Large deviations of convex hulls of self-avoiding random walks

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

A global picture of a random particle movement is given by the convex hull of the visited points. We obtained numerically the probability distributions of the volume and surface of the convex hulls of a selection of three types of self-avoiding random walks, namely, the classical self-avoiding walk, the smart-kinetic self-avoiding walk, and the loop-erased random walk. To obtain a comprehensive description of the measured random quantities, we applied sophisticated large-deviation techniques, which allowed us to obtain the distributions over a large range of support down to probabilities far smaller than $P = 10^{-100}$. We give an approximate closed form of the so-called large-deviation rate function Φ which generalizes above the upper critical dimension to the previously studied case of the standard random walk. Further, we show correlations between the two observables also in the limits of atypical large or small values.

I. INTRODUCTION

The standard random walk is a simple Markovian process which has a history as a model for diffusion. There are many exact results known [1]. If memory is added to the model, e.g., to interact with the past trajectory of the walk, analytic treatment becomes much harder. A class of self-interacting random walks that we will focus on in this study are *self-avoiding* random walks, which live on a lattice and do not visit any site twice. This can be used to model systems with excluded volume, e.g., polymers whose single monomers cannot occupy the same site at once [2]. There are more applications which are not as obvious, e.g., a slight modification of the *smart-kinetic self-avoiding walk* traces the perimeter of critical percolation clusters [3], while the *loop-erased random walk* can be used to study spanning trees [4] (and vice versa [5]).

One of the central properties of random walk models is the exponent ν, which characterizes the growth of the end-to-end distance r with the number of steps T, i.e., $r \propto T^\nu$. While this has the value $\nu = 1/2$ for the standard random walk, its value is larger for the self-avoiding variations, which are effectively pushed away from their past trajectory. In two dimensions, this value (and other properties) can often be obtained by the correspondence to Schramm-Loewner evolution [6–9]. But between two dimensions and the upper critical dimension, above which the behavior is the same as the standard random walk, Monte Carlo simulations are used to obtain estimates for the exponent ν.

Here we want to study the convex hulls of a selection of self-avoiding walk models featuring larger values of ν. The convex hull allows one to obtain a global picture of the space occupied by a walk without exposing all details of the walk. As an example, convex hulls are used to describe the home ranges of animals [10–12] or the spatial extent of animal epidemics [13]. In physics, they have been proposed to be applied for the analysis of surface diffusion or the detection of binding of molecules [14]. Here, more fundamentally, we will look at the *smart-kinetic self-avoiding walk* (SKSAW), the classical *self-avoiding walk* (SAW), and the *loop-erased random walk* (LERW), since they span a large range of ν values and are well established in the literature. About the convex hulls of standard random walks, we already know plenty of properties. The mean perimeter and area have been known exactly for over 20 years [15,16] for large walk lengths T, i.e., the Brownian motion limit. Since then simpler and more general methods were devised based on Cauchy's formula which relates the support function of a curve to the perimeter and the area enclosed by the curve [17,18]. More recently, also the mean hypervolume and surface for arbitrary dimensions was calculated [19]. For discrete-time random walks with jumps from an arbitrary distribution, the perimeters of the convex hull for finite (but large) walk lengths T were computed explicitly [20]. For the case of Gaussian jump lengths, even an exact combinatorial formula for the volume in arbitrary dimensions is known [21]. For the variance there is an exact result for Brownian bridges [22]. Concerning the full distributions, no exact analytical results are available. Here sophisticated large-deviation simulations were used to numerically explore a large part of the full distribution, i.e., down to probabilities far smaller than 10^{-100} [23–25]. Numerical studies of this kind, which are able to obtain the distribution over a wide range including the extreme tails, are useful to check predictions about, e.g.,

---

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

large-deviation rate functions such as discussed in Ref. [26]. Furthermore, they can explore new territory and stimulate other studies of large-deviation properties. For example Ref. [23] shows numerical results that the distribution of area and perimeter of the convex hulls of planar standard random walks obey the large-deviation principle, which was later proven by Ref. [27].

Despite this increasing interest in the convex hulls of standard random walks, there seem to be no studies treating the convex hulls of self-avoiding walks. To fill this void, we use Markov chain Monte Carlo sampling to obtain the distributions of some quantities of interest over their whole support. To connect to previous studies [23–25] we also compare the aforementioned variants to the standard random walk on a square lattice (LRW). We are mainly interested in the full distribution of the area A and the perimeter L of $d = 2$ dimensional hulls for walks in the plane, since the effects of the self-interactions are stronger in lower dimensions, but we will also look into the volume V in the $d = 3$ dimensional case. In the past study on standard random walks [25] we found that the full distribution can be scaled to a universal distribution using only the exponent ν and the dimension for large walk lengths T. For the present case, where a walk might depend on its full history, one could expect a more complex behavior. Nevertheless, our results presented below show convincingly that also for self-interacting walks the distributions are universal and governed mainly by the exponent ν, except for some finite-size effects, which are probably caused by the lattice structure. Furthermore, we use the distributions to obtain empirical large-deviation rate functions [28], which suggests that a limiting rate function is mathematically well defined. We also give an estimate for the rate function, which is compatible with the known case of standard random walks and with all cases under scrutiny in this study.

II. MODELS AND METHODS

This section gives a short overview over the models and methods used, with references to literature more specialized on the corresponding subject. Where we deem adequate, also technical details applicable for this study are mentioned.

A. Sampling scheme

To generate the whole distribution of the area or perimeter of the convex hull of a random walk over its full support, a sophisticated Markov chain Monte Carlo (MCMC) sampling scheme is applied [29,30]. The Markov chain is here a sequence of different walk configurations. The fundamental idea is to treat the observable S, i.e., the perimeter, area, or volume, as the energy of a physical system which is coupled to a heat bath with adjustable "temperature" Θ and to sample its equilibrium distribution using the Markov chain. This can be easily done using the classical Metropolis algorithm [31]. Therefore the current walk configuration is changed a bit. (The precise type of change is dependent on the type of walk we are looking at and is explained in the following sections.) The changes must be designed in a way that any configuration can be reached from any other configuration in finite time, i.e., ergodicity must be given. The change is accepted with the acceptance
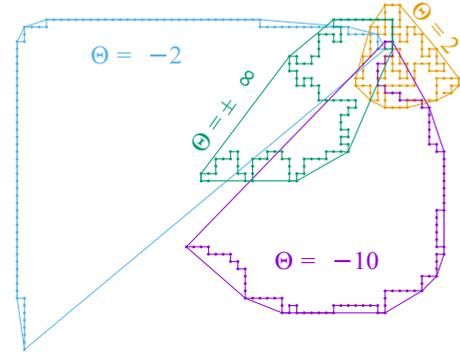

FIG. 1. Typical SAW configurations with $T = 200$ steps and their convex hulls at different temperatures Θ. $\Theta = \pm\infty$ corresponds to a typical configuration without bias.

probability

$$p_{\mathrm{acc}} = \min\{1, \mathrm{e}^{-\Delta S/\Theta}\} \tag{1}$$

and rejected otherwise, which fulfills detailed balance. This means, at long times the Markov process yields configurations $\mathcal{C}$ from its equilibrium distribution $Q_\Theta(\mathcal{C}) = \frac{1}{Z(\Theta)} Q(\mathcal{C}) \mathrm{e}^{-S(\mathcal{C})/\Theta}$, with the partition function $Z(\Theta)$. For the distribution of the observable $P(S)$ this means

$$P_\Theta(S) = \sum_{\{\mathcal{C}|S(\mathcal{C})=S\}} Q_\Theta(\mathcal{C}) \tag{2}$$

$$= \sum_{\{\mathcal{C}|S(\mathcal{C})=S\}} \frac{\exp(-S/\Theta)}{Z(\Theta)} Q(\mathcal{C}) \tag{3}$$

$$= \frac{\exp(-S/\Theta)}{Z(\Theta)} P(S). \tag{4}$$

That means the "temperature" Θ will bias the configuration towards specific ranges of the "energy" S. Configurations at small and negative Θ will show larger than typical S; small and positive Θ show smaller than typical S and large values independent of the sign sample configurations from the peak of the distribution. Figure 1 shows typical walk configurations of the self-avoiding walk at different values of Θ.

In a second step, histograms of the equilibrium distribution $P_\Theta(S)$ are corrected for the bias introduced via Θ. Using Eq. (4), we can easily remove this bias and arrive at the unbiased distribution

$$P(S) = \mathrm{e}^{S/\Theta} Z(\Theta) P_\Theta(S). \tag{5}$$

The free parameter $Z(\Theta)$ can be obtained by enforcing continuity and normalization of the distribution. This necessitates that we perform this sampling at multiple Θ such that there are good statistics over the whole range and overlapping histograms from which to choose $Z(\Theta)$, so that the overlapping regions coincide, i.e., the distribution is continuous. This, at the same time, serves as a quality estimate of the Markov process, since the overlaps will only coincide cleanly over their whole range if the samples were taken in equilibrium. So a clean coincidence is a strong hint at a good quality of the data. Further details and examples can be found in several other articles, where it has been applied and explained for specific models [26,29,30,32–35], but also in a very general form [36].

In particular, the algorithm was already used successfully in other studies looking at the large deviation properties of convex hulls of random walks [23,24].

B. Lattice random walk

All of the self-interacting random walks, which are the focus of this study, are typically treated on a lattice. Hence, we will start by introducing the simple, i.e., noninteracting, isotropic random walk on a lattice. For simplicity we will use a square lattice with a lattice constant of 1. A realization consists of T randomly chosen discrete steps $\boldsymbol{\delta}_i$. Here we use steps between adjacent lattice sites, i.e., d-dimensional Cartesian base vectors $\boldsymbol{e}_i$, which are drawn uniformly from $\{\pm \boldsymbol{e}_i\}$. The realization can be defined as the tuple of the steps $(\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_T)$ and the position at time τ as

$$\boldsymbol{x}(\tau) = \boldsymbol{x}_0 + \sum_{i=1}^{\tau} \boldsymbol{\delta}_i. \tag{6}$$

Here we set the start point $\boldsymbol{x}_0$ at the coordinate origin. The set of visited sites is therefore $\mathcal{P} = \{\boldsymbol{x}(0), \dots, \boldsymbol{x}(T)\}$.

The central quantity of the LRW is the average end-to-end distance

$$r = \sqrt{\langle (\boldsymbol{x}(T) - \boldsymbol{x}_0)^2 \rangle}, \tag{7}$$

where $\langle \dots \rangle$ denotes the average over the disorder. It grows polynomially and is characterized by the exponent ν via $r \propto T^\nu$. For the LRW it is $\nu = 1/2$, which is typical for all diffusive processes.

As the change move for the Metropolis algorithm, we replace a randomly chosen $\boldsymbol{\delta}_i$ by a new randomly drawn displacement. This way we can clearly reach any possible configurations, i.e., ergodicity holds. Since our quantity of interest is the convex hull, i.e., a global property of the walk, we do not profit much from local moves, e.g., crankshaft moves. Thus we use this simple, global move.

C. Smart-kinetic self-avoiding walk

The smart-kinetic self-avoiding walk [3,37] is probably the most naive approach to a self-avoiding walk. It grows on a lattice and never enters sites it already visited. Since it is possible to get trapped on an island inside already-visited sites, this walk needs to be *smart* enough to never enter such traps.

In $d = 2$ it is possible to avoid traps using just local information in constant time using the *winding angle* method [37]. In conjunction with hash table backed detection of occupied sites, a realization with T steps can be constructed in time $\mathcal{O}(T)$.

This method will typically yield longer stretched walks than the LRW due to the constraint that it needs to be self-avoiding. This can be characterized by the exponent ν, which is larger than $1/2$ in $d = 2$.

The sketch of Fig. 2 shows that this ensemble does not contain every configuration with the same probability but prefers closely winded configurations. This is also visible in Fig. 3(b). This is characterized by the exponent $\nu = 4/7$ [9], which is larger than ν for the LRW but smaller than for the SAW. Also note that it is conjectured that the upper critical dimension is $d = 3$ [37], i.e., $\nu = 1/2$ for all $d \geqslant 3$—possibly
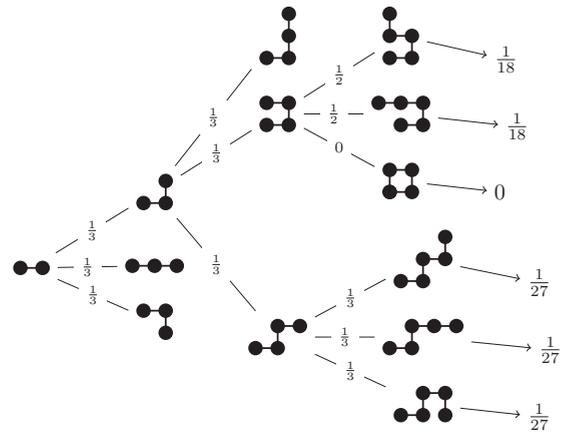

FIG. 2. Decision tree visualizing the probability to arrive at certain configurations following the construction rules of the SKSAW. Not all possible configuration have the same probability, and hence this rule defines a different ensemble than SAW.

with logarithmic corrections in $d = 3$. Therefore only $d = 2$ is simulated in this study.

While it is easy to draw realizations from this ensemble uniformly, i.e., simple sampling, it is not so straightforward to apply the MCMC changes. If one just changes single steps like for the LRW and accepts if it is self-avoiding or rejects if it is not, one will generate all self-avoiding walk configurations with equal probability. Our approach to generate realizations according to this ensemble handles the construction of the walk as a *black box*. It acts on the random numbers used to generate a realization from scratch. During the MCMC at each iteration one random number is replaced by a new random number and a SKSAW realization is regenerated from scratch using the modified random numbers [36]. Since every configuration of underlying random numbers can occur this way, every possible SKSAW configuration can be constructed, such that this protocol is ergodic. This change is then accepted according to Eq. (1) and undone otherwise.
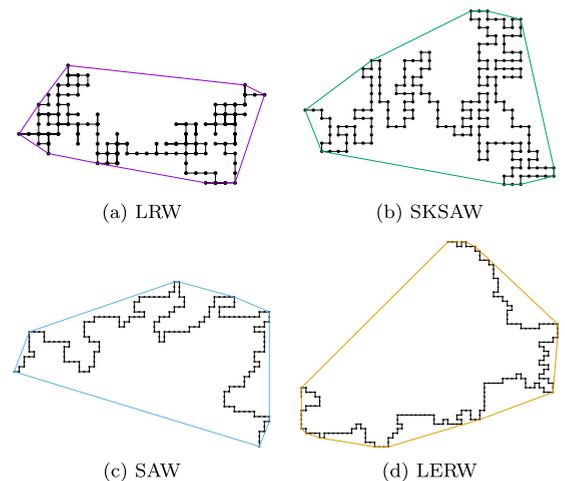

FIG. 3. Typical configurations with $T = 200$ steps, drawn uniformly from the corresponding ensembles, of all types of random walks under scrutiny in this study with their convex hulls.

D. Self-avoiding random walk

While the above-mentioned SKSAW does produce self-avoiding walks, SAW denotes another ensemble, the ensemble where realizations are drawn uniformly from the set of all self-avoiding configurations. It is not trivial to sample from this distribution efficiently. The black-box method used for SKSAW is not feasible, since the construction of a SAW takes time exponential in the length with simple methods like dimerization [2,38]. It is possible to perform changes directly on the walk configuration and accept them according to Eq. (1), but their rejection rate is typically quite high and the resulting configurations are very similar [2], which makes this inefficient. The state-of-the-art method to sample SAW is the *pivot algorithm* [2]. It chooses a random point and uses it as the pivot for a random symmetry operation, i.e., rotation or mirroring. If the resulting configuration is not self-avoiding, it is rejected. Otherwise we accept it with the temperature-dependent acceptance probability Eq. (1).

As mentioned previously, the exponent $\nu = 3/4$ [7] is larger than for the SKSAW. Since the upper critical dimension for SAW is $d = 4$, this study will also look at $d = 3$, where an exact value of ν is not known and the best estimate is $\nu = 0.587\,597(7)$ [39], though our focus is on $d = 2$ for this type.

While there are highly efficient implementations of the pivot algorithm [39,40], the time complexity of the problem at hand is dominated by the time needed to construct the convex hull; thus we go with the simple hash-table-based $\mathcal{O}(T)$ approach [2].

E. Loop-erased random walk

The LERW [41] uses a different approach to achieve the self-avoiding property. It is built as a simple LRW, but each time a site is entered for the second time, the loop that is formed, i.e., all steps since the first entering of this site, is erased. While this ensures no crossings in the walk, the resulting ensemble is different from the SAW ensemble and the walks are longer stretched out, as characterized by the larger exponent $\nu = 4/5$ [5,8,42]. Similar to the SAW, the upper critical dimension is $d = 4$ and an estimate for $d = 3$ is $\nu = 0.615\,76(2)$ [43].

For construction—similar to SKSAW—we need to keep all used random numbers and change them in the MCMC algorithm. This leads to a dramatically higher memory consumption than simple sampling, where each loop can be discarded as soon as it is closed.

F. Convex hulls

We will study the *convex hulls* $\mathcal{C}$ of the sites visited by the random walk $\mathcal{P}$. The convex hull of a point set $\mathcal{P}$ is the smallest polytope containing all points $P_i \in \mathcal{P}$ and all line segments (P_i, P_j). Some example hulls are shown in Fig. 3.

Convex hulls are one of the most basic concepts in computational geometry,[1] with noteworthy application

---

[1]Three of the first four examples for static problems of computational geometry in Wikipedia can utilize convex hulls for their solution (https://en.wikipedia.org/wiki/Computational_geometry, 12.01.2018).

in the construction of Voronoi diagrams and Delaunay triangulations [44].

For point sets in the $d = 2$ plane, we use Andrew's *monotone chain* [45] algorithm for its simplicity and *Quickhull* [46] as implemented by QHULL [47] for $d = 3$. Both algorithms have a time complexity of $\mathcal{O}(T \ln T)$. In $d = 2$ Andrew's monotone chain algorithm results in ordered points of the convex hull. Adjacent points (i, j) in this ordering are the line segments of the convex hull. Quickhull results in the simplicial facets of the convex hull.

To obtain the perimeter of a $d = 2$ convex hull, we sum the lengths of its line segments L_{ij}. To calculate the area and the volume, we use the same fundamental idea. In both cases we subdivide the area/volume into simplexes, i.e., triangles for the area and tetrahedra for the volume. Therefore we choose an arbitrary fixed point p_0 inside of the convex hull and construct a simplex for each facet f_m, i.e., for $d = 2$ each line segment of the hull $f_m = (i, j)$ forms a triangle (i, j, p_0), and each triangular face $f_m = (i, j, k)$ of a $d = 3$ dimensional polyhedron forms a tetrahedron with p_0. The volume of a triangle is trivially

$$A_{ijp_0} = \tfrac{1}{2} \operatorname{dist}(f_m, p_0) L_{ij},$$

where $\operatorname{dist}(f_m, p_0)$ is the perpendicular distance from a facet f_m to a point p_0. Since the union of all triangles built this way is the whole polygon, the sum of their areas is the area of the polygon. Similarly, the volume of a polyhedron is the sum of the volumes of all tetrahedra constructed from its faces. The volume of the individual tetrahedra is given by

$$V_{ijkp_0} = \tfrac{1}{3} \operatorname{dist}(f_m, p_0) A_{ijk}.$$

For random walks on a lattice with T steps of length 1 in d dimensions the maximum volume is

$$S_{\max} = \frac{(T/d_e)^{d_e}}{d_e!} \tag{8}$$

for T divisible by the effective dimension d_e of the observable, e.g., 2 for the area of a planar hull or 3 for the volume in three dimensions. For example, the configuration of maximum area corresponds to an L shape, i.e., $A_{\max} = \frac{T^2}{8}$. This form can be derived by the general volume of a d-dimensional simplex defined by its $d + 1$ vertices $\boldsymbol{v}_i$ [48]:

$$V = \frac{1}{d!} \det (\boldsymbol{v}_1 - \boldsymbol{v}_0, \ldots, \boldsymbol{v}_d - \boldsymbol{v}_0). \tag{9}$$

Without loss of generality, we set $\boldsymbol{v}_0$ to be the coordinate origin. To achieve maximum volume all $\boldsymbol{v}_i, i > 0$ need to be orthogonal and of equal length. Thus a random walk going T/d steps along some base vector $\boldsymbol{e}_i$ and continuing with T/d steps in direction $\boldsymbol{e}_{i+1}$ has a convex hull defined by the tetrahedron specified by $\boldsymbol{v}_i = \sum_{j=1}^{i} \frac{T}{d} \boldsymbol{e}_j$. The matrix $M = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d)$ is thus triangular and its determinant is the product of its diagonal entries $M_{ii} = \frac{T}{d}$, which leads directly to Eq. (8). An exception occurs in $d = 2$, where the perimeter is $L_{\max} = 2T$.

III. RESULTS

The focus of this work lies on $d = 2$ dimensional SAW and LERW. The results for higher dimensions and for SKSAW are generated with less numerical accuracy. The LRW results also

have a lower accuracy, as their purpose is mainly to scrutinize the effect of the lattice structure underlying all considered walk types in comparison to the nonlattice results from [25]. Also, not all combinations are simulated but only those listed with a value in Table II.

The same raw data is evaluated for equidistant bins and logarithmic bins. And the respective variants are shown according to the scaling of the x axis.

A. Correlations

To get an intuition for how the configurations with atypical large areas A or perimeters L look like, we visualize the correlation between these two observables as scatter plots in Fig. 4.

Since the smallest possible SAW is an (almost) fully filled square, there cannot be instances below some threshold, which explains the gaps on the left side of the scatter plots and of the distributions shown in the following section. In the center of the scatter plots, which is already in probability regions far beyond the capabilities of simple sampling methods, the behavior becomes strongly dependent on the bias.

If biasing for large perimeters (top), the area shows a nonmonotonous behavior. First, somehow larger perimeters come along typically with larger areas for entropic reasons, i.e., there are fewer configurations which are long and thin, and more bulky, which have a larger area. Though, for the far-right tail, the only configurations with extreme large perimeters are almost linelike and thus have a very small area. Also note that the excluded volume effect of the SAW leads to overall larger areas at the same perimeters.

On the other hand, when biasing for large areas (bottom) the configurations with largest area, which are L-shaped

(cf. Fig. 1), unavoidably have quite large perimeters; hence the scatter plots show an almost linear correlation between area and perimeter. Since the configurations of large areas naturally avoid self-intersections, since steps on already visited points do not enlarge the convex hull, the differences between LRW and SAW diminish in the right tail. Note that with the large-area bias, no walks with the very extreme perimeters exist, for the reason already mentioned.

Note, however, that these scatter plots are very dependent on which observable we are biasing for. In principle we observe that small perimeters are strongly correlated with small areas, while for large but not too large perimeters, there is a broad range of area sizes possible. For extremely large perimeters, the area must be small. For a comprehensive analysis, one would need a full two-dimensional histogram, which could be obtained using Wang-Landau sampling but which is beyond the scope of this study and would require a much larger numerical effort. Nevertheless, from looking at Fig. 4 one can anticipate that the two-dimensional histogram would exhibit a strong correlation for small values of L and a broad scatter of the accessible values of A for larger but not too large values of L.

B. Moments and distributions

The distributions of the different walk types differ considerably. This can be observed in Fig. 5, where distributions of the area A for all types with $T = 1024$ steps are drawn. The main part of the distribution shifts to larger values for larger values of ν as expected, and the probability of atypically large areas is boosted even more in the tails.

In the right tail, the distributions seem to bend down. Below, where we show results for different walk sizes T, we see that this is a finite-size effect of the lattice structure and the fixed step length. This can be seen also as follows: Since the lattice together with the fixed step length sets an upper bound on the area, the probability plummets near this bound for entropic reasons, i.e., there are for any walk length T only eight configurations with maximum area (due to symmetries) such that all self-avoiding types will meet at this point (not visible because the bins are not fine enough).
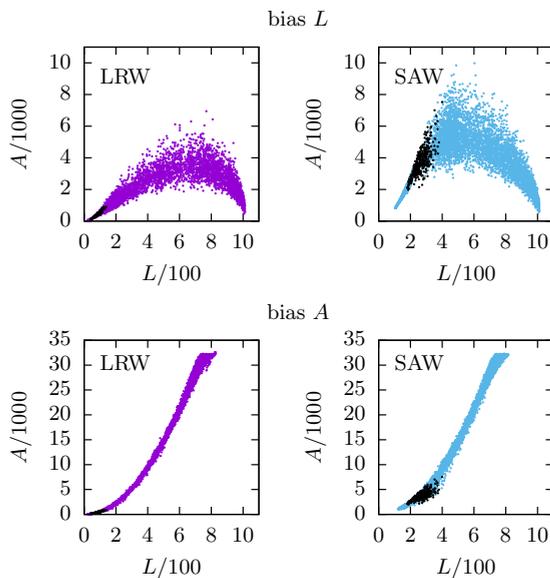

FIG. 4. The top row shows data from simulations biasing towards larger (and smaller) than typical perimeters L. The bottom row biases the area A. The left column shows data from LRW and the right from SAW both with $T = 512$ steps. The results of simple sampling are shown in black. Note that only very narrow parts are covered by simple sampling for the LRW.
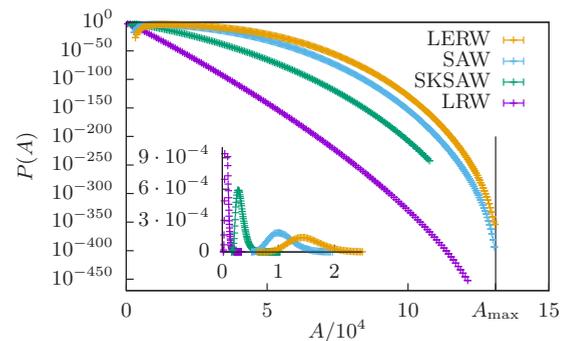

FIG. 5. Distribution of all scrutinized walk types with $T = 1024$ steps. The vertical line at $A_{\max} = 131\,072$ denotes the maximum area [Eq. (8)], i.e., SAW and LERW are sampled across their full support and SKSAW and LRW are not. The inset shows the peak region. The gap on the left is due to excluded volume effects, i.e., there are no configurations with area below some threshold, since this would require self-intersection.

This is supported from Ref. [23], which shows that the distribution $P(A)$ for standard random walks with Gaussian jumps, i.e., without lattice or fixed step length, do not bend down and have an exponential right tail. We conclude that the deviations from this are thus caused by this difference.

First we will look at the rescaled means $\mu_S = \langle S \rangle / T^{d_e \nu}$, where S is an observable and d_e its effective dimension, as introduced above in Eq. (8). The scaling is a combination of the scaling of the end-to-end distance $r \propto T^\nu$ and the typical scaling that a d-dimensional observable scales as r^d with a characteristic length r.

Nevertheless, due to finite-size corrections, the ratios $\mu_S = \langle S \rangle / T^{d_e \nu}$ will still depend on the walk length. Thus, the measured estimates $\mu_S = \mu_S(T)$ at specific walk lengths T need to be extrapolated to get an estimate of the asymptotic value $\mu_S^\infty = \lim_{T \to \infty} \mu_S(T)$. For the extrapolation we use [25]

$$\mu_S(T) = \mu_S^\infty + C_1 T^{-1/2} + C_2 T^{-1} + o(T^{-1}). \quad (10)$$

This choice is motivated by a large-T expansion for the area A ($d_e = 2$) of the convex hulls of standard random walks ($\nu = 1/2$) with Gaussian jumps [20],

$$\frac{\langle A \rangle}{T} = \frac{\pi}{2} + \gamma \sqrt{8\pi}\, T^{-1/2} + \pi(1/4 + \gamma^2)\, T^{-1} + o(T^{-1}), \quad (11)$$

where the constant $\gamma = \zeta(1/2)/\sqrt{2\pi} = -0.582\,59\ldots$. A natural guess for a generalization to observables of a different effective dimension d_e [25] and different walk types would be a similar behavior with different coefficients like Eq. (10).

Indeed, using this form to estimate the asymptotic means μ_S^∞ of the observable S yields good fits, as visible in Fig. 6. In fact, for the fit quality we obtain $\chi_{\rm red}^2$ values between 0.4 and 1.7. (The fit ranges for SKSAW begin at $T = 512$, and for LRW, SAW, and LERW at $T = 128$, hinting at more severe corrections to scaling for the former.) We assume that the scaling is thus valid for arbitrary random walk types. The resulting fit parameters are shown in Table I.

For standard random walks with Gaussian jumps the asymptotic means $\mu_{S,\rm Gaussian}^\infty$ are known [19]. These results can be used to predict the corresponding values for LRW. First consider the following heuristic argument for a $d = 2$ square lattice. On average a random walk takes the same amount of steps in x and y direction such that on average two steps
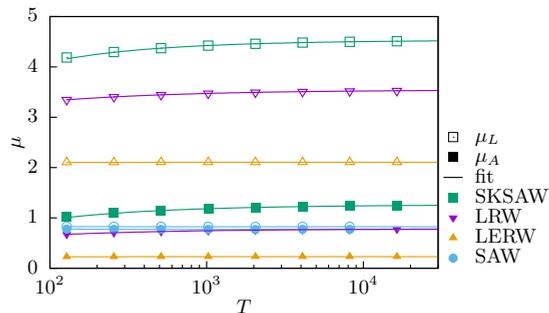

FIG. 6. Scaled means $\mu_A = \langle A \rangle / T^{2\nu}$ and $\mu_L = \langle L \rangle / T^\nu$ for different walk types. The lines are fits to extrapolate the asymptotic values shown in Table I according to Eq. (10). Error bars of the values are smaller than the line of the fit and are not shown for clarity.

TABLE I. Asymptotic mean values extrapolated from simulational data and the exactly known values for the standard random walk (LRW). The columns labeled with μ_L^∞ and μ_A^∞ are for $d = 2$, those labeled with $\mu_{\partial V}^\infty$ and μ_V^∞ are for $d = 3$. For $d = 3$ we did not simulate the SKSAW, see Sec. II C. Also, SAW has lower accuracy because of fewer samples in $d = 3$.

	μ_L^∞	μ_A^∞	$\mu_{\partial V}^\infty$	μ_V^∞
LRW (exact)	3.5449...	0.7854...	2.0944...	0.21440...
LRW	3.5441(7)	0.7852(2)	2.0945(4)	0.21445(4)
SKSAW	4.5355(12)	1.2642(5)		
SAW	0.8233(7)	0.7714(1)	2.070(2)	0.1998(2)
LERW	2.1060(3)	0.2300(1)	1.6436(2)	0.13908(3)

displace the walker by $\sqrt{2}$, i.e., the diagonal of a square. In contrast, a Gaussian walker with variance 1 will be displaced on average by 1 every step. To make both types comparable, we can increase the lattice constant to $\sqrt{2}$, which leads to an average displacement of 1 per step for the LRW. Using the same argumentation for higher dimensions, we can use the trivial scaling with the lattice constant S^{d_e} and the length of the diagonal of a unit hypercube $d^{1/2}$ to derive a general conversion:

$$\mu_{S,\rm LRW}^\infty = \mu_{S,\rm Gaussian}^\infty / d^{d_e/2}. \quad (12)$$

These known results are listed next to our measurements in Table I and are within error bars compatible with our measurements.

Since we have data for whole distributions, a natural question is whether this scaling does apply over the whole support of the distribution. There is evidence that this is true for the convex hulls of standard random walks [23] in arbitrary dimensions [25]. That means the distributions of an observable S for different walk lengths T should collapse onto one universal function

$$P(S) = T^{-d_e \nu} \widetilde{P}(S T^{-d_e \nu}). \quad (13)$$

Figure 7 shows the distributions of the $d = 2$ area of all considered random walk types scaled according to Eq. (13). The curves collapse well in the peak region and in the intermediate-right tail. In the far-right tail, clear deviations from a universal curve are obvious, which are the mentioned finite-size effects caused by the lattice.

The distributions look qualitatively similar, though with weaker finite-size effects, i.e., a better collapse, for the perimeter L (not shown). In $d = 3$, where we have studied the volume, the results also look similar but exhibit stronger finite-size effects (not shown).

Using the full distributions at different values of the walk length P_T, we can test if it obeys the large-deviation principle, i.e., if Φ exists, such that the distribution scales as

$$P_T \approx e^{-T\Phi} \quad (14)$$

for large values of T [28]. To simplify comparison, the support of the rate function is usually normalized to [0,1]. Here we achieve this by using the maximum Eq. (8). Solving Eq. (14) for Φ results in

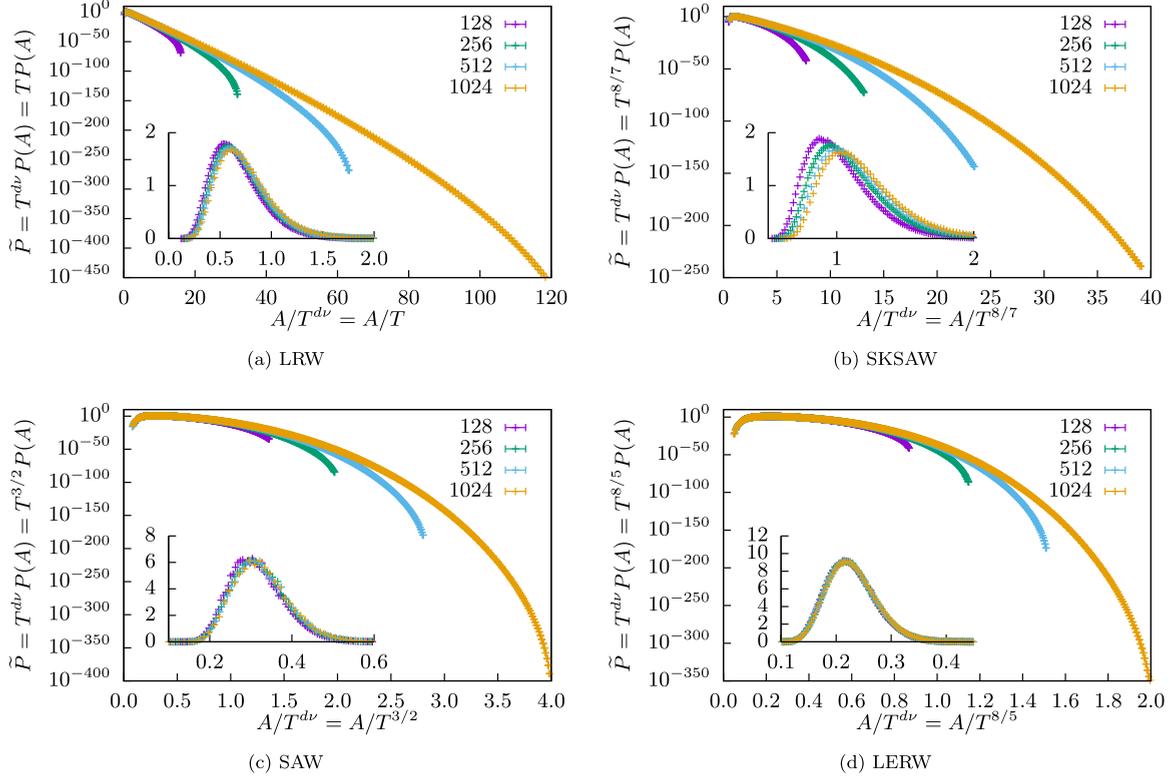$$\Phi(S/S_{\rm max}) = -\frac{1}{T} \ln P(S/S_{\rm max}). \quad (15)$$

FIG. 7. Distributions of the area of different types of random walks scaled according to Eq. (13) for different walk lengths T.

We plot this for a selection of our results in Fig. 8. From these plots, Φ seems to approximately follow a power law in the intermediate-right tail, while the finite-size effects caused by the lattice play a major role in the far-right tail, which consequently "bends up."

Assuming that the rate function behaves approximately as a power law, which seems consistent with our data shown in Fig. 8, i.e.,

$$\Phi(s) \propto s^{\kappa}, \tag{16}$$

the exponent κ can be estimated by combining the definition of Φ Eq. (14) with the scaling assumption Eq. (13) as follows; note that for clarity we use here $S_{\max} \propto T^{d_e}$:

$$\exp(-T\,\Phi(S/T^{d_e})) \approx \frac{1}{T^{\nu d_e}} \widetilde{P}(S/T^{\nu d_e}). \tag{17}$$

The $1/T^{\nu d_e}$ term on the right-hand side can be ignored next to the exponential. Since the right-hand side is a function of $S/T^{\nu d_e}$, the left-hand side must also be dependent only on $S/T^{\nu d_e}$. This can be achieved by assuming $-\nu d_e \kappa + d_e \kappa = 1$, as one can easily see by using Eq. (16):

Starting from the left-hand side

$$\exp(-T^1 \Phi(S/T^{d_e}))$$

$$\propto \exp(-T^1 (S/T^{d_e})^{\kappa})$$

$$= \exp(-T^{-\nu d_e \kappa + d_e \kappa}(S/T^{d_e})^{\kappa})$$

$$= \exp(-(S/T^{\nu d_e})^{\kappa}).$$

From this we can conclude

$$\kappa = \frac{1}{d_e(1-\nu)}, \tag{18}$$

which simplifies to the case of the standard random walk above the critical dimension of the given walk type [25],

$$\kappa_g = \frac{2}{d_e}.$$

To compare this crude estimate with the results of our simulations, we do a pointwise extrapolation of the empirical rate functions for fixed walk lengths T as done before in Refs. [23–25]. For the pointwise extrapolation, we use measurements Φ_T for multiple values of the walk length T at fixed values of $S/S_{\max}$. Since our data are discrete due to binning, the values of Φ_T are obtained by cubic spline interpolation. With these data points, which can be thought of as vertical slices through the plots of Fig. 8, we extrapolate the $T \to \infty$ case with a fit to a power law with offset

$$\Phi = aT^b + \Phi_{\infty}. \tag{19}$$

The extrapolated values are marked with black dots in Fig. 8. Since finite-size effects have a major impact on the tails due to the lattice structure, we expect that our estimate is only valid for the intermediate right tail of our simulational data. To estimate sensible uncertainties, we fit different ranges of our data and give the center of the range of the obtained κ as our estimate with an error including the extremes of the obtained κ. The black lines in Fig. 8 are our expected values, which are in all examples compatible with some range of our extrapolated data.
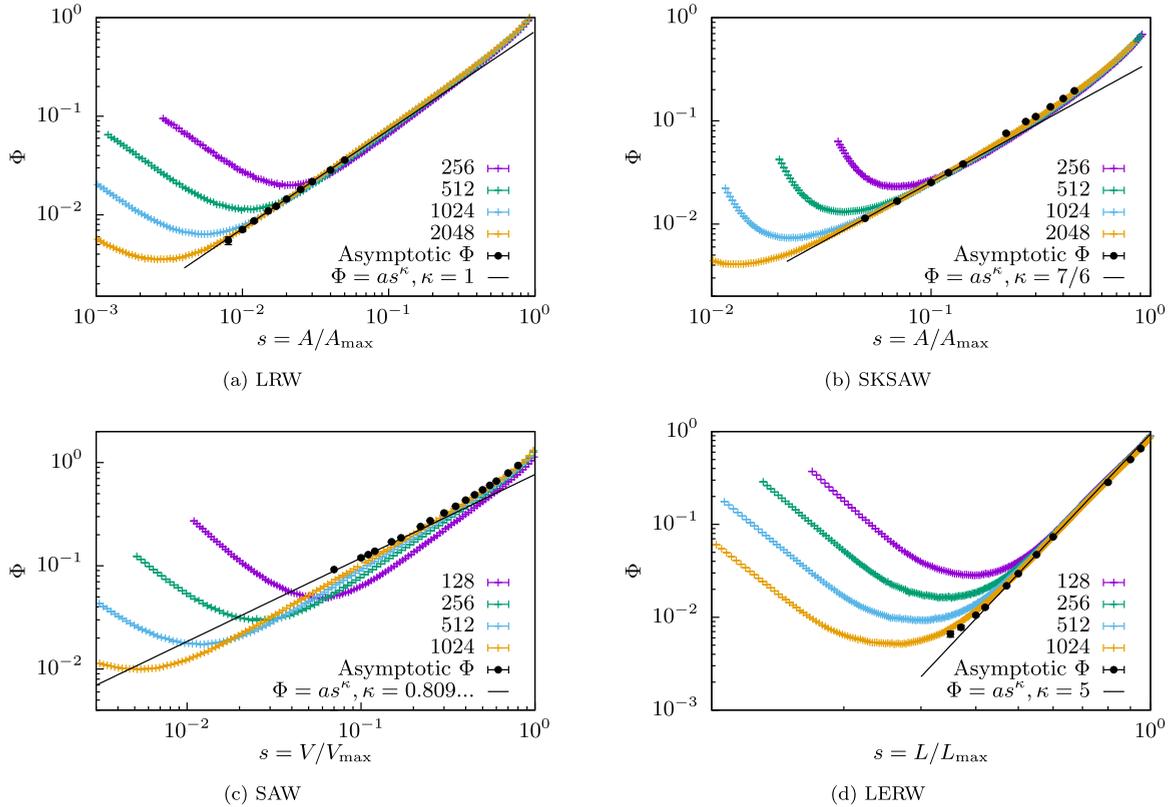
(a) LRW

(b) SKSAW

(c) SAW

(d) LERW

FIG. 8. Selection of asymptotic rate functions extrapolated from our data and our expected exponent κ of the rate function Φ.

All exponents κ we calculated, together with our expectations, are listed in Table II. A more detailed discussion of the examples shown in Fig. 8 follows.

In Fig. 8(a) the LRW is shown, which is equivalent to Brownian motion in the large-T limit for which Refs. [23] and [25] showed the rate function to behave like a power law with exponent $\kappa = 1$ for the area in $d = 2$. Using the above-mentioned procedure we obtain $\kappa = 0.99(2)$, which is in perfect agreement with the expectation $\kappa = 1$.

Figure 8(b) shows the same for the SKSAW. The obtained asymptotic rate function's exponent $\kappa = 1.28(12)$ is compatible with our expectation, though the stronger finite-size effects lead to larger uncertainties of our estimate.

Figure 8(c) shows the same but for the volume of the SAW in $d = 3$ dimensions. The finite-size effects are apparently

stronger for the volume in $d = 3$, as the slope of the right-tail rate function gets less steep with increasing system size.

Figure 8(d) shows the same for the perimeter of a $d = 2$ dimensional LERW. In contrast to the area and volume, the far-right tail of the perimeter seems to bend down instead of up, albeit slightly. Though in the intermediate right tail, the rate function seems to behave as expected.

In general, our data supports the convergence to a limiting rate function, which, mathematically speaking, means that the *large-deviation principle* holds. This means that the distributions are somehow well behaved and might be accessible to analytical calculations, although the estimate for what the rate function Φ actually is can possibly be improved. However, since our estimate for κ is always compatible with our measurements it appears plausible that also for interacting walks the distribution of the convex hulls is governed by the scaling behavior of the end-to-end distance, as given by the exponents ν.

TABLE II. Comparison of expected and measured rate function exponent κ. The value is the center of multiple fit ranges and the error is chosen such that the largest and the smallest result is enclosed.

	V		∂V	
	Eq. (18)	κ	Eq. (18)	κ
LRW	1	0.99(2)	2	
SKSAW	$\frac{7}{6}$	1.28(12)	$\frac{7}{3}$	
SAW	2	2.2(4)	4	4.11(14)
SAW $d = 3$	0.809...	0.92(11)	1.214...	
LERW	$\frac{5}{2}$	2.57(24)	5	4.82(19)
LERW $d = 3$	0.867...	0.89(9)	1.299...	

IV. CONCLUSIONS

We numerically studied the area and perimeter of the convex hulls of different types of self-avoiding random walks in the plane and to a lesser degree the volume of their convex hulls in $d = 3$ dimensional space. By applying sophisticated large-deviation algorithms, we calculated the full distributions, down to extremely small probabilities like 10^{-400}. We also obtained corresponding rate functions of these observables. Our data support a convergence of the rate functions, which means

the large-deviation principle seems to hold. We observed a generalized scaling behavior, which was before established for standard random walks. Thus, although the self-avoiding types of walk exhibit a more complicated behavior as compared to standard random lattice walks, and although the limiting scaled distributions of their convex hull's volume and surface look quite different for the various walk cases, in the end the convex hull behavior seems to be still governed by the single end-to-end distance scaling exponent ν.

We also observed, rather expectedly, that the two observables area and perimeter are highly correlated for small values. For large but not too large values of the perimeter, many different values of the area are possible but are statistically dominated by rather small values of the area. Extremely large values of the perimeter are only feasible with shrinking area.

Finally, we gave estimates for the large-T asymptotic mean values of the mentioned observables. These might be of interest for attempts to calculate these values analytically.

For future studies it could be interesting to look closer into the correlations between different observables that we briefly noted. For a more thorough study, it would be useful to obtain full two-dimensional histograms.

ACKNOWLEDGMENTS

[1] B. D. Hughes, *Random Walks and Random Environments* (Clarendon Press, Oxford, UK, 1996).

[2] N. Madras and G. Slade, Analysis of Monte Carlo methods, in *The Self-Avoiding Walk* (Springer, New York, 2013), pp. 281–364.

[3] A. Weinrib and S. A. Trugman, Phys. Rev. B **31**, 2993 (1985).

[4] S. S. Manna, D. Dhar, and S. N. Majumdar, Phys. Rev. A **46**, R4471 (1992).

[5] S. N. Majumdar, Phys. Rev. Lett. **68**, 2329 (1992).

[6] J. Cardy, Ann. Phys. **318**, 81 (2005).

[7] G. F. Lawler, O. Schramm, and W. Werner, arXiv:math/0204277.

[8] G. F. Lawler, O. Schramm, and W. Werner, Conformal invariance of planar loop-erased random walks and uniform spanning trees, in *Selected Works of Oded Schramm*, edited by I. Benjamini and O. Häggström (Springer, New York, 2011), pp. 931–987.

[9] T. Kennedy, J. Stat. Phys. **160**, 302 (2015).

[10] C. O. Mohr, American Midland Naturalist **37**, 223 (1947).

[11] B. J. Worton, Ecol. Model. **38**, 277 (1987).

[12] S. A. Boyle, W. C. Lourenco, L. R. da Silva, and A. T. Smith, Folia Primatol. **80**, 33 (2009).

[13] E. Dumonteil, S. N. Majumdar, A. Rosso, and A. Zoia, Proc. Natl. Acad. Sci. USA **110**, 4239 (2013).

[14] Y. Lanoiselée and D. S. Grebenkov, Phys. Rev. E **96**, 022144 (2017).

[15] L. T. Gérard Letac, Am. Math. Mon. **87**, 142 (1980).

[16] G. Letac, J. Theoret. Probab. **6**, 385 (1993).

[17] J. Randon-Furling, S. N. Majumdar, and A. Comtet, Phys. Rev. Lett. **103**, 140602 (2009).

[18] S. N. Majumdar, A. Comtet, and J. Randon-Furling, J. Stat. Phys. **138**, 955 (2010).

[19] R. Eldan, Electron. J. Probab. **19**, 1 (2014).

[20] D. S. Grebenkov, Y. Lanoiselée, and S. N. Majumdar, J. Stat. Mech. (2017) 103203.

[21] Z. Kabluchko and D. Zaporozhets, Trans. Am. Math. Soc. **368**, 8873 (2016).

[22] A. Goldman, Probab. Theory Relat. Fields **105**, 57 (1996).

[23] G. Claussen, A. K. Hartmann, and S. N. Majumdar, Phys. Rev. E **91**, 052104 (2015).

[24] T. Dewenter, G. Claussen, A. K. Hartmann, and S. N. Majumdar, Phys. Rev. E **94**, 052120 (2016).

[25] H. Schawe, A. K. Hartmann, and S. N. Majumdar, Phys. Rev. E **96**, 062101 (2017).

[26] A. K. Hartmann, P. L. Doussal, S. N. Majumdar, A. Rosso, and G. Schehr, Europhys. Lett. **121**, 67004 (2018).

[27] A. Akopyan and V. Vysotsky, arXiv:1606.07141.

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

[29] A. K. Hartmann, Phys. Rev. E **65**, 056102 (2002).

[30] A. K. Hartmann, Eur. Phys. J. B **84**, 627 (2011).

[31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[32] A. Engel, R. Monasson, and A. K. Hartmann, J. Stat. Phys. **117**, 387 (2004).

[33] A. K. Hartmann, Phys. Rev. Lett. **94**, 050601 (2005).

[34] S. Wolfsheimer, B. Burghardt, and A. K. Hartmann, Algorithms Mol. Biol. **2**, 9 (2007).

[35] P. Fieth and A. K. Hartmann, Phys. Rev. E **94**, 022127 (2016).

[36] A. K. Hartmann, Phys. Rev. E **89**, 052103 (2014).

[37] K. Kremer and J. W. Lyklema, Phys. Rev. Lett. **54**, 267 (1985).

[38] K. Suzuki, Bull. Chem. Soc. Jpn. **41**, 538 (1968).

[39] N. Clisby, Phys. Rev. Lett. **104**, 055702 (2010).

[40] N. Clisby, J. Stat. Phys. **140**, 349 (2010).

[41] G. F. Lawler, Duke Math. J. **47**, 655 (1980).

[42] A. J. Guttmann and R. J. Bursill, J. Stat. Phys. **59**, 1 (1990).

[43] D. B. Wilson, Phys. Rev. E **82**, 062102 (2010).

[44] K. Q. Brown, Inf. Process. Lett. **9**, 223 (1979).

[45] A. Andrew, Inf. Process. Lett. **9**, 216 (1979).

[46] A. Bykat, Inf. Process. Lett. **7**, 296 (1978).

[47] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, ACM Trans. Math. Software **22**, 469 (1996).

[48] P. Stein, Am. Math. Mon. **73**, 299 (1966).