Convex hulls of random walks in higher dimensions: A large-deviation study

Hendrik Schawe* and Alexander K. Hartmann†
Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany and LPTMS, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France

Satya N. Majumdar‡
LPTMS, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France
(Received 11 September 2017; revised manuscript received 14 November 2017; published 1 December 2017)

The distribution of the hypervolume V and surface ∂V of convex hulls of (multiple) random walks in higher dimensions are determined numerically, especially containing probabilities far smaller than $P = 10^{-1000}$ to estimate large deviation properties. For arbitrary dimensions and large walk lengths T, we suggest a scaling behavior of the distribution with the length of the walk T similar to the two-dimensional case and behavior of the distributions in the tails. We underpin both with numerical data in $d = 3$ and $d = 4$ dimensions. Further, we confirm the analytically known means of those distributions and calculate their variances for large T.

DOI: 10.1103/PhysRevE.96.062101

I. INTRODUCTION

The random walk (RW) is first mentioned [1] with this name in 1905 by Pearson [2] as a model, where at discrete times, steps of a fixed length are taken by a single walker in a random direction, e.g., with a random angle on a plane in two dimensions. This was later generalized to random flights in three dimensions [3] and RWs on a lattice in d dimensions [4]. A few decades later even more generalized models appeared, e.g., introducing correlation [5–7] or interaction with its past trajectory [8–10], its environment [11–15], or other walkers [16,17]. Despite the plethora of models developed for different applications, still simple isotropic RWs are used as an easy model for Brownian motion and diffusion processes [11,15,18], motion of bacteria [19,20], financial economics [21], detecting community structures in (social) networks [22,23], epidemics [24], polymers in solution [25–27], and home ranges of animals [28,29].

The most important quantity that characterizes RWs is the end-to-end distance and how it scales with the number of steps, giving rise to an exponent ν, i.e., the inverse fractal dimension. To describe the nature of different RW models more thoroughly, other quantities can be used. Here, we are interested in analyzing the "volume" and the "surface" of the RW, which can be conveniently defined by the corresponding quantities of the convex hulls of each given RW. These quantities are used, usually in two dimensions, to describe home ranges of animals [30,31]. But also, very recently, to detect different phases in intermittent stochastic trajectories, like the run and tumble phases in the movement of bacteria [32]. The convex hull of a RW is the smallest convex polytope containing the whole trace of the RW, i.e., it is a nonlocal characteristic that depends on the full history of the walker, namely all visited points.

The most natural statistical observables associated to the convex hull of a random trajectory are its (hyper-) volume and its (hyper-) surface. The full statistics of these two random variables are nontrivial to compute even for a single Brownian motion in two or higher dimensions. Even less is known on the statistics of these two random variables for a discrete-time random walk with a symmetric and continuous jump distributions. In fact, most publications concentrate on the area and perimeter of convex hulls for two-dimensional RWs. The mean perimeter and the mean area of a single random walk in a plane, as a function of the number of steps (in the limit of large number of steps with finite variance of step lengths where it converges to a Brownian motion), are known exactly since more than 20 years [33,34]. These results for the convex hull of a single Brownian motion in a plane have recently been generalized in several directions in a number of studies. These include the exact results for the mean perimeter and mean area of the convex hull for multiple independent Brownian motions and Brownian bridges in a plane [35,36], for the mean perimeter of the convex hull of a single Brownian motion confined to a half plane [37], and for the mean volume and surface of the convex polytopes in arbitrary dimensions d for a single Brownian motion and Brownian bridge [38–40]. Much less is known for discrete-time random walks with arbitrary jump length distributions. Very recently the mean perimeter of the convex hull for planar walks for finite (but large) walk lengths and arbitrary jump distributions were computed explicitly [41]. For the special case of Gaussian jump lengths, an exact combinatorial formula for the mean volume of the convex hull in d-dimensions was recently derived [39]. In $d = 2$, the asymptotic (for large number of steps) behavior of the mean area for Gaussian jump lengths was derived independently in Ref. [41]. Also the convex hulls of other stochastic processes like Lévy flights [42,43], random acceleration processes [44], or branching Brownian motion with absorption [24] were under scrutiny recently.

Analytical calculations of the variance or higher moments turned out to be much more difficult [45,46]. In absence of any analytical result for the full distribution of the volume and surface of the convex hull of a random walk, a sophisticated large-deviation algorithm was recently used to compute numerically the full distribution of the perimeter and the area of the convex hull of a single [47] and multiple [48]

*hendrik.schawe@uni-oldenburg.de
†a.hartmann@uni-oldenburg.de
‡satya.majumdar@u-psud.fr

random walks in two dimensions. Amazingly, this numerical technique was able to resolve the probability distribution down to probabilities as small as, e.g., 10^{-300} [47,48]. In this work, we will use simulations to obtain the distribution of the volume V and surface ∂V of the convex hull of a single random walk with Gaussian jump length distribution in dimensions $d \in \{3,4\}$ over a large range of its support. In particular, this range is large enough to include large deviations, here down to probability densities far smaller than $P(V) = 10^{-1000}$. While previous work [47,48] suggested that the area and perimeter distribution obeys the large deviation principle in $d = 2$, which was later proven for the perimeter [49], our results suggest that the same holds for higher dimensions. Regarding the scaling behavior of the mean and of the variance, we also study higher dimensions up to $d = 6$. Also we generalize scaling arguments to higher dimensions which were previously used to estimate the properties of these distributions for $d = 2$ [47].

The remainder of the paper is organized as follows. We will first introduce the RW model, give an overview for the calculation of convex hulls in higher dimensions, and describe the sampling technique used to reach the regions of sufficiently small probabilities in Sec. II. The presentation of our results is split into two parts. Section III A compares our numerically obtained means with the analytically derived values from Refs. [38,39] to check that our results are consistent with the literature. Also values for the variances for single and multiple RWs are presented. The behavior of the distributions, especially in their tails, is presented in Sec. III B. Section IV concludes and gives a small outlook to still open questions.

II. MODELS AND METHODS

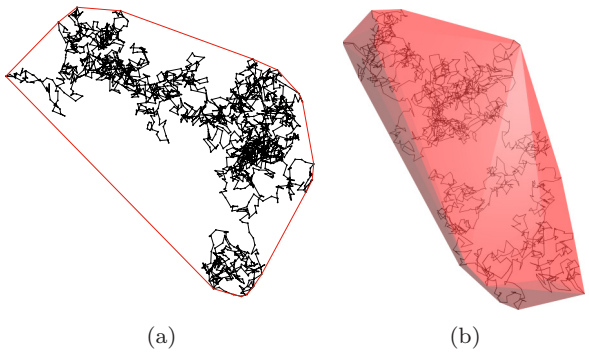
A. Random walks

A *random walk* [2,4] in d dimensions consists of T step vectors $\boldsymbol{\delta}_i$ such that its position at time τ is given as

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

where $\boldsymbol{x}_0$ is the starting position and chosen in the following always as the origin of the coordinate system. Thus, a realization of a walk can be characterized as a tuple of the displacements $(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_T)$. We will denote the set of visited points as $\mathcal{P} = \{\boldsymbol{x}(0), \ldots, \boldsymbol{x}(T)\}$. We draw the steps $\boldsymbol{\delta}_i$ from an uncorrelated multivariate Gaussian distribution with zero mean and unit width $G(0,1)$, i.e., d independent random numbers per step. Two examples for dimensions $d = 2$ and $d = 3$ are visualized in Fig. 1. While walks on a lattice show finite-size effects of the lattice structure [47], especially in the region of low probabilities, the Gaussian displacements lead to smooth distributions. Note that in the limit $T \to \infty$ Gaussian and lattice RWs do not behave differently. Both converge to the continuous-time Brownian motion [36].

The RW is very well investigated [1], especially it is known that the end-to-end distance r, and in fact every one-dimensional observable, scales as $r \propto T^{\nu}$ with $\nu = 1/2$. This exponent ν is the same in any dimension and characteristic for diffusion processes.



(a) (b)

FIG. 1. Examples for Gaussian random walks in $d = 2$ and $d = 3$. Their convex hull is visualized in red. (a) $d = 2$, $T = 2048$ and (b) $d = 3$, $T = 2048$.

B. Convex hulls

For a given point set $\mathcal{P}$ its *convex hull* $\mathcal{C} = \mathrm{conv}(\mathcal{P})$ is the smallest convex polytope enclosing all points $P_i \in \mathcal{P}$, i.e., all points P_i lie inside the polytope and all straight line segments (P_i, P_j) lie inside the polytope. In Fig. 1 two examples for $d = 2$ and $d = 3$ are shown.

Convex hulls are a well-studied problem with applications from pattern recognition [50] to ecology studies [51]. They are especially important in the context of computational geometry, where next to a wide range of direct applications [52,53] the construction of Voronoi diagrams and Delaunay triangulations [54] stand out, which in turn are useful in a wide range of disciplines [55]. Note that a lower bound for the worst-case time complexity of an exact convex hull algorithm for $T = |\mathcal{P}|$ points is $\Omega(T^{\lfloor d/2 \rfloor})$ [56–58], which is the order of possible facets, i.e., exponential in the dimension. Although, there are approximate algorithms [59,60] that probably would make the examination of higher-dimensional convex hulls feasible, we are only examining the convex hulls up to $d = 6$ using exact algorithms.

We measure the *(hyper-) volume* V, e.g., in $d = 3$ the volume, and the *(hyper-) surface* ∂V, e.g., in $d = 3$ the surface area. Determining surface and volume of a high-dimensional convex polytope is trivial given its facets f_i, which are $(d-1)$-dimensional simplexes. Choosing an arbitrary fixed point p inside the convex polygon, one can create a d-dimensional simplex from each facet f_i, such that their union fills the entire convex hull (cf. Fig. 2(a) for a $d = 2$ example). Therefore, the volume can be obtained by calculating

$$V = \sum_i \mathrm{dist}(f_i, p) a_i / d,$$

where $\mathrm{dist}(f, p)$ is the perpendicular distance from the facet f_i to the point p and a_i is the surface of the facet. The surface of a $(d-1)$-dimensional facet is its $(d-2)$-dimensional volume, which can be calculated with the same method recursively, until the trivial case of one dimensional facets, i.e., lines. Determining the surface uses the same recursion, by calculating $\partial V = \sum_i a_i$.

To foster intuition, this method is pictured for $d = 2$ in Fig. 2(a). Here, the facets are lines and the volume of the simplex is the area of the triangle. The perpendicular distances are visualized as dashed lines.

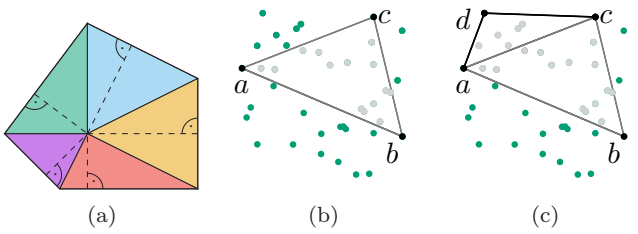


FIG. 2. Visualization of (a) the idea to calculate the volume of a convex polygon given its facets and an interior point, perpendicular distances are shown with dashed lines. In (b) and (c) examples of two consecutive recursive steps of the quickhull algorithm are shown. The point d is left of and farthest away from (a,c). Parts of the convex hull are black, discarded points are light gray.

In the scope of this study, we use the *quickhull* algorithm [61–63], and its excellent implementation in the *Qhull* library [64]. Quickhull is a divide-and-conquer algorithm applicable in arbitrary dimensions. For clarity, the algorithm will be explained for $d = 2$, since it makes the central idea clear. The technical details and the generalization to higher dimensions are well explained in Ref. [64].

Start with two points a,b on the convex hull, e.g., the points with minimum and maximum x-coordinate. Determine the point c left (when "looking" $a \to b$) of and farthest away from the edge (a,b) and discard all points inside the polygon (a,b,c). Repeat this step recursively with the edges (a,c) and (c,b) until there are no points on the left side of the current edge. All edges created in this way on the bottom level of the recursion are part of the convex hull. Two steps of this recursion are pictured in Figs. 2(b) and 2(c). The same process is repeated recursively with the point c' left of and farthest away from the inverse edge (b,a).

C. Sampling

We performed Markov chain Monte Carlo simulations to examine the distributions of the volume V and the surface ∂V of the convex hull of RWs in dimensions $d \in \{3,4\}$. To collect *large-deviation* statistics, i.e., obtain not only the peak, but also the tails of the distribution, we use both the classic *Wang-Landau* (WL) sampling [65,66] and a modified Wang-Landau sampling [67–69] with a subsequent entropic sampling [70,71] run. In contrast to similar studies [47,48] no temperature-based sampling scheme was used, since the difficulties to find suitable temperatures and regarding equilibration mentioned in Ref. [47] are even worse in higher dimensions.

Both sampling techniques generate Markov chains of *configurations*, where here configurations are realizations, each given by the tuple of RW displacements $(\boldsymbol{\delta}_1,..,\boldsymbol{\delta}_T)$. One only needs a function yielding an "energy" of a configuration and a way to change a configuration to a similar configuration. As energy we simply use the observable of interest S, i.e., either the volume V or the surface ∂V. To change a configuration, we replace a randomly chosen step $\boldsymbol{\delta}_i$ of the RW with a new randomly drawn step. Because all points $\boldsymbol{x}(\tau)$ for $\tau \geqslant i$ change, this is a global change of the walk. Though, this does not lead to a severe computational overhead, because after the update the convex hull has to be calculated again from scratch in any case.

For both WL versions at first a lower and upper bound of the observable S needs to be defined and the range in between is subdivided in overlapping *windows*, depending on system size T. While the windows can introduce errors in the results, which can lead to neighboring windows that overlap does not match, this phenomenon was not observed in this study. Also, small systems in low dimensions were sampled using the temperature-based method from Ref. [47], which does not use windows, and showed no noteworthy deviations from either of the WL variations. Also, for the present work it was sufficient to sample each window independently in parallel. Therefore, it was not necessary to apply a replica-exchange enhancement [72].

In the beginning, we start with an arbitrary configuration c_i of the walk. Afterwards we repeatedly propose random changes each leading to a new configuration c_{i+1} and accept each with the Metropolis acceptance probability,

$$p_{\text{acc}}[S(c_i) \to S(c_{i+1})] = \min\left(\frac{g[S(c_i)]}{g[S(c_{i+1})]}, 1\right), \qquad (1)$$

where g is an estimate for the density of states—basically the wanted distribution. If g equals the true density of states this will result in every S being visited with the same probability, i.e., a flat histogram of S. Since we do not know the true density of states in advance, WL iteratively improves the estimate g. Therefore, every time a value of S is visited, $g(S)$ is increased. The original article suggests to multiply $g(S)$ with a fixed factor f to perform the increase, i.e., $g(S) \mapsto g(S)f$, and after an auxiliary histogram fulfills some flatness criterion reduce this factor $f \mapsto \sqrt{f}$. This is repeated until f falls below some beforehand defined threshold f_{final} (here, $f_{\text{final}} = 10^{-8}$). Since the acceptance ratio changes during the simulation, detailed balance does not hold, such that systematic errors are introduced. To mitigate this, a better schedule to modify g is introduced in [68], which reduces the systematic errors. Basically, the flatness criterion is removed and the factor by which to increase $g(S)$ when visiting S is a function of the *Monte Carlo time* t of the simulation, i.e., $\ln[g(S)] \mapsto \ln[g(S)] + t^{-1}$. The sampling terminates as soon as $t^{-1} \leqslant f_{\text{final}}$ (here, $f_{\text{final}} = 10^{-5}$). This has the added benefit that the simulation time does not depend on some flatness criterion, which is hard to predict, but is at most $1/f_{\text{final}}$ Monte Carlo sweeps.

To remove the systematic error completely, one can use entropic sampling [70,71], i.e., fix the so far obtained estimate g and sample the system using the same acceptance as before from Eq. (1). The entropic sampling pass, which due to the fixed g obeys detailed balance and is thus not subject to the systematic error of the WL sampling, will calculate corrections for the initial estimate g. If the estimate g is close enough to the density of states, it should be able to mitigate the (small) systematic error. Finally, one creates a histogram H of the visited S to arrive at a corrected $\widetilde{g}(S) = g(S)H(S)/\langle H\rangle$ [71], where $\langle H\rangle$ is the average number of counts of the histogram.

During this simulation, the value S of the configuration may not leave its window, thus changes to configurations outside of the window are rejected. This also means that the first configuration must be within the window and is

therefore obtained via a greedy heuristic. The final distribution is obtained as follows: For mutually overlapping windows, the corresponding densities are multiplied by factors such that in the overlapping regions the densities agree as much as possible. Finally, the density obtained in this way is normalized yielding the whole distribution. To estimate the errors of the distribution, this simulation is done a couple of times and the standard error of the single bins is used as an error estimate.

For the results, which we will present in the following section, we used data from both sampling techniques and in some cases merged them. Comparisons of both techniques showed that the errors introduced by WL have no considerable influence on our results (not shown).

For the determination of mean and variances of convex hull volume and surface the contribution of the tails are negligible, thus we used simple sampling, which enables the simulation of longer walks, i.e., larger values of T, in a larger range of dimensions $d = 2, \ldots, 6$,

III. RESULTS

A. Mean and variance

At first, we will verify our simulations by comparing with some analytically known results [33,38] for the mean volume V and surface ∂V scaled appropriately as $\mu_V = \langle V\rangle / T^{d\nu}$ and $\mu_{\partial V} = \langle \partial V\rangle / T^{(d-1)\nu}$. The scaling comes from the $r \propto T^{\nu}$ scaling of the RW end-to-end, distance, with $\nu = \frac{1}{2}$, in combination with the typical scaling $V \propto r^d$ and $\partial V \propto r^{d-1}$. For large T it is known [38] that

$$\mu_V^{\infty} = \left(\frac{\pi}{2}\right)^{d/2} \Gamma\left(\frac{d}{2}+1\right)^{-2}, \tag{2}$$

$$\mu_{\partial V}^{\infty} = \frac{2(2\pi)^{(d-1)/2}}{\Gamma(d)}. \tag{3}$$

This simulation uses simple sampling to sample 10^6 (fewer for $d = 6$ resulting in larger uncertainties) sufficiently long walks from $T_{\min} = 128$ up to $T_{\max} = 262\,144$.

There is an exact result for the mean Volume of the convex hull for finite T [39]:

$$\langle V\rangle = \frac{2^{-d/2}}{\Gamma(d/2+1)} \sum_{n_1,\ldots,n_d} \frac{1}{\sqrt{n_1 \ldots n_d}} I(n_1,\ldots,n_d), \tag{4}$$

where $1 \leqslant n_i \leqslant T$ are integers and

$$I(n_1,\ldots,n_d) = \begin{cases} 1 & \text{if } n_1 + \ldots + n_d \leqslant T \\ 0 & \text{else .} \end{cases}$$

For example, for $d = 2$ and $d = 3$ this results in

$$\langle V_2\rangle = \frac{1}{2} \sum_{i=1}^{T} \sum_{j=1}^{T-i} \frac{1}{\sqrt{ij}}, \tag{5}$$

$$\langle V_3\rangle = \frac{2^{3/2} \times 4}{3\sqrt{\pi}} \sum_{i=1}^{T} \sum_{j=1}^{T-i} \sum_{k=1}^{T-i-j} \frac{1}{\sqrt{ijk}}, \tag{6}$$

respectively. The number of elements in the sums grows with $O(T^d)$ in the number of steps T and the dimension d, such that a numerical evaluation is only feasible for rather small T and d. We calculated some exact values to ensure the quality of our simulations and the extrapolation. These are marked with crosses in Fig. 3.

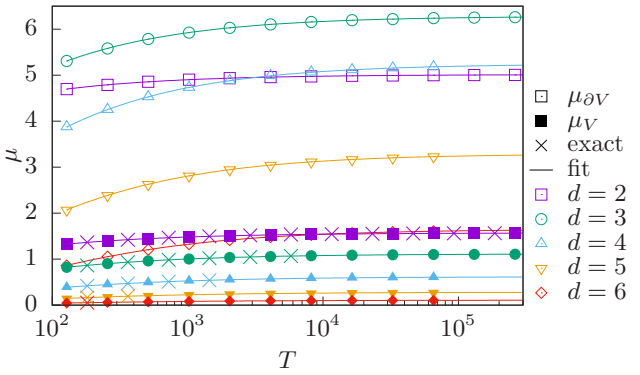


FIG. 3. Scaled mean of the surface $\mu_{\partial V} = \langle \partial V\rangle / T^{(d-1)\nu}$ (open symbols) and volume $\mu_V = \langle V\rangle / T^{d\nu}$ (solid symbols) for different dimensions (different shapes) and walk lengths T obtained by 10^6 samples each. Lines are fits [cf. Eq. (8)] to extrapolate for $T \to \infty$. Crosses are exact values [cf. Eq. (4)] and show very good agreement with the extrapolation. The asymptotic values are shown in Table I. Fits disregard small walk lengths for higher dimensions, since the expansion is valid for large T. To be precise, the fit ranges are $d \geqslant 5$: $T \geqslant 256$ for the surface and $d \geqslant 5$: $T \geqslant 512$ for the volume. They are chosen such that χ^2_{red} reaches a plateau, i.e., does not change significantly if even larger T are ignored (same ranges for the variances). The goodness of fit χ^2_{red} is between 0.3 and 1.2 for all fits. Error bars are smaller than the line of the fit.

To estimate the $T \to \infty$ limit asymptotic value μ_V^{∞}, it is necessary to extrapolate measurements for different lengths T of the walk. Recently in Ref. [41], the asymptotic large T expansion of the mean area of the convex hull of a 2D Gaussian random walk was worked out explicitly. For Gaussian jump distribution with zero mean and unit variance, it was found that the mean area $\langle A\rangle$ of the convex hull of a walk of T steps has the asymptotic expansion for large T,

$$\frac{\langle A\rangle}{T} = \frac{\pi}{2} + \gamma\sqrt{8\pi}\, T^{-1/2} + \pi(1/4+\gamma^2)\, T^{-1} + o(T^{-1}), \tag{7}$$

where the constant $\gamma = \zeta(1/2)/\sqrt{2\pi} = -0.58259\ldots$. This exact result in 2D leads to a natural guess in higher dimensions for the asymptotic large T expansion of the mean volume of the convex hull, namely,

$$\frac{\langle V\rangle}{T^{d\nu}} = \mu_V + C_1\, T^{-1/2} + C_2\, T^{-1} + o(T^{-1}). \tag{8}$$

This guess produces very good fits, shown in Fig. 3, with values in very good agreement with the expectations. We use the same function for the surface and the variances. Though small values of T need to be excluded from the fits, especially for high dimensions. The precise fit ranges are listed in the caption of Fig. 3.

The obtained asymptotic values are listed in Table I. Mind, that the error estimates are purely statistical and do not take into account higher order terms than those present in Eq. (8). To make matters worse, not the same large system sizes could be reached for higher dimensions due to the exponentially increasing time complexity [56].

TABLE I. Analytically expected (top, rounded to four decimal places) and from measurements extrapolated (bottom) asymptotic mean and variance of volume, respectively, surface. Analytical values for the variances are unknown (except for Brownian bridges [46]). Though for the perimeter ($d = 2$) rigorous bounds [73] are known $\sigma^{\infty 2}_{\partial V} \in [2.65 \times 10^{-3}, 9.87]$. Error estimates for the last column are obtained by Gaussian error propagation.

d	μ^{∞}_{V}	$\mu^{\infty}_{\partial V}$	$\sigma^{\infty 2}_{V}$	$\sigma^{\infty 2}_{\partial V}$	$\frac{\sigma^{\infty}_{V}}{\mu^{\infty}_{V}}$
2	1.5708	5.0132			
3	1.1140	6.2832			
4	0.6168	5.2499			
5	0.2800	3.2899			
6	0.1077	1.6493			
2	1.5705(3)	5.0127(5)	0.3078(3)	1.077(1)	0.3532(2)
3	1.1139(2)	6.2832(9)	0.1778(2)	3.093(3)	0.3785(2)
4	0.6164(1)	5.2473(10)	0.05882(7)	2.808(3)	0.3932(2)
5	0.2801(1)	3.2909(9)	0.01274(2)	1.279(2)	0.4032(3)
6	0.1077(1)	1.6492(6)	0.00193(1)	0.351(1)	0.4080(5)

Also, we looked at the average volume $\mu_V = \langle V \rangle / T^{dv}$ and surface $\mu_{\partial V} = \langle \partial V \rangle / T^{(d-1)v}$ of the convex hulls of multiple RWers with $n \in \{2,3,10,100\}$ independent RWs in $d = 3$ dimensions, which are tabulated in Table II. We determined the listed values in the same way as before with a fit to Eq. (8) (no figure shown) within the same ranges as single walks.

Since the single steps $\boldsymbol{\delta}_i$ are independent, two walkers, i.e., the $n = 2$ case, can be joined at the origin to one walk with twice the number of steps [74], thus $\mu^{\infty}_{V_2} = 2^{dv}\mu^{\infty}_{V}$ and $\mu^{\infty}_{\partial V_2} = 2^{(d-1)v}\mu^{\infty}_{\partial V}$ are the exact mean values for this case. The numerical data is within statistical errors compatible with this expectation. Though, for $n > 2$ this is not as easy anymore. We are not aware of any other published expectations for $d \geqslant 3$.

We have performed the same analysis (no figure shown) for the variances $\sigma^2_V = \mathrm{Var}(V)/T^{2dv}$ and $\sigma^2_{\partial V} = \mathrm{Var}(\partial V)/T^{2(d-1)v}$ and the same remarks apply.

For the ratio between standard deviation and mean,

$$\lim_{d\to\infty} \frac{\sigma^{\infty}_{V}}{\mu^{\infty}_{V}} = 0$$

is conjectured [38]. Our data shows no downward trend for this ratio as shown in the last column of Table I.

The argument of Ref. [38] is that the expectation of the second moment factorizes to the square of the first moment,

TABLE II. Analytically expected (top) and from measurements extrapolated (bottom) mean and variance of the volume, respectively, surface of the convex hull of n independent RWs in $d = 3$ dimensions. Analytical values for the variances are unknown. The quality of fit χ^2_{red} for all fits is between 0.4 and 1.7.

n	μ^{∞}_{V}	$\mu^{\infty}_{\partial V}$	$\sigma^{\infty 2}_{V}$	$\sigma^{\infty 2}_{\partial V}$
2	3.151	12.566		
2	3.153(1)	12.572(2)	1.427(1)	12.40(1)
3	5.332(1)	17.644(2)	3.796(4)	21.66(2)
10	17.695(2)	37.528(3)	22.54(3)	48.65(4)
100	66.233(7)	85.563(5)	68.65(10)	56.44(7)

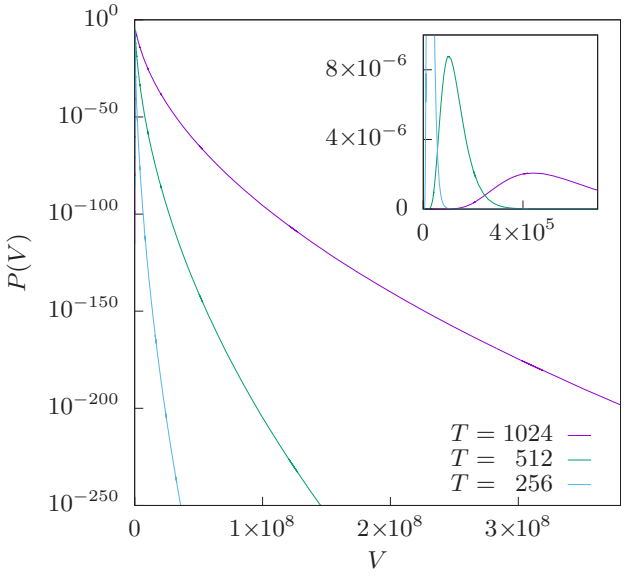


FIG. 4. Distribution of the volume of a $d = 4$ RW for different system sizes T. The inset shows the peak region in linear scale.

if all facets are orthogonal to each other. Since in high dimensions random vectors are with very high probability almost orthogonal, this suggests that the difference of these quantities, which is the variance, should be far smaller than each of them, i.e., the squared mean. Though, in the dimensions under scrutiny, i.e., $d \leqslant 6$ this effect seems not to dominate, because the facets have nonnegligible parallel components.

It is, of course, hard to estimate for which dimension the orthogonality starts to dominate. As a crude non-rigorous argument we take the scalar product of two random normalized vectors, which approximate the normal vectors of the facets. While its mean value is zero, its variance is $v = 1/d$. This variance is a measure for how parallel the two vectors are. Intuitively, it is clear that in $d = 2$ ($v = 1/2$) and $d = 3$ ($v = 1/3$) most facets have quite large and certainly not negligible parallel components. Then, for $d = 6$ the variance of $v = 1/6$ is not significantly smaller. We would assume that the factorization could dominate the other effects if the parallel component is far smaller—say, 1/20 or 1/100.

Therefore, to draw any conclusions, one should gather results for $d \gg 6$, which may be possible using some fast approximation scheme for convex hulls in high dimensions, though this is beyond the scope of this study.

B. Distributions

In addition to the first moments shown in the previous section, here we look at the actual distribution over a large part of the support. Since the Gaussian distribution, from which the steps are drawn, is not bounded, V and ∂V of a walk consisting of such steps are not bounded, either. Therefore, not the whole support, but a reasonably large part is sampled. Especially, it is large enough to investigate the large-deviation properties of the distribution. As an example, a part of the distribution for the volume of a convex hull of RWs in $d = 4$ dimensions is shown in Fig. 4.

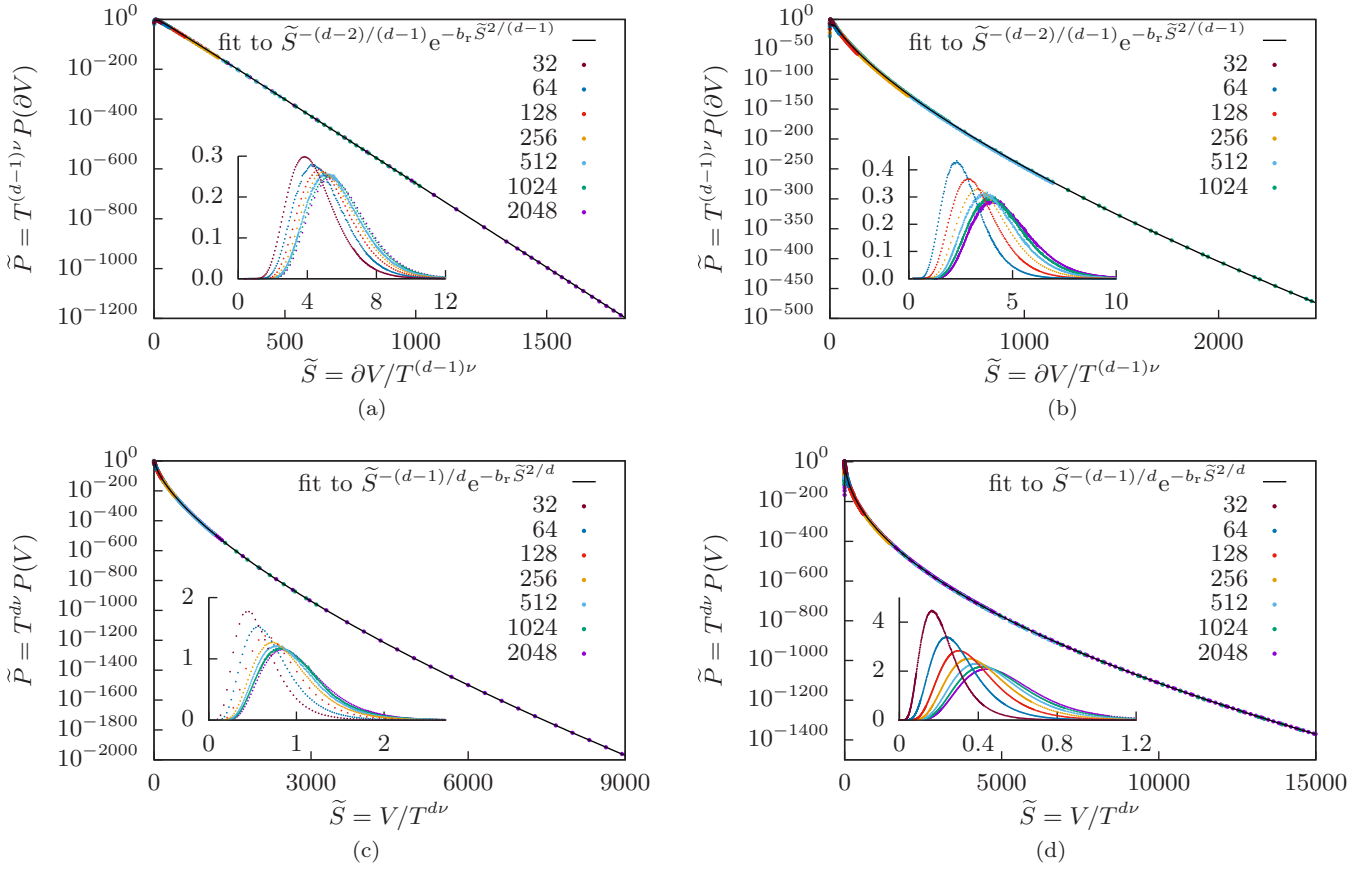


FIG. 5. Distributions of the surface (top) and volume (bottom) for $d \in \{3,4\}$ scaled according to Eq. (9). Statistical errors are smaller than the symbols. The scaling indeed collapses the distributions on one scaling function $\widetilde{P}$. Fits are shown for the largest system size. The inset shows the peak region in linear scale. For larger values of T the collapse works better. (Only a small fraction of all data points are visualized.) (a) $d = 3$, $\widetilde{S} > 500$, $b_r = 1.56$, $\chi^2_{\text{red}} = 2.5$, (b) $d = 4$, $\widetilde{S} > 200$, $b_r = 6.33$, $\chi^2_{\text{red}} = 1.2$, (c) $d = 3$, $\widetilde{S} > 500$, $b_r = 10.61$, $\chi^2_{\text{red}} = 0.8$, and (d) $d = 4$, $\widetilde{S} > 2500$, $b_r = 26.60$, $\chi^2_{\text{red}} = 1.1$.

As we mentioned in the previous section, $\langle V\rangle$ and $\langle \partial V\rangle$ scale for large values of T as $T^{d_e\nu}$ where d_e is the effective dimension of the observable, i.e., $d_e = d$ for the volume and $d_e = d-1$ for the surface. A natural question is, if the whole distribution does scale according to $T^{d_e\nu}$. Reference [47] already shows that this is true for $d = 2$. For higher dimension we arrive analogously at the scaling assumption for the distribution of the observable S,

$$P(S) = T^{-d_e\nu}\widetilde{P}(ST^{-d_e\nu}). \tag{9}$$

Figure 5 shows the distributions of the volume and surface of the convex hulls of RWs in $d \in \{3,4\}$ dimensions scaled according to Eq. (9). Apparently the scaling works very well in the right tail of larger than typical V. The inset shows that in the peak region there are major corrections to the assumed scaling for small values of T, but it also shows that those corrections rapidly get smaller for larger values of T. A power-law fit with offset to the position of the maxima of the distributions (no figure) with increasing walk length T, confirms convergence for large values of T, i.e., the peaks do collapse on one universal curve for $T \to \infty$.

In fact, the scaling for the distribution of the *span* s, which is the distance between the leftmost and rightmost point, of a one dimensional Brownian motion is known [1,75] to be

$$P(s,T) = (4DT)^{-\nu} f\left(\frac{s}{(4DT)^{\nu}}\right),$$

with some diffusion constant D and

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

which has the following asymptotic behavior [47]:

$$f(x) = 2\pi^2 x^{-5}\, e^{-\pi/4x^2}, \qquad \text{for } x \to 0,$$
$$f(x) = \frac{8}{\sqrt{\pi}}\, e^{-x^2}, \qquad \text{for } x \to \infty.$$

With this known $d = 1$ result for the span s, we can construct a guess for the higher dimensional observables, like the volume. Since the one-dimensional projection of a high dimensional RW has the same properties as a one dimensional RW (for this Gaussian model), we use the naive approach $S \propto s^{d_e}$, e.g., $V \approx s^d$. Substituting this into the known result leads to a guess for the expected behavior of the tails with

$$\widetilde{P}(\widetilde{S}) \propto \widetilde{S}^{-(d_e+4)/d_e}\, e^{-b_l\widetilde{S}^{-2/d_e}}, \qquad \text{for } \widetilde{S} \to 0, \tag{10}$$
$$\widetilde{P}(\widetilde{S}) \propto \widetilde{S}^{-(d_e-1)/d_e}\, e^{-b_r\widetilde{S}^{2/d_e}}, \qquad \text{for } \widetilde{S} \to \infty, \tag{11}$$

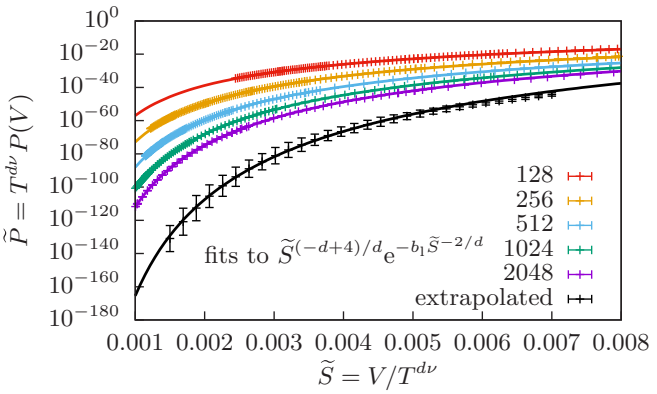


FIG. 6. Fit of the exponential Eq. (10) to the left tail. Note, that the prefactor is constant in this case (for clarity, only every tenths data point is plotted).

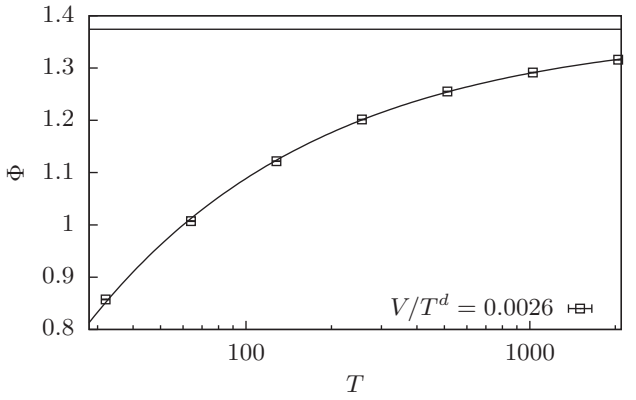


FIG. 7. Point-wise extrapolation of the value of the rate function at a fixed value V/T^d to $T \to \infty$ with a power law, here for a $d = 4$-dimensional volume. The power-law fit seems to be a reasonable approximation.

where a rescaled $\widetilde{S} = ST^{-\nu d_e}$ is introduced for clarity and with free parameters b_l and b_r. The $\frac{ds}{dS} \propto S^{-(d_e-1)/d_e}$ factors are introduced by the substitution.

For all values of T, the expected distribution for the left tail Eq. (10) fits well to the sampled data, shown for the example of the volume in $d = 4$ in Fig. 6. We extrapolated the curve point-wise to $T \to \infty$ assuming a power-law scaling, resulting in the limit curve in Fig. 6. Similar to the main region of the distribution (shown in Fig. 5), smaller values of T show larger deviations from the limit curve. Note that also the limiting curve fits Eq. (10) (with a suitable values for b_l and the prefactor).

The same analysis for the right tails is shown in Fig. 5, where Eq. (11) is fitted to the right tail of the distributions of the volume and surface in $d \in \{3,4\}$. The good χ^2_{red} values suggest that this is a good estimate of the asymptotic behavior indeed.

To determine whether a distribution P satisfies the large deviation principle [76], i.e., whether it scales as

$$P_T \approx e^{-T\Phi} \tag{12}$$

for some large parameter T, we look if the *rate function* Φ does exist in the $T \to \infty$ limit [76]. Comparing Eq. (12) to the behavior of the right tail [cf. Fig. 5 and Eq. (11)], the rate function seems to be a power law with an exponent $\kappa = 2/d_e$, i.e.,

$$\Phi(S) \propto S^{\kappa} = S^{2/d_e}. \tag{13}$$

Since we have numerical results for the distribution P, we can determine an empirical rate function Φ of the volume/surface S by extrapolation, (cf. Fig. 7) of

$$\Phi(S/S_{\max}) = -\frac{1}{T} \ln P(S/S_{\max}) \tag{14}$$

to the large T limit. While Φ is usually normalized to $\Phi \in [0,1]$, here S and thus Φ is not bounded. To get a rate function Φ comparable to other publications, we assume $S_{\max} = T^{d_e}$ like Ref. [47]. For the extrapolation we take values of different walk lengths T at multiple values of $S/S_{\max}$, e.g., V/T^d in Fig. 7. These can be thought of as vertical slices through the plot shown in Fig. 8. We use the measured values of Φ to extrapolate it point-wise to $T \to \infty$ using a power law with offset as shown in Fig. 7. Note that since for different walk lengths T we used different histogram bins, we obtain the intermediate values between the discrete bins by cubic spline interpolation. The extrapolation leads to an asymptotic rate function estimate. This shows that the rate function exists and this distributions satisfies the large-deviation principle. This holds for $d = 3$ and $d = 4$, for both volume and surface.

Fitting the power law Eq. (13) through the extrapolated points, as shown in Fig. 8 for the distribution of the volume in $d = 4$, confirms the expectation of $\kappa = 2/d_e$. This holds also for the other cases we considered (not shown as a figure). All measured values of κ are tabulated in Table III and are in reasonable agreement with the expectations. Further, the good quality of the fit χ^2_{red} and the good agreement of the exponents with the expectations, suggests that the power law is a reasonable ansatz and systematic errors due to deviations from this power law or finite-size effects are minor. Hence the given statistical errors should be reasonable. Since the same arguments are applicable for multiple walks, this procedure is

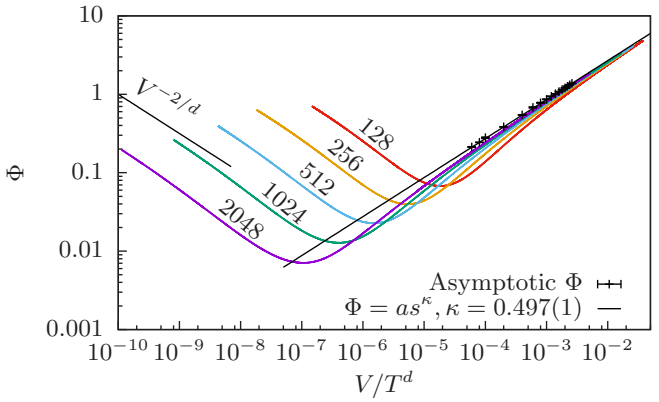


FIG. 8. Rate function of the distribution of the $d = 4$-dimensional volume of the convex hull of RWs for different walk lengths T. Crosses mark the $T \to \infty$ extrapolated values of the asymptotic rate function as shown in Fig. 7. To those a power law is fitted yielding an estimate for the rate function consistent with the guess in Eq. (13). Further, the expected power law behavior of the left tail is approached.

TABLE III. Comparison of expected and measured rate function exponent κ.

d	Volume V		Surface ∂V	
	Expected κ	Measured κ	Expected κ	Measured κ
2	1	0.994(4)	2	1.996(2)
3	2/3	0.665(1)	1	0.994(2)
4	1/2	0.497(1)	2/3	0.647(5)

tested for the distributions of $m = 3$ multiple walkers in $d = 3$ dimensions, which does also yield within errorbars the same exponent $\kappa = 0.642(17)$ as for the single walker (no figure).

Also note that the power-law relation for the left tail becomes visible, in the far left tail. The expected slope of the left tail $\Phi \propto s^{-2/d}$ [cf. Eq. (10)] is visualized in the far left tail in Fig. 8 and seems to be a reasonable approximation.

IV. CONCLUSIONS

We studied the volume and surface of convex hulls of RWs in up to $d = 6$ dimensions for which we confirmed the analytically known asymptotic means and we estimated the asymptotic variances.

Further, using sophisticated large-deviation sampling techniques we obtained large parts of the distributions P in up to $d = 4$ dimensions down to probability densities far smaller than $P = 10^{-1000}$. The distributions collapse over large ranges of the support (right tail) onto a single curve when being rescaled with the asymptotic behavior of the means. For the left tail, we observe a convergence to a limiting function. Even more, we used our results to confirm the expected functional shapes of the distributions in the left and the right tails, for finite and extrapolated values of T, respectively.

We used simple arguments and numerical simulations to determine the scaling behavior, as well as the asymptotic behavior also for both tails of the rate function $\Phi_r(S) \propto S^{2/d_e}$ for $d \in \{3,4\}$ and are confident that it is valid in arbitrary dimensions.

For future studies, it would be interesting to investigate the properties of the convex hulls of other types of random walks, exhibiting non-trivial values of ν, like self-avoiding walks or loop-erased RWs.

ACKNOWLEDGMENTS

We thank T. Dewenter and S. Zambare for interesting discussions and performing some preliminary test simulations. This work was supported by the German Science Foundation (DFG) through Grant No. HA 3169/8-1. H.S. and A.K.H. thank the LPTMS for hospitality and financial support during one- and two-month visits, respectively, where a considerable part of the projects were performed. The simulations were performed at the HPC clusters HERO and CARL, both located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Programme (INST Grant No. 184/108-1 FUGG and INST Grant No. 184/157-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State.

[1] B. D. Hughes, *Random Walks and Random Environments* (Clarendon Press, Oxford, 1996).
[2] K. Pearson, Nature **72**, 294 (1905).
[3] J. W. Strutt, London Edinburgh Dublin Philos. Mag. J. Sci. **37**, 321 (1919).
[4] G. Pólya, Math. Ann. **84**, 149 (1921).
[5] C. S. Patlak, Bull. Math. Biophys. **15**, 311 (1953).
[6] P. M. Kareiva and N. Shigesada, Oecologia **56**, 234 (1983).
[7] P. Bovet and S. Benhamou, J. Theor. Biol. **131**, 419 (1988).
[8] N. Madras and G. Slade, *The Self-avoiding Walk* (Springer, New York, NY, 2013), Chap. Analysis of Monte Carloimethods, pp. 281–364.
[9] G. F. Lawler, Duke Math. J. **47**, 655 (1980).
[10] A. Weinrib and S. A. Trugman, Phys. Rev. B **31**, 2993 (1985).
[11] M. V. Smoluchowski, Ann. Phys. **353**, 1103 (1916).
[12] W. Alt, J. Math. Biol. **9**, 147 (1980).
[13] P. J. van Haastert and M. Postma, Biophys. J. **93**, 1787 (2007).
[14] B. Weesakul, Ann. Math. Stat. **32**, 765 (1961).
[15] M. Kac, Am. Math. Mon. **54**, 369 (1947).
[16] M. E. Fisher, J. Stat. Phys. **34**, 667 (1984).
[17] G. Schehr, S. N. Majumdar, A. Comtet, and J. Randon-Furling, Phys. Rev. Lett. **101**, 150601 (2008).
[18] T. A. Witten and L. M. Sander, Phys. Rev. B **27**, 5686 (1983).
[19] D. W. Schaefer, Science **180**, 1293 (1973).
[20] E. A. Codling, M. J. Plank, and S. Benhamou, J. R. Soc., Interface **5**, 813 (2008).
[21] E. F. Fama, Financ. Anal. J. **21**, 55 (1965).
[22] M. Rosvall and C. T. Bergstrom, Proc. Natl. Acad. Sci. USA **105**, 1118 (2008).
[23] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, in *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 (ACM, New York, NY, 2013), pp. 505–514.
[24] E. Dumonteil, S. N. Majumdar, A. Rosso, and A. Zoia, Proc. Natl. Acad. Sci. USA **110**, 4239 (2013).
[25] W. Kuhn, Kolloid-Zeitschrift **68**, 2 (1934).
[26] E. Helfand, J. Chem. Phys. **62**, 999 (1975).
[27] C. Haber, S. A. Ruiz, and D. Wirtz, Proc. Natl. Acad. Sci. USA **97**, 10792 (2000).
[28] F. Bartumeus, M. G. E. da Luz, G. M. Viswanathan, and J. Catalan, Ecology **86**, 3078 (2005).
[29] L. Börger, B. D. Dalziel, and J. M. Fryxell, Ecol. Lett. **11**, 637 (2008).
[30] B. J. Worton, Biometrics **51**, 1206 (1995).
[31] L. Giuggioli, J. R. Potts, and S. Harris, PLoS Comput. Biol. **7**, e1002008 (2011).
[32] Y. Lanoiselée and D. S. Grebenkov, Phys. Rev. E **96**, 022144 (2017).
[33] L. T. Gérard Letac, Am. Math. Mon. **87**, 142 (1980).
[34] G. Letac, J. Theoret. Probabil. **6**, 385 (1993).
[35] J. Randon-Furling, S. N. Majumdar, and A. Comtet, Phys. Rev. Lett. **103**, 140602 (2009).
[36] S. N. Majumdar, A. Comtet, and J. Randon-Furling, J. Stat. Phys. **138**, 955 (2010).

[37] M. Chupeau, O. Bénichou, and S. N. Majumdar, Phys. Rev. E **91**, 050104 (2015).
[38] R. Eldan, Electron. J. Probab. **19**, 1 (2014).
[39] Z. Kabluchko and D. Zaporozhets, Trans. Amer. Math. Soc. **368**, 8873 (2016).
[40] V. Vysotsky and D. Zaporozhets, arXiv:1506.07827 (2015).
[41] D. S. Grebenkov, Y. Lanoiselée, and S. N. Majumdar, J. Stat. Mech. (2017) 103203.
[42] J. Kampf, G. Last, and I. Molchanov, Proc. Am. Math. Soc. **140**, 2527 (2012).
[43] M. Luković, T. Geisel, and S. Eule, New J. Phys. **15**, 063034 (2013).
[44] A. Reymbaut, S. N. Majumdar, and A. Rosso, J. Phys. A: Math. Theor. **44**, 415001 (2011).
[45] T. L. Snyder and J. M. Steele, Proc. Am. Math. Soc. **117**, 1165 (1993).
[46] A. Goldman, Probab. Theory Relat. Fields **105**, 57 (1996).
[47] G. Claussen, A. K. Hartmann, and S. N. Majumdar, Phys. Rev. E **91**, 052104 (2015).
[48] T. Dewenter, G. Claussen, A. K. Hartmann, and S. N. Majumdar, Phys. Rev. E **94**, 052120 (2016).
[49] A. Akopyan and V. Vysotsky, arXiv:1606.07141 (2016).
[50] R. Duda, P. Hart, and D. Stork, *Pattern Classification* (Wiley, New York, 2012).
[51] W. K. Cornwell, D. W. Schwilk, and D. D. Ackerly, Ecology **87**, 1465 (2006).
[52] F. P. Preparata and M. I. Shamos, Convex hulls: Basic algorithms, in *Computational Geometry: An Introduction* (Springer, New York, NY, 1985), pp. 95–149.
[53] M. Jayaram and H. Fleyeh, Amer. J. Intell. Syst. **6**, 48 (2016).
[54] K. Q. Brown, Info. Process. Lett. **9**, 223 (1979).
[55] F. Aurenhammer, ACM Comput. Surv. **23**, 345 (1991).
[56] V. Klee, Archiv der Mathematik **34**, 75 (1980).
[57] R. Seidel, A convex hull algorithm optimal for point sets in even dimensions, Ph.D. thesis, University of British Columbia, 1981.
[58] K. L. Clarkson and P. W. Shor, Discrete Comput. Geom. **4**, 387 (1989).
[59] Z.-B. Xu, J.-S. Zhang, and Y.-W. Leung, Appl. Math. Comput. **94**, 193 (1998).
[60] T. L. V. Hossein Sartipizadeh, Computing the approximate convex hull in high dimensions, (2016), arXiv:1603.04422.
[61] W. F. Eddy, ACM Trans. Math. Softw. **3**, 398 (1977).
[62] A. Bykat, Info. Process. Lett. **7**, 296 (1978).
[63] E. Mücke, Comput. Sci. Eng. **11**, 54 (2009).
[64] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, ACM Trans. Math. Softw. **22**, 469 (1996).
[65] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[66] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
[67] B. J. Schulz, K. Binder, M. Müller, and D. P. Landau, Phys. Rev. E **67**, 067102 (2003).
[68] R. E. Belardinelli and V. D. Pereyra, Phys. Rev. E **75**, 046701 (2007).
[69] R. E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007).
[70] J. Lee, Phys. Rev. Lett. **71**, 211 (1993).
[71] R. Dickman and A. G. Cunha-Netto, Phys. Rev. E **84**, 026701 (2011).
[72] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013).
[73] A. R. Wade and C. Xu, Stoch. Processes Appl. **125**, 4300 (2015).
[74] A. Engel, personal communications (2016).
[75] A. Kundu, S. N. Majumdar, and G. Schehr, Phys. Rev. Lett. **110**, 220602 (2013).
[76] H. Touchette, Phys. Rep. **478**, 1 (2009).