The convex hull of the run-and-tumble particle in a plane

Alexander K Hartmann [1], Satya N Majumdar[2], Hendrik Schawe[1,3], and Grégory Schehr[2]

[1] Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany
[2] LPTMS, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France
[3] Laboratoire de Physique Théorique et Modélisation, UMR-8089 CNRS-Université Cergy Pontoise, France

E-mail: alexander.hartmann@uol.de

E-mail: satya.majumdar@u-psud.fr

E-mail: hendrik.schawe@u-cergy.fr

E-mail: gregory.schehr@u-psud.fr

Abstract. We study the statistical properties of the convex hull of a planar run-and-tumble particle (RTP), also known as the "persistent random walk", where the particle/walker runs ballistically between tumble events at which it changes its direction randomly. We consider two different statistical ensembles where we either fix (i) the total number of tumblings n or (ii) the total duration t of the time interval. In both cases, we derive exact expressions for the average perimeter of the convex hull and then compare to numerical estimates finding excellent agreement. Further, we numerically compute the full distribution of the perimeter using Markov chain Monte Carlo techniques, in both ensembles, probing the far tails of the distribution, up to a precision smaller than 10^{-100}. This also allows us to characterize the rare events that contribute to the tails of these distributions.

1. Introduction

Random walks belong probably to the most thoroughly studied class of stochastic processes [1, 2, 3] with a wide variety of applications ranging from finance [4] to biology [5] or online search [6]. In its simplest form a random walk consists of the sum of independent and identically distributed random jumps, which converges, in the long time limit when suitably scaled, to Brownian motion (provided the jump distribution has a finite variance). However, to be a useful model for applications, several variants of this simple model have been introduced and studied. For example, one may introduce correlations between the steps to model animal movement [7, 8] or polymers [9]. One can also consider interactions of the walker with its environment to model organisms that are driven by concentration gradients [10, 11].

Here, we will focus on yet another variant, which has found interesting applications in modeling active matter and active particles. The term "active particle" refers to a class of self-propelled particles which, in contrast to "passive" particles such as Brownian motion, can generate dissipative directed motion by consuming energy directly from their environment [12, 13, 14]. Examples of active matter arise in a wide variety of biological and soft matter systems, including bacterial motion [5, 15, 16, 17], cellular tissue behavior [18], formation of fish schools [19, 20] as well as flock of birds [21, 22], amongst others. In this context, one of the most studied model is the run-and-tumble particle (RTP) [23, 24], initially introduced under the name of the "persistent random walk" [25, 26]. As illustrated in Fig. 1, an RTP performs a ballistic motion along a certain direction at a constant speed $v_0 \geq 0$ ("run") during a certain "time of flight" after which it "tumbles", i.e., chooses a new direction uniformly at random. Then it performs a new run along this new direction again with speed v_0 and so on. The tumblings occur instantaneously at random times with constant rate , i.e., the of different runs are independently distributed via an exponential distribution $p( ) = e^{-\gamma\tau}$. Despite its simplicity, this RTP model exhibits complex interesting features such as clustering at boundaries [13], non-Boltzmann distribution in the steady state in the presence of a confining potential [23, 27, 28, 29, 30], motility-induced phase separation [24], jamming [31], etc.

An interesting question related to the study of animal movements concerns the home range of an animal, i.e., the two-dimensional territory it covers while searching for food during a certain period of time [32]. This is a particularly useful information for ecologists to decide and design habitat-conservation planning. A convenient way to estimate this home range is to construct the convex hull of the trajectory, i.e., the smallest convex polygon containing every point visited by the walker (see Fig. 1). The perimeter and/or the area of this convex hull provide quantitative estimates of this home range.

For Brownian motion, the statistics of the convex hull is a classical problem in probability theory [33] and random convex geometry [34]. Quite recently, exploiting a beautiful connection to extreme value statistics, exact results have been obtained for the mean perimeter and mean area for the convex hull of multiple planar Brownian motions [35, 36], as well as for the convex hull of a single randomly accelerated particle (also called "the integrated Brownian motion") in two dimensions [37]. Extensions of these studies of the convex hull for Brownian motion in higher dimensions have also been discussed, mainly in the mathematics literature [38, 39]. The convex hull of random walks consisting of a finite number of discrete jumps drawn from some distribution has also been widely studied, both in the mathematics literature

[40, 41, 42, 43, 44] and more recently in physics [45, 46, 47]. In particular, using sophisticated sampling methods, higher moments, like the variance, as well as the full distributions of the perimeter and area of the convex hull of random walks in the plane [46] and in higher dimensions [47] were recently obtained numerically.

The convex hull of several variants of the random walk has been studied, including Lévy flights [48], continuous-time random walks [49], branching Brownian motion [51] or self-avoiding random walks [52, 53]. However, to the best of our knowledge nothing is known about the convex hull of a planar RTP, which, given the wide range of applications of this model, is clearly an interesting and important observable. The purpose of the present paper is to provide a detailed study, both analytical and numerical, of the perimeter of the convex hull of an RTP in the plane.

The rest of the paper is organized as follows. In Section 2 we briefly introduce the RTP model and present our main results. We proceed with Section 3 where we present the details of our analytical computations. In Section 4 we outline the numerical methods we have used to simulate RTP, to calculate convex hulls as well as how to obtain the desired distributions for the perimeter of the convex hull even in the tails down to extremely small probabilities, namely smaller than 10^{-100}. Finally, we present our short conclusions in Section 5.

2. Model: The run-and-tumble particle and its convex hull

We consider a single run-and-tumble particle in two dimensions (see Fig. 1). The dynamics occurs in continuous time and is defined as follows. The particle starts at the origin with an initial velocity of fixed magnitude v_0 and chooses a direction at random, i.e., the angle specifying the direction is chosen uniformly from $[0; 2\ ]$. Subsequently, in a small time dt, with probability dt , the particle changes its direction of flight by choosing a new angle uniformly in $[0; 2\ ]$—this is the "tumbling". Otherwise with the complementary probability $1 -\ dt$, the particle continues to move ballistically with speed v_0 in its current direction. Thus v_0 and are the only two parameters in this model. The parameter denotes the rate of tumbling. The distance covered between two successive tumblings is called a "run". We count the starting point 0 as a tumble and hence the number of tumblings is the same as the number of runs n and by definition $n \geq 1$. A typical trajectory for an RTP in the plane is shown in Fig. 1.

We consider two different ensembles: (i) an ensemble where we consider exactly n runs, i.e., we stop the process exactly after n runs—here the total time spent by the particle fluctuates from sample to sample while the number of runs n is fixed and (ii) an ensemble where the total observation time t is fixed—here the number of runs n fluctuates from sample to sample while the total duration t is fixed. In this paper, we are interested in the convex hull containing the trajectory of the RTP as an intuitive measure of the geometric size of the RTP. The convex hull is the smallest convex polygon enclosing a set of points, in this case the set of points visited by the RTP—imagine a rubber band around the points: it will contract and become their convex hull (see Fig. 1).

For both ensembles, we consider each trajectory of the RTP, draw the unique convex hull associated to it and compute the mean perimeter of the convex hull by averaging over all trajectories. Our main analytical results are exact results for the mean perimeter $\langle L_n \rangle$ in Eq. (19) for the ensemble (i) and $\langle L(t) \rangle$ in Eqs. (43) and (44) for the ensemble (ii). Besides, we perform simulations of this model to sample trajectories and measure the distribution of the perimeter of their hulls. In particular,

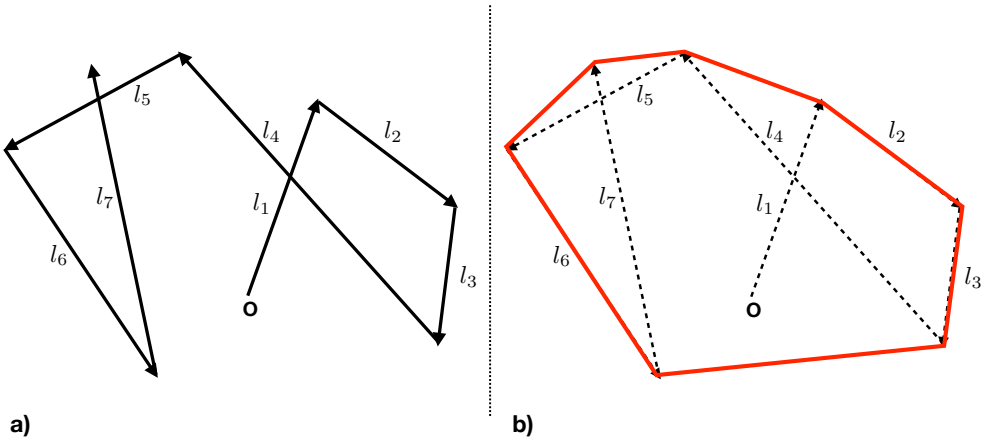

Figure 1. **a**) A typical trajectory of an RTP in two dimensions with $n = 7$ steps. The particle starts at the origin O, chooses a random direction and moves ballistically in that direction a distance $l_1 = v_0\,\tau_1$ where v_0 is constant and τ_1 is a random time drawn from the exponential distribution $p(\tau) = \gamma\, e^{-\gamma\tau}$. At the end of this first flight, the particle tumbles instantaneously and chooses a new random direction and again moves ballistically a distance $l_2 = v_0\,\tau_2$ with τ_2 drawn independently from $p(\tau) = \gamma\, e^{-\gamma\,\tau}$. Then the particle tumbles again and so on. The schematic figure shows a trajectory of an RTP with n tumblings, with the starting point is counted as a tumble. **b**) Same trajectory of an RTP where we have depicted, in red, the convex hull.

our numerical results for the mean perimeter are in very good agreement with our analytical predictions (see Fig. 4).

3. Exact mean perimeter of the convex hull of a run-and-tumble particle in two dimensions

To compute the mean perimeter of the convex hull of a $2\mathsf{d}$ RTP, in either fixed-n or fixed-t ensemble, we use the strategy developed in Refs. [35, 36]. It was shown in Refs. [35, 36] how Cauchy's formula [54] for the perimeter of an arbitrary convex curve in two dimensions can be applied to calculate the mean perimeter of the convex hull of a generic $2\mathsf{d}$ stochastic process. Using this procedure, the problem of computing the mean perimeter of the convex hull of an arbitrary $2\mathsf{d}$ stochastic process can be mapped to computing the maximum of the one dimensional component process [35, 36]. Let us briefly outline the key idea. Consider an arbitrary convex domain $\mathcal{D}$ in two dimensions with its boundary $\mathcal{C}$ parametrized as $\{\mathcal{X}(\mathsf{s}); \mathcal{Y}(\mathsf{s})\}$ with s denoting the arc distance along the boundary contour $\mathcal{C}$. Cauchy's formula states that the perimeter of the convex domain $\mathcal{D}$ is given by

$$\mathsf{L} = \int_0^{2\pi} \mathsf{M}(\ )\,\mathsf{d}\ ; \tag{1}$$

where $\mathsf{M}(\ )$ is the so called support function

$$\mathsf{M}(\ ) = \max_{s} \left[\mathcal{X}(\mathsf{s})\cos(\ ) + \mathcal{Y}(\mathsf{s})\sin(\ )\right] : \tag{2}$$

The quantity $M(\ )$ can be interpreted as follows: Consider the projections of all points of the boundary curve $\mathcal{C}$ along the direction and take the maximum of those projections.

Consider now an arbitrary set of n vertices $\{(X_i; Y_i);\ i = 1; 2; \ldots; n\}$ in $2d$ (they may represent the positions of a stochastic process in $2d$ at successive times in a given realization) and construct the convex hull $\mathcal{C}$ of these vertices. The perimeter of the convex hull is given by Cauchy's formula in Eq. (1). To apply this formula, we need to first evaluate $\{\mathcal{X}(s); \mathcal{Y}(s)\}$ of the convex hull $\mathcal{C}$ and then compute its maximum over s which seems to be a formidable task. The key observation of Refs. [35, 36] that bypasses this step is that the support function $M(\ )$ of the convex hull can be obtained directly from the underlying vertices (without the need to first compute $\{\mathcal{X}(s); \mathcal{Y}(s)\}$ of $\mathcal{C}$ and then maximizing over s) as

$$M(\ ) = \max_{1 \le i \le n} [X_i \cos(\ ) + Y_i \sin(\ )] : \tag{3}$$

Next we average Eq. (1) over all realizations of the stochastic process, i.e., over different realizations of the vertices $\{(X_i; Y_i)\}$ to get

$$\langle L_n \rangle = \int_0^{2\pi} \langle M(\ ) \rangle\, d\ : \tag{4}$$

Moreover, if the $2d$ process is isotropic (e.g. the RTP process in $2d$ is isotropic), $\langle M(\ ) \rangle$ can not depend on —hence, we may as well put $= 0$. This then simplifies Cauchy's formula and amounts to computing just the expected maximum of the one-dimensional component process [35, 36]

$$\langle L_n \rangle = 2 \quad \langle M_n \rangle \quad \text{where} \quad M_n = \max [X_1; X_2; \ldots; X_n] : \tag{5}$$

This mapping holds for any arbitrary $2d$ isotropic stochastic process. In recent years, this procedure has been successfully used to compute the mean perimeter for several $2d$ stochastic processes. These include a single/multiple planar Brownian motions [35, 36], planar random acceleration process [37], $2d$ branching Brownian motion with absorption in the context of epidemic outbreak [51], anomalous diffusion processes in $2d$ [49], a $2d$ Brownian motion confined in the half-plane [55, 56], discrete-time $2d$ random walks, Lévy flights [45], etc. Below we demonstrate that this procedure is also suitable to compute exactly the mean perimeter of the convex hull of a $2d$ RTP, both in the fixed-n and the fixed-t ensemble.

3.1. Fixed-n ensemble

In this ensemble, the total number of runs n of the RTP is fixed, but the duration t fluctuates from sample to sample. Since the RTP process is isotropic, we can use the general result in Eq. (5). For this, we need to first evaluate the probability distribution of the coordinates $\{X_1; X_2; \ldots; X_n\}$ of the x-component of the $2d$ RTP. To proceed, consider a particular run, say the i-th run in Fig. 1. The length of the run is $l_i = v_0\ _i$, where $_i$ is distributed exponentially $p(\ _i) = e^{-\gamma \tau_i}$. When projected along the x-axis, this corresponds to an increment $x_i = l_i \cos(\ _i)$ in the x-direction, where the angle $_i$ is distributed uniformly over $_i \in [0; 2\ ]$. Let us write, $x_i = v\ _i$, where $v = v_0 \cos(\ _i)$. Given the uniform distribution of $_i$, it is easy to compute the distribution $P(v)$ of v using $P(v)dv = d\ _i/(2\ )$ and we get

$$P(v) = \frac{1}{\sqrt{v_0^2 - v^2}}; \qquad -v_0 \le v \le v_0 : \tag{6}$$

Consequently, the joint distribution $p(x;\ )$ of the increment $x_i = v\ _i$ (along the x-direction) and the duration $\ _i$ of the i-th run is given by

$$p(x;\ ) = \text{Prob.}[x_i = x;\ \ _i = \ ] = \int_{-v_0}^{v_0} dv\ \ (x - v\ )\ \frac{1}{\sqrt{v_0^2 - v^2}}\ \ e^{-\gamma\ \tau}\ : \tag{7}$$

Integrating over v gives

$$p(x;\ ) = \frac{e^{-\gamma\ \tau}}{\sqrt{v_0^2\ \ ^2 - x^2}}\ \ \left(\ \ - \frac{x}{v_0}\right)\ ; \tag{8}$$

where $\ (z)$ is the standard Heaviside step function: $\ (z) = 1$ if $z > 0$ and $\ (z) = 0$ if $z < 0$.

Integrating further over $\ $, one sees that the increment x_i is distributed via the marginal probability density

$$f\ (x) = \text{Prob.}[x_i = x] = \int_0^\infty p(x;\ )\, d\ \ = -\int_{x/v_0}^\infty \frac{d\ \ e^{-\gamma\ \tau}}{\sqrt{v_0^2\ \ ^2 - x^2}} = \frac{}{v\ _0}\, K\ _0\left(\frac{|x|}{v_0}\right)\ ; \tag{9}$$

where $K\ _\nu(z)$ is the modified Bessel function with index $\ $. One can check easily that $f\ (x)$ is normalized to unity: $\int_{-\infty}^{\infty} f\ (x)dx = 1$. Furthermore, the variance $\ ^2$ of this distribution is finite and is given by

$$\ ^2 = \int_{-\infty}^{\infty} x^2\, f\ (x)\, dx = \frac{2}{}\left(\frac{v_0}{}\right)^2 \int_0^\infty y^2\, K\ _0(y)\, dy = \left(\frac{v_0}{}\right)^2\ : \tag{10}$$

Thus, each run of the $2d$ RTP gives rise to an independent increment x in the x-direction which is distributed via the symmetric, continuous, normalized to unity probability distribution function (PDF) $f\ (x)$ in Eq. (9). Thus the projected x-component process is just a one dimensional random walk, $X\ _i = X\ _{i-1} + x_i$ with independent and identically distributed (i.i.d.) increments x_i, each drawn from $f\ (x)$ in Eq. (9), and starting at $X\ _0 = 0$. Hence, we need to now compute the expected maximum $\langle M\ _n\rangle$ of this reduced one dimensional random walk process of n steps. This can be conveniently computed using a formula originally due to Kac [40] (see also Ref. [41])

$$\langle M\ _n\rangle = \frac{1}{2}\ \sum_{m=1}^{n} \frac{\langle|X\ _m|\rangle}{m}\ : \tag{11}$$

To compute $\langle|X\ _m|\rangle$, we need to know the distribution $P\ _m(X\ ) = \text{Prob.}(X\ _m = X\ )$ of the position of the $1d$ random walker at step m. This can be easily computed as follows. Clearly, $X\ _m = x_1 + x_2 + \cdots + x_m$ where x_i are i.i.d. random variables each drawn from $f\ (x)$. Hence

$$P\ _m(X\ ) = \int_{-\infty}^{\infty}\ \ \left(X\ - \sum_{i=1}^{m} x_i\right)\ \prod_{i=1}^{n} f\ (x_i)\, dx_i\ : \tag{12}$$

Taking the Fourier transform gives

$$\tilde{P}_m(k) = \int_{\infty}^{\infty} P_m(X\ )\, e^{i\ k\ X}\, dX\ = \left[\tilde{f\ }(k)\right]^m\ ; \tag{13}$$

where the Fourier transform of the jump distribution $f\ (x)$ in Eq. (9) turns out to be rather simple

$$\tilde{f\ }(k) = \frac{}{v\ _0}\ \int_{-\infty}^{\infty} e^{i\ k\ x}\, K\ _0\left(\frac{|x|}{v_0}\right)\, dx = \frac{1}{\sqrt{1 + \frac{v_0^2\ k^2}{\gamma^2}}}\ : \tag{14}$$

Substituting this in Eq. (13) and inverting the Fourier transform (fortunately it can be done explicitly) we get

$$\begin{aligned} P_m(X) &= \int_{-\infty}^{\infty} \left[1 + \frac{v_0^2 k^2}{2}\right]^{-m/2} e^{-i k X}\, dk \\ &= \frac{2^{(1-m)/2}}{v_0 \sqrt{\Gamma(m/2)}} \left(\frac{|X|}{v_0}\right)^{(m-1)/2} K_{(1-m)/2}\left(\frac{|X|}{v_0}\right) : \end{aligned} \quad (15)$$

where $\Gamma(z)$ is the standard Gamma function.

From this exact PDF of X_m in Eq. (15), one can easily compute $\langle |X_m| \rangle$

$$\langle |X_m| \rangle = \int_{-\infty}^{\infty} |X|\, P_m(X)\, dX = \frac{v_0}{\sqrt{\Gamma(m/2)}} \int_0^{\infty} y^{(m+1)/2} K_{(1-m)/2}(y)\, dy \quad (16)$$

$$= \frac{2 v_0}{\sqrt{\Gamma(m/2)}} : \quad (17)$$

Substituting this result in Kac's formula (11) and using Eq. (5) gives, for any $n \geq 1$,

$$\langle L_n \rangle = \frac{v_0 \sqrt{\ }}{} \sum_{m=1}^{n} \frac{\Gamma\left(\frac{m}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{m}{2} + 1\right)} : \quad (18)$$

This sum can be done explicitly and we get

$$\langle L_n \rangle = \frac{v_0}{} \left[-(\ \ + 2) + 2\sqrt{\ } \left\{ \frac{\Gamma\left(2 + \lfloor \frac{n-1}{2} \rfloor\right)}{\Gamma\left(\frac{3}{2} + \lfloor \frac{n-1}{2} \rfloor\right)} + \frac{\Gamma\left(\frac{3}{2} + \lfloor \frac{n}{2} \rfloor\right)}{\Gamma\left(1 + \lfloor \frac{n-1}{2} \rfloor\right)} \right\} \right] ; \quad (19)$$

where $\lfloor z \rfloor$ denotes the integer part of z. It is easy to check that for large n one gets the asymptotic behavior

$$\langle L_n \rangle \to \frac{v_0}{} \sqrt{8\, n} \quad \text{as } n \to \infty : \quad (20)$$

Using $= v_0/$ from Eq. (10), we see that for large n, $\langle L_n \rangle \to \sqrt{8\ n}$ which coincides with the mean perimeter of the convex hull of a discrete two dimensional random walk of n steps for any jump distribution with a finite variance ^{2} [45].

3.2. Fixed-t ensemble

Here, we consider the total duration t fixed, but the number of runs n during t may fluctuate from sample to sample and hence n is a random variable. Note that the duration $_n$ of the last interval (i.e., following the n-th tumbling) traveled by the particle before the epoch t is yet to be complete. Consequently, its distribution is $e^{-\gamma \tau_n}$ (denoting the probability of no tumbling during the interval of duration $_n$) is not normalized to unity. This is in contrast to the already completed preceding intervals each of which is distributed independently according to the normalized distribution $p(\ ) = \ e^{-\gamma \tau}$. For each of these preceding intervals, the joint distribution of the x-component and the duration of the interval is given by $p(x;\ \ )$ in Eq. (8) and each interval is independent of the other. The corresponding joint distribution for the last (incomplete run) interval, in contrast, is different by a factor $1/$

$$p_{\rm last}(x;\ \ ) = \frac{e^{-\gamma \tau}}{\sqrt{v_0^2\ ^2 - x^2}} \quad \left(\ \ - \frac{x}{v_0} \right) = \frac{1}{} p(x;\ \ ) ; \quad (21)$$

where $p(x;\ \ )$ is given in Eq. (8). Note however that the sum of the durations of the $(n-1)$ completed runs and the last incomplete run is fixed to be t. Hence, the grand

joint distribution of the x-increments $\{x_i\}$, their associated durations $\{\ _i\}$ (where $i = 1; 2; \ldots; n$) and the number of runs n is given by

$$P\ [\{x_i\};\ \{\ _i\};\ n|t] = \frac{1}{-}\ \left[\prod_{i=1}^{n} p(x_i;\ \ _i)\right]\ \ \left(\sum_{i=1}^{n}\ _i - t\right)$$
$$= \frac{1}{-}\ \left[\prod_{i=1}^{n} \frac{e^{-\gamma\,\tau_i}}{\sqrt{v_0^2\ \ ^2 - x_i^2}}\ \ \left(\ _i - \frac{x_i}{v_0}\right)\right]\ \ \left(\sum_{i=1}^{n}\ _i - t\right) : \tag{22}$$

The presence of the delta function on the right hand side (rhs) of Eq. (22) naturally signals that it is convenient to work in the Laplace space conjugate to t. Taking the Laplace transform of Eq. (22) with respect to t and integrating over the $\ _i$ variables, we obtain the Laplace transform of the marginal joint distribution $P\ [\{x_i\}; n|t]$ in a factorized form

$$\int_0^\infty P\ [\{x_i\};\ n|t]\ e^{-s\,t}\,dt = \frac{1}{-}\ \left[\prod_{i=1}^{n}\int_{x_i/v_0}^{\infty} \frac{e^{-(\gamma+s)\,\tau_i}}{\sqrt{v_0^2\ \ ^2 - x_i^2}}\,d\ _i\right]$$
$$= \frac{1}{-}\ \left[\prod_{i=1}^{n} \frac{}{v_0}\,K_0\left(\frac{(\ \ +s)\,|x_i|}{v_0}\right)\right]\ ; \tag{23}$$

where we used the result from Eq. (9). To proceed further, it is convenient to define a normalized (to unity) PDF $f_s(x)$ parametrized by s as follows

$$f_s(x) = \frac{K_0\left(\frac{(\gamma+s)\,|x|}{v_0}\right)}{\int_{-\infty}^{\infty} K_0\left(\frac{(\gamma+s)\,|x|}{v_0}\right)\,dx} = \frac{+s}{v_0}\,K_0\left(\frac{(\ \ +s)\,|x|}{v_0}\right) : \tag{24}$$

In terms of this normalized PDF $f_s(x)$, we can rewrite Eq. (23) in a convenient form as

$$\int_0^\infty P\ [\{x_i\};\ n|t]\ e^{-s\,t}\,dt = \frac{1}{-}\ \left(\frac{}{+s}\right)^n \prod_{i=1}^{n} f_s(x_i) : \tag{25}$$

We next invert the Laplace transform formally as

$$P\ [\{x_i\};\ n|t] = \int_\Gamma \frac{ds}{2\,i}\,e^{s\,t}\,\frac{1}{-}\ \left(\frac{}{+s}\right)^n \prod_{i=1}^{n} f_s(x_i)\,; \tag{26}$$

where Γ denotes a vertical Bromwich contour (to the right of all singularities of the integrand) in the complex s plane. Note that the increments x_i are correlated since the rhs of Eq. (26) does not factorize.

So, once again, the projected x-increments $\{x_i\}$ ($i = 1; 2 : \ldots; n$) form a random walk process in $1d$ where the position X_i of the walker evolves as $X_i = X_{i-1} + x_i$, starting from $X_0 = 0$. However, unlike in the fixed-n ensemble, the increments in the fixed-t ensemble are not independent random variables, but are correlated as in Eq. (26). To compute the mean perimeter $\langle L\,(t)\rangle$ of the convex hull for fixed t using Eq. (5), we need to compute the expected maximum of the $1d$ process X_i with fixed t whose increments and number of steps n are jointly distributed via Eq. (26), and finally sum over all n. It is convenient to define the following quantity

$$Q(M; n\,|t) = \int_{-\infty}^{\infty} dx_1 : \ldots \int_{-\infty}^{\infty} dx_n\ \text{Prob.}\,[X_1 < M; X\ _2 < M; \ldots; X\ _n < M; n\ |t]\ ; \tag{27}$$

$$\text{with} \qquad X_i = \sum_{j=1}^{i} x_j\ ; \tag{28}$$

that denotes the joint probability that the maximum of the random walk process is less than M and the number of steps is exactly n, for a given t. Taking a derivative with respect to M gives the joint PDF of M and n: $P(M;n\,|t) = \partial_M Q(M;n\,|t)$. Using Eq. (5) we then have

$$\langle L(t)\rangle = 2 \quad \sum_{n=1}^{\infty} \int_0^{\infty} dM\, M\, \partial_M Q(M;n\,|t)\,: \tag{29}$$

Using the joint distribution in Eq. (26) and the definition $Q(M;n\,|t)$ in Eq. (27) we then get formally

$$\langle L(t)\rangle = 2 \quad \sum_{n=1}^{\infty} \int_{\Gamma} \frac{ds}{2\,i}\, e^{s\,t}\, \frac{1}{} \left(\frac{}{+s}\right)^{n} \langle M_s(n)\rangle\,; \tag{30}$$

with

$$\langle M_s(n)\rangle = \int_0^{\infty} dM\, M\, \partial_M Q_s(M;n\,) \tag{31}$$

where

$$Q_s(M;n\,) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} f_s(x_i)\, dx_i\right] \text{Prob.}\,[Y_1 < M;\, Y_2 < M;\, \ldots;\, Y_n < M\,] \tag{32}$$

$$\text{with} \qquad Y_i = \sum_{j=1}^{i} x_j\,: \tag{33}$$

Thus, $Q_s(M;n\,)$ in Eq. (33) can be interpreted as the cumulative distribution of the maximum of an auxiliary $1d$ random walk process $\{Y_i\}$ of n-steps, $Y_i = Y_{i-1} + x_i$, with i.i.d. increments x_i each distributed via $f_s(x)$ given in Eq. (24). Consequently, $\langle M_s(n)\rangle$ in Eq. (31) is simply the expected maximum of this auxiliary n-step $1d$ process $\{Y_i\}$ parametrized by s. However, as mentioned before, the expected maximum of any $1d$ random walk process with i.i.d. increments can be computed using Kac's formula in Eq. (11), provided we can compute the expected absolute value of the process $\langle |Y_m|\rangle$ at step m. Then, we get

$$\langle M_s(n)\rangle = \frac{1}{2} \sum_{m=1}^{n} \frac{\langle |Y_m|\rangle}{m} \tag{34}$$

The average of $|Y_m|$ can be computed explicitly as in the case of the fixed-n ensemble discussed before. Using $Y_m = x_1 + x_2 + \cdots x_m$ where x_i are i.i.d. variables each drawn from the PDF $f_s(x)$, the PDF $P_m(Y\,)$ of Y_m can be computed in terms of the Fourier transform of the jump distribution $f_s(x)$ as in Eq. (13)

$$P_m(Y\,) = \int_{-\infty}^{\infty} \left[\tilde{f_s}(k)\right]^m\, e^{-i\,k\,Y}\, \frac{dk}{2}\,; \tag{35}$$

where

$$\tilde{f_s}(k) = \int_{-\infty}^{\infty} f_s(x)\, e^{i\,k\,x}\, dx = \frac{1}{\sqrt{\left(\frac{v_0}{\gamma+s}\right)^2 k^2 + 1}}\,: \tag{36}$$

We have used the explicit expression of $f_s(x)$ from Eq. (24) and the result in Eq. (14) where we replaced γ by $\gamma + s$. Substituting Eq. (36) in Eq. (35) and performing the integral (as in Eq. (15) with γ replaced $\gamma + s$) we get

$$P_m(Y\,) = \frac{(\gamma+s)\, 2^{(1-m)/2}}{v_0\, \sqrt{\pi}\, \Gamma(m/2)} \left(\frac{(\gamma+s)\,|Y\,|}{v_0}\right)^{(m-1)/2} K_{(1-m)/2}\left(\frac{(\gamma+s)\,|Y\,|}{v_0}\right)\,: \tag{37}$$

This then gives

$$\langle |Y_m| \rangle = \int_{-\infty}^{\infty} |Y|\, P_m(Y)\, dY = \frac{2 v_0}{+ s}\, \frac{\Gamma\left(\frac{1+m}{2}\right)}{\sqrt{\ }\,\Gamma(m{=}2)}\,: \tag{38}$$

Substituting this result in Eq. (34) we get

$$\langle M_s(n) \rangle = \frac{v_0}{(\ \ + s)\sqrt{4\ }} \sum_{m=1}^{n} \frac{\Gamma\left(\frac{m}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{m}{2}+1\right)} \tag{39}$$

and consequently Eq. (30) yields

$$\langle L(t) \rangle = v_0 \sqrt{\ } \sum_{n=1}^{\infty} \ \ ^{n-1} \left[\sum_{m=1}^{n} \frac{\Gamma\left(\frac{m}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{m}{2}+1\right)}\right] \int_{\Gamma} \frac{ds}{2\,i}\, e^{s\,t}\, \frac{1}{(\ \ + s)^{n+1}}\,: \tag{40}$$

The Bromwich integral can now be trivially performed using

$$\int_{\Gamma} \frac{ds}{2\,i}\, e^{s\,t}\, \frac{1}{(\ \ + s)^{n+1}} = \frac{t^n\, e^{-\gamma\, t}}{\Gamma(n+1)}\,: \tag{41}$$

This then gives us the formula for $\langle L(t) \rangle$ in the fixed-t ensemble

$$\langle L(t) \rangle = \frac{v_0 \sqrt{\ }}{\ } e^{-\gamma\, t} \sum_{n=1}^{\infty} \frac{(\ t\ )^n}{n!} \left[\sum_{m=1}^{n} \frac{\Gamma\left(\frac{m}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{m}{2}+1\right)}\right]\,: \tag{42}$$

The result in Eq. (42) is valid at any time t. Furthermore, it turns out that the double summation can be performed explicitly and $\langle L(t) \rangle$ can be written in a scaling form

$$\langle L(t) \rangle = \frac{v_0}{\ } H(\ \ t\ )\,; \tag{43}$$

where the scaling function $H(z)$ is given exactly by

$$H(z) = e^{-z}\ [2 - (\ \ + 2)\, e^{z} + 2\, z + \ \ \ (1 + z)\, (I_0(z) + L_0(z)) + \ \ z\ \ (I_1(z) + L_1(z))] \tag{44}$$

In Eq. (44), $I_\nu(z)$ is the standard modified Bessel functions of index \ \ , while $L_\nu(z)$ denotes the Struve function of index \ \ defined as

$$L_\nu(z) = \left(\frac{z}{2}\right)^{\nu+1} \sum_{k=0}^{\infty} \frac{1}{\Gamma\left(\frac{3}{2}+k\right)\Gamma\left(\frac{3}{2}+k+\ \ \right)} \left(\frac{z}{2}\right)^{2k}\,: \tag{45}$$

The function $H(z)$ in Eq. (44) can be plotted using *Mathematica*—a plot of $\langle L(t) \rangle$ vs. t for $v_0 = 1$ and \ \ $= 1$ is shown in Fig. 2. The scaling function $H(z)$ has the following asymptotic behaviors

$$H(z) = \begin{cases} 2\, z + \dfrac{-\ 4}{4}\, z^2 + O(z^3)\,; & \text{as } z \to 0\ \ , \\ \sqrt{8\ \ z\ } - (\ \ + 2) + O\left(\dfrac{1}{\sqrt{z}}\right)\,; & \text{as } z \to \infty\ \ . \end{cases} \tag{46}$$

Consequently, the mean perimeter $\langle L(t) \rangle$ initially grows ballistically for $t \ll \ \ ^{-1}$ and eventually for $t \gg \ \ ^{-1}$, it crosses over to the diffusive behavior

$$\langle L(t) \rangle \to \begin{cases} 2\, v_0\, t + O(t^2)\,; & \text{for } t \ll \ \ ^{-1} \\ \dfrac{v_0}{\ } \left[\sqrt{8\ \ t\ } - (\ \ + 2) + O\left(\dfrac{1}{\sqrt{t}}\right)\right] & \text{for } t \gg \ \ ^{-1}\ . \end{cases} \tag{47}$$

We remark that in the limit $v_0 \to \infty$, \ \ $\to \infty$, but with the ratio $v_0^2/(2\ \ ) = D_{\rm eff}$ kept fixed, it is well known that the RTP converges to a standard Brownian motion

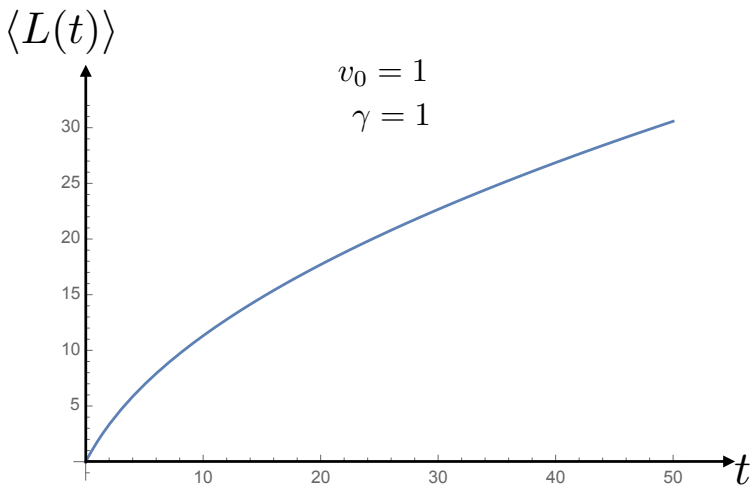

Figure 2. The mean perimeter of the convex hull of an RTP $\langle L(t)\rangle$ in Eqs. (43) and (44), plotted as a function of t (using Mathematica) for parameter values $v_0 = 1$ and $\gamma = 1$.

with an effective diffusion constant $\mathsf{D}_{\rm eff}$. In this limit, we get from the second line of Eq. (47), $\langle \mathsf{L}(\mathsf{t})\rangle \to \sqrt{16\ \mathsf{D}_{\rm eff}\,\mathsf{t}}$ and thus coincides with the well known result for the mean perimeter of the convex hull of a 2d Brownian motion of duration t and diffusion constant $\mathsf{D}_{\rm eff}$ (see e.g. the review [36]). Thus, at short times $\mathsf{t} \ll \ ^{-1}$, the behavior of $\langle \mathsf{L}(\mathsf{t})\rangle$ for the RTP is clearly different from that of a passive (ordinary) Brownian motion. However, at late times, the signature of activity is present not in the leading term (which effectively behaves as in the passive Brownian motion), but in the subleading (second term in the last line of Eq. (47)) nontrivial constant $-(\ +2)\mathsf{v}_0 =\ .$

To summarize, our principal results in this section concern the exact formulae for the mean perimeter of the convex hull of an RTP, both in the ensemble of fixed number n of runs (as given in Eq. (19) which is valid for all n) and in the fixed time t ensemble (given in Eq. (43) which is valid for all t). Later, in subsection (4.2) we compare the high precision simulation results to these exact formulae for the mean perimeter derived in this section. In Figs. 4(a) and 4(b) we compare the simulation results to the analytical formula for these two ensembles and find excellent agreement.

4. Numerically estimated mean perimeter and higher moments

Here we present the numerical part of this study. This section is split in three parts: Subsection 4.1 is a rather technical part of interest for the reader wanting to reproduce the results, which may be safely skipped when only interested in the results. Subsection 4.2 verifies the analytic results and subsection 4.3 studies the full distribution of the perimeter of the RTP's convex hull.

4.1. Methods

Modeling the RTP in simulation is straightforward. One draws random directions $_i$ uniformly from $[0; 2\ ]$ and the duration of the runs, i.e., the run times between tumble events, $_i$ from the exponential distribution

$$\mathsf{p}(\ ) = \ \mathrm{e}^{-\gamma\tau}; \tag{48}$$

where the latter task is easily achieved using the standard *inversion method* [57]. This is repeated, depending on the ensemble we want to simulate, until we either have drawn n random numbers of each or until the total time $\sum_i\ _i \geq \mathsf{t}$, in which case the last waiting time is truncated to result in an equality.

The convex hull of point sets in a $2\mathsf{d}$ plane is easy to obtain using well established methods, such as Andrew's Monotone Chain algorithm [58], which exploit the fact that a convex hull in $2\mathsf{d}$ is defined by the order of its vertices and find the hull of a point set of size n in time $\mathsf{O}(\mathsf{n}\log(\mathsf{n}))$. For our purposes a further speedup can be achieved by preprocessing the point set with Akl's heuristic [59]. The main idea behind this heuristic is that all points which are inside the convex hull of a subset, are also inside of the convex hull of the actual set and therefore not part of the convex hull. Using the points with minimal and maximal x- and y-coordinates, as well as points with extreme $\mathsf{x}+\mathsf{y}$ and $\mathsf{x}-\mathsf{y}$ as vertices for the subset hull, usually allows to discard the majority of points in time $\mathsf{O}(\mathsf{n})$ such that the exact algorithm can operate on a much smaller point set.

Given the set of n points in $2\mathsf{d}$, let $\mathsf{m} \leq \mathsf{n}$ denote the number of vertices of the associated convex hull. Given these m vertices in order by their Cartesian coordinates $\{(\mathsf{X}_i; \mathsf{Y}_i)\}$, the calculation of its perimeter is trivial

$$\mathsf{L} = \sum_{i=0}^{m-1} \sqrt{(\mathsf{X}_i - \mathsf{X}_{i+1})^2 + (\mathsf{Y}_i - \mathsf{Y}_{i+1})^2}; \tag{49}$$

with cyclical indexes, i.e., $\mathsf{X}_0 = \mathsf{X}_m$.

We employ two types of simulation to obtain our numerical results. To estimate the mean values and variances, we use *simple sampling*. That is, we generate 10^5 independent realizations of the RTP ensemble naively, construct their convex hulls and calculate the perimeter of each. These samples can be used to estimate the mean and variance. To obtain the distributions of the perimeter close to its typical values, it is sufficient to create a histogram from the collected samples. But note that the tails of this histogram will only contain events which occur with a probability, for our sample size, of around 10^{-5} or higher.

To reach the tails of the distribution containing extremely rare events with probabilities of, say, 10^{-100}, we need to employ a more sophisticated sampling method. Therefore, we use a *Monte Carlo* (MC) method based on Markov chains. Each state of the Markov chain consists of one realization of an RTP trajectory. The general idea is to use a Markov chain to generate realizations whose appearance probability is weighted with a known weight depending on the observable of interest, here the perimeter L. This weight has to exhibit a free parameter, which is used to bias the realizations towards different regions in L-space, such that good statistics can be sampled for atypical values of L. Afterwards the knowledge about the bias is used to calculate the wanted distribution $\mathsf{P}(\mathsf{L})$, from the biased realizations.

An example of a Markov chain of realizations of RTP trajectories is sketched in Fig. 3. To transition from the current state to the next state of the chain, we have

to define a *change move*. To decide for a simple yet efficient change move, we have to look at the representation of an RTP trajectory realization. Here we define it by a tuple of times between tumble events ($_1$; $_2$;:::) and a tuple of directions chosen at the tumble events ($_1$; $_2$;:::). A change move is constructed by replacing one random entry of the tuples by a new random waiting time $_i'$ from the exponential distribution $\mathsf{p}(\ )$ or by a new random direction from the corresponding uniform distribution $_i'$. Note that if this change was always accepted, it would lead to a Markov chain whose entries would be distributed uniformly over all RTP trajectory realizations.

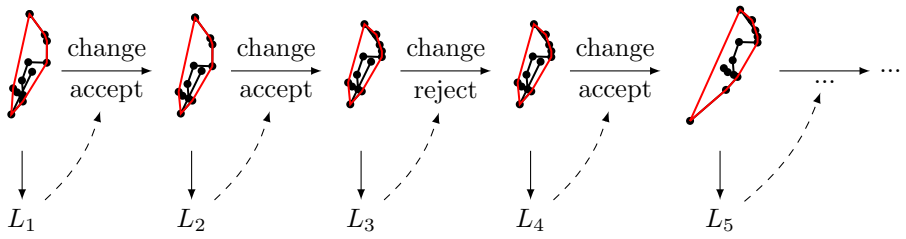

Figure 3. Sketch of the Markov chain of RTP realizations. At each step a change is proposed and either accepted, leading to a new entry in the chain or rejected, leading to a duplicate of the old entry in the chain. In this example a negative temperature leading to larger than usual hulls is used for the fixed-n ensemble with $n = 12$.

To introduce the bias, we use the classical Metropolis algorithm [60] which generates in its original formulation realizations of the canonical ensemble at a given temperature. To foster intuition, in the canonical ensemble at low temperatures one encounters realizations near the ground state, i.e., very low energies, and at high temperatures realizations with high energies. Here, we identify the observable of interest L with the energy appearing in the original Metropolis algorithm and can use the temperature Θ as a free parameter to bias the resulting distribution. Therefore we accept each proposed change with the probability $\mathsf{p}_{\text{acc}} = \min\left\{1; \mathsf{e}^{-\Delta L/\Theta}\right\}$; where $\Delta\mathsf{L}$ is the difference of the perimeter in the current and the proposed state.

If a proposed change is rejected, the current step will be repeated in the Markov chain. This way a sample of the states appearing in the Markov chain will eventually be distributed such that realizations $\mathcal{C}$ appear distributed according to $\mathsf{Q}_\Theta(\mathcal{C}) = \frac{1}{Z(\Theta)}\mathsf{Q}(\mathcal{C})\mathsf{e}^{-L(\mathcal{C})/\Theta}$, where $\mathsf{Q}(\mathcal{C})$ is the distribution of the realizations, i.e., when drawing uniformly from all realizations of RTP trajectories. The quasi-Boltzmann factor biases the samples to large or small values of L depending on the artificial temperature Θ, which can assume positive and negative values. Note that negative Θ will lead to realizations with atypically large perimeter.

Summing the distribution $\mathsf{Q}(\mathcal{C})$ over all realizations with the same perimeter results in the wanted distribution $\mathsf{P}(\mathsf{L})$, and after some elemental algebra we obtain the relation

$$\mathsf{P}(\mathsf{L}) = \mathsf{e}^{L/\Theta}\mathsf{Z}(\Theta)\mathsf{P}_\Theta(\mathsf{L}) \tag{50}$$

between the distribution we measure in the biased ensembles $\mathsf{P}_\Theta(\mathsf{L})$ and the wanted distribution $\mathsf{P}(\mathsf{L})$. The ratios of the two unknown constants $\mathsf{Z}(\Theta_i)$ and $\mathsf{Z}(\Theta_j)$ can be obtained by enforcing that the two estimates of the corresponding biased distributions $\mathsf{P}_\Theta(\mathsf{L})$ must coincide in overlapping regions. The absolute value of $\mathsf{Z}(\Theta)$ is obtained

by normalization of the entire distribution. This way, the wanted distribution $P(L)$ is obtained over a very large range.

Due to the small changes the single members of the Markov chain are correlated. Therefore, it will take some time to reach equilibrium and to forget the initial condition, which usually have a dramatically different value of the observable L than the typical realizations for the biased ensemble at the given Θ. The links in the chain until equilibration have to be discarded. Also in equilibrium subsequent samples will be correlated, which could lead to an underestimation of the statistical error. Therefore we use the integrated autocorrelation time [61] to only consider statistically uncorrelated samples for our results.

This technique was used before to study diverse problems ranging from traffic models [62], over sequence alignment [63] to the ground state energy distribution of a random energy model [64]. Even more general formulations of this algorithm exist for problems, where correct change moves are not easily defined, e.g., growth processes [52, 53, 65].

For the results obtained here (see Figs. 7(a) and 7(b)), we used about 20 temperatures per system size, and chains with a length of 10^5 sweeps, each sweep being n respectively $\lceil t \rceil$ change move proposals. Equilibration was always achieved in less than 1000 sweeps and did not pose any problems. The autocorrelation times range from less than 10 sweeps to around 100 for temperatures close to 0, such that even our very far tails consist of multiple thousand independent samples.

4.2. Mean and variance of the perimeter of the convex hull of RTP

First, we compare the analytical result for the mean perimeter for the fixed-n ensemble Eq. (18) and the fixed-t ensemble Eq. (43) to our simulations. These comparisons are shown in Fig. 4. The agreement between numerical and analytical results is excellent for all values of n, and respectively for all t. Figure 5 shows that this agreement spans over 6 orders of magnitude of our simulations.

Next, we give an outlook for the mean perimeter of the convex hulls in both ensembles, as well as the variances of the perimeter in both ensembles, in Fig. 5. Here, we will scale the results in an unusual way, to enable the visualization of different values of the parameter and a very large range of sizes in a way which enables a qualitative comparison of the different ensembles. Since we are mainly interested in the behavior at finite sizes and less in the asymptotic behavior, which should converge to the known case of Brownian motion, we will remove the asymptotic growth by showing $\langle L_n \rangle = \sqrt{n}$, respectively $\langle L(t) \rangle = \sqrt{t}$. Note that this way, for large sizes the curves will converge to a limit value and therefore compress the y-axis to allow the observation of fine details. For different values of these limit values are different, but connected by a simple relation depending on the variance of the jump-length distribution $_j$, i.e., $_j = \frac{v0}{\gamma}$ (cf. Eq. (10)) for fixed n and $_j = \frac{v0}{\sqrt{\gamma}}$ for fixed t. Therefore, we also scale with this factor which will lead to all values for the perimeter converging to the same limit value, which makes it possible to directly observe the qualitative influence of on the finite-size behavior. Also note the logarithmic x-axis to allow for a higher visual resolution at small times. Also, the variance is scaled appropriately with ^2_j and reproduces the limit values which were also measured before [46, 47], but are not known exactly even for the Brownian motion case (the only currently known exact result for variance is for the perimeter of the convex hull of Brownian bridges [66].)

The finite-size behavior of the mean perimeter shown in the left panel of Fig. 5

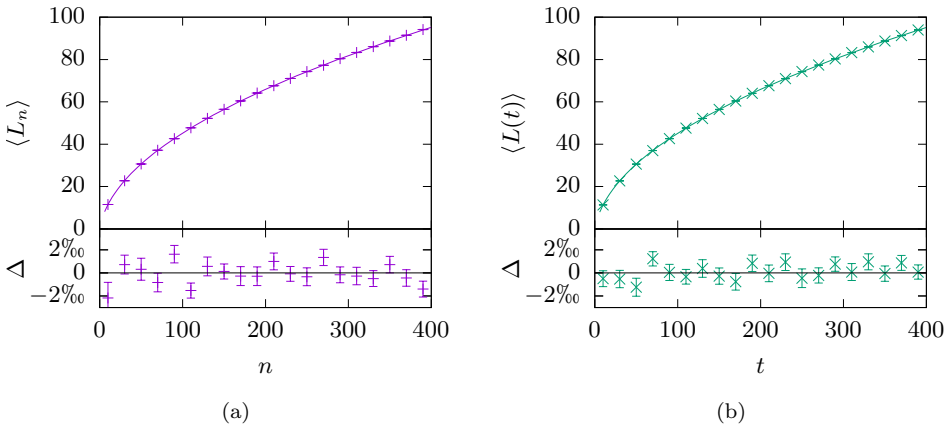

Figure 4. Left: Comparison of the exact mean perimeter $\langle L_n \rangle$ of Eq. (18) (lines) for the fixed-n ensemble with simulation results (symbols). Right: Comparison of the exact mean perimeter $\langle L(t) \rangle$ of Eq. (43) (lines) with simulation results (symbols). The errorbars are the standard errors of our simulation results and indicate excellent agreement, which is especially well visible when considering the relative difference Δ between our analytical and numerical result shown at the bottom of both plots. The shown data are for $v_0 = 1$ and $\gamma = 1$.

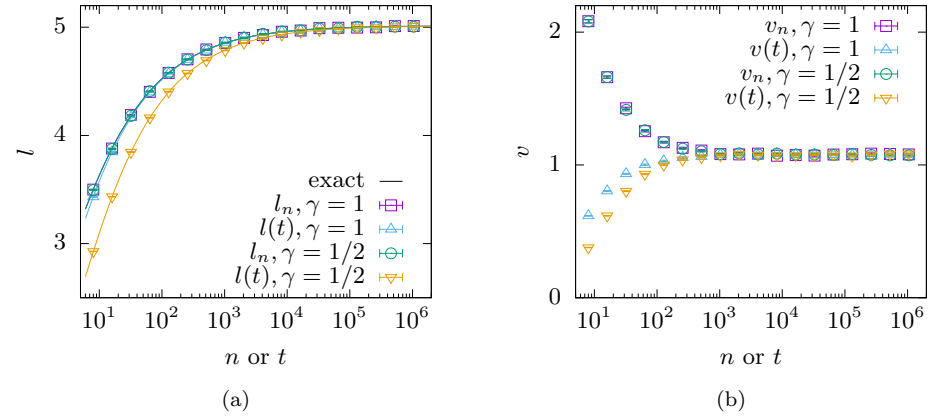

Figure 5. Left: Behavior of the mean perimeter of the convex hull. The exact values from Eqs. (18) and (43) are shown as lines. Symbols show values for the scaled mean perimeter of both ensembles $l_n = \langle L_n \rangle / (\sqrt{n}\sigma_j)$ and $l(t) = \langle L(t) \rangle / (\sqrt{t}\sigma_j)$. Right: Behavior of the variance of the perimeter of the convex hull. Symbols show values for the scaled variance of the perimeter of both ensembles $v_n = \mathrm{Var}(L_n)/(n\sigma_j^2)$ and $v(t) = \mathrm{Var}(L(t))/(t\sigma_j^2)$. Two values $\gamma = 1$ and $\gamma = 1/2$ are shown. For motivation of the unusual scaling see text. Errorbars are far smaller than the symbols.

mirrors the result obtained in the previous section for the perimeter: The behavior of the fixed-t ensemble changes with at small times, and converges to Brownian motion for long times. The fixed-n ensemble becomes indistinguishable for all values of when using this rescaling. This is consistent with the result stated in Eq. (18).

The behavior of the variance in the right panel shows the same effect. But interestingly, in contrast to the mean, where the fixed-n and fixed-t ensemble behave similarly, the variance approaches for the perimeter the asymptotic value from different directions depending on whether t is fixed (from below) or n is fixed (from above).

We also estimated the asymptotic values $\langle L_n \rangle / \sqrt{n} \to \quad_\infty$ for the means and $\mathrm{Var}(L_n)/n \to \quad^2_\infty$ for the variances and analogously for the fixed-t ensemble. All coincide within statistical errors with the current best estimates for Brownian motion, as expected. This is mainly a cross check, to establish the good quality of our data. They are obtained by fits to the form

$$L(t) = \quad_\infty + at^{-0.5} + bt^{-1}; \tag{51}$$

and analogously for fixed n, as well as for $\quad^2_\infty$. They are shown in table 1 in comparison to the current best values for them.

Table 1. Table of the asymptotic values in comparison to the known values for Brownian motion (BM). Note that the $\gamma = 1/2$ case corresponds for the fixed-n ensemble to $\sigma_j = 2$ and for the fixed-t ensemble to $\sigma_j = \sqrt{2}$. Dividing the shown values with the corresponding power of σ_j, leads to values compatible with Brownian motion. The asymptotic values are obtained by fits to Eq. (51) for $n \geq 100$ respectively $t \geq 100$ and show goodness of fit χ^2_{red} between 0.5 and 1.4.

	L	
	$\quad_\infty$	$\quad^2_\infty$
BM [38, 47],	5.0132..	1.077(1)
fixed-n, $= 1$;	5.014(1)	1.078(2)
fixed-t, $= 1$;	5.011(1)	1.076(2)
fixed-n, $= 1/2$;	10.027(3)	4.306(10)
fixed-t, $= 1/2$;	7.092(2)	2.152(5)

4.3. Full distributions

Here, we study the whole distribution of the perimeter of the convex hulls of RTP. We concentrate on two regions: First, we examine the region around the mean describing typical fluctuations, where we show the distribution for a large range of sizes. Second, we investigate the intermediate and far tail, which requires a much larger numerical effort, and is therefore restricted to small sizes.

When visualizing the distribution, we will scale it with the expected perimeter. The reasoning is similar to the unusual scaling of Fig. 5: First, this enables us to compare the distribution for values of n and t of different magnitudes. Second, it was observed previously that the distributions of standard random walks for different sizes collapse on a common curve when scaled with the mean perimeter in the limit of Brownian motion [46, 47].

Unsurprisingly the distributions for the perimeter of the RTP's convex hull to collapse in the limit of large sizes, which corresponds to Brownian motion, on the same curve when rescaled with the exact mean perimeter according to Eqs. (18) and (43). This is shown in Fig. 6. While small sizes show significant deviations from this limit shape, it is interesting that the fixed-t ensemble is apparently converging much faster.

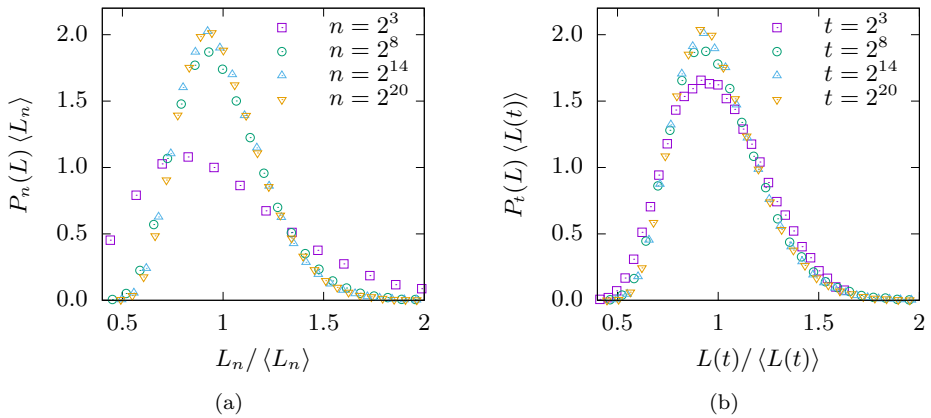

Figure 6. Distribution of the perimeter L of the RTP with $v_0 = 1$ and $\gamma = 1$. Both fixed-n and fixed-t are shown in comparison and scaled with the leading order of their mean. Note that a scaling with their exact mean does also not lead to a collapse of the small sizes (not shown).

With the large-deviation approach, we are able to sample the distribution of the perimeter over more than one hundred decades, see Figs. 7(a) and 7(b). Examining the tails of the distribution, one sees that the scaling of the distribution with its mean value works not only in the high probability region, but also in the intermediate tails, as shown in the insets. In the far tails this collapse ceases to work. Also in the far tail, we can observe a curious difference between the fixed-t and fixed-n ensemble. For fixed-n smaller numbers of tumblings lead to a larger value of the scaled probability for extremely large hulls, while for fixed-t a longer total time leads to a lower value of the scaled probability for extremely large hulls.

As expected due to previous results on convex hulls of random walks, which converge to Brownian motion in the long time limit, we test whether a *large deviation principle* holds [67]. That is, the *rate function* Φ, defined by $\mathsf{P}_n(\mathsf{L}) \approx \exp\left(-\mathsf{n}\Phi(\mathsf{L}_n)\right)$, respectively $\mathsf{P}_t(\mathsf{L}) \approx \exp\left(-\mathsf{t}\Phi(\mathsf{L}(\mathsf{t}))\right)$ (without explicit dependency on n or t) for large n respectively t, exists in the right tail. As an intuitive interpretation, the rate function determines how fast the probability density decays in the tails with increasing n respectively t. For the present distribution we can verify this directly using the tails of the distributions we obtained: The *empirical rate functions* $\Phi_n(\mathsf{L}{=}\mathsf{n})$, respectively $\Phi_t(\mathsf{L}{=}\mathsf{t})$ calculated from our data (not shown here) collapse on a power law with exponent 2 over a large range (including the full right tail) where they are independent of the system size, which is a strong hint that these forms are the actual n, respectively t independent rate functions. The exponent 2 is also compatible with previous results [46, 47].

To understand what characterizes the instances of extremely large and extremely small convex hulls, we measured during the large-deviation Markov chain MC simulations for the fixed-n ensemble the total time traveled t and for the fixed-t ensemble the number of tumble events n, for each realization encountered, respectively. To end up with a very large hull, the RTP can either make very large steps, which become exponentially rare with their length, or can take steps persistently into one direction, which also becomes exponentially improbable with the number of tumble

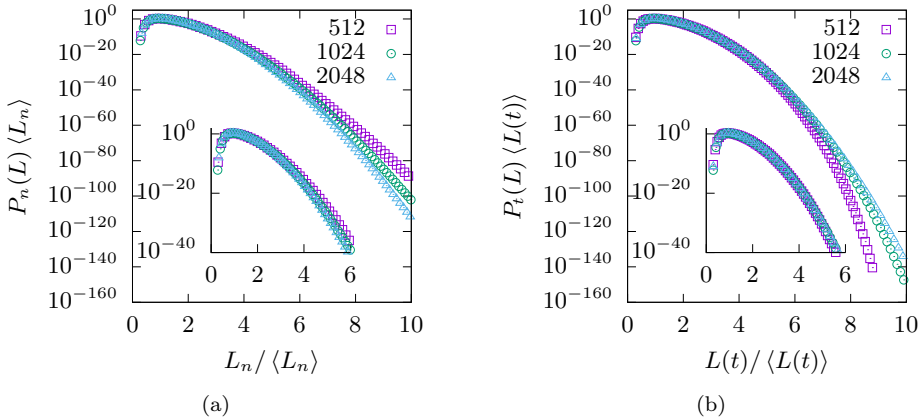

Figure 7. Probability density function of the perimeter of the convex hull around RTP including high quality statistics for extremely rare configurations with very large and very small hulls. Left for the fixed-n ensemble, right for the fixed-t ensemble.

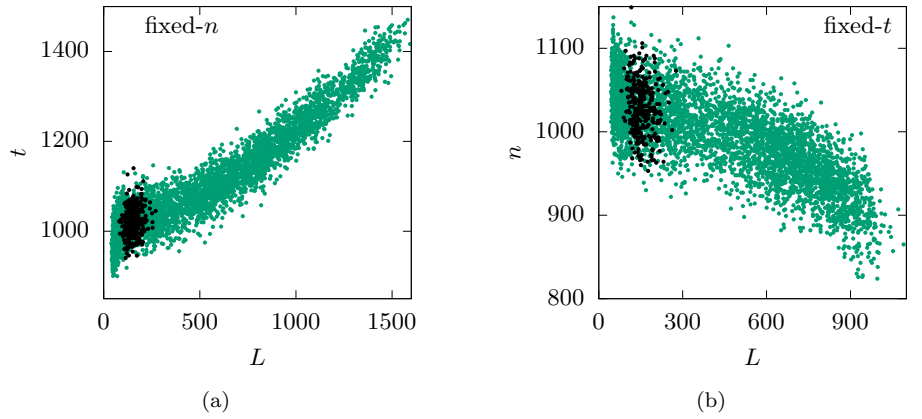

Figure 8. Left: Strong correlation of the total time t with the perimeter for the fixed-n ensemble. Right: Anticorrelation of the number of tumble events n with the perimeter for the fixed-t ensemble. Data for $\gamma = 1$ und $n = 1024$, respectively $t = 1024$. Events with a high probability (gathered using simple sampling) are marked black. Note that the density of points does not mirror the actual probability of the events.

events. If we look in Fig. 8 at the correlations between the perimeter of the hull and the number of tumble events n (for fixed-t), respectively the total time of the walk t (for fixed-n), we can identify the mechanism in both ensembles leading to extremely large hulls. In the fixed-n ensemble the perimeter of the hull is strongly correlated with the total time, i.e., the size of the hull is inflated by taking longer steps. The question is whether this correlation is sufficient to explain the large-deviation behavior. For a configuration of fixed "shape" (i.e. directions), if one blows up all traveling times by a certain factor, this would result in a growth of the perimeter by the same factor.

Now, the extreme configurations exhibit about a ten times higher perimeter, but only about a factor of 1.4 longer time. Hence, rare configurations are not only characterized by atypical large times between two tumble events, but also by rare combination of chosen directions.

Correspondingly, in the fixed-t ensemble, the number of steps is anticorrelated with the perimeter. A smaller number of tumble events, while having the same total traveling time, leads intuitively to more extended RTP trajectories, i.e., to larger perimeters. Nevertheless, since the extreme perimeter values are much larger than the typical ones and the number of tumble events is only moderately decreased, the large-deviation behavior is also for this ensemble influenced by rare combination of chosen directions.

5. Conclusion

In this paper we have studied the statistics of the perimeter of the convex hull of a single run-and-tumble particle in two dimensions. This run-and-tumble particle moves with constant speed (ballistically) during an exponentially distributed run time and changes its direction of motion at tumble events instantaneously by choosing a new direction at random. We derived exact expressions for the mean perimeter of the convex hull of its trajectory for two ensembles, with a fixed number of tumblings and with a fixed total time. Our numerical simulations in both ensembles are in excellent agreement with analytical results. For higher moments of the perimeter we could not derive exact results, but we obtained numerical results with very high precision. In particular, for both ensembles we presented detailed numerical studies of the variances as well as the full probability distributions of the perimeter of the convex hull. Deriving analytically the higher moments as well as the full distribution of the perimeter remains a challenging open problem. In particular, it would be interesting to understand in more detail what leads to the smaller variance of the perimeter of the fixed-t ensemble in comparison to the fixed-n ensemble for finite sizes.

Another closely related observable is the area of the convex hull of the trajectory. For Brownian motion and other stochastic processes, the mean area of the associated convex hull has been calculated analytically (for a review see [36]). In fact, a second formula due to Cauchy for the area of an arbitrary convex curves in $2\mathsf{d}$ was applied to convex hulls [35, 36] that allows one to compute the mean area of the convex hull of an arbitrary stochastic process in $2\mathsf{d}$. Let $\{\mathsf{X}_i; \mathsf{Y}_i\}$ (with $\mathsf{i} = 1; 2; \colon\colon\colon; \mathsf{n}$) denote the n vertices of a stochastic process in $2\mathsf{d}$ where i labels the time. Let $\mathcal{C}$ denote the convex hull of these vertices. Then the mean area of the convex hull of these vertices is given by [35, 36]

$$\langle \mathsf{A}_n \rangle = \quad \left[\langle \mathsf{M}_n^2 \rangle - \langle \mathsf{Y}_m^2 \rangle \right] : \tag{52}$$

where $\mathsf{M}_n = \max\left[\mathsf{X}_1; \mathsf{X}_2; \colon\colon\colon; \mathsf{X}_n\right]$ denotes the maximum of the x-coordinates which is achieved at step m. In formula (52) Y_m denotes the y-coordinate at step m, i.e, at the time when the x-coordinate achieves its maximum. This formula (52) is valid for any set of random points in $2\mathsf{d}$ [36], and hence it is also valid for the $2\mathsf{d}$ RTP process. However this analytic computation of the mean area is not that easy and we have not done it yet. However, we could easily obtain the mean area, as well as its higher moments and the full distribution numerically, using the same method that we used for the perimeter (not presented in this paper). Hence, it would be interesting if this mean area can be computed analytically to compare with our simulation data.

Finally, in this paper, we have considered the simplest version of the run-and-tumble model where the walker tumbles instantaneously. In more realistic models, there is an additional waiting time after each run during which the particle tumbles, before starting a new run. It would be interesting to see how the mean perimeter of the convex hull gets affected by the finite waiting time. Evidently, the result for the fixed-n ensemble will not depend on the waiting time. However, in the fixed-t ensemble, the mean perimeter will certainly depend on the waiting time, and it would be interesting to compute this.

Acknowledgments

HS and AKH thank the German Science Foundation (DFG) for financial support through the grant HA 3169/8-1. The simulations were performed at the HPC cluster CARL, located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Programme (INST 184/108-1 FUGG and INST 184/157-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State.

References

[1] Chandrasekhar S 1943 *Review of Modern Physics* **15** 1
[2] Feller W 1968 *An introduction to Probability Theory and its Applications, Vol. I and II* (New York: Wiley)
[3] Hughes B 1968 *Random walks and random environments* (Oxford: Clarendon Press)
[4] Fama E F 1965 *Financial Analysts Journal* **21** 55–59
[5] Berg H C and Brown D A 1972 *nature* **239** 500–504
[6] Page L, Brin S, Motwani R and Winograd T 1999 The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66 Stanford InfoLab
[7] Kareiva P M and Shigesada N 1983 *Oecologia* **56** 234–238 ISSN 1432-1939
[8] Bovet P and Benhamou S 1988 *Journal of Theoretical Biology* **131** 419 – 433 ISSN 0022-5193
[9] Madras N and Slade G 2013 *The Self-Avoiding Walk* (New York, NY: Springer New York) pp 281–364 ISBN 978-1-4614-6025-1
[10] Codling E A, Plank M J and Benhamou S 2008 *Journal of The Royal Society Interface* **5** 813–834
[11] Schaefer D W 1973 *Science* **180** 1293–1295 ISSN 0036-8075
[12] Marchetti M C, Joanny J F, Ramaswamy S, Liverpool T B, Prost J, Rao M and Simha R A 2013 *Review of Modern Physics* **85** 1143
[13] Bechinger C, Leonardo R D, Löwen H, Reichhardt C, Volpe G and Volpe G 2016 *Review of Modern Physics* **88** 045006
[14] Ramaswamy S 2017 *Journal of Statistical Mechanics: Theory and Experiment* 054002
[15] Alt W 1980 *Journal of Mathematical Biology* **9** 147–177 ISSN 1432-1416
[16] Berg H 2004 *E. Coli in Motion* (Heidelberg: Springer Verlag)
[17] Cates M E 2012 *Rep. Prog. Phys.* **75** 042601
[18] Trepat X, Wasserman M R, Angelini T E, Millet E, Weitz D A, Butler J P and Fredberg J J 2009 *Nature Physics* **5** 426
[19] Vicsek T, Czirók A, Ben-Jacob E, Cohen I and Shochet O 1995 *Phys. Rev. Lett.* **75** 1226
[20] Hubbard S, Babak P, Sigurdsson S T and Magnússon K G 2004 *Ecological Modelling* **174** 359
[21] Toner J, Tu Y and Ramaswamy S 2005 *Annals of Physics* **170** 359
[22] Kumar N, Soni H, Ramaswamy S and Sood A K 2014 *Nature Communications* **5** 4688
[23] Tailleur J and Cates M E 2008 *Phys. Rev. Lett.* **100**(21) 218103
[24] Cates M E and Tailleur J 2015 *Annu. Rev. Condens. Matter Phys.* **6** 219
[25] Weiss G H 2002 *Physica A* **311** 381
[26] Masoliver J and Lindenberg K 2017 *Eur. Phys. J B* **90** 107
[27] Dhar A, Kundu A, Majumdar S N, Sabhapandit S and Schehr G 2019 *Physical Review E* **99** 032132
[28] Sevilla F J, Arzola A V and Cital E P 2019 *Physical Review E* **99** 012145

[29] Mallmin E, Blythe R A and Evans M R 2019 *Journal of Statistical Mechanics: Theory and Experiment* P013204
[30] Basu U, Majumdar S N, Rosso A, Sabhapandit S and Schehr G 2019 *preprint arXiv:1910.10083*
[31] Slowman A B, Evans M R and Blythe R A 2016 *Physical Review Letters* **116** 218101
[32] Worton B J 1995 *Biometrics* **51** 1206–1215 ISSN 0006341X, 15410420
[33] Gérard Letac L T 1980 *The American Mathematical Monthly* **87** 142–142 ISSN 00029890, 19300972
[34] Hug D 2013 *Random polytopes, In Stochastic geometry, spatial statistics and random fields* (Heidelberg: Springer)
[35] Randon-Furling J, Majumdar S N and Comtet A 2009 *Phys. Rev. Lett.* **103**(14) 140602
[36] Majumdar S N, Comtet A and Randon-Furling J 2010 *Journal of Statistical Physics* **138** 955–1009 ISSN 1572-9613
[37] Reymbaut A, Majumdar S N and Rosso A 2011 *Journal of Physics A: Mathematical and Theoretical* **44** 415001
[38] Eldan R 2014 *Electron. J. Probab.* **19** no. 45, 1–34 ISSN 1083-6489
[39] Kabluchko Z and Zaporozhets D 2016 *Transactions of the American Mathematical Society* **368** 8873–8899
[40] Kac M 1954 *Duke Math. J.* **21** 501–509
[41] Spitzer F 1956 *Transactions of the American Mathematical Society* **82** 323–339
[42] Snyder T L and Steele J M 1993 *Proceedings of the American Mathematical Society* **117** 1165
[43] Kabluchko Z, Vysotsky V and Zaporozhets D 2017 *Advances in Mathematics* **320** 595
[44] Kabluchko Z, Vysotsky V and Zaporozhets D 2017 *Geometric and Functional Analysis* **27** 880
[45] Grebenkov D S, Lanoiselée Y and Majumdar S N 2017 *Journal of Statistical Mechanics: Theory and Experiment* **2017** 103203
[46] Claussen G, Hartmann A K and Majumdar S N 2015 *Phys. Rev. E* **91**(5) 052104
[47] Schawe H, Hartmann A K and Majumdar S N 2017 *Phys. Rev. E* **96**(6) 062101
[48] Kampf J, Last G and Molchanov I 2012 *Proceedings of the American Mathematical Society* **140** 2527–2535
[49] Luković M, Geisel T and Eule S 2013 *New Journal of Physics* **15** 063034
[50] Luković M, Geisel T and Eule S 2013 *New Journal of Physics* **15** 063034
[51] Dumonteil E, Majumdar S N, Rosso A and Zoia A 2013 *Proceedings of the National Academy of Sciences* **110** 4239–4244
[52] Schawe H, Hartmann A K and Majumdar S N 2018 *Phys. Rev. E* **97**(6) 062159
[53] Schawe H and Hartmann A K 2019 *Journal of Physics: Conference Series* **1290** 012029
[54] Cauchy A 1832 *Mémoire sur la rectification des courbes et la quadrature des surfaces courbées* (Paris)
[55] Chupeau M, Bénichou O and Majumdar S N 2015 *Phys. Rev. E* **91**(5) 050104
[56] Chupeau M, Bénichou O and Majumdar S N 2015 *Phys. Rev. E* **92**(2) 022145
[57] Hartmann A K 2015 *Big Practical Guide to Computer Simulations* (Singapore: World Scientific)
[58] Andrew A 1979 *Information Processing Letters* **9** 216 – 219 ISSN 0020-0190
[59] Akl S G and Toussaint G T 1978 *Information Processing Letters* **7** 219 – 222 ISSN 0020-0190
[60] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *The Journal of Chemical Physics* **21** 1087–1092
[61] Newman M and Barkema G 1999 *Monte Carlo Methods in Statistical Physics* (Clarendon Press) ISBN 9780198517979
[62] Staffeldt W and Hartmann A K 2019 *Phys. Rev. E* **100**(6) 062301
[63] Fieth P and Hartmann A K 2016 *Phys. Rev. E* **94**(2) 022127
[64] Schawe H, Hartmann A K, Majumdar S N and Schehr G 2018 *EPL (Europhysics Letters)* **124** 40005
[65] Hartmann A K 2014 *Phys. Rev. E* **89**(5) 052103
[66] Goldman A 1996 *J. Prob. Theor. Relat. Fields* **105** 57–83 ISSN 0178-8051 (print), 1432-2064 (electronic)
[67] Touchette H 2009 *Physics Reports* **478** 1 – 69 ISSN 0370-1573