

# Convex Hulls of Random Walks: Large-Deviation Properties

Gunnar Claussen\* and Alexander K. Hartmann†

*Institut für Physik, Universität Oldenburg, Carl-von-Ossietzky-Straße 9–11, 26111 Oldenburg, Germany*

Satya N. Majumdar‡

*Laboratoire de Physique Théorique et Modèles Statistiques (UMR 8626 du CNRS),  
Université de Paris-Sud, Bâtiment 100, 91405 Orsay Cedex, France.*

(Dated: April 23, 2015)

We study the convex hull of the set of points visited by a two-dimensional random walker of  $T$  discrete time steps. Two natural observables that characterize the convex hull in two dimensions are its perimeter  $L$  and area  $A$ . While the mean perimeter  $\langle L \rangle$  and the mean area  $\langle A \rangle$  have been studied before, analytically and numerically, and exact results are known for large  $T$  (Brownian motion limit), little is known about the full distributions  $P(A)$  and  $P(L)$ . In this paper, we provide numerical results for these distributions. We use a sophisticated large-deviation approach that allows us to study the distributions over a larger range of the support, where the probabilities  $P(A)$  and  $P(L)$  are as small as  $10^{-300}$ . We analyse (open) random walks as well as (closed) Brownian bridges on the two-dimensional discrete grid as well as in the two-dimensional plane. The resulting distributions exhibit, for large  $T$ , a universal scaling behavior (independent of the details of the jump distributions) as a function of  $A/T$  and  $L/\sqrt{T}$ , respectively. We are also able to obtain the rate function, describing rare events at the tails of these distributions, via a numerical extrapolation scheme and find a linear and square dependence as a function of the rescaled perimeter and the rescaled area, respectively.

PACS numbers: 02.50.-r, 75.40.Mg, 89.75.Da

Keywords: random walk, convex hull, large-deviation properties, rare events

## I. INTRODUCTION

Random walks, originally introduced in 1921 by G. Pólya [1], have since been a vital field of research. They are ubiquitous models for physical, biological and social processes [2–4]. Example applications from biology include self-propelled motion of bacteria and the diffusion of nutrients [3], as well as animal motion in general [5, 6] or during the marking of territories or description of home ranges [7–9]. For the latter case a strong increase of the amount of experimentally available data occurred after the introduction of automated radio/GPS tagging of animals [10, 11]. The usage of minimum convex polygons, called *convex hulls*, bordering the trace of an animal [7, 12] is a simple yet versatile [13] way to describe the home range and can be used for any type of (random-walk) data. In two dimensions, the convex hull of a point set is the minimum subset whose elements form a convex polygon in such a way that (a) all points of the set and (b) the connecting lines between all possible pairs lie inside the polygon. The convex hull is a suitable measure for time-discretized random walks because it can easily be attributed with geometrical quantities, i.e., hull area  $A$  and perimeter  $L$ . Also, for numerical calculations, simple and already quite fast algorithms with running times  $\mathcal{O}(N \log N)$  exist [14–16].

On the analytical side, much progress has been made for asymptotic results at long times, when a random walk (with a finite variance for step sizes) converges to the continuous-time Brownian motion (for a recent review see Ref. [17]). For a single Brownian motion of length  $T$  in two dimensions, the mean perimeter [18, 19] and the mean area [20] were computed long back.

Recently, adapting Cauchy’s formula [21] for convex curves in two dimensions to random curves, it was shown [17, 24] that the problem of computing the mean perimeter and the mean area of the convex hull of an arbitrary two dimensional stochastic process can be mapped to computing the extremal statistics of the one dimensional component of the process. This procedure was successfully applied recently to compute the mean perimeter and the mean area of several two dimensional stochastic processes such as  $N$  independent Brownian motions in 2-d [17, 24], random acceleration process in 2-d [25], 2-d branching Brownian motions with absorption with applications to epidemic outbreak [26] and 2-d anomalous diffusion processes [27]. Very recently, this method was also successfully used to compute the exact mean perimeter of the convex hull of a planar Brownian motion confined to a half-space [28]. Finally, using different methods, the mean perimeter and the mean area of the convex hull of a single Brownian motion, but in arbitrary dimensions, have been computed recently in the mathematics literature [29, 30].

The question naturally arises regarding the higher moments or even the full distribution of the perimeter and the area of the convex hulls of two dimensional random walks. Computing analytically even the second moment

---

\*Electronic address: [gunnar.claussen@uni-oldenburg.de](mailto:gunnar.claussen@uni-oldenburg.de)

†Electronic address: [alexander.hartmann@uni-oldenburg.de](mailto:alexander.hartmann@uni-oldenburg.de)

‡Electronic address: [satya.majumdar@u-psud.fr](mailto:satya.majumdar@u-psud.fr)

(and hence the variance), for the perimeter and the area of the convex hull of a single two dimensional Brownian motion, turns out to be highly difficult [22, 23]. Since so far no analytical results are available concerning the full distributions, performing numerical studies is a natural approach. Within a straightforward implementation (“simple sampling”), one generates many times independently a random walk of  $T$  steps and constructs the hull polygon and, subsequently computes its area  $A$  and perimeter  $L$ . Then one records histograms of these quantities. Nevertheless, this approach only gives insight into a small portion of the actual distributions  $P(A)$  and  $P(L)$ , i.e., configurations with  $P \propto K^{-1}$ , the inverse of the number  $K$  of samples. This demands for specific calculations with respect to the large-deviation properties of this model. So, we use a particular large-deviation approach [31, 32] which randomly alters steps of the random walk, compares the effected change in terms of either of the quantities and accepts the alteration according a Metropolis criterion involving an artificial Monte Carlo “temperature”  $\Theta$ . This is performed for various values of  $\Theta$ , giving access to different ranges of the distribution under scrutiny. Afterwards, the resulting distributions  $P_{\Theta}(A)$  or  $P_{\Theta}(L)$  can be merged according to a given scheme. In the end, we are able to obtain distributions  $P(A)$  and  $P(L)$  which cover a range in principle only limited by the numerical precision of our computers.

The rest of the paper is organized as follows: Sec. II will introduce the used random walk model, the associated convex hull and then will shortly allude to the algorithms and some heuristics. Following this, in Sec. III we will introduce the large-deviation method. Our results are presented in Sec. IV, and we will finally conclude in Sec. V, giving an outlook to possible future work.

## II. RANDOM WALKS, CONVEX HULLS AND ALGORITHMS

A time-discretized random walk consists of  $T$  step vectors  $\vec{\delta}_i$ , and the position  $\vec{x}(\tau)$  at timestep  $\tau < T$  is the sum of all steps up to  $\tau$ , i.e.:

$$\vec{x}(\tau) = \vec{x}_0 + \sum_{i=1}^{\tau} \vec{\delta}_i \quad (1)$$

The walk configuration itself is then the set  $\mathcal{W} = \{\vec{\delta}_1, \vec{\delta}_2, \dots, \vec{\delta}_T\}$  of steps [33]. The step  $\vec{\delta}_i = (\delta_{x,i}, \delta_{y,i})$  itself denotes a displacement of the particle by  $\delta_{x,i}$  in  $x$ -direction and  $\delta_{y,i}$  in  $y$ -direction. We consider two different types of steps:

1. A time-discrete approximation to a Brownian walk, i.e., both  $\delta_{x,i}$  and  $\delta_{y,i}$  are, for each  $i$ , drawn randomly from a Gaussian distribution with zero mean and variance one.
2. A walk with discrete steps corresponding to motion on a square lattice of spacing  $J$ . This

corresponds to the four possible steps  $\vec{x} \in \{(0, J), (0, -J), (-J, 0), (J, 0)\}$ , with each of these possible up/down/left/right steps having probability  $\frac{1}{4}$ .

Note that in the limit of long walks  $T \rightarrow \infty$  Gaussian walks and discrete lattice walks should statistically agree when setting  $J = \sqrt{2}$ . We also consider closed random walks which return to the origin, i.e.,  $\vec{x}(T) = \vec{x}(0)$ :

1. For the lattice case, we generate only the first half of the walk randomly. The second half is filled up with inverse steps  $\vec{\delta}_{i+\frac{T}{2}} = -\vec{\delta}_i$ . Finally the order of the steps is randomized via swapping randomly selected pairs of steps.
2. For the Gaussian case, we consider Brownian bridges [34], i.e., from the walk  $\vec{x}(\tau)$  we construct

$$\vec{x}_b(\tau) = \vec{x}(\tau) - \frac{\tau}{T} \vec{x}(T) \quad (2)$$

which fulfill  $\vec{x}_b(T) = \vec{x}(0)$

The walks generated in both cases have all necessary statistical properties of random walks, i.e., they agree with an imaginary (yet completely inefficient) approach where one generates all random walks and picks out the closed ones.

The convex hull  $\mathcal{C} = \text{conv}(\tilde{\mathcal{P}})$  of a two-dimensional point set  $\tilde{\mathcal{P}} = \{\tilde{P}_i\}, \tilde{P}_i \in \mathbb{R}^2$  is described through a convex set over  $\tilde{\mathcal{P}}$ . The points  $P$  within  $\mathcal{C}$  are given by all possible combinations  $P = \sum \alpha_i \tilde{P}_i$  with  $\tilde{P}_i \in \tilde{\mathcal{P}}$  and  $\sum_i \alpha_i = 1$  and  $\alpha_i \in \mathbb{R}_0^+$  (definition given according to [35]). This means:

1. All points  $P_i \in \mathcal{P}$  lie within  $\mathcal{C}$ .
2. All lines  $\overline{P_i P_j}; P_i, P_j \in \mathcal{P}$  also lie within  $\mathcal{C}$ .

The boundary of the convex set is a polygon which connects a subset  $\mathcal{P} \subset \tilde{\mathcal{P}}$  of  $H$  points from the point set, i.e.,  $\mathcal{P} = \{P_0, P_1, \dots, P_{H-1}\}$ , with  $P_i = (x_i, y_i)$  ( $i = 0, \dots, H-1$ ). The hull is attributed with area  $A$  and perimeter  $L$  according to (identifying  $i = H$  with  $i = 0$ ):

$$A(\mathcal{C}) = \frac{1}{2} \sum_{i=0}^{H-1} (y_i + y_{i+1})(x_i - x_{i+1}) \quad (3)$$

$$L(\mathcal{C}) = \sum_{i=0}^{H-1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \quad (4)$$

For our work, we determined the polygons bordering convex hulls (for which one uses shortly the term “convex hull”) numerically. Note that all convex hulls are represented in a computer, thus we can use dimensionless quantities subsequently. The most common convex hull algorithms operate in  $\mathcal{O}(N \log N)$ . We used Andrew’s variant [16] of the “Graham Scan” algorithm [14], which

constructs the convex hull by first drawing a dividing line through the point set and then by sorting out those points which don't form monotone chain of clockwise/counter-clockwise turns on each side of the line. In usual cases, the application of convex hull algorithms can be accelerated by usage of pre-selection heuristics, such as the one introduced by Akl and Toussaint [36]. This heuristic looks up extreme points of the set (i.e., those of maximum and minimum  $x$ - and  $y$ -coordinates) and discards all points which lie inside the quadrilateral formed by these points. We use a custom refinement of this heuristic, which is based on iterating the heuristic under rotation of the coordinate origin, which eliminates another fraction of inert points per each iteration.

### III. LARGE-DEVIATION SCHEME

For simple-sampling results, walk configurations  $\mathcal{W}$  are generated randomly, and the according convex hulls  $\mathcal{C}$  are calculated through the algorithm, resulting in a multitude values of  $A$  and  $L$ . As mentioned above, obtaining histograms of these values only gives access to the high probability regime, where the convex-hull-properties of typical random walks are measured. However, in order to obtain values of these quantities with especially low probabilities, allowing us the measure the distributions  $P(A)$  and  $P(L)$  over a large range of the support, a certain Markov-Chain Monte Carlo (MCMC) scheme can be used [31, 32].

The MCMC consist of an evolution of random walks  $\mathcal{W}(t)$ ,  $t$  being another discrete “time” parameter, not to be confused with the time parameter  $\tau$  of the random walks. For the walks, we measure the property  $S(t)$ , i.e., the area ( $S = A$ ) or perimeter ( $S = L$ ), depending on which distribution we are aiming at. The initial configuration  $\mathcal{W}(0)$  is any walk configuration, e.g., a randomly chosen one. At each Monte Carlo step  $t$ , the walk  $\mathcal{W}(t)$  is altered to  $\mathcal{W}^*$  by replacing a randomly selected step  $\vec{\delta}_i$  ( $i \in \{1, 2, \dots, T\}$ ) with a newly generated step  $\vec{\delta}'_i$ . The new step is generated according the same distribution as all random walks steps of the corresponding type, e.g., Gaussian. The convex hull of  $\mathcal{W}^*$  is calculated, leading to quantity  $S^*$ . The alteration  $\mathcal{W}^*$  is *accepted* ( $\mathcal{W}(t+1) = \mathcal{W}^*$ ) according to the Metropolis probability:

$$p_{\text{Met}} = \min \left[ 1, e^{-(S^* - S(t))/\Theta} \right]. \quad (5)$$

Here,  $\Theta$  is the (artificial) Monte Carlo “temperature”, which is just a parameter used to set the range of the sampled values. If the alteration is not accepted, it is *rejected*, i.e.,  $\mathcal{W}(t+1) = \mathcal{W}(t)$ . Now, over the course of the MC steps,  $A$  or  $L$  develops according to the effected changes and  $\Theta$ . Examples are shown in Fig. 1. As usual, the MCMC time is measured in terms of *sweeps*, i.e., number of steps per system size, which is here the walk length  $T$ .

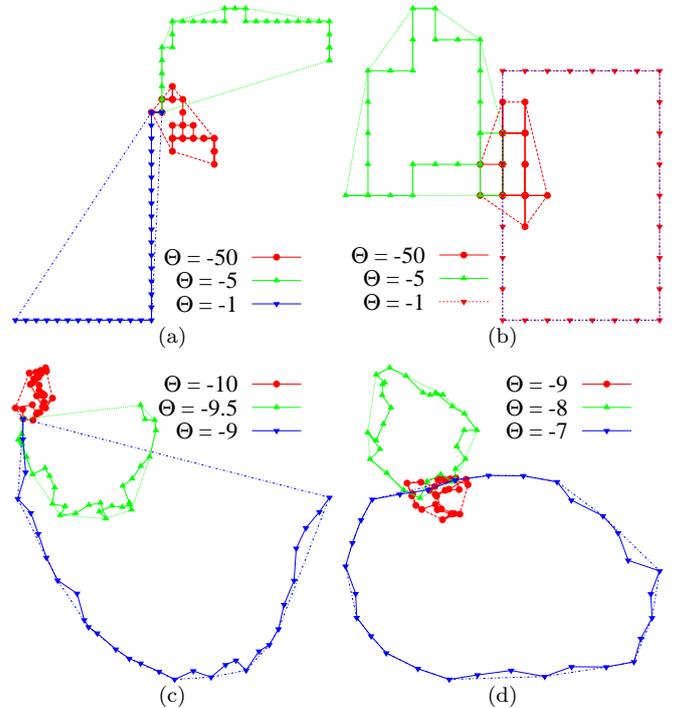


FIG. 1: (color online) Different renditions of closed walks of length  $T = 30$  after  $T_{\text{MC}} = 3 \times 10^5$  Monte Carlo steps, i.e.,  $T_s = 10^4$  sweeps, at the given temperatures. The top row features lattice steps, while the bottom row shows Gaussian steps, plus the according convex hulls (indicated by the dashed lines). The walks to the left are open and those to the right closed. Higher temperatures inhibit the acceptance of step alterations which lower the hull area  $A$ , thus the walks and their hulls experience drastic growth.

Like in any MCMC simulation one must equilibrate the simulation, i.e., discard the initial part of the measured quantities until “typical” values are found. Equilibration was found to be difficult but possible for simulations regarding hull area  $A$ . This was in particular difficult for Gaussian-distributed  $\vec{\delta}_i$  components, probably due to the fact that there exists no upper limit for  $A$  and  $L$ , in contrast to the discrete case. As demonstrated in Fig. 2  $\Theta = -16.25$  it still took several  $t_s \propto 10^5$  MC sweeps until equilibration. Opposed to this, simulations regarding perimeter  $L$  equilibrated rather quickly, typically within less than  $t_s \approx 10^3$  sweeps. This differing behavior seems to result from the fact that the replacement of one single step  $\vec{\delta}_i$  may affect much larger changes in  $A$  than in  $L$ , thereby making the whole model much more sensitive, especially for small value of  $T$ .

For a given walk type (discrete/Gaussian, open/closed), a given walk length  $T$  and a given quantity ( $A$  or  $L$ ) we performed simulations for different values of  $\Theta = \{\Theta_1, \dots, \Theta_K\}$ . The number  $K$  of temperatures and the actual values, positive and negative ones, depend heavily on the model, the walk length and the measured quantity, see below how the temperatures

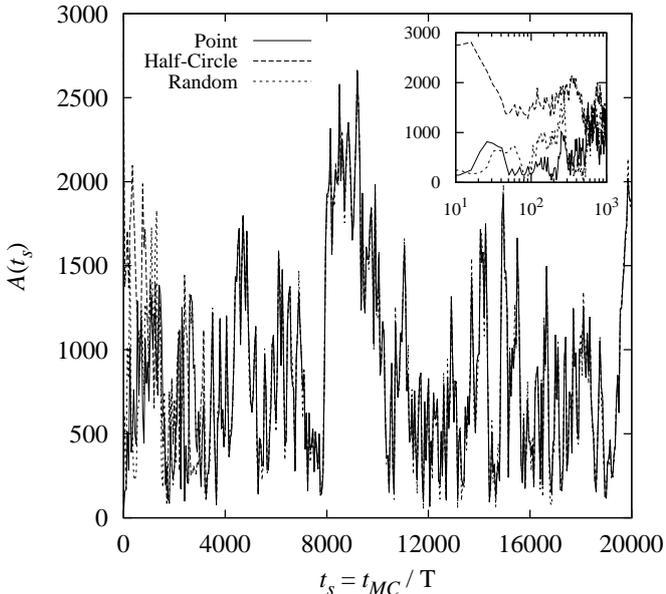


FIG. 2: Equilibration of the hull area  $A$  of open Gaussian walks over the number of Monte Carlo steps, normalized to  $t_s$  by walk length  $T = 50$ . For  $\Theta = -16.25$ , the model is fully equilibrated after ca.  $t_s = 5 \times 10^4$  sweeps, and heedless of the differing initial configurations (see inset), i.e.,  $A$  behaves similarly within the range of fluctuations.

are chosen. Thus, we got different distributions  $P_\Theta(S)$  which are related with the actual distribution  $P(S)$  according to the following relation [31]

$$P(S) = e^{S/\Theta} Z(\Theta) P_\Theta(S) \quad (6)$$

Note that “distribution” here either means “probability” or “density”, depending on whether the possible values of  $S$  are discrete or continuous. For different values of  $\Theta$ , different ranges of the measured value  $S$  were obtained. This allows for a picewise reconstruction of  $P(S)$ . It only requires knowledge of the normalization constants  $Z(\Theta)$ . They can be calculated through inversion of this formula whenever for two values  $\Theta_1$  and  $\Theta_2$  the ranges of the sampled values of  $S$  overlap. Thus, the temperatures were chosen such that for neighbouring temperatures the measured histograms sufficiently overlap. In principle, for each value of  $S$  where the measured values  $P_{\Theta_1}(S) > 0$  and  $P_{\Theta_2}(S) > 0$  one estimate of  $Z(\Theta_1)/Z(\Theta_2)$  is obtained. Note that in case of equilibration, this ratio is more or less constant for all values of  $S$  where the two histograms overlap. On the other hand, for non-equilibrated cases a systematic dependence on  $S$  is seen. In this way we have another convenient criterion to verify equilibration. Via these pairwise comparisons, all ratios  $Z(\Theta_k)/Z(\Theta_{k+1})$  can be determined. Finally, the overall determination of the normalization constants is obtained from the global normalization constraint  $\sum_S P(S) = 1$  (or  $\int P(S) dS = 1$  for the continuous case).

## IV. RESULTS

### A. Independent Points

To verify our simulations, we first simulated convex hulls of  $n$  independently distributed points in the unit square  $[0, 1]^2$ . For this case some mathematical results are known [37–39]. In particular the remaining area  $\tilde{A}_n = 1 - A(n)$  outside the convex hull is considered. The distribution of the rescaled remaining area

$$(\tilde{A}_n - 4b_n)/2c_n, \quad (7)$$

where  $b_n = \frac{2}{3} \frac{\log n}{n}$  and  $c_n = \sqrt{\frac{28}{27} \frac{\log n}{n^2}}$ , should converge for  $n \rightarrow \infty$  in distribution to the standard normal distribution  $\mathcal{N}(0, 1)$ .

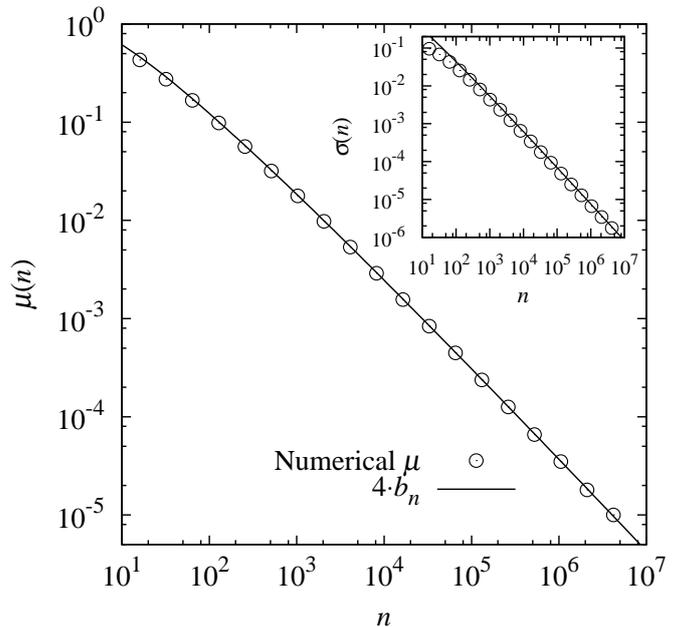


FIG. 3: Average  $\mu$  and variance  $\sigma^2$  (inset) of the remaining area  $\tilde{A}_n$  of a unit square in presence of a convex hull of  $n$  randomly distributed points. The symbols show the numerical results while the lines show  $4b_n$  and  $2c_n$ , respectively.

Thus, for the unscaled data  $A(n)$ , the mean  $\mu$  and the standard deviation  $\sigma$  should follow  $4b_n$  and  $2c_n$ , respectively. This is the case for our simulations, as visible in Fig. 3, where we indeed find a  $4b_n$  behavior for the mean and for larger sizes a  $2c_n$  behavior for the variance.

In a similar way, we attempted to visualize the predicted convergence of the distributions towards  $\mathcal{N}(0, 1)$  by plotting  $P(\tilde{A}_n)\sigma$  over  $(\tilde{A}_n - \mu)/\sigma$ , cf. Fig. 4. As visible, the measured density is very close to  $\mathcal{N}(0, 1)$  but no clear convergence is visible. Probably this would be visible only for much larger number  $n$  of points, thus the convergence is very slow. Note that these deviations become more pronounced in the small-probability tails, which we also obtained within the MCMC approach for

selected numbers  $n$  (not shown here). Anyway, for the purpose of verifying our simulations, the agreement with the predictions is sufficient.

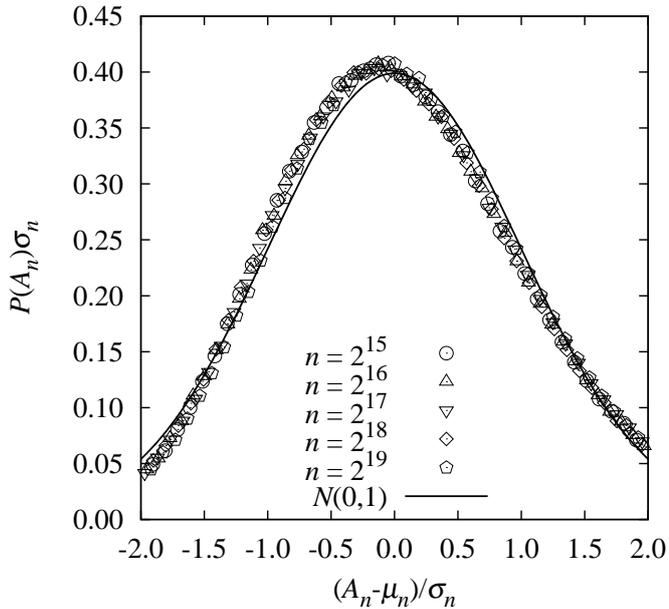


FIG. 4: As predicted by Cabo & Groeneboom [37], the distributions  $P(A_n)$  of the remaining area *not* covered by the convex hull over  $n$  random points inside the unit square should converge towards  $\mathcal{N}(0,1)$  when being rescaled according to Eq. 7.

### B. Random Walks

For each type of walk and either of the two quantities  $A$  and  $L$ , we considered walks lengths  $T \in [30, 2000]$  with ca.  $K = 10^6$  samples per temperature  $\Theta$ . Through sampling multiple values of  $A$  or  $L$  per MC run (i.e., each  $t_s = 10$  sweeps), it was possible to obtain the results for each case within several days of CPU time on a average-size multi-core cluster (using up to few dozens of cores per case, corresponding to the number of temperatures  $\Theta$ ).

An example of the resulting distributions,  $P(A)$  for the case of Gaussian open walks of various lengths, is shown in Fig. 5. Evidently, we were able to obtain the distributions over a large range of the support, with probability densities as small as  $10^{-300}$ .

As the systematic behavior of the distributions  $P(A)$  for different walk lengths  $T$  hints a common origin at a distribution  $\tilde{P}(A)$  (cf. Fig. 5), the first property we put under scrutiny is the scaling behavior. The scaling behavior [17] of the mean  $\langle A \rangle \sim T$  suggests that distributions  $P(A)$ , which in fact depends on the walk length  $T$ , scales according to the following relation:

$$P(A) = \frac{1}{T} \tilde{P} \left( \frac{A}{T} \right), \quad (8)$$

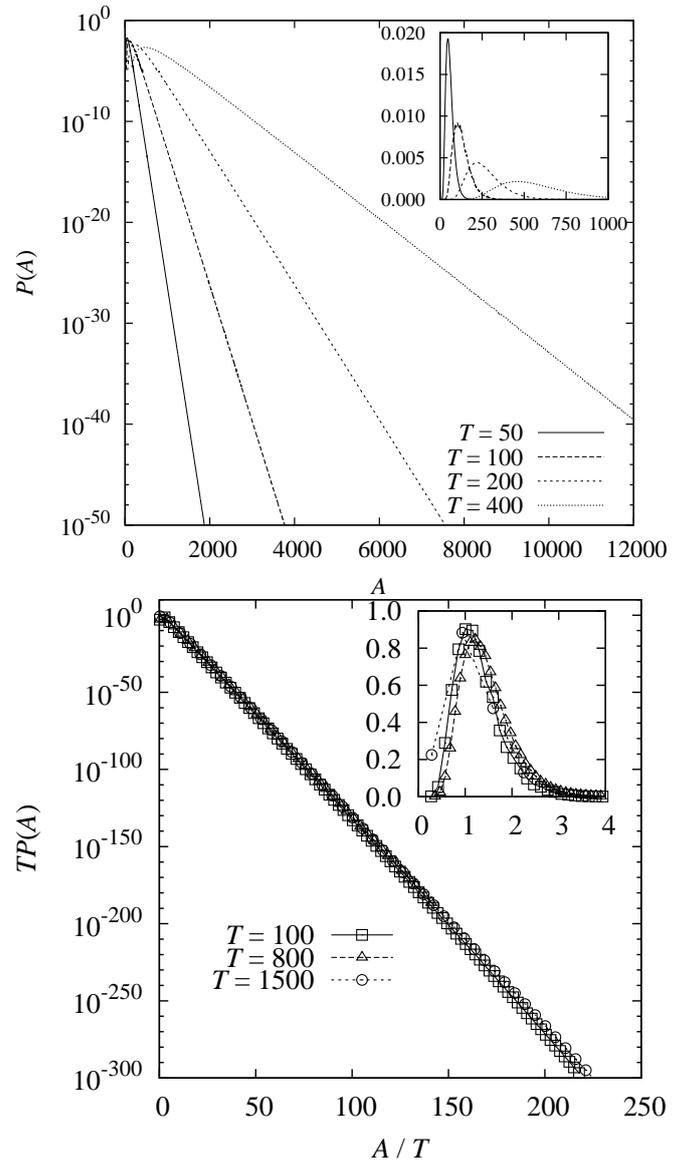


FIG. 5: Distributions  $P(A)$  of area  $A$  of the convex hull over Gaussian walks of different lengths  $T$ . (Top) “original” distribution  $P(A)$  for selected small values of the walk length  $T$ . Main: plot log scale, inset: peak region in linear scale. (Bottom) rescaling of  $P(A)$  according to Eq. (8). Main plot: log scale; inset: peak region in linear scale

where  $\tilde{P}$  is a  $T$ -independent distribution. The application of the scaling assumption leads to a reliable collapse of the distributions  $P(A)$  towards  $\tilde{P} \left( \frac{A}{T} \right)$  with the exception of finite-size effects within the tail, as shown in the bottom of Fig. 5.

For perimeter distributions, we use a similar scaling assumption, which is based upon  $\langle L \rangle \propto \sqrt{T}$  [17]:

$$P(L) = \frac{1}{\sqrt{T}} \tilde{P} \left( \frac{L}{\sqrt{T}} \right) \quad (9)$$

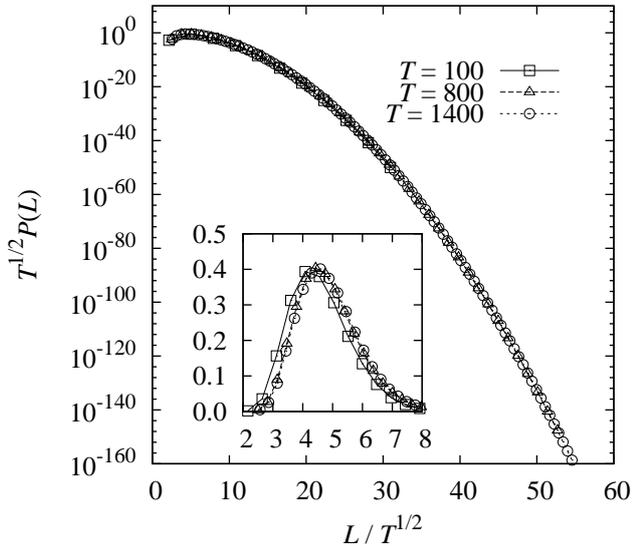


FIG. 6: Distributions  $P(L)$  of perimeter  $L$  of the convex hull over Gaussian walks of different lengths  $T$ : Rescaling of  $P(A)$  according to Eq. (9). Main plot: log scale; inset: peak region in linear scale.

The application of this relation to open walks, shown in Fig. 6, reveals an almost perfect collapse even for large values of  $T$  and the latter parts of the rare-event tail. We found similar scaling (not shown) for the discrete cases. Opposed to this, closed walks seem to be more strongly affected by finite-size effects, i.e., to tend towards  $\tilde{P}\left(\frac{L}{\sqrt{T}}\right)$  rather slowly. Nevertheless, in summary for all 8 cases the corresponding scaling is supported by our data.

Next, to extrapolate the leading behavior of the distribution in the large-deviation tail, we calculate the empirical rate function  $\Phi(s)$  [40] and guard its behavior over walk length  $T$ . Under the assumption that away from the maximum the the leading behavior of the probability is exponentially small in  $T$ , i.e., the rate function is defined as:

$$\Phi(s) \equiv -\frac{1}{T} \log P(s). \quad (10)$$

Distributions where the rate function is well defined, i.e., follow Eq. (10) are said to obey a “large-deviation principle”. To allow a comparison and extrapolation of the rate function,  $s$  is usually a quantity normalized with the maximum possible value, such that  $s \in [0, 1]$ . Thus, one would define  $s_A = \frac{A}{A_{\max}}$  and  $s_L = \frac{L}{L_{\max}}$ , accordingly. For the lattice walk model with  $J = \sqrt{2}$ , these maximum possible values were found to be  $A_{\max} = \frac{T^2}{4}$  and  $L_{\max} = 2\sqrt{2}T$  for open walks as well as  $A_{\max} = \frac{T^2}{8}$  and  $L_{\max} = \sqrt{2}T$  for closed walks. But for Gaussian walks no actual “maximum” values of  $A$  and  $L$  exist, so in remembrance of the lattice cases to obtain the same scaling, we choose  $s_A = \frac{A}{T^2}$  or  $s_L = \frac{L}{T}$ , respectively.

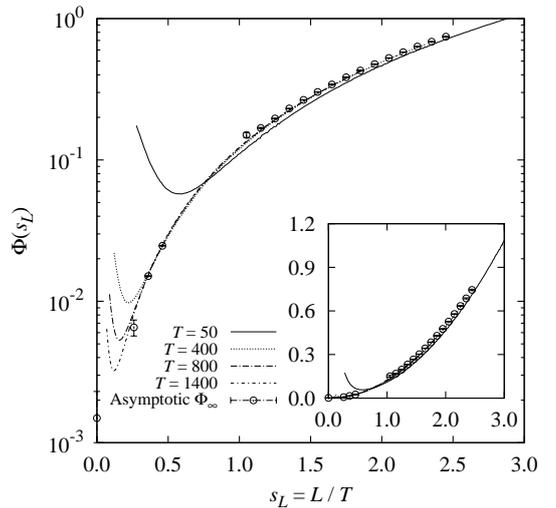


FIG. 7: The rate functions  $\Phi(s_L)$  of open Gaussian walks for different walk lengths (lines). The symbols show the extrapolated values ( $\Phi_{\infty}(s)$  see text and Fig. 8). The inset shows the same on normal scale.

An example of the resulting rate functions, for the case of the perimeter of the hull of open Gaussian walks, is shown in Fig. 7. Apparently, for large values of  $s_L$ , the rate functions for different walks lengths agree very well, while for small values of  $s_L$  strong finite length effects are visible. Thus, for small values of  $L/T$  the convergence to the limiting rate function is slower. Anyway, from the visual impression, the rate function seems to converge for  $T \rightarrow \infty$  pointwise to a limiting function, which starts at the origin. Thus, it seems that the distribution obey the large-deviation principle. To quantify this, we performed fits of the form

$$\Phi(s, T) = \Phi_{\infty}(s) - \xi T^{-\gamma} \quad (11)$$

for many values of  $s$ , as it is demonstrated in Fig. 8. We obtained similar results for all other walk types and measured quantities. We found that  $\Phi_{\infty}(s)$  appears to grow according to power laws  $s^{\kappa}$ . (The parameters  $\xi$  and  $\gamma$  are not of interest and show no systematic behavior with  $s$ ). The results are listed in Tab. I. The values for the area are very close to  $\kappa = 1$  and for the perimeter very close to  $\kappa = 2$ , in particular for the Gaussian walks.

Model	$\kappa_A, P(A)$	$\kappa_L, P(L)$
Gaussian, open	0.999(1)	2.03(2)
Gaussian, closed	1.06(2)	2.06(1)
lattice, open	1.18(1)	2.13(1)
lattice, closed	1.12(2)	2.12(1)

TABLE I: Resulting exponents  $\kappa$  which determine the power-law growth of the asymptotic value  $\Phi_{\infty}(s)$  of the rate function  $\Phi(s)$  with  $s_A$  or  $s_L$ , respectively.

This can be already seen when one combines the rate function for the scaled variables with the scaling forms

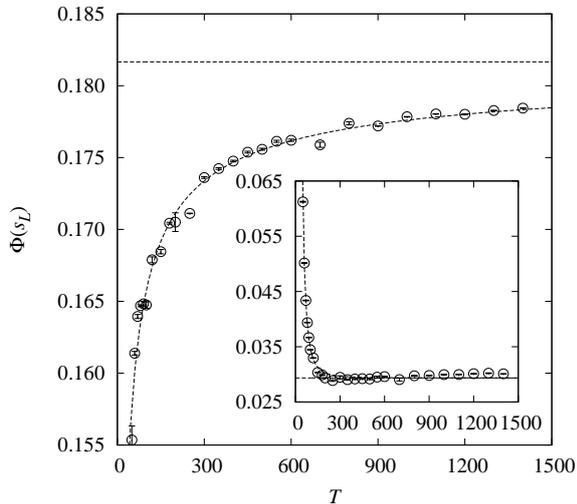


FIG. 8: Dependence of rate function  $\Phi(s_L)$  (for open Gaussian walks) on walk length  $T$  at selected values  $s = 1.2$  and  $s = 0.5$  (inset). A power law was used to determine the  $T \rightarrow \infty$  convergence towards an asymptotic value  $\Phi_\infty$ , indicated by horizontal lines.

Eqs. (8) and (9). By equating the two expressions for the distribution of the area we obtain

$$e^{-\Phi(A/T^2)T} \sim \frac{1}{T} \tilde{P}(A/T).$$

To make the argument of the exponential a function of  $A/T$  (the factor  $1/T$  is of lower order and can be ignored)  $\Phi(s) \sim s$  must hold, i.e.,  $\kappa_A = 1$ . Correspondingly for the perimeter we get

$$e^{-\Phi(L/T)T} \sim \frac{1}{\sqrt{T}} \tilde{P}(L/\sqrt{T}),$$

which results in  $\phi(s) \sim s^2$ , i.e., indeed  $\tilde{P}(L/\sqrt{T}) \sim e^{-(L^2/T^2)T} = e^{-(L/\sqrt{T})^2}$ , and  $\kappa_L = 2$ .

The fact that for Gaussian walks, the exponents  $\kappa$  agree very well with  $\kappa_A = 1$  and  $\kappa_L = 2$  but not quite for the lattice walks is probably due to the discrete structure of the lattice and due to the limited accessible area. These have a strong influence for the given limited length  $T$  of the walks and influence the rate function in particular for small and large values of  $s$ . In fact the rate functions for the lattice case are continuously bending up in a log-log plot (not shown), such that one easily can find a short interval where the expected exponent  $\kappa$  appears. Thus, to conclude this part, the data is well described by the rate function, i.e., the distribution obeys the above large-deviation principle. Hence, in connection with the fulfilled scaling forms Eqs. (8) and (9), we see that to leading order and asymptotically the distributions are given by  $P(A) \sim e^{-A/T}$  and  $P(L) \sim e^{-L^2/T}$ .

The Gaussian and the exponential tails, respectively for the distributions of the perimeter and the area, may

be guessed using very simple heuristic arguments. Indeed, the perimeter  $L$  of the convex hull of a two dimensional stochastic process morally scales as  $s$ , where  $s$  represents the span of the one dimensional component process. For the 2-d Brownian motion,  $s$  is simply the span of a one dimensional Brownian motion of duration  $T$ . The probability distribution  $P(s, T)$  of the 1-d Brownian motion of duration  $T$  and diffusion constant  $D$  is well known[4, 41] and has a scaling form

$$P(s, T) = \frac{1}{\sqrt{4DT}} f\left(\frac{s}{\sqrt{4DT}}\right) \quad (12)$$

where the scaling function  $f(x)$  is given exactly by

$$f(x) = \frac{8}{\sqrt{\pi}} \sum_{m=1}^{\infty} (-1)^{m+1} m^2 e^{-m^2 x^2}. \quad (13)$$

The scaling function has the following asymptotic behavior

$$f(x) \rightarrow \begin{cases} 2\pi^2 x^{-5} e^{-\pi^2/4x^2} & x \rightarrow 0 \\ \frac{8}{\sqrt{\pi}} e^{-x^2} & x \rightarrow \infty \end{cases} \quad (14)$$

and is plotted in Fig.9.

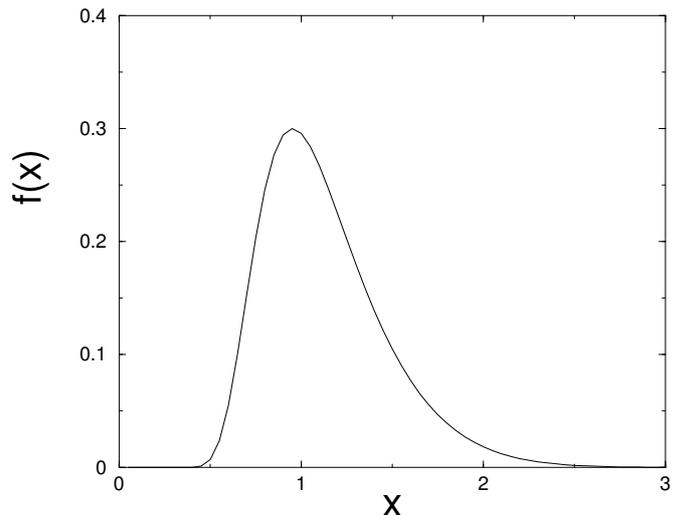


FIG. 9: The scaling function  $f(x)$  in Eq. (13) for the span distribution for a 1-d Brownian motion, plotted as a function of  $x$ .

Thus, for large  $s$ ,  $P(s, T) \sim \exp[-s^2/4DT]$  has a Gaussian tail. Consequently since  $L \sim s$ , one would expect a Gaussian tail for the distribution  $P(L)$ . Note that for small  $S$  (and hence for small  $L$ ), one would expect an essential singular behavior  $P(L) \sim \exp[-b/L^2]$  from Eq. (14), where  $a$  is some constant. The scaling function  $f(x)$  in Eq. (13), and hence  $P(L)$  is almost flat near  $L = 0$ , and falls off as a Gaussian for large  $L$  with a peak in

between, around which is has an approximate Gaussian shape (see Fig. 9)

$$P(L) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(L-\mu)^2}{2\sigma^2}} \quad (15)$$

Subsequently, for the area we can approximately assume  $A \propto L^2$ , and thus the distribution of  $\sqrt{A}$  should also be Gaussian, including a factor  $1/\sqrt{A}$  originating from  $dL/dA \sim 1/\sqrt{A}$ :

$$P(A) = \frac{1}{\sqrt{2\pi\sigma_A^2 A}} e^{-\frac{(\sqrt{A}-\mu_A)^2}{2\sigma_A^2}} \quad (16)$$

Had the assumption of a Gaussian distribution for the perimeter been entirely correct, it would be sufficient to take the average and the variance and plot the resulting Gaussian together with the data. First, we concentrate on the case of the hull perimeter for closed planar Brownian motions, where mean and variance are available analytically [23]. The mean scaled perimeter is given by [23]

$$\mu \equiv E\left(\frac{L}{\sqrt{T}}\right) = \sqrt{\frac{\pi^3 T}{2}} \frac{1}{\sqrt{T}} = \sqrt{\frac{\pi^3}{2}} \approx 3.937 \quad (17)$$

where  $E$  stands for the expectation value. For the variance, we use Goldman's result for the second moment [23],

$$E((L/\sqrt{T})^2) = \frac{\pi^2}{3} \left( \pi \int_0^\pi \frac{\sin u}{u} du - 1 \right) \quad (18)$$

The value of the integral of  $\int_0^\pi du \sin(u)/u$  is approximately 1.852, known as the Wilbraham-Gibbs constant. This then leads to the variance:

$$\begin{aligned} \sigma^2 &= E((L/\sqrt{T})^2) - E(L/\sqrt{T})^2 \\ &= \frac{\pi^3}{3}(1.852\pi - 1) - \frac{\pi^3}{2} \approx 0.348 \end{aligned}$$

Before comparing the actual distribution, we compare average and variance obtained from our numerical results. By fitting power laws plus a constant, similar to (11), we found  $\mu$  to converge asymptotically towards  $\mu_\infty = 3.937(1)$ , while  $\sigma^2$  fluctuated slightly around an average of  $\sigma_\infty^2 = 0.347(1)$ , see Fig. 10. Both values agree with the aforementioned predictions. The results of the extrapolations for all considered cases of the different walk models are listed in Tab. II, and it can be seen that the asymptotic values of the parameters of the distributions agree rather well between lattice and Gaussian walks, as expected. Note that for the fitting the results the stated error bars are only of statistical nature, thus do not include systematic contributions due to the unknown scaling behavior. Hence, the real error bars should be considerably larger. Thus the results for the averages  $\mu_\infty$  can be considered to match well the corresponding calculations in Refs.[17, 24] which read:

$$\langle A_{\text{op}} \rangle = \frac{\pi T}{2} \Leftrightarrow \mu_\infty = \frac{\pi}{2} \approx 1.571 \quad (19)$$

$$\langle A_{\text{cl}} \rangle = \frac{\pi T}{3} \Leftrightarrow \mu_\infty = \frac{\pi}{3} \approx 1.047 \quad (20)$$

$$\langle L_{\text{op}} \rangle = \sqrt{8\pi T} \Leftrightarrow \mu_\infty = \sqrt{8\pi} \approx 5.013 \quad (21)$$

Model	Gaussian open	Gaussian closed	lattice open	lattice closed
Area distributions $P(A)$				
$\mu_\infty$	1.585(1)	1.049(1)	1.577(1)	1.049(1)
$\sigma_\infty^2$	0.312(1)	0.088(1)	0.309(1)	0.089(1)
Perimeter distributions $P(L)$				
$\mu_\infty$	5.023(2)	3.937(1)	5.015(1)	3.938(1)
$\sigma_\infty^2$	1.075(1)	0.347(1)	1.075(1)	0.348(1)

TABLE II: Resulting asymptotic values of average  $\mu$  and variance  $\sigma^2$  of rescaled hull area  $A$  or perimeter  $L$  of the mentioned walk models. These values have been obtained by fitting to the results of  $\mu(T)$  and  $\sigma^2(T)$  with growing walk length, hereby extrapolating the  $T \rightarrow \infty$  case. For equal walk topologies (open/closed walks) and quantities (area/perimeter), the asymptotic values agree heedless of the actual step type, i.e., the walk geometry.

However, when actually plotting the numerical data together with a Gaussian parametrized by the analytic values, one sees that distribution only matches the numerics in the main region of the measured distributions  $P(L)$ , while the tail exhibits a different behaviour. On the other hand it is indeed possible to fit Eq. (15) to the tail of  $P(L)$  with good precision. This means the shape of the distribution is Gaussian there. Nevertheless, the resulting value of  $\sigma^2 \approx 1.35$  is considerably larger than the aforementioned, and the corresponding value of  $\mu \approx 0.1$  for these fits seems useless. In the same way, for the other cases of perimeter distributions (not shown), one can fit Gaussians well either to the main region of the data or to the tails. But it is not possible to fit the data using one single Gaussian over the full support. Thus, the approximations which lead to the assumption of a Gaussian distribution were a bit too strong and the true distributions seem to be very Gaussian-like, but slightly different.

Regarding the area distribution  $P(A)$ , the mentioned properties are demonstrated in Fig. 12. Given the previous results for the perimeter, we here just performed fits in a similar way. A Gaussian of the form of Eq. (16) can indeed be employed to fit the area distribution either in the high-probability region or everywhere away from the central region of the distribution, but again not over the full support. Thus, the same conclusion as for the perimeter distribution holds.

Although the focus of our simulations was on the right tails of the distributions, i.e., the large range of support for larger than typical values, we have also performed some simulations to study very small values of

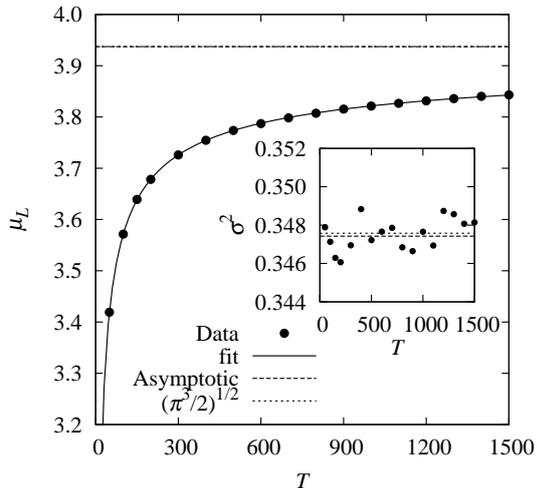


FIG. 10: Average (rescaled) perimeter  $\mu$  and variance  $\sigma^2$  for closed Gaussian as a function of the walk length  $T$ . A convergence to the analytical expectations  $\mu \approx 3.937$  and  $\sigma^2 \approx 0.348$  is visible.

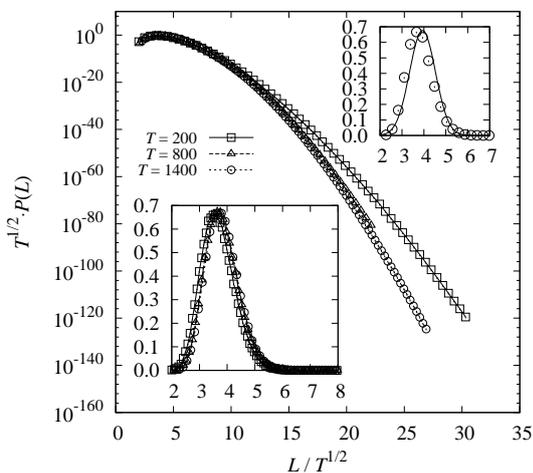


FIG. 11: Scaled distributions of the perimeter  $L$ . While the scaling assumption collapses the distributions  $\tilde{P}(L)$  onto similar behavior within the tail as well as in the main region (large inset), the Gaussian distribution according to Eq. (15) (small inset) with the values of  $\mu$  and  $\sigma^2$  taken from the analytics suits only in the main region.

the perimeter, to verify scaling function (14) leading to an expected  $\sim e^{-b/(L^2/T)}$  behavior.

The main plot of Fig. 13 shows, as example, the resulting rescaled distribution for open Gaussian walks of length  $T = 800$  together with a fit to this functional form. Apparently, the data follows the predicted form very well in the range of small perimeters. We have furthermore studied the asymptotic behavior of the factor  $b$  in the exponent and found (see inset of Fig. 13) that the behavior is compatible with a convergence to a limiting value.

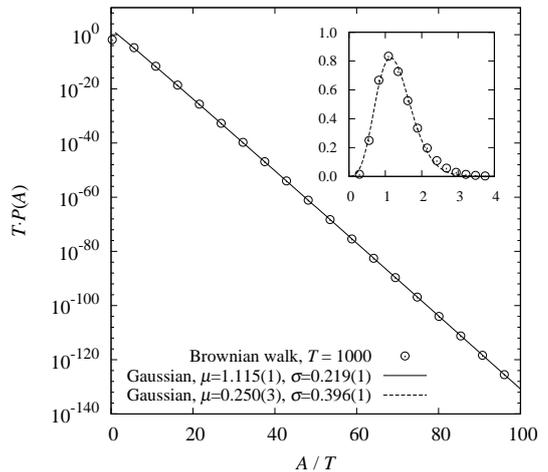


FIG. 12: Fit of a Gaussian according to Eq. (16) to  $P(A)$  with a focus on either the tail (large plot) or the region around the average (inset). It was not possible to fit one single Gaussian over the full support.

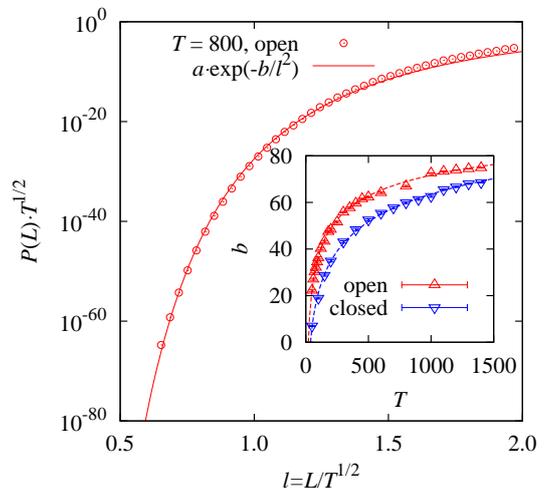


FIG. 13: (color online) Main plot: rescaled distribution (symbols) of the perimeter for the open Gaussian walks for small values of the perimeter, length  $T = 800$  together with a fit to the function  $a e^{-b/(L^2/T)}$ . Inset: Scaling behavior of the parameter  $b$  as function of walk length, for open and closed Gaussian walks, respectively. The lines represent fits to functions of the form  $b_\infty + cb^\eta$ , resulting in  $b_\infty = 129(9)$  (open) and  $b_\infty = 250(50)$  (closed), with a very slow convergence ( $\eta = -0.21(3)$  and  $\eta = -0.09(2)$ ).

## V. CONCLUSIONS

We succeeded with the application of the above explained large-deviation scheme to the study of properties of convex hulls of random walk models. We could obtain the corresponding distributions over broad ranges of convex hull area  $A$  and perimeter  $L$ , down to densities and probabilities as small as  $10^{-300}$ .

The scaling behavior of these distributions turned out to be linked to walk length  $T$  in the same way as the

averages  $\langle A \rangle \sim T$  and  $\langle L \rangle \sim \sqrt{T}$ . Also, the examination of the empirical rate functions resulted in simple power laws, which are compatible with the scaling behavior. This leads to the natural assumption that the perimeter  $L$  and the square root  $\sqrt{A}$  of the area are Gaussian distributed. Indeed either peak region of the data or the tails region fit this form very well, but it is not possible to fit over the full range of the support one single function. Nevertheless, for almost all our results we found asymptotic agreement between the Gaussian and the lattice walk case, as expected. For further studies, it could be interesting to obtain analytical estimates for the distributions for finite walk lengths  $T$  and measure the speed of convergence towards the limiting distribution.

Since the application of large-deviation simulation approaches for this problem turned out to be very useful, it should be applied for similar problems. For future work, besides obvious extensions like considering higher dimensions, we want to consider other random walk models,

particularly multiple interacting walkers [42, 43], in simple self-avoiding walks as well as in loop-erased random walks[44, 45] and, most importantly, for applications to biology like the formation of animal territories[9].

### Acknowledgments

We acknowledge personal support by O. Melchert and M. Manssen. Our simulations were carried out on the HERO cluster of Carl-von-Ossietzky Universität Oldenburg, which has been funded by the DFG through its Major Research Instrumentation Programme (INST 184/108-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State. SNM acknowledges support by ANR grant 2011-BS04-013-01 WALKMAT and in part by the Indo-French Centre for the Promotion of Advanced Research under Project 4604-3.

- 
- [1] G. Polya, *Math. Ann.* **84**, 149 (1921).
- [2] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1992).
- [3] H. C. Berg, *Random Walks in Biology* (Princeton University Press, Princeton, 1993).
- [4] B. D. Hughes, *Random walks and random environments* (Oxford University Press, Oxford, 1996).
- [5] P. Bovet and S. Benhamou, *J. Theor. Biol.* **131**, 419 (1988).
- [6] F. Bartumeus, M.G.E. Da Luz, G.M. Viswanathan and J. Catalan, *Ecology* **86**, 3078 (2005).
- [7] C. O. Mohr, *American Midland Naturalist* **37**, 223 (1947).
- [8] B. J. Worton, *Biometrics* **51**, 1206 (1995).
- [9] L. Giuggioli, J. R. Potts, and S. Harris, *PLoS Comput. Biol.* **7**, e1002008 (2011).
- [10] R. E. Kenward, *Wildlife Radio Tagging* (Academic Press, London, 1987).
- [11] E. Landguth and F. Huettmann, *Spatial Complexity, Informatics, and Wildlife Conservation* (Springer, Heidelberg, 2010).
- [12] B. J. Worton, *Ecol. Model.* **38**, 277 (1987).
- [13] S. A. Boyle, W. C. Lourenco, L. R. da Silva, and A. T. Smith, *Folia Primatol.* **80**, 33 (2009).
- [14] R. L. Graham, *Inform. Process. Lett.* **1**, 132 (1972).
- [15] A. Bykat, *Inform. Process. Lett.* **7**, 296 (1978).
- [16] A. M. Andrew, *Inform. Process. Lett.* **9**, 216 (1979).
- [17] S. N. Majumdar, A. Comtet, and J. Randon-Furling, *J. Stat. Phys.* **138**, 955 (2010).
- [18] L. Takács, *Amer. Math. Month.* **87**, 142 (1980).
- [19] G. Letac, *J. Theor. Prob.* **6**, 385 (1993).
- [20] M. El Bachir, *L'enveloppe convexe du mouvement brownien* (Ph.D thesis, Université Paul Sabatier, Toulouse, France, 1983).
- [21] A. Cauchy, *La rectification des courbes* (Mémoire de l'Académie des Sciences, Paris, 1832).
- [22] T. L. Snyder, and J. M. Steele, *Proc. Amer. Math. Soc.* **117**, 1165 (1993).
- [23] A. Goldman, *Probab. Theor. Relat. Fields* **105**, 57 (1996).
- [24] J. Randon-Furling, S. N. Majumdar, and A. Comtet, *Phys. Rev. Lett.* **103**, 140602 (2009).
- [25] A. Reymbaut, S. N. Majumdar, and A. Rosso, *J. Phys. A: Math. Theor.* **44**, 415001 (2011).
- [26] E. Dumonteil, S. N. Majumdar, A. Rosso, and A. Zoia, *Proc. Nat. Acad. Sci. USA*, **110**, 4239 (2013).
- [27] M. Luković, T. Geisel, and S. Eule, *New. J. Phys.* **15**, 063034 (2013).
- [28] M. Chupeau, O. Benichou, and S. N. Majumdar, arXiv:1412.2649
- [29] R. Eldan, arXiv:1211.2443
- [30] Z. Kabluchko, and D. Zaporozhets, arXiv:1404.6113
- [31] A. K. Hartmann, *Phys. Rev. E* **65**, 056102 (2002).
- [32] A. K. Hartmann, *Eur. Phys. J. B* **84**, 627 (2011).
- [33] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume I* (John Wiley & Sons, New York, 1950).
- [34] R. Mansuy and M. Yor, *Aspects of Brownian Motion* (Springer, Berlin, 2008).
- [35] F. P. Preparata, and M. I. Shamos, *Computational Geometry – An Introduction* (Springer, Berlin, 1985).
- [36] S. G. Akl, and G. T. Toussaint, *Inform. Process. Lett.* **7**, 219 (1978).
- [37] T. Cabo, and P. Groeneboom, *Probab. Theor. Relat. Fields* **100**, 31 (1994).
- [38] C. Buchta, *Discrete Comput. Geom.* **33**, 125 (2005).
- [39] P. Groeneboom, *Adv. Appl. Prob.* **44**, 330 (2012).
- [40] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [41] A. Kundu, S. N. Majumdar, and G. Schehr, *Phys. Rev. Lett.* **110**, 220602 (2013).
- [42] L. Acedo and S. B. Yuste, *Rec. Res. Develop. in Stat. Phys.* **2**, 83 (2002).
- [43] A. Kundu, S. N. Majumdar, and G. Schehr, *J. Stat. Phys.* **157**, 124 (2014).
- [44] G. F. Lawler, *J. Phys. A: Math. Gen.* **20**, 4565 (1987).
- [45] S. N. Majumdar, *Phys. Rev. Lett.* **68**, 2329 (1992).