Persistent random motion with maximally correlated fluctuations

Adam Svenkeson[1] and Bruce J. West[2]
[1]*Vehicle Technology Directorate, Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, Maryland 21005, USA*
[2]*Information Science Directorate, Army Research Office, Research Triangle Park, North Carolina 27703, USA*

(Received 10 May 2019; published 15 August 2019)

How often should a random walker change its direction of motion in order to maximize correlation in velocity fluctuations over a finite time interval? We address this optimal diffusion problem in the context of the one-dimensional persistent random walk, where we evaluate the correlation and mutual information in velocity trajectories as a function of the persistence level and the observation time. We find the optimal persistence level corresponds to the average number of direction reversals asymptotically scaling as the square root of the observation time. This square-root scaling law makes the relative growth between the average number of direction reversals and the persistence length invariant with respect to changes in the overall time duration of the random walk.

DOI: 10.1103/PhysRevE.100.022119

I. INTRODUCTION

The persistent random walk is a basic model for correlated diffusion with several applications [1], including the motion of tracer particles in a turbulent flow [2], microscopic dynamics of active matter [3,4], and scattering of waves [5–7]. In the one-dimensional persistent random walk [8,9], the walker's step at any discrete time n will be preferentially taken in the same direction as the previous step, with probability α (Fig. 1). We assume there is no underlying directional bias, which is to say space is homogeneous and isotropic. The persistence level α can be thought of as an inertial control parameter of the diffusion model that characterizes the tendency for a particle (walker) to remain in its current state of motion. If $\alpha = 1/2$, then the walker tosses a fair coin to determine the direction of each step and a simple random walk is recovered. If $0 < \alpha < 1/2$, then the walker tends to reverse its direction of motion when given an opportunity, leading to an antipersistent random walk; whereas if $1/2 < \alpha < 1$, then the direction of the walker changes less frequently in time and the random walk is persistent. The latter is the case discussed herein.

A fundamental result in persistent random-walk theory [2,8], due to a variant of the central limit theorem for correlated random variables, is that ordinary diffusion is recovered in the long-time limit for any persistence level $1/2 < \alpha < 1$. With that long-time limit in mind the particular value of α has little importance. The specific value within that range has no effect on the qualitative nature of walker trajectories; it only sets the timescale at which the transition from ballistic to diffusive motion occurs. The correlation time then determines the magnitude of the diffusion coefficient through the Taylor-Green-Kubo formula [10] (also known as just the Green-Kubo formula in general transport processes). However, if the diffusion process is limited to a finite time, as determined by the total number of steps N, then the persistence level takes on new importance. The essential nature of the correlated random-walk trajectory can be changed from being strongly diffusive to being strongly ballistic by varying α from one half to one. We emphasize that this statement holds for any finite observation time N, even in the asymptotic limit $N \to \infty$.

Traditionally the persistence level is viewed as a fixed property of the diffusion phenomenon of interest, whereas the observation time is a quantity varied by the experimenter. But in some cases the opposite notion may be more appropriate, where there may be a natural time interval over which to observe the process, and the desideratum is to vary, or engineer, the parameters of the system to achieve certain transport behavior during that time interval. This setting falls within the framework of optimal diffusion in which natural choices for the observation time interval may be dictated by the characteristic properties of the system, for example, the system's linear size divided by the average particle velocity [11] or the average lifetime of a walker [12].

A variety of objectives arise when framing optimal diffusion problems. For example, there is the optimal search problem where a random walker searches a physical space for a specified target. An optimal solution is found by minimizing the mean first passage time to reach the target [13,14]. Another example objective is exploration, which can be optimized by maximizing the mean area covered or convex hull of the random walk [15,16]. A third example is fast relaxation to equilibrium, where the optimal solution is that which minimizes the relaxation time [17]. Additionally, some variations of an optimal diffusion problem might involve time-dependent control parameters, like engineered swift equilibration or shortcuts to adiabaticity [18,19].

Outside of these well-founded optimal diffusion problems there are many conceivable objective functions that might be studied without corresponding *a priori* to a clear physical interpretation in terms of the evolution of trajectories or probability density functions. For instance, there is the objective of maximizing the velocity autocorrelation function

$$C(\alpha, N) = \overline{[v_n - \overline{v}(N)][v_{n+1} - \overline{v}(N)]}, \tag{1}$$

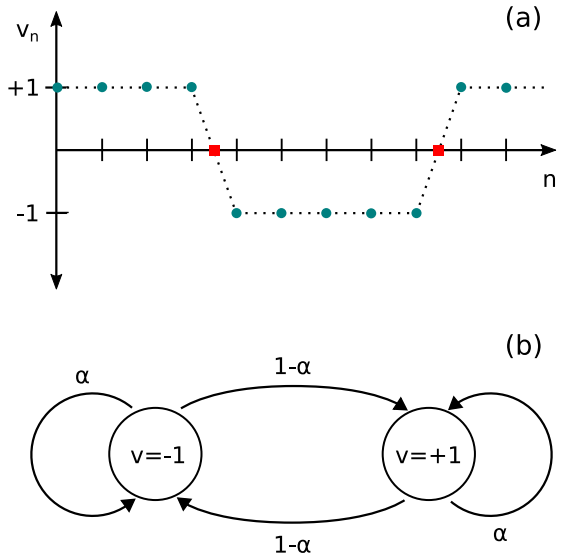

FIG. 1. Illustrations of the persistent random walk. (a) Example velocity trajectory. Green circles indicate the random walker's velocity at each time step. Red squares mark the occurrence of direction-changing events. (b) Markov chain representation of the walker's velocity.

or the mutual information function

$$I(v_n, v_{n+1}; N) = \sum_{v_n} \sum_{v_{n+1}} p(v_n, v_{n+1}) \ln \left[\frac{p(v_n, v_{n+1})}{p(v_n) p(v_{n+1})} \right], \quad (2)$$

which is the kind of optimized diffusion problem we address herein. It is important to note that Eqs. (1) and (2) are defined in this context to be functionals of the random-walk velocity trajectory $[v_0; v_1, \ldots, v_N]$ generated with a given persistence level α, with the overline in Eq. (1) denoting a time average and the probabilities in Eq. (2) evaluated empirically from time averages along single trajectories. These objective functions measure, to some extent, the amount of structure or complexity existing within a process and at least qualitatively indicate that the optimal trajectories containing highly correlated fluctuations should strike a balance between disordered direction-changing events and ordered ballistic motion. Equations (1) and (2) can be approximately understood by considering a generic objective function F in the form

$$F(\alpha, N) = F_1(\alpha) - F_N(\alpha), \quad (3)$$

where F_1 is a measure of local order at the level of one time step, and F_N is a measure of long-range order at the level of N time steps, the longest timescale accessed by the trajectory. F_1 can be interpreted as a reward for correlated motion, and F_N a penalty for rarely changing direction (or, conversely, a reward for random motion). The objective then is to maximize correlation while retaining fluctuations in the velocity trajectory. We seek to quantify this order-disorder balance and build intuition about the physical consequences of maximizing the correlation and mutual information objective functions, Eqs. (1) and (2). We point out that it is necessary to consider a finite observation time interval when formulating the optimization problem. Maximizing the asymptotic mutual information ($N = \infty$) entails setting the persistence level to $\alpha = 1$, which is a trivial solution without fluctuations. Similar arguments apply to the correlation function, Eq. (1), which measures correlation in velocity fluctuations, rather than velocity itself, by way of subtracting the time-averaged velocity. These considerations ensure the existence of competing terms in the objective function at short and long timescales, F_1 and F_N in Eq. (3).

Maximizing information-theoretic measures has been identified as an advantageous approach for generating self-organized behavior in robotic systems [20,21]. In particular, it was found by Ay *et al.* [22] that a protocol of maximizing predictive information in the wheel velocity trajectories of a simple two-wheeled robot results in the robot exploring space efficiently. The optimal solution that maximized predictive information was for the control parameter to be tuned to a critical region near a pitchfork bifurcation. The two-wheeled robot exhibits a run-and-tumble motion slightly more complicated than a persistent random walk.

We study the one-dimensional persistent random walk to determine the optimal persistence level $\hat{\alpha}(N)$ where the walker's motion has maximally correlated fluctuations. As a Markov process with only two states and one control parameter (Fig. 1), it is perhaps the simplest setting in which to study optimal diffusion trajectories identified by an information maximization criterion. The central problem can be rephrased as the question: How often should a random walker switch directions in order to maximize correlation in its velocity trajectory? We find that the optimal strategy corresponds to the average number of direction-changing events in a trajectory scaling as the square root of the observation time.

In Sec. II we discuss the persistence time and average number of events in persistent random-walk trajectories. This is followed in Sec. III by a discussion of a scaling ansatz for the optimal diffusion problem to be expected in the limit of a long observation time interval. Following these preliminary sections, we carry out an analysis of the correlation function, Eq. (1), which behaves in a manner similar to mutual information but is easier to analyze. In Sec. V the mutual information, Eq. (2), is calculated analytically and its scaling behavior is compared to the correlation function. The paper concludes in Sec. VI with a discussion of open problems related to maximally correlated diffusion, including universality of optimal strategies and connections to existing results for optimal search and exploration by random walkers. We also mention a potential application of the maximum information problem for Markov chains to the thermodynamics of information engines.

II. BASIC PROPERTIES OF PERSISTENT RANDOM-WALK TRAJECTORIES

The analysis presented herein is based on realizations of a one-dimensional persistent random walk in discrete time. The walker's speed is set to $|v| = 1$ for simplicity and velocity trajectories are generated over N time steps (e.g., see Fig. 1). For each trajectory $[v_0; v_1, \ldots, v_N]$ the initial velocity v_0 is randomly selected from $+1$ and -1 with equal probability, and subsequent velocity values are randomly generated

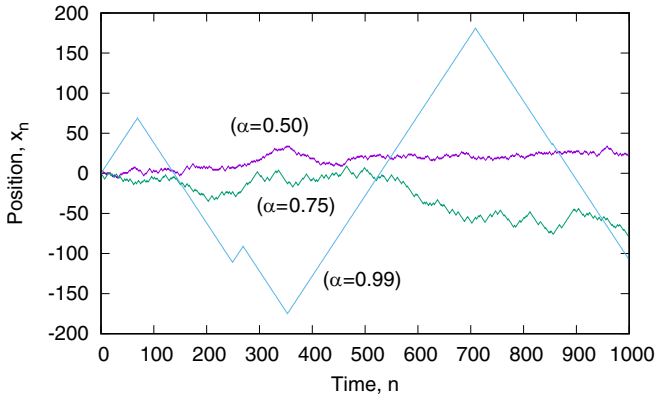

FIG. 2. Persistent random-walk trajectories. The number of events (changes of direction) in a trajectory decreases and the persistence time increases as α approaches 1.

according to the update rule

$$v_{n+1} = \begin{cases} v_n & \Pr = \alpha \\ -v_n & \Pr = 1-\alpha \end{cases}.$$

Persistent random-walk trajectories consist of laminar regions of ballistic motion separated by direction-changing events (Fig. 2). Hereafter an event refers to a change of direction by the random walker. Trajectories can be characterized on average by the number of events and the duration of the persistent regions between events.

Formally, the velocity trajectory is a realization of a Poisson or Bernoulli process where the rate of events ϵ is related to the persistence level by

$$\epsilon = 1 - \alpha. \tag{4}$$

More specifically the process is a symmetric two-state Markov chain, depicted in Fig. 1. The average number of events $\langle N_e \rangle$ in a trajectory is given by the product of the rate of events and the time duration,

$$\langle N_e \rangle = \epsilon N = (1-\alpha)N. \tag{5}$$

The persistence time τ_α of the random walk is the average time between consecutive events. The probability distribution function $\psi(n)$ for an event to occur exactly n time steps after the previous event occurred is

$$\psi(n) = \alpha^{n-1}(1-\alpha). \tag{6}$$

From this waiting-time distribution, the average time between events is

$$\tau_\alpha = \sum_{n=1}^{\infty} n\psi(n) = \frac{1}{1-\alpha}. \tag{7}$$

The persistence time, as determined from the persistence level α, is an inherent timescale of the stochastic process. When analyzing observables of the process the observation time should be compared with the persistence time. Ballistic motion is observed on relatively short timescales $N \ll \tau_\alpha$. Diffusive motion is observed on relatively long timescales $N \gg \tau_\alpha$, see Fig. 2.

The persistence length l_p of the random walk is the average distance traveled by the walker between consecutive events. Since we assumed the walker moves with speed $|v| = 1$, the persistence length is equal to the persistence time defined in Eq. (7),

$$l_p(\alpha) = \tau_\alpha = \frac{1}{1-\alpha}. \tag{8}$$

The persistence length is an inherent length scale of the stochastic process.

III. SCALING ANSATZ

Before going ahead, it is useful to discuss the expected behavior of solutions to finite-time optimization problems in the persistent random walk and introduce a scaling ansatz. We are looking for the optimal persistence level $\hat{\alpha}(N)$ that maximizes a chosen objective function, a functional of random-walk trajectories of duration N. We can expect the limiting behavior:

$$\lim_{N\to\infty} \hat{\alpha}(N) = 1; \tag{9}$$

otherwise, the optimal trajectories will become too diffusive as the observation time is increased. For large N, a reasonable scaling assumption is of the inverse power law (IPL) form:

$$\hat{\alpha}(N) \sim 1 - \frac{1}{N^\nu}, \tag{10}$$

where the IPL index is positive ($\nu > 0$) to ensure Eq. (9) is satisfied.

This scaling assumption can be alternatively expressed in terms of the optimal rate of events $\hat{\epsilon}(N)$, which vanishes in the limit of long observation times:

$$\lim_{N\to\infty} \hat{\epsilon}(N) = 0, \tag{11}$$

and scales as the IPL:

$$\hat{\epsilon}(N) \sim N^{-\nu}. \tag{12}$$

For a finite observation time N, there exists a persistence level $\alpha_{\mathrm{CLT}}(N)$ at which central limit theorem approximations of the random-walk process begin to break down. See Fig. 3 for example. This breakdown, which is indicative of the transition between diffusive motion ($\langle N_e \rangle \gg 1$) and ballistic motion ($\langle N_e \rangle \ll 1$), occurs when the number of events in a trajectory is on the order of one, $\langle N_e \rangle \sim 1$. Equation (5) implies the following relation between the number of steps of the walker and the persistence parameter:

$$N \sim \frac{1}{1-\alpha} = \tau_\alpha, \tag{13}$$

where we also made use of the definition of the persistence time in Eq. (7). Therefore, for observation times on the order of the persistence time τ_α the central limit theorem is not yet applicable. Alternatively, we can rewrite Eq. (13) in terms of the persistence level, explicitly denoting the central limit theorem condition on α:

$$\alpha_{\mathrm{CLT}}(N) \sim 1 - \frac{1}{N}. \tag{14}$$

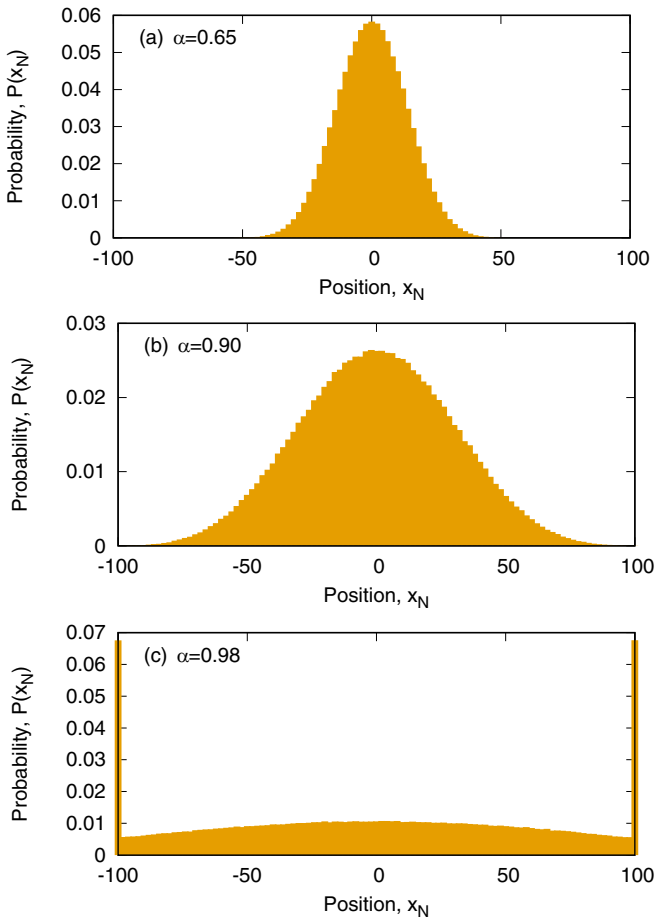

FIG. 3. Empirical probability distributions for the final position x_N reached by the random walker after $N = 100$ steps for various persistence levels (generated from an ensemble of 10^6 trajectories). The central limit theorem treatment of the random-walk process breaks down and ballistic peaks become predominant as α approaches 1.

In other words, trajectories look diffusive over the finite observation time when $\alpha \ll \alpha_{\mathrm{CLT}}(N)$.

Equation (14) provides the correct scaling behavior for the transition between diffusive and ballistic motion, but it does not provide an accurate ad hoc estimate of the persistence level below which the random-walk process is sufficiently diffusive for central limit theorem arguments to be safely adopted. The problem is that the transition takes place over a range of α, and we would like to establish a more conservative estimate for α_{CLT} that lies near the diffusive end of that range. To do this we write:

$$\alpha_{\mathrm{CLT}}(N) \approx 1 - \frac{\kappa}{N}, \tag{15}$$

where κ is a constant related to the average number of events needed for the central limit theorem approximation to reach a certain degree of accuracy. Equation (15) is equivalent to the statement that central limit theorem arguments hold when $\langle N_e \rangle \geqslant \kappa$. We adopt $\kappa = 10$, which satisfies the diffusive limit $\langle N_e \rangle \gg 1$ by an order of magnitude. The implications of this choice will become clearer following calculations in the next section. For now we point out that ballistic trajectories are governed by the probability for the random walker not to have changed direction up to time N, $P = (1-\epsilon)^N$. In the limit of large N and small ϵ this probability becomes $P = \exp(-\epsilon N)$. Therefore, the condition of setting $\langle N_e \rangle \geqslant \kappa$ is analogous to the statement that the normal approximation of the probability distribution for the time-averaged velocity $\overline{v}(N)$, or for the final position $x(N)$ as shown in Fig. 3, ceases to be accurate when the two ballistic peaks in the distribution exceed a threshold of $P = \frac{1}{2}\exp(-\kappa)$; for $\kappa = 10$ the threshold is $P \approx 2.27\times10^{-5}$ (with the factor of $1/2$ coming from the existence of two equiprobable initial directions, $v_0 = \pm 1$).

We make the ansatz that optimal solutions, having maximum correlation and mutual information, follow the scaling function Eq. (10) with an exponent $\nu < 1$. When this is the case, the relation

$$\tfrac{1}{2} < \hat{\alpha}(N) < \alpha_{\mathrm{CLT}}(N) < 1 \tag{16}$$

holds if N is sufficiently large. For short observation times Eq. (16) may be violated because of the different constants showing up in the exact expressions for the optimal persistence level in Eq. (10) and the value of κ in Eq. (15). When sending N to infinity with $\nu < 1$, a separation of scales occurs between the central limit theorem boundary $\alpha_{\mathrm{CLT}}(N)$ and the optimal solution $\hat{\alpha}(N)$, because the latter quantity tends to one more slowly. The optimal solution lies in the diffusive regime and central limit theorem arguments apply when calculating the optimal persistence level. These conditions lead to universality and perhaps also to wider applicability of the solution beyond physical diffusion processes. This ansatz is founded on the fact that we are looking for maximally correlated fluctuations, and the process must be on the diffusive side rather than on the ballistic side in order for large fluctuations to exist.

IV. AUTOCOVARIANCE

First we study the autocovariance function:

$$C(\alpha, N) \equiv \overline{v_n v_{n+1}} - \overline{v}^2(N) \tag{17}$$

as a measure of correlated fluctuations in a single velocity trajectory, such that the overline denotes a time average over the observation time window. Operationally, the two-point correlation is a sum over the steps in the random walk:

$$\overline{v_n v_{n+1}} = \frac{1}{N-1}\sum_{n=1}^{N-1} v_n v_{n+1}, \tag{18}$$

and the time-averaged velocity is also a sum over steps:

$$\overline{v}(N) = \frac{1}{N}\sum_{n=1}^{N} v_n. \tag{19}$$

We stress that the time-averaged velocity depends on the observation time. Since there is no directional bias and the magnitude of each step is of unit size, $\overline{v}(N)$ vanishes in the limit $N \to \infty$ as events accumulate and erase the memory of the initial condition v_0. However, we are interested in the finite-time fluctuations in observables of the random walk. Moreover, we will study the behavior of the autocovariance function along with its maximum value as a function of N. The two-point correlation function $\overline{v_n v_{n+1}}$ is a local quantity

and therefore does not depend as strongly on N, in contrast to the transient relaxation of $\overline{v}^2$. We focus analysis on the time lag of one step because of the Markovian nature of the process.

Equation (17) is a sample autocovariance function for adjacent velocity values. As an objective function, Eq. (17) rewards motion continued along the current direction through the first term on the right-hand side, but through the second term it penalizes motion continued in the same direction during a majority of the entire time. In this way the function favors trajectories with large, correlated fluctuations. Analysis of the autocovariance $C(\alpha, N)$ serves as a simple route to an optimal solution $\hat{\alpha}(N)$ that turns out to be similar to mutual information. A similar autocorrelation approach has been adopted for time-series analysis of brain activity [23,24], as well as for phase transition dynamics [25]. Near phase transitions, correlations in time and space are slowly decaying IPL functions related by scaling critical exponents [26]. Consequently, maximum correlation or information protocols often identify optimal regions in parameter space near critical points, when they exist [27–30].

For any finite observation time the covariance will be a stochastic quantity, and so in addition to the time average within a given trajectory we further consider the average over an ensemble of trajectories:

$$\langle C \rangle = \langle \overline{v_n v_{n+1}} \rangle - \langle \overline{v}^2(N) \rangle. \tag{20}$$

Equation (20) is the objective function for the optimal diffusion problem addressed in this section.

The first term on the right-hand side of Eq. (20), after explicitly carrying out the two types of averages, is

$$\langle \overline{v_n v_{n+1}} \rangle = \frac{1}{N-1} \sum_{n=1}^{N-1} \langle v_n v_{n+1} \rangle = 2\alpha - 1, \tag{21}$$

since the ensemble average makes the sum over time redundant and the probability for consecutive velocity values to be equal is, by definition, the persistence level α. The second term on the right-hand side of Eq. (20) is more difficult to evaluate because of the N dependence; the result is given in Eq. (29). While the steps in the derivation of Eq. (29) can be found in tutorial discussions of the mean-squared displacement for the persistent random walk, we repeat them here to make the discussion self-contained. The square of the time-averaged velocity is given by the double sum:

$$\overline{v}^2(N) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} v_i v_j. \tag{22}$$

Thinking of the double sum in terms of a matrix whose elements a_{ij} are products of instantaneous velocity values $v_i v_j$, it becomes evident that the matrix is symmetric and each diagonal element is 1. This realization leads to the simplification:

$$\overline{v}^2(N) = \frac{1}{N^2} \left[N + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} v_i v_j \right], \tag{23}$$

which readily simplifies on taking an ensemble average. On bringing the ensemble average inside the summation, the matrix of velocity pairs is converted into a matrix of correlation coefficients $c_k = \langle v_i v_j \rangle$ with $k = j - i$. In this way straightforward combinatorics yields:

$$\langle \overline{v}^2(N) \rangle = \frac{1}{N^2} \left[N + 2 \sum_{k=1}^{N-1} (N-k) c_k \right]. \tag{24}$$

Adopting an event picture, it can be shown that the correlation coefficients are given by the sum:

$$c_k = \sum_{l=0}^{k} (-1)^l P(l, k), \tag{25}$$

where $P(l, k)$ is the probability for l events to occur in k time steps, i.e., the binomial probability:

$$P(l, k) = \binom{k}{l} (1-\alpha)^l \alpha^{k-l}. \tag{26}$$

The sums over even and odd values of l in Eq. (25) are separately evaluated and then recombined to obtain the correlation coefficient:

$$c_k = (2\alpha - 1)^k \equiv r^k, \tag{27}$$

where to simplify notation we have introduced the new variable:

$$r = 2\alpha - 1. \tag{28}$$

Note that r can be thought of as a velocity correlation level: $r = 0$ corresponds to zero correlation reducing the analysis to the simple random walk and $r = 1$ corresponds to ballistic motion. Substituting the correlation coefficient into Eq. (24) and evaluating the sum leads finally to the analytic expression:

$$\langle \overline{v}^2(N) \rangle = \frac{1}{N^2} \left\{ N + 2 \left[\frac{(N-1)r - Nr^2 + r^{N+1}}{(1-r)^2} \right] \right\}. \tag{29}$$

Most random-walk processes imply a fairly large observation time, $N \gg 1$. To account for this we manipulate the expression in Eq. (29) for the second moment by collecting terms in increasing powers of $1/N$:

$$\langle \overline{v}^2(N) \rangle = \frac{1}{N} \frac{1+r}{1-r} - \frac{2r}{N^2(1-r)^2} (1 - r^N). \tag{30}$$

Since we are interested in the regime of large but finite N, we drop the second term on the right-hand side of this last expression. This must be done with some caution, however, since we have to allow for the possibility that the region of interest $\alpha \approx \hat{\alpha}(N)$ is very close to 1, which indeed happens when $N \to \infty$. It is useful to recall from Eq. (5) that the average number of events in a trajectory of length N is given by $(1-\alpha)N$. Accordingly, the second moment can be expressed as an expansion in powers of $\langle N_e \rangle^{-1}$,

$$\langle \overline{v}^2(N) \rangle = \frac{\alpha}{\langle N_e \rangle} - \frac{2\alpha - 1}{2\langle N_e \rangle^2} [1 - (2\alpha - 1)^{N+1}]. \tag{31}$$

The large-N limit is more appropriately viewed as a limit of a large average number of events, $\langle N_e \rangle \gg 1$. When $\langle N_e \rangle$ is large

this expression is approximately

$$\langle \overline{v}^2(N)\rangle \approx \frac{\alpha}{\langle N_e\rangle} = \frac{\alpha}{1-\alpha}\frac{1}{N}, \tag{32}$$

which is compatible with the central limit theorem.

When treating the persistence parameter α and the observation time N as independently adjustable quantities, for example, fixing N large and then varying α over the entire interval $[1/2, 1]$, it is important to keep in mind that the approximation yielding Eq. (32) will fail in the limit $\alpha \to 1$ if N is finite. The condition for failure of the approximation can be stated as $\langle N_e\rangle < \kappa$, or α exceeding the central limit theorem value of the persistence parameter given by Eq. (15), which was discussed in Sec. III. More precisely, Eq. (31) can be written as

$$\langle \overline{v}^2(N)\rangle = \frac{\alpha}{\langle N_e\rangle} - \Delta^2, \tag{33}$$

where the deviation Δ^2 from the central limit theorem expression, Eq. (32), is bounded by the average number of events:

$$\Delta^2 \equiv \frac{2\alpha-1}{2\langle N_e\rangle^2}[1-(2\alpha-1)^{N+1}] < \frac{1}{\langle N_e\rangle^2}. \tag{34}$$

Therefore the estimate for α_{CLT} in Eq. (15) based on the average number of events $\langle N_e\rangle \geqslant \kappa$ is equivalent to setting a threshold on the accuracy of the central limit theorem approximation of the time-averaged velocity squared: $\Delta^2 < 1/\kappa^2$ when $\alpha < \alpha_{\mathrm{CLT}}$. Looking ahead, we rely on the scaling ansatz from Sec. III and assume that the central limit theorem approximation is useful for the purpose of calculating the optimal value of the persistence parameter. Comparing the end result we find for $\hat{\alpha}(N)$ with the central limit theorem boundary verifies the inequality in Eq. (16). This suggests that the central limit theorem approximation is valid in the vicinity of the maximum of the covariance function, though it deteriorates for persistence values in $\alpha_{\mathrm{CLT}} < \alpha < 1$.

To better see that Eq. (32) is an expression of the central limit theorem recall that the displacement of the walker after N steps can be recovered from the mean velocity:

$$x(N) = \sum_{i=1}^{N} v_i = N\overline{v}(N). \tag{35}$$

Squaring this quantity and taking an ensemble average yields, after inserting Eq. (32), the familiar asymptotic expression

$$\langle x^2(N)\rangle \approx \frac{\alpha}{1-\alpha}N \tag{36}$$

for the second moment of the displacement growing linearly in time. If we write:

$$\langle x^2(n)\rangle = 2D_\alpha n, \tag{37}$$

for a one-dimensional diffusion process, then evidently the diffusion coefficient D_α for the persistent random walk is given in terms of the persistence level:

$$D_\alpha = \frac{1}{2}\frac{\alpha}{1-\alpha}. \tag{38}$$

Equation (38) clearly indicates that the diffusion coefficient increases rapidly as α approaches one. This divergence occurs because the persistence time or length, i.e., the effective step size of the rescaled random walk, becomes infinite in this limit. If we relate a large diffusion coefficient to more efficient exploration of space, enough time must elapse (much greater than the persistence time) in order for a region of space to be truly explored, rather than just drifted through without any changes in direction on the part of the walker. This provides a better picture of how the persistence level α and the observation time N need to be balanced against one another for optimal diffusion. Note that with Eq. (35), in addition to Eq. (21), we can alternatively write the velocity autocovariance in terms of the displacement:

$$\langle C\rangle = \langle v_n v_{n+1}\rangle - \frac{1}{N^2}\langle x^2(N)\rangle. \tag{39}$$

Eq. (39), with only ensemble averages, emphasizes that the analysis of the persistent random walk based on time averages of velocity, namely Eq. (20), can be interchanged with an analysis based on related ensemble-averaged quantities or vice versa.

Returning to the main calculation, from Eqs. (21) and (29) we find the average of the autocovariance function, Eq. (20), for a persistent random-walk trajectory:

$$\langle C\rangle = r - \frac{1}{N^2}\left\{N + 2\left[\frac{(N-1)r - Nr^2 + r^{N+1}}{(1-r)^2}\right]\right\}. \tag{40}$$

Assuming that $\langle N_e\rangle \gg 1$ and adopting the approximate expression for the time-averaged velocity squared, Eq. (32), results in the asymptotic velocity autocovariance:

$$\langle C(\alpha, N)\rangle \approx 2\alpha - 1 - \left(\frac{\alpha}{1-\alpha}\right)\frac{1}{N}. \tag{41}$$

The optimal persistence level $\hat{\alpha}$ is obtained from the extremum of the average velocity autocovariance obtained by taking the derivative of Eq. (41) with respect to α:

$$\frac{\partial}{\partial\alpha}\langle C(\alpha, N)\rangle = 2 - \frac{1}{(1-\alpha)^2}\frac{1}{N}. \tag{42}$$

Setting this derivative to zero yields the optimal value of the persistence level:

$$\hat{\alpha}(N) \approx 1 - \left(\frac{1}{2N}\right)^{1/2}. \tag{43}$$

Equation (43) is the key result of this section. It provides our first estimate of the optimal persistence level where fluctuations are maximally correlated. Looking back to the scaling assumption in Eq. (10) it is evident that the scaling exponent is $\nu = 1/2$. The scaling solution of $\hat{\alpha}(N)$ with the exponent $\nu = 1/2$ is referred to as the optimal strategy for generating persistent random motion with maximally correlated fluctuations over a finite time window N.

Figure 4 shows the behavior of the autocovariance function and its maximum value. When N is relatively small, the central limit theorem approximation is not yet accurate around the maximum value of $\langle C\rangle$. In fact, for $N = 100$ we have $\alpha_{\mathrm{CLT}} < \hat{\alpha}$ in violation of the inequality in Eq. (16). However, as N is increased α_{CLT} becomes greater than $\hat{\alpha}$ and the central limit theorem approximation becomes very good in the vicinity of the optimal persistence level. This validates

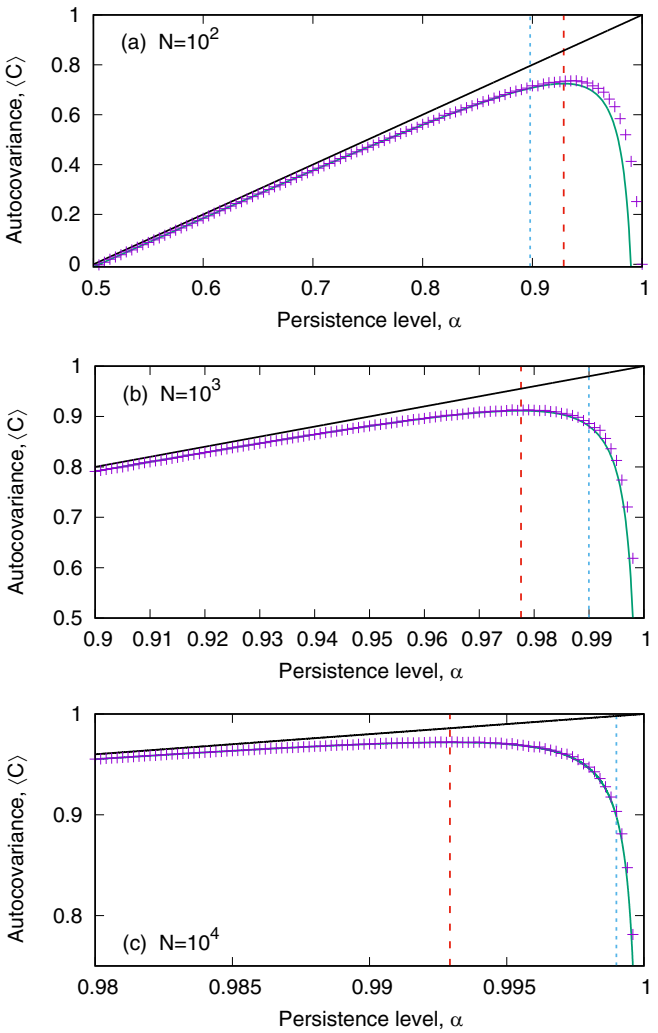

FIG. 4. Autocovariance function for increasing values of N. The approximation of Eq. (41) (solid green line) is compared to numerical simulation of the persistent random walk (purple markers). The optimal persistence level in Eq. (43) (dashed red vertical line) is included along with α_{CLT} from Eq. (15) (short dashed cyan line). The central limit theorem approximation becomes accurate around the optimal persistence level $\hat{\alpha}$ for observation times $N > 1000$.

the scaling ansatz, as well as the optimal strategy found for large N in Eq. (43). We note that the central limit theorem approximation, although adequate for our analysis here of large N, can be improved on by using large deviation theory [31–33] or exact solutions known for the persistent random walk and telegrapher's equation [8,34].

More insight can be gained by considering the average number of events in a trajectory when $\alpha = \hat{\alpha}(N)$. Substituting the optimal persistence value into Eq. (5) leads to:

$$\langle N_e(\hat{\alpha})\rangle \approx \left(\frac{N}{2}\right)^{\frac{1}{2}}, \tag{44}$$

which is a restatement, or corollary, of the key result Eq. (43). It is evident that for trajectories to have maximally correlated fluctuations the average number of events should be proportional to the square root of the number of time steps in the random walk.

At the optimal persistence level the persistence length defined in Eq. (8) is

$$l_p(\hat{\alpha}) \approx (2N)^{\frac{1}{2}}. \tag{45}$$

Note that this entails the relation

$$N = l_p\langle N_e\rangle \tag{46}$$

for consistency.

The scaling exponent of $\nu = 1/2$ appearing in Eqs. (44) and (45) is important, perhaps more so than finding the exact values of $\langle N_e(\hat{\alpha})\rangle$ and $l_p(\hat{\alpha})$ because the objective function $C(\alpha, N)$ is not unique, and slight changes to the objective function will inevitably lead to a different optimal solution. For example, if the autocovariance time lag is k steps such that

$$\langle C_k\rangle = \langle \overline{v_n v_{n+k}}\rangle - \langle \overline{v}^2(N)\rangle, \tag{47}$$

then based on Eq. (27) the objective function is approximately

$$\langle C_k\rangle \approx (2\alpha - 1)^k - \frac{\alpha}{1-\alpha}\frac{1}{N}, \tag{48}$$

and in the dual limit of large N and $\alpha \to 1$ the optimal solution is

$$\hat{\alpha}(N) \approx 1 - \left(\frac{1}{2kN}\right)^{\frac{1}{2}}. \tag{49}$$

The scaling behavior is unchanged, but the constant prefactor depends on the time lag k. To further see that $\nu = 1/2$ is to some extent a universal exponent that characterizes an optimal strategy for a class of objective functions, in the next section we perform calculations where mutual information is the objective function. The exponent $\nu = 1/2$ in one sense marks the boundary between overly deterministic trajectories and overly random trajectories. If the rate of events were to go to zero relatively quickly as the observation time increases to infinity, that is, if $\nu > 1/2$ in Eq. (12), then the random-walk trajectories would become more ordered or ballistic than what is optimal. If the rate of events were to decay relatively slowly with N, such that $\nu < 1/2$ in Eq. (12), then the trajectories would become more diffusive or disordered than what is optimal.

To place the optimal solution in a proper context consider the following basic scaling arguments. Assume that the average number of events scales with the observation time as

$$\langle N_e(\hat{\alpha})\rangle \sim N^a, \tag{50}$$

and the persistence length scales as

$$l_p(\hat{\alpha}) \sim N^b, \tag{51}$$

then according to Eq. (46) the scaling exponents must satisfy the condition:

$$a + b = 1. \tag{52}$$

From the scaling assumption for the persistence level, Eq. (10), $a = 1 - \nu$ and $b = \nu$. Based on Eqs. (50) and (51), it is useful to introduce an additional scaling equation for the relative growth between the average number of events and the

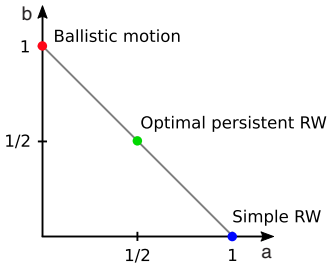

FIG. 5. Comparing exponents in the scaling laws Eqs. (50) and (51) for different kinds of motion. The optimal strategy for maximally correlated persistent random motion corresponds to the scaling exponents $(a = 1/2, b = 1/2)$, which lies half way in between the two extremes of ballistic motion and uncorrelated random motion on the line $a + b = 1$.

persistence length:

$$\frac{\langle N_e(\hat{\alpha})\rangle}{l_p(\hat{\alpha})} \sim N^{a-b} = N^{1-2v}. \tag{53}$$

The optimal solution corresponds to $a = b = 1/2$. Physically this means that the average number of events and the persistence length both diverge as the observation time is increased, but their relative growth is left invariant, $a - b = 0$ in Eq. (53). This central result should be compared with the extreme cases of the simple random walk and ballistic motion. For the simple random walk ($\alpha = 1/2$), the average number of events scales linearly with the observation time:

$$\langle N_e\rangle = \frac{N}{2}, \tag{54}$$

and the persistence length is constant:

$$l_p = 2. \tag{55}$$

Therefore, scaling analysis for the simple random walk yields $a = 1$ and $b = 0$. For ballistic motion ($\alpha = 1$), the average number of events effectively vanishes:

$$\langle N_e\rangle < 1, \tag{56}$$

while the persistence length exceeds the observation time:

$$l_p > N. \tag{57}$$

Therefore, ballistic motion corresponds to the scaling exponents $a \approx 0$ and $b \approx 1$. Evidently the scaling exponents for the optimal solution lie exactly in the middle of these two extreme points on the line $a + b = 1$. Figure 5 illustrates these scaling arguments.

Interestingly, the optimal strategy consisting of Eqs. (44) and (45) appears to be different from that derived in the optimal search problem, where a persistent random walker is trying to minimize the first passage time to a target [14]. Therein it was suggested that the optimal strategy corresponds to ballistic motion (in terms of scaling exponents):

$$l_p \sim L, \tag{58}$$

where L is the linear size of the exploration space in which the walker performs the search. The constant in the exact expression for Eq. (58) is, however, smaller than 1, so there are events in the trajectory. Also, the scaling result of Eq. (58) was found for a two-dimensional persistent random walk in a bounded space, so further analysis of the strategies for maximizing correlation in similar situations is required to make a fair comparison. In an attempt to compare to Eq. (45) presently, we can relate the observation time to the system size by $N \sim L^{\delta}$. If the observation time is naturally determined from the system's size divided by the walker's speed, then $N \sim L$ ($\delta = 1$), and the two optimal strategies are different. If instead the observation time can be freely adjusted and the exponent δ selected through design of the optimal diffusion experiment, then a correspondence may be established if, for example, $N \sim L^2$ ($\delta = 2$). In that case trajectories might look ballistic for observation times on the order of the linear size L but diffusive for observation times on the order of the area L^2. This heuristic correspondence does not account for boundary effects that usually go along with setting a finite system size (i.e., periodic boundary conditions are implicit in the argument).

V. MUTUAL INFORMATION

The application of information theory made herein involves calculating the average predictive information in a single persistent random-walk velocity trajectory. The predictive information is a measure of the information contained in the history of a process that is useful for predicting future states; it is defined as the mutual information between the history of a process and its future [35]. Since the velocity in the persistent random walk is a Markov process, the predictive information reduces to the mutual information between consecutive velocity values [22]. We can calculate the mutual information between consecutive velocity values through the entropy relation:

$$I(v_n, v_{n+1}) = H(v_n) + H(v_{n+1}) - H(v_n, v_{n+1}), \tag{59}$$

where, with nats as the unit of measurement, the Shannon entropy of the instantaneous velocity is defined:

$$H(v_n) = -\sum_{v_n} p(v_n) \ln p(v_n) \tag{60}$$

and the joint entropy is defined:

$$H(v_n, v_{n+1}) = -\sum_{v_n}\sum_{v_{n+1}} p(v_n, v_{n+1}) \ln p(v_n, v_{n+1}). \tag{61}$$

Equation (59) is often referred to as a delayed mutual information, auto mutual information, or time-lagged mutual information [36,37], where in this case the time lag is one time step.

Two problems are commonly encountered when numerically calculating mutual information from a time series [38,39]: One is binning the data and the other is estimating probabilities from a finite number of samples. The binning problem is avoided here for the binary velocity trajectory. However, the problem of interpreting probabilities as empirical rates holds paramount importance for this work. In order to reproduce an information-theoretic analysis similar to that for the correlation above, we work in the experimental or applied setting where quantities entering into the equations are empirical rates instead of the theoretical setting where quantities are well-defined probabilities.

We first consider the entropy in Eq. (60). To evaluate the two terms in this sum, estimates for the single-point probabilities $p(v_n = +1)$ and $p(v_n = -1)$ are required. The direct approach to estimating a probability from a trajectory is to count the number of instances occurring along the trajectory and divide by the observation time to obtain a rate. For example, if N_+ is the number of time steps where $v_n = +1$, then:

$$p(v_n = +1) \approx \frac{N_+}{N} = n_+. \tag{62}$$

However, this approximation only works if N_+ is very large. In other words, rates converge to probabilities in the limit of an infinite observation time:

$$\lim_{N\to\infty} \frac{N_+}{N} = p(v_n = +1), \tag{63}$$

the law of large numbers.

To evaluate the entropy, typically one would generate a sufficiently long time series (based on timescales inherent in the stochastic process) to effectively reach the infinite-time regime $N \approx \infty$ where rates can be freely interchanged with probabilities as in Eq. (62). We are interested in the behavior of the random-walk process and in the corresponding information measures with increasing N. Consequently, we fix the observation time N to a finite value, and we replace the probabilities with empirical rates, which are stochastic quantities, fluctuating from one trajectory to another. The information measures then become stochastic quantities as well. Finally we take an ensemble average of the single-trajectory information measures to get a consistent estimate of the finite-time random-walk behavior. Generally N cannot be made large enough to neglect fluctuations in Eq. (62). That is, finite-time fluctuations play a fundamental role in the optimization problem addressed, even in the limit $N \to \infty$.

The entropy $H(v_n)$ is based on the residence time of the velocity trajectory in each of the two states, and is essentially the same as $H(v_{n+1})$. The finite-time (stochastic) entropy of the velocity is written as:

$$H(v;N) = -[n_+ \ln n_+ + n_- \ln n_-]. \tag{64}$$

We stress that the entropy $H(v;N)$ in Eq. (64) is a functional of a single velocity trajectory. Note the trajectory-level entropy in Eq. (64) based on empirical rates (time averages) is different from the trajectory-level stochastic entropy $s(n) = -\ln p(v_n, n)$ defined in stochastic thermodynamics [40,41]. The stochastic rates n_+ and n_- are related by the normalization condition:

$$n_+ + n_- = 1. \tag{65}$$

It is convenient to reintroduce here the time-averaged velocity previously defined in Eq. (19):

$$\overline{v}(N) = n_+ - n_-, \tag{66}$$

and change variables by expressing the rates in terms of the time-averaged velocity:

$$n_\pm = \tfrac{1}{2}(1 \pm \overline{v}). \tag{67}$$

Then the entropy, Eq. (64), is also expressed in terms of the time-averaged velocity:

$$H(v;N) = -\tfrac{1}{2}[(1+\overline{v})\ln(1+\overline{v}) + (1-\overline{v})\ln(1-\overline{v}) - 2\ln 2]. \tag{68}$$

Note that the time-averaged velocity $\overline{v}$ is restricted to the interval $[-1, +1]$. Furthermore, it is expected to vanish in the long-time limit $N \to \infty$. This limiting behavior suggests an expansion of the natural logarithm terms in the entropy, which leads to the expression:

$$H(v;N) = \ln 2 - \sum_{n=1}^{\infty} \frac{1}{2n(2n-1)} \overline{v}^{2n}. \tag{69}$$

If we assume N is sufficiently large to warrant keeping only the first two terms in an expansion of the natural logarithm:

$$\ln(1 \pm \overline{v}) \approx \pm\overline{v} - \tfrac{1}{2}\overline{v}^2, \tag{70}$$

then this large-N approximation leads to a simple expression for the finite-time entropy of the instantaneous velocity:

$$H(v;N) \approx \ln 2 - \tfrac{1}{2}\overline{v}^2(N). \tag{71}$$

The ensemble average of the stochastic entropy is to this order of approximation given by:

$$\langle H(v;N)\rangle \approx \ln 2 - \frac{1}{2}\langle \overline{v}^2(N)\rangle \approx \ln 2 - \frac{1}{2}\frac{\alpha}{1-\alpha}\frac{1}{N}, \tag{72}$$

where in the last step we utilized Eq. (32).

The structure of the finite-time corrections to the entropy is evident from Eq. (72). The first term corresponds to the limit of an infinite observation time:

$$\langle H(v;N=\infty)\rangle = H(v;N=\infty) = \ln 2, \tag{73}$$

where rates become equivalent to probabilities, and each velocity trajectory spends an equal amount of time in the $v = +1$ and $v = -1$ states, such that $p(v_n) = \frac{1}{2}$ in Eq. (60). The second term in Eq. (72) is the $O(\frac{1}{N})$ correction to the infinite (theoretical) entropy which is a consequence of the central limit theorem approximation for the persistent random walk.

Corrections such as Eq. (72) to entropy and other information-theoretic quantities due to a finite number of samples have been discussed in the literature [42–46]. Here we are dealing with a correlated time series, and the correlation time τ_α plays an important role in the finite-time corrections [46], as can be seen from the dependence in Eq. (72) on α. It is important to note that the convention in information theory literature is to interpret the finite-time fluctuations as a source of statistical error or bias in estimating the "true" entropy $H(v;\infty)$ of the process [43–46]: The finite-time entropy Eq. (72) underestimates the true entropy, and therefore measurements of $\langle H(v;N)\rangle$ are corrected by adding on an $O(\frac{1}{N})$ term to extrapolate from the finite-time measured value to the infinite-time result. We adopt a different interpretation where the finite-time measurements are taken as empirical entropies having physical significance, rather than being systematically biased estimations of an idealized asymptotic entropy $H(v;N=\infty)$. Because all measurements are finite in practice, the finite-time information measures are

arguably more physically relevant to correlated fluctuations than their infinite-time theoretical counterparts.

Adopting this physical perspective we proceed to calculate the joint entropy, Eq. (61). The double sum results in four terms representing the possible combinations of two consecutive velocity values. The finite-time (stochastic) joint entropy can be written as:

$$H(v_n, v_{n+1}; N) = \sum_{i=1}^{4} H_i, \tag{74}$$

where the H_1 term represents the "++ " combination where both velocity values are positive, H_2 represents "+−," H_3 represents "−+," and H_4 represents "−−." For a generic term H_i we write:

$$H_i = -n_i \ln n_i. \tag{75}$$

For example, when $i = 2$ we have

$$p(v_n = +1, v_{n+1} = -1) \approx \frac{N_{+-}}{N} = n_{+-} = n_2, \tag{76}$$

and Eq. (75) becomes

$$H_2 = -n_2 \ln n_2 = -n_{+-} \ln n_{+-} = H_{+-}. \tag{77}$$

Rather than utilizing the time-averaged velocity, as we did with Eq. (67) for the single-point entropy calculation, here we adopt a more general approach based on the central limit theorem for Markov chains. (See the Appendix for a discussion of the Markov chain central limit theorem.) We assume the rate n_i in Eq. (75) can be treated as a normal random variable with mean μ_i and variance σ_i^2:

$$n_i \sim \mathcal{N}(\mu_i, \sigma_i^2). \tag{78}$$

In other words, we assume:

$$n_i \approx \mu_i + \sigma_i \xi = \mu_i \left(1 + \frac{\sigma_i}{\mu_i} \xi\right), \tag{79}$$

where ξ is a Gaussian random variable with zero mean and unit variance, $\xi \sim \mathcal{N}(0, 1)$. Then we substitute this expression into Eq. (75), which yields:

$$H_i \approx -(\mu_i + \sigma_i \xi) \left[\ln \mu_i + \ln\left(1 + \frac{\sigma_i}{\mu_i} \xi\right)\right]. \tag{80}$$

We assume N is sufficiently large that σ_i, which is proportional to $\frac{1}{\sqrt{N}}$, is much smaller than the mean μ_i. Then we expand the natural logarithm, keeping only the first two terms as we did in Eq. (70) to obtain the expression:

$$H_i \approx -(\mu_i + \sigma_i \xi_i) \left[\ln \mu_i + \frac{\sigma_i}{\mu_i} \xi - \frac{1}{2}\left(\frac{\sigma_i}{\mu_i} \xi\right)^2\right]. \tag{81}$$

After multiplying out this expression and taking an ensemble average we are left with:

$$\langle H_i \rangle \approx -\mu_i \ln \mu_i - \frac{1}{2} \frac{\sigma_i^2}{\mu_i}, \tag{82}$$

noting that odd moments of ξ vanish and even moments are equal to 1.

Next we calculate the mean and variance of the stochastic rates in each case. The persistent random walk is defined by the probability α for the velocity to not change from one step to the next. If half of these cases where no change in velocity occurs in the trajectory correspond to two positive values, we obtain $n_{++} = \alpha/2$ or, equivalently:

$$\mu_1 = \frac{\alpha}{2}. \tag{83}$$

The variance is more difficult to calculate and details are relegated to the Appendix where we discuss the central limit theorem approximation for Markov chains. It turns out that the variance is

$$\sigma_1^2 = \frac{\alpha(\alpha^2 - 2\alpha + 2)}{4(1-\alpha)} \frac{1}{N}. \tag{84}$$

The associated entropy term is

$$\langle H_1 \rangle \approx -\frac{\alpha}{2} \ln \frac{\alpha}{2} - \frac{\alpha^2 - 2\alpha + 2}{4(1-\alpha)} \frac{1}{N}. \tag{85}$$

Due to symmetry between "++" and "−−," we have:

$$\langle H_1 \rangle = \langle H_4 \rangle. \tag{86}$$

On the other hand, events occur in the trajectory with probability $1 - \alpha$. For the "+−" case, the mean is

$$\mu_2 = \frac{1-\alpha}{2}, \tag{87}$$

and the variance is

$$\sigma_2^2 = \frac{\alpha(1-\alpha)}{4}. \tag{88}$$

The corresponding term in the joint entropy is

$$\langle H_2 \rangle = -\frac{1-\alpha}{2} \ln \frac{1-\alpha}{2} - \frac{\alpha}{4} \frac{1}{N}, \tag{89}$$

and once more by symmetry we have:

$$\langle H_2 \rangle = \langle H_3 \rangle. \tag{90}$$

Combining the four terms, the finite-time joint entropy can be written:

$$\langle H(v_n, v_{n+1}; N) \rangle \approx \ln 2 - \alpha \ln \alpha - (1-\alpha) \ln(1-\alpha) - \frac{2-\alpha}{2(1-\alpha)} \frac{1}{N}. \tag{91}$$

Note, when $\alpha \approx 1$ we have:

$$\langle H(v_n, v_{n+1}; N) \rangle \approx \ln 2 - \alpha \ln \alpha - (1-\alpha) \ln(1-\alpha) - \frac{1}{2} \frac{\alpha}{1-\alpha} \frac{1}{N}, \tag{92}$$

which connects back to the expressions involving the time-averaged velocity $\langle \overline{v}^2(N) \rangle$. Taking the limit $N \to \infty$ in Eq. (91) recovers the asymptotic expression for the joint entropy:

$$H(v_n, v_{n+1}; \infty) = \ln 2 - \alpha \ln \alpha - (1-\alpha) \ln(1-\alpha). \tag{93}$$

Using the above expressions for the entropies, the approximate ensemble-averaged finite-time mutual information is

$$\langle I(v_n, v_{n+1}; N) \rangle \approx \ln 2 + \alpha \ln \alpha + (1-\alpha) \ln(1-\alpha) - \frac{1}{2} \frac{\alpha}{1-\alpha} \frac{1}{N}. \tag{94}$$

Note that from now forward we assume that α is close to 1. Taking the limit $N \to \infty$ in Eq. (94) recovers the asymptotic expression for the mutual information:

$$I(v_n, v_{n+1}; \infty) = \ln 2 + \alpha \ln \alpha + (1-\alpha)\ln(1-\alpha). \quad (95)$$

Next we evaluate the optimal persistence level as a function of the observation time by taking the derivative of the mutual information:

$$\frac{\partial}{\partial \alpha}\langle I \rangle = \ln \alpha - \ln(1-\alpha) - \frac{1}{2}\frac{1}{(1-\alpha)^2}\frac{1}{N}. \quad (96)$$

Setting this equation to zero establishes the problem to solve, that being the transcendental equation:

$$\ln \alpha - \ln(1-\alpha) - \frac{1}{2}\frac{1}{(1-\alpha)^2}\frac{1}{N} = 0. \quad (97)$$

We neglect the contribution from the $\ln \alpha$ term since the optimal value of α tends to 1 as N tends to infinity. In terms of the small parameter $\epsilon = 1 - \alpha$ we must solve:

$$\ln \epsilon = -\frac{1}{2N\epsilon^2}. \quad (98)$$

By change of variables ($\epsilon \to e^{z/2}$ and $N \to -1/c$) this equation can be manipulated into an equation of the form $ze^z = c$ whose solution is, by definition, the Lambert W function [47], $z = W(c)$. Consequently, the solution to Eq. (98) can be written in terms of the Lambert W function (the lower branch, W_{-1}):

$$\hat{\alpha}(N) \approx 1 - \exp\left[\frac{1}{2}W_{-1}\left(-\frac{1}{N}\right)\right] \quad (99)$$

for the optimal persistence level or, alternatively,

$$\hat{\epsilon}(N) \approx \exp\left[\frac{1}{2}W_{-1}\left(-\frac{1}{N}\right)\right]. \quad (100)$$

Equations (99) and (100), as solutions to the maximum information optimal diffusion problem, represent the key results of this section. To gain insight into the optimal solution and to connect it with the earlier results for autocovariance, note that the Lambert W function can be approximated by

$$W_{-1}(x) \approx \ln(-x) - \ln[-\ln(-x)] \quad (101)$$

when x is small [47]. Substituting Eq. (101) into Eq. (100) leads to the expression:

$$\hat{\epsilon}(N) = 1 - \hat{\alpha}(N) \approx \frac{1}{\sqrt{N \ln N}}. \quad (102)$$

It is noteworthy that in this approximation $\hat{\epsilon}(N)$ is a regularly varying function. In other words, we can write:

$$\hat{\epsilon}(N) \approx \left(\frac{1}{N}\right)^{\frac{1}{2}} g(N), \quad (103)$$

where

$$g(N) = \frac{1}{\sqrt{\ln N}} \quad (104)$$

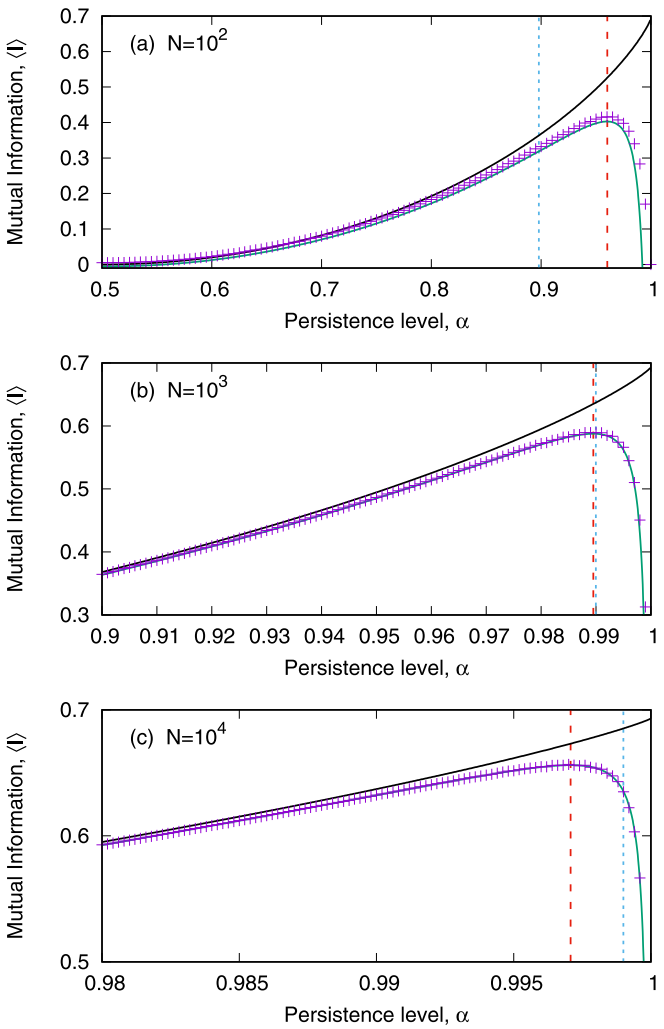

FIG. 6. Mutual information for increasing values of N. The approximation of Eq. (94) (solid green line) is compared to numerical simulation of the persistent random walk (purple markers). The infinite-N expression from Eq. (95) (convex black line) is included for reference. The optimal persistence level in Eq. (99) (dashed red vertical line) is included along with α_{CLT} from Eq. (15) (short dashed cyan line).

is a slowly varying function of the observation time N. If we treat the slowly varying function $g(N)$ as approximately constant, we recover the square-root scaling behavior:

$$\hat{\alpha}(N) \sim 1 - \frac{1}{\sqrt{N}}. \quad (105)$$

Note the mutual information objective function leads essentially to the same optimal strategy as the autocovariance objective function studied in Sec. IV. These optimal strategies have in common the scaling exponent $\nu = 1/2$. For mutual information this is no longer an exact scaling behavior, but an approximation $\nu \approx 1/2$ of the more complicated regularly varying function in Eq. (102). In practice the exponent ν appears slightly larger than $1/2$ because of the slow variation of the natural logarithm factor in Eq. (102) as N goes to infinity.

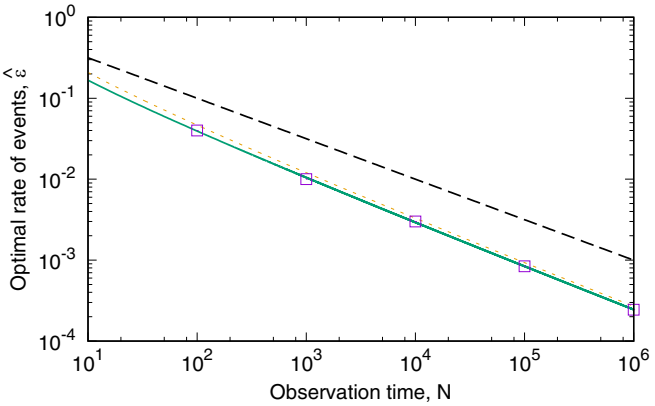

FIG. 7. Scaling behavior of the optimal rate of events that maximizes mutual information. The approximate optimal solution given by Eq. (100) (solid green line) is compared to numerical simulation of the persistent random walk (purple squares). The approximation of Eq. (102) (short dashed orange line) and the square-root scaling (dashed black line), Eq. (103) with $g(N) = 1$, are included for reference.

Figures 6 and 7 show that the results of the theoretical analysis for the mutual information agree with numerical simulations of the persistent random walk. The mutual information behaves in a way that is qualitatively similar to the autocorrelation of Fig. 4. The central limit theorem approximation of Eq. (94) becomes very good for persistence levels exceeding the optimal value $\hat{\alpha}$ as N increases toward infinity. This validates the theoretical result in Eq. (99) for the optimal persistence level when N is large.

VI. DISCUSSION

The optimal strategy for a persistent random walker to maximize correlation or mutual information is to switch directions at a rate which scales as one over the square root of the overall time duration N. This results in fluctuating trajectories that asymptotically are not biased toward either direction, yet retain a highly correlated motion. Since the optimal trajectories are diffusive the central limit theorem plays a key role in the analysis and may lead to universality of the optimal strategy, i.e., scaling behavior of optimal solutions with exponent $\nu = 1/2$, for a class of objective functions that includes both the mutual information and the autocovariance functions studied here. Objective functions associated with this class approximately have the form of Eq. (3) where local order increases linearly with the persistence level and long-range order is inversely proportional to the average number of events, in the dual limit of $N \to \infty$ and $\alpha \to 1$. Universality suggests that the results for the persistent random walk may be applicable to other correlated stochastic processes displaying a transition from ballistic to diffusive scaling.

The one-dimensional persistent random walk can be reinterpreted as a one-dimensional Ising model, which is an example application of the Hammersley-Clifford theorem or Markov-Gibbs equivalence [48]. Additionally, the autocovariance function $C(\alpha, N)$ defined in Eq. (17) can be mapped to an Ising-like Hamiltonian,

$$\mathcal{H} = -J \sum_{i=1}^{N-1} \sigma_i \sigma_{i+1} - K \left(\sum_{i=1}^{N} \sigma_i \right)^2, \tag{106}$$

where the σ_i are spin variables that replace the velocity values of the random walk. The first term represents the nearest-neighbor interaction energy, while the second term, proportional to the mean field squared, represents the energy cost associated with a nonvanishing mean field. The magnetic-field-squared term in the Hamiltonian is found also in Ising formulations of partition problems whose solutions involve subsets with an equal number of elements [49]. The problem of maximizing autocovariance in persistent random walk trajectories is transformed into the statistical mechanics problem of minimizing the free energy in the spin system. Note that with both terms present in the Hamiltonian, the system is a mixture of a one-dimensional Ising model (a spin chain) and a Curie-Weiss model (a fully connected or complete graph). When matching with the autocovariance of the persistent random walk, the interactions on the one-dimensional chain are ferromagnetic, $J > 0$, and these nearest-neighbor interactions compete with long-range antiferromagnetic interactions forming the complete graph. The map existing between the persistent random walk and the Ising model opens up a variety of analysis tools. For example, renormalization group analysis of persistent random motion can be understood in terms of the well-developed renormalization group theory for the Ising model and vice versa [48]. Studying the generic Ising-like Hamiltonian in Eq. (106) itself along with renormalization and the mapping to the persistent random walk may provide further insight into the maximum correlation optimization problem and the universality class of objective functions with index $\nu = 1/2$. It is worth noting that the persistent random-walk limit $\alpha \to 1$ is analogous to the zero-temperature limit in the Ising model, and the exponent of $1/2$ shows up also in the scaling analysis of the free energy density and correlation length for the zero temperature phase transition [50], when the reduced temperature variable is defined as $t = \exp(-4J/k_B T)$ from the energy cost associated with flipping a single spin in an ordered configuration.

The optimal strategy for maximizing correlation identified here appears to differ from existing results for optimal search and cover by persistent random walkers [14,51]; a persistence length on the order of the system size was found to be most efficient when searching for a target. However, it is difficult to make a fair comparison presently because we only addressed the simplest case of unbounded motion in one dimension, while the search and cover problems are often formulated in two- and three-dimensional bounded spaces. Since the central limit theorem plays a key role in our analysis, we expect that the scaling solution Eq. (10) with exponent $\nu = 1/2$ will hold also in higher dimensions if the space is not bounded. The argument is that the second moment of position, entering into the objective functions as in Eq. (39), for example, grows linearly in time for a normal diffusion process regardless of the dimension. Changing the dimension only changes the constant prefactor in Eq. (37). However, it would be important to check this argument through further analysis and numerical simulations in higher dimensions and modify the analysis

accordingly for bounded spaces, which at the same time would facilitate a better comparison between the different objectives of maximizing correlation and minimizing search and cover times.

To establish a more complete picture of optimal diffusion and universality classes of optimal solutions, it would be interesting to consider the maximum correlation problem for other types of random walks, including run-and-tumble motion [52], Lévy walks [53], and different multistate random walks [8,54], e.g., where the speed is no longer constant. The analysis could be extended to continuous time and space or to account for noisy measurements of the velocity. In general, optimization could act on the waiting-time distribution that governs the occurrence of events. For example, a truncated IPL distribution would be of particular interest because in that case an intermediate asymptotic scaling regime can arise where the diffusion process is anomalous. Lévy-type motion is often cited as a way to enhance search strategies [53], and one could check how it affects correlation and mutual information. Additionally, correlated random motion in more general environments could give insight into the connections between the maximum information protocols and optimal control. For example, the environment might contain obstacles, where it was shown that maximizing information can lead to an efficient exploration of space [22]. The scaling result we have found should provide a reference point for the more complicated random motion in random environments and perhaps could be useful in a perturbation theory or homogenization approach valid for environments with weak disorder.

Maximizing correlation and mutual information in a Markov process of finite time duration is likely to have applications beyond physical diffusion processes and the persistent random walk. For an example we mention recent developments in the thermodynamics of information engines. Correlations are a fuel source. However, an information engine must be designed in a way that allows it to extract work from correlations [55–57]. In particular, an information engine may be designed to extract work from a Markov chain of bits that contains correlations but no overall bias [57]. Therefore the optimal diffusion strategy might be useful also as a strategy for engineering a fuel source for certain types of information engines, since the protocol effectively maximizes correlation between adjacent entries in a finite tape without introducing an overall bias. This example further suggests that finite-time fluctuations in information-theoretic measures may have physical significance for engineering applications rather than being interpreted as a source of error in a statistical analysis with insufficient data.

APPENDIX

In this Appendix we discuss the Markov chain central limit theorem and variance calculations. We closely follow Refs. [58,59] and recommend consulting these references for additional details about Markov chains that are not included here.

Consider a regular Markov chain with r states $\{s_1, s_2, \ldots, s_r\}$ and an $r\times r$ transition matrix P. For a realization of the Markov chain, the state at time n is denoted by $S(n)$. A key notion of Markov chains is that the joint probability for the process to be in state s_i at time n and state s_j at a later time m is determined by iterating the transition matrix from the initial probability. That is, for the joint probability:

$$\Pr[S(n)=s_i, S(m)=s_j] = \Pr[S(n)=s_i]p_{ij}^{(m-n)}, \tag{A1}$$

where $p_{ij}^{(k)}$ stands for element (i, j) of the k-step transition matrix P^k. The Markov chain converges to equilibrium, such that:

$$\lim_{N\to\infty} P^N = W. \tag{A2}$$

Each row of the matrix W is equal to the equilibrium distribution for the r states or $W_{ij} = w_j$ where w_j is the equilibrium probability for the chain to be in state s_j.

The matrix product $M^k = (P-W)^k = P^k - W$ tends to the zero matrix as k goes to infinity, and the series $I + M + M^2 + \cdots$ converges to the matrix inverse $(I-M)^{-1}$. The fundamental matrix Z is defined:

$$Z = \sum_{k=0}^{\infty}(P-W)^k = (I-P+W)^{-1}, \tag{A3}$$

and is useful for calculating many asymptotic properties of the Markov chain. Note that I is the identity matrix. Importantly, the fundamental matrix is Cesaro summable:

$$Z = \lim_{N\to\infty}\sum_{k=0}^{N-1}\frac{N-k}{N}(P-W)^k, \tag{A4}$$

which can be written alternatively as:

$$Z = I + \lim_{N\to\infty}\sum_{k=1}^{N-1}\frac{N-k}{N}(P^k-W). \tag{A5}$$

A stochastic observable $A(N)$ of interest is the fraction of time that the chain spends in state s_i during the first N time steps. If N_i is the total number of time steps spent in state s_i up to time N, then:

$$A(N) = \frac{N_i}{N} = \frac{1}{N}\sum_{n=1}^{N}\chi_n, \tag{A6}$$

where χ_n is a random variable or indicator function defined by:

$$\chi_n = \begin{cases} 1 & \text{if } S(n)=s_i \\ 0 & \text{otherwise} \end{cases}. \tag{A7}$$

We are interested in calculating the limiting variance of $A(N)$. The variance is defined:

$$\mathrm{Var}[A(N)] = \langle A^2(N)\rangle - \langle A(N)\rangle^2. \tag{A8}$$

Substituting Eq. (A6) into Eq. (A8) yields:

$$\mathrm{Var}[A(N)] = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\langle\chi_n\chi_m\rangle - \left(\sum_{n=1}^{N}\langle\chi_n\rangle\right)^2, \tag{A9}$$

or, equivalently,

$$\mathrm{Var}[A(N)] = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\langle\chi_n\chi_m\rangle - w_i^2\right), \tag{A10}$$

after bringing the mean value inside the double sum. The correlation coefficient $\langle \chi_n \chi_m \rangle$ for the Markov chain is determined by the probability to be in state s_i at a certain time step, taken to be the equilibrium probability w_i, multiplied by the transition probability to return to state s_i in $k = |n - m|$ time steps:

$$\langle \chi_n \chi_m \rangle = w_i p_{ii}^{(k)}. \tag{A11}$$

Note that $p_{ii}^{(0)} = 1$ by definition. From Eq. (A11), the expression for the variance in Eq. (A10) becomes

$$\mathrm{Var}[A(N)] = \frac{w_i}{N}\left[(1 - w_i) + 2\sum_{k=1}^{N-1} \frac{N-k}{N}\left(p_{ii}^{(k)} - w_i\right)\right]. \tag{A12}$$

In the limit $N \to \infty$ the sum converges and is related to the fundamental matrix Z by Eq. (A5):

$$\lim_{N\to\infty} \sum_{k=1}^{N} \frac{N-k}{N}\left[p_{ii}^{(k)} - w_i\right] = z_{ii} - 1. \tag{A13}$$

The asymptotic variance is expressed in terms of the equilibrium probability w_i and the diagonal element z_{ii} of the fundamental matrix:

$$\mathrm{Var}[A(N)] = \frac{1}{N}\left(2w_i z_{ii} - w_i - w_i^2\right). \tag{A14}$$

For N large the variance can be approximated by a constant factor divided by N. We mention that the expression for the constant factor in Eq. (A14) is stated in Ref. [59] without any derivation. The complete derivation (with some typos) can be found in Ref. [58].

In summary the Gaussian or central limit theorem approximation for the fraction of time spent in state s_i is given by:

$$A(N) = \frac{N_i}{N} \approx w_i + \sqrt{\frac{2w_i z_{ii} - w_i - w_i^2}{N}}\,\xi, \tag{A15}$$

where $\xi \sim \mathcal{N}(0, 1)$ is a normalized Gaussian random variable.

Example 1. Two-state Markov chain. Let the transition matrix be given by:

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}. \tag{A16}$$

The row vector $w = [w_1 \quad w_2]$ of equilibrium probabilities that solves the equation:

$$wP = w \tag{A17}$$

under the normalization condition:

$$\sum_{i=1}^{2} w_i = 1 \tag{A18}$$

has the simple form:

$$[w_1 \quad w_2] = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}. \tag{A19}$$

The diagonal elements of the fundamental matrix, found by substituting Eq. (A16) and Eq. (A19) into Eq. (A3), are

$$Z_{11} = \frac{b(a+b)+a}{(a+b)^2}, \tag{A20}$$

$$Z_{22} = \frac{a(a+b)+b}{(a+b)^2}. \tag{A21}$$

From Eq. (A14), the variance for the fraction of time spent in state 1 for a trajectory of duration N is

$$\mathrm{Var}\left[\frac{N_1}{N}\right] = \frac{ab(2-a-b)}{(a+b)^3}\frac{1}{N}. \tag{A22}$$

Although the equilibrium probabilities of the two states may be different, the variance for the fraction of time spent in state 2 is the same as state 1, $\mathrm{Var}[\frac{N_2}{N}] = \mathrm{Var}[\frac{N_1}{N}]$. Therefore we have the central limit theorem approximation:

$$\frac{N_1}{N} \approx \frac{b}{a+b} + \sqrt{\frac{ab(2-a-b)}{(a+b)^3}\frac{1}{N}}\,\xi, \tag{A23}$$

$$\frac{N_2}{N} \approx \frac{a}{a+b} + \sqrt{\frac{ab(2-a-b)}{(a+b)^3}\frac{1}{N}}\,\xi, \tag{A24}$$

where $\xi \sim \mathcal{N}(0, 1)$.

The transition matrix for the persistent random-walk Markov chain, Fig. 1, has the form:

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix}. \tag{A25}$$

Letting $a = b = 1 - \alpha$ in the above equations for the general two-state Markov chain, we end up with:

$$\mathrm{Var}\left[\frac{N_1}{N}\right] = \mathrm{Var}\left[\frac{N_2}{N}\right] = \frac{1}{4}\frac{\alpha}{1-\alpha}\frac{1}{N}. \tag{A26}$$

Recall that the time-averaged velocity:

$$\overline{v}(N) = \frac{N_1 - N_2}{N} \tag{A27}$$

is an observable of interest that was introduced in Eq. (19). Its variance in terms of the residence times is given by:

$$\mathrm{Var}[\overline{v}(N)] = \mathrm{Var}\left[\frac{N_1}{N}\right] + \mathrm{Var}\left[\frac{N_2}{N}\right] - 2\mathrm{Cov}\left[\frac{N_1}{N}, \frac{N_2}{N}\right]. \tag{A28}$$

Since $\frac{N_2}{N} = 1 - \frac{N_1}{N}$, the covariance between residence times takes the form:

$$\mathrm{Cov}\left[\frac{N_1}{N}, \frac{N_2}{N}\right] = -\mathrm{Var}\left[\frac{N_1}{N}\right]. \tag{A29}$$

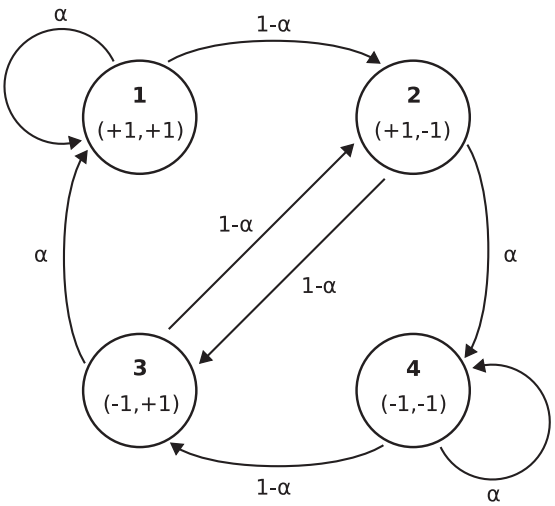

FIG. 8. Four-state Markov chain diagram representing the transitions between each velocity pair (v_n, v_{n+1}).

Therefore the asymptotic variance of the time-averaged velocity is

$$\mathrm{Var}[\overline{v}(N)] = \langle \overline{v}^2(N) \rangle = \frac{\alpha}{1-\alpha}\frac{1}{N}. \tag{A30}$$

This recovers the earlier result in Eq. (32) by way of the fundamental matrix in the Markov chain representation of the persistent random-walk velocity trajectory.

Example 2. Four-state Markov chain. Here we discuss the four-state Markov chain depicted in Fig. 8. Each state represents one of the possible velocity pairs (v_n, v_{n+1}): State 1 corresponds to $(v_n = +1, v_{n+1} = +1)$, state 2 corresponds to $(v_n = +1, v_{n+1} = -1)$, state 3 corresponds to $(v_n = -1, v_{n+1} = +1)$, and state 4 corresponds to $(v_n = -1, v_{n+1} = -1)$. From the two-state Markov chain for the persistent random-walk velocity trajectory we can construct the four-state Markov chain for consecutive velocity pairs by moving a window of size two along the original velocity trajectory. For example, the velocity trajectory $[+1, +1, -1, -1, +1]$ becomes $[(+1, +1), (+1, -1), (-1, -1), (-1, +1)]$, and the four-state Markov chain follows the transition sequence $1 \to 2 \to 4 \to 3$. The 4×4 transition matrix for this Markov chain has the form:

$$P_4 = \begin{bmatrix} \alpha & 1-\alpha & 0 & 0 \\ 0 & 0 & 1-\alpha & \alpha \\ \alpha & 1-\alpha & 0 & 0 \\ 0 & 0 & 1-\alpha & \alpha \end{bmatrix}. \tag{A31}$$

The row vector $w = [w_1 \quad w_2 \quad w_3 \quad w_4]$ of equilibrium probabilities is defined by the equilibrium condition:

$$wP_4 = w. \tag{A32}$$

The equilibrium probabilities can be found by solving the above system of equations supplemented with the normalization condition:

$$\sum_{i=1}^{4} w_i = 1, \tag{A33}$$

and the symmetry conditions $w_1 = w_4$ and $w_2 = w_3$ for an unbiased process. We find the values:

$$w = \left[\frac{\alpha}{2} \quad \frac{1-\alpha}{2} \quad \frac{1-\alpha}{2} \quad \frac{\alpha}{2}\right]. \tag{A34}$$

From P_4 and w the fundamental matrix can be constructed according to Eq. (A3). For its diagonal elements we find:

$$Z_{11} = Z_{44} = \frac{3\alpha - 4}{4(\alpha - 1)}, \tag{A35}$$

$$Z_{22} = Z_{33} = \frac{3}{4}. \tag{A36}$$

From Eq. (A14) the asymptotic variance for the fraction of time spent by the process in each of the states is

$$\mathrm{Var}\left[\frac{N_1}{N}\right] = \mathrm{Var}\left[\frac{N_4}{N}\right] = \frac{\alpha(\alpha^2 - 2\alpha + 2)}{4(1-\alpha)} \tag{A37}$$

and

$$\mathrm{Var}\left[\frac{N_2}{N}\right] = \mathrm{Var}\left[\frac{N_3}{N}\right] = \frac{\alpha(1-\alpha)}{4}. \tag{A38}$$

The results for the equilibrium vector, Eq. (A34), and asymptotic variances, Eqs. (A37) and (A38), were used in the calculation of the time-lagged mutual information in Sec. V.

[1] G. H. Weiss, Physica A **311**, 381 (2002).
[2] G. I. Taylor, Proc. Lond. Math. Soc. **20**, 196 (1922).
[3] Z. Sadjadi, M. R. Shaebani, H. Rieger, and L. Santen, Phys. Rev. E **91**, 062715 (2015).
[4] D. Escaff, R. Toral, C. Van den Broeck, and K. Lindenberg, Chaos **28**, 075507 (2018).
[5] M. Boguñá, J. M. Porrà, and J. Masoliver, Phys. Rev. E **59**, 6517 (1999).
[6] M. F. Miri and H. Stark, Phys. Rev. E **68**, 031102 (2003).
[7] Z. Sadjadi, M. F. Miri, M. R. Shaebani, and S. Nakhaee, Phys. Rev. E **78**, 031121 (2008).
[8] G. H. Weiss, *Aspects and Applications of the Random Walk* (North-Holland, Amsterdam, 1994).
[9] M. Kac, Rocky Mt. J. Math. **4**, 497 (1974).
[10] T. Gilbert and D. P. Sanders, J. Phys. A: Math. Theor. **43**, 035001 (2010).
[11] H. Tennekes and J. L. Lumley, *A First Course in Turbulence* (The MIT Press, Cambridge, MA, 1972).
[12] D. Campos, E. Abad, V. Méndez, S. B. Yuste, and K. Lindenberg, Phys. Rev. E **91**, 052115 (2015).
[13] O. Bénichou, C. Loverdo, M. Moreau, and R. Voituriez, Rev. Mod. Phys. **83**, 81 (2011).
[14] V. Tejedor, R. Voituriez, and O. Bénichou, Phys. Rev. Lett. **108**, 088103 (2012).
[15] S. N. Majumdar, A. Comtet, and J. Randon-Furling, J. Stat. Phys. **138**, 955 (2010).
[16] M. Luković, T. Geisel, and S. Eule, New J. Phys. **15**, 063034 (2013).

[17] S. Jafarizadeh, IEEE Control Syst. Lett. **2**, 465 (2018).
[18] I. A. Martínez, A. Petrosyan, D. Guéry-Odelin, E. Trizac, and S. Ciliberto, Nat. Phys. **12**, 843 (2016).
[19] M. Chupeau, B. Besga, D. Guéry-Odelin, E. Trizac, A. Petrosyan, and S. Ciliberto, Phys. Rev. E **98**, 010104(R) (2018).
[20] N. Ay, H. Bernigau, R. Der, and M. Prokopenko, Theory Biosci. **131**, 161 (2012).
[21] G. Martius, R. Der, and N. Ay, PLoS ONE **8**, e63400 (2013).
[22] N. Ay, N. Bertschinger, R. Der, F. Güttler, and E. Olbrich, Eur. Phys. J. B **63**, 329 (2008).
[23] F. Lombardi, H. J. Herrmann, D. Plenz, and L. de Arcangelis, Sci. Rep. **6**, 24690 (2016).
[24] C. Meisel, K. Bailey, P. Achermann, and D. Plenz, Sci. Rep. **7**, 11825 (2017).
[25] M. Luković, F. Vanni, A. Svenkeson, and P. Grigolini, Physica A **416**, 430 (2014).
[26] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (CRC Press, Boca Raton, FL, 2018).
[27] T. Bossomaier, L. Barnett, and M. Harré, Complex Adapt. Syst. Model. **1**, 9 (2013).
[28] R. T. Wicks, S. C. Chapman, and R. O. Dendy, Phys. Rev. E **75**, 051125 (2007).
[29] S.-J. Gu, C.-P. Sun, and H.-Q. Lin, J. Phys. A: Math. Theor. **41**, 025002 (2008).
[30] H. Matsuda, K. Kudo, R. Nakamura, O. Yamakawa, and T. Murata, Int. J. Theor. Phys. **35**, 839 (1996).
[31] M. Kahlen, A. Engel, and C. Van den Broeck, Phys. Rev. E **95**, 012144 (2017).
[32] H. Touchette, Phys. Rep. **478**, 1 (2009).
[33] H. Touchette, in *Modern Computational Science 11: Lecture Notes from the 3rd International Oldenburg Summer School*, edited by R. Leidl and A. K. Hartmann (University of Oldenburg Press, Oldenburg, 2011).
[34] V. Méndez, D. Campos, and F. Bartumeus, *Stochastic Foundations in Movement Ecology* (Springer, Heidelberg, 2014).
[35] W. Bialek, I. Nemenman, and N. Tishby, Neural Comput. **13**, 2409 (2001).
[36] S. Kullback, *Information Theory and Statistics* (Dover, Mineola, NY, 1959).
[37] F. von Wegner, H. Laufs, and E. Tagliazucchi, Phys. Rev. E **97**, 022415 (2018).
[38] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[39] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[40] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[41] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[42] B. Harris, Colloq. Math. Societatis Janos Bolyai **16**, 323 (1975).
[43] W. Li, J. Stat. Phys. **60**, 823 (1990).
[44] M. S. Roulston, Physica D **125**, 285 (1999).
[45] J. A. Bonachela, H. Hinrichsen, and M. A. Muñoz, J. Phys. A: Math. Theor. **41**, 202001 (2008).
[46] A. Lesne, J.-L. Blanc, and L. Pezard, Phys. Rev. E **79**, 046208 (2009).
[47] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, Adv. Comput. Mech. **5**, 329 (1996).
[48] M. Yin, Physica A **392**, 1347 (2013).
[49] A. Lucas, Front. Phys. **2**, 5 (2014).
[50] R. K. Pathria and P. D. Beale, *Statistical Mechanics* (Butterworth-Heinemann, Oxford, UK, 2011).
[51] M. Chupeau, O. Bénichou, and R. Voituriez, Nat. Phys. **11**, 844 (2015).
[52] J.-F. Rupprecht, O. Bénichou, and R. Voituriez, Phys. Rev. E **94**, 012117 (2016).
[53] V. Zaburdaev, S. Denisov, and J. Klafter, Rev. Mod. Phys. **87**, 483 (2015).
[54] J. Masoliver and K. Lindenberg, Eur. Phys. J. B **90**, 107 (2017).
[55] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Phys. Rev. X **8**, 031036 (2018).
[56] E. Stopnitzky, S. Still, T. E. Ouldridge, and L. Altenberg, Phys. Rev. E **99**, 042115 (2019).
[57] R. A. Brittain, N. S. Jones, and T. E. Ouldridge, New J. Phys. **21**, 063022 (2019).
[58] J. G. Kemeny and J. Laurie Snell, *Finite Markov Chains* (D. Van Nostrand Company, Inc., Princeton, NJ, 1960).
[59] C. M. Grinstead and J. Laurie Snell, *Introduction to Probability* (American Mathematical Society, Providence, RI, 1997).