather than individual letters. To quantify the benefit of this pair encoding, let $T(i \to j)$ be the conditional probability that the i-th character is followed by the j-th character. Then the string of letters produced by this model is a discrete-time (1$^{\text{st}}$ order) Markov process, with T the transition probabilities. The theoretical limit for how much the text described by this model can be compressed is given by the *first-order entropy rate $h^{(1)}$*,

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

measuring the conditional entropy of knowing the next symbol given the previous one, and it can be easily verified that $h^{(1)} \le H$. (The entropy rate here is information-theoretic and is not to be confused with the thermodynamic entropy production rate, which quantifies a non-equilibrium system's heat exchange with its environment.) Shannon then proceeded to generalize this idea to higher-order correlations by computing the conditional probability that each character appears given the two previous characters (a 2$^{\text{nd}}$ order Markov process with entropy rate $h^{(2)}$), and so on. The limit $h = \lim_{k \to \infty} h^{(k)}$ estimates the true information content (in bits per character) of the language.

A key observation is that $h^{(k+1)} \le h^{(k)}$: a higher-order estimate of the entropy rate always yields a lower value than a lower-order estimate. Knowing more preceding characters can only help us to guess the next character more accurately, and thus the information revealed by the next character

is lower. For similar reasons, the entropy rate of a non-Markov process is always bounded above by the entropy rate $h^{(k)}$ of the k-th order Markov approximation of that process.

Although Shannon accomplished the remarkable feat of estimating the entropy of the English language without computers, we now can automate the same task by employing lossless compression algorithms. Given a sample path of a stationary stochastic process, the size (in bits per step) of the compression algorithm output is known to converge, asymptotically in the limit of large path size, to the theoretical entropy rate h of the process[25]. So, for example, to detect whether a piece of text – or a single-molecule trajectory – can be explained by first-order Markovian dynamics, we can on the one hand compress the text and compute the implied entropy rate h of the process from the compressed size; and on the other hand, measure the character-pair transition probabilities $T(i \rightarrow j)$ in the text and compute the first-order entropy rate $h^{(1)}$ using Eq. 2. In principle, $h \approx h^{(1)}$ precisely when the original process is first-order Markov. In practice, of course, we need to worry that the compression algorithm is imperfect: it will overestimate h, and if the error is sufficiently large, the above procedure may fail to differentiate between Markov and non-Markov processes.

In what follows, we explore whether this "compression test" can reveal memory effects when applied to several families of non-Markov models commonly used to describe single-molecule dynamics. We find that this method can indeed differentiate between Markov and non-Markov dynamics. We further show that it can differentiate between static and dynamic disorder and thus narrow down the choice of a model to describe anomalous-diffusion-type phenomena. The rest of the paper is organized as follows: in Section 2 we describe the technical details of the method. Section 3 reports on its applications to several stochastic models. Section 4 concludes with further remarks and a discussion of future applications.

2. Methods.

2.1. Core algorithm. We assume that the experimental time-dependent observable $i(t)$ takes N possible discrete states, and that the process $i(t)$ is stationary. We then present a general algorithm, based on the compression test outlined above, to determine whether the process is likely Markov or non-Markov. Our method can further distinguish between k-th order Markov processes, for modest k, and non-Markov processes with longer-term memory effects. We assume for now a discrete-time

process and extend our approach to determining the Markovianity of continuous-time processes in section 2.2.

First, we estimate the entropy rate of the process. We compute a sample path $i(1), i(2), \ldots, i(L)$ of length L and represent it as an L-character string with an alphabet of size N. We compress the string using LZMA2 lossless encoding, as implemented in the xz library (https://tukaani.org/xz/). The size of the resulting compressed data is measured in bits and divided by L to obtain an estimate $\tilde{h}$ of the entropy rate. In our experiments we chose $L \approx 10^8$; see the supplemental material for full methodological details.

The entropy rate estimate is affected by compressor error: it will differ from the process's "true" entropy rate h by an error $|\tilde{h} - h| = \epsilon_h$ depending on the implementation details of the compression algorithm (which stores metadata within the output file) and the inability of the algorithm to perfectly compress the sample data. Although ϵ_h is guaranteed to converge to zero in the limit of an infinitely long sample path $L \to \infty$, in practice the error remains significant even for paths containing on the order of a billion steps.

Next, we construct an approximate k-th order Markovian description of the process $i(t)$. For any sequence of k consecutive states $S = i(m), i(m+1), \ldots, i(m+k-1)$ encountered in the sampled path and for any i we can estimate the (conditional) transition probability $T(S \to i)$ that S is followed by i from the frequency of the subsequence S, i within the sampled string. We then sample a path $j(1), j(2), \ldots, j(L)$ from the k-th order Markov process defined by these probabilities and compress it as described above, yielding entropy rate estimate $\tilde{h}^{(k)}$. Note that this estimate does not exactly match the true entropy rate of the k-th order Markov process,

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

since again the estimate suffers from compressor error $|\tilde{h}^{(k)} - h^{(k)}| = \epsilon_h^{(k)}$. By comparing $\tilde{h}$ to $\tilde{h}^{(k)}$, rather than to $h^{(k)}$ directly, we exploit favorable cancellation of error: *if* the original random process is indeed k-th order Markov, then ϵ_h and $\epsilon_h^{(k)}$ are drawn from the same distribution, so that $|\tilde{h}^{(k)} - \tilde{h}|$ is small even for values of L for which the individual errors remain large.

Markovianity detection therefore amounts to checking if $\tilde{h} \approx \tilde{h}^{(1)}$. When $\tilde{h} \ll \tilde{h}^{(1)}$ the process is likely non-Markov; higher-order Markov processes with longer memory can be discerned by

repeating the experiment for different k (though the cost of doing so grows exponentially with k). We show below that this procedure does indeed clearly distinguish between Markov, second-order Markov, and non-Markov processes.

2.2. Generalization to continuous-time processes. The core algorithm can also be applied to continuous-time processes. We use kinetic Monte Carlo simulations[26-28] to obtain stochastic trajectories $i(t)$ reporting on the location of the system in the discrete state space as a function of time t. Depending on the problem (and as described below for each individual case in the next Section) we then either sample the continuous-time trajectory at finite (and sufficiently small) time intervals Δt, which results in discrete time series $i(0), i(\Delta t), i(2\Delta t)$ … approximating the original trajectory, or describe the trajectory as a succession of states $i_0, i_1, i_2, …$, where i_k is the first site visited by the continuous-time process after leaving the previous site i_{k-1}. We will refer to the former description as "continuous-time" (even though i is sampled at finite time intervals) and the latter as "discrete time." By construction, self-transitions $i \rightarrow i$ are not allowed in discrete-time processes, since each new state visited by the walk is different from the previous one. In contrast, the continuous-time description allows for self-transitions, where the state i remains the same during the sampling interval Δt, i.e., $i(t + \Delta t) = i(t)$. It is of course possible that the process visits a state other than i and returns to the original state, resulting in $i(t + \Delta t)$ being equal to i, though the probability of such events can be made arbitrarily small by choosing a sufficiently small value of Δt.

An important family of continuous-time signals consists of processes $i(t)$ whose discrete-time description is Markovian but continuous-time description is not. Such processes are known as semi-Markov or renewal processes[29], and they will be further discussed in Section 3.7.
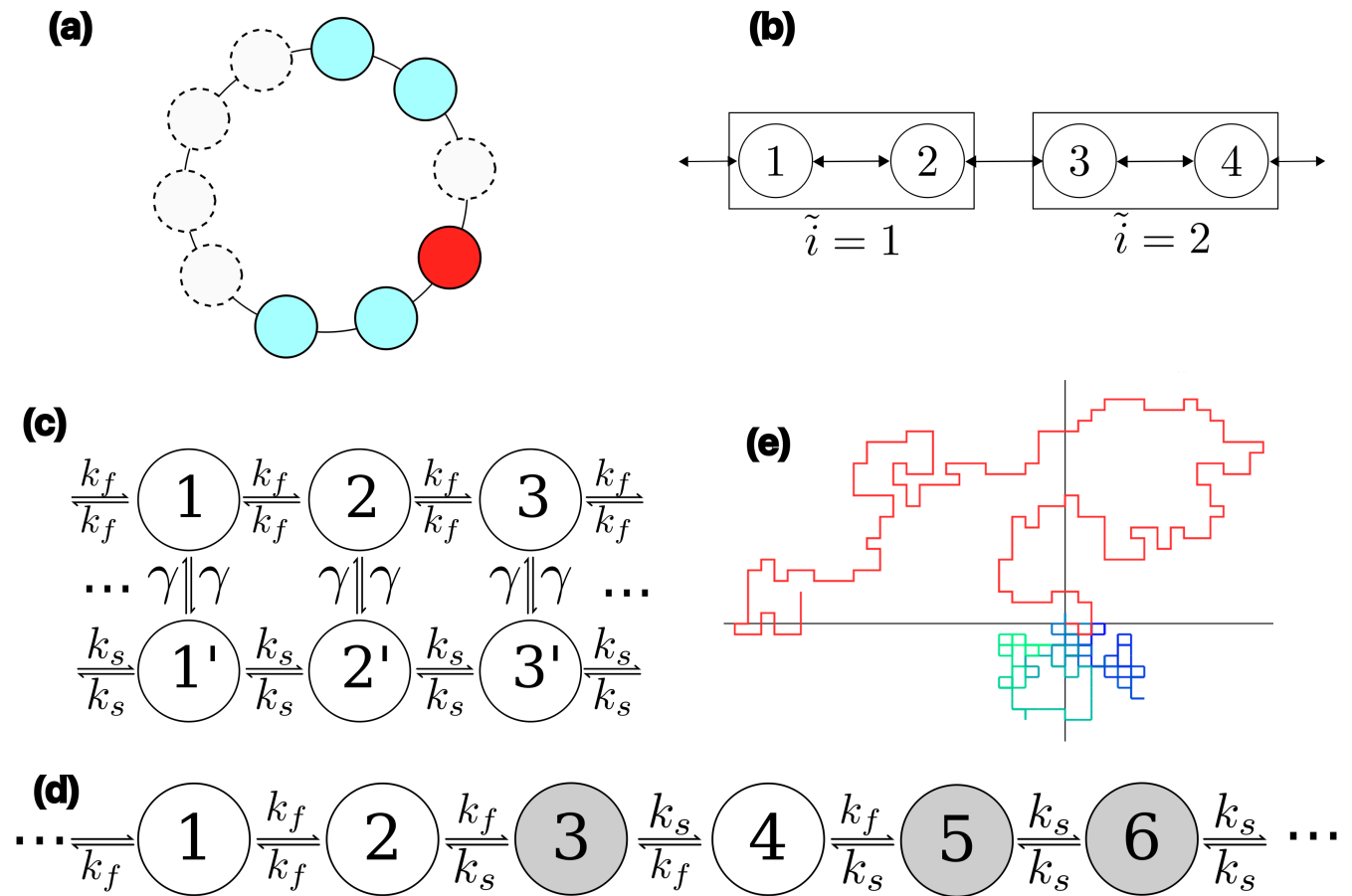
Figure 1: The models under study in this paper. (a) Single-file diffusion on a ring lattice. Each particle (filled circle) can move to an adjacent lattice site only when the latter is vacant (empty circle). The observer monitors the position of a single tagged particle (red). Here the number of lattice sites is $R = 10$ and the number of walkers is $M = 5$. (b) Coarse-grained random walk, in which every n adjacent lattice sites (n=2 here), are grouped into a single state, is a non-Markov process. The labels inside the circles enumerate lattice sites i, while the labels outside the boxes enumerate the coarse grained sates $\tilde{\iota}$. The observer can only tell which boxed state $\tilde{\iota}$ the particle is in. (c) A random walker with two internal states. The transition rates between pairs of states (1,2), (2,3), etc. have a higher value k_f than those between pairs (1',2'), (2',3'), k_s, and the walker can switch between internal states (i.e. between states n and n') with a rate γ. Thus the random walk is fast along the track formed by states 1,2,3,…, and slow along 1',2',3'….. The observer, however, cannot differentiate 1 from 1', 2 from 2', etc., and so the resulting observed process is non-Markovian. (d) 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 (e) Self-avoiding random walk on a 2D lattice. Shown is a trace of a self-avoiding walk (red) and a non-self-avoiding walk (blue-green) on a square lattice.

3. Results.

3.1. Single-file diffusion, Simple Case. A classic example of a random walk with (generally) long memory is single-file diffusion[30, 31] (Fig 1a). This model has applications, for example, as the prototype

of diffusion in crowded environment of a biological cell[5, 32], passage of multiple solute particles across a biological channel[33], and non-Markovian barrier crossing kinetics[34]. 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 $i(t)$ on the ring of a single tagged particle (shown in red in Fig. 1a), where the discrete time t enumerates successive steps of single-file diffusion. It is known that the motion of the tagged particle generally exhibits anomalous diffusion and thus $i(t)$ is not a Markov process.
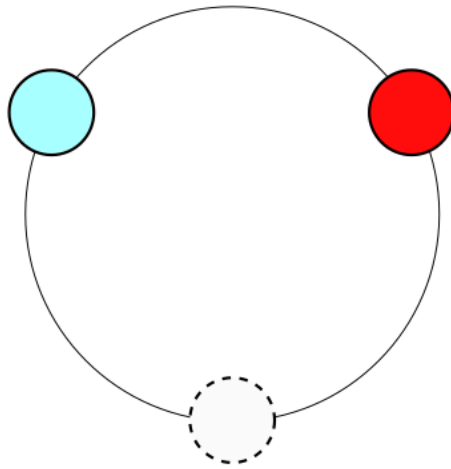
Figure 2. Single-file diffusion with $M = 2$ walkers and $R = 3$ sites on a ring, where one tagged walker (red) is being observed. The filled circles represent occupied sites, and the open circle depicts the vacancy (i.e. unoccupied site). At every timestep, a random walker is chosen to move, and moves into the vacancy, leaving its old site vacant.

It is instructive to consider the simple, analytically tractable, case with $R = 3$ and $M = 2$ illustrated in Figure 2. The entropy rate $h^{(1)}$ for the tagged particle using the 1$^{\text{st}}$ order Markov process approximation can be calculated analytically: at any step, one of the two walkers moves, and thus the probability that the tagged particle moves is ½. When it does, it moves either clockwise or counterclockwise with the same probability, and thus, at any step we have 3 outcomes: no move $\left(T(i \rightarrow i) = \frac{1}{2}\right)$, move clockwise $\left(T(i \rightarrow i + 1) = \frac{1}{4}\right)$, and move counterclockwise $\left(T(i \rightarrow i - 1) = \frac{1}{4}\right)$. Using Eq. 2, the entropy rate based on these transition probabilities is $-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{4}\log_2\frac{1}{4} = $ 1.5 bits/step.

The entropy rate $h^{(2)}$ of the 2nd order Markov process approximation can also be found analytically via a change of variable. Specifically, consider the process $\ell(t) = i(t) - i(t-1)$. This process may be viewed as a "step" representation of the particle's trajectory, as it informs about the direction and length of the step taken. Since there is a one-to-one correspondence between $i(t)$ and $\ell(t)$ assuming that the starting point $i(0)$ is known, the two processes have the same entropy rate. Suppose now that the tagged particle takes a step in the clockwise direction, $\ell = +1$. It easy to see (Fig. 2) that the next step in the same direction would be into a site that is occupied, so $T_\ell(+1 \to +1) = 0$, where we have written T_ℓ to emphasize that these are the transition probabilities of $\ell(t)$ instead of $i(t)$. The tagged and untagged particles have the same probability to move, and so we find $T_\ell(+1 \to -1) = \frac{1}{2}$, $T_\ell(+1 \to 0) = \frac{1}{2}$. By symmetry, we also find $T_\ell(-1 \to -1) = 0$, $T_\ell(-1 \to +1) = \frac{1}{2}$, $T_\ell(-1 \to 0) = \frac{1}{2}$. Using similar arguments, one can show that $T_\ell(0 \to \pm 1) = \frac{1}{4}$, $T_\ell(0 \to 0) = \frac{1}{2}$. The unconditional step probabilities are evaluated in a similar manner, $P_\ell(\pm 1) = \frac{1}{4}$, $P_\ell(0) = \frac{1}{2}$. Using Eq. 3 we now find $h_\ell^{(1)} = 1.25$ bits/step. Again, the subscript ℓ indicates that we are considering the process $\ell(t)$ rather than $i(t)$. Since $i(t)$ is recoverable from $\ell(t)$, and $\ell(t)$ can only depend on $i(t)$ and $i(t-1)$, $h_\ell^{(1)} = h^{(2)}$, and hence $h^{(2)}$ = 1.25 bits/step.

Note that we have derived the 2nd-order Markov entropy rate of the process $i(t)$ by computing the 1st-order entropy rate of a different representation $\ell(t)$ of the same physical process. This highlights the fact that different representations of the same system may have different Markov orders. In our examples below that take place on a structured lattice, we will continue to make use of the "step" representation $\ell(t) = i(t) - i(t-1)$ when convenient; the entropy rates for the two representations are related by $h_\ell^{(k)} = h_i^{(k+1)}$. We will often simplify notation and use $h^{(k)} \equiv h_i^{(k)}$ to denote the entropy rate of the process in the original "lattice site" representation (in accord with the usual convention in physics that a simple random walk depending only on the current site is regarded as a 1st order Markov process), but will use a subscript (e.g. i or ℓ) when we want to emphasize which representation is used to calculate the order.

Although $h^{(2)}$ provides an improved estimate of the single-file diffusion entropy rate, it is still an overestimate. The true entropy rate is h = 1 bit/step, as can be seen from the following argument: there is a one-to-one correspondence between sample paths $i(t)$ of the tagged particle, and sample

paths $v(t)$ of the vacant site (dashed circle in Fig. 2). But the latter is a Markovian, unbiased random walk (since at every step it exchanges places with one of its two occupied neighbors), and so its entropy rate is exactly $h = 1$ bit/step. Note that this example again shows how the same random process can be Markov in one representation (motion of the vacancy) and non-Markov in a different one (motion of the tagged particle).

Applying compression to a trace with $L = 10^9$ steps gives the estimates $\tilde{h}_i^{(1)} = 1.62$ bits/step, $\tilde{h}_i^{(2)} = 1.35$ bits/step, and $\tilde{h} = 1.08$ bits/step. The discrepancy between h and $\tilde{h}$ stems from the compressor error $\epsilon_h^{(k)}$ vanishing very slowly as a function of L. If we subtract a constant $\Delta h \approx 0.10$ from all three measurements, we find that $\tilde{h}_i^{(1)} = 1.52$ bits/step, $\tilde{h}_i^{(2)} = 1.25$ bits/step, and $\tilde{h}_i = 0.98$ bits/step, which is almost perfect agreement with the theoretical values.

3.2 Single-File Diffusion, General Case. For larger M and R, we use the lattice-site representation to obtain compression-based entropy-rate estimates $\tilde{h}^{(k)}$ for several values of k, and compare them to the compression-based estimate $\tilde{h}$ of the true entropy rate, as well as to the entropy rates $h^{(k)}$ (Eqs. 2-3) of the Markov approximations in the absence of any compressor error.

Results are shown in Figure 3a for $M = 7$ walkers and for different numbers of sites R. Here, as in the $M = 2$, $R = 3$ case, we observe a discrepancy between $h^{(k)}$ and $\tilde{h}^{(k)}$. As discussed in Section 2, it is most meaningful to compare the $\tilde{h}^{(k)}$ against each other and against $\tilde{h}$.

When $M \approx R$, we further observe that the true entropy rate $\tilde{h}$ is significantly lower than its finite-order estimates $\tilde{h}^{(k)}$, $k = 1,2,3$, indicating strongly non-Markovian character of single-file diffusion. Since $\tilde{h} < \tilde{h}^{(k)}$, the memory of the previous k states is insufficient to construct an adequate description of the process. 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. Consistent with this observation, the true entropy rate estimate $\tilde{h}$ and its k-order Markovian estimates $\tilde{h}^{(k)}$ converge to the same value as R increases. We also note that, as expected, the estimated entropy rates decrease with increasing order k, that is, $\tilde{h}^{(1)} > \tilde{h}^{(2)} > \tilde{h}^{(3)}$.
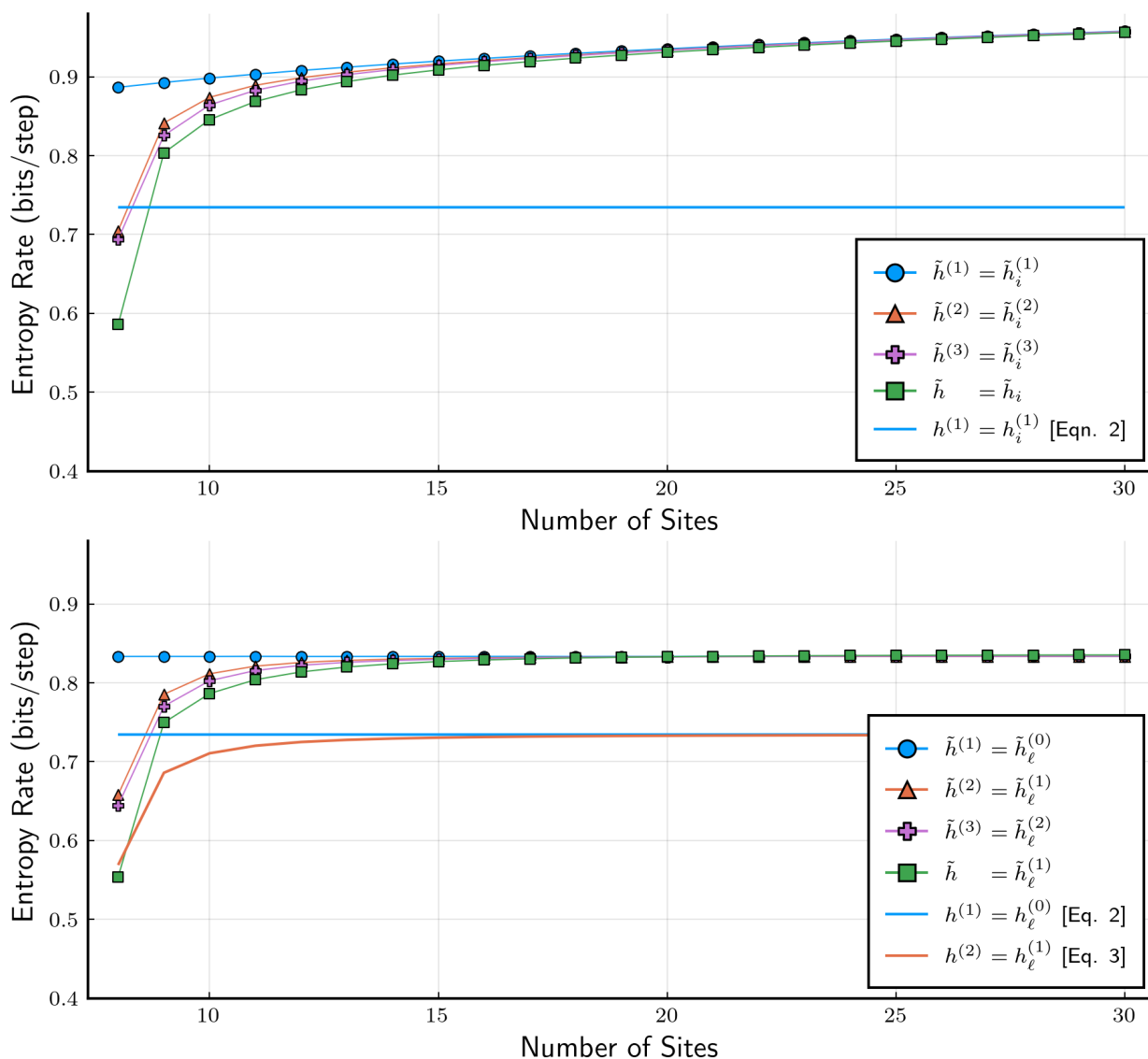
Figure 3. Entropy rates of a tracer particle in a single-file diffusion setup on a ring lattice shown as a function of the number R of lattice sites for a fixed number of random walkers, $M = 7$. Quantities with a tilde (e.g. $\tilde{h}_i^{(k)}$) were computed using compression, while quantities without a tilde (e.g. $h_i^{(k)}$) were obtained using Eqs 2 or 3 using numerically estimated transition probabilities. When the number of open sites $R - M$ is low, the system behaves in a very non-Markov manner. Once the number of sites greatly exceeds the number of particles, the system's behavior approaches Markov.(a) Entropy rates computed using the site-based $i(t)$ representation. (b) Entropy rates computed using the step $\ell(t)$ representation.

We now compare the entropy-rate estimates based on the site representation $i(t)$ and on the step representation $\ell(t)$. In theory, both processes should have the same entropy rate h, and,

moreover, we have $h_\ell^{(k)} = h_i^{(k+1)}$. The compression-based estimates of these quantities, which are greater than their true values, are however not guaranteed to be the same, and indeed they are not (compare Fig. 3b with Fig. 3a). Moreover, compression-derived entropy rates $\tilde{h}_\ell^{(k)}$ are significantly greater than their counterparts $h_\ell^{(k)}$ obtained from Eqs. 2-3 using numerically estimated transition probabilities (Fig. 3b). Importantly, however, the *relative* orderings between $\tilde{h}^{(k)}$ are preserved. Thus examination of either Figure 3a and Figure 3b unambiguously leads to the representation-independent conclusions (1) that single-file diffusion is a process with long memory that cannot be described by k-order Markov process with $k \leq 3$ when most of the lattice sites are occupied ($M \approx R$) and (2) that the process becomes increasingly close to a Markov process as R grows much larger than M.

3.3. Memory in coarse-grained random walks. A fundamental source of dynamical memory is coarse-graining[1]. An experiment cannot resolve each of the individual microscopic states of the molecule, so multiple microscopic states are effectively lumped into collective observable states. To illustrate this effect, we use a simple toy model in which the true dynamics are an unbiased discrete-time random walk along a one-dimensional lattice, whose sites are enumerated by i. Suppose that the spatial resolution of the experiment only allows one to resolve a length of n adjacent sites. The observations thus form a random process on groupings of n adjacent sites (Fig. 1b). For example, sites $i = 1, 2, \ldots n$ correspond to a collective state $\tilde{\imath} = 1$, sites $i = n+1, n+2, \ldots 2n$ to $\tilde{\imath} = 2$, etc. The new variable $\tilde{\imath}$ also undergoes an unbiased (in the sense that the unconditional probabilities of making steps left and right are the same) discrete-time random walk, but this walk is no longer Markovian: as will be seen below, the conditional probability of making a step left or right depends on the direction of the preceding step.

We cannot directly apply our method to compute $h_i^{(k)}$, as we did in Section 3.2, since (as stated in Section 2) our method requires a random process on finitely-many states. However, we can once again define $\ell(t) = \tilde{\imath}(t) - \tilde{\imath}(t-1)$, which has the same entropy rate as $\tilde{\imath}(t)$, and examine $h_\ell^{(k-1)}$ to determine $h_i^{(k)}$. Given that the probabilities of stepping left $\tilde{\imath} \to \tilde{\imath} - 1$ or right $\tilde{\imath} \to \tilde{\imath} + 1$ are the same, we have $P_\ell(-1) = P_\ell(+1) = 1/2$, the same as for the original random walk. Therefore, the first-order Markov approximation of $\ell(t)$ has entropy rate of $h^{(1)} = \log_2 2 = 1$ bit per step (Eq. 2).

Intuitively, at every step one needs to specify whether the walker moves left or right, which occur with equal probability. The coarse-grained walk, however, is not Markovian because the direction of each step is correlated with the direction of the previous step. For example, for $n = 2$, one can show that the conditional probability of making the next step in the same direction as the previous step is $1/3$ while the probability of making a step in the opposite direction is $2/3$. In other words, we have

$$T_\ell(+1 \to +1) = T_\ell(-1 \to -1) = \frac{1}{3}$$

$$T_\ell(+1 \to -1) = T_\ell(-1 \to +1) = \frac{2}{3}.$$

Moreover, it is easy to see that the direction of a step only depends on that of the previous step and not on earlier steps. In other words, the 1st -order Markov model in the step representation (2nd-order in the site representation) is exact for our coarse-grained random walk.

Using Eq. 2 and the fact that the (unconditional) probabilities to make a step in either direction are $P_\ell(\pm 1) = 1/2$, we find an entropy rate $h^{(2)}(n = 2) = h_\ell^{(1)}(n = 2) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} \approx 0.918$ bits/step. In contrast, when we applied the compression method to a synthetic 2nd order Markov walk generated according to the above transition probabilities (using $L = 10^8$ steps) we obtained a value of $\tilde{h}^{(2)} = \tilde{h}_\ell^{(1)} \approx 0.996$ bits/step. Once again, the compressor overestimates the true entropy rate— in fact, applying the compressor to the original, ungrouped random walk (which has an entropy of exactly 1 bit/step) yields a result of $\tilde{h} \approx 1.074$ bits per step, a value of $\epsilon_h = 0.074$ bits/step higher than expected. However, $\tilde{h}^{(2)}(n = 2)$ and $\tilde{h}(n = 2)$ show excellent agreement, and this comparison reveals that the grouped random walk is second-order Markov (see Fig. 4a).

More generally, for coarse states containing n adjacent sites, one can show that

$$T_\ell(+1 \to +1) = T_\ell(-1 \to -1) = \frac{1}{n+1},$$

$$T_\ell(+1 \to -1) = T_\ell(-1 \to +1) = \frac{n}{n+1}.$$

Intuitively, a walker entering a group of n states from the left is much more likely to escape it back to the left than to cross this group and escape to the right. As a result, the entropy rate of the second-order Markov approximation is

$$h_i^{(2)}(n) = -\frac{1}{n+1}\log_2\frac{1}{n+1} - \frac{n}{n+1}\log_2\frac{n}{n+1}. \qquad (4)$$

Figure 4 compares the compression-derived entropy rate $\tilde{h}$ with the Markovian approximations calculated in two ways: (1) using the exact theoretical values known in this case (i.e., $h^{(1)}(n) = 1$ bit/step and $h^{(2)}(n)$ given by Eq. 4 and (2) using our compression method. As with the single-file diffusion system, the compression algorithm does not attain the theoretical compression ratio and overestimates the entropy rate in both cases, but comparison of $\tilde{h}$ and $\tilde{h}^{(k)}$ reveals that the system is second-order Markov. Moreover, one observes from Figure 4a that $\tilde{h}^{(2)}(n)$ differs from $h^{(2)}(n)$ by a nearly constant offset. Subtracting this offset from $\tilde{h}(n)$ results in an estimate for the entropy of the coarse-grained random walk that is virtually indistinguishable from the true value (Fig. 4b).
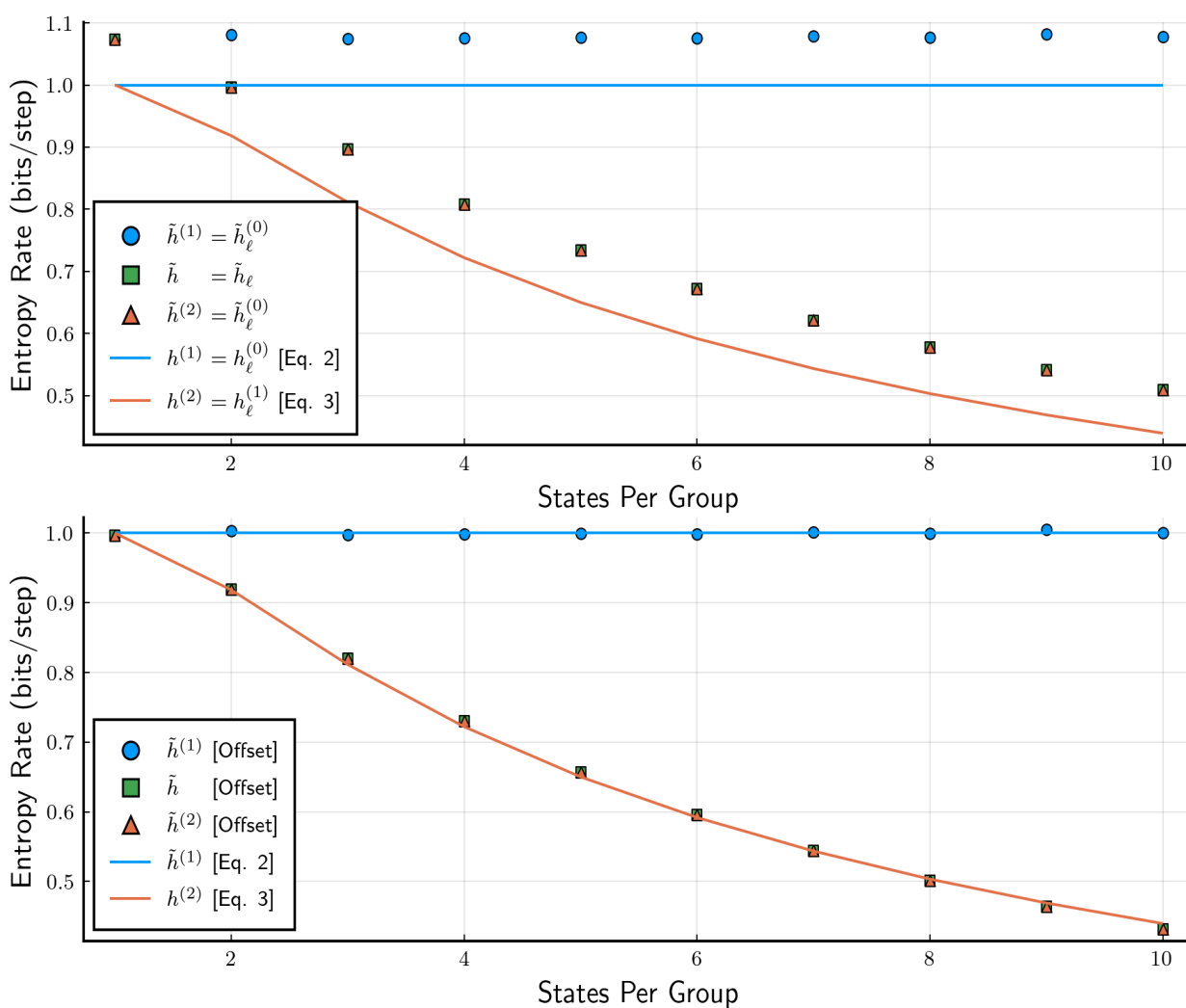
Figure 4. (a) Entropy rate (per step) $\tilde{h}$ of a coarse-grained random walk as a function of the number of states n forming a single coarse-grained state. This entropy rate is compared with the compression-derived entropy rates $\tilde{h}^{(1)}$ and $\tilde{h}^{(2)}$ of the 1st and 2nd order approximating Markov processes, as well as with their theoretical values $h^{(1)} = 1$ bits/step and $h^{(2)}$ given by Eq. 4. (b) The same plots after subtracting the offset Δh from the compressor-derived entropy rates.

3.4. Random walker with internal states. Another example of coarse-graining is the model shown in Fig. 1c. Unlike the previous case, we are now considering a continuous-time process, where a random walker can be in one of the two internal states; in one, its diffusion is fast (quantified by a jump rate k_f) and in the other it is slow (jump rate k_s). The walker can switch between the two internal states stochastically, with a switching rate γ. The kinetic scheme (Fig. 1c) thus consists of a "fast" and a "slow" track, with the walker randomly switching between the two. According to this scheme, the time evolution of the probabilities $P(i, t)$ and $P(i', t)$ to occupy sites i and i' obey the continuous-time master equation,

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

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

which describes a Markov process. The states i and i', however, appear indistinguishable to the observer and are thus lumped to a single coarse state characterized by the apparent position $i = 1,2,...$ of the walker along the track. Unless $k_f = k_s$, the time evolution of the probability $P(i,t) + P(i',t)$ to be at position i cannot be described by a master equation like Eq. 5a or 5b: the observed random walk is non-Markovian.

Models of this type have been used, e.g., to describe the dynamics of biomolecular motors traveling along their tracks[4, 35, 36], even though in our case the walker has no directional bias and thus does not undergo unidirectional motion characteristic of molecular motors.

The stochastic time evolution corresponding to Eq. 5 was simulated using the kinetic Monte Carlo method (see, e.g., refs.[26-28]) and the entropy rate was estimated using the "continuous time" approach described in Section 2.2, using a time step Δt chosen to satisfy the inequalities $k_f \Delta t \ll 1, \gamma \Delta t \ll 1$, to guarantee that the probability of the walker changing state more than once during Δt is

negligibly small. As with the coarse-grained random walk, we computed entropy rate using the step representation (and make use of $\tilde{h}^{(k)} = \tilde{h}_\ell^{(k-1)}$) to avoid dealing with infinite sites.

The compression-derived entropy rate of this walk is shown in Figure 5a as a function of the switching rate γ for the case $k_s = 0.1 k_f$, and compared to the entropy rates of the first- and second-order Markov approximations. For all values of the switching rate γ, the true entropy rate of the walker is always lower than that of the two reference Markov processes (whose entropy rates are indistinguishable), indicating non-Markovianity of the underlying process. Despite the significant statistical noise observed in Fig. 5a, particularly at slow switching rates, the difference between the true entropy rate and that of the first-order reference Markov process (shown in Fig. 5b) is always negative and is less noisy than the absolute entropy rates shown in Fig. 5a – thus the compression method detects the non-Markovianity of the underlying dynamics reliably even when the simulations have not fully converged.

The dependence of the entropy rate on the switching rate also agrees with what one expects on physical grounds. Specifically, in the limit of high switching rates, $\gamma \gg k_f$, the process becomes effectively Markovian, with the effective transition rate coefficients between adjacent positions on the track equal to the average rate, $k_{\text{eff}} = (k_s + k_f)/2$, with a corresponding entropy rate of h_{eff}. Note this Markov process is identical to the 1$^{\text{st}}$ order reference Markov process, which is the same regardless of the switching rate. Consistent with this observation, we observe that (1) the compression-derived value $\tilde{h}^{(1)}$ is independent of the switching rate (aside from statistical noise) and (2) the true entropy rate approaches the entropy rate $h^{(1)}$ of the reference Markov process at large values of γ (Fig. 5a,b).

Consider now the opposite case of slow switching rate, $\gamma \ll k_s$. In this case we expect that the trajectory of the random walker will consist of long segments where the walker stays on the fast (or slow) track, jumping between neighboring lattice sites with a rate of k_f (or k_s). Since the fractions of time spent on each track are $\frac{1}{2}$, we expect the entropy rate to approach an average value of $h = (h_f + h_s)/2$, where h_f (h_s) is the entropy rate of a Markovian random walk along the fast (slow) track (i.e. a 1D random walk with a transition rate k_f (k_s). Indeed, this is the behavior observed in Fig. 5b,

with a compression-derived estimate of $(h_f + h_s)/2$ agreeing with the compression-derived true entropy rate $\tilde{h}$ at low switching rates.
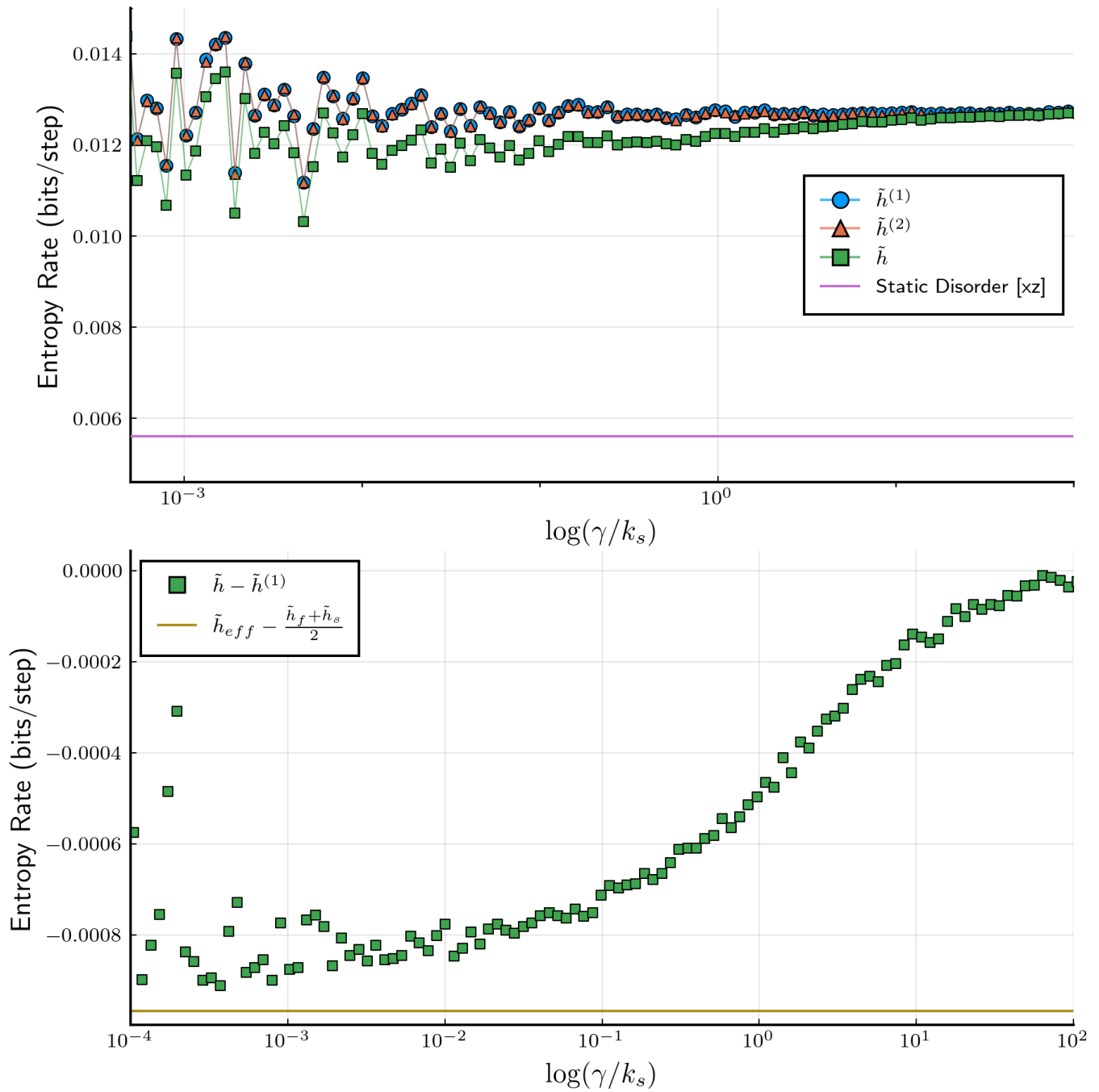
Figure 5. (a) Entropy rate for the random walk with two internal states (Fig. 1c), plotted as a function of the switching rate γ between the fast and slow tracks and compared with the reference 1st and 2nd order Markov processes. Solid line represents the compression-derived entropy rate for the random walk with static disorder (Fig. 1d). The kinetic parameters are $k_s = 0.1$, $k_f = 1.0$. (b) The difference $h - h^{(1)}$ between the compression-derived true entropy rate and that of the Markovian approximation plotted as a function of the switching rate. At low switching rates, $\gamma \to 0$, the entropy rate (measured relative to its $\gamma \to \infty$ limit) is seen to approach the expected value equal to the mean of the (compression-derived) entropy rates of two Markovian processes, the slow one (with the jump rate k_s) and the fast one (jump rate k_f).

3.5. Compression differentiates between static and dynamic disorder.

The two-track model considered in Section 3.4 is an example of a model with dynamical 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) that is itself undergoing a stochastic process switching between the slow and fast states. By construction, the probabilities of finding the walker in the slow and fast states are equal. It is instructive to consider a seemingly similar model with static (quenched) disorder. In this model of a 1D random walk (Fig. 1d), each lattice site is randomly assigned to have either long or short average dwell time. Transitions that leave the former occur with rate k_s, and transitions that leave the latter occur with rate k_f. Since in both models the random walker visits the "slow" and the "fast" sites with equal probabilities, one often uses these two types of models interchangeably, but the two models are not equivalent[37]. More generally, various anomalous-diffusion-type phenomena are often interpreted in terms of random walks on a network of "traps", with a certain distribution of the trap energies leading to a distribution of trapping times[5]. One often models such a random walk as a renewal process (see Section 3.7 for further discussion of such processes), in which the times spent on each trap are statistically independent, but such an assumption is not correct when the trap properties are "frozen" in time[5, 37].

While differentiating between static and dynamic disorder in anomalous diffusion phenomena is often a challenge, the compression-based method explored here achieves this goal rather straightforwardly, at least for the model system considered. Specifically, the entropy rate corresponding to the process with static disorder is always lower than that for dynamic disorder, as can be observed in Fig. 5a. It is easy to understand why: every time the random walker visits a new lattice

site, the information gained consists of the site identity 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 this site's dwell time from the time spent on this site in the previous visit.

3.6 Self-avoiding random walks. A random walk that is not allowed to cross its prior path offers an interesting example of a non-Markov process with infinite (or, more precisely, as long as the walk itself) memory. Such random walks have been studied extensively in the context of polymer physics, where they serve as models of polymer backbones. Moreover, a compression-based approach to the mathematically equivalent problem of computing the entropy of a polymer has been recently studied by Avinery, Kornreich and Beck[38].

Here we consider self-avoiding 2D walks on a square lattice (Fig. 1e). Because of the symmetry between up/down and left/right directions, the probabilities of making step in one of the 4 possible directions are equal, $T(i \rightarrow i') = 1/4$, for any of the 4 sites i' adjacent to i. Thus the 1st order Markov model of the process is simply the random walk with the self-avoidance condition removed. For such a walk the number of possible step directions at each step is $4 = 2^2$, and thus its entropy rate is $h^{(1)} = \log_2 4 = 2$ bits per step.

The 2nd order Markov model of the self-avoiding walk clearly must account for the fact that any step cannot be the reverse of the previous step. Since there are only 3 possible steps left, a naïve estimate of the entropy rate would be $h^{(2)} = \log_2 3 \approx 1.585$ bits/step, assuming that all three possible directions are equally probable. A more precise estimate for this value, obtained numerically from Eq. 3 using transition probabilities measured from simulation data, is slightly lower, $h^{(2)} \approx 1.578$. Note that this model is locally self-avoiding, since by demanding that a step cannot be reversed, we automatically ensure self-avoidance for 3 steps in a row. Thus, only long-range correlations (longer than 3 steps) may account for the difference between $h^{(2)}$ for the 2nd order Markov model and the entropy rate of the true self-avoiding walk. As we show below, the compression-based method is still capable of differentiating between these two models.

Figure 6 shows the entropy rates estimated numerically as functions of the walk length. To apply our method we sampled 250,000 length-N self-avoiding random walks using Sokal's pivoting

algorithm[39]. For sufficiently large values of N, the entropy rate becomes independent of the length of the walk. This property is an established fact[40] but not immediately obvious, as the memory of the walk is as long as the length of the walk itself. Specifically, the total number of self-avoiding random walks asymptotically grows as

$$\Omega(N) \propto \mu^N \tag{6}$$

as $N \to \infty$, where numerical estimates exist[41] for lower and upper bounds for μ, $2.625622 < \mu < 2.679193$. We can interpret μ as the average number of possible step directions available at each step of the walk, and thus its entropy rate is

$$h \approx \log_2 \mu \approx 1.4, \tag{7}$$

which we take as the "exact" entropy rate in this case. As seen in Figure 6, the compression method differentiates among the two reference Markov approximations and true dynamics, indicating that the actual observed process has more memory than that of one previous step (in fact of 3 previous steps given that the self-avoidance is guaranteed for any 3 consecutive steps in this model).

While the compression-derived estimate of the connectivity parameter, $\mu = 2^{\tilde{h}} \approx 3$, exceeds the best currently-known upper bound on the theoretical value, this estimate can be improved by accounting for the compression error: for the second-order Markov reference process we can compute $\epsilon_h^{(2)} = \tilde{h}^{(2)} - h^{(2)} \approx 0.1 \text{bits}$ from Eq. 3 Assuming that $\epsilon_h \approx \epsilon_h^{(2)}$ and applying this same correction to estimate $h \approx \tilde{h} - 0.1 \approx 1.5$ bits/step and $\mu \approx 2.8$, still somewhat greater than the known bound.
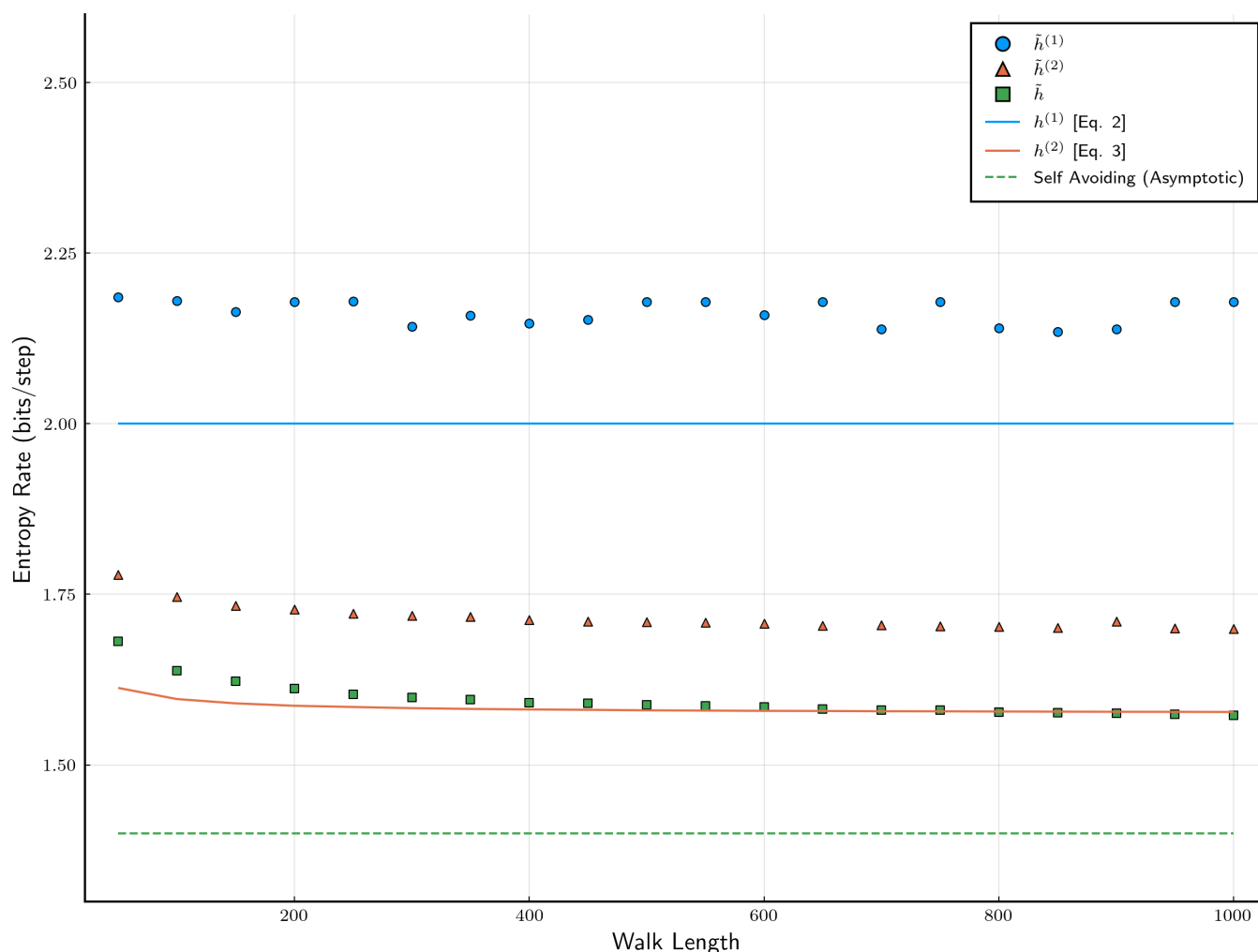
Figure 6. Entropy rates for self-avoiding walks of various lengths, compared with the with the entropy rates of the corresponding 1st order and 2nd order Markov models. Literature value for the entropy of the self-avoiding random walk[40] is shown as a dashed line.

3.7. Continuous-time random walks (CTRW) constitute an important class of non-Markov models used widely to describe anomalous-diffusion-type phenomena[6, 29, 42]. Milestoning, a method for computing long-time molecular dynamics, also uses a CTRW-type representation[43]. CTRW is an example of a semi-Markov or renewal process. A continuous-time random walker makes a step from a current site i to new sites j according to the conditional probabilities $T(i \rightarrow j)$, with the sojourn time at i before making this step drawn from a specified probability density $\rho(t)$ (the case where this probability density is itself state dependent is a straightforward generalization). Thus the random walk can be

viewed as a combination of two processes: the first one is the sequence of sites visited $i_1, i_2, \dots, i_n, \dots$ and the other is the sequence of sojourn times $t_1, t_2, \dots, t_n, \dots$ spent on each site. The first of the two is a discrete-time Markov process. The total entropy rate is the sum of the entropy rate for the discrete-time Markov process and the entropy rate associated with the sequence of the sojourn times. We will now focus on the latter. As the sojourn times are statistically independent, the entropy rate is just the entropy of the distribution $\rho(t)$. More precisely, if we discretize the time by binning the temporal data into sufficiently fine time bins of equal size Δt, then the entropy rate is given by

$$\eta[\rho] \approx -\int_0^\infty dt\rho(t)(\log_2 \rho(t) \quad -\log_2 \Delta t), \tag{8}$$

where the last term accounts for the dependence of the entropy on the resolution with which the time is measured.

Using the method of Lagrange multipliers, one finds that of all the distributions $\rho(t)$ with a fixed first moment, $\langle t \rangle = \int_0^\infty t\rho(t)dt$, the one that maximizes $\eta[\rho]$ is the exponential distribution,

$$\rho(t) = \langle t \rangle^{-1} e^{-t/\langle t \rangle}, \tag{9}$$

for which

$$\eta[\rho] = \frac{1}{\ln 2} - \log_2 \frac{\Delta t}{\langle t \rangle}. \tag{10}$$

This distribution corresponds to a Poisson process, which is Markovian. This illustrates the general observation that Markovian processes maximize the entropy rate, provided that the average frequency of the transitions is fixed.

4. Discussion

Reconstruction of the underlying models of single-molecule dynamics from experimental observables has received much recent attention (see, e.g., ref.[44] for a review) and remains a challenge in the field. The problem is further confounded by experimental constraints such as finite temporal and spatial resolution. In all of the examples studied here, compression-derived entropy rate estimates could differentiate between simulated Markov and non-Markov time series, as well as between models with dynamical 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 signal and its

candidate model (e.g. Markov description). We note that this method does not require a high data sampling rate that may be unattainable experimentally: indeed, the entropy rate at an arbitrary data sampling rate can be compared with that of the corresponding reference Markov process, the latter being always computable via Eqs. 2-3.

Markovian approximations to the non-Markov systems studied here provide upper bounds on the true entropy rate. In general, one can construct the kth-order Markov reference model of the process and to estimate its entropy rate $h^{(k)}$; if $h^{(k)}$ converges toward the entropy rate h of the experimental signal as k increases sufficiently fast, this result is a practical estimate of the number of previous steps needed to be included to produce an accurate description of the signal. For example, for the coarse-grained random walk with grouped states (Section 3.3) the 1$^{\text{st}}$ order Markov process does not reproduce the "experimental" entropy rate, but the 2$^{\text{nd}}$ order one (which happens to be the exact description in this case) does (Fig. 2). In contrast, the 2$^{\text{nd}}$ order Markov model does not provide a noticeable improvement over the 1$^{\text{st}}$ order one for the random walk from Section 3.6, suggesting long memory (as compared to the sampling time step). In such a case, a different description of the signal is desirable.

This work focused on signals where the observed variable is discrete. Such signals are common in single-molecule fluorescence studies in which the arrival of an individual photon can be viewed as a discrete event[45-48]. In contrast, force spectroscopy studies measure continuous variables such as a molecule's extension or position of a molecular motor along its track[4, 49, 50]. Likewise, single-particle tracking studies measure continuous signals, see, e.g., ref.[51]. Can the compression method be applied to a continuous signal? A naïve answer to this is to digitize 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]. For stochastic signals, $h(\epsilon)$ is known to diverge as $\epsilon \to 0$, and its behavior for certain stochastic models has been studied[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.

References

1.   R. Zwanzig, *Nonequilibrium Statistical Mechanics*. (Oxford University Press, 2001).
2.   K. Neupane, A. P. Manuel and M. Woodside, Nature Physics **12**, 700-703 (2016).
3.   H. A. Kramers, Physica **7**, 284-304 (1940).
4.   A. B. Kolomeisky, *Motor Proteins and Molecular Motors*. (CRC Press, 2015).
5.   I. M. Sokolov, Soft Matter **8**, 9043-9052 (2012).
6.   R. Metzler, J. H. Jeon, A. G. Cherstvy and E. Barkai, Phys Chem Chem Phys **16** (44), 24128-24164 (2014).
7.   P. Debnath, W. Min, X. S. Xie and B. J. Cherayil, J Chem Phys **123** (20), 204903 (2005).
8.   I. Grossman-Haham, G. Rosenblum, T. Namani and H. Hofmann, Proc Natl Acad Sci U S A **115** (3), 513-518 (2018).
9.   X. Hu, L. Hong, M. D. Smith, T. Neusius, X. Cheng and J. C. Smith, Nature Physics **12**, 171-174 (2016).
10.   S. C. Kou and X. S. Xie, Phys Rev Lett **93** (18), 180603 (2004).
11.   G. Luo, I. Andricioaei, X. S. Xie and M. Karplus, J Phys Chem B **110** (19), 9363-9367 (2006).
12.   W. Min, G. Luo, B. J. Cherayil, S. C. Kou and X. S. Xie, Phys Rev Lett **94** (19), 198302 (2005).
13.   H. Yang, G. Luo, P. Karnchanaphanurach, T. M. Louie, I. Rech, S. Cova, L. Xun and X. S. Xie, Science **302** (5643), 262-266 (2003).
14.   H. Y. Aviram, M. Pirchi, Y. Barak, I. Riven and G. Haran, J Chem Phys **148** (12), 123303 (2018).
15.   J. N. Taylor, M. Pirchi, G. Haran and T. Komatsuzaki, J Chem Phys **148** (12), 123325 (2018).
16.   J. S. t. Bryan, I. Sgouralis and S. Presse, J Chem Phys **152** (12), 124106 (2020).
17.   Z. Kilic, I. Sgouralis, W. Heo, K. Ishii, T. Tahara and S. Presse, Cell Rep Phys Sci **2** (5) (2021).
18.   Z. Kilic, I. Sgouralis and S. Presse, Biophys J **120** (3), 409-423 (2021).
19.   I. V. Gopich and A. Szabo, J Phys Chem B **113** (31), 10965-10973 (2009).
20.   A. M. Berezhkovskii and D. E. Makarov, J Phys Chem Lett **9** (9), 2190-2195 (2018).
21.   D. Hartich and A. Godec, (arXiv:2011.04628 . Accepted for publication in Phys. Rev. X, 2021).
22.   R. Satija, A. M. Berezhkovskii and D. E. Makarov, Proc Natl Acad Sci U S A **117** (44), 27116-27123 (2020).
23.   A. Lapolla and A. Godec, Physical Review Research **3** (2), L022018 (2021).
24.   C. E. Shannon, Bell System Technical Journal **January**, 57-64 (1951).
25.   A. D. Wyner and J. Ziv, Proc. IEE **82**, 872-877 (1994).
26.   R. Elber, D. E. Makarov and H. Orland, *Molecular Kinetics in Condense Phases: Theory, Simulation, and Analysis*. (Wiley and Sons, 2020).
27.   D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
28.   D. E. Makarov, in *Reviews in Computational Chemistry*, edited by A. L. Parrill and K. B. Lipkowitz (John Wiley & Sons, 2017), Vol. 30.

29.	J. Klafter and I. M. Sokolov, *First steps in random walks : from tools to applications*. (Oxford University Press, Oxford ; New York, 2011).
30.	T. E. Harris, Journal of Applied Probability **2** (2), 323-338 (1965).
31.	J. L. Lebowitz and J. K. Percus, Physical Review **155** (1), 122-138 (1967).
32.	J. Shin, A. M. Berezhkovskii and A. B. Kolomeisky, J Phys Chem Lett **11** (11), 4530-4535 (2020).
33.	G. Hummer, J. C. Rasaiah and J. P. Noworyta, Nature **414** (6860), 188-190 (2001).
34.	A. Lapolla and A. Godec, J Chem Phys **153** (19), 194104 (2020).
35.	M. L. Mugnai, M. A. Caporizzo, Y. E. Goldman and D. Thirumalai, Biophys J **118** (7), 1537-1551 (2020).
36.	A. M. Berezhkovskii and D. E. Makarov, Journal of Chemical Physics **151**, 065102 (2019).
37.	S. Burov and E. Barkai, Phys Rev Lett **106** (14), 140602 (2011).
38.	R. Avinery, M. Kornreich and R. Beck, Phys Rev Lett **123** (17), 178102 (2019).
39.	A. D. Sokal, in *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, edited by K. Binder (Oxford University Press, New York, 1995).
40.	J. Des Cloizeaux and G. Jannink, *Polymers in solution: their modelling and structure*. (Clarendon Press, Oxford, 1990).
41.	R. Bauerschmidt, H. Duminil-Copin, J. Goodman and G. Slade, in *Probability and Statistical Physics in Two and More Dimensions (Clay Mathematics Proceedings)*, edited by D. D. Ellwood, D. Newman, V. Sidoravicius and W. Werner (Clay Mathematics Institute, 2010), pp. 395-467.
42.	R. Metzler and J. Klafter, Physics Reports **339**, 1-77 (2000).
43.	R. Elber, J Chem Phys **144** (6), 060901 (2016).
44.	D. E. Makarov, J Phys Chem B **125** (10), 2467-2476 (2021).
45.	B. Schuler and W. A. Eaton, Curr Opin Struct Biol **18** (1), 16-26 (2008).
46.	H. S. Chung and I. V. Gopich, Phys Chem Chem Phys **16** (35), 18644-18657 (2014).
47.	H. S. Chung and W. A. Eaton, Curr Opin Struct Biol **48**, 30-39 (2018).
48.	F. Sturzenegger, F. Zosel, E. D. Holmstrom, K. J. Buholzer, D. E. Makarov, D. Nettels and B. Schuler, Nat Commun **9** (1), 4708 (2018).
49.	D. E. Makarov, *Single Molecule Science: Physical Principles and Models*. (CRC Press, Taylor & Francis Group, Boca Raton, 2015).
50.	D. B. Ritchie and M. T. Woodside, Curr Opin Struct Biol **34**, 43-51 (2015).
51.	N. Zijlstra, D. Nettels, R. Satija, D. E. Makarov and B. Schuler, Phys. Rev. Letters **125**, 146001 (2020).
52.	G. Boffetta, M. Cencini, M. Falcioni and A. Vulpiani, Physics Reports **356** (6), 367-474 (2002).
53.	P. Gaspard and X.-J. Wang, Physics Reports **235** (6), 291-343 (1993).