Giant leaps and long excursions: fluctuation mechanisms in systems with long-range memory

Robert L. Jack[1,2] and Rosemary J. Harris[3]

[1]*Department of Applied Mathematics and Theoretical Physics,*
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
[2]*Department of Chemistry, University of Cambridge,*
Lensfield Road, Cambridge CB2 1EW, United Kingdom
[3]*School of Mathematical Sciences, Queen Mary University of London,*
Mile End Road, London E1 4NS, United Kingdom

We analyse large deviations of time-averaged quantities in stochastic processes with long-range memory, where the dynamics at time t depends itself on the value q_t of the time-averaged quantity. First we consider the elephant random walk and a Gaussian variant of this model, identifying two mechanisms for unusual fluctuation behaviour, which differ from the Markovian case. In particular, the memory can lead to large-deviation principles with reduced speeds, and to non-analytic rate functions. We then explain how the mechanisms operating in these two models are generic for memory-dependent dynamics and show other examples including a non-Markovian simple exclusion process.

I. INTRODUCTION

Memory effects and long-range temporal correlations are important in many physical systems [1–5], and in other scientific fields ranging from biology to telecommunications to finance [6, 7]. It is particularly notable that long-ranged memory can change fluctuation behaviour qualitatively, compared to Markovian (memory-less) cases. Demonstrations of this include non-Markovian random walks [8–13], models of cluster growth [14–16], and agent-based models where decisions depend on past experience [17]. The distinction between Markovian and non-Markovian systems is also important when formulating general theories. For example, a large deviation theory of dynamical fluctuations is now established for Markovian systems [18–23], but memory can lead to new effects which cannot be captured by the standard theory [16, 24–27]. In particular, one finds [24, 25] a breakdown of the standard large-deviation principles (LDPs) that hold quite generically in finite Markovian systems [22].

In this work, we consider non-Markovian systems where the dynamics depend explicitly on a time-averaged current, whose value at time t is denoted by q_t. This is a simple type of memory that occurs in a wide range of physical models [8, 12, 16, 17, 24, 25]. Using methods of large-deviation theory [18–22, 28], we show how this long-range memory can lead to anomalous fluctuations of q_t. We explain that much of this behaviour can be understood by considering two generic fluctuation mechanisms, where memory plays an intrinsic role. These general mechanisms are useful for classifying previous results for non-Markovian systems, and for identifying new phenomena.

We illustrate these mechanisms by analysing current fluctuations in the elephant random walk (ERW) of [8, 10, 29, 30], and a related process which we call the Gaussian elephant random walk (GERW). A key differ-

ence from Markovian systems is that large (rare) fluctuations in these models are associated with currents that are strongly time-dependent [16, 24, 25, 28] – a large current at early times biases the subsequent evolution and can trigger anomalous fluctuations that persist for large times. The two specific mechanisms that we discuss are: (i) a very large initial current flow in a finite time interval, which results in anomalously large deviations (specifically, an LDP for q_t with a speed that is less than t [24, 25]); and (ii) a large initial current that occurs over a sustained time interval, which leads to a breakdown of the central limit theorem (CLT) for q_t [8] and an LDP which generically has a non-analytic rate function [16]. The GERW illustrates mechanism (i), which we refer to as an *initial giant leap* (IGL); the ERW illustrates mechanism (ii) which we call a *long initial excursion* (LIE). We also describe several other examples of systems in which these mechanisms occur.

The structure of the paper is as follows. Sec. II introduces the models that we analyse, and some relevant theory. In Sec. III we describe the large-deviation behaviour of the ERW and GERW models. Sec. IV gives a general theory for the IGL and LIE mechanisms and Sec. V describes how this theory plays out in several other models, to illustrate its applicability. Sec. VI gives a summary of the main conclusions and open questions. Additional details of calculations are given in Appendices.

II. MODELS AND METHODS

A. Definitions of ERW and GERW

The ERW is a random walk model, in discrete time [8, 9]. The position of the elephant after step t is x_t. In the variant of the model that we consider here, x_t takes values in a finite domain $\{0, 1, \ldots, L-1\}$, with periodic boundaries. (The choice of periodic boundaries does not

change the physical behaviour but it is useful when comparing the large-deviation behaviour with that of Markovian systems, see Sec. II B. In the following we do not distinguish our periodic variant from the original ERW, except in the rare cases where this is necessary.) We take $x_0 = 0$ and denote the displacement of the elephant on step t by Δx_t. Hence the time-averaged current is

$$q_t = \frac{1}{t} \sum_{\tau=1}^{t} \Delta x_\tau \qquad (1)$$

with $q_0 = 0$. The dynamical rule is that

$$\Delta x_t = \pm 1 \quad \text{with probability} \quad \frac{1 \pm a q_{t-1}}{2}, \qquad (2)$$

where $a \in (-1, 1)$ is a parameter that corresponds to $2p - 1$ in [8].

The GERW is similar to the ERW, except that the position x_t is a real number in $[0, L)$, still with periodic boundaries. The dynamical rule is that Δx_t is a Gaussian-distributed real number with mean $a q_{t-1}$ and variance unity. Hence the mean value of Δx_t (conditional on q_{t-1}) is the same as for the ERW. At first glance, one might also expect fluctuations in the two models to behave similarly but in fact their memory-induced large-deviation behaviour is very different. Note that although the GERW is rather simple to analyse, it is useful to study in detail as a contrast to the ERW and to illustrate the strong effects of memory.

B. Large deviations for Markovian and non-Markovian dynamics

For the ERW and GERW, we consider the probability density for q_t at large times, which we denote by $p_t(q)$. We will be particularly concerned with the tails of this probability distribution and the associated fluctuation mechanisms, which are characterised by *large deviation theory* [18–22]. This theory describes rare fluctuations, outside the range of CLTs and their generalisations [10]. In recent years, it has been applied to time-averaged quantities in many physical systems, yielding important new insights [20, 31–34].

For finite Markov chains, there is a well-established large deviation theory due originally to Donsker and Varadhan (DV) [35–38], see for example [23, 39] for recent summaries. Within this theory, time-averaged quantities such as q_t obey LDPs of the form

$$p_t(q) \simeq \exp[-tI(q)], \qquad (3)$$

where t is called the *speed* of the LDP, and I the *rate function*. More generally, one may also consider LDPs of the form

$$p_t(q) \simeq \exp\left[-t^\theta I(q)\right]. \qquad (4)$$

with $\theta \neq 1$. In some of the non-Markovian models considered here we find $0 < \theta < 1$ so the speed t^θ is reduced, compared to the Markovian case. Physically, this means that the memory makes large fluctuations less rare [24, 25]. See [40–42] for some other examples where LDPs with reduced speed are associated with enhanced fluctuations.

For large deviations, an important quantity is the cumulant generating function for q_t:

$$G(\lambda, t) = \log\langle e^{\lambda t q_t} \rangle . \qquad (5)$$

To analyse the limit of large t, we consider the scaled cumulant generating function (SCGF) which can be defined generally for LDPs with speed t^θ:

$$\psi_\theta(\lambda) = \lim_{t \to \infty} \frac{1}{t^\theta} \log\langle e^{\lambda t^\theta q_t} \rangle . \qquad (6)$$

For the usual case $\theta = 1$ we omit the subscript and write simply $\psi(\lambda)$. If the limit (6) exists and certain other technical conditions are met then the Gartner-Ellis theorem states that the LDP (4) holds with

$$I(q) = \sup_\lambda [\lambda q - \psi_\theta(\lambda)] . \qquad (7)$$

The classical (DV) theory deals with LDPs of speed t. Under suitable assumptions, the rate function can be shown to be analytic and strictly convex. For processes on discrete state spaces, it is sufficient that (i) the model is Markovian; (ii) transition rates (or transition probabilities) are independent of time; (iii) the model is finite and irreducible (and, for discrete-time systems, aperiodic); (iv) the contribution of each transition to the sum in (1) is fully determined by its initial and final state. For the ERW, condition (i) is violated, but the other conditions still hold. [The ERW was defined on a finite (periodic) domain so that assumption (iii) is valid.] This enables a clear comparison with the classical theory. The striking result of this comparison is that the memory effect in the ERW leads to a rate function that is (generically) singular at $q = 0$, as we show below. Such behaviour is strictly forbidden in the classical case and is directly attributable to the memory effect, via the breaking of assumption (i).

By contrast, the GERW does not allow such a clear comparison with the classical case. The model is defined on a compact domain, and assumption (iii) can be generalised to account for this, while still ensuring an analytic rate function. However, the GERW allows for jumps with $|\Delta x_t| > L$, in which case the contribution to (1) is not fully-determined by the initial and final states (due to the periodic boundaries). In principle, assumption (iv) might be generalised to account for this effect, but the memory effect means that the typical jump size can diverge as $q_t \to \infty$, which would not be allowed in the classical case. In this sense, the GERW violates the classical assumptions more strongly than the ERW. We show below that this strong violation can lead to an LDP with reduced speed, $\theta < 1$.

In fact, the large-deviation behaviour that we will observe also has implications for a particular class of *Markovian* models. To see this, note that both the ERW and GERW can be formulated as Markovian models for either the current q_t or (equivalently) the displacement $Q_t = tq_t$. The case of the displacement is more natural: in this case the dynamical rule of the ERW is that $Q_{t+1} = Q_t \pm 1$ with probabilities $(1 \pm aQ_t/t)/2$. In this formulation, one sees that the transition probabilities depend explicitly on time. The GERW may be formulated in a similar way. Representing the models in this way, the Markovian assumption (i) above is now obeyed, but assumption (ii) is violated. It follows that the behaviour presented here can be viewed as either an extension of the classical theory to a particular class of non-Markovian models, or an extension to a class of Markovian models with explicit time-dependence in the rates.

In terms of methods, it is notable that the SCGF $\psi(\lambda)$ in the classical theory can be characterised as the largest eigenvalue of a matrix, the tilted generator [22]. Hence the rate function is available, via (7). Such a determination of $\psi(\lambda)$ is not possible for the models considered here, and other methods must be used, for example the theory of Dupuis-Ellis [43] as in [28], or arguments based on separation of time scales [24, 25].

We close this section by noting that since the (G)ERW models are defined on finite periodic domains, they *cannot* be formulated as Markovian processes for x_t, even with time-dependent rates. The transition rates (or transition probabilities) would need to depend on the winding number around the periodic boundaries.

C. General models

Although we use the ERW and GERW as motivating examples, we stress that the mechanisms they illuminate have much wider applicability. To this end, we introduce here a general notation for describing a broad class of models with similar memory-induced phemenonlogy. Several examples are given in Sec. V.

We consider models in which the time t may be continuous or discrete. Let $\mathcal{C}_t$ denote the configuration of the (general) model at time t, this corresponds to x_t in the (G)ERW. This $\mathcal{C}_t$ may come from a finite set as in the periodic ERW, or it may indicate a vector in some compact domain such as $[0, L]^d$. The periodic GERW is in this latter class with $d = 1$.

All models considered are jump processes. We define a time-averaged quantity that generalises (1):

$$q_t = \frac{1}{t} \sum_{\text{jumps } j} \alpha_j \qquad (8)$$

where the sum is over all jumps up to time t and α_j depends on the properties of jump j. (In the ERW there is one jump on each time step and $\alpha_j = \pm 1$ coincides with Δx_t.) In discrete time the model is specified by the conditional distribution of $\mathcal{C}_{t+1}$ given $(\mathcal{C}_t, q_t)$, supplemented by a rule specifying the contribution α_j for each jump. In continuous time the model is specified by a set of jump rates [dependent on $(\mathcal{C}_t, q_t)$], and a rule specifying the α_j. We assume that the dynamical rules do not depend explicitly on time, but only on $(\mathcal{C}_t, q_t)$.

All the results that we present can be straightforwardly generalised to the case where q_t is a time-average of a state-dependent quantity (for example $\frac{1}{t} \int_0^t b(\mathcal{C}_\tau)d\tau$ as in [23]) but we restrict here to the form (8). The models that we consider have scalar q_t but the analysis is easily extended to vectorial q_t. For continuous-time models then some regularisation may be required for (8) at short times, see for example Sec. V A.

Consistent with Sec. II B we observe that these general models can be formulated as Markov processes $(\mathcal{C}_t, q_t)$ but they are not Markovian for $\mathcal{C}_t$. Independent of this mathematical distinction, the physical role of q is to capture the role of memory: its definition depends on the full history of the process.

III. FLUCTUATIONS IN THE GERW AND ERW

In this section we describe the large-deviation behaviour of (G)ERW models. For the ERW we draw on results of [8, 10, 28] and we characterise the relevant fluctuation mechanisms. For the GERW then $p_t(q)$ can be computed quite straightforwardly and leads to an LDP with reduced speed, we describe the relevant fluctuation mechanisms in this case too.

A. Preliminary results

We summarise here some preliminary results for the ERW and GERW with further detail given in Appendices A and B. For $0 < a < 1/2$ the memory is relatively weak in these models and at large times they both obey a CLT, where the variance behaves asymptotically as

$$\langle q_t^2 \rangle \simeq \frac{1}{t(1 - 2a)}. \qquad (9)$$

Our work focusses on $a > 1/2$ where the memory effect is strong, and both ERW and GERW exhibit superdiffusive behaviour:

$$\langle q_t^2 \rangle \simeq \frac{\chi}{t^{2(1-a)}} \qquad (10)$$

where χ is an a-dependent constant which we denote by χ_{E} and χ_{G} for the two models. For the ERW we have $\chi_{\mathrm{E}} = 1/[(2a-1)\Gamma(2a)]$ from [8].

For the GERW the total displacement is a sum of Gaussian-distributed increments so $p_t(q)$ is Gaussian at all times (although the increments are neither independent nor identically distributed). As shown in Appendix A, for large times one has

$$p_t(q) \propto \exp\left[-\frac{q^2 t^{2(1-a)}}{2\chi_{\mathrm{G}}}\right], \qquad (11)$$

where χ_G can be obtained from the limit of a series solution. (The proportionality sign is used because a t-dependent normalisation constant has been omitted. We use this notation in cases where the normalisation is clear from the context.) For large times, the corresponding CGF is

$$G(\lambda,t) \simeq \frac{\lambda^2 t^{2a}\chi_G}{2} \ . \tag{12}$$

We now turn to the ERW. Since the second derivative of the CGF gives the variance of tq_t (and using that the distribution is symmetric) one obtains from (10) an expansion in powers of λ (at fixed $t \gg 1$):

$$G(\lambda,t) \simeq \frac{\lambda^2 t^{2a}\chi_E}{2} + O(\lambda^4) \ , \tag{13}$$

which is similar to (12). However, there is no CLT for q_t [44, 45] which has consequences for the correction terms in (13) and their scaling with t.

Baur and Bertoin [10] considered typical fluctuations of q at large times (that is, fluctuations with probabilities of order unity). Their theorem 3 states that $p_t(q)$ is a scaling function of qt^{1-a}, as in the GERW. Hence, G is described by a scaling form at large t

$$G(\lambda,t) \simeq g(\lambda t^a) \ , \tag{14}$$

which holds as $t \to \infty$, with the argument of g held fixed. In a recent mathematical study, Franchini [28] considered large deviations in Pólya urn models, which can be mapped onto the ERW [10]. Corollary 12 of [28] establishes that $p_t(q)$ follows an LDP, and that (14) extends into the large-deviation regime: taking $t \to \infty$ with fixed $\lambda \ll 1$, one has

$$g(\lambda t^a) \simeq c_E t|\lambda|^{1/a} \tag{15}$$

for some constant c_E (dependent on a). This result is discussed in Appendix B and a formula for c_E is given in (B10). Hence for $q \ll 1$ and $t \to \infty$ one has the large-deviation result

$$p_t(q) \propto \exp\left(-\kappa_E t|q|^{1/(1-a)}\right) \ . \tag{16}$$

with $\kappa_E = (1-a)(a/c_E)^{a/(1-a)}$. For larger q there are deviations from the scaling form; the full SCGF is given in (B4), as derived in [28]. Equ. (16) is an LDP with speed t and rate function $I(q) \simeq \kappa_E|q|^{1/(1-a)}$. As advertised above, the long-ranged memory has resulted in a rate function that is non-analytic (at $q = 0$), except in exceptional cases where $1/(1-a)$ happens to be an even integer. From a physical perspective, note that if q_t obeyed a CLT then its asymptotic variance would be $1/I''(0)$: here we have $I''(0) = 0$, which shows that the scaling is superdiffusive, consistent with (10).

We emphasise that the distributions (11,16) are sharply peaked as $t \to \infty$. In this sense, both systems are ergodic [11].
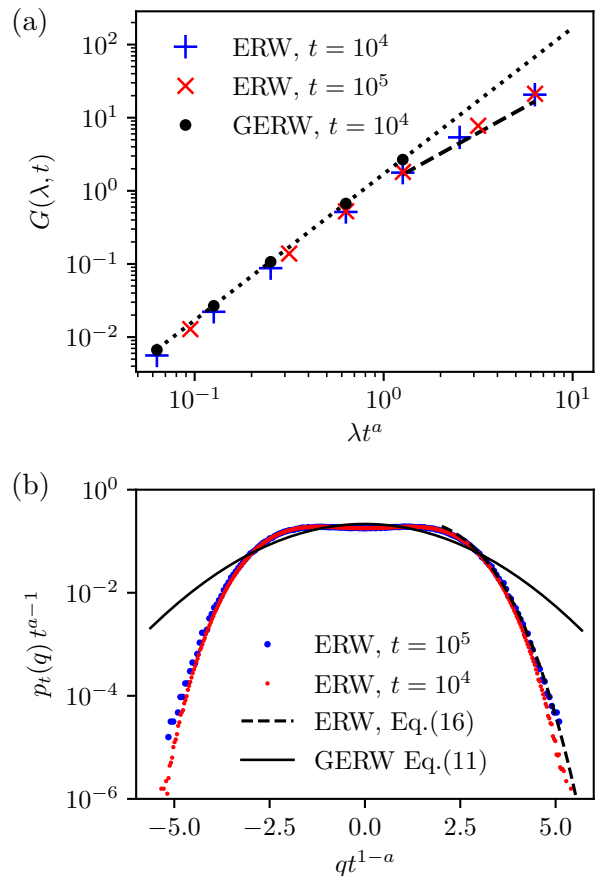

FIG. 1. Numerical data for ERW and GERW with $a = 0.7$. (a) CGF for ERW and GERW, plotted as a scaling function of λt^a. The dotted line is the analytical (large-t) GERW result (12) with χ_G obtained from Appendix A. The dashed line corresponds to (15) and agrees well with the data, given that it is an asymptotic prediction that assumes that both t and λt^a are large. There are no fitting parameters. (b) Distribution of q for the ERW, which collapses to a scaling function of qt^{1-a}, as predicted by [10]. (The collapse is not quite perfect, which we attribute to finite-t corrections to scaling.) The solid line is the analytical (large-t) Gaussian distribution of the GERW, for comparison; the dashed line is the prediction (16) for the tail of the distribution; the constant κ_E is derived using results from Appendix B while the proportionality constant is determined by fitting to the data.

Fig. 1 shows numerical data, which illustrates these preliminary results. For small values of λt^a the CGF for the ERW is proportional to $|\lambda t^a|^2$ and matches the large-t GERW result. For larger λt^a, the CGF for the ERW matches the large-deviation form (15) without any fitting [the value of c_E is given in (B10)]. For both ERW and GERW, the distribution $p_t(q)$ is a scaling function of qt^{1-a}. The ERW result (16) is shown in Fig. 1(b) with a dashed line: the value for κ_E is derived from c_E but the proportionality constant in (16) is used as a fitting parameter.

We close this section with a result for conditional av-

erages. The models have fixed initial conditions and averages over the dynamics are denoted by $\langle \cdot \rangle$. Define also $\langle \cdot \rangle_{q_\tau}$ as an average that is conditioned on the value of q_τ. For both GERW and ERW, averaging over the possibilities in a single step gives

$$(\tau + 1)\langle q_{\tau+1} \rangle_{q_\tau} = (\tau + a)q_\tau . \tag{17}$$

It follows [8] that for $t > \tau$,

$$\langle q_t \rangle_{q_\tau} = q_\tau \frac{\Gamma(t+a)\Gamma(\tau+1)}{\Gamma(t+1)\Gamma(\tau+a)} . \tag{18}$$

Hence for large t :

$$\langle q_t \rangle_{q_\tau} = \eta \frac{q_\tau}{t^{1-a}} . \tag{19}$$

with $\eta = \Gamma(\tau+1)/\Gamma(\tau+a)$. That is, if the elephant is conditioned to have a non-typical value of q_τ, its subsequent evolution involves regression to the mean (zero) as a power law with exponent $1 - a$. This result will be used in the following to rationalise the large-deviation behaviour of these models. As a point of comparison, time-averaged quantities in finite Markovian systems generically have power-law relaxation with exponent 1.

B. IGL mechanism for large deviations in the GERW

We now analyse large deviations for the GERW, in detail. Consider a discrete-time trajectory with t steps which we represent using its sequence of increments: $\boldsymbol{X} = (\Delta x_1, \Delta x_2, \ldots, \Delta x_t)$. This trajectory occurs with probability $P(\boldsymbol{X})$ which is a multivariate Gaussian distribution, so all correlations can be computed exactly (at fixed t). Specifically,

$$P(\boldsymbol{X}) \propto \exp[-\mathcal{S}(\boldsymbol{X})/2] \tag{20}$$

with

$$\mathcal{S}(\boldsymbol{X}) = \sum_{\tau=0}^{t-1} [(\tau+1)q_{\tau+1} - q_\tau(\tau+a)]^2 \tag{21}$$

where we recall that q_τ is related to the increments Δx by (1), with $q_0 = 0$.

To characterise large deviations, the most likely path that achieves $q_t = q$ can be derived, by conditioning $P(\boldsymbol{X})$ on this rare event. Collecting terms in (21), one obtains

$$\mathcal{S}(\boldsymbol{X}) = t^2 q_t^2 + \sum_{\tau=1}^{t-1} q_\tau^2 \left(2\tau^2 + a^2 + 2a\tau\right)$$
$$- 2\sum_{\tau=1}^{t-1} q_\tau q_{\tau+1}(\tau+1)(\tau+a) . \tag{22}$$

Conditioning on q_t, we arrive at a Gaussian distribution for the $(t-1)$-dimensional vector $\boldsymbol{q} = (q_1, q_2, \ldots, q_{t-1})$. This is

$$P_{\text{micro}}(\boldsymbol{q}|q_t) \propto \exp\left(hq_t q_{t-1} - \frac{\boldsymbol{q}^T M \boldsymbol{q}}{2}\right) , \tag{23}$$

where $h = t(t+a-1)$, and M is a matrix whose elements can be read from (22). The subscript "micro" recalls that conditioning on $q_t = q$ is analogous to considering a microcanonical ensemble in thermodynamics. Completing the square in the exponent, one obtains

$$P_{\text{micro}}(\boldsymbol{q}|q_t) \propto \exp\left[-\frac{(\boldsymbol{q} - hq_t\boldsymbol{\mu})^T M (\boldsymbol{q} - hq_t\boldsymbol{\mu})}{2}\right] \tag{24}$$

where $\boldsymbol{\mu}$ is given by the $(t-1)$th column of M^{-1}. Hence the most likely path with $q_t = q$ is given by

$$\langle q_\tau \rangle_{\text{micro}} = \mu_\tau q h . \tag{25}$$

This path depends on the value of q_t and on the associated time t. For finite t, the path can be straightforwardly computed numerically (for all $\tau < t$).

It is also possible to construct an optimally-controlled process (or auxiliary process) whose typical dynamics generate the most likely path to $q_t = q$. This is similar to the Doob-transformed dynamics of [22, 46], see also [47–51] and Appendix C 1 of this work for the general theory. For the GERW, we have derived the optimally-controlled process, see Appendix C 2. Its average path is $\langle q_\tau \rangle_{\text{con}} = \langle q_\tau \rangle_{\text{micro}}$ (for $\tau = 1, 2, \ldots, t$). Fig. 2(a) shows results illustrating the average path under the controlled dynamics, also compared with the ERW (see below). The mechanism for achieving a rare value of q_t is that the GERW makes very large hops on the first few steps, after which q_τ decreases towards q_t.

The results so far are valid for any finite time but we are interested in large deviations as $t \to \infty$. In this case the problem may be simplified. We characterise the most likely path as the minimum of the exponent in (24). Writing the matrix product as a sum over time steps [similar to (21)], we fix some K and separate the sum into terms with $\tau \leq K$ and $\tau > K$. For $\tau > K$ we make the replacement $q_\tau \to \tilde{q}(\tau)$ where $\tilde{q}$ is a smooth function of τ; this allows the sum to be estimated by an integral. Fixing values for q_K and q_t, the action can be minimised (exactly) over the function $\tilde{q}$, which is equivalent to solving the instanton equation in [25]. One finds

$$\tilde{q}(\tau) = C_1 \tau^{-a} + C_2 \tau^{-(1-a)} \tag{26}$$

where C_1 and C_2 are fixed by the boundary conditions at $\tau = K, t$. Writing $u = K/t$ (so $0 < u < 1$), the contribution to $\mathcal{S}$ from this path is [25]

$$\mathcal{S}_1 = \frac{2a-1}{u^{1-2a}-1}\left(q_t - q_K u^{1-a}\right)^2 t . \tag{27}$$

For this optimal path (21) then reduces to

$$\mathcal{S} = \sum_{\tau=0}^{K-1} [(\tau+1)q_{\tau+1} - q_\tau(\tau+a)]^2 + \mathcal{S}_1 \tag{28}$$
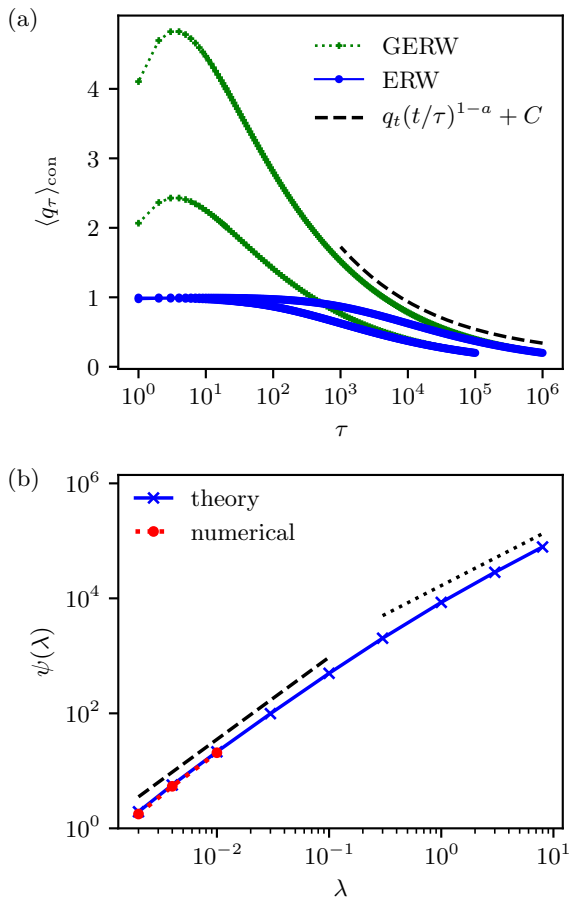
FIG. 2. (a) Averaged paths $\langle q_\tau \rangle_{\text{con}}$ of the controlled dynamics with $t = 10^5, 10^6$ and $\langle q_t \rangle_{\text{con}} = 0.2$, for the ERW and GERW with $a = 0.7$. These illustrate the IGL and LIE mechanisms. The dashed line indicates the long-time behaviour (29) which is common to both ERW and GERW. (In the plot, this has been offset by $C = 0.14$, for clarity.) (b) Theoretical estimate for $\psi(\lambda)$ in the ERW again with $a = 0.7$, derived at $t = 10^4$ and compared with numerically exact results for small λ. The dashed and dotted lines indicate the predicted power laws $\psi \propto \lambda^{1/a}$ and $\psi \propto \lambda$ respectively.

which is to be minimised over $q_1, q_2, \ldots, q_K$. This is the procedure used to obtain the GERW paths in Fig. 2(a). We typically take $K = 40$, this choice does not strongly affect the results because replacing the discrete sum by an integral is accurate for $K \gg 1$.

Remembering that we focus throughout on the case $a > 1/2$, the behaviour of (26) for large t, τ gives

$$\langle q_\tau \rangle_{\text{con}} \approx q_t (t/\tau)^{1-a} , \qquad (29)$$

similar to [25]. Comparing with (19), we see that the long-time behaviour of the optimally-controlled dynamics matches the natural regression to the mean.

Extrapolating (29) back to $\tau = 1$ indicates that for (rare) paths that end at q_t, the first hop should have size $q_t t^{1-a}$, which diverges as $t \to \infty$. In fact the early-time behaviour is more complex but the size of the first hop

is indeed of this order. The diverging hop is the reason that we call this mechanism an *initial giant leap* (IGL). It applies in the Gaussian elephant for all fluctuations with $q_t = O(1)$ as $t \to \infty$.

Two comments are in order. First, the analysis here for $\tau > K$ recovers exactly that of [25], the fact that the distribution of q_t is sharply-peaked under the controlled dynamics can be used to justify the so-called temporal additivity assumption in that work, for $\tau \geq K \gg 1$. However for the early part of the trajectory with $\tau < K$, it is important that the model evolves by discrete time steps and that q_t can change significantly in a single step. This means that the temporal additivity assumption is not valid in this regime. For this reason, quantitative results for $p_t(q)$ require a more detailed analysis of early times, without the temporal additivity assumption. We accomplish this here by analysing numerically the sum of terms with $\tau < K$. The second comment is that we use the language of a giant leap, but we note that the GERW makes (on average) very large jumps on several of the early time steps. We explain below that we are using IGL to refer to any divergent displacement q^* in a finite time interval τ^*, see Sec. IV B.

C. LIE mechanism for large deviations in the ERW

As discussed in Sec. III A, large-deviation properties of the ERW are available from [28]. In particular, there is an LDP for q_t with speed t whose rate function behaves for small q as

$$I(q) \simeq \kappa_{\text{E}} |q|^{1/(1-a)} . \qquad (30)$$

Correspondingly,

$$\psi(\lambda) \simeq c_{\text{E}} |\lambda|^{1/a} , \qquad (31)$$

for small λ. [Recall Equs. (15,16).]

We characterise here the mechanism responsible for (31), by deriving a controlled process which captures the behaviour of the relevant conditioned path ensemble, see Appendices C 1 and C 3. This controlled process is similar to the original process, but now $\Delta x_\tau = \pm 1$ with time-dependent probabilities $(1 \pm b_\tau)/2$ where $(b_1, b_2, \ldots, b_t)$ are variational parameters that we optimise, to reproduce the large-deviation mechanism.

This analysis yields a controlled process for which $\langle q_\tau \rangle_{\text{con}}$ is shown in Fig. 2(a): for early times, typical paths have $q_\tau \approx 1$ which is the maximum possible value in the ERW. This behaviour persists over a finite fraction of the trajectory, which motivates the name, *long initial excursion* (LIE). For larger times, q_τ decreases. Fig. 2(b) shows our theoretical estimate of $\psi(\lambda)$ obtained by a variational analysis at finite t, compared with numerically exact results from direct simulation. The theoretical estimate (i) matches the exact result in the region where numerical results are available; (ii) is consistent with (31) for $t^{-a} \ll \lambda \ll 1$; (iii) recovers $\psi(\lambda) \simeq |\lambda|$ for large λ,

which is the exact result (since $q \leq 1$). The controlled dynamics give a good description of the true $\psi(\lambda)$.

It can also be shown that the averaged paths in Fig. 2(a) capture the true fluctuation mechanism. We sketch the argument. At the level of large deviations, the true mechanism is the path measure P^{con} that achieves equality in (C4). From [28], the large-deviation event $q_t = q$ is associated with a single path, in the sense that the conditional distribution of $q_{\alpha t}$ is sharply-peaked as $t \to \infty$ for all $\alpha \in (0,1]$. Our ansatz for the controlled process is sufficiently general to capture this path, so minimising over all controlled paths is sufficient to make (C4) an equality, and hence to obtain the true mechanism. This argument also justifies the temporal additivity principle of [24] in this case.

Ref. [25] used that principle together with a quadratic expansion of the action about $q = 0$, for (symmetric) models similar to the ERW. This predicts dominant paths similar to (26). The results presented here show that such an expansion is not generically valid: for all $q_t \neq 0$, large-deviation events involve initial excursions far from $q_t = 0$, and the quadratic expansion breaks down. Nevertheless, if $q_t \ll 1$ then the quadratic expansion is applicable at large times and can be used to show that the optimally-controlled process behaves the same as the GERW for large τ, that is $\langle q_\tau \rangle_{\mathrm{con}} \approx (t/\tau)^{1-a} q_t$ as in (29). [For the ERW, this result is valid for $a > 1/2$ and $q_t \ll 1$ with $t, \tau \to \infty$ such that also $\langle q_\tau \rangle_{\mathrm{con}} \ll 1$]. A very similar case is analysed in Sec III.C of [16], for a cluster growth model.

Comparing (29) with (19), we see that the long-time behaviour of the optimally-controlled dynamics matches the natural regression to the mean, for both ERW and GERW. In other words, the controlled dynamics is almost that of the original model, when τ is sufficiently large. In Sec. IV C below, we exploit this fact to show that the scaling $\psi(\lambda) \sim |\lambda|^{1/a}$ of (31) is generic if optimally-controlled processes have (i) $\langle q_t \rangle_{\mathrm{con}} \approx 1$ until some time $\tau^* \sim t$, and (ii) $\langle q_\tau \rangle_{\mathrm{con}} \sim \tau^{-(1-a)}$ for long times. This is the sense in which the ERW is a prototype for a general fluctuation mechanism.

IV. GENERIC FLUCTUATION MECHANISMS

A. Overview of method

We have explained that the large-deviation behaviour of the ERW and GERW is different from that expected in Markov chains. Fluctuations in these models occur by mechanisms where the particle makes a large *excursion* from the origin at early times, which biases all future motion in the same direction, via the memory effect. This leads to a reduced speed in the LDP of the GERW and to a singular rate function in the ERW. The difference between ERW and GERW arises from the different characters of their initial excursions (a giant leap over a finite time for the GERW and a long excursion scaling with tra-

jectory length for the ERW). In this section we explain that such phenomena are relevant for a broad class of non-Markovian models. We provide general conditions under which excursions can occur, and explain their consequences for LDPs.

We consider models where q_t converges to its mean as $t \to \infty$ (to be precise, this is convergence in probability). We denote this mean value by

$$q_\infty = \lim_{t \to \infty} \langle q_t \rangle . \qquad (32)$$

For simplicity we discuss deviations with $q_t > q_\infty$; the opposite case is a straightforward analogue. We consider excursions which extend over a time period between $t = 0$ and some time τ^*. The probability $p_t(q)$ can be bounded from below by restricting to paths where the size of the excursion is at least q^*, that is $q_{\tau^*} \geq q^*$. By conditional probability:

$$\log p_t(q) \geq \log p_t(q | q_{\tau^*} \geq q^*) + \log P(q_{\tau^*} \geq q^*) \qquad (33)$$

where $P(q_{\tau^*} \geq q^*)$ is the probability of the excursion and $p_t(q | q_{\tau^*} \geq q^*)$ is the corresponding conditional probability density for q_t.

The inequality (33) is valid for all q^*, τ^*. Now suppose that q, t are given and we seek a useful bound on $p_t(q)$: this requires that we choose suitable values for q^*, τ^*. To this end, we introduce the notation $\langle \cdot \rangle_{q^*, \tau^*}$ for averages that are conditioned on $q_{\tau^*} \geq q^*$. Then we choose q^*, τ^* such that $\langle q_t \rangle_{q^*, \tau^*} = q$ and we further assume that the conditional distribution of q_t is sharply-peaked at this value. This means that if we consider trajectories where a suitable excursion has already taken place before τ^*, then following the natural dynamics of the model for $t > \tau^*$ will result in $q_t \approx q$ with a probability close to unity.

Under these assumptions, (33) reduces to

$$\log p_t(q) \gtrsim \log P(q_{\tau^*} \geq q^*) . \qquad (34)$$

In other words, we now have a more explicit lower bound on $p_t(q)$ which is valid if

$$\langle q_t \rangle_{q^*, \tau^*} = q. \qquad (35)$$

(The additional requirement that the conditional distribution is sharply peaked is always obeyed in the following.)

The strategy in Secs. IV B and IV C below is to characterise situations in which (34) can be used to establish LDPs that differ from those expected in finite Markovian models. In particular, we now establish a sufficient condition for memory to have a strong effect on the large-t behaviour. Physically, the idea is that after the excursion, the time-averaged current relaxes to its steady-state value as a power law with exponent $1 - a$, as established in (19) for the (G)ERW. Finite Markovian systems relax generically as t^{-1}, so a encodes the effects of memory, this is related to the fixed-point stability analysis of [25]. The condition that we will require is that for $t > \tau^*$,

$$\langle q_t - q_\infty \rangle_{q^*, \tau^*} \simeq (q^* - q_\infty) \mathcal{F}(q^*, \tau^*) \left(\frac{\tau^*}{t} \right)^{1-a} \qquad (36)$$

for some function $\mathcal{F}$, and some number $a \in (0,1)$.

We then arrive at the following method for deriving bounds on $p_t(q)$. We must first establish (36) for a particular model, at least for q^* in some range. To bound $p_t(q)$ for specific values of q, t, we must then find a combination q^*, τ^* such that (36) holds, with $q_t = q$. As long as this is possible, the constraint (35) is satisfied and the resulting q^*, τ^* can be substituted into (34) to obtain a bound on $p_t(q)$. Note that the combination q^*, τ^* depends in general on t; the final step is to take $t \to \infty$ in order to characterise large deviations that occur in this limit. This strategy is similar to those used in [16].

B. Generic IGL mechanism

We now show how a generic IGL mechanism leads to a useful bound. We achieve this by laying out the properties that a model should have, in order that this mechanism is relevant. A defining feature of the IGL is that it takes place over a finite time period τ^* and that the size of the excursion diverges in the limit $t \to \infty$.

The first requirement is that the model of interest supports very large excursions. To characterise their probability, we require that there exists some τ^* such that for $q^* \to \infty$ we have

$$\log P(q_{\tau^*} \geq q^*) \simeq -\gamma |q_* - q_\infty|^\beta , \qquad (37)$$

with $\gamma, \beta > 0$. Since we consider divergent excursions, we require that (36) remains valid even as $q^* \to \infty$. In the following we take τ^* to be a fixed parameter, the choice of its value is discussed below. We define

$$f_*(\tau^*) = \lim_{q^* \to \infty} \mathcal{F}(q^*, \tau^*) \qquad (38)$$

which we require to be strictly positive. These requirements place strong constraints on the range of models for which the IGL mechanism will determine the large deviations but, as we demonstrate, such models do indeed exist. Then (33,36) with $q = \langle q_t \rangle_{q^*, \tau^*}$ yield

$$-\log p_t(q) \lesssim t^{\beta(1-a)} |q - q_\infty|^\beta \kappa_{\mathrm{IGL}} \qquad (39)$$

with $\kappa_{\mathrm{IGL}} = \gamma f_*(\tau^*)^{-\beta} \tau_*^{-\beta(1-a)}$.

Equ. (39) corresponds to an LDP with speed $t^{\beta(1-a)}$. If this speed is less than t, fluctuations are qualitatively larger than one finds in generic Markovian systems. In principle the bound (39) can be optimised over τ^*. However, (39) already establishes that the speed of the LDP can be less than t, without any requirement for optimisation over τ^*. This is the central result. In this sense, the specific value of τ^* is not crucial.

The GERW satisfies all the requirements for the IGL mechanism, with $\beta = 2$; one may take $\tau^* = 1$. The applicability of (36) was already shown in (19). The resulting bound is consistent with the exact result (11), it gives the right scaling with t and the correct general mechanism. However, the constant κ_{IGL} obtained from this generic argument does not coincide with the prefactor in the exponent of (11): obtaining that result requires the more detailed (model-dependent) calculation of Sec. III B.

We note in passing that some arguments of [25] are similar to those of this section, but the connection between the giant leap and the reduced speed of the LDP was neglected in that work. In particular, the requirement that (36) must hold as $q^* \to \infty$ means that some care is required when applying the arguments of [25] to generic models; they are not valid in the ERW, for example.

C. Generic LIE mechanism

The LIE mechanism is generically associated with excursions that have finite q_* but diverging τ_* (proportional to t). This may be compared with the IGL, which has fixed τ^* and diverging q^*. The LIE mechanism has two central requirements, which must hold for some q^*, different from q_∞. First, (36) must hold asymptotically for $1 \ll \tau^* \ll t$. Second,

$$f_\ddagger(q^*) = \lim_{\tau^* \to \infty} \mathcal{F}(q^*, \tau^*) \qquad (40)$$

must be strictly positive. Comparing with (38), the roles of q^*, τ^* are reversed.

Under these conditions, we assume that there is an LDP with speed t as in (3), verify the self-consistency of this assumption, and establish a bound on the rate function $I(q)$ for $|q - q_\infty| \ll 1$. Since τ^* is proportional to t, this means that

$$P(q_{\tau^*} \geq q^*) \simeq \exp[-\tau_* I(q_*)] \qquad (41)$$

which is analogous to (37). Using this with (34,36) yields

$$-\log p_t(q) \lesssim t \kappa_{\mathrm{LIE}} |q - q_\infty|^{1/(1-a)} \qquad (42)$$

with

$$\kappa_{\mathrm{LIE}} = I(q^*) \left(\frac{1}{|q_* - q_\infty| f_\ddagger(q^*)} \right)^{1/(1-a)} . \qquad (43)$$

The result (42) is consistent with the assumption of an LDP with speed t, but it shows (for $a > 1/2$) that the rate function increases from zero more slowly than any quadratic function. As noted above, this means that $I''(0) = 0$, corresponding to superdiffusive scaling.

In addition, by Varadhan's lemma [a standard result in large deviation theory [18, 21], which amounts to the inverse Legendre transform of (7)], one obtains

$$\psi(\lambda) \gtrsim \sup_q [\lambda q - |q - q_\infty|^{1/(1-a)} \kappa_{\mathrm{LIE}}] \qquad (44)$$

which gives

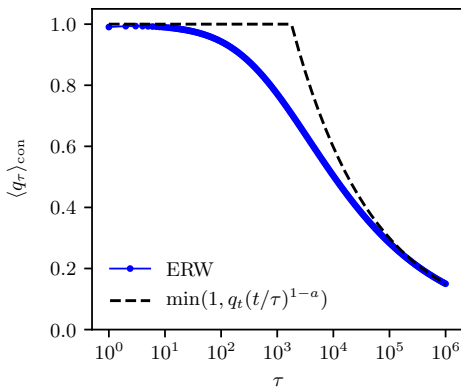$$\psi(\lambda) \gtrsim \lambda q_\infty + |\lambda|^{1/a} c_{\mathrm{LIE}} \qquad (45)$$

FIG. 3. Comparison between the optimally controlled path for the ERW (similar to Fig. 2) and the corresponding generic LIE path used to derive (42). We take $a = 0.7$ with $q_t = 0.15$ and $t = 10^6$. The generic LIE path has an excursion with $q^* = 1$, after which q_τ relaxes back towards zero, as the system follows its natural dynamics (19). The generic LIE path does not capture the details of the optimally-controlled (instanton) path which means that the coefficient κ_{LIE} does not match κ_{E} in (16), but the generic LIE argument is sufficient to capture the non-quadratic form of the rate function at $q = 0$.

with

$$c_{\mathrm{LIE}} = a \left(\frac{1-a}{\kappa_{\mathrm{LIE}}} \right)^{(1/a)-1}. \tag{46}$$

All these generic arguments are consistent with the behaviour of the ERW, which has $q_\infty = 0$. In particular, the requirement for (36) to hold asymptotically follows from (19).

Moreover, the results of Sec. III C indicate that the true fluctuation mechanism for the ERW is an LIE with $q^* = 1$. Also $p_t(1) = (1+a)^{t-1}/2^t$ because all hops have $\Delta x_t = 1$ in this case, so $I(1) = \log[2/(1+a)]$. The coefficient in (19) is $\eta \simeq \tau^{1-a}$ as $\tau \to \infty$, which means $f_{\ddagger}(q^*) = 1$. Hence the bound (39) holds with $\kappa_{\mathrm{LIE}} = \log[2/(1+a)]$. The exact result for the ERW can be obtained from $\kappa_{\mathrm{E}} = (1-a)(a/c_{\mathrm{E}})^{a/(1-a)}$ as quoted in Sec. III A, together with (B10).

For the representative case $a = 0.7$, we find $\kappa_{\mathrm{E}} = 0.04$ while $\kappa_{\mathrm{LIE}} = 0.16$. Given that the generic LIE argument is much simpler than the full calculation of κ_{E}, this level of agreement is reasonable. The generic LIE argument is based on a simple path (or equivalently a simple controlled process) that includes a long excursion: the path is illustrated in Fig. 3, where it is compared with the optimal LIE path discussed in Sec. III C. The generic LIE path captures the correct qualitative behaviour and matches the optimal path for small and large times. However, the agreement is not quantitative, and the difference between κ_{LIE} and κ_{E} reflects this.

D. Discussion of generic IGL and LIE mechanisms

We summarise the difference between the IGL and LIE mechanisms. The IGL makes a giant (divergent) excursion in a finite time and leads to an LDP with reduced speed. The LIE makes a finite excursion over a long (divergent) time period; it leads to an LDP with speed t, and to a rate function with $I''(q_\infty) = 0$ which is (generically) non-analytic at q_∞. In all the examples that we have managed to construct, the IGL mechanism relies on microscopic transition rates that diverge as $q_t \to \infty$, in order to satisfy (36).

The IGL mechanism has an interesting analogy with condensation in interacting-particle systems [52, 53]: to achieve $q_t = q$ the system must support an excess current which may be distributed over a macroscopic fraction of the time period (as in the LIE), or *condensed* into a finite time interval (the IGL). A similar phenomenon is described by the "single-big-jump" principle for sums of random variables (including certain types of correlated process) [54]; the particular history-dependence in our models, with $a > 0$, constrains the condensation to take place at the beginning of the time period.

We close this section by noting that (39,42) are both lower bounds on the probability $p_t(q)$. Physically, this means that fluctuations *can* take place by IGL and LIE mechanisms, so fluctuations of a given size q are *at least as likely* as (39,42) predict. We have not ruled out competing mechanisms that might allow fluctuations of the same size to occur in a more likely way. As a simple example, an LIE bound can be obtained for the GERW but does not accurately describe the probability of rare fluctuations, because the IGL mechanism is available and occurs with (much) higher probability. [Indeed it is easy to see that the IGL mechanism, if available, will always dominate the LIE mechanism if $\beta(1-a) < 1$.] To rule out competing mechanisms, one would need a matching upper bound on the probability; this seems to require more detailed (model-dependent) analysis.

V. EXAMPLE MODELS EXHIBITING IGLS AND LIES

By considering IGLs and LIEs, we have established simple and generic requirements which enable bounds on the probabilities of large-deviation events. It is straightforward to construct (or identify) other models that exhibit these mechanisms. In this section we give a brief discussion of three such cases. Similar to the ERW in Sec. III C, we establish bounds on the probabilities of large excursions by using arguments based on optimal control theory, these computations then enable us to check conditions for the IGL and LIE mechanisms. Our main purpose here is not to describe the model behaviour in detail, but rather to illustrate the general relevance of the identified mechanisms.

A. IGL in unidirectional hopping model

As an example in continuous time, we modify the unidirectional walker model of [24]. Similarly to the ERW, we consider a particle with integer-valued position x_t which we identify with the configuration $\mathcal{C}_t$. The particle always hops in the same direction so $\Delta x_t = 1$. We define q_t as the total time-averaged displacement which corresponds to (8) with $\alpha_j = 1$ for all jumps. In the variant of the model that we consider, the particle makes its first jump at time t_0; subsequent jumps occur with rate

$$r(q_t) = aq_t , \qquad (47)$$

where $0 < a < 1$. The regularisation parameter t_0 is important because if one allows jumps to occur at arbitrarily early times then q_t in (8) can become arbitrarily large after just one jump; combined with (47), this can lead to pathological fluctuations.

The results of [24] indicate that large deviations with $q_t > 0$ involve a giant leap of size $q^* \sim t^{1-a}$, leading to an LDP with speed t^{1-a}. However, that work made an assumption of temporal additivity which (strictly-speaking) is valid only for $t_0 \gg 1$. Here we discuss the case where t_0 takes any positive value; we show that the IGL mechanism operates, and $p_t(q)$ can be bounded as in (39), which is consistent with an LDP with speed t^{1-a}.

To analyse the IGL we take $\tau^* = 2t_0$. In this case we show in Appendix D that

$$\log P(q_{\tau^*} \geq q^*) \gtrsim -\gamma_{\text{uni}}q^*t_0 , \qquad (48)$$

with $\gamma_{\text{uni}} = O(1)$ as $q^* \to \infty$. That is, the probability of a large excursion to q^* in a finite time decays at most exponentially in q^*. This establishes the requirement (37) for an IGL.

Moreover, after the excursion the average displacement obeys

$$\tau \frac{\partial}{\partial \tau} \langle q_\tau \rangle_{q^*, \tau^*} = (a - 1)\langle q_\tau \rangle_{q^*, \tau^*} , \qquad (49)$$

which follows directly from the master equation of the model. Similar to (19), integrating this equation yields $\langle q_\tau \rangle_{q^*, \tau^*} = q^*(\tau_*/\tau)^{1-a}$ which is exactly the required condition (36) with $q_\infty = 0$ and $\mathcal{F}(q^*, \tau^*) = 1$. Note that this holds even as $q^* \to \infty$, which is related to the fact that $r(q^*)$ diverges in this limit. To apply (34) requires that the conditional distribution of q_t after the excursion is sharply-peaked: this is easily verified.

Hence, the conditions for an IGL are in place and we have established (39) with $\beta = 1$ and $f_*(\tau^*) = 1$, that is

$$-\log p_t(q) \lesssim t^{1-a}\kappa_{\text{uni}}q , \qquad (50)$$

with $\kappa_{\text{uni}} = 2^{a-1}\gamma_{\text{uni}}t_0^a$, using $\tau^* = 2t_0$, from above. This corresponds to an LDP with speed t^{1-a} as shown in [24, 25] by arguments based on an assumption of temporal additivity. Our analysis avoids any such assumption; it also shows that the unusual speed of the LDP arises because the fluctuation mechanism is an IGL.

The result (50) applies to the unidirectional model with $r(q) = aq$ but, in fact, the main ingredient required in the analysis was $\lim_{q \to \infty}[r(q)/q] = a$ (with $0 < a < 1$). We therefore expect the IGL mechanism to operate for a broad class of models where this assumption holds.

B. LIE in cluster growth models

We consider a model of a growing cluster as in [14–16]. The cluster contains two types of particles (for example, red and blue) whose numbers at time t are n_t^R and n_t^B. The cluster evolves in discrete time and a single particle is added on each step, so $n_t^R + n_t^B = t$. (This is the irreversible model of [14], in that particles are added but never removed.) The configuration is given by $\mathcal{C}_t = (n_t^R, n_t^B)$ and we take $q_t = (n_t^R - n_t^B)/t$ which means that $\alpha_t = \pm 1$ in (8) according to whether a red or blue particle is added.

On step t, the added particle is red ($+$) or blue ($-$) with probability $(1 \pm \tanh Jq_{t-1})/2$ where $J > 0$ is a parameter that reflects the difference in energy on adding either a red or blue particle. In this case, the dynamics of the quantity $m_t = (n_t^R - n_t^B)$ is similar to that of the ERW position x_t, but with the nonlinear tanh function replacing the linear function in (2). This nonlinearity leads to a symmetry-breaking transition: for $J < 1$ then $q_t \approx 0$ at long times ("mixed" clusters) but for $J > 1$ then $q_t \approx \pm\bar{m}$, which corresponds to spontaneous demixing. Here $\bar{m}$ is the order parameter for the underlying phase transition [14]. Large deviations in this model were discussed previously in [15, 16], it may be also formulated as an urn model so the results of [28] are applicable.

In the mixed (one-phase) regime, the behaviour of this model is qualitatively similar to the ERW. It can be analysed similarly to Sec. III C, using the same (general) controlled model: red/blue particles are added with probabilities $(1 \pm b_t)/2$. The theoretical arguments of Appendix C 1 can then be applied. Indeed, these ideas were already applied to the growth model in [16]: for $1/2 < J < 1$ this led to a result analogous to (13), with $a = J$. However, that paper did not come to a definitive conclusion about the speed of the LDP in this regime. The general results of the present work can be used to resolve this open question, and to understand the rare-event mechanism. We outline the argument below (again for $1/2 < J < 1$).

The results of [28] prove that the LDP for this model must have speed t, so one may expect an LIE mechanism, similar to the ERW. Moreover, Ref. [16] showed that (19) holds in this system for relaxation as $t \to \infty$ after an initial excursion. However, contrary to the ERW, this result is now valid only for $\langle q_t \rangle_{q_\tau} \ll 1$. This establishes that (36) holds, but only for small values of q^*.

We therefore fix some small value for this parameter and construct the LIE, using (36) as in Sec. IV C to fix $\tau^* = t(q^*f_\ddagger(q^*)/q_t)^{-1/(1-a)}$ so that the natural dynamics after the excursion arrives at q_t with proba-

bility 1. By [28], this excursion has $\text{Prob}(q_{\tau^*} \geq q^*) \simeq \exp[-\tau^* I(q^*)]$, although the rate function I is not known explicitly. These results can be used with (34) to obtain

$$- \log p_t(q) \lesssim t \kappa_{\text{LIE}} |q|^{1/(1-a)} , \qquad (51)$$

as in (42). Since the validity of (36) is restricted to small q^*, this construction is restricted to small q (strictly positive and fixed as $t \to \infty$). Still, this generic bound is sufficient to establish the non-analytic behaviour of the rate function at $q = 0$. The use of a fixed small value of q^* is convenient for this argument but is not expected to be optimal for the large-deviation mechanism; in fact we anticipate the true large-deviation mechanism to involve an excursion with $q^* = 1$, as for the ERW. This means that the prefactor κ_{LIE} is likely to be far from optimal, but the scaling (51) is expected to be robust.

The overall picture is that for small values of q_t (fixed as $t \to \infty$), the cluster growth model with $1/2 < J < 1$ behaves similar to an ERW with $a = J$, exhibiting an LIE fluctuation mechanism and a rate function that increases from zero with exponent $1/(1-a)$. Physically, the similarity can be explained by an argument similar to the fixed-point stability analysis of [25], because the exponent that appears in the LIE bound only depends on the asymptotic (long-time) dynamics close to the fixed point. For models that can be formulated as urns [28], we therefore expect these similarities with the ERW to be generic, based on an expansion of the urn function about the fixed point.

C. LIE in a non-Markovian exclusion process

We consider a non-Markovian simple exclusion process (SEP) where N particles hop in continuous time on a periodic one-dimensional lattice of L sites, subject to the constraint that each site may contain at most one particle. We define $n_i = 1$ if site i contains a particle, and $n_i = 0$ otherwise. A configuration is specified as $\mathcal{C} = (n_1, n_2, \ldots, n_L)$. The time-averaged current is $q_t = (Lt)^{-1} \sum_{\text{jumps } j} \Delta x_j$, as in (8), where the sum is over all particle hops, with $\Delta x_j = \pm 1$ according to whether the hop is to the right or the left. Large deviations of q_t have been studied extensively in the Markovian case [55, 56]. For non-Markovian models, similar quantities have been studied in [25, 57]. Given the connections between exclusion processes and traffic modelling [58], the generalisation of such models to include memory of previous flow (current) is quite natural [25].

We introduce here a memory of mean-field type, so that every particle hops either right $(+)$ or left $(-)$ with rate $w_\pm = [1 \pm \tanh(\nu q_t)]/2$, as long as the destination site is empty. The non-linearity in this model is similar to that of the cluster growth model which leads to some similar phenomenology.

It is useful to note that detailed balance is broken in this model (except for $q_t = 0$), but the dynamical rules for
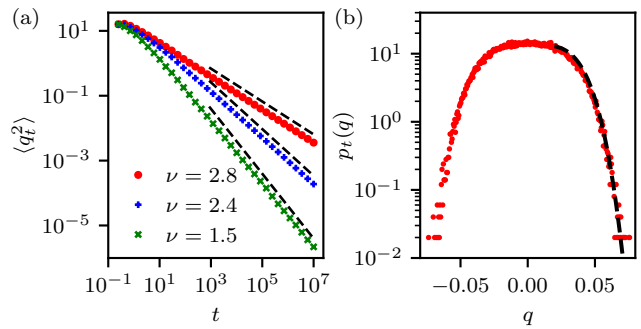

FIG. 4. Numerical results for a non-Markovian simple exclusion process with $(N, L) = (8, 16)$, so $\nu_c = 3.75$. (a) For $\nu = 2.4, 2.8$, particle motion is superdiffusive so the variance of q_t decays as a power law consistent with (10), dashed lines indicate power-law behaviour with the theoretically-predicted exponent $a = \nu/\nu_c$. For $\nu = 1.5$ the behaviour is diffusive, $\langle q_t^2 \rangle \propto t^{-1}$, since $\nu < \nu_c/2$. (b) The distribution of q_t for $\nu = 2.8$ at $t = 10^5$ is similar to the ERW in Fig. 1, the dashed line is a fit to (16) with $a = \nu/\nu_c$.

any given q_t correspond to an asymmetric simple exclusion process with periodic boundaries, whose stationary state has all particles distributed independently (subject to the exclusion constraint). Assuming that the system is in such a stationary state at time t, and its time-averaged current is q_t, the (average) rate for accepted particle hops is

$$\left\langle L \frac{d}{dt}(t q_t) \right\rangle_{q_t} = N \frac{L-N}{L-1} \tanh(\nu q_t) . \qquad (52)$$

Here the factor of $(L-N)/(L-1)$ is the probability that a site adjacent to a given particle is vacant. Expanding the tanh about $q_t = 0$ shows that the zero-current state $\langle q_t \rangle = 0$ is stable only if $\nu < \nu_c$ with

$$\nu_c = \frac{L(L-1)}{N(L-N)} . \qquad (53)$$

We identify ν_c as a phase-transition point, directly analogous to the cluster-growth model.

For $\nu < \nu_c$, expansion of (52) about $q_t = 0$ yields (36) with $a = \nu/\nu_c$, which is again similar to the growth model and indicates that the LIE scenario is applicable, at least for small q^*. As a controlled model, we consider a (Markovian) asymmetric simple exclusion process with a time-dependent asymmetry parameter, so hops in the $(\pm)$-direction have $w_\pm = (1 \pm b_t)/2$. This controlled model also has particles distributed independently at all times. In this case the KL divergence may be computed similarly to (C9). This allows numerical optimisation of the controlled dynamics – the optimal behaviour is similar to the ERW and cluster growth models, showing an LIE mechanism. An explicit LIE bound may also be derived by following exactly the same steps as used for the cluster growth model in Sec. V B.

Contrary to the other models considered here, we do not expect this controlled model to fully capture the

large-deviation mechanism, because it neglects interparticle correlations which are important for large deviations in exclusion processes [55]. This effect might be captured by combining the temporal additivity principle [25] with results for large deviations in Markovian exclusion processes [55], but such an analysis is beyond the scope of the present work. However, we expect the general features to be robust: a large excursion at early times and a rate function scaling as (42). Numerical results confirming the similarity between this non-Markovian SEP and other LIE models are shown in Fig. 4. This analysis illustrates that the generic fluctuation mechanisms described in this paper are not limited to simple one-particle systems.

As a final observation, note that since particles do not pass each other in exclusion processes, trajectories with time-averaged current $q_t = c$ at long times must have single-particle currents whose time averages all converge to c also. For this reason, we would expect similar behaviour if each particle had an individual memory of its own individual displacements, in contrast to the simple (mean-field) case considered here, where the motion of each particle is affected by the memory of the whole system.

VI.   OUTLOOK

We have presented two mechanisms by which large deviations can occur in non-Markovian processes, leading to generic bounds (39,42) on the probabilities of these rare events. To prove that these bounds give the right scaling in specific cases requires more detailed analysis, as illustrated here for the simple ERW and GERW models. (Such analyses are necessary to rule out competing mechanisms with larger probability then the IGL and LIE.) Our results indicate that the LIE mechanism operates in a non-Markovian exclusion process, and the general mechanistic insights have enabled us to clarify and extend several other results from the literature [16, 24, 25]. This understanding is also relevant in socioeconomic decision models that can be approximated by generalised urn/elephant models [17]; by revealing fluctuation mechanisms in these systems, our analysis may be utilised to predict and control their long-term fluctuations. We look forward to future work exploiting these new insights, in order to elucidate the rich fluctuation behaviour of non-Markovian systems.

Appendix A: Typical fluctuations in the GERW

We derive (10) for the GERW. Suppose that after t steps $Q_t = tq_t$ has a Gaussian distribution with mean zero and variance v_t. Then $Q_{t+1} - Q_t$ is normally distributed with mean aq_t and variance 1 so the distribution of Q_{t+1} is

$$p(Q_{t+1}) = \frac{1}{z_t} \int \exp\left[-\frac{[Q_{t+1} - (1 + \frac{a}{t})Q_t]^2}{2} - \frac{Q_t^2}{2v_t} \right] \mathrm{d}Q_t \tag{A1}$$

with $z_t = \sqrt{4\pi^2 v_t}$. This distribution is normal with mean zero and variance $v_{t+1} = 1 + v_t(1 + a/t)^2$. From this recursion relation one finds a series solution for v_t in terms of the gamma function:

$$v_t = \frac{\Gamma(a+t)^2}{\Gamma(a+1)^2\Gamma(t)^2} \left(\sum_{n=1}^{t-1} \frac{\Gamma(a+1)^2\Gamma(n+1)^2}{\Gamma(a+n+1)^2} + 1 \right). \tag{A2}$$

The form of the large-t behaviour can be obtained directly from the recursion by writing $v_t = v(t)$ so that $v'(t) \approx 1 + 2av(t)/t$. Hence $v(t) \approx t/(1 - 2a) + ct^{2a}$ and so the variance of q_t is

$$\mathrm{Var}(q_t) = \frac{v(t)}{t^2} \approx \frac{1}{t(1-2a)} + c_a t^{-2(1-a)} \tag{A3}$$

where subleading terms at higher order in t^{-1} have been omitted. The second term is dominant for $a > 1/2$ and the constant c_a corresponds to χ_{G} in the asymptotic variance; its value can be extracted as a limit from the series solution. For $a = 0.7$ as used in Fig. 1 numerical evaluation of the limit yields $\chi_{\mathrm{G}} \approx 3.4$.

Appendix B: Large deviations in ERW by mapping to urn model

For large deviations of q_t in the ERW, the SCGF $\psi(\lambda)$ can be obtained exactly by adapting results of [28]. We state the equations and characterise the behaviour at small λ.

The ERW can be interpreted as an urn model [10]. If the fraction of $+$ steps before time t is s_t then the probability that $\Delta x_{t+1} = +1$ is $\pi(s_t)$ where

$$\pi(s) = \frac{1 + a(2s - 1)}{2} \tag{B1}$$

is the corresponding urn function [28]. Given this urn function, the parameters (a, b) of Corollary 12 of [28] correspond to $((1-a)/2, a)$ in the notation of this work. Since $q_t = 2s_t - 1$ then

$$G(\lambda, t) = \log\langle \mathrm{e}^{\lambda t(2s_t - 1)} \rangle. \tag{B2}$$

Define $\tilde{\psi}(\mu) = \lim_{t\to\infty} t^{-1} \log\langle \mathrm{e}^{\mu t s_t} \rangle$ as the SCGF of [28], denoted in that work by ψ. Then (6,B2) yield

$$\psi(\lambda) = \tilde{\psi}(2\lambda) - \lambda. \tag{B3}$$

Hence by Corollary 12 of [28] one has for $\lambda > 0$ that

$$\psi(\lambda) = -\log\left[1 - we^{-2w\lambda}y^{1/a}\mathcal{B}(w, -2w, y)\right] - \lambda \ , \quad (B4)$$

where we introduced shorthand notation $w = \frac{1-a}{2a}$ and $y = 1 - e^{-2\lambda}$ (used only within this Appendix), and where

$$\mathcal{B}(w, v, y) = \int_y^1 (1-t)^{w-1}t^{v-1}dt \quad (B5)$$

is a particular case of the incomplete Beta function.

We now compute the behaviour of ψ at small λ, observing that $y \simeq 2\lambda$ in this limit. Our regime of interest is $1/2 < a < 1$ so that $0 < w < 1/2$. In this case $\mathcal{B}(w, -2w, y)$ diverges as $y \to 0$. To extract the nature of this divergence, introduce a factor of $1 = t + (1-t)$ into the integrand of (B5) to yield

$$\begin{aligned}\mathcal{B}(w, v, y) &= \int_y^1 \left[(1-t)^{w-1}t^v + (1-t)^w t^{v-1}\right]dt \\ &= \frac{v+w}{v}\int_y^1 (1-t)^{w-1}t^v dt - \frac{y^v(1-y)^w}{v} \ , \end{aligned}$$
$$(B6)$$

where the second line used an integration by parts, with the assumption that $w > 0$. There is no such assumption on v, the case of interest is $-1 < v < 0$. The limiting behaviour at small y can now be extracted: for $v > -1$ and $y \to 0$ then

$$\mathcal{B}(w, v, y) \simeq \frac{v+w}{v}B(w, 1+v) - \frac{y^v}{v} + o(1) \ . \quad (B7)$$

Here $B(x, y)$ is the (complete) Beta function which is given for $x, y > 0$ by $\int_0^1 t^{x-1}(1-t)^{y-1}dt$, it is finite and positive. Moreover, the relation $B(w, v) = \frac{v+w}{v}B(w, 1+v)$ extends the Beta function to negative arguments. Using these results with (B4) and identifying $-2w = (a-1)/a$ gives

$$\psi(\lambda) = $$
$$-\log\left[1 - wy^{1/a}\left(B(w, -2w) + \frac{y^{(a-1)/a}}{2w} + o(1)\right)\right]$$
$$- \lambda \ . \quad (B8)$$

Finally, noting that $y \simeq 2\lambda$ and using that ψ is an even function:

$$\psi(\lambda) = \frac{1-a}{2a}|2\lambda|^{1/a}B\left(\frac{1-a}{2a}, \frac{a-1}{a}\right)[1 + o(1)] \ . \quad (B9)$$

Hence

$$c_E = \frac{2^{(1/a)-1}(1-a)}{a}B\left(\frac{1-a}{2a}, \frac{a-1}{a}\right) \quad (B10)$$

in (15,31). Analysing the subleading term shows that in fact the first correction to (B8) is $\psi(\lambda) = c_E|\lambda|^{1/a} + O(\lambda^2)$.

Recall that we assumed here $1/2 < a < 1$, since this is the regime of interest for this work. However (B4) also applies for $0 < a < 1/2$ – similar analysis can also be carried out in that case. For $a < 0$ the corresponding result is given in [28], the resulting ψ is analytic.

Appendix C: Controlled dynamics

1. Outline of general theory

As discussed in the main text, one method for analysing fluctuation mechanisms is to construct controlled processes whose typical trajectories reproduce the rare-event behaviour of interest. Such processes can be analysed variationally.

We work in the generic framework where the configuration of the system at time t is $\mathcal{C}_t$. A trajectory or sample path is denoted $\mathcal{C}$ and its probability $\mathcal{C}$ in the original model is denoted by $P(\mathcal{C})$. Throughout our analysis, we fix t as the trajectory length and we use τ to indicate a generic time within the trajectory. Now let $P_{con}(\mathcal{C})$ be the probability of $\mathcal{C}$ in some controlled model, which has different dynamics. Optimal-control theory provides the following general inequality [43]

$$G(\lambda, t) \geq \lambda t\langle q_t\rangle_{con} - \mathcal{D}(P_{con}||P) \quad (C1)$$

where $\langle \cdot \rangle_{con}$ indicates an average in the controlled model, and $\mathcal{D}(Q||P)$ is the Kullback-Leibler (KL) divergence between the distributions Q and P. To prove (C1) define

$$P_{cano}(\mathcal{C}) = e^{\lambda t q_t - G(\lambda, t)}P(\mathcal{C}) \quad (C2)$$

which is a normalised probability distribution, by definition of G. (The subscript "cano" indicates that this definition is analogous to that of the canonical ensemble in thermodynamics.) Then by definition of the KL divergence, the right-hand side of (C1) can be expressed as

$$\lambda t\langle q_t\rangle_{con} - \mathcal{D}(P_{con}||P) = G(\lambda, t) - \mathcal{D}(P_{con}||P_{cano}) \ . \quad (C3)$$

The KL divergence is non-negative so the right-hand side is less than or equal to $G(\lambda, t)$, and (C1) follows. Moreover, there is equality in (C1) if and only if $P_{con} = P_{cano}$.

In addition, setting $\theta = 1$ in the definition (6) we obtain $\psi(\lambda) = \lim_{t\to\infty} t^{-1}G(\lambda, t)$ so (C1) yields

$$\psi(\lambda) \geq \lim_{t\to\infty}\left[\lambda\langle q_t\rangle_{con} - \frac{1}{t}\mathcal{D}(P_{con}||P_{cano})\right] \ . \quad (C4)$$

If this bound is saturated then the controlled process gives an accurate representation of the rare event of interest, see also below. We emphasise that for non-Markovian processes as considered here, the limit in (C4) involves controlled processes where the dynamical rule at time τ depends both on τ and on the total trajectory length t; accurate bounds require controlled processes with time-dependent rates.

2. GERW

We construct the optimally-controlled process for large deviations of q_t in the GERW. Using (C2) one obtains a distribution for the trajectory $\boldsymbol{X}$, as defined in Sec. III B:

$$P_{\text{cano}}(\boldsymbol{X}) \propto \exp\left[\lambda t q_t - G(\lambda, t) - \frac{\mathcal{S}(\boldsymbol{X})}{2}\right] . \quad \text{(C5)}$$

where q_t also depends on $\boldsymbol{X}$ though (1). This distribution is Gaussian for the increments and for the q_τ, and one has an analogue of (25) which is

$$\langle q_\tau \rangle_{\text{cano}} = \mu_\tau \langle q_t \rangle_{\text{cano}} h \quad \text{(C6)}$$

where h, μ_τ are the same quantities that appear in (25). That is, choosing λ in the canonical ensemble fixes $\langle q_t \rangle_{\text{cano}}$. Then the average path in this ensemble coincides with the average path in a corresponding microcanonical ensemble with $q_t = \langle q_t \rangle_{\text{cano}}$.

Since P_{cano} in (C5) is Gaussian, it is possible to construct exactly an optimally-controlled process that generates trajectories according to this distribution. This process achieves equality in (C1) and captures the mechanism by which large rare fluctuations occur in the GERW. This is similar to the Doob transform, as discussed in [22], with time-dependent rates as in [16]. Within the controlled system, the displacement on step τ is Gaussian with mean $aq_{\tau-1} + b_\tau$ and variance unity. This means that $P_{\text{con}}(\boldsymbol{X}) = \exp(-\tilde{\mathcal{S}}(\boldsymbol{X})/2)$ with

$$\tilde{\mathcal{S}}(\boldsymbol{X}) = \sum_{\tau=0}^{t-1}[(\tau+1)q_{\tau+1} - q_\tau(\tau+a) - b_{\tau+1}]^2 , \quad \text{(C7)}$$

analogous to (21). Hence

$$\tilde{\mathcal{S}}(\boldsymbol{X}) = \mathcal{S}(\boldsymbol{X}) - 2tq_t b_t + 2\sum_{\tau=1}^{t-1} q_\tau[(\tau+a)b_{\tau+1} - \tau b_\tau] + \sum_{\tau=1}^{t} b_\tau^2 . \quad \text{(C8)}$$

The optimally-controlled process has $P_{\text{con}} = P_{\text{cano}}$ [recall (C3)], which is achieved by setting $b_t = \lambda$ and using $b_{\tau-1} = b_\tau(1 + \frac{a}{\tau-1})$ iteratively to fix the b_τ. For the CGF this identification yields $G(\lambda, t) = \frac{1}{2}\sum_{\tau=1}^{t} b_\tau^2$.

3. ERW

For the ERW, a variational characterisation of $\psi(\lambda)$ is available following [28]. This construction also allows computation of the dominant paths shown in Fig. 2.

We outline the approach, which is to define a controlled process that *almost* achieves equality in (C1), up to a correction that vanishes on taking the limit in (C4). The typical path of this controlled model captures the mechanism of the (rare) fluctuations that achieve $q_t = q$ in the ERW. (Specifically, for large t and any $u > 0$, the conditional distribution of q_{ut} for paths that achieve $q_t = q$ is sharply peaked at $\langle q_{ut} \rangle_{\text{con}}$, see [28].)

We use (C1) with the controlled dynamics described in the main text for which $(b_1, b_2, \ldots, b_t)$ are variational parameters. The KL divergence between P_{con} and P is

$$\mathcal{D} = \frac{1}{2}\sum_{\tau=1}^{t}[(1+b_\tau)\log(1+b_\tau) + (1-b_\tau)\log(1-b_\tau)]$$
$$-\frac{1}{2}\sum_{\tau=1}^{t}(1+b_\tau)\langle\log(1+aq_{\tau-1})\rangle_{\text{con}}$$
$$-\frac{1}{2}\sum_{\tau=1}^{t}(1-b_\tau)\langle\log(1-aq_{\tau-1})\rangle_{\text{con}} , \quad \text{(C9)}$$

and we have

$$\langle q_\tau \rangle_{\text{con}} = \frac{1}{\tau}\sum_{k=1}^{\tau} b_k . \quad \text{(C10)}$$

Moreover, the variance of q_τ in this controlled process is at most $1/\tau$ so it is consistent to assume that q_τ is sharply peaked for almost all terms in the sums in (C9). Hence $\mathcal{D} \approx \hat{\mathcal{D}}$ with

$$\hat{\mathcal{D}} = \frac{1}{2}\sum_{\tau}[(1+b_\tau)\log(1+b_\tau) + (1-b_\tau)\log(1-b_\tau)]$$
$$-\frac{1}{2}\sum_{\tau}(1+b_\tau)\log(1+\langle aq_{\tau-1}\rangle_{\text{con}})$$
$$-\frac{1}{2}\sum_{\tau}(1-b_\tau)\log(1-\langle aq_{\tau-1}\rangle_{\text{con}}) . \quad \text{(C11)}$$

Using (C10) this is an explicit function of the b_τ variables, so the right-hand side of (C1) can be maximised numerically, which yields a numerical estimate of $G(\lambda, t)$ and hence (by considering large but finite t) one may estimate $\psi(\lambda)$.

For numerical work we use a similar method to that for the GERW: we split the sums in (C11) into contributions from small τ and large τ and we approximate the sum over large-τ contributions by an integral (which is also estimated numerically). This combination of sum and integral is maximised numerically to obtain estimates of $\psi(\lambda)$ and of the corresponding (average) path (C10). This yields the results of Fig. 2.

Appendix D: IGL mechanism in unidirectional hopping model

This Appendix establishes (48), which means that (37) holds for the model of Sec. V A, with $\beta = 1$. For this condition, it is sufficient to consider a finite-time interval between t_0 and τ^* (there is no large-time limit because we are focussing on the excursion that occurs at early times). For a compact notation we work on the interval $(t_0, \tau]$ and we write k for a generic time within this interval.

Consider a controlled process where the first hop is at time t_0 (as for the original model), after which hops take

place with a time-dependent rate $b(\tau)$. Then $(\tau q_\tau - 1)$ is Poissonian with mean $\int_{t_0}^{\tau} b(k)\mathrm{d}k$ and so

$$\tau\langle q_\tau\rangle_{\mathrm{con}} = 1 + \int_{t_0}^{\tau} b(k)\mathrm{d}k . \tag{D1}$$

The KL divergence of (C1) is

$$\mathcal{D} = \int_{t_0}^{\tau} \left\{ b(k)\left\langle \log\frac{b(k)}{aq_k}\right\rangle_{\mathrm{con}} - b(k) + \langle aq_k\rangle_{\mathrm{con}} \right\} \mathrm{d}k , \tag{D2}$$

similar to (C9). In addition to (C1), the KL divergence also allows a bound on the probability distribution of q_t. Roughly speaking, if one can construct a controlled process such that the large-deviation event occurs with probability one, $P_{\mathrm{con}}(q_\tau \geq q) = 1$, then the probability of this event in the original model can be bounded from below:

$$-\log P(q_\tau \geq q) \leq \mathcal{D}(P_{\mathrm{con}}\|P) . \tag{D3}$$

This may be proved by Jensen's inequality; a more precise statement is given (for example) in Equs. (14,15) of [16]. Hence we seek an upper bound on $\mathcal{D}$.

To achieve this, we use $\log(1/x) \leq (1/x) - 1$ with $x = q_k/\langle q_k\rangle_{\mathrm{con}}$ to write

$$\mathcal{D} \leq \int_{t_0}^{\tau} \left\{ b(k)\log\frac{b(k)}{a\langle q_k\rangle_{\mathrm{con}}} - 2b(k) + \langle aq_k\rangle_{\mathrm{con}} \right.$$
$$\left. + b(k)\langle q_k\rangle_{\mathrm{con}}\left\langle\frac{1}{q_k}\right\rangle_{\mathrm{con}} \right\}\mathrm{d}k . \tag{D4}$$

For a Poisson random variable X with mean $\overline{x}$, one has $\langle\frac{1}{1+X}\rangle = \mathrm{e}^{-\overline{x}}\sum_{n=0}^{\infty} \overline{x}^n/(n+1)! = (1-\mathrm{e}^{-\overline{x}})/\overline{x}$. Since $(kq_k - 1)$ is Poissonian, we obtain

$$\mathcal{D} \leq \int_{t_0}^{\tau} \left\{ b(k)\log\frac{b(k)}{\langle aq_k\rangle_{\mathrm{con}}} - 2b(k) + \langle aq_k\rangle_{\mathrm{con}} \right.$$
$$\left. + b(k)\langle kq_k\rangle_{\mathrm{con}}\frac{1-\mathrm{e}^{-\langle kq_k-1\rangle_{\mathrm{con}}}}{\langle kq_k-1\rangle_{\mathrm{con}}} \right\}\mathrm{d}k . \tag{D5}$$

To recover the results of [25] one should assume that $kq_k \gg 1$ throughout the integration range, so that the second line of the integrand reduces to $b(k)$. This is valid for $t_0 \gg 1$. Then one sets $\tau = t$ and minimises the resulting KL divergence over the path $\hat{q}(k) = \langle q_k\rangle_{\mathrm{con}}$, using (D1) to replace $b(k) \to (\partial/\partial k)(k\hat{q}(k))$. The optimal path behaves for short times as $k\hat{q}(k) = 1 + A[(k/t_0)-1]$ where A is proportional to the size of the giant excursion [25].

Our approach here does not require t_0 to be large: we retain all terms in (D5), and use (D3) with $\tau = \tau^*$ to establish (37). To obtain a convenient bound we set $\tau^* = 2t_0$ and choose $b(k)$ such that $\langle kq_k\rangle_{\mathrm{con}} = 1 + Ax$ with $x = (k/t_0) - 1$ and $A = 2q^*t_0 - 1$. This requires $b(k) = A/t_0$ and ensures that $\langle q_{\tau^*}\rangle_{\mathrm{con}} = q^*$. [Note, $b(k)$ is only independent of k for $k < \tau^*$ (i.e., during the excursion), the controlled process reverts to the natural dynamics of the model for $k > \tau^*$.] Then (D5) with $\tau = \tau^*$ becomes

$$\mathcal{D} \leq A\int_0^1 \left\{ \log\frac{A(1+x)}{a(Ax+1)} - 2 + \frac{a(Ax+1)}{A(1+x)} \right.$$
$$\left. + (1+Ax)\frac{1-\mathrm{e}^{-Ax}}{Ax} \right\}\mathrm{d}x . \tag{D6}$$

We are concerned with the limit $q^* \to \infty$ which corresponds to $A \to \infty$. The integral can be evaluated in this limit and the KL divergence scales as

$$\mathcal{D} \lesssim \gamma_{\mathrm{uni}}q^*t_0 \tag{D7}$$

with $\gamma_{\mathrm{uni}} = 2[\log(4/a) + a(1 - \log 2) - 1]$. To apply (D3) we require additionally that $P_{\mathrm{con}}(q_{\tau^*} \geq q^*) \to 1$ as $q^* \to \infty$: this holds because the distribution of q_{τ^*} is Poissonian with a diverging mean equal to q^*, so it is sharply peaked at q^*. Hence (D3) is applicable with KL divergence (D7) and the probability of the excursion obeys (48), as required.

---

[1] N. C. Keim and S. R. Nagel, Phys. Rev. Lett. **107**, 010603 (2011).

[2] C. Scalliet and L. Berthier, Phys. Rev. Lett. **122**, 255502 (2019).

[3] J. Kappler, J. O. Daldrop, F. N. Brünig, M. D. Boehle, and R. R. Netz, J. Chem. Phys. **148**, 014903 (2018).

[4] P. Van Mieghem and R. van de Bovenkamp, Phys. Rev. Lett. **110**, 108701 (2013).

[5] J. Zhang and T. Zhou, Proc. Natl. Acad. Sci. USA **116**, 23542 (2019).

[6] G. Rangarajan and M. Ding, eds., *Processes with Long-Range Correlations: Theory and Applications*, Lecture Notes in Physics, Vol. 621 (Springer-Verlag, Berlin Heidelberg, 2003).

[7] J. Beran, Y. Feng, S. Ghosh, and R. Kulik, *Long-Memory Processes: Probabilistic Properties and Statistical Methods*, berlin heidelberg ed. (Springer-Verlag, 2013).

[8] G. M. Schütz and S. Trimper, Phys. Rev. E **70**, 045101 (2004).

[9] S. Hod and U. Keshet, Phys. Rev. E **70**, 015104 (2004).

[10] E. Baur and J. Bertoin, Phys. Rev. E **94**, 052134 (2016).

[11] A. A. Budini, Phys. Rev. E **94**, 022108 (2016).

[12] A. A. Budini, Phys. Rev. E **95**, 052110 (2017).

[13] A. Rebenshtok and E. Barkai, Phys. Rev. Lett. **99**, 210601 (2007).

[14] K. Klymko, J. P. Garrahan, and S. Whitelam, Phys. Rev. E **96**, 042126 (2017).

[15] K. Klymko, P. L. Geissler, J. P. Garrahan, and S. Whitelam, Phys. Rev. E **97**, 032123 (2018).

[16] R. L. Jack, Phys. Rev. E **100**, 012140 (2019).

[17] R. J. Harris, New J. Phys. **17**, 053049 (2015).

[18] F. den Hollander, *Large deviations* (American Mathematical Society, Providence, RI, 2000).

[19] V. Lecomte, C. Appert-Rolland, and F. van Wijland, J. Stat. Phys. **127**, 51 (2007).

[20] B. Derrida, J. Stat. Mech. **2007**, P07023 (2007).

[21] H. Touchette, Phys. Rep. **478**, 1 (2009).

[22] R. Chétrite and H. Touchette, Ann. Henri Poincaré **16**, 2005 (2015).

[23] R. L. Jack, Eur. Phys. J. B **93**, 74 (2020).

[24] R. J. Harris and H. Touchette, J. Phys. A **42**, 342001 (2009).

[25] R. J. Harris, J. Stat. Mech. **2015**, P07021 (2015).

[26] C. Maes, K. Netočný, and B. Wynants, J. Phys. A **42**, 365002 (2009).

[27] A. Faggionato, arXiv:1709.05653 (2017).

[28] S. Franchini, Stoch. Process. Appl. **127**, 3372 (2017).

[29] B. Bercu, J. Phys. A **51**, 015201 (2017).

[30] V. M. Kenkre, arXiv:0708.0034.

[31] J. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).

[32] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, Phys. Rev. Lett. **98**, 195702 (2007).

[33] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Science **323**, 1309 (2009).

[34] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Phys. Rev. Lett. **116**, 120601 (2016).

[35] M. D. Donsker and S. R. S. Varadhan, Comm. Pure Appl. Math **28**, 1 (1975).

[36] M. D. Donsker and S. R. S. Varadhan, Comm. Pure Appl. Math **28**, 279 (1975).

[37] M. D. Donsker and S. R. S. Varadhan, Comm. Pure Appl. Math **29**, 389 (1976).

[38] M. D. Donsker and S. R. S. Varadhan, Comm. Pure Appl. Math **36**, 183 (1983).

[39] H. Touchette, Physica A **504**, 5 (2018).

[40] D. Nickelsen and H. Touchette, Phys. Rev. Lett. **121**, 090602 (2018).

[41] G. Gradenigo and S. N. Majumdar, J. Stat. Mech. **2019**, 053206 (2019).

[42] B. Meerson, Phys. Rev. E **100**, 042135 (2019).

[43] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations* (Wiley, 1997).

[44] F. N. C. Paraan and J. P. Esguerra, Phys. Rev. E **74**, 032101 (2006).

[45] M. A. A. da Silva, J. C. Cressoni, G. M. Schütz, G. M. Viswanathan, and S. Trimper, Phys. Rev. E **88**, 022115 (2013).

[46] R. Chétrite and H. Touchette, J. Stat. Mech. **2015**, P12001 (2015).

[47] C. Maes and K. Netočný, EPL **82**, 30003 (2008).

[48] A. Simha, R. M. L. Evans, and A. Baule, Phys. Rev. E **77**, 031117 (2008).

[49] D. Simon, J. Stat. Mech. **2009**, P07017 (2009).

[50] R. L. Jack and P. Sollich, Prog. Theor. Phys. Supp. **184**, 304 (2010).

[51] R. L. Jack and P. Sollich, Eur. Phys. J.: Spec. Topics **224**, 2351 (2015).

[52] S. Grosskinsky, G. M. Schütz, and H. Spohn, J. Stat. Phys. **113**, 389 (2003).

[53] M. R. Evans and B. Waclaw, J. Phys. A: Math. Theor. **47**, 095001 (2014).

[54] A. Vezzani, E. Barkai, and R. Burioni, Phys. Rev. E **100**, 012108 (2019).

[55] C. Appert-Rolland, B. Derrida, V. Lecomte, and F. van Wijland, Phys. Rev. E **78**, 021122 (2008).

[56] V. Lecomte, J. P. Garrahan, and F. van Wijland, J. Phys. A **45**, 175001 (2012).

[57] M. Cavallaro and R. J. Harris, J. Phys. A **49**, 47LT02 (2016).

[58] K. Nagel, Phys. Rev. E **53**, 4655 (1996).