
Polymers in Fractal Disorder

Von der Fakultät für Physik und Geowissenschaften
der Universität Leipzig
genehmigte DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt von Dipl.-Phys. Niklas Fricke
geboren am 09.01.1984 in Warburg.

Gutachter: Prof. Dr. Kurt Kremer
Prof. Dr. Wolfhard Janke

Tag der Verleihung: 23.05.2016

Contents

Preface	2
1. Introduction	11
2. Background and theory	15
2.1. Polymers and self-avoiding walks	15
2.1.1. Polymers in solution: universality and modeling	15
2.1.2. Scaling properties of the self-avoiding walk	16
2.1.3. Analytical approaches	17
2.2. Critical percolation clusters	19
2.2.1. The percolation transition	19
2.2.2. Fractal dimensions	20
2.2.3. Cluster structure: backbone and dangling-ends	21
2.2.4. Diffusion and conductance — dynamical exponents	24
2.3. Self-avoiding walks on percolation clusters	25
2.3.1. The model	25
2.3.2. Quenched vs. annealed average and other ambiguities	26
2.3.3. End-to-end distances and the exponent ν	27
2.3.4. The number of conformations Z	29
2.3.5. Open questions	32
3. The taming of the SAW	33
3.1. Existing numerical methods for SAWs on CPCs	33
3.1.1. Chain-growth Monte Carlo	33
3.1.2. Exact enumeration	36
3.2. Scale-free enumeration (SFE)	37
3.2.1. Factorizing the enumeration	38
3.2.2. Difficulties and optimizations	50
3.2.3. Mean end-to-end and chemical distances	52
3.2.4. Spatial distributions	53
3.2.5. Thermal averages	54
3.3. Partitioning the clusters	56
3.3.1. Evaluating cell hierarchies	57
3.3.2. Finding the right hierarchy	60
3.4. Checking and benchmarking the SFE method	66
3.4.1. Preliminary considerations	66
3.4.2. Performance of the SFE method	67

Contents

3.5. Analyzing PERM	72
3.6. Discussion: Scope and performance of SFE compared to other methods	77
4. Scaling behavior of self-avoiding walks on critical percolation clusters	79
4.1. End-to-end distances	79
4.1.1. Results in two dimensions:	80
4.1.2. Results in three dimensions	83
4.1.3. Results in four and five dimensions	86
4.1.4. Results in six and seven dimensions	89
4.1.5. Chemical distances	91
4.1.6. All cluster (AC) averages	93
4.1.7. Discussion	94
4.2. Number of conformations	96
4.2.1. Median, mean, deviations and bias	96
4.2.2. Probability-density distributions	99
4.2.3. Scaling behavior of the average entropy $[\ln Z_N]$	102
4.2.4. Log-normal approximation	104
4.2.5. All cluster (AC) averages	106
4.3. A closer look at individual clusters	109
4.3.1. Density distributions	111
4.3.2. Summary and Ideas	114
5. Variations of the model	117
5.1. Away from criticality	117
5.1.1. Above p_c : SAWs on supercritical clusters	117
5.2. Self-attracting self-avoiding walks	124
5.2.1. Full temperature range	125
5.2.2. Fixed temperatures	129
5.3. Kinetic-growth walks	133
5.4. Further ideas	135
5.4.1. Modifying the walks	135
5.4.2. Changing the medium	136
6. Conclusion	137
A. Generating and analyzing critical percolation clusters	141
A.1. Setup	141
A.1.1. Creating the clusters	142
A.2. Estimating the fractal dimensions d_f and d_{\min}	142
A.3. Extracting the backbone and measuring d_{BB}	144

List of Figures

2.1. Percolation clusters below, at, and above the percolation threshold . . .	20
2.2. Backbone of a 3D cluster wrapping around the lattice.	22
2.3. Bi-connected components of critical percolation clusters (CPCs).	23
2.4. Conformation of a self-avoiding walk (SAW) on a CPC	25
3.1. Real-space renormalization on the Sierpinski gasket	38
3.2. Partitioning of a 2D CPC into cell hierarchy	39
3.3. Possible states for a cell with three links	41
3.4. Maximum number of states vs. number of links of a cell	42
3.5. State transitions for a cell with three links	43
3.6. Snapshot of the enumeration procedure	45
3.7. Pseudo code of the enumeration routine for normal cluster sites	46
3.8. Pseudo code of the enumeration routine for decimated <i>child sites</i>	47
3.9. Pseudo code of the enumeration routine for the external <i>parent site</i>	48
3.10. Conformation of a self-attracting SAW a 2D CPC	56
3.11. Average enumeration time as function of the number of <i>faces</i> of a cell	58
3.12. Average time as function of the maximum number of state combinations	59
3.13. Primary pieces and connectivity graph for the cluster from Fig. 3.2	61
3.14. Active pieces during the amalgamation process	62
3.15. Amalgamation procedure for a backbone of 500 chemical shells	62
3.16. Runtime of the scale-free enumeration (SFE) program vs. number of steps N	69
3.17. Peak memory usage of the SFE method	69
3.18. Bi-connected components for percolation clusters at varying concentration	71
3.19. Runtime for the SFE method on supercritical clusters	71
3.20. PERM estimates for the mean squared end-to-end distance of SAWs on a CPC	73
3.21. PERM estimates for the end-point densities on a CPC	73
3.22. PERM estimates for the conformation densities on a CPC	74
3.23. Convergence behavior of PERM estimates on CPCs.	75
3.24. Average bias of PERM estimates for the squared end-to-end distance	75
3.25. Bias of the PERM estimator for the quenched averages of the end-to-end distances	76
3.26. Distribution of number of PERM tours needed until convergence	77
4.1. Quenched average of the mean squared end-to-end distance vs. number of SAW steps N on CPCs in $D = 2 - 7$	80

List of Figures

4.2. Mean squared end-to-end distances vs. N on $2D$ CPCs and backbones	81
4.3. Estimates of the scaling exponent ν_{2D} as function of the fit range	82
4.4. Estimates of ν from successive slopes	83
4.5. Mean squared end-to-end distances vs. N on $3D$ CPCs and backbones	85
4.6. Initial behavior of the mean squared end-to-end distances for $3D$ SAWs on CPCs	85
4.7. Estimates for the exponent ν on $3D$ CPCs as function of the fit range	86
4.8. Mean squared end-to-end distances vs number of SAW steps on CPCs in $4D$ and $5D$	87
4.9. Estimates for ν on CPCs in $4D$ and $5D$ as function of the fit range	88
4.10. Mean squared end-to-end distances vs. N steps on CPCs in $4D$ and $5D$	89
4.11. Estimates for ν on CPCs in $6D$ and $7D$ as function of the fit range	90
4.12. Mean chemical distances for SAWs on CPCs and backbones in $D = 2, 3$	92
4.13. Estimates for the chemical-distance exponent ν_l on CPCs and backbones in $2D$ and $3D$ as function of the fit range	92
4.14. All cluster (AC) average of the mean squared end-to-end distances for SAWs on $2D$ CPCs.	93
4.15. Numerical estimates for ν on CPCs and backbones in different dimensions compared to analytical predictions	95
4.16. Maximum, mean, and median value of the number of SAW conformations Z_N as function of the size of a random sample of $3D$ CPCs	97
4.17. Maximum, mean, and median of Z_N vs. N for a random sample of $3D$ CPCs	98
4.18. Probability densities of the average entropy $\ln Z_N$ of SAWs on CPCs and backbones in $2D$ and $3D$	100
4.19. Skewness and excess kurtosis of the probability densities of $\ln Z_N$ in $2D$ and $3D$	100
4.20. Variances of the distribution of $\ln Z_N$ on CPCs and backbones in $D = 2-7$	101
4.21. Average entropy per SAW step on CPCs and backbones in $2D$ and $3D$ fitted by two different approaches.	103
4.22. Logarithm of the measured mean number of conformations, mean entropy, and log-normal approximation for SAWs on CPCs and backbones in $D = 2, 3$	105
4.23. Fraction of $2D$ percolation clusters that support at least one SAW conformation as function of the number of steps	106
4.24. All cluster (AC) average of the mean number of SAW conformation on percolation clusters compared to exact predictions	107
4.25. AC average of the mean entropy per SAW step on $2D$ fitted by two different approaches	108
4.26. Variance of distribution of the entropy for SAWs on percolation clusters in the AC ensemble	109
4.27. Mean squared end-to-end distances and excess entropies for SAW on single $2D$ CPCs	110
4.28. Mean squared end-to-end distances and excess entropies for SAW on single $3D$ CPCs	111

4.29. Mean squared end-to-end distance and excess entropy as functions of N for some exemplary $2D$ clusters	112
4.30. Distributions of SAW end-points and conformation densities after varying numbers of steps on a $2D$ CPC	113
4.31. Distributions of SAW end-points and conformation densities after 1600 steps on some $2D$ CPCs	114
5.1. Mean squared end-to-end distance for SAWs on $2D$ percolation clusters above the percolation threshold	118
5.2. Mean squared end-to-end distance for SAWs on $3D$ percolation clusters above the percolation threshold	119
5.3. Distribution of end-point locations for SAWs on $2D$ percolation clusters at different concentrations	122
5.4. Conformation densities for SAWs on $2D$ supercritical percolation clusters at different concentrations	123
5.5. Average energy per monomer vs. temperature for SASAWs of different length on CPCs	126
5.6. Average specific heat vs. temperature for SASAWs of different length on CPCs	127
5.7. Ratio of the mean squared end-to-end distances of SASAWs and athermal SAWs on CPCs vs. temperature	127
5.8. Thermal derivative of the mean squared end-to-end distance for SASAWs on CPCs	128
5.9. Temperature dependence of the specific heat C_V on single CPCs	128
5.10. Temperature dependence of the mean squared end-to-end distance on single CPCs	129
5.11. Distribution of the locations of specific heat maxima	129
5.12. Mean squared end-to-end distances for SASAWs on CPCs at different temperatures as functions of the number of steps	130
5.13. Rescaled mean squared end-to-end distances for SASAWs on CPCs at different temperatures as functions of the number of steps	130
5.14. Rescaling of the mean squared end-to-end distances as function of the number of steps for SASAWs at different temperatures on single CPCs	132
5.15. Mean squared end-to-end distance for kinetic growth walks on CPCs in $2D$ and $3D$	134
A.1. Average mass vs lattice extension for CPCs in different dimensions	143
A.2. Average chemical distance on CPCs vs. lattice extension.	144
A.3. Backbone mass vs. lattice size	145

List of Tables

2.1. Percolation threshold for hypercubic lattices	19
2.2. Fractal dimensions of critical percolation clusters	22
2.3. Backbone dimension of critical percolation clusters	23
2.4. Walk dimension of critical percolation clusters and backbones	24
3.1. Number of SAW segments in cell E of the example from Fig. 3.2	40
3.2. Number of SAW segments in cell D of the example from Fig. 3.2	41
3.3. Transition matrices for a cell with three links	43
3.4. Number of SAW segments for all states of cell C from Fig. 3.2	44
3.5. Number of SAW segments within cell B from Fig. 3.2	49
3.6. Number of SAW segments within the root cell A from Fig. 3.2	50
3.7. Exponents for time and memory increase of the SFE method	70
3.8. Length limits for the SFE method at varying concentration	71
3.9. Fraction of percolation clusters where PERM estimators did not converge	76
4.1. Estimates of the scaling exponent ν_{2D} as function of the fit range	82
4.2. Estimates for ν on $3D$ CPCs and backbones	86
4.3. Estimates for ν on CPCs and backbones in $4D$ and $5D$	88
4.4. Estimates for ν on CPCs and backbones in $4D$ and $5D$	90
4.5. Chemical-distance exponent ν_l on CPCs and backbones in $D = 2, 3$	92
4.6. Numerical estimates for ν on CPCs and backbones in different dimen- sions compared to analytical predictions	94
4.7. Exponents and amplitudes describing the scaling of the variances of $\ln Z_N$ with N	102
4.8. Estimates for various parameters describing the scaling of the average entropy	104
4.9. Estimates for the connectivity constant of SAWs on CPCs and backbones	105
5.1. Estimates for the exponent ν on supercritical $2D$ clusters	120
5.2. Estimates for the exponent ν on supercritical $3D$ clusters	120
5.3. Estimates Θ -temperatures and critical scaling exponents ν^Θ for SA- SAWs on CPCs found in the literature	125
5.4. Estimates for the exponent ν for SASAWs at different temperatures on CPCs in $2D$ (top) and $3D$ (bottom)	131
A.1. Lattice extensions L used for different dimensions and numbers of steps.	141
A.2. Estimates for the fractal dimension d_f of CPCs.	143

List of Tables

A.3. Estimates for the shortest-path dimension d_{\min} of CPCs. 144
A.4. Estimates for the backbone dimension d_{BB} 145

1. Introduction

Polymers are ubiquitous in our world. From synthetic compounds such as plastics to the very building block of life: fibers, proteins, and DNA — these long, chain-like macromolecules feature in a variety of roles. And yet it was relatively recently in our history that we became aware of their existence. In 1920 Hermann Staudinger proposed the structure of small repeating units (*monomers*) linked by covalent bonds to explain the high molecular weight measured for certain materials, and, after some initial skepticism, these ideas became generally accepted.

Today, our understanding of polymers is still very limited. They are usually much too complex, with too many degrees of freedom and different kinds of interactions. In fact, there is little chance that we will ever develop a general theory able to properly predict, say, the native structure of a protein or by which pathways it will fold. Studying the effects of specific configurations of chemically active groups may be of some help in that regard, but such an approach is not very satisfying from a theoretical-physics point of view as it does not yield a deeper understanding. A more physical way of approaching polymers is to identify what aspects can be generalized and investigate these by means of simplified models. In other words, we concentrate on those properties that arise solely from the characteristic geometrical structure and are independent of chemical details. This field of theoretical polymer physics has been pioneered by Paul Flory in the 1930s, who realized that some universal properties of flexible polymers in good solvent condition can be explained by applying Kuhn's concept of excluded volume to the random walk process. This *self-avoiding walk* (SAW) is today one of the most successful models in statistical physics. When fitted with additional nearest-neighbor attractions, it even captures the behavior of so-called Θ -polymers, which collapse to a globular state at low temperatures. Our understanding of the SAW has deepened substantially since de Gennes' discovery of its fundamental relationship to spin systems and the theory of critical phenomena in 1972. Not only did this intriguing insight consolidate the importance of the SAW model in physics, it also opened it up to the newly emerged machinery of field-theoretical renormalization group theory. This led to a revolution in the field, driven most prominently by the works of Edwards, Zinn-Justin, and Des Cloizeaux. Besides theoretical advances, drastic improvements of computational capabilities have also greatly furthered our knowledge of the SAW and similar models. These improvements consisted in massive increases in computation power on the one hand and evermore sophisticated algorithms on the other. On the whole, the SAW model is today reasonably well understood — though mathematically rigorous results remain elusive.

In reality, polymers are not always immersed in a perfectly homogeneous solvent; their environment is often highly irregular across many length scales. The interior of a biological cell, for instance, is teeming with a variety of macromolecules and or-

1. Introduction

ganelles of very different shapes and sizes. The paradigm for studying the influence of such structural disorder on polymer conformations in a controlled manner is the model of a self-avoiding walk on a lattice with randomly placed defects. Most intriguing is the situation where the concentration of sites is equal to the percolation threshold of the lattice, i.e., when a macroscopically large cluster of non-defect sites can just barely exist. Such a *critical percolation cluster* is a fractal object, thus representing the limiting case of disorder on any length scale. The model of SAWs on critical percolation clusters may seem somewhat artificial, but it continues the spirit of approaching reality from a simplistic and well-defined limiting case. As a combination of two of the most fundamental models in statistical physics, this interplay of fractals is also of considerable academic appeal. It has therefore been studied extensively in the past. A variety of tools, both analytical and numerical, have been employed: analytical approaches included mean-field (“Flory”) approximations, real-space renormalization-group methods, and perturbative field theory, while numerical studies used exact enumeration as well as various Monte Carlo methods. However, the model proved much more challenging than the pure SAW, and many of its aspects are still today far from clear. Analytical investigations have remained largely qualitative, often relying on uncontrolled or heuristic approximations, and have yielded conflicting results. Numerical approaches, on the other hand, could only handle relatively small systems, where the asymptotic behavior is obscured by finite-size effects. This was because Monte Carlo simulations of SAWs tend to struggle with the strong confinement presented by the medium, while the computational cost for exact enumeration generally increases exponentially with the length of the walks.

Fortunately, however, the fractal structure of a critical percolation cluster lends itself to a hierarchical factorization of the problem, and it is thus possible to exactly enumerate SAWs very efficiently. As part of my diploma thesis, I had presented a proof of concept for this idea, which has by now matured into a working method. It allows complete enumeration of SAWs of well over 10^4 steps on critical percolation clusters of different dimensionality, beating even the most efficient Monte Carlo methods.

Equipped with this new tool, I could look at the problem of self-avoiding walks on critical percolation clusters in unprecedented detail. Besides the presentation and discussion of the methods itself, the results of these investigations constitute the main part of this work. More specifically, the remainder of the thesis is organized as follows:

- Chapter 2 aims to provide an overview of the relevant theoretical concepts. It also acquaints the reader with the model that is the main subject of this thesis and lays out the current state of research. After a quick dive into polymer physics with special focus on the self-avoiding walk model, I turn to the theory of percolation and discuss the fractal geometry of critical percolation clusters. Finally the model of SAWs on critical percolation clusters is introduced and defined, and the existing knowledge relevant for my work is briefly reviewed.
- Chapter 3 begins with a short review of existing numerical methods for studying SAWs on critical percolation clusters, as far as they are relevant for this thesis. Its main part consists of a detailed elaboration of the newly developed “scale-free

enumeration” (SFE) method and its variants: first the key ideas are outlined, then the actual enumeration algorithm is explained, and finally a description is given of how the clusters are partitioned into a hierarchy of nested “cells” as is required to factorize the enumeration procedure. This is followed by a discussion of the method’s performance, scope, and limitations, substantiated through benchmark results. In addition, a quantitative comparison with the most capable Monte Carlo method, the pruned-enriched Rosenbluth method (PERM), under different conditions is presented. This analysis also served to test the capacity and reliability of PERM, which was used to produce some of the results for Chapter 5.

- The main results, concerning the asymptotic scaling behavior of SAWs on critical percolation clusters, are presented in Chapter 4. Most notably, they comprise accurate estimates of the scaling exponent of the end-to-end distance on critical percolation clusters and their backbones in dimensions two to seven, as well as empirical evidence for unexpected scaling behavior of the average entropy and number of SAW conformations. This second aspect involves an analysis of the distribution of the numbers of conformations, which exhibits extreme-value statistics. Finally, a close-up look at the spatial distribution of walks on a few individual clusters is presented, supplying qualitative clues to help explain the observed scaling behavior. The chapter also encompasses comparison with numerical results and analytical predictions taken from previous studies.
- In Chapter 5, the model is expanded in different directions.
Using both PERM and the SFE method, I investigated SAWs on percolation clusters at concentrations above the critical threshold. This was done in order to test the predictions of the Meir-Harris model, which is the widely accepted theory for SAWs in disorder. The rest of the chapter is dedicated to variations of the SAW, namely, importantly SAWs with nearest-neighbor interactions and kinetic growth walks.
- In Chapter 6, I summarize the main findings and conclusions of the thesis and discuss perspectives and challenges for future investigations.
- The Appendix 6 outlines the creation and analysis of the critical clusters. It furthermore contains technical details which needed to be included for the sake of reproducibility but would have impaired the readability.

2. Background and theory

This chapter gives a more detailed introduction to the topic, outlines the relevant theoretical concepts, and describes the history and state of research of the main subject. In addition, it serves to introduce most of the definitions and nomenclature that I use. Apart from some comment and the argumentation in Section 2.3, the chapter does not represent original research of mine but is simply a compilation of theories and results that are relevant for this thesis.

2.1. Polymers and self-avoiding walks

Polymers are macromolecules consisting of repeating units called monomers. These are typically made up of a small number of atoms and are connected by covalent bonds. The number of monomers in a polymer, also known as the *degree of polymerization*, is typically denoted by N . For many common polymers, it can be impressively large, e.g., up to $N \approx 10^6$ for linear polyethylene (HDPE) and $N \approx 10^8$ for human DNA. One commonly distinguishes between homopolymers, for which the elemental units are identical, and heteropolymers, for which they vary. Many synthetic polymers, such as polyethylene, are in the first category, while most biopolymers, notably DNA and proteins, are in the second. However, this distinction is not important here as the microstructure shall be neglected anyway. Another classifier is a polymer's architecture: in general, polymers come in a number of different shapes such as branches, combs, rings or stars. Here I shall only be concerned with the simplest case of a linear chain. More comprehensive discussions of polymer statistics can be found, e.g., in Refs. [1–3].

2.1.1. Polymers in solution: universality and modeling

Monomers from different parts of a polymer molecule (and from different molecules) attract one another by electro-static or van-der-Waals forces, depending on the specific microstructures. In a good solvent, however, these interactions are screened, simplifying the problem and allowing for a more generic description. Another simplification and generalization is achieved through *coarse-graining*: Polymers differ in their stiffness as the bonds between monomers have different capacities to allow for rotation or bending. This is quantified by the *persistence length* l_p , the length scale of the decay of bond-angle correlations along the chain. By considering segments of length $2l_p$, called the Kuhn length, instead of the elementary chemical units, we obtain a chain of monomers that is effectively *freely-jointed*. The only remaining free parameter in this description is the chain length in units of Kuhn lengths (also typically denoted by N). Finally, by considering the *thermodynamic limit*, $N \rightarrow \infty$, even this last parameter

2. Background and theory

is eliminated. Properties that characterize a polymer in such a generic description then apply to a range of different systems, a phenomenon called *universality*. The most important example of such a universal quantity for polymers is the exponent ν , which describes how the average size (end-to-end distance or radius of gyration) asymptotically grows with N .

To study universal quantities of polymers, we can thus use a simplistic model system that represents the respective *universality class*. The simplest model is the (*discrete*) *ideal chain*, where the polymer is represented as a *random walk* on a regular lattice. The root-mean-square (RMS) end-to-end distance of a random walk increases with the square root of the number of steps:

$$\sqrt{\langle R_N^2 \rangle} \sim N^\nu = N^{1/2}, \quad (2.1)$$

and the probability distribution of the distances is Gaussian:

$$P_N(\vec{R}) \sim \left(\frac{3}{2\pi N}\right)^{3/2} \exp\left(-\frac{3\vec{R}^2}{2N}\right). \quad (2.2)$$

The number of possible conformations Z , by contrast, increases exponentially:

$$Z_N = \mu^N. \quad (2.3)$$

The *connectivity constant* μ is *not* universal but equal to the coordination number of the lattice. Its meaning cannot be directly translated to continuous systems.

The *ideal chain* is a useful test-tube model, and it even correctly describes the behavior of flexible polymers in very dense solutions. In dilute solutions, where polymers are largely on their own and each monomer will be almost exclusively surrounded by solvent molecules, another model applies: the *self-avoiding walk* (SAW). One can think of a SAW as a random walk that can visit each position only once. However, the picture of walker taking successive steps in random directions is slightly misleading in the context of polymers as this would lead to different statistical weights for different conformations. As explained later on (Section 5.3), this would give the so-called kinetic growth walk (KGW). For the standard SAW, by contrast, all conformations are equally likely.

2.1.2. Scaling properties of the self-avoiding walk

Despite its simplicity, relatively little is rigorously known about the SAW. (The mathematical theory of SAWs is discussed in Ref. [4]) We know that a scaling law of the form of Eq. (2.1) applies for its end-to-end distance, but not even the “obvious” bounds $1/2 \leq \nu^{\text{SAW}} < 1$ are rigorously proven in $D = 2-4$. We furthermore know that the number Z_N of SAW conformations asymptotically also increases exponentially, i.e., that the limit

$$\lim_{N \rightarrow \infty} Z_N^{1/N} := \mu \quad (2.4)$$

2.1. Polymers and self-avoiding walks

exists. Its exact value is only known for the honeycomb lattice, but relatively tight bounds have been derived for hypercubic lattices. The asymptote μ^N is approached from above by Z_N , with

$$\mu^N \leq Z_N \leq \begin{cases} \mu^N e^{KN^{1/2}}, & D = 2 \\ \mu^N e^{KN^{2/(D+2)}}, & D = 3, 4. \end{cases} \quad (2.5)$$

Much knowledge about SAWs stems from non-rigorous sources such as conformal field theory, renormalization-group analysis, and numerical simulations. In the physics community, there is in fact very little doubt that the scaling of the number of conformations is described by

$$Z_N \sim \mu^N N^{\gamma-1}, \quad (2.6)$$

with a universal parameter γ called the *enhancement exponent*. The value of γ is strongly suspected to be $\frac{43}{32}$ in $2D$, while the best estimate in $3D$ is currently $\gamma = 1.15698(34)$ [5]. The exponent describing the scaling of the root-mean-square end-to-end distance,

$$\sqrt{\langle R_N^2 \rangle} \sim N^\nu, \quad (2.7)$$

is in $2D$ exactly given by the Flory value of $\nu = 3/4$ (see next section), and the most accurate numerical $3D$ estimate is $\nu = 0.587597(7)$ [6].

In $D \geq 4$, self-avoidance is negligible and the SAW asymptotically behaves like the ideal chain, i.e., $\nu = 1/2$ and $\gamma = 1$. However, in $D = 4$, the *upper critical dimension* of the system, there are still strong logarithmic corrections which can obscure this behavior to some extent.

2.1.3. Analytical approaches

A simple way to approximate the exponent ν is given by the Flory theory [7], which is based on minimizing a mean-field estimate of the free energy $F_N(R)$. Ignoring correlations, we assume that the self-avoidance results in an effective repulsion energy that increases with the squared density times the volume:

$$E_N(R) \sim V\rho^2 \sim \frac{N^2}{R^D}. \quad (2.8)$$

To estimate the entropy, we use the radial distribution of the (non-self-avoiding) random walk given by Eq. (2.2) integrated over the surface:

$$S_N(R) = \ln Z_N(R) \approx \ln(Z_N P_N(R)) \sim \frac{R^2}{N} - (D-1) \ln R + C_N, \quad (2.9)$$

where the term C_N is independent of R . In leading order of R , we hence get

$$F_N(R) = E - S \sim \frac{N^2}{R^D} + \frac{R^2}{N}. \quad (2.10)$$

2. Background and theory

Plugging in $R \sim N^\nu$ and minimizing with respect to ν then yields the famous Flory formula:

$$\nu = \frac{3}{D+2}. \quad (2.11)$$

This prediction is surprisingly accurate: the $2D$ value of $\nu = 4/3$ is in fact exactly right (though this has not yet been proven rigorously), and numerical estimates proved the $3D$ value to be less than 1% off the mark. The Flory formula even gives the correct result for $D = 4$, even though the above derivation is actually only valid for $D < 4$, where the terms on the right-hand side of Eq. (2.10) are dominant.

As was first realized by de Gennes [8], there is an interesting connection between the SAW and the $O(n)$ -model for spin systems [9] (for details see for instance Refs. [1, 2, 10, 11]). The $O(n)$ -model, which includes the Ising model ($n = 1$), the xy -model ($n = 2$), and the Heisenberg model ($n = 3$), is defined by the partition function

$$Z_{O(n)} = \text{Tr} \prod_{\langle i,j \rangle} e^{K \vec{s}_i \cdot \vec{s}_j}, \quad (2.12)$$

where $K = -J/(k_B T)$ is a reduced coupling constant and \vec{s}_i, \vec{s}_j are vectors on an n -dimensional unit sphere that are nearest neighbors in a lattice. Using a high-temperature expansion and taking the limit $n \rightarrow 0$ (which does not seem to have a geometrical interpretation), one obtains for the spin-spin correlation function:

$$G_{O(n \rightarrow 0)}(k, l) = \sum_{k \rightsquigarrow l} z^N, \quad (2.13)$$

where $z = \tanh K$ and the sum goes over all SAWs connecting sites k and l (N is the respective number of steps). Summing over all end points l then reveals that the magnetic susceptibility of the $O(n \rightarrow 0)$ -model, $\chi_{O(n \rightarrow 0)}$, corresponds to the grand canonical partition function of the self-avoiding walk:

$$\chi_{O(n \rightarrow 0)} = \sum_l G_{O(n \rightarrow 0)}(k, l) = \sum_{k \rightsquigarrow} z^N = \sum_N Z_N z^N = \mathcal{Z}_{SAW}. \quad (2.14)$$

$\mathcal{Z}_{SAW}(z)$ is the generating function for the series of the numbers of conformations Z_N . According to Eq. (2.6), it diverges at $z_c = \mu^{-1}$ as

$$\lim_{z \rightarrow z_c} \mathcal{Z}_{SAW}(z) \sim (z - \mu^{-1})^{-\gamma}. \quad (2.15)$$

The *critical fugacity* z_c corresponds to a critical (temperature) point in the $O(n \rightarrow 0)$ -model, where γ characterizes the divergence of the magnetic susceptibility. Correspondences between other exponents can also be established: the critical exponent α of the specific-heat divergence is connected to the number of self-avoiding polygons (SAPs) (which are SAWs in the limit $R \rightarrow 0$), and the exponent ν in Eq. (2.7) is indeed the critical exponent for the correlation length in the $O(n \rightarrow 0)$ -model:

$$\xi_{O(n \rightarrow 0)} \sim |T - T_c|^{-\nu}. \quad (2.16)$$

This relationship between two seemingly very different models has been extremely useful and allowed much of the extensive knowledge and methodology developed for spin systems to be carried over to SAWs. In particular, it allowed analyzing the SAW by means of perturbative renormalization-group methods [11].

Various heuristic approaches to the SAW have also been undertaken, which shall not be discussed here. They involve real-space renormalization [12] and transfer-matrix [13] techniques as well as studies of related models that are easier to calculate such as directed self-avoiding walks (DSAWs) [14] or SAWs on finitely-ramified deterministic fractals; see Section 2.3.3.

2.2. Critical percolation clusters

Percolation is probably the simplest model in statistical physics to exhibit interesting critical behavior. I shall discuss the topic only superficially, focusing on aspects that are relevant here. For a proper treatment, see, for instance, Refs. [15–17].

2.2.1. The percolation transition

If sites (or bonds) of a lattice are activated at random, there will be a point where a cluster of active sites spans across the whole system. If the lattice is infinitely large, this will occur at a sharply defined concentration p_c , the *percolation threshold*. The value of p_c depends on the lattice type and is only for a few cases exactly known. They include the bond-diluted square lattice and the site-diluted triangular lattice but none of the site-diluted square or (hyper-) cubic lattices, which are used in this work. However, accurate numerical estimates for p_c are available for these cases. The values I used are listed in Table 2.1.

Table 2.1.: Values for the percolation thresholds on hypercubic lattices used in this work.

D	p_c
2	0.59274621(13) [18]
3	0.31160768(15) [19]
4	0.1968861(14) [20]
5	0.1407966(15) [20]
6	0.109017(2) [20]
7	0.0889511(9) [20]

When the threshold p_c is crossed, the system undergoes a second-order phase transition, characterized by the continuous emergence of an order parameter, the fraction of sites in the largest cluster, $P(p) \sim |p - p_c|^{-\beta}$, and a diverging correlation length: $\xi \sim |p_c - p|^{-\nu}$. The critical exponents β and ν are universal, i.e., their values do not depend on details such as the lattice type. To visualize this transition, clusters of active sites are shown in Fig. 2.1 for three different concentrations: below, at (or very

2. Background and theory

close to), and above p_c . Directly at the threshold, the system is scale invariant and clusters exist of all sizes. More precisely, the probability distribution of cluster sizes s per site (*cluster numbers*) is described by a power-law:

$$n_s \sim s^{-\tau}, \quad (2.17)$$

where τ is the so-called Fisher exponent. These *critical percolation clusters* (CPCs), if sufficiently large compared to the lattice constant, are themselves scale-invariant (self-similar) objects, whose geometry is not properly described by Euclidean dimensions. Conceptually, one usually considers critical clusters of infinite size even though strictly speaking every individual cluster is still finite at p_c (a truly infinite cluster only emerges for $p > p_c$). This is justified because of the scale-free distribution of cluster sizes, which implies that the clusters can be arbitrarily large.

In numerical studies lattices are necessarily finite, so one often takes the largest cluster to represent an “infinite” cluster. To get a bit “closer to infinity”, one can furthermore only consider clusters that percolate the lattice, e.g., by connecting to opposite edges of the lattice or by *wrapping* around it (assuming periodic boundary conditions). Depending on the context, I shall use the term *incipient cluster*, *percolating cluster* or *critical cluster* for these. In this work the most generous version of the *wrapping criterion* is used to define percolation: a cluster must close back on itself in at least one dimension. The wrapping criterion produces less finite-size effects than the spanning criterion [18], and it also has the benefit of producing statistically homogeneous clusters without boundary distortions.

2.2.2. Fractal dimensions

The mass (or volume) M of any regular object scales with its linear extension (or radius) L as $M \sim L^D$, where D is an integer number: the Euclidean dimension. For critical percolation clusters, the exponent is non-integer, in general not even rational.

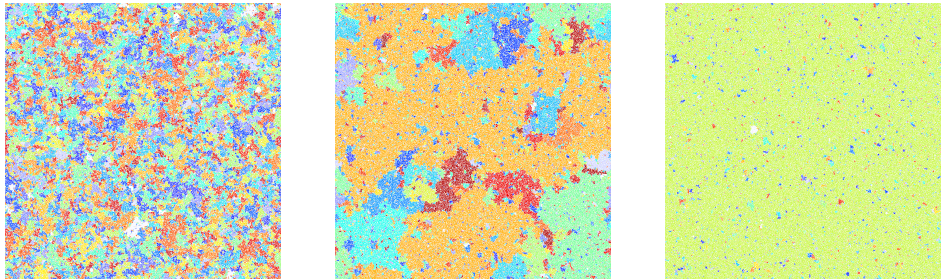


Figure 2.1.: Clusters of connected active sites (randomly colored) at three different concentrations: $p = 0.55$, $p = 0.592746 \approx p_c$, and $p = 0.65$ (from left to right).

2.2. Critical percolation clusters

However, it is a well-defined number, commonly denoted by d_f . It is the most important example of a fractal dimension and is usually (and here) simply called *the* fractal dimension. As a definition I use the size-mass relation,

$$M \sim L^{d_f}. \quad (2.18)$$

The fractal dimension can be defined more rigorously as the Hausdorff-Besicovitch dimension [21], but Eq. (2.18) is sufficient here. Note that SAWs are themselves also fractals, and that Eq. (2.7) corresponds to Eq. (2.18) with $d_f = 1/\nu$.

Like the critical exponents, the fractal dimension of the critical clusters is universal, i.e., independent of the lattice type. In fact, d_f can be regarded as a geometrical (“static”) critical exponent, and it is linked to the “thermal” exponents via the hyper-scaling relation

$$d_f = D - \beta/\nu, \quad (2.19)$$

where D is the Euclidean dimension of the lattice and β and γ characterize the divergence of the order parameter and the correlation length, respectively. For the $2D$ case, the thermal exponents have been determined exactly [22–24] yielding $d_f = 91/48 \approx 1.896$. These methods do not work for $D > 2$, where one has to resort to numerical estimation instead. The best current estimates for d_f in dimensions $D = 2-6$ are listed in Table 2.2, alongside estimates of other fractal dimensions introduced below. Some of the values for higher dimensions are my own results. They were obtained as a byproduct of this work since I had to generate a large number of critical clusters anyway. These measurements are briefly described in the Appendix A. $D = 6$ is the upper critical dimension for percolation, where all critical exponents and fractal dimensions assume their mean-field values. Besides d_f , several other fractal dimensions can be defined. The *chemical dimension* d_l , also known as graph dimension, topological dimension, or spreading dimension, describes the scaling of the mass within a *chemical distance* of l steps:

$$M(l) \sim l^{d_l}. \quad (2.20)$$

For non-fractals, d_l also coincides with the Euclidean dimension D . Unlike d_f , d_l cannot (yet?) be related to the thermal exponents, and even its $2D$ value is only approximately known. Closely related to d_l is the *shortest-path dimension* d_{\min} , which describes how the (average) Euclidean distance scales with the (average) chemical distance:

$$l \sim R^{d_{\min}} = R^{d_f/d_l}. \quad (2.21)$$

Numerical estimates for d_{\min} and d_l are also given in Table 2.2.

2.2.3. Cluster structure: backbone and dangling-ends

The most important substructure of a percolation cluster is its so-called *backbone*, the part without singly-connected *dangling ends*. The backbone carries the cluster’s connectivity and is widely considered to be the only relevant part for the asymptotic statistics of SAWs on CPCs. (That issue is investigated in Chapter. 4.)

A common definition of the backbone is as the area through which current would flow

2. Background and theory

Table 2.2.: Estimates for various fractal dimensions of critical percolation cluster. My own estimates (see Appendix A.2) are marked with an asterisk. The values for d_l were obtained via d_f/d_{\min} .

D	d_f	d_{\min}	$d_l (d_f/d_{\min})$
2	91/48 [23]	1.13077(2) [25]	1.6766(3)
3	2.5230(2) [26]	1.3756(3) [26]	1.834(6)
4	3.044(2) *	1.604(3) *	1.90(4)
5	3.517(7) *	1.813(3) *	2.04(2)
≥ 6	4	2	2

if voltage were attached to two distant sites [27] or, almost equivalently, as the union of sites/bonds transversed by all self-avoiding walks connecting two distant cluster sites [28]. The choice of these two “seed sites”, however, is somewhat arbitrary. For clusters that percolate by wrapping around the lattice the backbone is therefore more elegantly defined as the largest *bi-connected component* that wraps around the system. The term “*bi-connected component*” refers to subsets of cluster sites that cannot be disconnected by removing a single site. As can be seen the example of a $3D$ percolating component shown in Fig. 2.2, the backbone is very thinly connected, so that removing a single (“red”) site would be enough to stop it from percolating. Figure 2.3 shows all bi-connected components of a percolating (wrapping) cluster and of a finite (non-wrapping) cluster in $2D$. For finite clusters, the distribution of sizes of bi-connected components is also scale free [29], while for a percolating cluster, the backbone tends to be much larger than the other bi-connected components. The distribution of the singly-connected *dangling ends* (areas that remain if the backbone is removed) is also described by a power law [30].

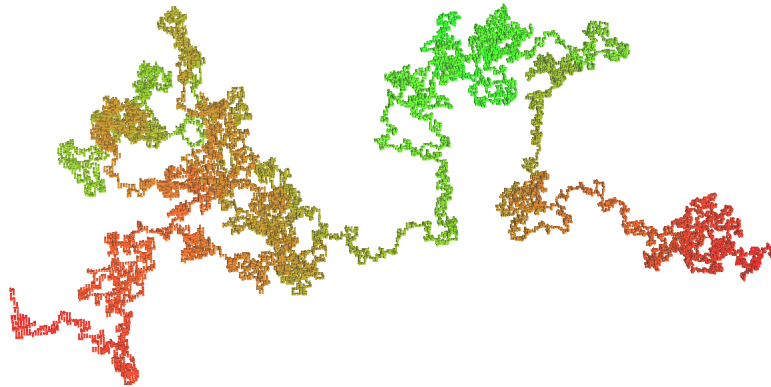


Figure 2.2.: Backbone of a $3D$ cluster of 300^3 sites. Note that the two red ends are connected across periodic boundary conditions.

2.2. Critical percolation clusters

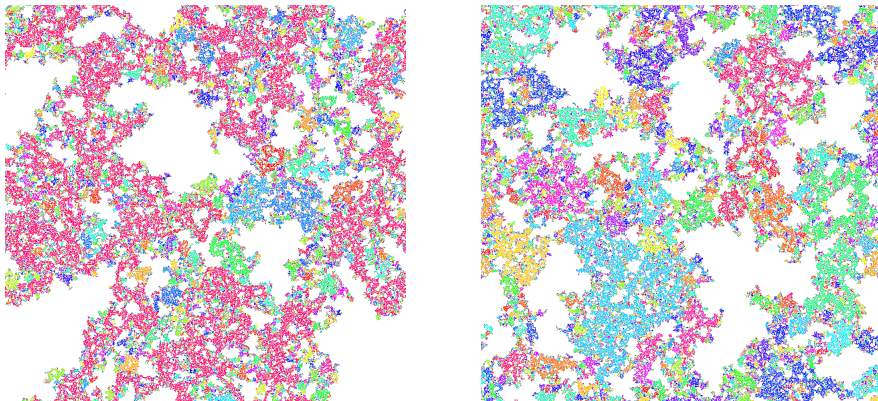


Figure 2.3.: Bi-connected components (randomly colored) of two $2D$ percolation clusters. The one on the left side percolates by wrapping around the lattice, while the one to the right does not. For the percolating cluster, the backbone (red) is the largest component.

Table 2.3.: Numerical estimates for the backbone dimension for different Euclidean dimensions. My own results (see Appendix A.3) are marked with an asterisk.

D	2	3	4	5	≥ 6
d_{BB}	1.6434(1) [19]	1.8736(5) *	1.932(8) *	1.93(16) [28]	2

The backbone's fractal dimension is called the *backbone dimension* of the cluster and denoted here by d_{BB} . For critical clusters its value is always between one and two ($d_{\text{BB}} = 2$ for $D \geq 6$). For supercritical clusters ($p > p_c$), it coincides with the Euclidean dimension of the lattice, just as the fractal dimension of the full cluster. Efforts to find an exact analytical expression for d_{BB} below $6D$ have so far proved unsuccessful even in the simplest case ($2D$). Numerical estimates for d_{BB} are listed in Table 2.3.

The scale-free distribution of bi-connected components implies an important property of CPCs: they are *finitely ramified*. This means that parts of the cluster of any size can be disconnected by removing a finite number of sites (or bonds) [31]. This property is shared by some famous deterministic fractals such as the Sierpinski gasket or the Mandelbrot-Given fractal; see, e.g., Refs. [32, 33]. For these, finite ramification implies that one can perform exact RSRG transformations to solve many problems analytically. They are therefore often taken as qualitative models for CPCs.

2. Background and theory

2.2.4. Diffusion and conductance — dynamical exponents

Transport processes on critical clusters (and other fractals) are characterized by so-called dynamical exponents. Most relevant here is the *walk dimension* d_w describing the spread of random walks (diffusion) in the medium:

$$\sqrt{\langle R^2 \rangle} \sim N^{1/d_w} \quad \left(= N^{\nu_{pc}^{RW}} \right). \quad (2.22)$$

Closely related is the *spectral dimension* (aka *fracton dimension*), $d_s = 2d_f/d_w$. It describes how much area is covered by a RW, its probability of returning to the origin, as well as the density of excitation frequencies (hence the name). Interestingly, the value of d_s is very close to $4/3$ in all dimensions. However, exact equality (the famous Alexander-Orbach conjecture [34]), which would have related the static and the dynamical exponents, has been disproved numerically [35].

The walk dimension is also related to the (total) conductance exponent $\tilde{\mu}$ and the (total) resistance exponent $\tilde{\zeta}$, defined via

$$\sigma \sim L^{-\tilde{\mu}} \quad \text{and} \quad \rho \sim L^{\tilde{\zeta}}, \quad (2.23)$$

where σ and ρ denote the total conductivity and resistance, respectively. This connection is expressed by the ‘‘Einstein relation’’:

$$d_w = d_f - D + 2 + \tilde{\mu} = d_f + \tilde{\zeta}. \quad (2.24)$$

Since conductivity and resistance are determined by the backbone alone, this implies

$$d_{w, \text{BB}} - d_{\text{BB}} = d_w - d_f. \quad (2.25)$$

The walk dimension plays a role in this work as it is thought to be related to the scaling behavior of SAWs; see Section 2.3.3. As can be seen in Table 2.4, numerical estimates for d_w and d_w are unfortunately not very precise, except for the two-dimensional case.

Table 2.4.: Estimates for the walk dimensions of critical clusters and backbones. The results in the last row were calculated from d_w (first row) and the values from Tables 2.3 and 2.2 via Eq. (2.25).

D	2	3	4	5	≥ 6
d_w	2.8784(8) [36]	3.88(3) [17]	4.68(8) [37]	5.50(6) [37]	6
$d_{w, \text{BB}}$	2.62(3) [38]	3.09(3) [38]	-	-	4
$d_{w, \text{BB}}$ (2.25)	2.6260(9)	3.23(3)	3.57(8)	3.9(2)	4

2.3. Self-avoiding walks on percolation clusters

In this section, the model constituting the principal topic of this thesis is introduced more closely. I shall state the main questions, briefly review the most important previous findings, and discuss the various analytical theories and numerical approaches that have been employed.

2.3.1. The model

At the core this work are SAWs on diluted hypercubic lattices with random site dilution. Special focus is on the situation where the level of dilution corresponds to the percolation threshold of the lattice. On each *disorder configuration* C_i (i.e., realization of a dilute lattice), all SAWs start at a designated site s_0 , called *origin*. I always consider ensemble averages over all possible *conformations* of the same number of steps N , i.e., all trajectories of length N that start at s_0 , only pass through sites that are present in C_i , and visit no site more than once. An example of a conformation of $N = 200$ steps on a critically dilute 2D lattice is shown in Fig. 2.4. All such conformations are considered as equally likely, meaning that they contribute uniformly to the ensemble average.

I was not the first to investigate this model; the number of articles dedicated to it (or very close variants, e.g., with a different kind of lattice) is indeed quite daunting [39–86] (I may well have missed some). The following discussion is not meant as a complete review of the topic but shall focus on the (from my perspective) most essential points.

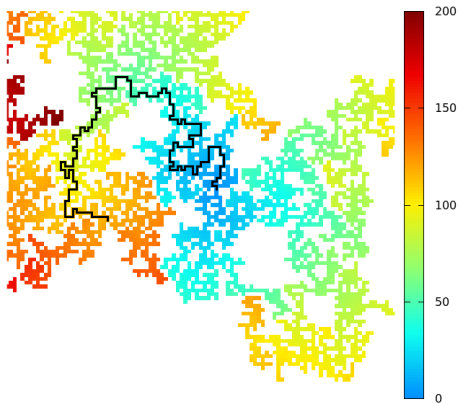


Figure 2.4.: Typical conformation of a 200-step SAW (black) on a critical two-dimensional percolation cluster. Coloring of cluster sites encodes their chemical (shortest-path) distance to the origin (center).

2. Background and theory

2.3.2. Quenched vs. annealed average and other ambiguities

For disordered systems, one has to deal with two averages: one over (all) different disorder configurations (replicas) and one for each individual realization. There are two different ways how this can be done: separately (*quenched*) or simultaneously (*annealed*). The theory of disordered systems (mainly in the context of spin glasses) is discussed in detail in Ref. [87]. For the problem at hand, quenched averaging means that we first take the *conformational average* of an observable by averaging over all Z_C possible walk conformations ω on one *disorder configuration* C , e.g.,

$$\langle R^2 \rangle_C = \frac{\sum_{C,\omega} R_{C,\omega}^2}{Z_C}. \quad (2.26)$$

Secondly, we uniformly take the *quenched disorder average* over all these conformational averages:

$$[\langle R^2 \rangle] := \frac{\sum_C \langle R^2 \rangle_C}{\sum_C 1} \quad (2.27)$$

The term disorder configuration here refers a percolation cluster with a fixed origin *on the percolating cluster* for the SAWs (I will later sometimes simply write “cluster” instead). Every disorder configuration contributes equally to the quenched average, irrespective of Z_C , the number of walks conformations it supports. In turn, single walk conformations on denser configurations that support many SAWs will contribute more with respect to those on sparser ones. The *annealed average*, by contrast, weights all SAW conformations equally, so that the contribution from each disorder configuration is proportional to Z_C :

$$\{R^2\} := \frac{\sum_C \sum_{C,\omega} R_{C,\omega}^2}{\sum_C Z_C} = \frac{\sum_C Z_C \langle R^2 \rangle_C}{\sum_C Z_C}. \quad (2.28)$$

The quenched and the annealed average behave differently in general, and some confusion has been caused by the fact that the distinction was not made very clear in some earlier studies. It is important to stress that the starting point (*origin*) must be fixed on each cluster for the quenched average. Performing the ensemble average over all SAW conformations from different starting points on the same cluster gives an ill-defined mixture of the two averages, which will converge to the annealed average with increasing cluster size.

Although fixing the starting point may seem less physical as it corresponds to polymers that are pinned at one end, the quenched average has attracted much more interest than the annealed one. The reason is that it is more intriguing from a theoretical point of view: while the annealed average is in some ways trivial (as explained below), quenched systems are rich in challenging problems with relations to spin systems and a unique critical behaviors. For this work, I focused on the quenched system, not least because the annealed average is, incidentally, much harder to access numerically. This is because Z_C is distributed very broadly, so that a very small number of configurations will determine the average.

2.3. Self-avoiding walks on percolation clusters

For the average number of conformations Z , the distinction between quenched and annealed does not apply as it is not defined for a single walk¹, but one still has to be mindful of the exact definition of the average: Some studies use $\langle Z \rangle$ to denote the average over all disorder configurations, including those with no SAW conformations (e.g., when the start is on an empty site). I shall instead mainly consider clusters where the starting location is on the incipient cluster and use the notation $[Z]_{\text{IC}}$ for this average:

$$[Z]_{\text{IC}} := \frac{\sum_C Z_C}{\sum_C 1}. \quad (2.29)$$

For both the end-to-end distance and the number of conformations, one can also consider averages over all configurations that support at least one SAW conformation for a given length N . This is usually called the *all cluster* (AC) average in contrast to the *incipient* or *infinite cluster* (IC) average, and shall here be denoted by $[\dots]_{\text{AC}}$. IC and AC averages of the end-to-end distance and other observables are widely believed to share the same asymptotic behavior, and most researchers have so far focused on the IC case. For better comparability with the literature, I therefore also primarily investigated the IC average. In hindsight, however, the AC average might have been a better choice as it can be defined completely independent of boundary conditions, percolation criteria or lattice sizes.

2.3.3. End-to-end distances and the exponent ν

The most prominent question concerning SAWs on diluted lattices is whether (and how) the exponent ν it is affected by the concentration p . It was first raised by Chakrabarti and Kertesz [39], who claimed that dilution is always relevant (in the renormalization-group sense) for both the quenched and the annealed system. Their argumentation was based on the Harris criterion [88], which, in its simplest form, states that disorder is relevant if

$$2 - D\nu_c > 0 \quad (\alpha > 0), \quad (2.30)$$

where ν_c is the critical exponent for the correlation length of the undiluted system. However, as was pointed out by Aharony (see note added in proof of Ref [39]) and later by Kim [43] and Harris himself [44], this simple version of the criterion does not apply in the $n \rightarrow 0$ limit. Indeed, Harris showed that ν is completely independent of the p for the annealed average (taken over all disorder configurations). This is basically because in a randomly diluted system, every SAW conformation from the undiluted system will exist with uniform probability p^N . Since this factor is independent of the extension $\sqrt{R^2}$, the average cannot be affected:

$$\sqrt{\langle R^2 \rangle_p} = \sqrt{\langle R^2 \rangle_1} \sim N^{\nu_1} \quad (2.31)$$

However, the question about the behavior of quenched average the still remained unanswered. One simple argument for a change of ν at critical dilution but not above is

¹Though one could define the annealed averages as $\{Z\} = \frac{\sum_C Z_C^2}{\sum_C Z_C}$.

2. Background and theory

based on the fact that the critical behavior usually depends on the system's dimensionality. Since the fractal dimensions of critical clusters are smaller than the Euclidean dimension of the lattice, one should expect $\nu_{p_c} > \nu_1$, which is also supported by Flory arguments. Indeed, simply replacing D with the fractal dimension d_f in Eq. (2.11),

$$\nu = \frac{3}{d_f + 2} \quad (2.32)$$

gives plausible results, as noted in Refs [40, 49]. However, this is a bit too naive as the derivation from Section 2.1.3 does not work on critical clusters. Also, as pointed out in Ref. [48], d_f is probably not the relevant fractal dimension for the SAWs, and using the backbone dimension d_{BB} instead yields less plausible values. Several Flory approximations have subsequently been proposed [40, 48, 55, 62, 89], which I shall not all discuss. In Ref. [69] they are reviewed and compared to numerical results existing at the time. This analysis already practically excluded all suggestions except

$$\nu = \frac{2 + \alpha_{\text{BB}}}{d_{\text{BB}} + \alpha_{\text{BB}}d_{\text{w,BB}}}, \quad (2.33)$$

with

$$\alpha_{\text{BB}} = \frac{d_{\text{min}}}{d_{\text{w,BB}} - d_{\text{min}}} \quad [55]. \quad (2.34)$$

Eq. (2.33) is derived by minimization of the free energy assuming a radial distribution for random walks of the quite general form

$$P_N(R) \sim \exp - \left(R^{d_{\text{w,BB}}} / N \right)^{\alpha_{(\text{BB})}}, \quad (2.35)$$

and has in fact been proposed in several works though with different estimates for α .

There have also been several studies based on real-space renormalization-group (RSRG) methods [49, 53, 59, 65, 76, 90]. Drawing on the fractal geometry of the system, such approaches are intuitive and can yield qualitative understanding, but they have to rely on uncontrolled approximations. Unfortunately, the various studies lead to quite different conclusions, suggesting that the problem is too subtle for such heuristic approaches. Another strategy is looking at closely related models where the renormalization-group transformations can be formulated exactly, as was done, for instance, in Refs. [48, 91–95]. However, such approaches can also only yield qualitative insights.

A more systematic yet less intuitive way to investigate the problem is by Fourier-space RG analysis of the $O(n \rightarrow 0)$ model (see Section 2.1.3) as was done in Refs. [43, 59, 65, 80, 81, 96, 97]. However, this also produced some disagreement as the combination of replica formalism and $n \rightarrow 0$ -limit proved difficult. Most impact had the work by Meir and Harris [59], where the authors used ϵ -expansion (with $\epsilon = 6 - D$) supported by RSRG-arguments and numerical (exact-enumeration) data on the bond-diluted square lattice. According to what has become known as the Meir-Harris model, the RG flow in the p - z plane (z being the fugacity) has an unstable fixed point at $p = p_c = 1/2$ (and $z = z_c \approx 0.788$) and a stable one at $p = 1$ ($z \approx 0.366$). They

2.3. Self-avoiding walks on percolation clusters

concluded $\nu_p = \nu_1$ for $p > 1$ and $\nu_{p_c} \neq \nu_1$, and a first-order ϵ -expansion gave

$$\nu_{p_c} = 1/2 + \epsilon/42. \quad (2.36)$$

The Meir-Harris model was initially criticized [65], but eventually its predictions gained general acceptance. A following, more detailed investigation by Ferber et al. [80] confirmed Eq. (2.36) and added a second order term:

$$\nu_{p_c} = 1/2 + \epsilon/42 + \frac{110}{9261} \epsilon^2. \quad (2.37)$$

However, this study was later put into question by Janssen and Stenull [81, 97], who claimed that the renormalization could not be performed when using the static conformational average and argued that the kinetic average (corresponding to KGWs) should be considered instead. They even claimed that the conformational average does not have any asymptotic scaling, though without specifying what such a “nonscaling behavior” would look like. (Note that the conformational average of the end-to-end distance is well-defined and its asymptote bounded above and below by power-laws). Their results (for KGSAWs) also confirmed Eq. (2.36), but disagreed with Ref. [80] about the second-order term:

$$\nu_{p_c} = 1/2 + \epsilon/42 + \frac{677}{42} \left(\frac{\epsilon}{42} \right)^2. \quad (2.38)$$

While the scaling exponent for SAWs and KGSAWs is was shown to be the same on regular lattices [98–100], this is not clear in the presence of disorder.

Numerical studies, which used exact enumeration or Monte Carlo methods, have also generated controversy. Earlier works [40, 54, 58] were rather inconclusive about the question whether $\nu_{p_c} = \nu_1$, but later analyses consistently found $\nu_{p_c} > \nu_1$ [67, 72, 73, 78, 82], in agreement with the Meir-Harris prediction. However, there was still some disagreement as to the exact value of ν . The question whether $\nu_p = \nu_1$ holds for $p_c < p < 1$ has received much less attention. Presumably this was because following the publication of the Meir-Harris model, most researchers expected the equality to hold and that no interesting value could be determined. One study where the question was investigated [72] actually found evidence that $\nu_p > \nu_1$, but this finding seems to have largely been ignored. As part of my diploma thesis [101], I had investigated the issue for the $2D$ case, also finding evidence that $\nu_p > \nu_1$.

2.3.4. The number of conformations Z

The behavior of the average number of conformations has also been controversially discussed in a large number of publications. By the same argument that led to Eq. (2.31), it follows that the average of Z over *all* disorder configurations (including those where the origin is a defect site) only trivially depends on the concentration p [41, 44–46, 50]:

$$\langle Z_N \rangle_p \sim p^{N+1} Z_N^{(1)}, \quad (2.39)$$

2. Background and theory

and hence $\mu(p) = p\mu(1)$, according to Eq. (2.6).

However, the above argument does not directly apply for the average $[Z_N]_{\text{AC}}$ over all clusters that support at least one SAW (of length N) and not at all the incipient cluster average $[Z_N]_{\text{IC}}$. There are essentially three main questions concerning the behavior of these averages at the critical concentration p_c :

1. Does a scaling law of the form of Eq. (2.6) hold?
2. Does $\mu(p_c) = p_c\mu(1)$ hold as suggested in Ref. [67]? (And if not, what is $\mu(p_c)$?)
3. Assuming 1., is $\gamma(p_c) = \gamma(1)$? (Or how does γ change?)

Of course, one can also ask these questions for $p \neq p_c$, but the focus is usually (and also here) on the critical case². While 1. has always been assumed, 2. and 3. have been investigated in several studies for both averages. The results have been quite divergent, but the latest numerical studies to address the question [78, 84] found $\mu(p_c) = p_c\mu(1)$ for $[Z]_{\text{IC}}$.

In fact, for $[Z_N]_{\text{AC}}$, 1. and 2. should actually hold for the same reason as Eq. (2.39), as mentioned in Ref [74]. I shall elaborate the essential argument here somewhat more and furthermore show that $\gamma(p_c) > \gamma(1)$: Let $\{C\}_N$ denote the set of all disorder configurations of size $L \gg N$ and $\{C\}'_N$ the subset that allow for at least one N -step SAW. $[Z_N]_{\text{AC}}$ can now be written as

$$[Z_N]_{\text{AC}} = \frac{\sum_{\{C\}'_N} Z_N^C}{\sum_{\{C\}'_N} 1} = \frac{\sum_{\{C\}_N} Z_N^C}{\sum_{\{C\}'_N} 1}. \quad (2.40)$$

With Eq. (2.39) we then get

$$[Z_N]_{\text{AC}} = \frac{\sum_{\{C\}_N} Z_N^C}{\sum_{\{C\}_N} 1} \frac{\sum_{\{C\}'_N} 1}{\sum_{\{C\}'_N} 1} \sim p^N Z_N^{(1)} Q_N^{-1}, \quad (2.41)$$

where the factor $Q_N := \frac{\sum_{\{C\}'_N} 1}{\sum_{\{C\}_N} 1}$ is the probability a randomly chosen conformation allows for an N -step SAW. Due to the scale-free distribution of cluster sizes at p_c , this fraction can be expected to scale with a power law. Since a cluster must comprise at least $s = N$ sites to accommodate an N -step walk, an upper bound is asymptotically given by

$$Q_N \leq \int_N^\infty s n_s ds \sim \int_N^\infty s^{1-\tau} ds \sim N^{2-\tau}, \quad (2.42)$$

where n_s is the *cluster number* and the τ the Fisher exponent, which is always larger than two. Thus we can conclude

$$[Z_N]_{\text{AC}} \sim \mu_{p_c}^N N^{\gamma-1+\delta} \quad (2.43)$$

²Note that $[Z_N]_{\text{AC}}$ and $[Z_N]_{\text{IC}}$ have the same behavior for $p > p_c$, while the incipient cluster is not defined for $p < p_c$.

2.3. Self-avoiding walks on percolation clusters

with

$$\mu_{p_c} = p_c \mu_1 \quad \text{and} \quad \delta \geq +\tau - 2. \quad (2.44)$$

With $\tau_{2D} = 187/91$ and $\tau_{3D} = 2.18$ and the values for γ from Section 2.1.2 we thus get

$$\gamma 2D_{p_c} \geq \gamma_1^{2D} + \frac{5}{98} = \frac{2187}{1568} \quad \text{and} \quad \gamma_{p_c}^{3D} \gtrsim \gamma_1^{3D} + 0.18 \approx 1.337. \quad (2.45)$$

A very similar argument was made by Grassberger [72] for the average entropy. Note that Q_N can be measured much more easily than $[Z_N]_{AC}$ itself.

The above argumentation does not work for the incipient-cluster average. However, we can expect

$$[Z_N]_{IC} \geq [Z_N]_{AC} \quad \text{for } p = p_c, \quad (2.46)$$

$$[Z_N]_{IC} \sim [Z_N]_{AC} \quad \text{for } p > p_c, \quad (2.47)$$

where the second statement is due the fact that large clusters but finite clusters are exponentially suppressed above the percolation threshold. It follows that the leading factor for the scaling behavior of $[Z_N]_{IC}$ should also be $(p\mu^1)^N$: if it were larger —it cannot be smaller according to Eq. (2.46)— there would have to be a $p > p_c$ so that $[Z_N]_{IC}(p_c) > [Z_N]_{AC}(p) \sim (p\mu^1)^N N^{\gamma'-1}$ for some N , and hence, with Eq. (2.47), $[Z_N]_{IC}(p_c) > [Z_N]_{IC}(p)$ asymptotically, which is obviously false.

A quantity related to the average number of chains is the (quenched) average of the entropy $\ln Z_N$. Here even the average over all disorder conformations $\langle \ln Z_N \rangle$ (with $\ln 0 := 0$)³ is non-trivial. As first remarked by Derrida [46] it should behave differently from $\ln \langle Z_N \rangle$ due to the multifractal distribution of Z . The issue was subsequently investigated more closely, but the results were highly controversial: It was first argued that

$$\langle \ln Z_N \rangle \sim N \ln p\mu - \frac{1-p}{2p\alpha} N^\alpha, \quad (2.48)$$

with $\alpha = 2 - D\nu$ being the specific-heat exponent (of the undiluted system) [56, 64]. Later work based on analytical, Lifshitz-type arguments and exact enumeration results [68, 71], concluded

$$\langle \ln Z_N \rangle \simeq N(\ln \mu - 1.2 |\ln p|^{1/D\nu}) + (\gamma - 1) \ln N + \text{corrections} \quad (2.49)$$

(with some evidence for a stretched-exponential correction term), while a MC investigation [72] yielded

$$\langle \ln Z_N \rangle \sim N^{1-\delta}, \quad (2.50)$$

with $\delta = 0.075(5)$. In that last study, a linear increase with N was found for the average $[\ln Z_N]_{AC}$, which was related to $\langle \ln Z_N \rangle$ by essentially the same argument used above [see Eq. (2.41)]. The Incipient cluster average of the entropy, $[\ln Z_N]_{IC}$, has, to my best knowledge, not yet been directly investigated, but there have been investigations of the average over incipient-cluster backbones, $[\ln Z_N]_{BB}$, assuming

³Some authors use $\langle \ln(Z_N + 1) \rangle$ instead, but the distinction does not matter in the limit $N \rightarrow \infty$.

2. Background and theory

that the two should behave alike asymptotically [77, 78]. These studies suggested a scaling law as for $\ln Z_N$:

$$[\ln Z_N]_{\text{BB}} \sim \ln \mu_0 + (\gamma_0 - 1) \ln N, \quad (2.51)$$

with modified parameters:

$$\mu_0^{2D} = 1.456(5), \quad \gamma_0^{2D} = 1.26(5), \quad \mu_0^{3D} = 1.317(5), \quad \gamma_0^{3D} = 1.19(5). \quad (2.52)$$

2.3.5. Open questions

To end this chapter, I have compiled a list of issues regarding SAWs on percolation clusters that are, in my view, not fully settled, some of which will be addressed in this thesis.

- Is $\nu_p = \nu_1$ for $1 > p > p_c$? This is generally assumed to be the case based on analytical arguments [59], but the (few) numerical studies [72, 101] that addressed the issue found evidence against it.
- What is the value of ν_{p_c} ? While it has by now been convincingly shown that $\nu_{p_c} > \nu_1$, numerical and analytical estimates vary substantially (see Table 2 in Ref. [96]).
- Is the scaling behavior on the backbones the same as on full clusters, i.e., $\nu_{\text{IC}} = \nu_{\text{BB}}$? This assumption was often made [48, 60, 70, 74, 78, 83], but the argument on which it is based is flawed, in my view at least [101], and numerical evidence on the matter is so far inconclusive.
- Does averaging over all clusters that support SAWs lead to the same exponent as averaging over the largest (incipient) clusters, i.e., $\nu_{\text{IC}} = \nu_{\text{AC}}$? This is also usually assumed, but numerical results on the matter are sparse.
- Is ν for KGWs on critical clusters the same as for the standard SAWs?
- How does $[Z]_{\text{IC}}$ scale with N ? Does it, as is generally expected, follow a law like Eq. (2.6), and if so, are μ and γ the same as for the all cluster average $[Z]_{\text{AC}}$?
- How does the average entropy $[\ln Z_N]$ scale with N ? As mentioned, at least four very different conjectures regarding this issue have been put forward.
- Can ν (and γ) be linked, qualitatively at least, to the various fractal dimensions of the medium?
- How accurate are the existing analytical results for ν ?

3. The taming of the SAW

The SAW is difficult to treat numerically (or analytically) due to the “non-Markovian” long memory of the process. One either has to generate each single conformation or resort to chain-growth Monte Carlo methods such as the pruned-enriched Rosenbluth method (PERM) [102]. These work reasonably well for low or medium concentrations of defects, but they falter and become unreliable around the percolation threshold. At this particular point, however, it is actually possible to do exact enumeration quite efficiently via a new approach that makes use of the fractal nature of the medium [103–106]. The bulk of this chapter will be dedicated to presenting and discussing this novel method, through which most of this work’s results were obtained. First though, some commonly used methods shall briefly be discussed, as far as they are relevant here.

3.1. Existing numerical methods for SAWs on CPCs

For regular lattices, highly sophisticated algorithms have been developed to investigate SAWs: The most efficient enumeration methods have allowed for 71 steps in $2D$ [107] and 36 steps in $3D$ [5, 108]. Due to the exactness of the data, these results can be extrapolated substantially. The best current MC method uses Markov chain (“pivot”) updates that exploit the low density and fractal nature of the SAW and can thus handle several hundred million steps [109]. Unfortunately, neither approach seems to work for disordered systems. There, conformations must be enumerated directly, which allows for much fewer steps. Meanwhile, Markov-chain Monte Carlo methods become very inefficient as global moves are almost always rejected, so that one is better served with chain-growth Monte Carlo methods.

3.1.1. Chain-growth Monte Carlo

Simple sampling: The most straightforward Monte Carlo approach samples conformations with uniform probability. For athermal systems, this is theoretically the optimal way to sample, but it can be difficult to realize in practice. For the SAW, one has to grow a conformation by taking steps in random directions and discard it if a blocked site is picked (i.e., one that is defect or already visited). Thus, the probability to succeed will shrink exponentially with the number of steps taken, inhibiting the generation of long chains. Despite this *exponential attrition*, simple sampling has been used now and then to study SAWs on CPCs [40, 58, 67]. It allows for longer chains than (simple) exact enumeration (discussed in Section 3.1.2), and one does not have to worry about biases that can afflict more sophisticated MC methods.

3. The taming of the SAW

Rosenbluth-Rosenbluth method (RR): The attrition rate of the simple-sampling algorithm can be reduced by choosing only from neighbor sites that are available. However, conformations are then no longer generated with uniform probability but with

$$P_n \propto \prod_{i=1}^n \frac{1}{m_i}, \quad (3.1)$$

where m_i denotes the number of free neighbors for the i th step and n is the number of steps. The RR method does hence not sample the equilibrium SAW ensemble, where all conformations contribute equally, but the so-called Kinetic-growth walk (KGW) discussed in Section 5.3. To get back to the standard ensemble, however, we simply need to weight conformations with

$$w_n = \frac{1}{P_n} = \prod_{i=1}^n m_i. \quad (3.2)$$

This algorithm was first proposed in the very early days of computing [110]. Although it suffers much less attrition than the simple-sampling method, it is not actually more useful as it wastes most effort on low-weight conformations that do not contribute much to the average. This “go-with-the-losers” strategy is not only inefficient, it also causes a bias if samples are too small. This is because the weighting only works correctly if the total weight is properly assessed, i.e., if a sufficient number of (rare) conformations with large weights have been sampled. The RR method was therefore never applied to the problem at hand, but it is nonetheless important as it underlies the widely used pruned-enriched Rosenbluth method.

Pruned-enriched Rosenbluth method (PERM): The problems of the RR method can elegantly be overcome by a population-control strategy. The idea is to support conformations that have relatively large weights and suppress those with small weights: Whenever a weight w_n surpasses an upper threshold $W_n^>$, the conformation is duplicated (“enriched”) and the weight is halved. Conversely, if w_n falls short of a lower threshold $W_n^<$, it is doubled and the chain is terminated (“pruned”) with probability 1/2. These modifications keep all expectation values unchanged but constrain the weights between the thresholds. The tricky part is to adjust these thresholds such that the population of chains does neither explode nor collapse but remains roughly constant with increasing length. This is best done self-consistently by pegging the weights to the average total weight per started chain (“tour”):

$$W_n^> = C^> \frac{\sum_{t=1}^T \tilde{w}_{n,t}}{T} = C^> \frac{\sum_{t=1}^T \sum_{b_t=1}^{B_t} w_{n,b_t}}{\sum_{t=1}^T B_t}, \quad W_n^< = \frac{C^<}{C^>} W_n^>, \quad (3.3)$$

where $\tilde{w}_{n,t}$ is the accumulated weight of all branches B_t created (by enriching) during tour t , T denotes the number of tours, and $C^>$ is large than $C^<$. The constants $C^>$ and $C^<$ control the size of the population ($\approx 1/\sqrt{C^>C^<}$) and are best chosen smaller than one to give the population an initial boost. The exact values of the constants

3.1. Existing numerical methods for SAWs on CPCs

are not overly important; they can be tuned easily during a few test runs. For most simulations, I used $C := \sqrt{C^>C^<} = 0.001$, $C^>/C^< = 10$. The algorithm is realized efficiently by a simple depth-first recursion, which takes as arguments a position pos ($= \vec{R}$) on the lattice and the length n and weight w of the branch:

```
perm_step(n, w, pos){
  Rsq_acc[n] += w*Rsq[pos];
  if(n < N){
    options = number of available neighbors;
    if(options > 0){
      w *= options;
      w_acc[n] += w;
      if(w < C_bottom*w_acc[n]/tours){ //prune
        if(random_nr < 0.5){
          return;
        }
        w *= 2;
      }
      pos = blocked; //block visited site
      neighbor_pos1 = random available neighbor; //choose next position
      if(options > 1 && w > C_top*w_acc[n]/tours){ //enrich
        do{
          neighbor_pos2 = random available neighbor;
        }while(neighbor_pos2 == neighbor_pos1)
        w /= 2;
        prem_step(n+1, w, neighbor_pos2);
      }
      prem_step(n+1, w, neighbor_pos1);
      pos = available; //free visited site
    }
  }
}
```

The variables $\text{Rsq_acc}[n]$ ($\sum_{t=1}^T \sum_{b_t=1}^{B_t} w_{n,b_t} R_{n,b_t}^2$) and $\text{w_acc}[n]$ ($\sum_{t=1}^T \sum_{b_t=1}^{B_t} w_{n,b_t}$) accumulate the total end-to-end distances and weights of all branches of length n from all tours, respectively. The function $\text{perm_step}(0,1,\text{origin})$ is called repeatedly, incrementing tours until enough have been performed. In this implementation, branches are only enriched if there are more than one available neighbors, which is actually closer to the improved version of PERM (“nPERMss”) from Ref [111].

PERM was originally designed to simulate Θ -polymers, which it does extremely well [102], but it also works for normal SAWs in disorder [82, 84] and a range of other models; see Ref. [112]. Unlike the simple RR method, it follows a “go-with-the-winners” strategy where conformations with large weights receive more attention. It also overcomes the *exponential attrition*, which still afflicts the RR method to some extent: running into “dead-ends” will typically only result in the loss of short branches, and the algorithm can compensate this by increasing the rate of *enrichment*.

Still, PERM often struggles to explore the rugged landscape of a critical percolation cluster. Some cluster regions can be very difficult to enter even though they may

3. The taming of the SAW

be important for the average. This can lead to a bias similar to that of the RR method (though much less severe), which can easily go unnoticed [101], an issue that is investigated more closely in Section 3.5. In this work PERM was used in situations where the SFE method described in the next section does not (always) work, namely on supercritical clusters and for KGWs. For these systems, the aforementioned bias happens to be much less of a problem. The PERM version for KGWs is particularly simple as the weights are only determined by the attrition rate due to trapping and no pruning is required at all.

If one is only interested in the results for the maximum length, one can increase the efficiency of the program by closing off *dangling ends* that are too small to accommodate the remaining $N-n$ number of steps. As an upper limit for the maximal number of possible SAW steps within a dangling end, I took its mass minus the mass of all smaller dangling ends it contained plus the largest such upper limit among these smaller dangling ends. On critical clusters, this “trap avoiding” optimization increases the performance significantly, but it has little effect above p_c .

3.1.2. Exact enumeration

Besides Monte Carlo methods, exact enumeration (EE) can be used to study SAWs on percolation clusters (or any other graphs). In this approach, one calculates the conformational averages up to a fixed length on individual clusters by computationally generating all possible SAW conformations. This simple EE method was not used to produce results for this work, but it underlies the SFE method introduced in the next section. It also played an important role in verifying the correct implementation the other methods (SFE and PERM).

Essentially, the method consists of a single recursive routine that generates the chain conformations by a depth-first traversal of the cluster: At each call the routine receives as arguments the length n of the current chain and the position \mathbf{x} ($= \vec{R}$) of its end. A chain counter $Z[n]$ is incremented, the squared Euclidean distance to the origin, $\text{Rsq} [=(\vec{R} - \vec{R}_0)^2]$, is measured and added to a variable $\text{Rsq_acc}[n]$, and \mathbf{x} is blocked to future visits (self-avoidance). If the final length N is not yet reached, all nearest-neighbor sites are addressed in turn. For each neighbor that is available (i.e., neither visited nor defect) the routine calls itself with the incremented length and the neighbor’s position as arguments. At the end of each frame of the routine, the original position pos is unblocked. The pseudo code for this function is sketched below:

```
step(n, pos){
  Rsq_acc[n] += Rsq[pos];
  Z[n] += 1;
  if(n < N){
    x = blocked;
    for(all available neighbor_pos){
      step(n+1, neighbor_pos);
    }
    pos = available;
  }
}
```

3.2. Scale-free enumeration (SFE)

The function is initially called with the origin and a length of zero as arguments. It terminates automatically once all possible conformations have been generated, and the average end-to-end distance $\langle R_n^2 \rangle$ is obtained as $\text{Rsq_acc}[n]/Z[n]$. Apart from purely statistical errors on the disorder averages, one thus obtains exact results, avoiding the risk of any bias. Therefore, and because of the method's simplicity, EE is often preferred over MC methods and has widely been applied to study the problem at hand [61, 71, 73, 78, 113]. EE has, however, one huge drawback: The number of conformations increases exponentially with the number of steps, and so does the time needed to generate them. Thus, the affordable number of steps is quite limited: the latest study enumerated walks of up to 45 steps on three dimensional clusters [113], and even on a modern super computer this number could probably not be increased by a substantial factor.

3.2. Scale-free enumeration (SFE)

None of the methods discussed so far is particularly suited for the problem at hand. Exact enumeration is limited to very short chains due to the exponential complexity, seemingly inherent to the approach. Chain-growth MC methods (PERM), on the other hand, can handle somewhat longer chains but are much less efficient and reliable than they are on full lattices as the thin, fractal landscape of the critical clusters is difficult to explore. Fortunately, however, this fragile, self-similar nature of the medium can also be exploited to factorize the problem. More specifically, we can make use of the fact that critical percolation clusters are finitely ramified fractals; see Section 2.2.3. For regular, geometrical fractals with this property such as the Sierpinski gasket, Fig.3.1 (left), many problems can be solved by exact real-space renormalization. This is achieved through repeated application of a decimation procedure, making use of the scale symmetry. For instance, we can calculate the number of SAW conformations connecting two tips of a Sierpinski gasket after i iterations (one iteration combining three smaller triangle into a larger one) from the number of conformations connecting the tips after $i-1$ iterations. This decimation ("coarse graining") is depicted in Fig. 3.1 (right) from Ref. [92], where P'_1 denotes the number of SAWs connecting tips A and B without visiting C, P'_2 the number of SAWs connecting A and B via C, and the unprimed P s correspond to the respective numbers in the previous iteration.

Such an exact transformation scheme cannot be applied directly for disordered systems which lack the exact symmetry. However, there is still a statistical scale symmetry, which means that we can divide the cluster into nested cells on different length scales that have the same properties *on average*. Crucially, thanks to the finite ramification, the average number of interconnections between cells need not increase with their size. Hence, for any problem that involves only local (nearest-neighbor) interactions and can be solved numerically on a small scale, it should be possible to find a numerical decimation procedure such that a more general version of the method can handle cells of any size (consisting of smaller, decimated cells) with roughly the same effort.

Let us now turn to our specific problem of SAWs on CPCs. Within a small sepa-

3. The taming of the SAW

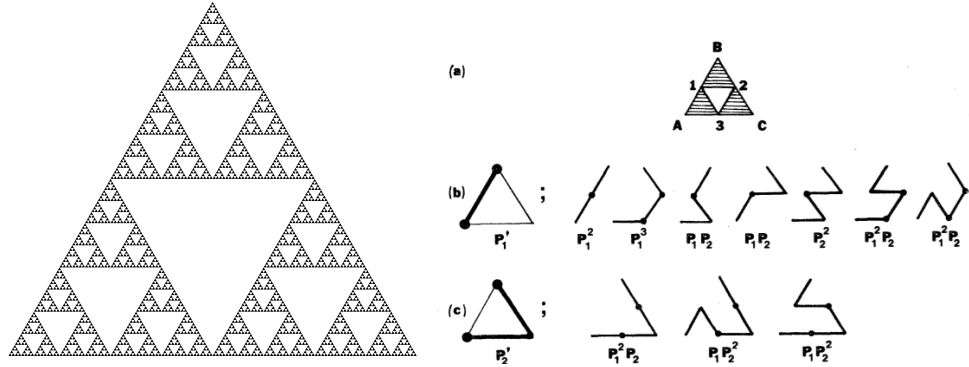


Figure 3.1.: Left: Sierpinski gasket fractal after six iterations. Right: Coarse graining procedure for SAWs on a Sierpinski gasket (with permission from Ref. [92]), where the numbers P'_1 (P'_2) of SAWs from A to B (via C) are constructed from the result P_1 and P_2 of the previous iteration.

rate region of space (“cell”)¹ one can easily determine the number of SAW *segments* connecting any pair of sites using the simple exact enumeration method from above. Imagine now that this small cell (“child”) is part of a larger one (“parent”), and that the number of site pairs which connect the child to the rest of the parent (“links”) is small (of the order of one). We can thus easily enumerate all *segments* that connect different links within the child. The crucial observation now is that when we enumerate the segments through the parent, the child can effectively be treated as a single, “decimated” site. Any walk that visits it is weighted with the number of segments that connect the respective *link sites* through the small cell, and the enumeration can thus be factorized. The finite ramification now implies that we should be able to partition the cluster into a hierarchy cells where each has relatively few links and is small enough (once all its children have been decimated) that we can exactly enumerate all SAW segment conformations connecting them.

The “scale-free enumeration” method will thus consist of two parts: the cluster is first partitioned into a suitable hierarchy of nested cells; then the enumeration procedure is performed cell-by-cell. Since understanding what kind of cell hierarchy is “suitable” requires detailed knowledge of the enumeration procedure, the two parts will be presented in reverse order.

3.2.1. Factorizing the enumeration

The basic ideas outlined above are relatively simple, but a thorough description of the method is challenging: due to its recursive nature, some aspects need to be understood before they can be properly explained. I therefore do not start with a general formula-

¹I had previously [101, 103, 104] used “blob” instead of *cell*, which was an unfortunate name as it could be confused other usages of that term.

3.2. Scale-free enumeration (SFE)

tion of the algorithm but demonstrate how it works using a simple example. General concepts are introduced along the way and condensed into the paragraph “Generic procedure”, to which the impatient reader may skip directly.

Let us now consider the following problem: We want to count the number of conformations of SAWs with up to $N = 15$ steps starting from a fixed point (origin) of a percolating cluster using the ideas from above. It is convenient to use such a short length (which could easily be handled by the standard method) to explain the basic principles. These remain unchanged for longer walks. I leave the end-to-end distance aside for the moment; its measurement is discussed later on. The first thing we can do is to crop off everything that cannot be reached within N steps, reducing the cluster to N chemical shells around the origin. Next, we partition this cropped cluster into a hierarchy of nested cells as shown in Fig. 3.2 for a specific example which shall be used throughout in the following. How the partitioning is done is explained in the next section. Note that each cell in the hierarchy may contain several other cells (“children”) as well as a number of “bare sites” that are not encapsulated in the children. Also note that the starting location is among the bare sites of the largest cell (“root”). The cells will be dealt with in *post-order*: E, D, C, B, A.

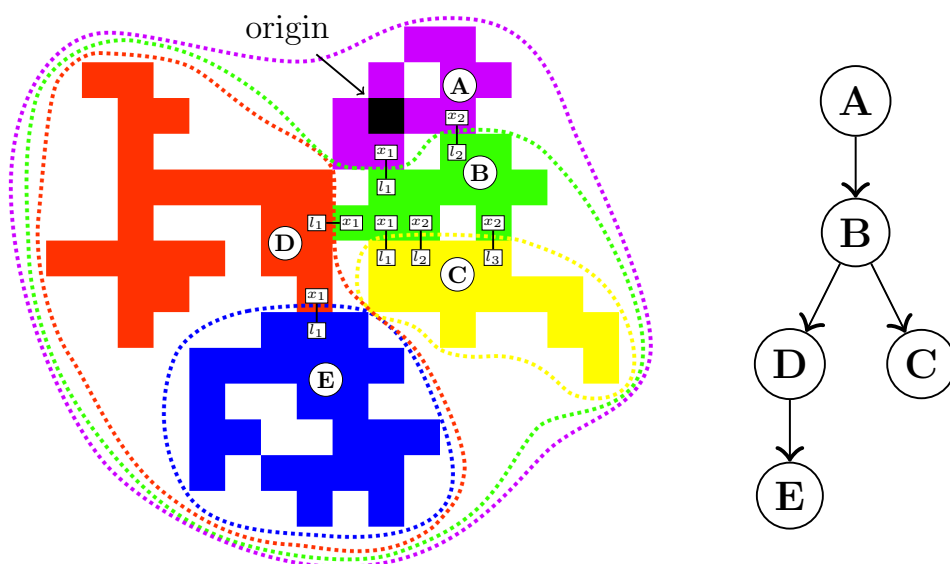


Figure 3.2.: Partitioning of an exemplary cluster into a tree hierarchy of nested cells. The cluster has been cropped to $N = 15$ chemical cells around the origin (black site in cell A), the maximal SAW length considered in this example.

3. The taming of the SAW

Cell E (one link, no children): We start by determining the number of SAW *segments*, $Z^E[n]$, of length up to $n_{\max} = N - d_{l_1} = 7$, where d_{l_1} is the chemical distance of *link site* l_1 in E. For this we essentially use the simple enumeration method from Section 3.1.2: we block the external site x_1 in D to confine the walks to cell E and initialize the routine with link site l_1 as the starting position and a length of $n = 0$. This yields the results listed in Table 3.1.

The term “*segment*” will in the following be used to denote a unique conformation of parts of SAWs within a cell. One segment can consist of several snippets which need not be connected within the cell. However, each must be connected to at least one link (external site) or to the origin (for the root cell).

Table 3.1.: Number of SAW segments within cell E of the example in Fig. 3.2

n	0	1	2	3	4	5	6	7
$Z^E[n]$	1	3	5	11	12	18	21	32

Cell D (one link, one child): Next we decimate, “renormalizing” cell E effectively to a single site “ e ”, and move on to enumerate the walk segments within its parent, cell D. Here we do essentially the same thing (with $n_{\max} = N - d_{l_1} = 10$), but we use two separate counters: $Z_{e_0}^D[n]$ for walks that have not visited “child site” e and $Z_{e_1}^D[n]$ for those that have. Apart from this, e behaves as any regular site. Once the counting process has terminated we can calculate the total number of segments within D via

$$Z^D[n] = Z_{e_0}^D[n] + \sum_{i=0}^n Z_{e_1}^D[i] Z^E[n-i]. \quad (3.4)$$

The sum term represents combinations of segments with a combined length of n . In the following, this “convolution product” shall be written as:

$$(Z^X * Z^Y)[n] := \sum_{i=0}^n Z^X[i] Z^Y[n-i]. \quad (3.5)$$

In practice, this operation is implemented as a double loop:

```
for(n_X=0; n_X <= N; n_X++){
  for(n_Y=0; n_Y <= N; n_Y++){
    Z_XY[n_X+n_Y] += Z_X[n_X]*Z_Y[n_Y];
  }
},
```

where n_X+n_Y corresponds to n from above.

When the summations are done, all information concerning cell E, in particular $Z^E[l]$, can be deleted to release precious memory space. The results for the counters

3.2. Scale-free enumeration (SFE)

$Z_{e_0}^D[n]$ and $Z_{e_1}^D[n]$ are listed in the first two rows of Table 3.2; the last row gives the results of the summation, Eq. (3.4).

Table 3.2.: Number of SAW segments within cell D for the example from Fig. 3.2. The first two rows have resulted from direct counting of segments that either connect (first row) or do not connect (second) to child site e . The last row gives the total numbers of segments, obtained via Eq. (3.4) with Z^E from Table 3.1.

n	0	1	2	3	4	5	6	7	8	9	10
$Z_{e_0}^D[n]$	1	3	5	6	5	5	4	5	5	12	14
$Z_{e_1}^D[n]$	0	0	0	1	0	1	0	1	0	0	0
$Z^D[n]$	1	3	5	7	8	11	18	23	37	50	75

Cell C (three links, no children): Next in line is cell C, which has three links. Things now become more complicated as we need to distinguish between a number (namely ten) of different classes of segments: three that start from one link but do not visit any other, three connecting a pair of links, and three that connect two links and reenter at the third. We shall in addition include the “empty” case without any segments. These classes of segments shall be associated with different “states” of the cell, which are represented diagrammatically as shown in Fig. 3.3. Note that states fall into two categories: Those that have a loose end, i.e., where the walk ends within the cell, are denoted as “terminal” (t -states), the others as “conductive” (c -states). Also note that the segments connecting two links are considered undirected.

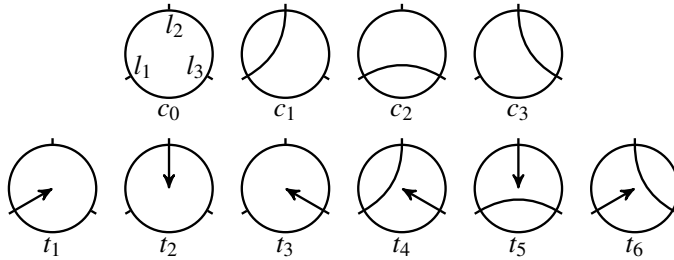


Figure 3.3.: All possible states for a cell with three links (l_1, l_2, l_3), representing, e.g., cell C from Fig. 3.2. (Figure reproduced from Ref. [104])

Maximum number of states: In order to understand the limitations of the approach and the requirements on the cell hierarchy, it helpful to know how the number of possible states of a cell Ω increases with the number of links L . To calculate $\Omega(L)$, we first consider all possible c -states for a cell with $2i$ connected and $L - 2i$ open

3. The taming of the SAW

links. Their number is equal to the number of states with no open links (as if $L = 2i$), denoted as $C[2i]$, times the $\binom{L}{2i}$ possibilities to intersperse the unconnected links. For $D \geq 3$, $C[2i]$ equals the number of possibilities to partition the $2i$ links into i pairs. The number of possible permutations of all links is $(2i)!$, but the order of the pairs ($i!$ possibilities) is irrelevant as are the 2^i possible permutations of links within the pairs, hence: $C[2i]_{D \geq 3} = \frac{(2i)!}{i!2^i}$. In $2D$ the number is smaller as the lines in the state diagrams may not cross. In that case states without unconnected links can be mapped one-to-one to binary trees with i nodes [114], for which the number of configurations is given by the Catalan numbers: $C[2i]_{2D} = \binom{2i}{i}/(i+1)$. For each c -state, there are $L - 2i$ t -states, one for each unconnected link. Summing over all $2i \leq L$ we hence obtain for the total number of states as

$$\Omega(L) = \Omega_c(L) + \Omega_t(L) = \sum_{i=0}^{\lfloor L/2 \rfloor} C[2i] \binom{L}{2i} (1 + L - 2i), \quad (3.6)$$

with

$$C[2i] = \begin{cases} \binom{2i}{i}/(i+1), & D = 2 \\ \frac{(2i)!}{i!2^i}, & D \geq 3. \end{cases} \quad (3.7)$$

These series increase rather quickly as can be seen in Fig. 3.4, where $\Omega(L)$ is plotted for $D = 2$ and $D \geq 3$. Although typically many of the states cannot be realized in practice due to constraints within a cell or limited length of the walks, cells with much more than ten links tend to be unmanageable.

Our goal now is to determine the number of segments for each c -state and t -state, $\{Z^{C,ci}[n]\}$ and $\{Z^{C,ti}[n]\}$. This need not be done for each state separately; instead, segments belonging to one state can be extended to become segments from another.

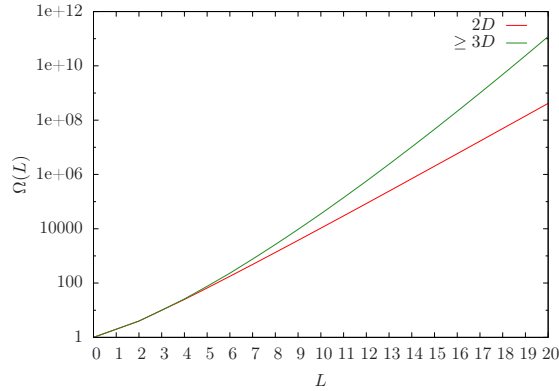


Figure 3.4.: Maximum number of states of a cell as function of the number of links [Eq. (3.6)] in $D = 2$ and $D \geq 3$.

3.2. Scale-free enumeration (SFE)

For instance, a segment from t_1 can grow into a c_1 segment by connecting to l_2 (see Fig. 3.3) and then become a t_4 segment by reentering at l_3 . This behavior can be integrated smoothly into the enumeration procedure: We treat the three external link sites (x_1, x_2, x_3 in Fig. 3.2) as one single “parent site” representing everything outside of our cell C. Whenever this parent site is visited by the walker, we switch from a t -state to a c -state and to another t -state when it is left; i.e., when the cell is reentered. Separate counters need to be kept for each individual state.

The transitions between c -states and t -states for a three-link cell are depicted in Fig. 3.5. The transition rules, telling us which next state will result from stepping over a link, are stored in matrices as shown in Table 3.3. How these matrices are determined, which is done once at the start of the program. Transitions where the entries are empty or in brackets are not allowed, either because they are impossible or because they would lead to double counting: since the loose end of a t -state is directed but segments (snippets) connecting a pair of links are not, different t -states could lead to the same c -state, which is prevented by blocking all but one of these transitions. In principle, it does not matter which of the $t \rightarrow c$ transitions is allowed.

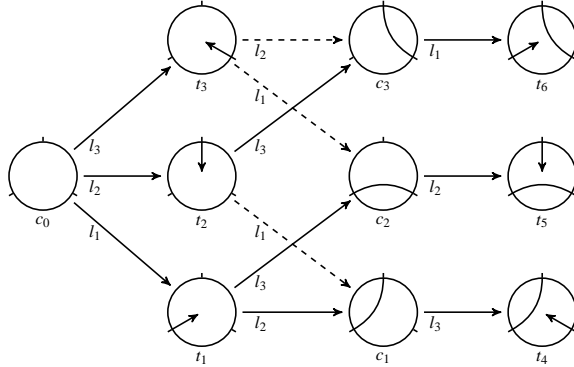


Figure 3.5.: Diagram of allowed transitions for a cell with three links. The labels on the arrows denote to which link a SAW segment must connect for the transition to occur. Dashed arrows represent transitions that are not allowed. (Figure reproduced from Ref. [101])

Table 3.3.: Matrices determining the transitions between states for a cell with three links, e.g., cell C from Fig. 3.2. The table on the left describes transitions from a c -state to a t -state, the one on the right those from a t -state to a c -state. Transitions are not allowed for empty entries or those in brackets.

	c_0	c_1	c_2	c_3
l_1	t_1	-	-	t_4
l_2	t_2	-	t_6	-
l_3	t_3	t_5	-	-

	t_1	t_2	t_3	t_4	t_5	t_6
l_1	-	(c_1)	(c_3)	-	-	-
l_2	c_1	-	(c_2)	-	-	-
l_3	c_2	c_3	-	-	-	-

3. The taming of the SAW

Table 3.4.: Number of SAW segments for all states of cell C of the example from Fig. 3.2.

n	0	1	2	3	4	5	6	7	8	9	10	11
$Z^{C,t_1}[n]$	1	2	3	6	11	14	13	12	9	8	4	4
$Z^{C,t_2}[n]$	1	3	5	9	9	11	8	8	6	3	3	0
$Z^{C,t_3}[n]$	1	3	4	9	10	13	10	7	3	0	0	0
$Z^{C,t_4}[n]$	0	1	2	4	9	8	8	5	0	0	0	0
$Z^{C,t_5}[n]$	0	0	0	0	0	2	1	0	0	0	0	0
$Z^{C,t_6}[n]$	0	0	1	1	4	4	3	1	0	0	0	0
$Z^{C,c_0}[n]$	1	0	0	0	0	0	0	0	0	0	0	0
$Z^{C,c_1}[n]$	0	1	0	1	0	1	0	1	0	0	0	0
$Z^{C,c_2}[n]$	0	0	0	1	0	6	0	1	0	0	0	0
$Z^{C,c_3}[n]$	0	0	1	0	3	0	2	0	0	0	0	0

The counting starts at the parent site with the cell in state c_0 and a length of $n = 0$. By convention n is not incremented for steps on and off the parent site. “Normal” sites are blocked when they are visited as before, but the parent site is not. Once the routine has terminated, we are left with the set of arrays listed in Table 3.4. Note that this table also includes the “empty” c_0 state, which had been omitted thus far.

Generic procedure (L links and k children with L_k links): The next target, cell B, has two links and two children, one with one link and the other with three. With this example in mind, let us now switch to a generic formulation: We want to determine the number of segments $Z^{X,s_X}[n]$ for each state s_X of a cell X that has L links and k children, x_1, \dots, x_k , with L_1, \dots, L_k links, respectively. Each state represents an ensemble of SAW *segments*, i.e., all distinct parts of viable SAW conformations crossing the cell. The c -states are uniquely defined by a specific set of pairs of connected links, t -states in addition have a loose end attached to one of the links (see Fig. 3.3). For each child and for the cell itself, we have matrices describing transitions between the states; see Table 3.3. These contain the information which t -state will result from attaching a loose end to a free link of a c -state, and which c -state will result from connecting the loose end of a t -state to a free link. For the children, all transitions to states that have occurred (when the segments through the child were enumerated) are allowed, whereas for the cell each c -state can only be accessed from one t -state (solid lines in Fig. 3.5 and non bracketed entries in Table 3.3). We now do a depth-first recursion on a graph consisting of the children (“child sites”), the sites that are directly in the cell (“bare sites”), and the external part of the cluster (“parent site”). Children and parent site behave point-like, but whenever they are visited or left by the walker the state of the respective child or the cell itself (for the parent site) changes according to the transition rules. Figure 3.6 represents a snapshot of the enumeration procedure for an exemplary cell with one link and two children. For each distinct combination of states (of cell

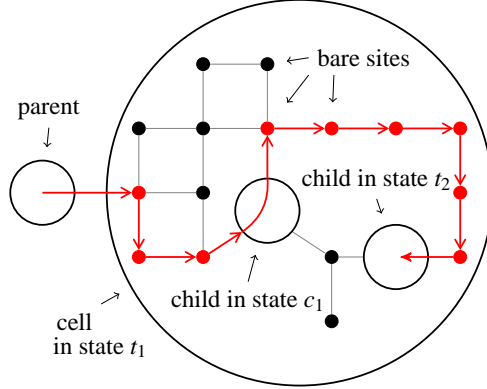


Figure 3.6.: Snapshot of the enumeration procedure for a cell with one link and children with three and two links, respectively. In this example, the walker has taken $n = 10$ steps (the one from the parent does not count). Cell and children are in states t_1 , c_1 , and t_2 , respectively, so the counter $Z_{c_1, c_2}^{t_1}[n]$ is incremented. (Figure reproduced from Ref. [104])

X and all children) that occurs, we now need a separate counter: $Z_{s_{i_1}, \dots, s_{i_k}}^{X, s_i}[n]$, where s_{i_j} denotes a specific state of child x_j and s_i a specific state of the cell. The counting is realized by three recursive functions, `cell_step`, `parent_step`, and `child_step`, which manage the enumeration steps on the different site types. Pseudo code for these functions is given in Fig. 3.7, Fig. 3.8, and Fig. 3.9, respectively. Note that lattice defects are implicitly realized by not listing the respective site as a neighbor. I have focused on readability rather than ultimate efficiency here, and the given pseudo code segments therefore differ somewhat from my actual implementation.

The enumeration process is initialized by calling the `parent_step` function with arguments `parent_link = none`, `n = 0`, and `state = [t0; c0, \dots, c0]`, where “none” means that all entries to the cell are open and the dummy cell state `t0` will transition to `c0`. The empty state `c_0` will thus always comprise exactly one segment of length zero. After the routine has terminated, the resultant arrays, $Z_{s_{i_1}, \dots, s_{i_k}}^{X, s_i}[n]$, need to be combined with those for the respective states of the children, $Z^{x_i, s_{i_j}}[n]$, which have been determined previously. We obtain the numbers of segments for each distinct state s_i of the cell:

$$Z^{X, s_i}[n] = \sum_{\{s\}_1} \dots \sum_{\{s\}_k} \left(Z_{s_{i_1}, \dots, s_{i_k}}^{X, s_i} * Z^{x_1, s_{i_1}} * \dots * Z^{x_k, s_{i_k}} \right) [n], \quad (3.8)$$

where $\{s\}_j$ means that the sum goes over all states s_{i_j} of the child x_j that have occurred, and “*” denotes the convolution product defined in Eq. (3.5). Since that operation is associative and commutative, the products can be evaluated successively in arbitrary order.

3. The taming of the SAW

```
cell_step(states, n, pos){
  // count segment
  Z[states][n] += 1;
  // measure end-to-end distance
  Rsq_acc[states][n] += Rsq[pos];
  // maximum length not yet reached?
  if(n < n_max){
    // block current position (self-avoidance!)
    pos = blocked;
    // go through neighboring sites
    for(all available neighbors){
      // neighbor is normal site?
      if(neighbor_pos within cell){
        cell_step(states, n+1, neighbor_pos)
      }
      // neighbor is link of child
      else if(neighbor_pos in child x){
        // identify link
        child_link = links[pos][neighbor_pos];
        // access allowed (child may transition to t-state)?
        if(allowed(cs_ts[x][child_link][states[c]]) == true){
          child_step(x, child_link, states, n+1);
        }
      }
      // neighbor outside of the cell (parent site)?
      else if(neighbor_pos external){
        // identify link
        parent_link = links[pos][neighbor_pos];
        // access allowed (cell may transition to c-state)?
        if(allowed(ts_cs[cell][parent_link][states[cell]]) == true){
          parent_step(parent_link, states, n);
        }
      }
    }
    // unblock current position
    pos = available;
  }
}
```

Figure 3.7.: Pseudo code of the function that is called whenever a normal site of the cell is accessed. The matrices `ts_cs` and `cs_ts` control the transitions from *t*-states to *c*-states and vice versa of the cell or a child. `child_step` and `parent_step` are given in Fig. 3.8 and Fig. 3.9, respectively.

3.2. Scale-free enumeration (SFE)

```
child_step(x, child_link, states, n){
  // backup current state of child x
  old_cstate=states[x];
  // transition to t-state
  states[x] = cs_ts[x][child_link][states[x]];
  // count segment
  Z[states][n] += 1;
  // go through (other) exits from child
  for(all other links of x){
    // exit allowed (child x may transition to respective c-state)?
    if(allowed(ts_cs[x][other_link][states[x]]) == true){
      // maximum length not yet reached?
      if(n < n_max){
        // get exit position
        neighbor_pos = positions[other_link];
        // child x transitions to next c-state
        next_states=states;
        next_states[x] = ts_cs[x][other_link][states[x]];
        // exit to normal site?
        if(neighbor_pos within cell){
          // site not yet visited?
          if(neighbor_pos == available){
            cell_step(next_states, n+1, neighbor_pos)
          }
        }
        // exit to another child?
        else if(neighbor_pos in child x2){
          // access allowed (other child may transition respective to t-state)?
          if(allowed(cs_ts[x2][other_link][states[x2]]) == true){
            child_step(x2, other_link, next_states, n+1);
          }
        }
        // exit to external (parent) site
        else if(external neighbor_pos){
          // access allowed (cell may transition to respective c-state)?
          if(allowed(ts_cs[cell][other_link][states[cell]]) == true){
            parent_step(other_link, states, n);
          }
        }
      }
    }
  }
  // child c reverts back to previous c-state
  states[c]=old_cstate;
}
```

Figure 3.8.: Pseudo code of the function that is called whenever a child is accessed.

3. The taming of the SAW

```

parent_step(parent_link, states, n){
  // backup current state of the cell
  old_tstate=states[cell];
  // transition to a c-state
  states[cell] = ts_cs[cell][parent_link][states[cell]];
  // count segment
  Z[states][n] += 1;
  // go through (other) entries to the cell
  for(all other links of cell){
    // entry allowed (cell may transition to respective t-state)?
    if(allowed(cs_ts[cell][other_link][states[cell]]) == true){
      // maximum length not yet reached?
      if(n < n_max){
        // get reentry position
        neighbor_pos = positions[other_link];
        // cell transitions to next c-state
        next_states=states;
        next_states[cell] = cs_ts[cell][other_link][states[cell]];
        // reentry to normal site?
        if(neighbor_pos within cell){
          // reentry site not yet visited?
          if(neighbor_pos == available){
            cell_step(next_states, n, neighbor_pos)
          }
        }
        // reentry to a child?
        else if(neighbor_pos in child x){
          // access allowed (child may transition respective to t-state)?
          if(allowed(cs_ts[x][other_link][states[x]]) == true){
            child_step(c, other_link, next_states, n);
          }
        }
      }
    }
  }
  // cell reverts back to previous t-state
  states[cell]=old_tstate;
}

```

Figure 3.9.: Pseudo code of the function that is called whenever the parent is accessed, i.e., when the walker leaves the cell.

Each summand in Eq. (3.8) represents the contribution from one specific state combination. Note that any c_0 -state has $Z[n] = \delta_{n,0}$, the identity element of the convolution product, so that children that were not accessed have simply no effect. Also note that for any c -state of the cell, all children also have to be in c -states, while for a t -state of the cell, at most one child may be in a t -state. For instance, for the number of c_1 segments through cell B (connecting both links) from Fig. 3.2 with children $x_1 = D$

3.2. Scale-free enumeration (SFE)

Table 3.5.: Number of SAW segments within cell B from Fig. 3.2.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$Z^{B,t_1}_{c_0, c_0}$	1	2	4	4	6	4	6	2	2	0	0	0	0	0
$Z^{B,t_1}_{c_0, t_3}$	1	2	4	4	6	4	6	2	2	0	0	0	0	0
$Z^{B,t_1}_{c_0, c_2}$	0	0	0	1	1	4	5	8	4	5	0	0	0	0
$Z^{B,t_1}_{c_0, t_5}$	0	0	0	0	0	0	0	0	2	0	2	0	0	0
$Z^{B,t_1}_{t_1, c_2}$	0	0	0	0	0	0	0	0	0	1	0	1	0	0
$Z^{B,t_1}_{c_0, c_3}$	0	0	0	0	2	2	7	6	6	2	1	0	0	0
.									.					
.									.					
.									.					
Z^{B,t_1}	1	2	5	9	20	31	56	67	112	123	191	208	275	289
Z^{B,t_2}	1	2	3	8	6	15	24	57	78	145	170	283	280	0
Z^{B,c_1}	0	0	1	0	2	0	2	0	8	0	28	0	24	0

and $x_2 = C$, Eq. (3.8) is simply (I drop “[n]” now):

$$Z^{B,c_1} = Z^{B,c_0} + \sum_{i=1}^3 (Z^{B,c_0,c_i} * Z^{C,c_i}), \quad (3.9)$$

while the number of t_1 segments (entry at l_1 of B) is given by

$$Z^{B,t_1} = \sum_{i=0}^3 Z^{B,t_1}_{c_0,c_i} * Z^{C,c_i} + \sum_{i=1}^6 Z^{C,t_i} * Z^{B,t_1}_{c_0,t_i} + \sum_{i=0}^3 Z^{D,t_1} * Z^{B,t_1}_{t_1,c_i} * Z^{C,c_i}. \quad (3.10)$$

The first factor of every summand in Eq. (3.10) represents the “*terminal*”, the location (the cell itself or a child) where the segments end for the respective combination of states.

After the children’s degrees of freedom have thus effectively been “integrated out”, all information concerning them can be deleted. Later on, the cell itself will assume the role of a child as we move on to enumerate the segments through the “grand parents”. For the example (cell B) a total of 33 different state combinations occur, some of which are represented in Table 3.5. Combining them with the arrays of the two children (Table 3.2 and Table 3.4) via Eq. (3.8) yields the numbers of segments for the three different states (now omitting c_0 again) of cell B shown at the bottom of Table 3.5.

Root cell (A): The root cell at the top of the hierarchy (A in the example) is a slightly special case: it has no parent, and the process is started by calling the `cell_step` function with the origin site (marked black in Fig. 3.2), length $n = 0$, and state t_1 as arguments. Otherwise, the routine proceeds as usual. For the example it produces four different arrays, one for each possible state of child B (c_0, t_1, t_2, c_1). These are

3. The taming of the SAW

then combined with the results from Table 3.5 via Eq. (3.8) to obtain the final results given in Table 3.6.

Table 3.6.: Number of SAW segments within the root cell A from Fig. 3.2, i.e., total number of conformations

n	0	1	2	3	4	...	12	13	14	15
$Z^{A,t_1}[n]$	1	4	4	6	10	...	468	535	813	895

3.2.2. Difficulties and optimizations

So far, the focus has been on the theoretical concepts of the algorithm. Now the more practical issues of the implementation shall be discussed, namely, how to keep computation time and memory requirements at bay. Regarding the computation time, there are two main bottlenecks. The enumeration process itself can be time consuming for cells that contain a large amount of *bare sites*, which will lead to a large number of segment conformations to be counted. As with the normal enumeration routine, the increase goes exponentially and can quickly get out of hand. The second bottleneck is calculating the convolution products in Eq. (3.8). This aspect is particularly important for longer SAWs, as the time increases with the square of the lengths of the arrays involved. The memory consumption will mainly be caused by the need to store the arrays for all the different state combinations that occur. These performance aspects are discussed further in Section 3.3.1.

The first paragraphs of this section discuss optimizations that proved critical for reaching the SAW lengths used in this work. Later paragraphs discuss optimizations that proved to be less important, or were not implemented at all.

Hashing state combinations: The large number of states that a cell might assume (see Fig. 3.4) is a central problem. Even (much) larger is the number of combinations of states that can occur during the enumeration process: for a cell with L links and k children with L_1, \dots, L_k links, respectively, the number of conceivable state combinations is

$$\Theta = (\Omega_c(L) - 1) \prod_{i=1}^k \Omega_c(L_i) + \Omega_t(L) \left(\sum_{i=1}^k \frac{\Omega_t(L_i)}{\Omega_c(L_i)} \prod_{j=1}^k \Omega_c(L_j) + \prod_{i=1}^k \Omega_c(L_i) \right), \quad (3.11)$$

with $\Omega_c(L)$ and $\Omega_t(L)$ from Eq. (3.6). This results from the observation that for any c -state of the cell, all children must also be in c -states, while for each t -state of the cell up to one child can be in a t -state. The “ -1 ” is because the c_0 state of the cell is not counted. Θ can quickly become dauntingly large: for instance, if a cell in $D > 2$ has 8 links and contains 3 children with 5, 6 and 7 links, respectively, (a realistic situation) there could be over 6.5×10^9 different state combinations. Hence, keeping one array (with $\sim N$ entries) for each possibility is clearly unfeasible. Fortunately though, it

3.2. Scale-free enumeration (SFE)

is also not necessary since only a small fraction can typically be realized in practice. For the given example, the number will more likely be around 10^5 . The best way to keep track of these arrays is by using hash tables: each combination that occurs is identified via a unique key index such as a vector whose entries are the states of cell and children, which is then hashed to a memory address. This hashing does produce some overhead, but the time for this operation is asymptotically constant with the number of entries.

Length caps: The fact that the length of SAWs considered is always restricted can be exploited at various points in the program. As mentioned, the maximum length `n_max` of the segments should take into account the (minimal) number of steps needed to reach the cell from the origin. More generally, it can incorporate the minimal number of (external) steps that would allow for the respective state, as well as the steps required for traversing the children. For instance, during enumeration of the segments within cell B from Fig. 3.2, any viable SAW realizing a state combination with $s^B = t_2$ and $s^C = c_2$ must take at least 3 steps outside of the cell (to get to l_2 in B) plus 3 steps for traversing child C (from l_3 to l_1), so that its length can be capped at `n_max=N-6=9`. We can easily include this capping by associating each state of the cell and each child with the minimum number of steps necessary to realize it and subtract these from `n_max`. The required steps within children are known exactly, but obtaining the minimal number of external steps is, in general, difficult and costly. Instead, I therefore only use lower limits, which are determined for all cells before the main program starts. The basic idea is to measure the shortest number of steps needed to externally connect each pair of links and then (for each state) to determine the combination of these pieces with the minimal total length that could be connected to a segment of the respective state. If this minimal external distance is larger than N (or if there is no connecting path at all), the transition to the state will simply be blocked. The limited length also plays a role when we connect the segments in Eq. (3.8): the longest segments for each state can usually only be combined with states with other segments that are relatively short without the maximum length being exceeded. Hence, we can often cap the maximum lengths when doing multiple convolutions.

Exploiting the bipartiteness: It is a specialty of the square lattice that the number of steps for paths connecting two specific sites must either all be odd or all be even. As a consequence, every second entry in the array for any c -state or any state combination where the segments terminate within a child are zero; see, for instance Table 3.5. By cutting out the empty slots and increasing the length by steps of two we can significantly reduce memory and time (mainly for the summations). One can even go one step further and only count ends on odd (or even) sites. In the end, one will thus only obtain the results for even (odd) lengths, which is not a big loss as we are mostly interested in the results for the maximum length. The gain for this last optimization did not turn out to be too significant for the enumeration of simple SAWs, and I ultimately only used it for SAWs with nearest-neighbor attraction (Section 3.2.5). In that case, it reduced the computation time and, more importantly, the required

3. The taming of the SAW

memory by almost a factor of two. One might exploit the bipartiteness even more by using the two-step method from Ref. [115]. This could speed up the counting process substantially but also make the code significantly more complicated than it already is.

Intermittent summations: On occasions, the memory required to store the arrays for all state combination can still become exceedingly large. It can therefore be useful to advance parts of the summation in order to free some of the memory. A good moment is whenever the enumeration program returns to the first instance of the `parent_step` function (Fig. 3.9), i.e., when all segments originating from one link are completed. At that point, none of the state combinations found so far is going to reoccur, and the partial summation can be done without hesitation. In desperate cases, it may be necessary to sum up somewhere during the enumeration process. This, however, is costly and the gain is uncertain. All previous combinations might reoccur shortly after and the memory can quickly fill up again. Still, this can make the difference and allow one to exactly enumerate all conformations on particularly difficult (e.g., supercritical or correlated) clusters.

Improved convolutions: The time needed for connecting the states [Eq. (3.8)] becomes more and more relevant with increasing chain length as each convolution involves $\sim n^2$ operations. This part should therefore be the main target of further optimizations. One possibility that I did not explore is the use of fast convolution algorithms. These rely on the circular convolution theorem and use the fast Fourier transform (FFT) algorithm; see, for instance, Ref. [116]. In principle, they should allow to reduce the complexity of the convolutions to $O(n \log n)$.

3.2.3. Mean end-to-end and chemical distances

So far, I have discussed how the SFE method can be used to determine the number of SAW conformations. Measuring other observables is straightforward as long as they depend on local properties of the conformation, i.e., if they can be evaluated separately in different spatial regions. Hence, one can easily measure the end-to-end distance but not the radius of gyration, which requires knowledge about the whole conformation at once.

Obtaining the mean squared end-to-end distance $\langle R^2 \rangle [n]$ is relatively simple and does not require much additional computational effort: We take the squared distances of the loose end for any segment conformation that terminates on a bare site of the cell and accumulate them into arrays: $R_{\text{acc } s_{i_1}, \dots, s_{i_k}}^{X, t_i} [n]$, where t_i is an arbitrary t -state of the cell and the meaning of the other indices is the same as in Eq. (3.8). This is done at each call of the `cell_step` function (see Fig. 3.7). As with the numbers of conformations, we need separate variables for each length and combination of states, but now only those combinations where the cell is in a t -state and all children are in c -states are relevant. After the segments through a cell are enumerated, the (not normalized) averages for the squared distances of the segments for each t -state of a cell are obtained by replacing the *terminal* factors in each summand in Eq. (3.8) with

3.2. Scale-free enumeration (SFE)

the respective accumulated end-to-end distances. Where the segments terminate on a bare site directly in the cell (all children are in c -states), $R_{\text{acc}c_{i_1}, \dots, c_{i_k}}^{X, t_i} [n]$ replaces $Z_{c_{i_1}, \dots, c_{i_k}}^{X, t_i} [n]$; where they terminate in child x_j , $Z^{x_j, t_{i_j}}$ is replaced by $R_{\text{acc}}^{x_j, t_{i_j}}$. For instance, the accumulated end-to-end distances for segments of the t_1 state of cell B from Fig. 3.2 [Eq. (3.10)] are obtained as

$$R_{\text{acc}}^{B, t_1} = \sum_{i=0}^3 R_{\text{acc}c_0, c_i}^{B, t_1} * Z^{C, c_i} + \sum_{i=1}^6 R_{\text{acc}}^{C, t_i} * Z_{c_0, t_i}^{B, t_1} + \sum_{i=0}^3 R_{\text{acc}}^{D, t_1} * Z_{t_1, c_i}^{B, t_1} * Z^{C, c_i}. \quad (3.12)$$

In the very end, we divide the results for the (only) state t_1 of the root cell A by their numbers to obtain normalized averages:

$$\langle R^2 \rangle [n] = \langle R^{2A, t_1} \rangle [n] = \frac{R_{\text{acc}}^{A, t_1} [n]}{Z^{t_1} [n]}. \quad (3.13)$$

The average chemical distances are obtained in the same manner.

3.2.4. Spatial distributions

For more detailed information than simple averages, such as the end-point distribution $P(\vec{R})$, one has to invest substantially more resources. In principle, we already count the segments ending at each single site, but keeping track of this information is onerous: for each state (combination) where the segments end within a cell, we will need two-dimensional arrays, $Z_{s_{i_1}, \dots, s_{i_k}}^{X, t_i} [n, \vec{R}]$, storing the number of ends on each site. This massively raises the required amount of memory, and one has to be thrifty in order to keep it manageable. The runtime also increases as the summation for a t -state [see Eq. (3.10)] needs to be performed for each site of the *terminal*. Of course, the other convolutions in Eq. (3.8), those non involving the terminal, do not depend on \vec{R} and should always be evaluated first. In the end, we normalize the distribution by dividing by the total number of conformations.

The sizes of the arrays storing the distributions depend linearly on the number of bare sites (typically < 50) for the state-combination arrays or (after decimation) by the total number of sites within the child (up to $O(M) \sim N^{d_i}$). If one does not need the probabilities as a function of \vec{R} but only of R^2 or of one coordinate, things will be a bit cheaper as sites with the same distance can be grouped together. One may also not need ultimate resolution, especially for very long walks, in which case one can make use of binning to further reduce the effort.

Another way to cut costs is by neglecting sites where relatively few walks end. This may sound somewhat sloppy, but if one wants to graphically investigate the distribution of SAW endpoints, neglecting sites whose values are below the resolution of the scale is not a real loss. To preserve the exactness for all other sites I proceeded in the following way: After the arrays for all states of a cell have been determined, I determine the maximal value of $Z^{X, t_i} [n, \vec{R}]$ (among all sites in the cell) for each state and length. I then go through all sites and “remove” them unless at least one of the values for the

3. The taming of the SAW

different states and lengths surpasses the respective maximal value times a threshold $t_{<} < 1$ which decides how much is neglected. Since the end-point distributions tend to be very strongly peaked (see Section 4.3.1) this trick can bring very significant gains.

One can also measure the fraction of SAW conformations passing through each single site, which I shall call the *conformation density* $P_C(\vec{R})$. This is slightly more complicated than measuring the end-point distribution and also requires a bit more computation resources (though the asymptotic complexity is the same). $P_C(\vec{R})$ is obtained in a similar fashion as $P(\vec{R})$: We use two-dimensional arrays, $C_{s_{i_1}, \dots, s_{i_k}}^{X, s_i} [n, \vec{R}]$, for each state combination (now including c -states of the cell) storing how many segments have passed through each (bare) site \vec{R} . For this we introduce a variable \mathbf{n}' that track of how many steps on bare sites have been performed by the current segment (\mathbf{n}' is smaller than \mathbf{n} from the code in Fig. 3.7 as it does not count steps on child sites) and an array $\mathbf{r}[\mathbf{n}']$ that stores these sites temporarily. Wherever $\mathbf{Z}[\mathbf{states}][\mathbf{n}]$ is incremented in the standard SFE routine, we now increment $\mathbf{C}[\mathbf{states}][\mathbf{n}][\mathbf{r}[\mathbf{i}']]$ for all $\mathbf{i} \leq \mathbf{n}'$. I shall not describe how to adapt the summation procedure for the conformation densities, which is fiddly but straightforward.

3.2.5. Thermal averages

So far it was assumed that each conformation ω contributes equally to the conformational average $\langle \dots \rangle$. This is the case as long as all have the same energy or if the temperature of the system is infinite. Otherwise, we have to weight conformations with their respective Boltzmann factors, $e^{-E_\omega/(k_B T)}$, where E_ω denotes a conformation's energy. For theoretical works, it is convenient to set $k_B := 1$, so that the thermal average of an observable, e.g. R^2 , is given by

$$\langle R^2 \rangle_T = \frac{\sum_{\omega} R_{\omega}^2 e^{-E_{\omega}/T}}{\sum_{\omega} e^{-E_{\omega}/T}}. \quad (3.14)$$

There are two different ways to obtain the thermal averages using exact enumeration: One is consider a fixed temperature T and weight each generated conformation according to its energy “on the fly”. The other approach is to count the number of conformations $Z[E_i]$ for each distinct (discrete) energy and their respective end-to-end distances, $R^2[E_i]$, and later calculate the average via

$$\langle R^2 \rangle_T = \frac{\sum_i R^2[E_i] Z[E_i] e^{-E_i/T}}{\sum_i Z[E_i] e^{-E_i/T}}. \quad (3.15)$$

This has the advantage that the results for any temperature can be obtained without redoing the enumeration and is therefore the obvious choice (see, for instance, Ref. [117]) when using the simple enumeration method from Section 3.1.2. However, for the SFE method, this comes at a high price: Each array storing the numbers or distances for a specific state or state combination (for each length) needs to be split into a set of arrays to keep track of different energy levels. This is in fact analogous to the handling of the spatial distributions discussed above. The summation process

3.2. Scale-free enumeration (SFE)

will also become significantly more demanding as any convolution [see Eq. (3.5)] will now involve a fourfold loop (over lengths and energies):

```
for(n_X=0; n_X <= N; n_X++){
  for(e_X=0; e_X <= E_max; e_X++){
    for(n_Y=0; n_Y <= N; n_Y++){
      for(e_Y=0; e_Y <= E_max; e_Y++){
        Z_XY[n_X+n_Y][e_X+e_Y] += Z_X[n_X][e_X]*Z_Y[n_Y][e_Y];
      }
    }
  }
}
```

Depending on the question, it may therefore be preferable to calculate the average on the fly for a fixed temperature. To do so, one simply weights at each incrementation (“+=”) in Figs. 3.7-3.9 with the Boltzmann factor for the respective segment.

Conformation energies: The SFE method can handle only energies that are defined locally. Long-ranged “Coulombic” interactions, for instance, do not permit the spatial factorization on which the method relies. However, there are a number of relevant cases that can be treated. In simplest one, each position is associated with an energy $\epsilon(\vec{R})$, which is straightforward to implement. This energy could, for instance, stem from interaction with defect sites.

One can also easily integrate stiffness by associating a bending energy with any change of direction, by passing the current direction as an argument to the `step` functions. This works for steps on and of child or parent sites just as for regular site, so that the program does not change otherwise.

More complicated (but also more interesting) is the case of SAWs with nearest-neighbor interactions, so-called self-attracting SAWs (SASAWs). Here each pair of adjacent monomers that are not consecutive in the chain contribute an energy $-\epsilon$. To enumerate SASAWs, we need to count the neighboring visited sites in each `step`, and increase a counter for the number (and average distance) of segments of the given length, state combination, and number of such *contacts*. The problem now is that walks segments may also interact across cell boundaries as depicted in Fig. 3.10 for a conformation of the familiar example from Fig. 3.2.

To accommodate for this, we have to split the states s_i up into different “*substates*” $s_{i,j}$, depending on which of the unused links is visited without exiting or entering the cell. In general, each c -state may thus split up into 2^x substates, where x is the number of pairs of neighboring, unused link sites, while for each t -state the number is (at most) 2^x times the number of remaining unused link sites. The substates are partially ordered by defining $s_{i,j} \geq s_{i,k}$ if all link sites visited in $s_{i,k}$ are also visited in $s_{i,j}$. The enumeration procedure is now modified in the following way: As before, a cell switches its state whenever the parent is accessed. In addition, it switches its substate whenever a link site within the cell is visited. Likewise, a child switches its state whenever it is visited and its substate whenever one of its link sites are passed

3. The taming of the SAW

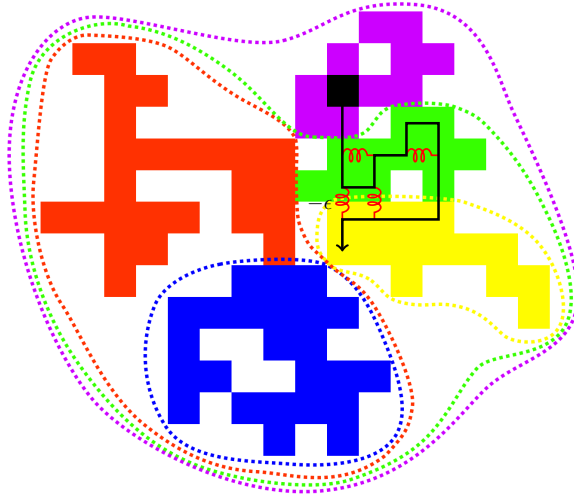


Figure 3.10.: Conformation of a SASAW (black) on the cluster from Fig. 3.2. The red coils symbolize attractions between adjacent monomers. Note that two of them cross the boundary between cell B (green) and cell C (yellow).

by from outside. We now have to keep track of different arrays for each combination of substates of cell and children, for which we count the numbers of segments (of different length and contacts numbers) and their average end-to-end distances. After the counting, the segments are connected in a similar manner as before [Eq. (3.8)]. Now, however, we combine (convolute) an array for a (sub)state combination where a child in substate $s_{i,k}$ with the sum of all substates $s_{i,j}$ of that child for which $s_{i,j} \geq s_{i,k}$, and we shift the energy by ϵ times the number of visited links of (the child's) substate $s_{i,k}$ to account for the cross-boundary interactions. In the same way, segments through one child need to be combined with those through another if they are bordering each other.

3.3. Partitioning the clusters

The factorization of the enumeration procedure described in the previous section requires the cluster to be partitioned into a tree hierarchy of nested cells as the one shown in Fig. 3.2. For a cluster consisting of a set $S = \{s_i\}$ of connected sites with SAW origin s_0 , I define a “cell hierarchy” \mathcal{H} as any set of subsets of connected sites with the following three properties:

1. $S \in \mathcal{H}$
2. $H_1 \not\subset H_2 \wedge H_2 \not\subset H_1 \implies H_1 \cap H_2 = \emptyset \quad \forall H_1, H_2 \in \mathcal{H}$

$$3. s_0 \in H \iff H = S \quad \forall H \in \mathcal{H}.$$

In other words, each cell must consist of connected (i.e., neighboring) sites that are not part of any other cell except the cell’s “ancestors”, i.e., cells that fully contain it, and its “offspring”, i.e., cells fully contained by it. This property establishes a partial order which defines the tree. At the top of this tree must be the full cluster (root cell), which is the only cell that contains the origin.

In theory, the SFE method works for any cell hierarchy, but there can be huge differences in the amount of time and memory it will require. If the hierarchy consists solely of the full cluster (which is valid), we essentially recover the basic enumeration method from Section 3.1.2. On the other hand, if it consist of cells with many *children* and many *links*, we have to deal with a large number of possible *state combinations* Θ [see Eq. (3.11)] demanding time [to evaluate Eq. (3.8)] and memory.

We hence need a method that can partition critical clusters into hierarchies for which the enumeration will need a limited amount of memory (depending on how much is available) and as little time as possible. Of course, the resources required by this “partitioning method” itself must also be taken into consideration. Before we can develop this method though, we first need to establish parameters by which we can estimate how good a particular hierarchy is, so that we know in which direction we want to go.

3.3.1. Evaluating cell hierarchies

The enumeration program deals with each cell separately. Hence, to estimate the total runtime, we should try to predict the time needed for the treatment of an individual cell, $\langle t_X \rangle$. This comprises the time for the actual enumeration, i.e., the runtime of the program outlined in Figs. 3.7, 3.8, and 3.9, and the time for connecting the segments [Eq. (3.8)]. The enumeration time will be roughly proportional to the total number of segment conformations generated, while the summation will roughly be determined by the total number of *state combinations* that have occurred. This second aspect will also govern the memory usage, so that minimizing the expected time should also reduce the required memory. The number of segment conformations correlates with the effective area of a cell, the number of *bare sites* plus the number of children. However, a better predictor turned out to be the number “*faces*” f , i.e., polygons formed by the edges of the graph on which the enumeration takes place. The graph in Fig. 3.6, for instance, has 5 *faces*. The heuristic motivation is that the faces correspond to possibilities for a path to split into two viable new paths. This would suggest an exponential dependence with a base of $b \approx 2$. The number of *faces* of a planar graph can be obtained immediately via

$$f = e - v + 1, \tag{3.16}$$

where v is the number of vertices (nodes), i.e., the number of bare sites plus the number of children and parent, and e denotes the number of edges between the nodes ($e = 23$ and $v = 19$ in Fig. 3.6). Eq. (3.16) can also be used for non-planar graphs (in $D > 2$), where the notion of *faces* is not clear otherwise. To test the dependence of $\langle t_X \rangle$ on

3. The taming of the SAW

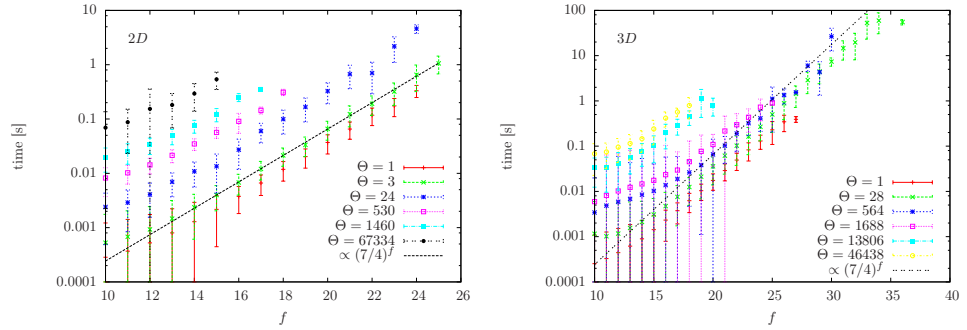


Figure 3.11.: Average time for the enumeration of SAW segments through a cell as a function of its number of faces in 2D (left) and 3D (right) plotted on a linear-log scale. The value of Θ from Eq. (3.11) characterizes the configuration of links and children.

f , I measured the average time needed to enumerate the SAW segments (with unlimited numbers of steps) for different cells with the same configurations of links and children (represented by the value of Θ) as a function of f . The results, which are plotted in Fig. 3.11², show that there is indeed an exponential dependence, albeit with $b \approx 1.75$ rather than $b = 2$ in both 2D and 3D (curiously so):

$$\langle t_X \rangle \sim 1.75^x, \quad (3.17)$$

Note that the error bars in the figure display the standard deviation from the mean and not the statistical error, so that they reflect the predictive power of the parameter f . They are relatively large for small f , where the times were too short to be properly clocked (hence also the flat start). Other parameters that I considered, namely the number of edges or vertices in a cell, lead to substantially larger deviations.

To estimate the effect of the number of links and children, let us assume that the time will, on average, increase monotonously with the number of *state combinations* that occur. This number is not known beforehand, but it correlates with the maximal number, Θ , which is given via Eq. (3.11). To investigate the significance of Θ , I measured the times (for enumeration and summation) for a sample of cells and calculated the averages for a fixed number of faces. The results can be seen in Fig. 3.12. Here the dependence seems roughly to be a power law, with an exponent that is also rather similar (≈ 0.6) in 2D and 3D:

$$\langle t_X \rangle \sim \Theta^x, \quad x \approx \begin{cases} 0.62, & D = 2 \\ 0.56, & D = 3. \end{cases} \quad (3.18)$$

²For this analysis I had to start with a preliminary, “best guess” version of the program, which was then refined iteratively. The data for the figures were produced with the final version of the program.

3.3. Partitioning the clusters

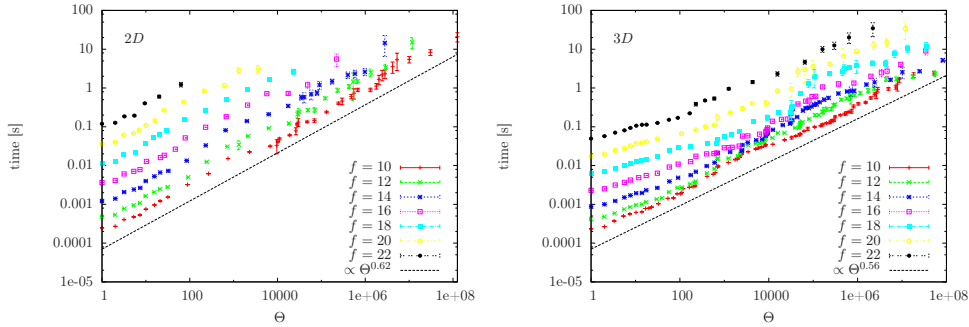


Figure 3.12.: Log-log plots of the average time for enumeration and summation of SAW segments through a cell with a fixed number of faces f as a function of the maximal number of state combinations Θ in 2D (left) and 3D (right).

My first guess for the total time was the sum of Eq. (3.17) and Eq. (3.18), which does work reasonably well. However, the fact that the data curves for different f in Fig. 3.18 (on a log-log scale) are roughly straight, equidistant parallels (in 2D at least) suggests that the product of Eq. (3.17) and Eq. (3.18) may actually be more relevant. The reason for this is not entirely clear, although one can see that the two terms cannot contribute independently: For a larger number of faces, more state combinations will be realized since the walks have more freedom to access the links. At the same time, the number of total segments being enumerated will approximately be multiplied by the number of such combinations. In practice, such a multiplicative approach lead indeed to better estimates. In the end, I used the following estimator to predict the time a cell X will require:

$$\langle t_X \rangle \sim b^{f_X} \Theta_X^x \quad (3.19)$$

with $b = 1.6$ and $x = 0.6$. These parameters proved to work well for both $D = 2$ and $D = 3$. The fact that $b = 1.75$ is not the optimal choice in practice is not too surprising, after all the above motivations are very heuristic and could not be expected to yield perfect predictions. In practice, I also use a lower cut-off for the factor b^{f_X} to reflect the constant overhead reflected in the leveling or the curves in Fig. 3.11 for low f .

The cases $D > 3$ did not warrant special treatment, as the enumeration is effort almost negligible for them; see Section 3.4.2. The parameters are adapted slightly for different versions of the program, e.g., when energy histograms are computed (Section 3.2.5), usually by increasing x and reducing b to promote larger cells with fewer links.

The total expected time for a particular cell hierarchy \mathcal{H} can now be estimated as

$$\langle T_{\mathcal{H}} \rangle = a \sum_{\{X_i\}_{\mathcal{H}}} \langle t_{X_i} \rangle, \quad (3.20)$$

giving us a measure by which to compare the quality of different hierarchies (the factor

3. The taming of the SAW

a does not matter for this). The next question is how to create a hierarchy for which $\langle T_{\mathcal{H}} \rangle$ is as small as possible.

3.3.2. Finding the right hierarchy

It is clear that any good hierarchy must reflect the self-similarity of the system. Cells on different levels should statistically look the same once the children are *decimated*. It is therefore natural to look for a partitioning method that also operates in a self-similar fashion. Broadly, one might think of two different strategies: a *top-down* approach where one starts by dividing the full cluster into weakly connected pieces, which are then subdivided in the same manner, or a *bottom-up* approach where one starts out with small pieces that are repeatedly merged into larger ones. I opted for the latter route, for no particular reason. The basic ideas for the algorithm outlined below.

Amalgamation strategy:

- We begin with a cluster fragmented into many small “*primary pieces*”³, which are not much larger than individual sites.
- We then consider all options to merge (a group of) neighboring *pieces* into a larger one. Each option is rated according to a set of criteria that reflect how tightly the pieces involved are bound together.
- The merging option with the best rating is carried out and all resulting new options are assessed. This step is repeated until only one piece is left. A tree is constructed during the process, where the children of each node (piece) are the smaller pieces from which it was formed.
- When only one piece is left, the tree is “uprooted”, so that the *primary piece* that contains (or is) the *origin* becomes the new *root*.
- In the end, some (actually most) of the *pieces* are dismantled: they are removed from the tree, and the children are adopted by their grand parent. In particular, all *primary pieces* are scrapped to become *bare sites*. The *Pieces* that remain then become the cells of the hierarchy.

Initialization: In principle, we could start with the individual sites as *primary pieces*. To speed things up a bit, however, sites with only one neighbor are directly joined to it. Sites with only two neighbors are joined to either one of them unless if this would close a loop, i.e., create a *face*. The process goes on until no more sites can be joined. For the familiar example from Fig. 3.2, this results in an initial setting as shown in Fig. 3.13. The shapes of the primary pieces depend on where one adds the “irrelevant” sites, but the corresponding connectivity graph is unequivocal. Note that a *primary piece* may contain a *face* if it is connected to one neighbor only, as

³These *pieces* are similar to our *cells*; some will indeed become *cells* eventually. However, the two concepts are not exactly identical, so I use a different term here.

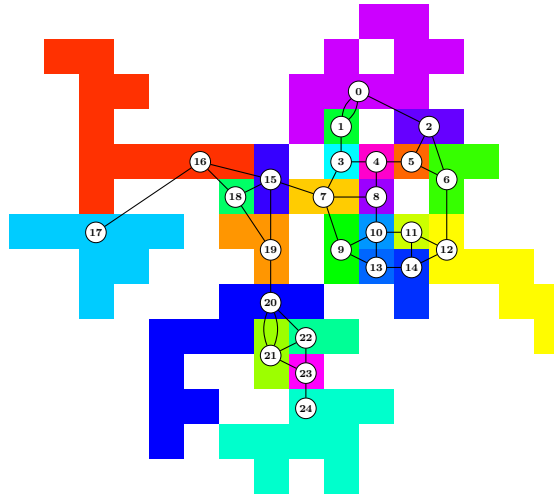


Figure 3.13.: *Primary pieces* (randomly colored) and corresponding connectivity graph for the exemplary cluster from Fig. 3.2.

is the case for pieces 17 (light blue) and 24 (turquoise) in the figure. For each *piece*, we now consider all options to merge it with any subset of neighboring pieces that are connected among themselves. All options are rated (see below) and sorted into a priority queue accordingly.

Rating the options: The parameters by which we rate the different merging options should favor the creation of *pieces* (and hence *cells*) with few links, i.e., few edges linking to the respective node in the connectivity graph. The most relevant parameter thus turned out to be the “link creation number” Δ , defined as the number of links of the resultant pieces minus the number of links of the piece from which the options are considered. $\Delta < 0$ means that the merger is good as it reduces the connectivity of the graph. For instance, merging piece 15 (dark blue, 4 links) with piece 7 (dark yellow, 4 links) in Fig. 3.13 would create a new piece with 6 links, so the choice is rather poor ($\Delta = 2$), while merging 0 and 1 would be good ($\Delta = -1$). A second important parameter, which in my implementation is consulted whenever there is no option with $\Delta \leq 0$ is the *height* h of the pieces in the emerging tree, i.e., how many generations of offspring the resultant piece would contain. Preferring mergers with a low h supports an evenly growth of the pieces, thus preventing that some small areas will remain very fragmented and obstruct the whole process. Other aspects can also be taken into account, but these two turned out to be most important by far. Each potential merger also receives a random number that decides the case when mergers are rated equally otherwise.

3. The taming of the SAW

Amalgamation process: The main part of the partitioning program consists of the following routine: We take the best-rated merger option from the beginning of our queue (see above) and check if all involved pieces are still *active*, i.e. if they have not already been merged. In that case the option is discarded and we pick the next best until we find one that is up to date. We then define the pieces involved as children of a larger piece consisting of their combined sites, and we determine the number of *faces* f_0 created within the new piece as the number of links between its children plus one minus the number of children [cf. Eq. (3.16)]. After the merger has been effected, we consider all options to merge the new piece with its neighbors, rate them, and sort them into our queue. Figure 3.14 shows the active pieces and the connectivity graph for the example after 2, 5, and 10 mergers have been performed. The procedure is repeated over and over, and similar pieces of ever larger sizes appear until only one is left. The self-similar nature is evident in Fig. 3.15, where the active pieces of the backbone of a larger cluster cutout (500 chemical shells) are shown during different stages of the process.

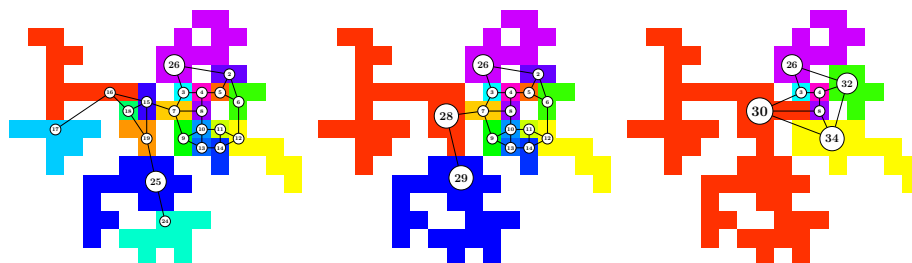


Figure 3.14.: Active pieces after 2, 5, and 10 amalgamation steps (mergers).

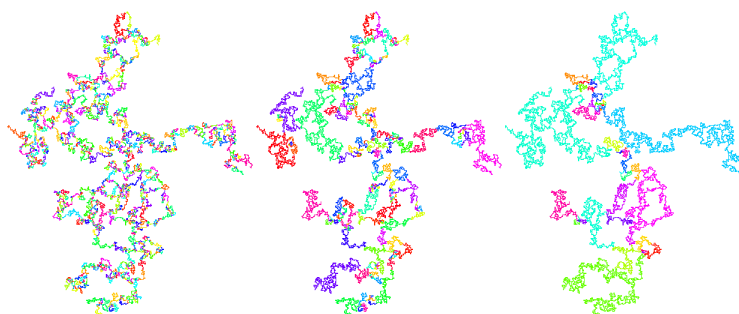


Figure 3.15.: Stages during the amalgamation process for the backbone of a CPC cutout of 500 chemical shells around the origin. Shown are the active pieces after 5000, 7500, and 7900 mergers have been performed (from left to right). The figure is taken from Ref. [103]

3.3. Partitioning the clusters

Uprooting the tree: One condition for the *cell hierarchy* is that (only) the root must contain the origin site, and so we have to rearrange our tree accordingly. The primary piece that contains the origin is put on top, and its former parent (the piece into which it was merged) and siblings (the pieces that were merged with it) are defined as its children. This means that the former parent piece is inverted: sites that used to belong to it now become the outside. Similarly, its former parent and siblings become its children. This inversion is carried out for the whole ancestry.

Finalizing: The tree is now almost ready to become our cell hierarchy and should already reflect the connectivity of the cluster on all length scales. However, only the primary pieces contain any *bare sites*. This area where the basic walk segments will actually be enumerated should be distributed more or less evenly throughout the hierarchy. To achieve this, some of the pieces will be *dismantled*, including all the primary ones. These simply become bare sites contained by their parents. For any other piece that is dismantled, we define that all its children and bare sites are passed on to its parent and remove it from the hierarchy. The number of faces in the dismantled piece are also added to the parent's. We choose the pieces to be dismantled such that the total estimated time [Eq. (3.20)] becomes minimal. This can be done efficiently [presumably $O(n \log n)$ in number of pieces] by making use of the tree structure: First, we dismantle all primary pieces. Then we estimate the time for all⁴ pieces that contain no other pieces (*leaves* of the tree) according to their number of *faces* and *links* via Eq. (3.19). For the pieces on the next level, we consider dismantling any combination of their children including all and none. For each combination the time is estimated from the number of faces and the Θ -value of the potential cell plus the times estimated for the remaining (“*spared*”) children. The combination with the lowest estimate is recorded; it represents the best choice as long as the piece itself is not dismantled. However, to find the global optimum for whole the hierarchy, we must also think of the scenario where the piece is dismantled and keep an “combination list” of all combinations that could somehow turn out to be optimal. The case where the piece is kept is stored as the “*zeroth entry*” of the combination list. Each entry of such a list encompasses the following items:

- a. The combination of children to be dismantled and references to entries in their respective *combination lists*. For instance, if a piece has three children, the combination $\{2, 1, 0\}$ would refer to the second entry from the list of the first child combined with the first entry of the second child's list. The value zero signifies that the third child is not to be dismantled. Such a combination fully specifies which pieces are dismantled within the entire branch below the piece. The combination of the *zeroth entry* represents the optimal choice if the respective piece is kept.
- b. The sum of the (best) time estimates for all *spared* offspring o_i , $\sum_{i=1}^k T_{o_i}$ (which includes the times for their respective offspring). For the *zeroth entry*, the time estimate for the piece itself [Eq. (3.19)] is also added.

⁴In practice, the pieces are actually dealt with in pre-order, too.

3. The taming of the SAW

- c. The number ϕ of faces that are passed on when the piece is dismantled, i.e., that are outside of the spared offspring. This value is set to zero for the zeroth entry.
- d. A number reflecting the possible c -state combinations of the spared offspring:

$$\theta^c := \prod_{i=1}^k \Omega_c(L_{o_i}), \quad (3.21)$$

where L_{o_i} denotes the number of links for the i th spared offspring, or $\theta^c := 1$ if the piece does not have children. For the zeroth entry, this is replaced by $\Omega_c(L)$, the number of c -states of the potential cell [Eq. (3.6)].

- e. A number reflecting the possible t -state combinations of the spared offspring:

$$\theta^t := \sum_{i=1}^k \left(\frac{\Omega_t(L_{o_i})}{\Omega_c(L_{o_i})} \prod_{j=1}^k \Omega_c(L_{o_j}) \right), \quad (3.22)$$

or $\theta^t := 0$ if there are no children. Together with θ^c , this number will be used to calculate the maximal number of state combinations Θ [Eq. (3.11)]. For the zeroth entry, it is replaced by $\Omega_t(L)$.

For any combination where the piece is dismantled (not the zeroth entry), the values b. and c. are obtained by simply summing the respective values from the lists of the children, while θ^c and θ^t are calculated via the following recursions:

$$\theta^c = \prod_{i=1}^k \theta_{x_i}^c \quad (3.23)$$

and

$$\theta^t = \sum_{i=1}^k \left(\frac{\theta_{x_i}^t}{\theta_{x_i}^c} \prod_{j=1}^k \theta_{x_j}^c \right), \quad (3.24)$$

where $\theta_{x_i}^c$, $\theta_{x_i}^t$ denote the values from the selected list entry of the i th child.

The time for the branch if the piece is spared (item b. of the zeroth entry) is obtained by summing the respective b.-values from the children plus the estimated time for the envisioned itself from Eq. (3.19). The number of faces required for this estimate is obtained as

$$f = \sum_{i=1}^k \phi_{x_i} + f_0 + L - 1, \quad (3.25)$$

where ϕ_{x_i} are taken from the children's lists (c.), L is the number of links to the parent, and f_0 was determined during the amalgamation process. The Θ -value is calculated via

$$\Theta = (\Omega_c(L) - 1) \prod_{i=1}^k \theta_{x_i}^c + \Omega_t(L) \left(\sum_{i=1}^k \frac{\theta_{x_i}^t}{\theta_{x_i}^c} \prod_{j=1}^k \theta_{x_j}^c + \prod_{i=1}^k \theta_{c x_i} \right), \quad (3.26)$$

3.3. Partitioning the clusters

which is equivalent to Eq. (3.11) when the θ s are substituted by the number of state combinations for the spared offspring. The zeroth entry of the *root* will in the end refer to the optimal combination of pieces to be dismantled/kept in the whole hierarchy.

The computational effort for this routine remains manageable since we do not need to store every combination; only those that could potentially be optimal. Hence, if all parameters b.-e. are equal or worse (i.e., larger) for one combination than for another one in the list, it can be discarded. In addition, we can introduce upper thresholds for the values, so that the length of the lists will remain bounded. This is permissible as there are practical limits anyway to size and complexity of cells. For instance, the number of faces in a cell should in practice not be much larger than thirty as the time to enumerate the segments will become much too long otherwise. If chosen generously, these bounds could hence only make us overlook an optimum that would still not be good enough anyway.

Factorizing the partitioning: Due to the randomness involved, one can repeat the amalgamation process a couple of times to improve the outcome. This works reasonably well so that clusters large enough for over a thousand SAW steps can be handled even in $2D$, where the decomposition is most difficult. There is, however, one flaw that becomes relevant if we want to push the length to the limit or make meaningful measurements of the asymptotic time complexity of the enumeration method: As the clusters get larger, the chances for finding an exceptionally good hierarchy by repeating the amalgamation with different seeds dwindle. In some areas of the cluster, the partitioning will be better in others worse, but on average it will probably be average. Hence, we would need to increase the number of trials exponentially with the cluster mass to retain the quality of the hierarchies when we increase the system size, which is not practical. On the other hand, fixing the number of trials will spoil the time scaling of the enumeration method.

The solution is to self-consistently factorize the amalgamation process itself. This can once again conveniently be achieved by a recursive approach: As before, we start amalgamating the smallest pieces. But whenever we create an “ideal” piece that we certainly want to keep (for example, with only one external link), we try to optimize the partitioning within that piece before we move on. For this we recursively apply the same routine several times within the restricted area of the ideal piece. (We can think of the whole cluster as the largest ideal piece.) When we retry the amalgamation within an ideal piece with a different seed (for the random number generator), we start with the largest ideal pieces contained by it instead of the *primary pieces*, significantly accelerating the process. Each ideal piece is thus optimized independently from the rest of the cluster, and with each repetition, the amalgamation will become faster as more ideal pieces are identified. I defined a piece as *ideal* if

1. more than half of the links of each of its constituents (pieces that are merged to form it) connects to other constituents,
2. it has fewer than L_{id} links, where L_{id} is initially 2 and slowly increased as the amalgamation routine is called repeatedly.

3. The taming of the SAW

Developing these criteria involved some heuristics and empirical trials, which I shall not elaborate here.

3.4. Checking and benchmarking the SFE method

In this section I discuss issues to do with the practical application of the SFE method: verifiability of the results, scope and limitations, problems, and performance.

3.4.1. Preliminary considerations

Error detection: Since the SFE method is relatively complex, there is a serious risk of bugs. Most would have obvious consequences, but others may corrupt the results more subtly. This raises the question of how far the results produced by the SFE method can be trusted, or how we can verify them given that no other method can handle similar chain lengths.

First, one can always check the results for the first thirty or so steps using the standard enumeration method from Section 3.1.2. This requires very little extra time and was therefore done at any execution of the program. Not a single mismatch occurred in over 10^8 runs with the final implementation, by which the results of this work were produced. This strongly indicates that it works correctly and that the results for more steps can also be trusted. (Note that there is in principle no difference between the handling of initial and later steps.) I also used PERM to directly verify the results for SAWs of a couple of hundred steps and found excellent agreement as long as sufficiently many PERM tours were started; see Section 3.5. However, for even longer chains PERM itself becomes unreliable. A better approach takes advantage of the fact that the results must be independent of the *cell hierarchy* into which a cluster is partitioned: The amalgamation method described above produces slightly different hierarchies when run with different seeds, and one can lend a hand by tweaking some parameters. The resulting differences will be strongest in highly connected cluster regions where cells have many links, i.e., where the enumeration is most complex and prone to errors. In case of an error, it is highly unlikely that the results for different hierarchies are affected in the same way. Hence, if the results are identical, they are probably correct. This check was carried out for several thousand of the largest systems and did not find any mismatches beyond numerical inaccuracies, which do not affect the first 14 digits for my choice of data types. (I use exact integers during the counting and 128-bit floats for the combined state arrays. It is possible to use exact data types throughout, but the cost in performance and memory is high and the gain irrelevant.) I am hence highly confident that my latest implementation and all the results generated by it are correct.

Uneven usage of resources: One peculiarity of the SFE method is that its time and memory requirements depend on the individual cluster configurations and can fluctuate quite strongly. This is because the resources needed to deal with one *cell* depend very sensitively on its properties. In theory, there will in fact always be some clusters

3.4. Checking and benchmarking the SFE method

that cannot be partitioned such that the enumeration (of SAWs of decent length) can realistically be performed. Fortunately, these cases practically never seem to occur if we are at the percolation threshold. Still, some clusters are always significantly more demanding than others, making resource allocation for the program somewhat awkward. To estimate the quenched averages for the largest systems ($N = 12800$) I had to restart the program with extra resources for some of the clusters. In very few cases, I had to try the partitioning with different parameters until the enumeration would not surpass the limitations of the system. Simply discarding such more difficult cases would introduce a bias, as will be seen in Section 3.4.2. However, in many situations this bias would be negligible in practice.

Theoretical complexity: Theoretical assessment of the asymptotic complexity of the SFE method is difficult, but one can at least estimate some upper and lower bounds. Obviously, the expected time has to increase with the area A covered by the SAWs, implying (see Section 2.2.2)

$$T \gtrsim A \sim N^{d_1}. \quad (3.27)$$

We can expect the number of cells in a hierarchy to increase $O(N^{d_1})$ as well since the total number of cells is always less than twice the number of cells on the lowest level of the tree. However, the time needed to deal with a cell should on average increase with its total area, i.e., the total number of its offspring. This is mainly because the convolution of SAW segments through a cell with the segments through its children, Eq. (3.8), increases roughly with the square of the total number steps a SAW can take within cell and children. Assuming that this number is given by the total area of the cell (including children), which is obviously an upper bound, and given that the total cell areas increase exponentially with the level (height) h in the hierarchy while their number decreases, we obtain

$$T \lesssim \int_0^{\ln A} e^{-h} A (e^h)^2 dh \sim A^2. \quad (3.28)$$

Together with Eq. (3.27) we thus have

$$N^{d_1} \lesssim T \lesssim N^{2d_1}. \quad (3.29)$$

It was assumed here that the cluster can be partitioned such that the quality of the cell hierarchies does not deteriorate when the system size is increased, i.e., that cells in larger system do not tend to have more links, children and faces. The partitioning method introduced above seems to achieve that (as long as $p = p_c$) although I have no theoretical proof for this. Of course, the time for the partitioning must also be factored in, but it does not seem to matter asymptotically as shown below.

3.4.2. Performance of the SFE method

The performance was analyzed empirically for the simplest and fastest version of the SFE, which measures the number of chain conformations and their mean end-to-end (or

3. The taming of the SAW

chemical) distance. The method relies explicitly on the properties of critical clusters (self-similarity and finite ramification), but up to some finite length it can also be used somewhat above p_c . The following discussion is limited to the cases $D \leq 5$; for $D = 6, 7$ the time for enumeration and partitioning is negligible compared to the time needed to create the clusters.

At criticality: I used the SFE method to enumerate SAWs of different length, $N = 50, 71, 100, \dots$ (increasing by factors of \sqrt{t}) on critical clusters in $D = 2-5$ on identical machines⁵ to benchmark the performance. The maximum length was $N = 12800$ in $2D$ and $3D$ and $N = 4525 / N = 3200$ in $4D/5D$, respectively. For each length and dimension, I used a sample of 10^3 randomly generated clusters. The average times required for partitioning a cluster and enumerating the SAWs are plotted in Fig. 3.16 as a function of the number of steps on a log-log scale. For comparison I also plotted the time needed to generate the clusters (see Appendix A.1.1). The straight slopes for the enumeration time clearly indicate power-law behavior rather than the exponential increase which is typical for exact enumeration methods. I used a simple least-squares fit to estimate the exponents; the results are given in Table 3.7. Curiously, the exponent for the enumeration time is very similar in all dimensions (≈ 2.4). I have currently no explanation for this finding; it may well be a coincidence. All values are within the theoretical boundaries from Eq. (3.29). The amplitudes, meanwhile, strongly diminish with increasing dimensionality as the clusters become more and more loopless. For $D > 3$, the time needed to generate the clusters clearly dominates in the range investigated. The same goes for the memory usage, which I therefore only investigated in $D \leq 3$. As can be seen in Fig. 3.17, the peak memory usage in $D \leq 3$ appears to increase polynomially as well. The estimated exponents are also listed in Table 3.7. The peak is usually reached during the enumeration part in $D = 2$ and during the partitioning in $D = 3$. There are some unfortunate cases where the required memory (and time) is considerably larger than the average. However, the deviations are not so extreme that a single configuration has significant impact on the average.

⁵Intel Xeon E5-2640 (2.5GHz)

3.4. Checking and benchmarking the SFE method

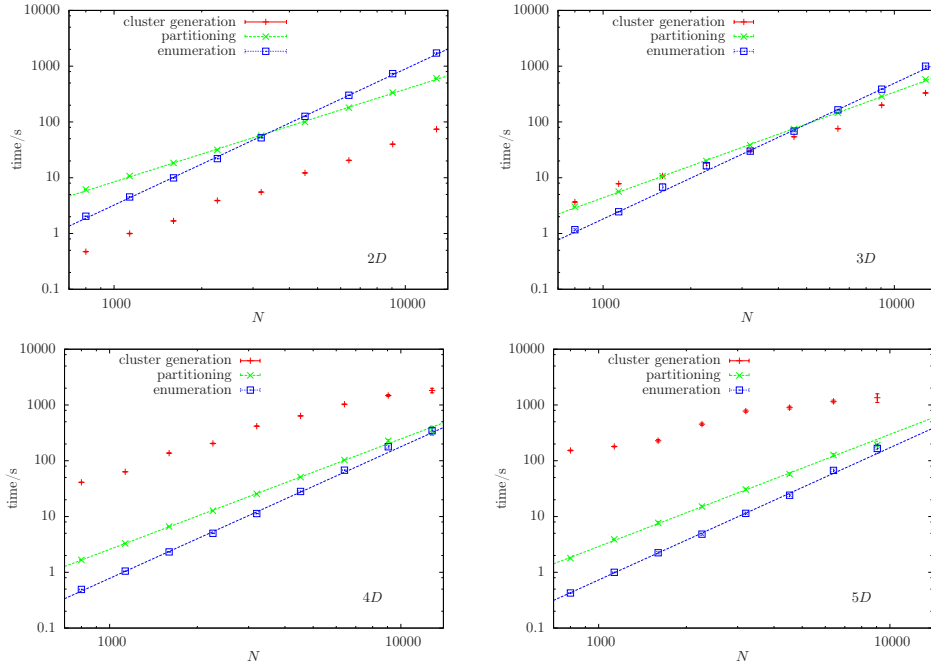


Figure 3.16.: Average computation time for the different parts of the SFE program vs. number of steps (for dimensions $D = 2-5$). The dashed lines are power-law fits; the exponents are given in Table 3.7. (For $D > 2$, the lattice sizes are increased sublinearly with N , hence the apparent saturation of the cluster generation times.)

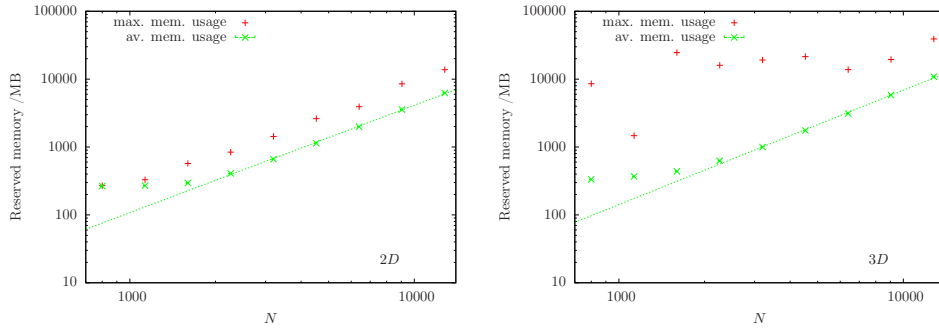


Figure 3.17.: Average and maximum peak memory usage of the SFE method measured for a sample of 10^3 clusters in $2D$ (left) and $3D$ (right).

3. The taming of the SAW

Table 3.7.: Measured exponents of the time and memory increase for the SFE method on critical clusters in different dimensions.

D	enumeration time	partitioning time	peak memory
2	2.44(2)	1.66(2)	1.58(2)
3	2.43(2)	1.89(2)	1.26(2)
4	2.36(4)	1.98(2)	-
5	2.37(4)	2.01(4)	-

Above criticality: Above p_c , the clusters lose their critical properties. This can be seen in Fig. 3.18, which shows the *bi-connected components* (see Section 2.2.3) of the relevant cutout regions for SAWs of $N = 400$ steps on $2D$ clusters at various concentrations p . While the sizes of these vary on all length scales at criticality, there is one dominant component above p_c , and with increasing system size more and more cuts will be required to disconnect the cluster into separate pieces of similar size. The SFE method can therefore no longer achieve polynomial complexity on these supercritical clusters. However, up to some point (line, actually) $N(p)$ it can still compete with other methods. As it turns out, we can even reach significantly into regimes where the fractal nature of the clusters is no longer apparent.

I analyzed the performance of the method on supercritical ($p > p_c$) clusters in $D = 2$ and $D = 3$, for concentrations of up to $p = 0.70$ ($\approx 1.18 \times p_c$) in $2D$ and $p = 0.35$ ($\approx 1.12 \times p_c$) in $3D$. The results (mean times for the enumeration) of this benchmarking are plotted in Fig. 3.18. For some cases, the method did not succeed to partition the clusters such that SFE could be carried out within the memory restrictions (64GB RAM) or within realistic time. The rate of these “failures” increases with concentration and length. This can be seen in Table 3.8, where the fail rates and median times are listed for the “border-line” cases where the rate is larger than zero but smaller than 50%. These failures are not regarded in the average times, which is why I listed the medians in the table instead. Discarding configurations where the method fails will not necessarily spoil the quenched averages as long as they are relatively few. However, if the fail rate is significant, one has to be very careful, in particular for small systems (with high p): the chances of success tend to be lower for clusters that are very compact, which correlates positively with Z and negatively with $\langle R^2 \rangle$. For instance, on the 67% of clusters where SFE succeeded for $D = 2$, $p = 0.7$, $N = 71$, the average squared end-to-end distance was determined as $[\langle R^2 \rangle] = 599(2)$, whereas for the remaining 33% I obtained (with PERM) $[\langle R^2 \rangle] = 572(3)$. This discrepancy is not due to inaccuracy of PERM, which yielded results perfectly consistent with those from SFE for the other 67%.

3.4. Checking and benchmarking the SFE method

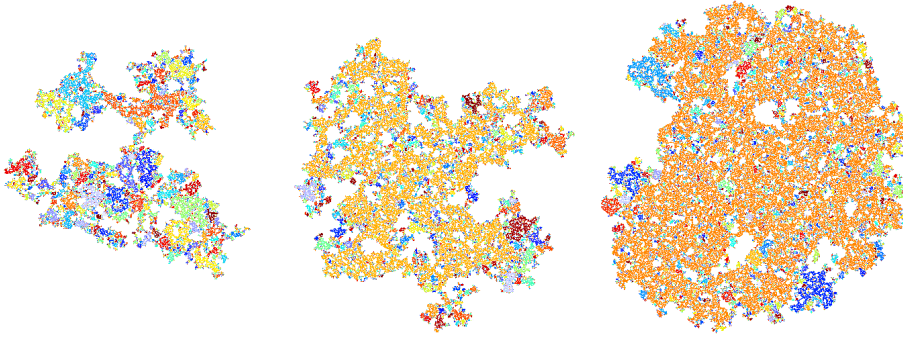


Figure 3.18.: Bi-connected components of the cluster region which is accessible within $N = 400$ steps for concentrations $p = 0.59274621 \approx p_c$, $p = 0.60$, and $p = 0.62$ (from left to right) for a typical incipient cluster on a 2D-lattice.

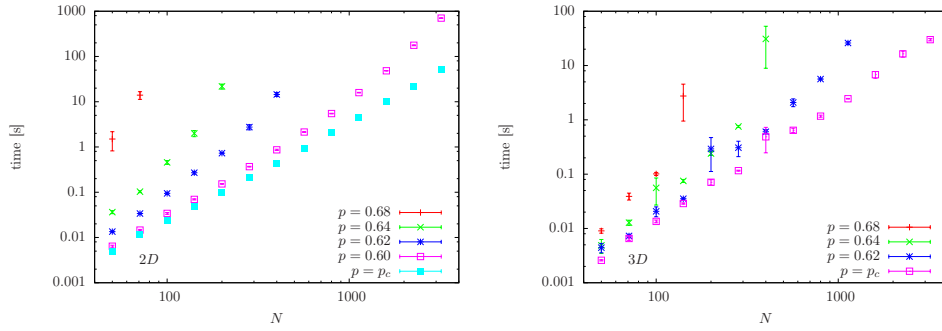


Figure 3.19.: Average time for the enumeration on supercritical clusters in 2D and 3D.

Table 3.8.: Lengths at which the SFE method starts to falter for different concentrations in 2D and 3D besides the respective fail rates and median times \hat{t} . At criticality, the method succeeded for all clusters investigated.

2D				3D			
p	N	fail rate	\hat{t} [s]	p	N	fail rate	\hat{t} [s]
p_c	12000	< 0.002%	2236	p_c	12000	< 0.002%	934
0.60	3200	5%	646	0.315	800	0.01%	7.1
0.61	800	0.01%	29.1	0.315	1131	2.6%	22
0.61	1131	12%	185	0.32	400	7.7%	3.0
0.62	400	0.01%	10.6	0.33	50	1.0%	0.42
0.62	566	18.4%	161	0.35	71	15%	0.53
0.64	200	3%	13.3				
0.68	71	12.2%	3.4				
0.70	50	1.6%	2.3				
0.70	71	33%	89				

3.5. Analyzing PERM

Having exact results to compare with allowed me to closely investigate the accuracy of PERM in dilute systems. Since PERM has become a standard MC method, such an analysis is justified in its own right, but I also had more concrete motivations: First, I wanted to be in a better position to compare the SFE method with PERM; see Section 3.6. Second, I wanted to have a clearer idea of how many tour starts are necessary to obtain reliable results with PERM, which was used to generate part of the results in Chapter 5.

There are two parameters by which the accuracy (and runtime) of PERM can be modified: the number of *tours* t and the population-control parameter $C = \sqrt{C^>C^<}$ [see Eq. (3.3)]. The program self-consistently adjusts the weights such that if it runs long enough, $1/C$ branches will on average be completed per tour. The runtime will therefore roughly increase with the *effective number of tours* $\tilde{t} := t/C$.

To analyze the algorithm, I fixed the population control to $C = 0.001$ and varied the number of starts. It should be noted that with the improved nPERMss version on which my implementation is based, lowering C is often more rewarding than increasing t : a small C means the initial part of the chains are exactly enumerated rather than sampled repeatedly, which enhances both performance and accuracy; see Ref. [111]. However, varying t is more convenient for the analysis here as it can be done in one run of the program.

The statistical errors on the conformational average estimated with PERM asymptotically vanishes like $\epsilon \sim \tilde{t}^{-1/2}$. However, this quantity is usually not very relevant as it will be flattened out by the disorder average. This is not the case for systematic errors, which can arise if too few tours are started. Their main cause is that PERM sometimes fails to discover areas that are difficult to enter even though their entropic contributions are relevant. This is exemplified for the cluster in Fig. 3.20 (left), on which I used PERM to estimate the average end-to-end distance. The right-hand side of the figure shows the results after varying numbers of tours. They can be seen to converge nicely towards exact values determined by SFE (black curve) for $N \leq 300$ but not for $N = 400$. The errors are estimators of statistical error (see Ref. [102]):

$$\epsilon = \sqrt{\frac{\langle \widehat{R^2}^2 \rangle - \langle \widehat{R^2} \rangle^2}{t} \frac{\langle \widehat{w}^2 \rangle}{\langle \widehat{w} \rangle^2}}, \quad (3.30)$$

where the hats denote averages over branches from individual tours. They clearly fail to cover the deviations as the fluctuations are not correctly assessed.

More detailed information is provided in the distribution of end-point locations and the *conformation densities* (see Section 3.2.4) shown in Fig. 3.21 and Fig. 3.22. The pictures on the right side were obtained by exact enumeration. They reveal that the majority of chains end in a region relatively close to the center. After the first 10^5 effective tours, PERM did in fact not produce any conformation ending in that region, and even for $\tilde{t} = 10^8$ other regions are emphasized more strongly as can be seen in the figures. Interestingly, the problem here is not that the chains fail to enter the region

(those of length $N=200-300$ mostly end there) but rather that they go there too early and are then forced out as they grow longer.

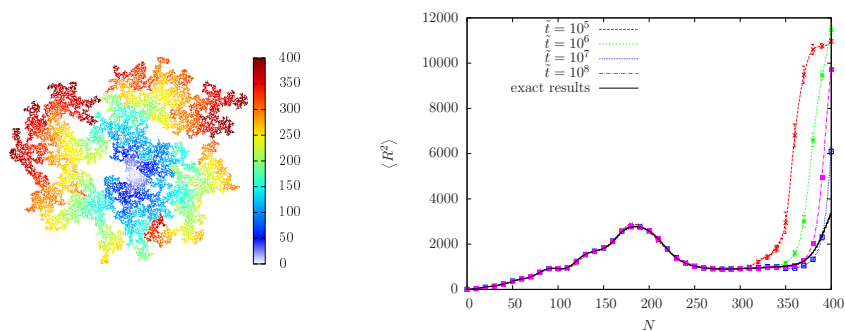


Figure 3.20.: Left: Chemical distance on a critical percolation cluster. The origin is in the light region near the middle. The cluster is trimmed to the area accessible by walks of $N = 400$ steps. Right: end-to-end distances vs. N measured on the cluster shown left. The black curve give the exact results obtained by enumeration, the other data were produced by PERM with varying numbers of started tours.

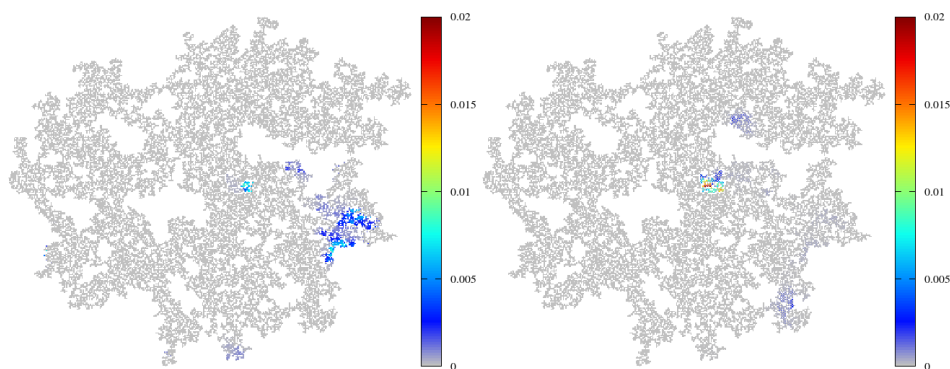


Figure 3.21.: End-point densities for $N = 400$ -step SAWs on the cluster from above (Fig. 3.20) produced by PERM within $\tilde{t} = 10^8$ effective tours (left) and exact densities obtained by SFE (right).

3. The taming of the SAW

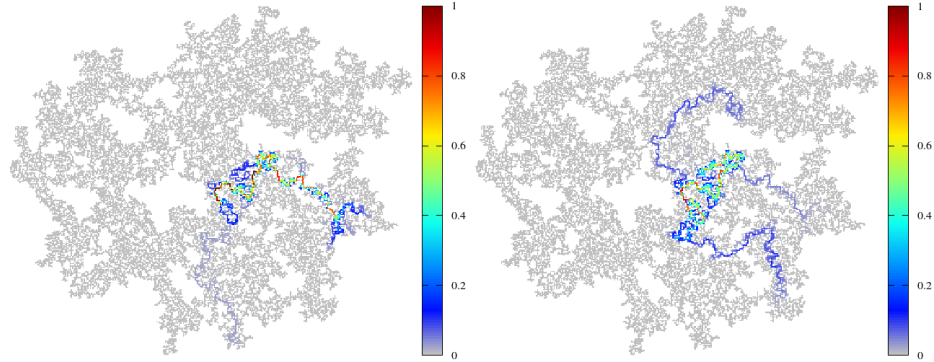


Figure 3.22.: SAW conformation density for $N = 400$ -step SAWs on the cluster from above (Fig. 3.20) produced by PERM within $\tilde{t} = 10^8$ effective tours (left) and exact densities obtained by SFE (right).

This problem is not as severe on every cluster, though. In Fig. 3.23, the magnitude of the estimator for the relative statistical error $|\epsilon/\langle R^2 \rangle|$ is plotted vs. the number of tours for a couple of arbitrarily chosen clusters in $2D$ and $3D$ besides the magnitude of the relative deviation from the exact value (bias):

$$|b_{R^2}| = \frac{|\langle R^2 \rangle - \overline{R^2}|}{\langle R^2 \rangle}, \quad (3.31)$$

where $\overline{R^2}$ is the PERM estimator and the angular brackets denote the exact average.

Fig. 3.24 shows the quenched average of the magnitude of the relative bias, $[|b_{R^2}|]$, as function of the number of tours. This quantity is a measure of the trustworthiness of PERM results for individual clusters. The average was taken over 10^4 cluster and for $N = 400$. At criticality, it vanishes roughly with $\tilde{t}^{-0.36}$ in both $2D$ and $3D$. For the improved version that avoids small dangling ends, the bias vanishes slightly faster, and the runtime is also improved. The decay is significantly faster for large p , but almost unchanged close to p_c . Indeed, for some reason, the decay in $3D$ is actually slowest a bit above p_c , at $p = 0.32$. Also shown the figure is the relative bias for the KGW, which vanishes almost instantly. This was to be expected since KGW conformations are naturally sampled according to their weights.

The bias is not always in the same direction, so that the actual effect on the quenched average is smaller. However, there is a tendency to underestimate both $\langle R^2 \rangle$ and Z , which results in a net bias of the estimators for the quenched averages. As I had already observed previously [101], this bias vanishes relatively quickly for concentrations significantly above p_c but seems very persistent ($\sim \tilde{t}^{-0.5}$) at criticality. Thanks to the new method, I could now test this more thoroughly. For the (relative) bias on the

3.5. Analyzing PERM

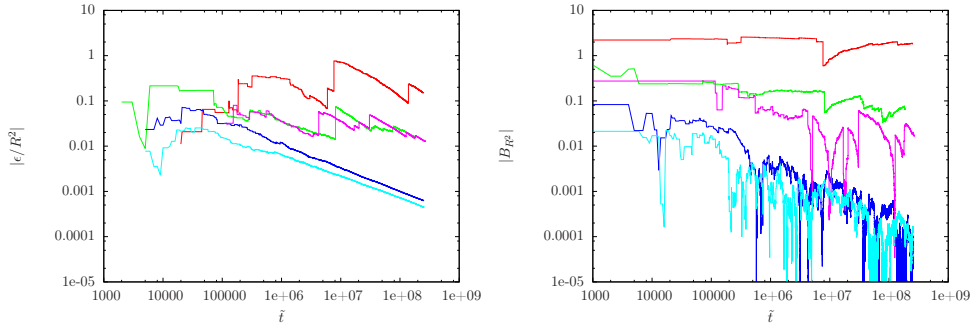


Figure 3.23.: Convergence behavior of PERM estimates for the end-to-end distance of 400-step SAWs on a couple of critical 2D clusters. The left plot shows the magnitude of the estimated relative statistical error and the right-hand plot the actual deviation from the exact values [Eq. (3.31)]. The red curve belongs to the cluster from Fig. 3.20

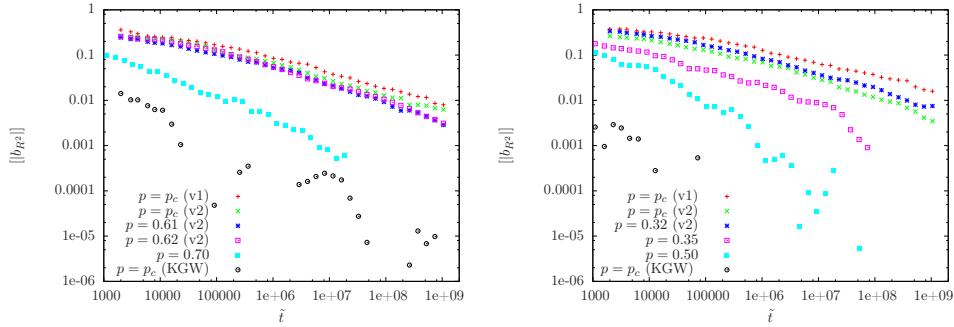


Figure 3.24.: Average relative bias of PERM estimates of the squared end-to-end distance as a function of the number of effective tours on clusters at different levels of concentration in 2D and 3D. At and close to the critical concentration ($p \leq 0.62$ in 2D and $p \leq 0.32$ in 3D), the deviations were measured with respect to the exact values, elsewhere with respect to the final PERM estimates. The different versions, v1 and v2, denote the use of the trap avoiding optimization (v2); see end of Section 3.1.1. The black circles show the bias for the KGWs, which is practically zero after a few tours.

quenched average,

$$B_{[R^2]} = \frac{[\langle R^2 \rangle] - [\overline{R^2}]}{[\langle R^2 \rangle]} \quad (3.32)$$

3. The taming of the SAW

shown in Fig. 3.25, the decay is indeed $\tilde{t}^{-0.5}$ or slower over a long range. Note that the irregular behavior for large \tilde{t} is likely due to statistical fluctuations as the number of clusters where the results have not properly converged becomes quite small. Again, the performance slightly above criticality, $p = 0.32$, in $3D$ is even slightly worse than at p_c .

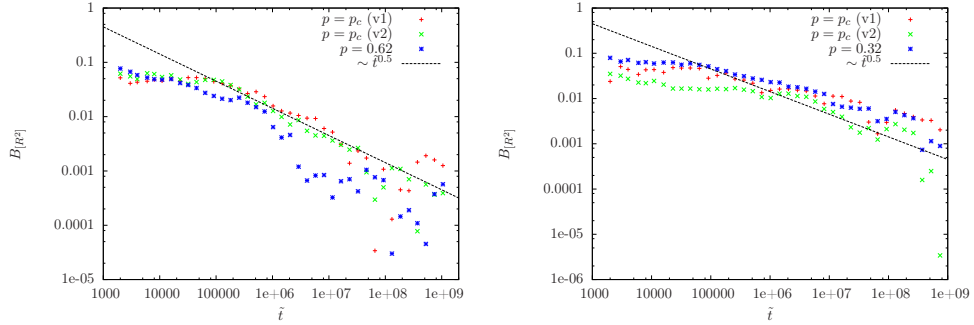


Figure 3.25.: Bias of the PERM estimator for quenched average of the end-to-end distance [Eq. (3.32)], obtained by sampling up to 2^{20} tours of 400-step SAWs on 10^3 clusters in $2D$ (left) and $3D$ (right).

To quantify the performance of PERM, I investigated the quenched average of the end-to-end distance for 400-step SAWs on 10^3 clusters at and closely above p_c on which I had determined the exact conformational averages using the SFE. I then measured the number of tours and the time it took until the results converged (and remained) within 1% of the exact value. The results for the critical cases are plotted as histograms Fig. 3.23. Interestingly, these histograms look very similar for $2D$ and $3D$. As can be seen, there is a significant fraction of clusters where the results did not converge within the maximum of $\tilde{t} = 10^3 \times 2^{20}$ tours I afforded. The rate of these “failures” is listed in Table 3.9 together with the median times and the average magnitude of the relative bias.

Table 3.9.: Fail (non-convergence) rates for PERM simulations of 400-step SAWs on $2D$ (left) and $3D$ (right) clusters at and closely above criticality. Also shown are the mean and median times till convergence as well as the average of the relative bias; see Eq. (3.31).

$2D$					$3D$				
p	fail rate	$\tilde{t}[s]$	$\bar{t}[s]$	$[b_{R^2}]$	p	fail rate	$\tilde{t}[s]$	$\bar{t}[s]$	$[b_{R^2}]$
p_c	13.8%	206	902	0.006	p_c	10.8%	141	958	0.0035
0.61	8.5%	189	1036	0.003	0.315	13.9%	321	1646	0.0051
0.62	8.6%	249	1155	0.003	0.32	15.7%	534	1622	0.0075

3.6. Discussion: Scope and performance of SFE compared to other methods

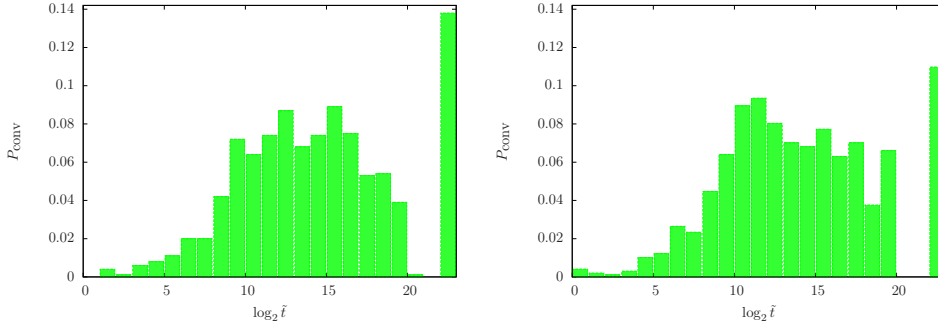


Figure 3.26.: Distribution of numbers of PERM tours above which error estimator and actual deviation of the end-to-end distance remain below 1% of the exact value. The left histogram was obtained for critical clusters in $D = 2$, the right one for $D = 3$. Both represent a sample of 10^3 clusters with SAWs of $N = 400$ steps. The rightmost column corresponds to clusters where the results did not converge within the limit of 2^{20} tours ($10^3 \times 2^{20}$ effective tours).

3.6. Discussion: Scope and performance of SFE compared to other methods

The scope of the SFE method is clearly somewhat limited. Where it can be used, however, it is vastly more efficient than the standard enumeration method. The benefit is most striking at criticality, where the complexity of the SFE seems to be polynomial rather than exponential. Hence, it can handle SAWs several orders of magnitude longer than the standard method. In terms of computation time, this gain sounds more impressive: Exactly enumerating 10^9 conformations (a typical number for SAWs of $N = 50$ steps on $2D$ CPCs) with standard EE takes about one minute, and there are typically 10^{1550} conformations of for SAWs of length $N = 10^4$. Hence, exact enumeration of all conformations of SAWs of 10^4 steps on a CPC would typically take more than 10^{1500} lifetimes of the universe.

Even with PERM it would take extremely long to get reliable results for SAWs of such length. I do not know how long exactly, but from my experience it is not feasible to simulate SAWs of much more than $N = 500$ steps on a critical cluster, and previous studies have not gone beyond $N = 200$. It is harder to say which of the two methods, SFE or PERM, is better when the concentration is somewhat above criticality. Indeed, quantitative comparison of the performances of SFE and PERM is not really possible: PERM being a Monte Carlo method, the estimators for the conformational averages become more accurate with the amount of statistics invested, i.e., the number of started *tours*, and obtaining exact results would take infinitely long. On the other hand, having exact conformational averages is not that important if we also need to take the quenched disorder average. In fact, the effects of any

3. The taming of the SAW

purely statistical errors on the conformational averages will eventually average out if the disorder sample is large enough, so that a small number of *tours* might actually suffice. However, on critical clusters there also tends to be a systematic error (*bias*) depending on the number of started tours, as was shown above. Hence, one would need to increase the number of *tours* with the disorder sample size to keep the bias negligible compared to the statistical error on the disorder average.

Still, based on the findings from the last two sections, Table 3.8 and Table 3.9 in particular⁶, it is fair to say that the SFE method is clearly more efficient while $p \leq 0.62$ and $p \leq 0.32$ in $2D$ and $3D$, respectively, while PERM is more useful for $p \geq 0.64$ and $p \geq 0.33$. In the intermediate regime, both methods have pros and cons: SFE gives exact results for most clusters but fails completely for others. By contrast PERM will always give (approximate) results, but one has to be careful to avoid bias effects. Apart from the slight worsening of the performance of PERM for concentrations closely above p_c , an effect that is odd but rather insignificant, the two methods show strikingly opposite trends: SFE gets worse where PERM gets better. On average this probably also holds true for different clusters at the same concentration. SFE works well when a cluster is weakly connected and can be decomposed by removing a small number of sites. These “bottlenecks” are exactly what makes exploring the cluster difficult for PERM, which therefore works much better on strongly connected clusters.

On balance, I thus come to the following, conciliatory conclusion: The two methods are not really in competition, they are complementary. For concentrations where both methods work to some extent, they should best be used in conjunction: One can start by doing SFE where possible, use these results to estimate the number of PERM *tours* needed for convergence, and then use PERM to estimate the results on those clusters where SFE had failed. These are typically more compact and will thus be easier cases for PERM, so that the determined *tour* limits will be generous. Conceivably, one might even mix the two algorithms, using PERM to estimate the results for cells where enumeration takes too long. However, I did not try this idea in practice.

⁶The fail rates can only be compared in so far that roughly similar maximum times were granted to the two methods. A more systematic comparison was done in Ref. [105] but with a less efficient version of the SFE program.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

The scale-free enumeration method allows for much more detailed investigations of SAWs on CPCs than previous numerical tools. This justified revisiting questions that have already been thoroughly investigated in the past but were never fully settled. These questions mainly regard the scaling behavior of the (quenched) average end-to-end distance and the number of conformations on critical clusters as well as cluster backbones.

4.1. End-to-end distances

I measured the conformational averages of the squared end-to-end distances $[\langle R^2 \rangle]$ for random samples of disorder configurations (incipient clusters) in order to assess the asymptotic scaling behavior of the quenched disorder average:

$$[\langle R^2 \rangle] \sim N^{2\nu} \quad (4.1)$$

The measurements were carried out independently for walks of varying length N , started with $N = 50$ and increasing by factors of $\sqrt{2}$. The maximum was set to $N = 12000$ in $2D$ and $3D$, while the chains in higher dimensions had to be shorter as creating the clusters became too demanding. An overview of the results for $[\langle R^2 \rangle]$ as a function of N is shown in Fig. 4.1 for systems of varying dimension on a double logarithmic scale. The power-law behavior predicted by Eq. (4.1) is reflected by constant slopes. With increasing dimension, these can be seen to approach the mean-field value of $\nu_{\text{MF}} = 1/2$, represented by the dashed line. The solid line corresponds to the exact prediction for the free SAW in $2D$ ($\nu_1 = 3/2$). As noted in previous works, the exponent on $2D$ CPCs is slightly larger.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

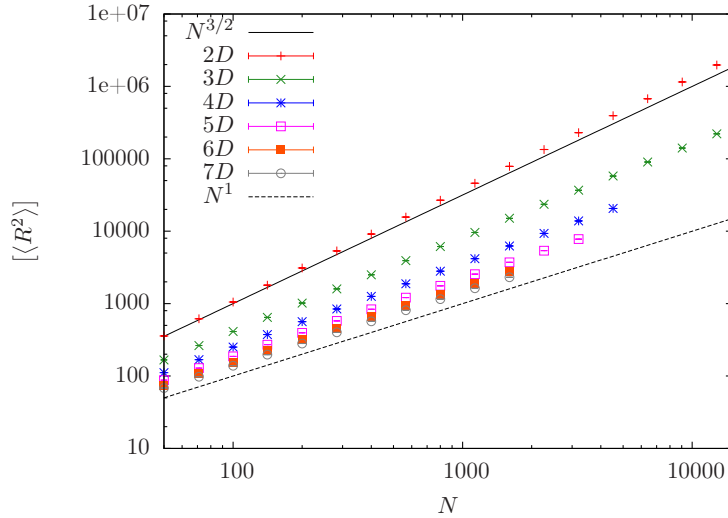


Figure 4.1.: Results for the mean squared end-to-end distance as a function of the number of SAW steps on critical clusters in $D = 2-7$. The lines correspond to the regular-lattice value, $\nu_1 = 3/4$, and the mean-field value, $\nu_{\text{MF}} = 1/2$.

4.1.1. Results in two dimensions:

The two-dimensional case is particularly important since this is where the two components of the system (SAWs and CPCs) are best understood individually. It has attracted most interest in the past, perhaps also because it is good to visualize and usually the easiest case to study. (As discussed in the previous chapter, the SFE method is an exception: It is least efficient in $2D$, where the critical clusters are more strongly connected than in higher dimensions.) After the $3D$ case, it has therefore also received most of my attention.

Directly plotting $[\langle R^2 \rangle]$ vs N as in Fig. 4.1 broadly shows the power-law behavior, but the range is too large to reveal details. In Fig. 4.2 I have therefore scaled the results for $[\langle R^2 \rangle]$ by a factor $N^{1.56} \approx N^{2\nu}$. Thus, the slope becomes close to zero and minor deviations can be discerned. The figure shows the results for the full incipient clusters and for backbones of incipient clusters. I chose very large sample sizes ($\approx 10^6$ clusters) and smaller intervals ($N_{i+1}/N_i \approx 2^{1/4}$) for the initial range of $N \leq 100$. This required little effort and allows for a better direct comparison with previous studies. The sample size in the range of larger N was at least 5×10^4 for each individual length. There are several things to note about Fig. 4.2: For both backbones and full clusters, the slope starts relatively low, and there appears to be a crossover to a larger value at about $N = 30$. In the region around $N = 150$, it decreases again before stabilizing eventually. This kind of finite-size behavior was quite unexpected, so I carefully checked that it is not due to varying lattice sizes. However, the choice of L had no noticeable influence as long as the lattice was large compared to the typical

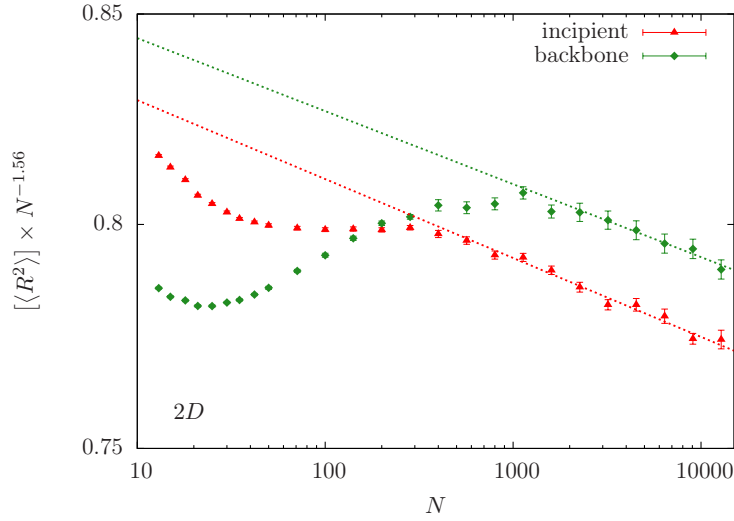


Figure 4.2.: Scaled mean squared end-to-end distance as a function of the number of SAW steps on incipient critical $2D$ clusters and cluster backbones. The values have been divided by $N^{1.56} \approx N^{2\nu}$ to enhance visibility. Straight lines show least-squares power-law fits to the data range $N = 800$ - 12000 .

extension of the SAWs. This can be guaranteed by choosing $L > 2N$, which is in $2D$ easily affordable.

To estimate the exponent ν , I used a simple least-squares fit of Eq. (4.1), adjusting the lower cutoff to optimize the quality of the fit. The dependence of the estimates on the fit range can be seen in Fig. 4.3. The values are also listed in Table 4.1 together with the parameters of the fits. Note that the data points in Fig. 4.3 are correlated as the fit ranges overlap. Based on the χ^2 values and the fit errors as well as my optical judgment of the data curves, I consider the results obtained with $N_{\min} = 800$ as my best estimates¹

$$\nu_{IC} = 0.7749(4), \quad \nu_{BB} = 0.7766(5) \quad (4.2)$$

The exponent for the backbone appears to be slightly larger, but the difference is not very significant and seems rather to diminish with the system size. This gives some support to the hypothesis that ν is the same on the backbones as on the full clusters, but it is not a clear confirmation.

¹The slightly different value I gave in Ref. [104] ($\nu_{IC} = 0.7745(15)$) had been obtained from smaller systems ($N = 1000$).

4. Scaling behavior of self-avoiding walks on critical percolation clusters

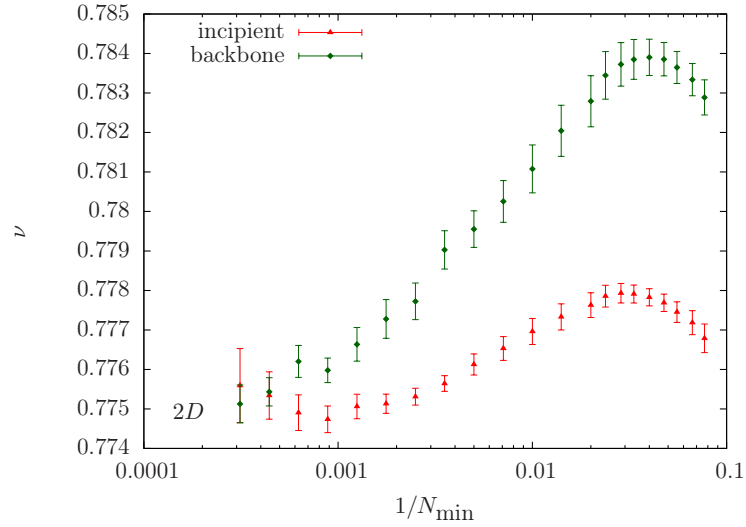


Figure 4.3.: Effect of the lower cutoff of the fit range on the estimate for ν on critical 2D clusters and backbones. The upper cutoff was $N = 12000$ for all points.

Table 4.1.: Estimates of the exponent ν on critical 2D clusters and backbones for different fit ranges. Also listed are the effective number of degrees of freedom (NDF) and the reduced χ^2 values.

N_{\min}	N_{\max}	NDF	ν_{IC}	χ^2_{IC}	ν_{BB}	χ^2_{BB}
30	100	4	0.7780(5)	2.04	0.7856(5)	2.00
13	12000	23	0.7767(4)	8.38	0.7773(9)	8.69
50	12000	15	0.7777(4)	3.98	0.7828(7)	6.61
100	12000	13	0.7770(4)	3.19	0.7811(7)	4.59
200	12000	11	0.7761(3)	2.02	0.7796(5)	2.68
400	12000	9	0.7752(3)	1.03	0.7777(5)	1.54
566	12000	8	0.7750(3)	1.02	0.7773(5)	1.40
800	12000	7	0.7749(4)	1.07	0.7766(5)	1.01
1600	12000	5	0.7749(6)	1.04	0.7763(5)	0.62
3200	12000	3	0.776(1)	0.963	0.7751(5)	0.39

As methods used in previous studies were limited to system sizes smaller than $N = 150$, they could not have revealed the asymptotic behavior. Moreover, the fairly constant slopes in the initial regimes ($N \approx 0-30$ and $N \approx 50-200$), might not have given much reason for suspicion. It is therefore not surprising that my estimates for ν differ somewhat from earlier ones, e.g., $\nu_{\text{BB}} = 0.778(15)$ [78] and $\nu_{\text{BB}} = 0.782(3)$ [82].

However, the difference to these results is very small as the initial slope for the backbone (which was considered in both these studies) happens to be very close to the final slope. When I use a similar range as in Ref. [82], see first row in Table 4.1, the results are fully consistent with both these previous estimates.

Unfortunately, the data is not sufficiently precise to allow fitting higher-order terms to Eq. (4.1). Capturing the initial inflection-point would in fact require at least two additional terms, introducing too many free parameters.

I also tried a *successive-slopes* analysis similar to the one used Ref. [78]. For each sample, I looked at the logarithmic slope of the last ten percent of steps:

$$\bar{\nu}_N = \frac{\ln [\langle R_N^2 \rangle] - \ln [\langle R_{N'}^2 \rangle]}{\ln N - \ln N'} \quad (4.3)$$

with $N' = 0.9N$. Since the values for N and N' are strongly correlated, the error was estimated by bootstrap resampling; see Ref. [118]. To estimate the asymptotic value of $\bar{\nu}$, I took the intersection with the y -axis of a linear fit to $\bar{\nu}_N$ vs. $1/N$; see Fig. 4.4. This gave

$$\nu_{\text{IC}} = 0.781(2), \quad \nu_{\text{BB}} = 0.778(2). \quad (4.4)$$

Since this approach did not seem to give better results than direct fitting, I did not pursue it further.

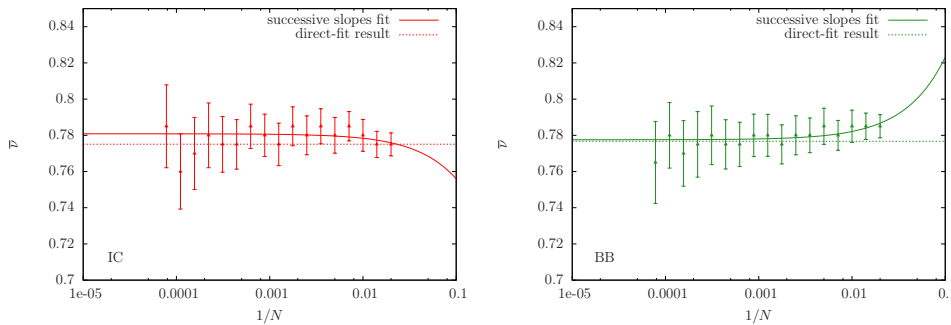


Figure 4.4.: “Running” estimates for ν from successive slopes [Eq. (4.3)] for incipient $2D$ clusters (left) and cluster backbones (right). The solid lines correspond to linear least-squares fits; the dashed lines show the result from direct fitting [Eq. (4.2)] for comparison.

4.1.2. Results in three dimensions

The three-dimensional case is physically most relevant. It has therefore also been widely studied in the past and represents the main focus of my work. As in $2D$, I considered walks of up to $N = 12000$ steps and used independent samples for different maximum lengths. Here however, I could not afford setting $L > 2N$, which would have been much too costly, but the clusters were still large compared to the typical

4. Scaling behavior of self-avoiding walks on critical percolation clusters

extensions. The exact sizes I used are listed in the Appendix A.1 (for all dimensions).

The results for the mean squared end-to-end distances are shown in Fig. 4.5. The values were again scaled by a factor of roughly $N^{2\nu}$ for the sake of visibility. Just as in $2D$, the asymptotic behavior only seems to set in after about $N = 800$ steps. However, there appears to be no inflection point this time. The slope is fairly constant up to $N = 100$, from where it begins to crossover to a smaller value. Accordingly, the estimate for ν may depend on the range of lengths one investigates as can be seen in Fig. 4.7 and Table 4.2. Judging by the quality of the fits, my best estimates from this approach are obtained with $N_{\min} = 800$ on the full clusters and $N_{\min} = 1131$ on the backbones²:

$$\nu_{\text{IC}} = 0.6462(4), \quad \nu_{\text{BB}} = 0.6467(7). \quad (4.5)$$

These two values are consistent within the error bars, again supporting the “backbone hypothesis”. However, the data curves in Fig. 4.5 initially show a small but clear difference, the backbone slope being somewhat larger than the one for full clusters, which starts out much closer to the asymptotic value. This is reflected in the findings from previous studies of SAWs on cluster backbones: $\nu_{\text{BB}} = 0.662(6)$ [78] and $\nu_{\text{BB}} = 0.667(3)$ [82]. Similar values can again be recovered by restricting the fit range accordingly (for the backbone data) as seen in Fig. 4.6 and the first rows in Table 4.2. Here, too, one could easily be misled to believe that the initial behavior already reflects the thermodynamic limit as power-law fits in the range $N < 100$ appear very convincing.

Unlike in $2D$, fitting the first confluent correction term to Eq. (4.1), $N^{2\nu-\Delta}$, works reasonably well for $D = 3$. Using the ansatz [115, 119]

$$[\langle R^2 \rangle] = a(N + \delta)^{2\nu} [1 + b/(N + \delta)^\Delta], \quad (4.6)$$

with a shift of $\delta = 1/2$ over the range 25 – 12000, I obtained the estimates:

$$\nu_{\text{IC}} = 0.644(2), \quad \Delta_{\text{IC}} = 0.51(5); \quad (4.7)$$

$$\nu_{\text{BB}} = 0.640(3), \quad \Delta_{\text{BB}} = 0.34(4) \quad (4.8)$$

The results for ν are compatible to those from simple power-law fits. They hardly depend on the shift parameter δ (as long as it is reasonably small), which provides for better convergence by smoothing out higher-order corrections. The fits are plotted as solid lines in Fig. 4.5. The reduced χ^2 -values are 1.22 and 1.59 for full clusters and backbones, respectively.

²These values were given slightly wrong in Ref. [106] due to an accidental slip. An erratum currently in print.

4.1. End-to-end distances

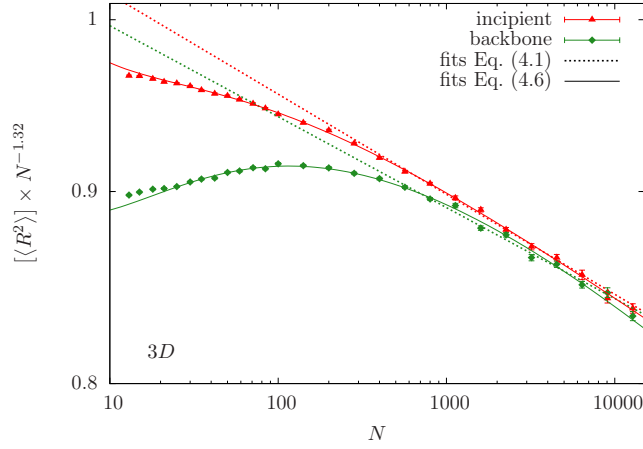


Figure 4.5.: Scaled mean squared end-to-end distance as a function of the number of SAW steps on critical $3D$ clusters and cluster backbones. The values have been divided by $N^{1.32} \approx N^{2\nu}$ to enhance visibility. Straight lines show least square power-law fits to the data range $N = 800-12000$. For the solid curves, the first confluent correction term to Eq. (4.1) was included [Eq. (4.6)]. Here, the full range $N \geq 25$ was fitted.

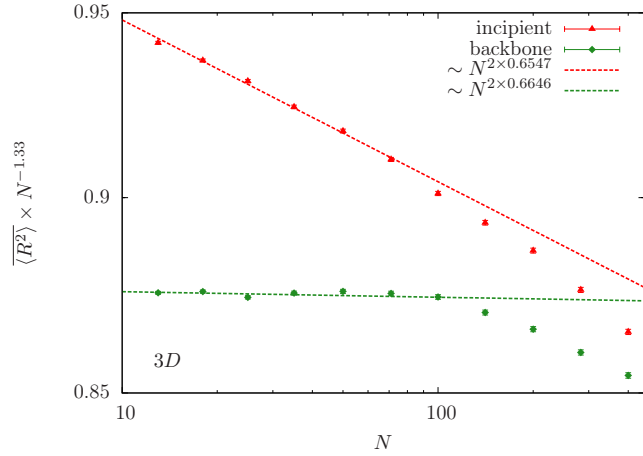


Figure 4.6.: Mean squared end-to-end distances for short SAWs on $3D$ CPCs. Lines are least-squares power-law fits to the range $N \leq 84$.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

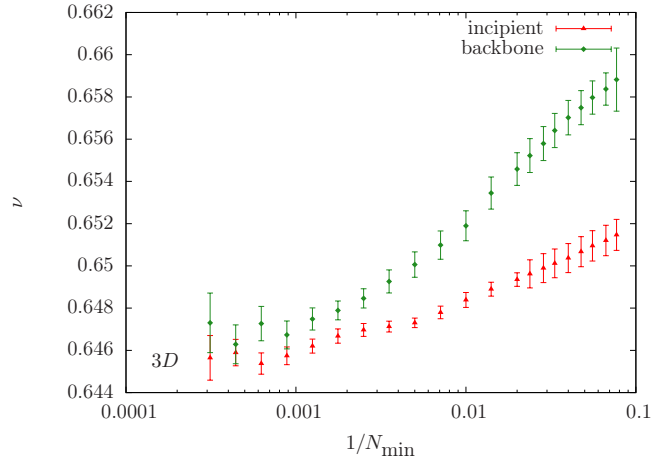


Figure 4.7.: Effect of the lower cutoff of the fit range on the estimate for ν on critical 3D clusters and backbones. The upper cutoff was $N_{\max} = 12000$ for all points.

Table 4.2.: Estimates of the exponent ν on critical 3D clusters and backbones from simple power-law fits over different ranges. Also listed are the effective number of degrees of freedom (NDF) and the reduced χ^2 values.

N_{\min}	N_{\max}	NDF	ν_{IC}	χ^2_{IC}	ν_{BB}	χ^2_{BB}
13	42	6	0.6553(6)	1.31	0.6645(6)	1.16
13	84	10	0.6547(5)	1.61	0.6646(4)	1.43
13	12000	25	0.6515(8)	7.84	0.659(2)	15.8
50	12000	15	0.6494(4)	4.05	0.6546(9)	9.74
100	12000	13	0.6484(4)	3.13	0.6519(8)	6.19
200	12000	11	0.6473(3)	1.45	0.6501(7)	3.91
400	12000	9	0.6470(4)	1.41	0.6485(5)	2.17
566	12000	8	0.6467(4)	1.31	0.6479(5)	1.77
800	12000	7	0.6462(4)	1.03	0.6475(5)	1.70
1131	12000	6	0.6457(4)	0.941	0.6467(7)	1.53
1600	12000	5	0.6459(7)	0.909	0.6473(9)	1.51
3200	12000	3	0.646(1)	0.974	0.647(2)	1.32

4.1.3. Results in four and five dimensions

While the practical significance of self-avoiding walks on hyper-cubic clusters is perhaps not obvious, they fill a key position in connecting with theory: The critical behavior in $4D$ and $5D$ is still non-trivial, but the proximity to the upper-critical dimension ($D_{\text{uc}} =$

6) means that predictions from perturbative field-theory should be most accurate. Also, since the differences between backbones and full clusters are more pronounced in higher dimensions, these cases should be helpful in clarifying the question of whether or not the SAWs' behavior is determined solely by the backbone.

I enumerated walks of up to $N = 4525$ steps in $4D$ and up to $N = 3200$ in $5D$. As discussed in Section 3.4.2, the bottleneck in high dimensions is the creation and storage of the clusters rather than the actual enumeration procedure. It therefore made sense to take multiple starting positions from each cluster (or backbone). I used 10 randomly chosen starting points in $D = 4$ and 100 in all higher dimensions. Since results from different starting sites on the same cluster are correlated, I binned the results from each cluster and measured the variances of the bin averages to obtain estimates for the statistical error.

The measured mean squared end-to-end distances, shown in Fig. 4.8, initially increase with a larger slope, similarly as in $D = 3$. However, finite-size effects seem to wear off more quickly and the curves straighten after about $N = 400$ steps. This is reflected in the estimates of the exponent ν for different fit ranges, which are given in Table 4.3 and plotted in Fig. 4.9. Again, the backbone slopes are initially larger (and thus further off) though the differences are quite small. The best results were obtained using a lower cutoff of $N_{\min} = 283$ in both $D = 4$ and $D = 5$:

$$\nu_{\text{IC}}^{4D} = 0.5769(5), \quad \nu_{\text{BB}}^{4D} = 0.5784(9) \quad (4.9)$$

$$\nu_{\text{IC}}^{5D} = 0.5371(6), \quad \nu_{\text{BB}}^{5D} = 0.5411(9) \quad (4.10)$$

In $4D$, the result for the backbones and the whole incipient clusters are consistent, further corroborating the ‘‘backbone hypothesis’’, but they do deviate in $5D$. However, my estimates for $D \geq 5$ should not be too strongly relied upon because the clusters might simply not have been sufficiently large.

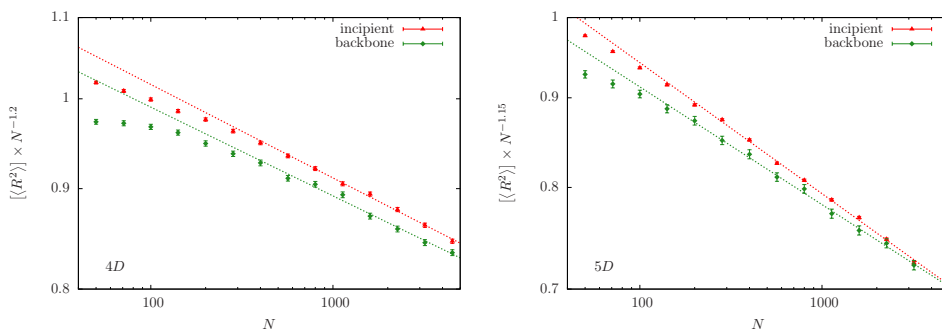


Figure 4.8.: Scaled mean squared end-to-end distance as a function of the number of SAW steps on critical clusters and cluster backbones in $4D$ (left) and $5D$ (right). The values have been divided by $\approx N^{2\nu}$ for better visibility. Straight lines show least square power-law fits to the data ranges $N = 283$ - 4525 and $N = 283$ - 3200 , respectively.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

Statistical fluctuations are also larger on the backbones, which is due to the fact that they are considerably less massive. Thus, the results from different starting points are more strongly correlated and the individual “bins” less representative of the quenched disorder average.

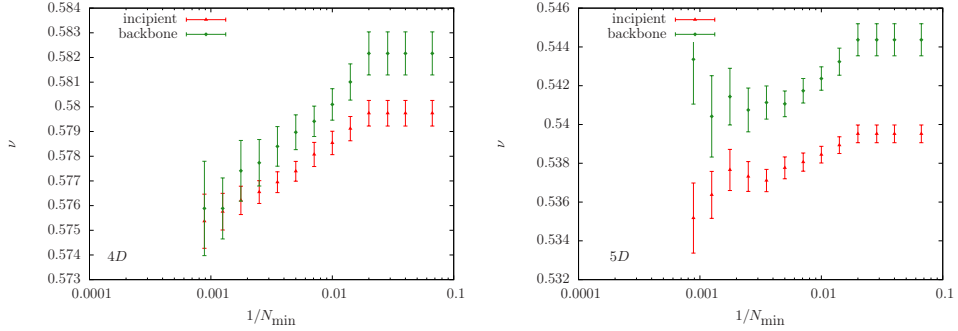


Figure 4.9.: Effect of the lower cutoff of fit range on the estimate for ν on critical clusters and their backbones in $D = 4$ (left) and $D = 5$ (right).

Table 4.3.: Estimates of the exponent ν on critical clusters and backbones in $4D$ (top) and $5D$ (bottom) for different fit ranges. Also listed are the effective number of degrees of freedom (NDF) and the reduced χ^2 values.

$4D$						
N_{\min}	N_{\max}	NDF	ν_{IC}	χ^2_{IC}	ν_{BB}	χ^2_{BB}
50	4525	12	0.5797(6)	2.64	0.5822(9)	2.91
100	4525	10	0.5785(5)	2.18	0.5801(8)	1.64
200	4525	8	0.5774(5)	1.14	0.5790(8)	1.35
283	4525	7	0.5769(5)	0.992	0.5784(9)	1.30
400	4525	6	0.5765(5)	0.905	0.578(1)	1.25
566	4525	5	0.5762(6)	0.907	0.577(2)	1.34
800	4525	4	0.5758(8)	0.911	0.576(2)	1.07
$5D$						
N_{\min}	N_{\max}	NDF	ν_{IC}	χ^2_{IC}	ν_{BB}	χ^2_{BB}
50	3200	11	0.5395(5)	2.96	0.5444(9)	1.41
100	3200	9	0.5389(5)	2.04	0.5424(7)	1.04
200	3200	7	0.5378(6)	1.91	0.5411(7)	0.611
283	3200	6	0.5371(6)	1.60	0.5411(9)	0.658
400	3200	5	0.5373(8)	1.72	0.541(2)	0.728
566	3200	4	0.538(2)	1.86	0.541(2)	0.777

4.1.4. Results in six and seven dimensions

I included the cases at and above the presumed upper critical dimension ($D_{\text{uc}} = 6$) to check whether the predicted random walk behavior can be observed. As finite-size effects for the clusters appear to be very strong in $D = 6$ (see Appendix A.2) and only relatively small clusters are affordable in $D = 7$, it is not at all clear that this should be the case. My upper limit for the lengths of SAWs was $N = 1600$ in both $D = 6$ and $D = 7$. I started walks from 100 randomly chosen sites per cluster and investigated 5×10^3 clusters for each length. The results for the end-to-end distances are shown in Fig. 4.10. Here the values were simply divided by N , so that they should become flat to recover the expected diffusive behavior. As in lower dimensions, the slope is initially larger. However, due to limited system sizes and stronger finite size effects, it is hard to tell whether the asymptotic behavior is reached. Indeed, it seems that the slopes would still decrease further for larger systems judging by Fig. 4.11 and the results Table 4.4, which describe the dependence on the chosen fit range. In $6D$ the distance to the theoretical mean-field value, $\nu_{\text{MF}} = 1/2$, is still quite large. In $7D$, by contrast, the slope does become very close to one though not quite consistent within the error bars. The asymptotic scaling behavior on the backbones appears again to be the same or very similar though the initial slopes are larger. The statistical fluctuations are also significantly stronger for the backbones for the reason described in the last subsection. The best numerical estimates for the exponent ν in $6D$ and $7D$ were obtained using a lower cutoff of $N_{\text{min}} = 200$:

$$\nu_{\text{IC}}^{6D} = 0.5153(5), \quad \nu_{\text{BB}}^{6D} = 0.5211(8) \quad (4.11)$$

$$\nu_{\text{IC}}^{7D} = 0.5068(6), \quad \nu_{\text{BB}}^{7D} = 0.509(2). \quad (4.12)$$

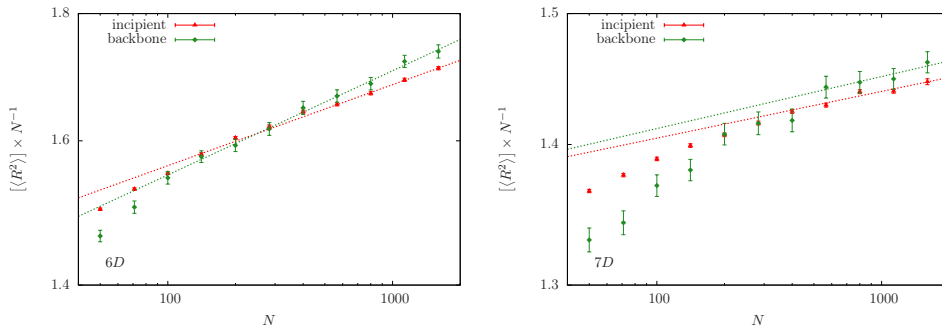


Figure 4.10.: Scaled mean squared end-to-end distance as a function of the number of SAW steps on critical clusters in $6D$ (left) and $7D$ (right) and cluster backbones. The values have been divided by N for better visibility. Straight lines show least square power-law fits to the data in the range $N \geq 200$.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

Table 4.4.: Estimates of the exponent ν on critical clusters and backbones in $6D$ (top) and $7D$ (bottom) for different fit ranges. Also listed are the effective number of degrees of freedom (NDF) and the reduced χ^2 values.

$6D$

N_{\min}	N_{\max}	NDF	ν_{IC}	χ^2_{IC}	ν_{BB}	χ^2_{BB}
50	1600	9	0.5187(8)	3.93	0.524(2)	1.35
100	1600	7	0.5170(7)	2.56	0.5211(6)	0.486
200	1600	5	0.5153(5)	1.04	0.5211(8)	0.495
283	1600	4	0.5152(6)	1.14	0.521(2)	0.524
400	1600	3	0.5148(9)	1.24	0.520(2)	0.539

$7D$

N_{\min}	N_{\max}	NDF	ν_{IC}	χ^2_{IC}	ν_{BB}	χ^2_{BB}
50	1600	9	0.5085(5)	2.94	0.513(1)	1.21
100	1600	7	0.5075(5)	1.93	0.511(1)	0.921
200	1600	5	0.5068(6)	1.72	0.509(2)	0.658
283	1600	4	0.5062(7)	1.59	0.509(2)	0.735
400	1600	3	0.5056(9)	1.56	0.509(3)	0.848

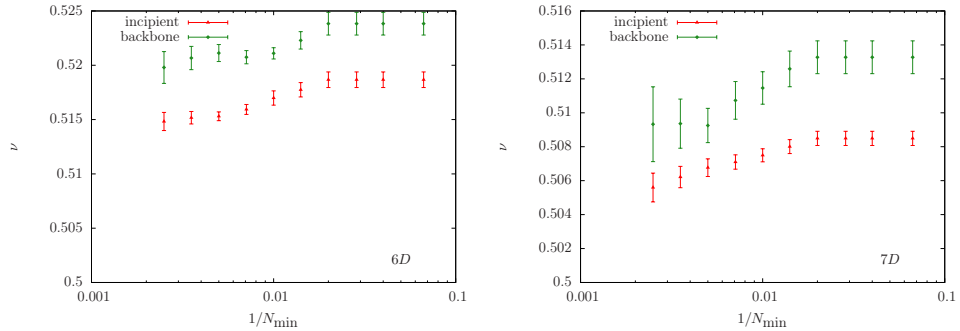


Figure 4.11.: Effect of the lower cutoff of fit range on the estimate for ν on critical clusters and their backbones in $D = 6$ (left) and $D = 7$ (right).

As already mentioned, the high dimensional cases are particularly difficult due to the unfortunate combination of strong finite-size effects and small affordable system sizes. The fact that the estimates differ from the theoretical value of $\nu^{\text{uc}} = 1/2$ can therefore neither be taken as evidence that mean-field behavior does not set in at $D = 6$, nor should it raise suspicion about the reliability of the estimates in lower dimensions.

4.1.5. Chemical distances

Like the Euclidean end-to-end distance, the quenched average of the chemical (end-to-end) distance is supposed to asymptotically follow a power-law:

$$[\langle l \rangle] \sim N^{\nu_l} \quad (4.13)$$

However, researchers (including myself) have paid much less attention to it, which may be somewhat unjust as the chemical distance is actually a more natural metric in fractal systems. It has been suggested that, as a consequence of Eq. (2.21), ν and ν_l should be related via

$$\nu_l = \nu \cdot d_{\min}, \quad (4.14)$$

which seemed to be supported by numerical data [78]. Since d_{\min} is known relatively precisely (in $2D$ and $3D$ at least), and since $\langle l \rangle$ fluctuates less than $\langle R^2 \rangle$ (cf. Ref [77]), this relation might actually allow for more precise determination of ν .

Measuring $\langle l \rangle$ in addition to $\langle R^2 \rangle$ is straightforward using the SFE (see Section 3.2.3) and requires very little extra effort. Unfortunately, however, I missed doing so while producing the bulk of my data, so that I can only present results for the $D = 2$ and $D = 3$ cases here. These were obtained from samples of at least 2×10^4 clusters and backbones for each considered length (again up to $N = 12000$) in $2D$ and at least 3×10^4 in $3D$.

As can be seen in Fig. 4.12, the behavior of $[\langle l \rangle]$ is analogous to that of $[\langle R \rangle]$: the slope is slightly larger initially and stabilizes around $N = 1000$ to a value which is very similar for the backbones and the full clusters. Estimates from least-squares fits are given in Table 4.5. They were obtained from the range $N = 800, \dots, 12000$. The values indeed turned out very close to those obtained via Eq. (4.14), but the differences for the full cluster values are a bit too large. This discrepancy is slightly disconcerting, since Eq. (4.14) seems to follow rather directly from the definitions. (Note, however, that Eq. (2.21) refers to uniform averages over all cluster sites, while Eq. (4.1) and Eq. (4.13) take uniform averages over all walk conformations.) It may be due to the fact that the logarithmic slopes for the chemical distances on the full clusters still have, for some reason, not fully converged in the fitted range as can be seen in Fig. 4.13, where the dependence on the fit range is shown. Curiously, this is contrary to the situation with the Euclidean distances, where the backbone slopes appeared to converge more slowly.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

Table 4.5.: Measured values for the exponent ν_l of the chemical distances [Eq. (4.13)] compared to estimates obtained from Eq. (4.14) on incipient clusters (IC) and backbones (BB) in 2D and 3D.

D	type	ν	d_{\min}	$\nu \cdot d_{\min}$	ν_l
2	IC	0.7749(4)	1.1318(2)	0.8770(6)	0.880(1)
2	BB	0.7767(5)	1.318(2)	0.8783(6)	0.878(1)
3	IC	0.6462(4)	1.3756(3)	0.8889(8)	0.8948(3)
3	BB	0.6467(7)	1.3756(3)	0.890(2)	0.8915(3)

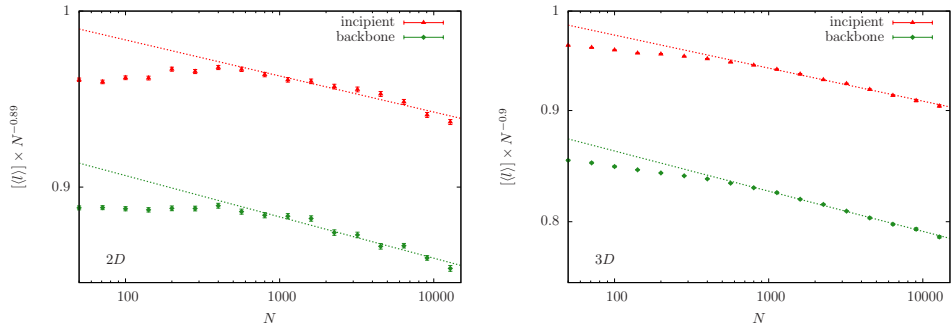


Figure 4.12.: Mean chemical distances of SAWs on CPCs (red) and backbones (green) in 2D (left) and 3D (right). The values are divided by $N^{0.9} \approx N^{\nu_l}$ for the sake of visibility. Dashed lines show least-squares power-law fits obtained from the range $800 \leq N \leq 12000$.

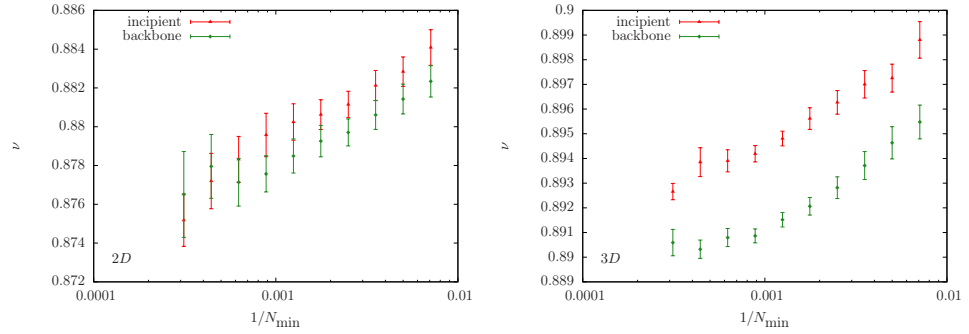


Figure 4.13.: Effect of the lower cutoff of the fit range on the estimate for ν on critical clusters and backbones in 2D and 3D. The upper cutoff is $N = 12000$ for all points.

4.1.6. All cluster (AC) averages

Instead of the incipient cluster (IC) average, where only percolating clusters are considered, one can also average over all clusters that support at least one SAW conformation of a given length N (see Section 2.3.2). This all cluster (AC) average is rarely discussed in the literature and has been assumed to yield the same results as the IC average in the thermodynamic limit [60]. This has been supported somewhat by numerical results [67] but is still far from certain.

While I predominantly investigated the IC average, in order to be consistent with the literature and because it should be less affected by finite-size effects, I decided to compare the two averages at least for the $2D$ case. I took samples of at least 3×10^4 clusters (of sufficient size) for each length in the range $N = 50, 71, \dots, 12000$. The results for the end-to-end distances are shown in Fig. 4.14 together with those obtained on incipient percolation clusters. Quite clearly, the slopes do not seem to converge to the same value but indicate

$$\nu_{AC} < \nu_{IC}. \quad (4.15)$$

My estimate from a least-squares fit over the range $N \geq 800$,

$$\nu_{AC} = 0.7681(5), \quad (4.16)$$

is indeed significantly below the incipient-cluster estimate [Eq. (4.2)], $\nu_{IC} = 0.7749(4)$, clearly disproving the presumed equality.

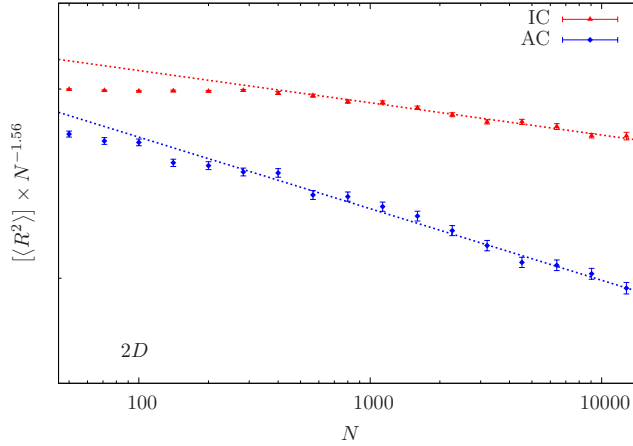


Figure 4.14.: All cluster (AC) average of the mean squared end-to-end distances of SAWs on $2D$ CPCs (blue) compared to the results on incipient clusters (red). The values were scaled by a factor of $N^{1.56}$ ($\approx N^{2\nu}$) for the sake of visibility. The lines correspond to least-squares fits over the range $N \geq 800$.

4.1.7. Discussion

My best estimates of the exponent ν for the quenched RMS end-to-end distance are collected in the second and third column of Table 4.6. The other columns of the table contain previous estimates obtained from Flory arguments [Eq. (2.33) plus Eq. (2.34)] and perturbative field theory [Eq. (2.37) and Eq. (2.38)]. I did not include the numerous analytical and numerical results from earlier studies as they are less precise and reliable, but an overview can be found, e.g., in Refs. [96, 120]. In Fig. 4.15 my results are plotted together with the other estimates from Table 4.6. The best agreement is found with the results from the Flory approximation. As to the field-theory estimates, the agreement with the results from Eq. (2.38) is clearly better than with those from Eq. (2.37) apart from the $2D$ case. There, however, one would expect the strongest discrepancy as the distance to the upper critical dimension is largest. Judging by these results, Eq. (2.38) therefore seems more credible. Note that using [1/2] Padé approximants for Eq. (2.37) and Eq. (2.38) as done in Ref. [85], the agreement with the numerical estimates is slightly worse as the values are larger. The discrepancy to the mean-field value at $D = 6$ is probably due to finite-size effects and logarithmic corrections.

As to the backbone hypothesis: In all dimensions investigated, the estimates for ν_{IC} and ν_{BB} are very close, though not always entirely consistent within the error bars. Still, at least for $D = 2, 3$, where the data are most reliable, the agreement is convincing. On balance, the findings hence support the “backbone hypothesis”, $\nu_{IC} = \nu_{BB}$. However, the results on the backbones are consistently “worse” in the sense that the slopes $\frac{d \ln \langle R^2 \rangle}{d \ln N}$ start further away from the asymptotic limit and converge more slowly towards it. This contradicts the picture that in the asymptotic limit SAWs only exist on the backbone [48, 60, 70, 74, 78, 83]. As I have previously argued [101], the idea that dangling-ends are irrelevant because they cannot support “infinite” SAW conformations is a flawed. As long as the clusters are large compared to the length

Table 4.6.: My numerical estimates for ν on incipient clusters and backbones in different dimensions (second and third column). Also listed are results from the Flory approximation Eq. (2.33) with α either from Eq. (2.34) (ν_{Flory1} [55]), using the highlighted values from Table 2.2 and Table 2.4, as well as from perturbative RG, namely from Eq. (2.37) (ν_{RG1} [80]) and Eq. (2.38) (ν_{RG2} [81]).

D	ν_{IC}	ν_{BB}	ν_{Flory1}	ν_{RG1}	ν_{RG2}
2	0.7749(4)	0.7767(5)	0.7592(7)	0.7853...	0.7414...
3	0.6462(4)	0.6467(7)	0.644(5)	0.6783...	0.6537...
4	0.5769(5)	0.5784(9)	0.584(4)	0.5951...	0.5842...
5	0.5371(6)	0.5411(9)	0.535(3)	0.5357...	0.5329...
6	0.5153(5)	0.5211(8)	0.5	0.5	0.5
7	0.5068(6)	0.509(2)	0.5	0.5	0.5

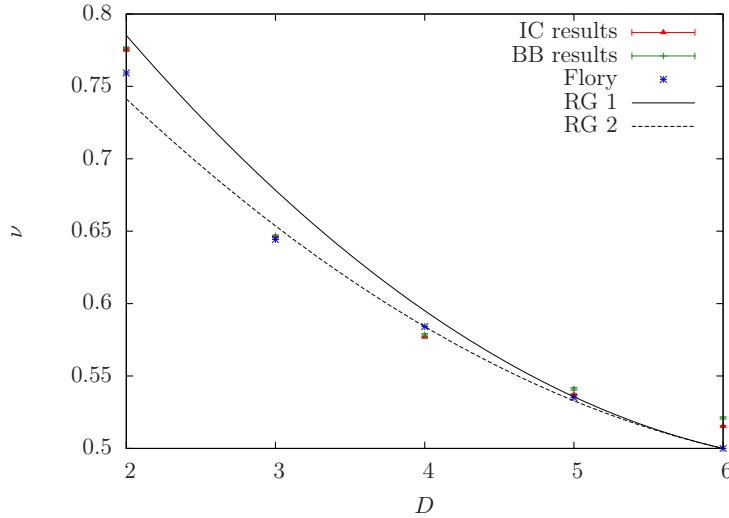


Figure 4.15.: Results for ν on CPCs (red) and backbones (green) for different dimensions juxtaposed with previous analytical estimates. The blue asterisks correspond to Flory estimates via Eq. (2.33) with α from Eq. (2.34). The lines represent the field-theory estimates, Eq. (2.37) (solid) and Eq. (2.38) (dashed).

of walks, the backbone, rather than the dangling ends, becomes negligible in the thermodynamic limit: its smaller fractal dimension implies that, within a fixed number of steps, almost none of the walks will visit it. In fact, the notion of an “infinite” SAW is misconceived to begin with. One should rather think finite SAWs whose length can be increased indefinitely, just as critical clusters can be indefinitely large but not, strictly speaking, infinite (see Section 2.2.1). Based on the data, I still conclude that $\nu_{IC} = \nu_{BB}$ is likely true even though the underlying argument is not. In any case, studying the backbone in place of the full clusters does not seem to pay off, as the asymptotic behavior is reached more slowly.

One important question that I did not investigate is whether there is multifractal behavior, i.e., whether the relations

$$[N^l] \sim |R|^{l/\nu_l} \quad (l \geq 1), \quad (4.17)$$

and

$$[|R|^l] \sim N^{l\nu_l} \quad (l \geq 1) \quad (4.18)$$

give rise to spectra of different exponents ν_l . The “standard” scaling law corresponds to ν_2 in Eq. (4.18), but the Meir-Harris model actually uses ν_1 of Eq. (4.17). In both Ref. [80] and Ref. [81], it was found that Eq. (4.17) leads to distinct exponents ν_l . However, it was claimed that Eq. (4.18) should not lead to different values [80].

4. Scaling behavior of self-avoiding walks on critical percolation clusters

A numerical study [82] seemed to confirm multifractality for Eq. (4.17) though the evidence is not very strong. This issue still needs to be clarified in the future.

4.2. Number of conformations

On a regular lattice, the number of SAW conformations increases with the number of steps asymptotically as

$$Z \sim \mu^N N^{\gamma-1} \quad (4.19)$$

where γ is a universal critical exponent and μ a system-dependent connectivity constant, see Section 2.1.2. It has long been assumed that the quenched average over all incipient critical clusters, $[Z_N]_{\text{IC}}$, follows a similar scaling law with modified values for γ and μ , but there is yet little evidence to support this.

The challenge in the last section was to extract the asymptotic behavior of the end-to-end distance by extrapolating averages from a finite number of steps. These averages could be estimated reliably and accurately with manageable effort. Although the standard deviations increased roughly as fast as the averages themselves, the number of replicas needed to maintain the same level of (relative) precision remained more or less constant with increasing N . This, unfortunately, is no longer the case for the average number of conformations. As shall be discussed in the following, estimating $[Z_N]$ for large N becomes a major challenge in itself due to fat tails in the distribution. As a consequence, assessing the quenched disorder average turned out to be much more difficult since little can be gained by sampling $[Z_N]$ directly even with the SFE method. However, information can be gleaned from the variance of the distribution Z_N in combination with the “average entropy” $[\ln Z_N]$, which is also an important quantity in its own right. In this section, I shall focus the discussions on the 2D and 3D cases. Unless stated otherwise, the behavior in higher dimensions was found to be qualitatively the same.

4.2.1. Median, mean, deviations and bias

In leading order the number of SAW conformations on a given cluster increases exponentially with the number of steps since for each conformation in step N there will be as many conformations in step $N + 1$ as there are options for the loose ends. For the quenched average $[Z_N]$ one can hence roughly expect

$$[Z_N] \sim m^N, \quad (4.20)$$

with m being the effective coordination number minus one. Actually, this is a lower bound since more spacious cluster regions contribute disproportionately to the increase of Z_N , just as more spacious clusters contribute disproportionately to the quenched average. In any case, the number of conformations quickly becomes quite large. For example, on a critical 3D cluster the number of 800-step conformations, Z_{800} , will typically be around 10^{99} . More precisely, this is the estimated median obtained from a sample of 10^6 clusters. The actual average, $[Z_{800}]$, is harder to assess. One can

4.2. Number of conformations

easily calculate the arithmetic mean value³, $\overline{Z_{800}} \approx 10^{116}$, but just as this value is much larger than the median, the actual expectation value will be much larger still. The reason is that typical deviations from the median also increase exponentially with N and the arithmetic mean therefore tends to be dominated by very few very large outliers (rare events). Hence, $\overline{Z_N}$ can only be relied upon as an estimator for $[Z_N]$ if those outliers are properly represented, which would require extremely large sample sizes. This problem is exemplified in Fig. 4.16, where the mean, median, and maximum value are plotted as function of the sample size. Even where the sample is already fairly large, the mean still increases quite suddenly as single large outliers completely alter the value. For insufficient sample sizes, the estimator for the average will almost certainly be biased, i.e., too small.

The longer the walks, the more severe does this bias problem become. In fact, unless sample sizes are increased drastically (probably exponentially) with N , the arithmetic sample mean will always be similar to the median rather than the true average. This can be seen in Fig. 4.17, where mean and median and maximum value are plotted as a function of N for a fixed sample size of 5×10^4 $2D$ clusters. Around $N = 100$, there is a visible change in the behavior of $\overline{Z_N}$, which is accompanied by larger error bars. These errors can only reflect the deviations from the arithmetic mean itself and thus fail to capture the bias. In the picture, one can make out three different regimes: For $N < 100$, the outliers are properly represented, and the estimator is more or less

³Throughout most of the work, I use the same notation for the theoretical expectation value and the measured average, but here I need to distinguish.

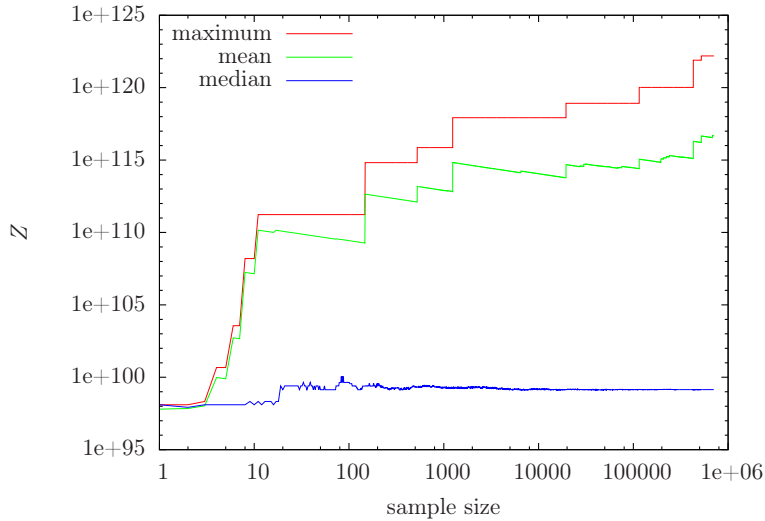


Figure 4.16.: Log-log plot of the maximal, mean, and median values of the number of conformations as function of the size of a random sample of $3D$ CPCs for SAWs of $N = 800$ steps.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

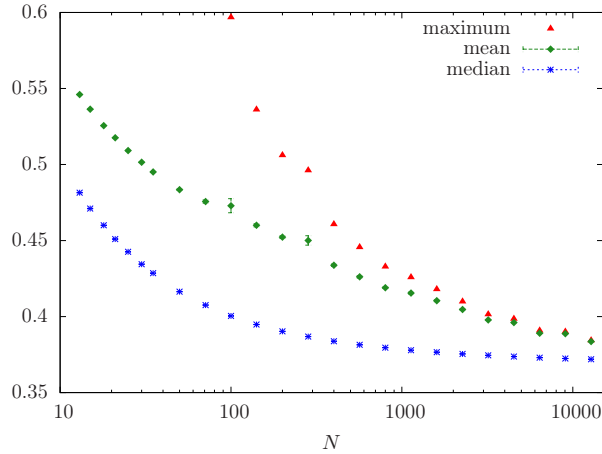


Figure 4.17.: Maximum, mean, median, and minimal numbers of conformations Z_N from samples of $S \approx 5 \times 10^4$ 2D CPCs.

unbiased; for $100 < N < 400$, the outliers are not fully represented which is manifested in strong fluctuations and large errorbars; and for $N > 400$, the relevant part of the distribution is missed completely, which is not reflected in the errors at all.

Could the bias be overcome? From a computational physics perspective, the problem with $[Z_N]$ is that we sample the clusters uniformly although the distribution of weights with which they contribute (i.e., Z) is very strongly peaked. The general strategy in such cases is importance sampling, which can be realized by Markov-chain Monte Carlo methods (e.g., Metropolis-Hastings [121]) or population-control growth algorithms (e.g., PERM). The first approach could be applied by evolving instead of growing the clusters. We would determine the number Z_N of SAW conformations on a cluster (using SFE) then suggest randomly removing or adding sites and accept the suggestion with a probability reflecting the resulting difference in Z_N . Crucially, one would have to modify the SFE method so that it does not need to start all-over each time a site is modified but could assess the change on the fly. It is credible but not certain that this might work. For the second approach, we would grow a population of clusters (using the Leath method described in Section A.1.1). Every so often, we would enumerate the numbers of (shortened) SAW conformations on each growing cluster. Thus we would determine the cluster “weights” and could prune and enrich them accordingly. Both approaches could potentially solve the problem but both would also be highly complicated.

4.2.2. Probability-density distributions

The probability of (randomly) generating a cluster allowing for Z_N to $Z_N + \Delta$ conformations can be written as

$$P(Z_N, \Delta) = \int_{Z_N}^{Z_N + \Delta} \rho_N(Z_N) dZ_N, \quad (4.21)$$

where it is convenient to consider $\rho_N(Z_N)$ as a continuous probability density function and Z_N as a continuous random variable. $\rho_N(Z_N)$ could in principle be approximated by random sampling to a histogram of bins whose size should increase exponentially to accommodate the large deviations described above. It is, however, more practicable to bin the distribution of the logarithms, $\rho'_N(\ln Z_N)$, to equally sized bins. The distribution $\rho_N(Z_N)$ can then be obtained via

$$\rho_N(Z_N) = \rho'_N(\ln Z_N) \frac{d \ln Z_N}{dZ_N} = \rho'_N(\ln Z_N) / Z_N. \quad (4.22)$$

The histograms pertaining to $\rho'_N(\ln Z_N)$ closely resemble Gaussian distributions as can be seen in Fig. 4.18 for the case of 3D critical clusters and their backbones. This would imply a lognormal distribution of Z_N , whose fat tails would indeed account for the large deviations. As was pointed out in Ref. [72], such a distribution can be motivated by regarding Z_N as a product of N (uncorrelated) random variables, namely the (weighted) average number of free neighbors per step. While this captures the essence of what is going on, it is only a qualitative explanation as these numbers would in fact be correlated. A closer look at the distributions for the full clusters (Fig. 4.18, top) does indeed reveal slight but systematic deviations from normal distributions with the same mean (also shown in the plots): the peaks are a bit too high and, more importantly, the tails (at large values) a bit too small. Indeed, the distributions on the full clusters have a slightly negative skewness, $\gamma_1 < 0$, and all have a slightly positive excess kurtosis $\gamma_2 > 0$ as shown in Fig. 4.19. Skewness and excess kurtosis are defined as usual:

$$\gamma_1 := \mu_3 / \sigma^3 \quad (4.23)$$

and

$$\gamma_2 := \mu_4 / \sigma^4 - 3 \quad (4.24)$$

with μ_i denoting the i 'th central moment of the distribution. They are zero for a normal distribution, and their values are a good measure of the ‘‘normality’’ of a distribution. Here they are relatively small in all four cases but especially so on the backbones, where the resemblance to a Gaussian is, for some reason, particularly strong.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

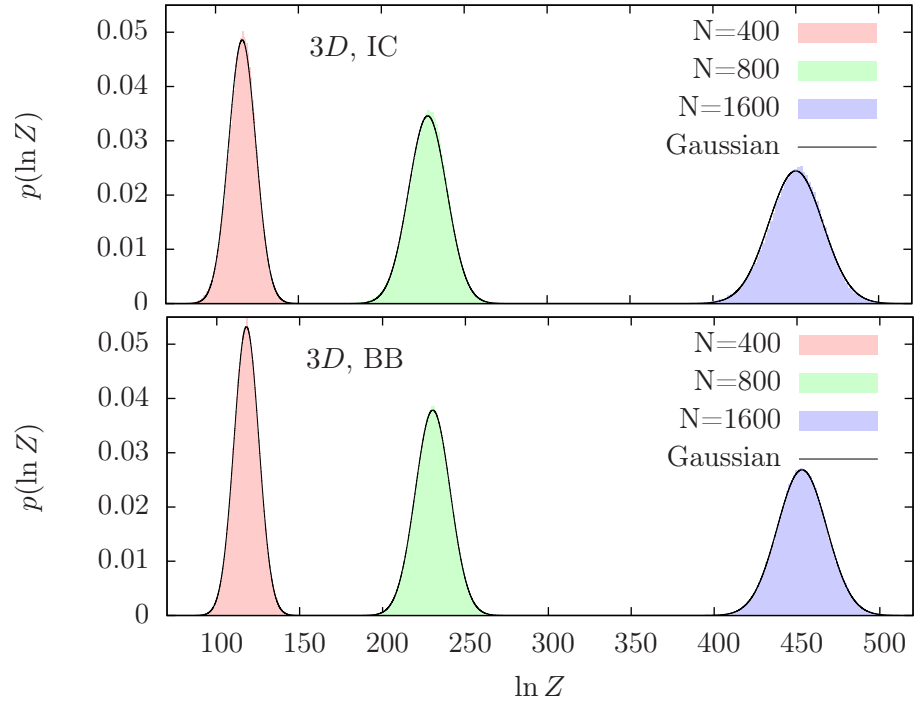


Figure 4.18.: Measured probability densities of the entropy $\ln Z_N$ on incipient CPCs (top) and backbones (bottom) for various lengths. The lines are normal distributions with the same mean values and variances (no fit involved).

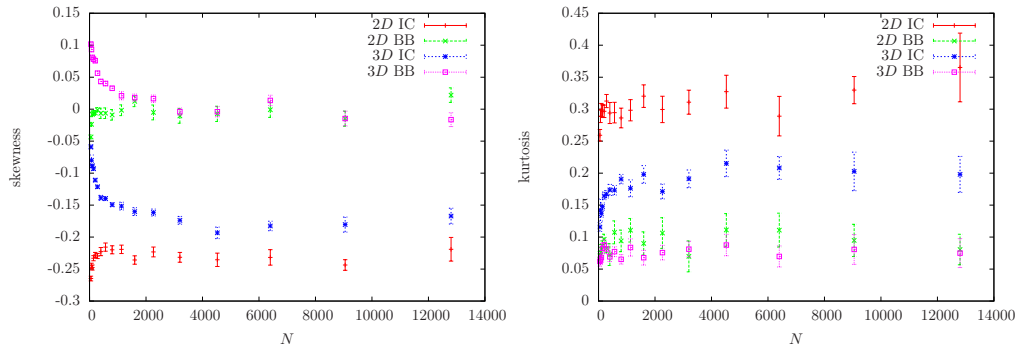


Figure 4.19.: Measured skewness (left) and excess kurtosis (right) of the distributions of $\ln Z_N$ on critical clusters and their backbones in $2D$ and $3D$.

4.2. Number of conformations

Interestingly, the increase of the variances of $\rho'_N(\ln Z_N)$ with N , appears to be very close to linear, both on backbones and full clusters in all dimensions studied:

$$\sigma_{\ln Z}^2 \sim N^\phi, \quad \phi \approx 1 \quad (4.25)$$

This can be seen in Fig. 4.20, where the variances divided by N are plotted against N . The slopes of all curves become roughly constant, suggesting $\phi \approx 1$. Indeed, the exponents obtained from least-squares fits listed in Table 4.7 are almost all consistent with $\phi = 1$, with the odd exception $D = 2$ (Note that the fit errors are perhaps too optimistic and that the estimates for ϕ seem to increase slightly but systematically when the range is shifted to larger N). The $2D$ case also stands out in that the variances are actually smaller than in $3D$ in the range of N that I studied even though the values of $\ln Z$ are typically much larger. The estimates of the statistical errors on the variances were obtained by bootstrap resampling. Unfortunately they are not correct for $D > 3$, especially on the backbones, which is due to the fact that multiple starting points were used from each cluster (see Section 4.1.3). These correlations caused an underestimation of the errors because the variances of different bootstrap resamples were less diverse. The effect is more drastic on the backbones, where the correlations are much stronger since they offer less room for different starting points. However, the impact on the actual estimates for A and ϕ (and their errorbars) is probably still negligible.

For $D = 2, 3$, $\phi \approx 1$ was already noted in previous studies [72, 73, 77] (another study found $\phi \approx 1.3$ [71]). It is very tempting to assume $\phi = 1$ exactly for $D > 2$, but so far there is no argument to explain this remarkable value, nor the $2D$ exception. Note that without self-avoidance, i.e., for the ideal chain, ϕ is significantly larger than one [122].

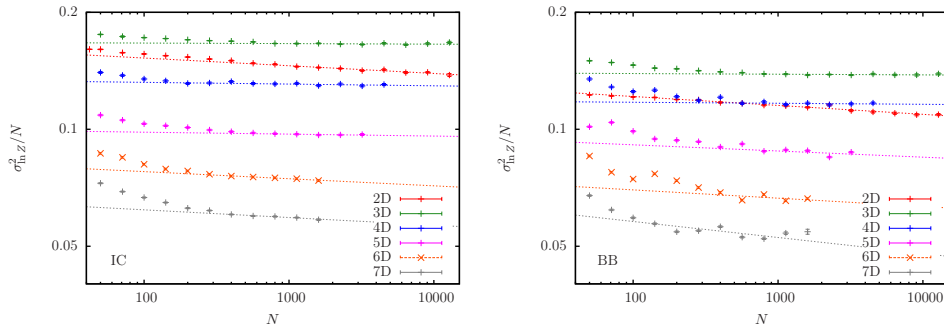


Figure 4.20.: Variances of the distribution of $\ln Z_N$ divided by N on CPCs (left) and their backbones (right) for systems of different dimensionality. (The errors are underestimated for $D > 3$.)

4. Scaling behavior of self-avoiding walks on critical percolation clusters

Table 4.7.: Measured exponents ϕ and amplitudes A for the variance of the entropy distribution as function the number of steps on CPCs (IC) and their backbones (BB) in different dimensions.

D	fit range	χ_{IC}^2	A_{IC}	ϕ_{IC}	χ_{BB}^2	A_{BB}	ϕ_{BB}
2	800-12800	1.03	0.167(3)	0.980(2)	1.39	0.135(4)	0.977(3)
3	800-12800	1.01	0.170(3)	0.997(3)	0.853	0.140(2)	0.998(2)
4	800-4525	0.965	0.135(5)	0.995(6)	1.08	0.119(5)	0.997(5)
5	566-3200	0.590	0.102(2)	0.993(3)	4.10	0.11(2)	0.97(2)
6	400-1600	0.508	0.085(2)	0.982(3)	4.42	0.08(1)	0.98(2)
7	400-1600	0.773	0.068(2)	0.980(3)	4.48	0.06(2)	0.97(3)

4.2.3. Scaling behavior of the average entropy $[\ln Z_N]$

The distribution of the entropies $\ln Z_N$ does not have fat tails, so that $[\ln Z_N]$ is much easier to study than $\ln [Z_N]$. The exponential version,

$$Z^{(0)} := e^{[\ln Z]}, \quad (4.26)$$

has also been investigated to understand the multifractal nature of the system, see Ref. [78]. There, the authors assumed a scaling form similar to (4.19) as well and estimated the “zeroth moment” exponent $\gamma^{(0)}$ and connectivity constant $\mu^{(0)}$ by fitting to their data (obtained through exact enumeration):

$$[\ln Z]/N = \ln a^{(0)}/N + \ln \mu^{(0)} + (\gamma^{(0)} - 1) \ln N/N. \quad (4.27)$$

In contrast to the situation with $\ln [Z]$, the SFE method now allows for a much larger range to be investigated. However, my results from fitting Eq. (4.27) turned out to strongly depend on the range, with fits over larger ranges being rather unconvincing. Empirically, a much better fitting ansatz was instead found to be⁴

$$[\ln Z]/N = \ln a^{(0)}/N + \ln \mu^{(0)}(1 + bN^{-\zeta}) \quad (4.28)$$

Fits of both approaches, Eq. (4.27) and Eq. (4.28), are shown in Fig. 4.21 for the 2D and 3D cases. The results from the second approach are listed in Table 4.8 (those from the first are meaningless as the quality of the fits was dismal). Despite the extra parameter in Eq. (4.28), the quality of the fit is much more significant; the fairly small errorbars in Fig. 4.21 are almost perfectly hit for all data points. For $D \geq 4$, the smaller data sets only allowed fitting Eq. (4.28) when the amplitude was fixed to $a^{(0)} = 1$. This fix may seem dubious, but $a^{(0)} \approx 1$ is a reasonable guess (given also the estimates for $a^{(0)}$ for 2D and 3D), and the influence of the term $\ln a^{(0)}/N$ is quite modest anyway. The results from these fits are also given in Table 4.8. Here, the error intervals were chosen to represent the changes of the parameters when $a^{(0)}$ is varied

⁴A different parametrization, $[\ln Z]/N = \ln a^{(0)}/N + \ln \mu^{(0)} + \ln cN^{-\zeta}$, would give Eq. (4.30) a cleaner form. However, I kept the above expression to remain consistent with Ref. [106].

4.2. Number of conformations

between 0.75 and 1.5. The errors from the fits are slightly smaller.

Although there is no theoretical justification for (4.28) (there actually is none for (4.27) either other than analogy), it must therefore be assumed to be correct. There might still be a power-law term with an exponent $\gamma^{(0)}$, but if so it could not be extracted from the data as the other terms are too dominant.

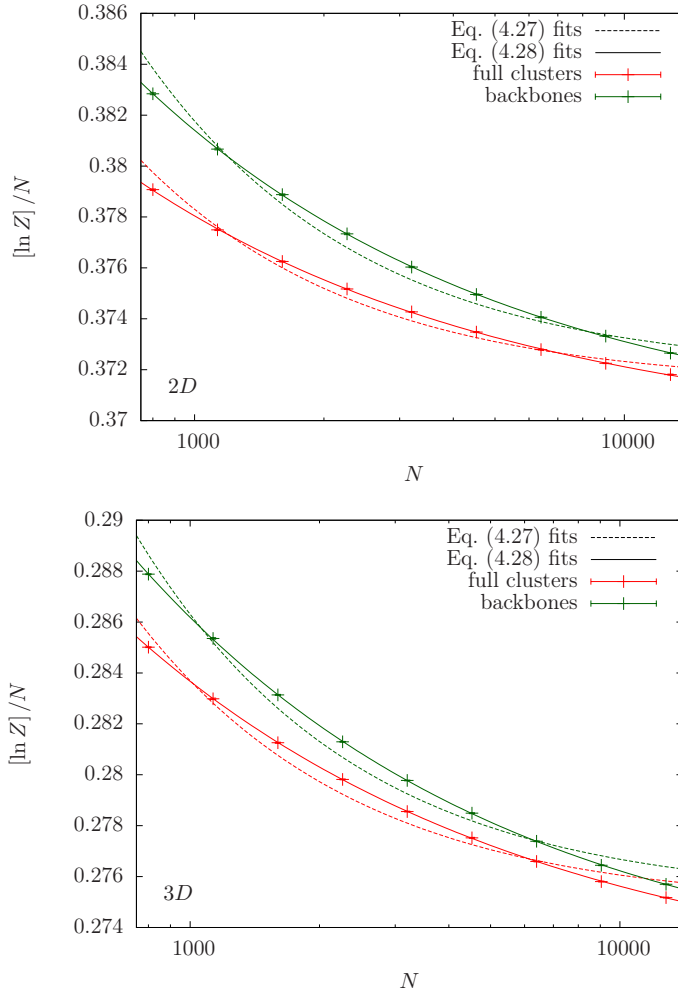


Figure 4.21.: Average entropy per step for SAWs on critical percolation clusters (red) and their backbones (green) in dimensions $D = 2$ (top) and $D = 3$ (bottom) plotted on log-linear scale. The curves correspond to fits of Eq. (4.27) (dashed) and (4.28) (solid).

4. Scaling behavior of self-avoiding walks on critical percolation clusters

Table 4.8.: Results obtained by fitting Eq. (4.28) to the average entropy per step $[\ln Z]$ for SAWs on critical percolation clusters (top) and backbones (bottom) in different dimensions. For $D \geq 4$, the amplitude was fixed to $a^{(0)} = 1$.

IC						
D	fit range	χ^2	$a^{(0)}$	$\mu^{(0)}$	b	ζ
2	566-12800	1.44	1.7(1.0)	1.4462(6)	0.5(1)	0.44(6)
3	566-12800	0.650	1.5(5)	1.3115(3)	0.9(2)	0.43(3)
4	400-4525	0.852	1 (fixed)	1.2143(3)	2.1(4)	0.49(3)
5	283-3200	4.21	1	1.1551(4)	3.4(7)	0.51(3)
6	200-1600	0.845	1	1.1189(6)	5(1)	0.55(4)
7	200-1600	0.626	1	1.0960(5)	8(2)	0.59(3)
BB						
D	fit range	χ^2	$a^{(0)}$	$\mu^{(0)}$	b	ζ
2	566-12800	0.485	1.0(6)	1.4467(3)	1.0(3)	0.50(3)
3	800-12800	1.18	0.8(7)	1.3121(4)	1.8(6)	0.51(4)
4	400-4525	0.453	1 (fixed)	1.2143(3)	2.3(4)	0.52(3)
5	283-3200	0.670	1	1.1561(2)	4.3(8)	0.62(4)
6	200-1600	2.84	1	1.1198(2)	7(3)	0.71(4)
7	200-1600	0.0941	1	1.0957(2)	5(2)	0.67(6)

4.2.4. Log-normal approximation

The mean entropy also gives some evidence regarding the scaling behavior of the average number of chains. If we assume the distribution of $\ln Z$ to be Gaussian, which is not exactly true but fairly accurate as an approximation, $[Z]$ can be obtained as the expected value of the corresponding (log-normal) distribution:

$$[Z] \approx Z_{\log n} := e^{[\ln Z] + \sigma_{\ln Z}^2 / 2}. \quad (4.29)$$

That this approximation is reasonable can be seen in Fig. 4.22, where $\ln \bar{Z}$, $[\ln Z]$, and $\ln Z_{\log n}$ are plotted as functions of N . $[\ln Z]$ is practically identical to the logarithm of the median in Fig. 4.17 and is significantly smaller than $\ln \bar{Z}$, at least initially. As already discussed, the mean value becomes biased for larger N , underestimating the true average. The log-normal approximation $Z_{\log n}$ agrees very well with $\ln \bar{Z}$ in the initial (unbiased) regime, in particular in $3D$ and on the backbones, where the distributions are closest to normal (see Fig. 4.19). However, in contrast to $\ln \bar{Z}$, it does not slump for larger N and seems a plausible continuation of the curve for $\ln \bar{Z}$. This suggests that $[Z]$ might well be governed by a scaling law analogous to an exponential version of Eq. (4.28). If we assume $\phi = 1$ in Eq. (4.25), combining Eq. (4.29) and Eq. (4.28) gives:

$$[Z] \approx Z_{\log n} \sim a^{(0)} \mu^N \tilde{\mu}^{N^{1-\zeta}}, \quad \mu = \mu^{(0)} e^{A/2}, \quad \tilde{\mu} = \mu^{(0)b} \quad (4.30)$$

4.2. Number of conformations

where A is the amplitude in Eq. (4.25). The estimates for μ are given in Table 4.9.

If correct, this finding for the scaling behavior of $[Z]_{\text{IC}}$ would be quite surprising and currently not explained by any theory, especially in view of the arguments concerning $[Z]_{\text{AC}}$ from Section 2.3.4. A note of caution, though: The stretched-exponential correction for $\exp[\ln Z]_{\text{IC}}$ is in itself not surprising. As was already remarked in Ref. [41], $\exp[\ln Z]$ should behave differently from $[Z]$ due to the width of the distribution of Z (though this particular behavior was not predicted anywhere yet). The numerical evidence for Eq. (4.28) is strong, but the argumentation for Eq. (4.30) hinges on the (asymptotic) accuracy of approximation Eq. (4.29), which is much less certain.

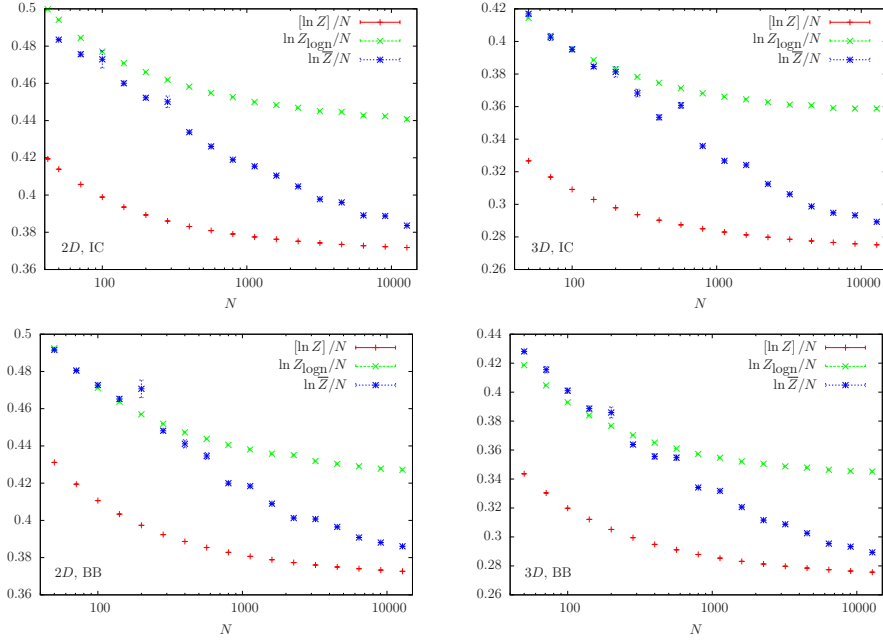


Figure 4.22.: Logarithm of the mean number of conformations $\ln \bar{Z}$ (blue), mean entropy $[\ln Z]$ (red), and log-normal approximation $\ln Z_{\text{logn}}$ (green) divided by N for SAWs on CPCs (top) and backbones (bottom) in 2D (left) and 3D (right).

Table 4.9.: Estimates for the connectivity constant obtained via $\mu = \mu^{(0)} e^{A/2}$, where the amplitude A stems from a fit of Eq. (4.25) with $\phi = 1$.

D	2	3	4	5	6	7
μ_{IC}	1.553(1)	1.4250(5)	1.2961(5)	1.2126(7)	1.162(1)	1.1292(8)
μ_{BB}	1.5306(5)	1.4061(7)	1.2871(5)	1.2080(4)	1.1578(4)	1.1255(3)

4.2.5. All cluster (AC) averages

As explained in Section 2.3.4, when the quenched average is taken over all disorder configurations that support at least one N -step SAW conformation it should scale as

$$[Z_N]_{\text{AC}} \sim (p\mu)^N N^{\gamma+\delta-1}, \quad (4.31)$$

where γ and μ are the same as for the free SAW [Eq. (2.6)]. The modification δ to the enhancement exponent relates to the probability Q_N that a random disorder can accommodate an SAW of N steps; see Eq. (2.41). Figure 4.23 shows the fraction Q_N of contributing disorder configurations as a function of N for $2D$ systems. (As mentioned, I only investigated the AC average for the $2D$ case). Each data point represents an independent sample of over 6×10^4 configurations (at least 3×10^4 contributing clusters). A simple power-law fit yielded the estimate

$$\delta = 0.0640(6) \quad (4.32)$$

with an amplitude $A_\delta = 0.628(3)$, in accordance with the lower bound of $5/98 \approx 0.051$ [Eq. (2.44)].

Since γ and μ are known accurately for the free $2D$ SAW (see Section 2.1.2), Eq. (4.31) can now be checked against the data. Of course, measurement of $[Z]_{\text{AC}}$ is also hampered by large deviations, so that agreement is only expected in the initial regime. Figure 4.24 shows the results for the average $\ln \bar{Z}_{\text{AC}}$ per monomer in the range

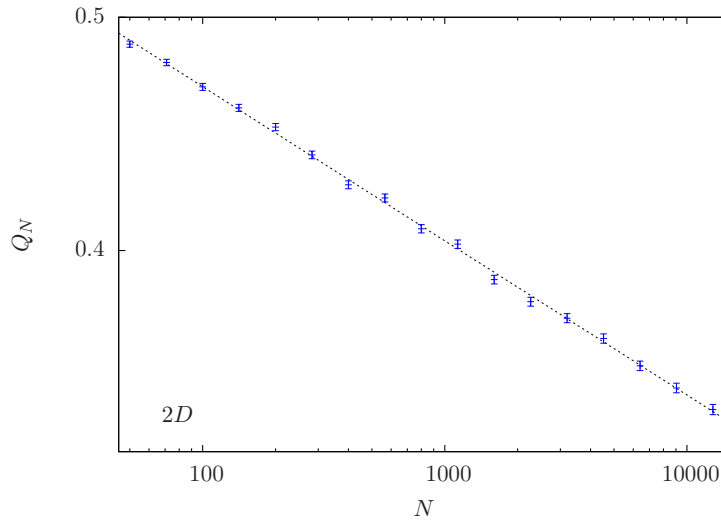


Figure 4.23.: Fraction of $2D$ disorder configurations that support at least one N -step SAW conformation. The dotted line corresponds to a least-squares power-law fit over the whole data range.

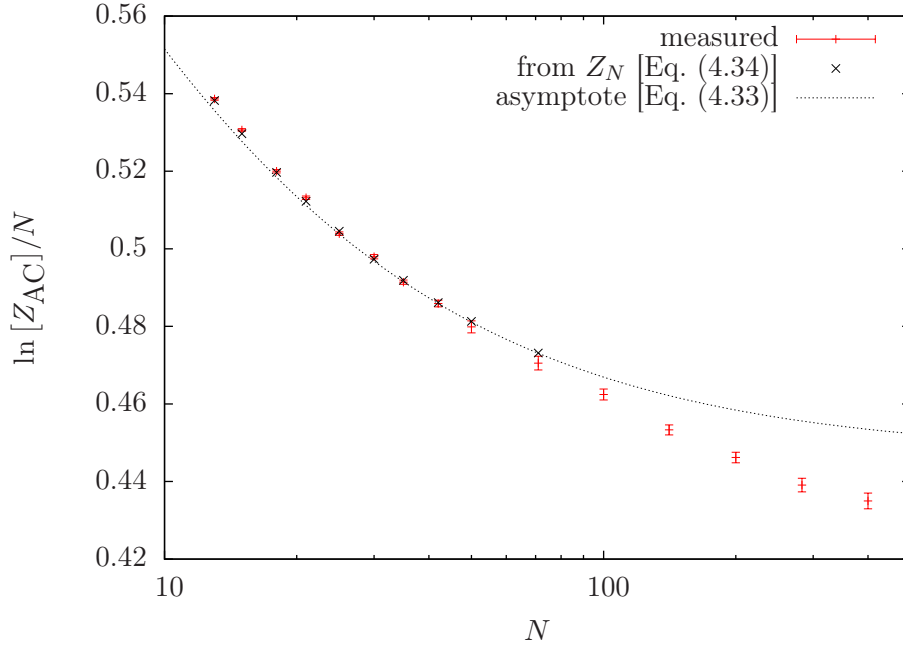


Figure 4.24.: Measured logarithm of the (AC) average number of conformations per monomer (red pluses), estimate based exact values for the free square-lattice SAW [Eq. (4.34)] (black crosses) and estimate of the asymptote [Eq. (4.33)] (dotted line).

$N \leq 400$. Also shown in the figure are the expected asymptote,

$$\ln [Z]_{\text{AC}} \sim (\ln A + N \ln(p\mu) + (\gamma + \delta - 1) \ln N)/N \quad (4.33)$$

with an amplitude of $A = pA_{\text{free}}/A_{\delta} \approx 0.438$ as well as the estimates from the exact values Z_N for the free SAW taken from Refs. [107, 123]:

$$\ln Z_N + (N + 1) \ln p + \delta \ln N - \ln A_{\delta} \quad (4.34)$$

As expected, the directly measured values initially agree well with the estimates based on the free SAW. There is in fact no significant difference to Eq. (4.33) for $N \leq 40$, strongly suggesting that the AC average is indeed given by Eq. (4.31). The deviance for larger N is due to the aforementioned bias, which seems to appear around $N = 50$. It is reasonable to assume that, for similar sample sizes, the $[Z]_{\text{IC}}$ is also unbiased for $N \leq 40$.

The AC average of the entropy behaves similarly to the IC average: As shown in Fig. 4.25, an ansatz of the form of Eq. (4.27) fits the data poorly, and it is much better

4. Scaling behavior of self-avoiding walks on critical percolation clusters

described by Eq. (4.28). However, the samples are a bit smaller as for the IC case and fluctuations are larger, so that a larger range had to be taken and the results are less accurate. My best estimate was obtained by fitting over the range $N \geq 50$:

$$[\ln Z]_{AC} = \ln a^{(0)} + N \ln \mu^{(0)} + bN^{1-\zeta} \quad (4.35)$$

with $a^{(0)} = 1.8(2)$, $\mu^{(0)} = 1.4390(5)$, $b = 0.5(1)$, and $\zeta = 0.46(4)$.

This result implies that the log-normal approximation, Eq. (4.29), does not work for the AC average as its prediction would contradict Eq. (4.31). However, the AC distributions of $\ln Z$ are also much less Gaussian, with a skewness of $\gamma_1 < -6$ and an excess kurtosis of $\gamma_2 > 50$. The failing of the approximation for the AC case does therefore not mean that it is also incorrect for the IC case. Another interesting difference to the IC average is that the variance of the distribution of $\ln Z$ increases more strongly. As can be seen in Fig. 4.26, the initial increase is roughly linear (this was also found in Ref. [72]), but then there appears to be a crossover to an exponent of almost two. Fitting over the range $N \geq 2263$, I obtained the estimate

$$\sigma_{\ln Z_{AC}}^2 \sim N^{\phi_{AC}}, \quad \phi_{AC} = 1.92(2). \quad (4.36)$$

The variance could be expected to be larger over all clusters, but this finding suggests that there is indeed a qualitative difference between the two averages.

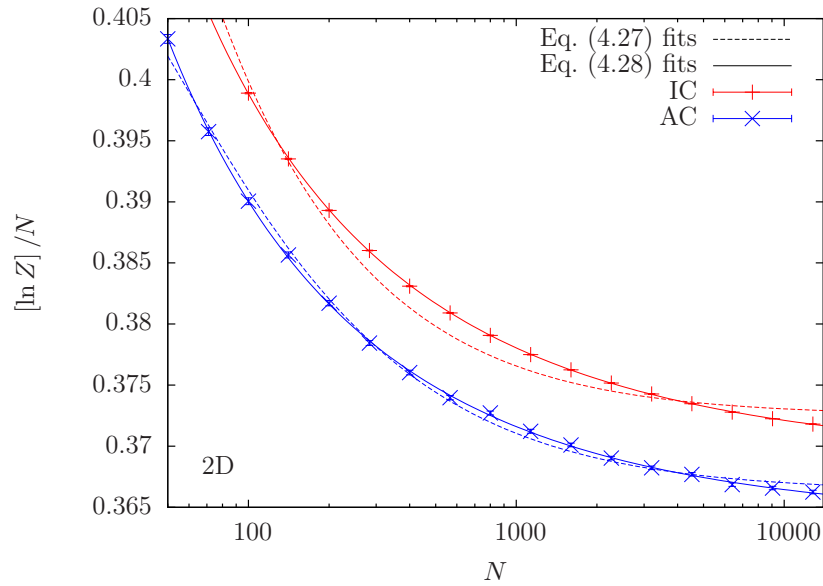


Figure 4.25.: AC average of the entropy $\ln Z$ (blue crosses) fitted via Eq. (4.27) (dashed line) and Eq. (4.28) (solid line). The IC results (red) are shown again for comparison.

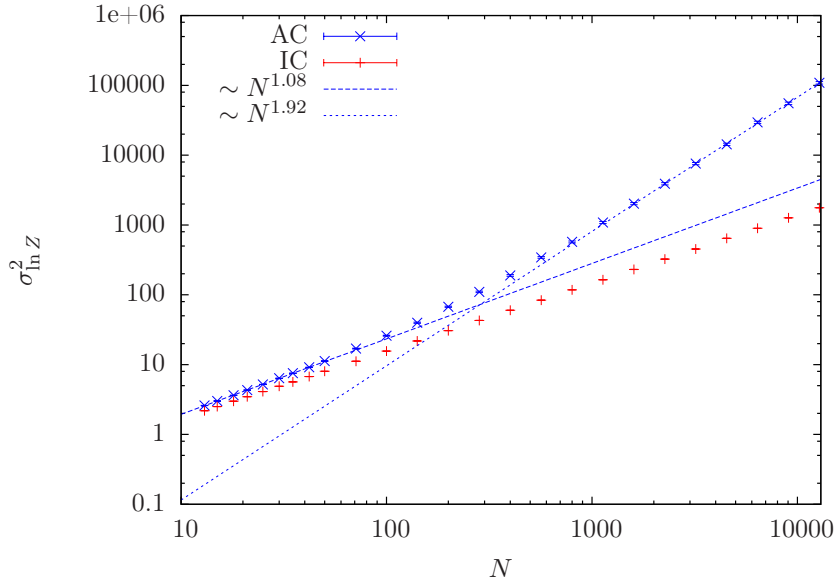


Figure 4.26.: Variance of the entropy $\ln Z$ for all contributing clusters (blue crosses). The dashed line correspond to a fit over the initial range ($N \geq 50$); the dotted one to a fit over the later range ($N > 2000$). The red crosses show the variance of the IC distributions for comparison.

4.3. A closer look at individual clusters

When investigating SAWs on CPCs, one is usually interested in the quenched disorder averages, and so far that has been the focus of the discussion. Zooming in on a few individual cluster configurations may seem arbitrary, but it can in fact reveal interesting details and thereby help understanding the general behavior. For a small number of configurations, I could afford pushing the chain length to $N = 2 \times 10^4$ on CPCs in $2D$ and $3D$. Plots of the measured mean squared end-to-end distances on linear and logarithmic scales are shown in Fig. 4.27 for $2D$ and Fig. 4.28 for $3D$. Also shown are the “excess entropies”,

$$S_e = \ln Z - N \ln \mu^{(0)}, \quad (4.37)$$

describing how much the increase of the number of conformations exceeds the (typical) connectivity constant to the power of N , thus characterizing the cluster’s individual structures. Note that the fluctuations in these plots are only visible when the precise value of $\mu^{(0)}$ (not μ) is used.

For both the end-to-end distance and the excess entropy, the curves are strikingly irregular. The magnitude of the fluctuations of $\langle R^2 \rangle$ is in fact roughly on the order of

4. Scaling behavior of self-avoiding walks on critical percolation clusters

the values themselves, and they are not only present from cluster to cluster but also when the length is increased on the same cluster. In the log-log plots of the end-to-end distances, one can nonetheless clearly discern a typical slope, whose value is close to the quenched average value of 2ν .

The shapes of the $\langle R^2 \rangle$ -curves are also remarkable: They show wide plateaus, where the value is fairly constant over a long range of N , separated by abrupt jumps and drops. There are also intervals where the changes are relatively smooth, but they seem to become less important with increasing length. Correlations between the end-to-end distances and the excess entropies are not immediately obvious, but careful observation reveals that S_e tends to rise when $\langle R^2 \rangle$ changes. The increase then typically continues through the initial parts of the $\langle R^2 \rangle$ -plateaus but falters or even reverses during the later parts.

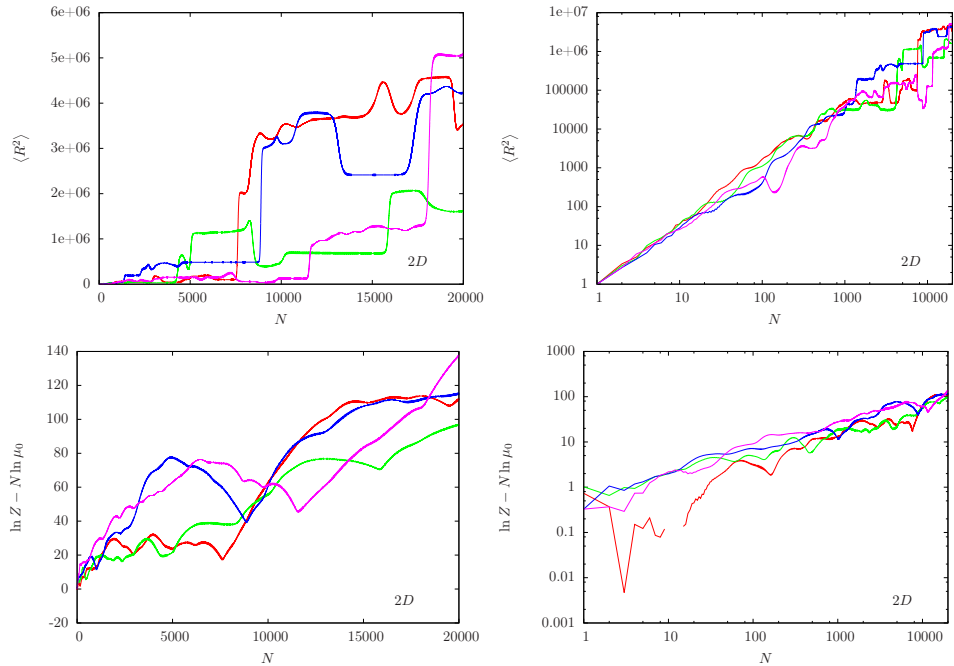


Figure 4.27.: Mean squared end-to-end distances (top) and excess entropy for SAWs of 2×10^4 steps on some randomly chosen 2D CPCs. The plots on the right are on a log-log scale where the average trend for the end-to-end distances is more clearly visible.

4.3. A closer look at individual clusters

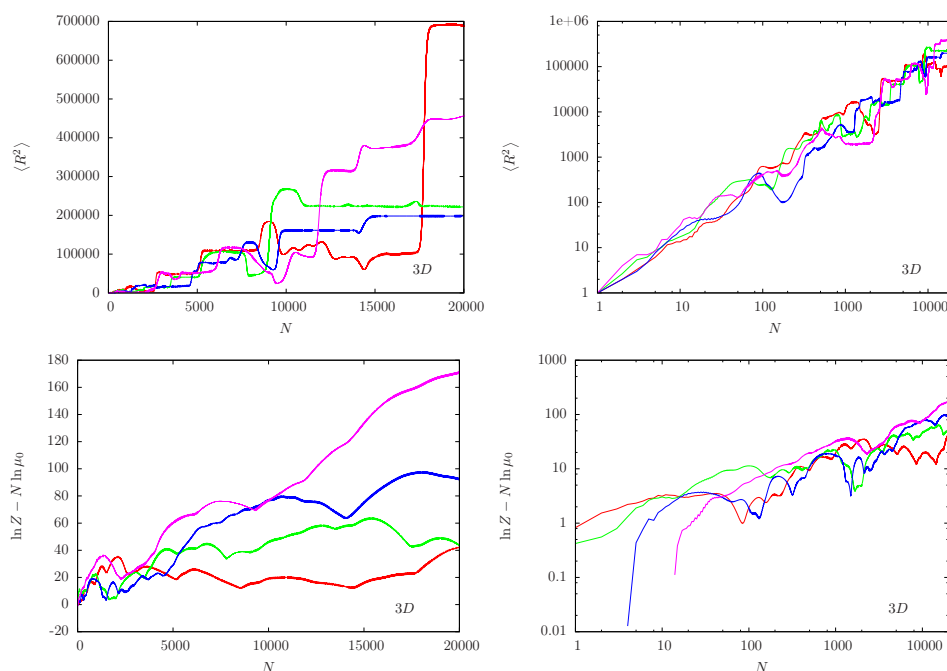


Figure 4.28.: Mean squared end-to-end distances (top) and excess entropy for SAWs of 2×10^4 steps on some randomly chosen 3D CPCs. The plots on the right are on a log-log scale where the average trend for the end-to-end distances is more clearly visible.

4.3.1. Density distributions

To understand the behavior, it is helpful to look at the spatial distributions of end points and the fraction of conformations that pass through each site, the SAW *conformation density*. As explained in Section 3.2.4, generating these distributions requires significantly more resources, so that the systems could not be made as large.

Fig. 4.29 shows the end-to-end distances for up to $N = 1600$ steps for four exemplary, randomly chosen 2D clusters. For these reduced lengths, the jumps and plateaus are much less distinct, but it is reasonable to assume that the steeper and flatter regions in these small systems are due to the same mechanisms as the clearer jumps and plateaus in large systems.

I systematically looked at the distributions of end-points as a function of N on about ten different clusters. As a fairly representative example, I shall discuss “cluster 3” from Fig. 4.29. Snapshots of the distributions at different lengths are shown in Fig. 4.30. Note that only the relevant region of the cluster is shown; the total area in reach (N chemical shells from the origin) is significantly larger. As can be seen, there tend to be a few “hot spots” where the vast majority of SAW conformations end,

4. Scaling behavior of self-avoiding walks on critical percolation clusters

while the rest of the cluster is fairly deserted. These *hot spots* correspond to plateaus of the end-to-end distances, whereas a jump means that one is suddenly supplanted by another. For instance, after 400 steps (not shown) the blue curve in Fig. 4.29 has a clear plateau and almost all walks end within a very small region. After 540 steps (Fig. 4.30, top row), this region has lost its appeal, and several other regions are competing for dominance. At about $N = 580$ (second row), there is again one dominating region corresponding to the next plateau, which is eventually succeeded by another one, corresponding to the plateau around $N = 800$ (third row). However, the conformation density plot (right) reveals that the walks still pass through the previously dominant region, i.e., its entropic contribution still matters. Interestingly, after $N = 1180$ steps (fourth row), most conformations again terminate there. This fact accounts for the trough in end-to-end distance (note that the value there is roughly the same as around $N = 600$). From there on, a more distant region steadily gains in relevance, becoming the strongly preferred ending spot after $N = 1600$ steps (bottom row). Note that the regions where most walks end after 580 and 800 steps are still entropically relevant, being traversed by almost all conformations. While such a lasting influence seems to be common, there are also regions that dominate the average end-to-end distance for a certain length, but become completely irrelevant for larger N .

The picture for $N = 1600$ appears to be very typical for longer walks, as can be seen in Fig. 4.31, where the final distributions for the other clusters from Fig. 4.29 are plotted. There is a very concentrated “flow” of conformations towards one specific area. The relevant cluster regions through which almost all conformations pass, have almost linear “pearl-necklace” structures. The density is close to one on some “bottleneck sites”, which appear in a dark red in the pictures. Between those, we can see blobs of yellow and blue colors in which the density is spread out. These are typically regions with strong entropic contributions, and they often correspond to the places where most walks ended for some smaller number of steps.

It is remarkable that, after a large number of steps, such a thin pearl-necklace does

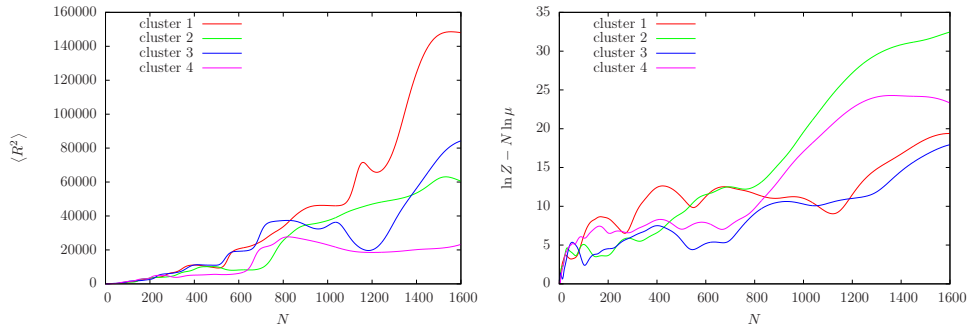


Figure 4.29.: Mean squared end-to-end distance (left) and excess entropy (right) as functions of N for some exemplary 2D CPCs. The blue curve corresponds to the distributions shown in Fig. 4.30.

4.3. A closer look at individual clusters

almost exclusively determine the statistics for the whole cluster. The distributions of conformation densities do not seem to disperse at all, nor do those of the end-points. This is in stark contrast to the behavior of diffusive random walks or SAWs in undiluted (regular) systems.

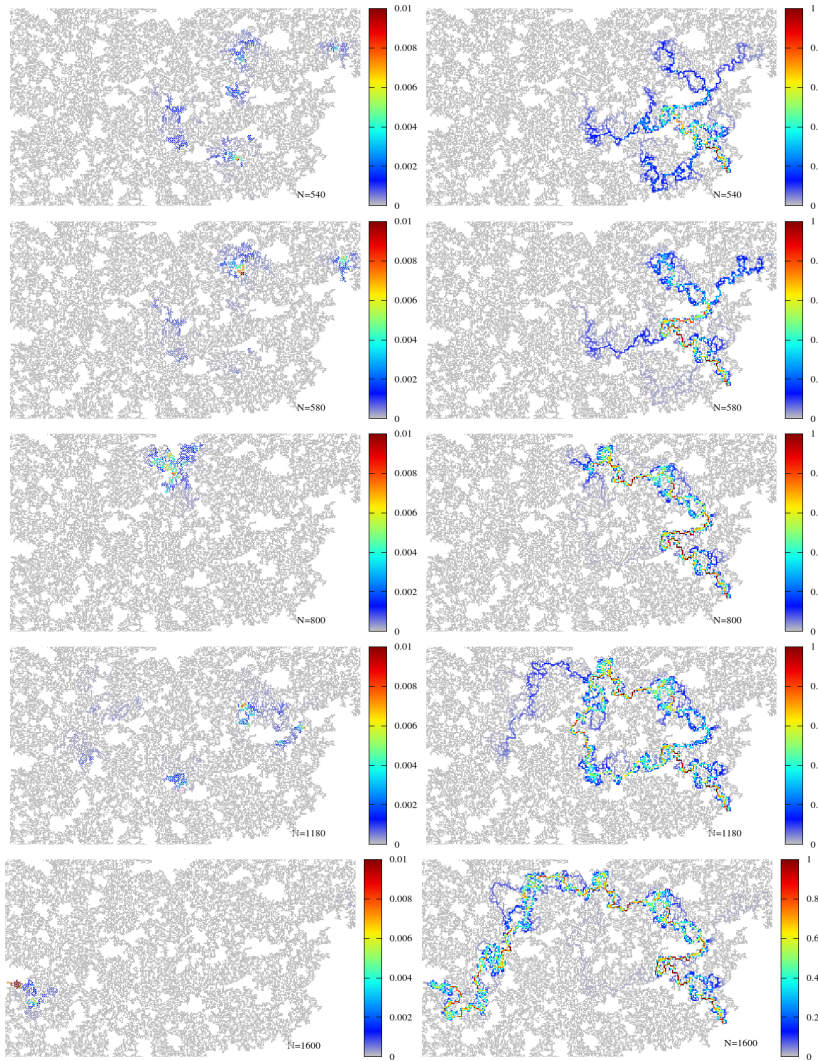


Figure 4.30.: Distributions of the end-points (left) and conformation densities (right) after 540, 580, 800, 1180 and 1600 steps (from top to bottom) on “cluster 3” from Fig. 4.29.

4. Scaling behavior of self-avoiding walks on critical percolation clusters

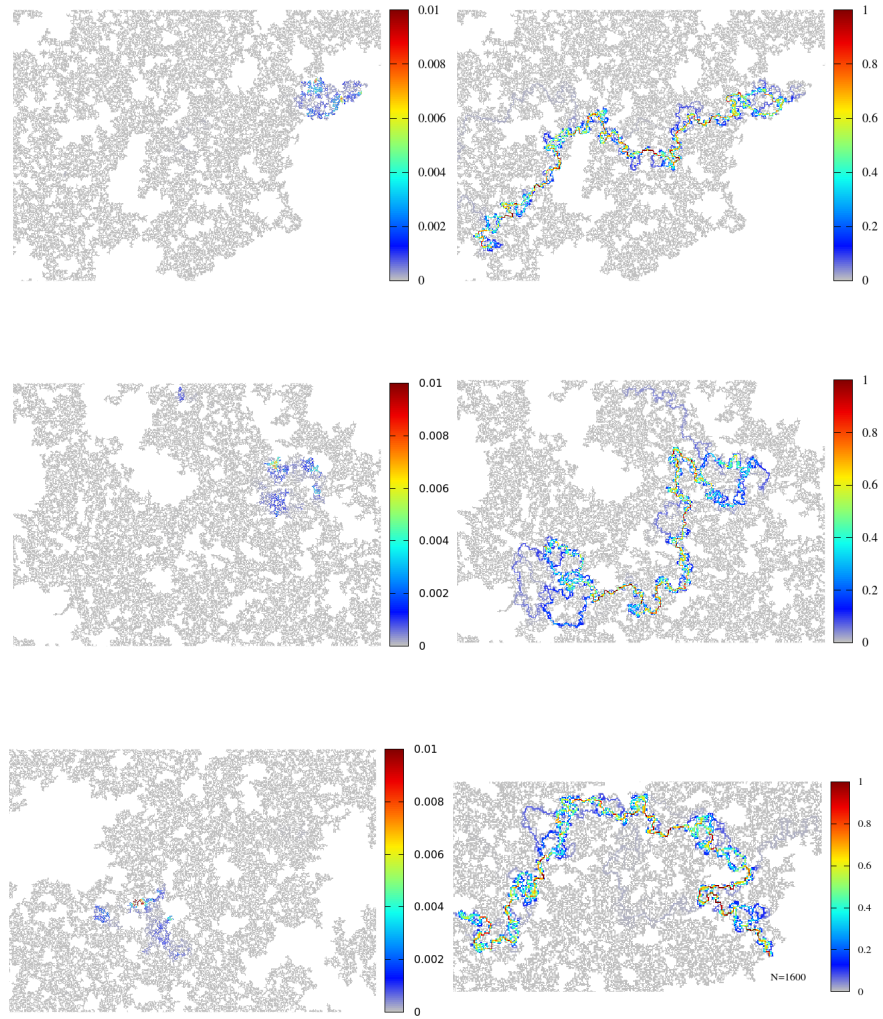


Figure 4.31.: End-point distributions (left) and conformation densities (right) for SAWs of $N = 1600$ steps on the ($2D$) clusters 1, 2 and 4 from Fig. 4.29.

4.3.2. Summary and Ideas

The key observations from above description can be summarized as follows:

- $\langle R^2 \rangle$ and S_e fluctuate strongly with changing N and from cluster to cluster.
- Curves of $\langle R^2 \rangle$ vs. N feature wide plateaus and sudden jumps.

4.3. A closer look at individual clusters

- The excess entropy $S_e(N)$ correlates with $\langle R^2 \rangle(N)$, typically rising after a jump in $\langle R^2 \rangle$ and falling towards the end of a plateau.
- The majority of SAW conformations (for fixed N) live on one thin perl-necklace region and end in one small area.
- Dominating end regions correspond to plateaus in the $\langle R^2 \rangle$ curves; jumps occur when one such region is replaced by another.
- When N is increased, regions that were previously preferred end locations often, but not always, are still traversed by most conformations.
- Distributions of the end-point densities and the conformation density do not seem to disperse with increasing N . For the end-points, the maximum remains roughly 0.01, while the conformation density remains ≈ 1 for all bottle-neck sites along the relevant perl-necklace region.

I interpret these observations within the following qualitative picture: The cluster is a network of largely separate “blobs”, which vary in size and density (i.e., connectivity). Let us assume, however, that they are of a typical size, like the “pearls” in Fig. 4.31. The fluctuations around this typical size (and connectivity) are probably Gaussian, but due to the exponential increase of SAW conformations, they cause large deviations in the number of SAWs that can pass through the blobs. If we consider different strings of blobs, the corresponding distribution of realizable SAW conformations will probably have a diverging variance, giving rise to non-self-averaging behavior: there is usually only one entropically favorable string of blobs through which almost all conformations will go. The blob where they terminate contributes most to the entropy as the “free end” of the SAW can assume much more conformations. The plateaus of the end-to-end distances correspond to favorable blobs where almost all chains within a certain range of lengths end. If the length is increased, the blobs through which the conformations pass eventually become too small and the end will be pushed to a blob further away. This first results in a decreasing excess entropy, then in a sudden change in the average end-to-end distance. It may also happen that with increasing length a different string of blobs becomes preferable and will dominate the statistics, which is typically accompanied by a rise of S_e .

This behavior is possibly connected to the observed scaling law for the average entropy, Eq. (4.28): the term corresponding to a simple exponential, $N \ln \mu^{(0)}$, characterizes the increase of conformations with length due to the typical number of options for each step, as if the cluster were homogeneous. The stretched exponential term, $\tilde{\mu}^{N^{1-\zeta}}$, may correspond to the increasing chance of finding entropically favorable (strings of) blobs.

5. Variations of the model

This chapter treats various of the basic model of SAWs on CPCs. The results were produced either with variants of the SFE method and/or PERM. The aim here is to get a step closer to realistic polymer models, by including short-range interactions, as well as getting a more general perspective by varying the substrate. Besides, the scope of applications of SFE shall be discussed a bit more. As these are mainly qualitative objectives, I limited the investigations to $D = 2$ and $D = 3$ systems. As far as I can tell from preliminary checks, the principle behavior in higher dimensions seems to be the same.

5.1. Away from criticality

The most obvious thing we can modify is the concentration of accessible sites, p . According to theory [59, 96], any concentration $p > p_c$ should be irrelevant for the asymptotic scaling limit of the end-to-end distance, i.e., $\nu_p = \nu_1$. However, numerical evidence for this assertion is lacking so far. Indeed, the results of the few studies where SAWs on supercritical clusters have been investigated have at best been inconclusive, or — in the case of my own diploma thesis [101] — even contrarian. The subcritical case, $p < p_c$, has received even less attention. Here the clusters, known as *lattice animals* are also fractals but with a different (smaller) fractal dimension, and the behavior of SAWs on them could be quite interesting. The only problem is that *lattice animals* have a typical size. Larger ones are suppressed exponentially, see for instance [16], so that obtaining them by simple sampling can take quite long. Moreover, requiring them to percolate (by wrapping or spanning the system) leads to results that depend on the lattice size as lattice clusters “percolating” larger lattices tend to be very stretched. My investigations of SAWs on percolating clusters below p_c did therefore not yield conclusive results. An interesting project future would be investigating the AC average (see Section 4.1.6) of SAWs on lattice animals. The supercritical case, on the other hand, has the problem that the clusters will no longer be fractals on larger length-scales, so the SFE method must fail eventually. Fortunately, however, this happens at concentrations where PERM is quite efficient (see Sect 3.5), so that the whole region $p_c \leq p \leq 1$ can be covered with relative ease.

5.1.1. Above p_c : SAWs on supercritical clusters

To investigate walks on supercritical clusters, I used a combination of SFE and PERM, depending on which proved better suited for the specific settings; see Section 3.5. Where possible, I cross-checked the PERM results against the exact results for some

5. Variations of the model

clusters to ascertain how many effective tours were needed in order to avoid bias effects. In regimes where only PERM could be used, I took care to start enough tours so that the results seemed no longer to decrease with increasing statistics. As already discussed, the bias problem is most severe in regimes where SFE works well, i.e., in the vicinity of p_c , so that the results obtained with PERM far above p_c can be trusted.

Figures 5.1 and 5.2 show the results for the mean squared end-to-end distances as function of N for various levels of concentration in the range $p_c \leq p \leq 1$ on $2D$ and $3D$ clusters, respectively. For $p_c < p < 1$, each data point was obtained from 10^4 cluster configurations. The values in the plots were scaled by the asymptote of the undiluted system, $N^{2\nu_{\text{full}}}$, which is exactly known in $2D$ and very precisely in $3D$; see Section 2.1.2. For the sake of completeness, the results of my own simulations on the full lattice were also included. It should be noted that I did not go to the limit of what is possible on weakly dilute lattices. Besides, PERM is probably not the best choice of algorithm for p close to one. Significantly larger systems should in principle be accessible.

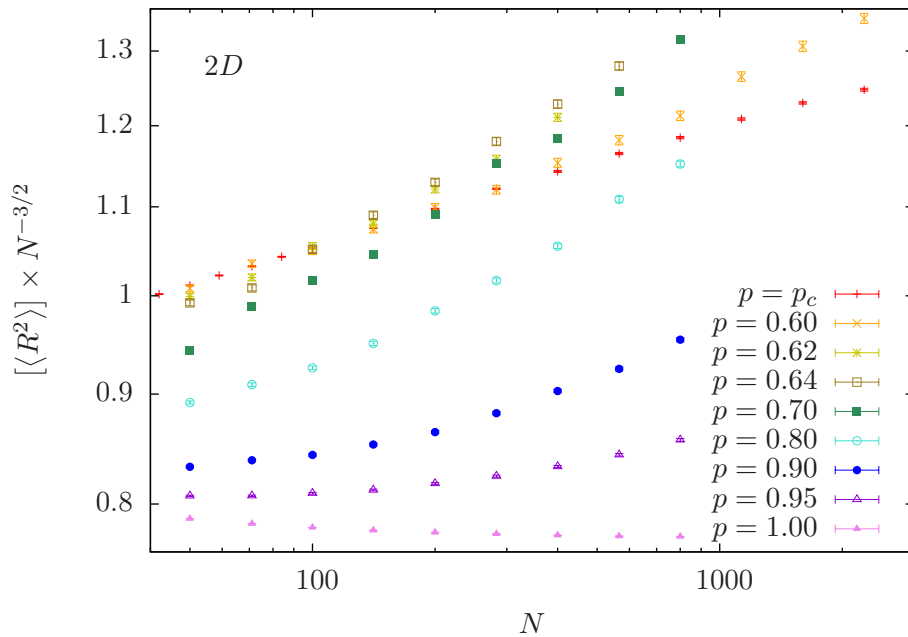


Figure 5.1.: Mean squared end-to-end distance for SAWs on percolation clusters at concentrations $p \geq p_c$ in $2D$. The values are divided by $N^{\nu_1} = N^{3/2}$ to enhance visibility.

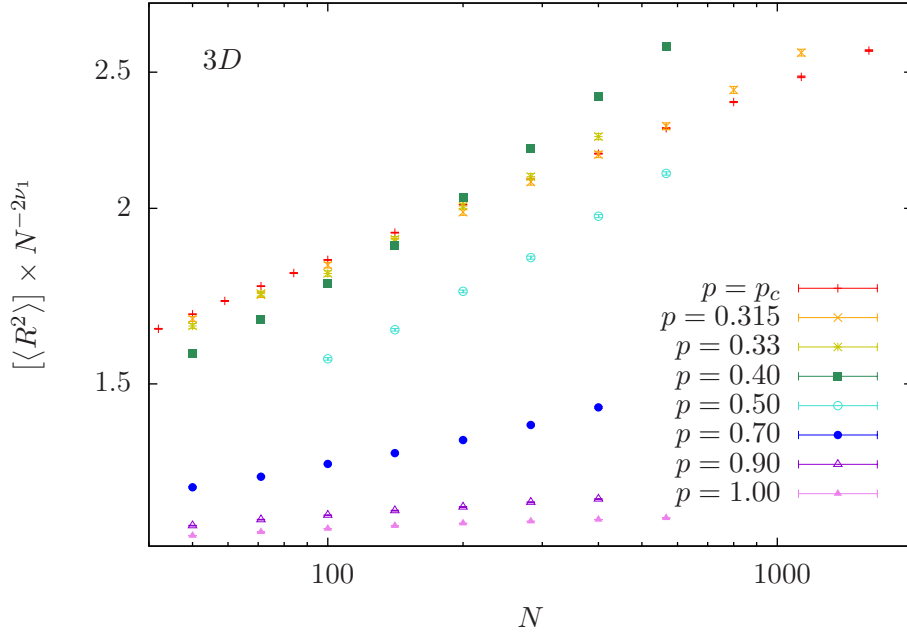


Figure 5.2.: Mean squared end-to-end distance for SAWs on percolation clusters at concentrations $p \geq p_c$ in $3D$. The values are divided by $N^{1.175} \approx N^{2\nu_1}$ for better visibility.

As can be seen, the results are clearly at odds with the theoretical prediction of the Meir-Harris model [59], according to which any concentration $p > p_c$ should give the asymptotic behavior of the undiluted case (Section 2.3.3): When the concentration is raised above p_c the slope increases significantly rather than decreasing towards ν_1 . Likewise, the slope clearly increases with respect to ν_1 when the concentration is lowered from $p = 1$. These increases even seem to become more pronounced with increasing N . Astoundingly, even the very weak dilution of $p = 0.9$ appears to raise the value of ν substantially in $2D$, and the same trend is evident in $3D$. For the $2D$ case, I had already noted this in my diploma thesis [101] though based on much less precise data. These new results appear to confirm my suspicion that the Meir-Harris theory does not correctly describe the model and suggest that this is no $2D$ anomaly. While the possibility that the ranges investigated still do not capture the asymptotic behavior can never be totally ruled out — a lesson learned from the results at criticality — this seems quite unlikely here. There is no indication whatsoever that the slope will diminish eventually, and it would be difficult to explain how it can even initially grow so much larger than at criticality. Based on these results, I therefore conclude

$$\nu_p > \nu_{p_c} > \nu_1 \quad \text{for } p_c < p < 1. \quad (5.1)$$

5. Variations of the model

Table 5.1.: Estimates for the exponent ν on supercritical 2D clusters. Also listed are the result for the critical case from Section 4.1.1 and the exact value for the free SAW.

p	fit range	method	χ^2	ν
0.59274621	800 - 12800	SFE	1.07	0.7749(4)
0.60	400 - 2263	SFE	1.03	0.796(2)
0.61	283 - 1131	SFE	0.874	0.799(2)
0.62	100-566	SFE	1.36	0.800(3)
0.64	71-566	SFE/PERM	0.673	0.807(1)
0.70	141-800	PERM	2.11	0.814(3)
0.80	283-800	PERM	1.45	0.811(3)
0.90	283-800	PERM	1.40	0.787(2)
0.95	283-800	PERM	1.50	0.768(2)
1.00	-	exact	-	0.75

Table 5.2.: Estimates for the exponent ν on supercritical 3D clusters. Also listed are the result for the critical case from Section 4.1.2 and the most accurate literature result for the free SAW.

p	fit range	method	χ^2	ν
0.31160768	12800	SFE	1.03	0.6462(4)
0.315	400-1131	SFE	1.02	0.668(4)
0.32	100-400	SFE/PERM	0.923	0.667(3)
0.33	100-400	SFE/PERM	1.16	0.684(3)
0.35	141-566	PERM	1.19	0.702(3)
0.40	141-566	PERM	0.792	0.705(2)
0.50	141-566	PERM	2.01	0.679(3)
0.70	141-400	PERM	1.87	0.623(2)
0.90	141-400	PERM	1.05	0.5969(4)
1.00	-	MCMC [6]	-	0.587597(7)

Table 5.1 and Table 5.2 show estimates for ν at different values of concentration obtained by least-squares fits. Note that the values do not yet seem to have stabilized but might still rise substantially, in particular those close to p_c and 1. It is therefore quite plausible that universal fixed-point values of $\nu_{2D}^* \approx 0.81$ and $\nu_{3D}^* \approx 0.71$ are valid for any p in the regime $p_c < p < 1$. Both estimates are substantially above the critical values, which are already larger than those for the regular systems.

I have no explanation as to why the exponent on diluted lattices (above p_c) is larger than on regular ones [Eq. (5.1)], but at least it is clear that

$$[\langle R^2 \rangle]_p > \langle R^2 \rangle_1 \quad (5.2)$$

5.1. Away from criticality

must hold for any $p < 1$. This is because the annealed average is exactly equal to the full lattice average [Eq. (2.31)], and in order to get from the annealed to the quenched average, one has to weight every SAW conformation ω with the inverse of the number of conformations on the respective disorder configuration, $Z_{C(\omega)}^{-1}$ (see Section 2.3.2). Hence, configurations with few SAWs contribute more strongly to the quenched average than to the annealed average. On average, these also have more defects around the origin since this is where the density of possible conformations is highest. The SAWs are hence more stretched out as central defects select against compact conformations. However, it is not clear whether this effect is relevant in the asymptotic limit $N \rightarrow \infty$ or whether it is averaged out. Numerical investigation of the correlations between Z and $\langle R^2 \rangle$ might therefore be an interesting project for the future.

End-point distributions and conformation density: For SAWs on critical clusters, the statistics are determined by a few entropically favorable channels, and the conformation density does hardly disperse with increasing length; see Section 4.3. To see whether this behavior is a particularity of the critical system or whether it might also be responsible for the unexpected behavior on supercritical clusters, I studied SAW end-point distributions and conformation densities for a couple of clusters at different concentrations: $p = 0.7, 0.8, 0.9, 1.0$. The results for one example with $N = 400$ steps are shown in Fig. 5.3 and Fig. 5.4 alongside the critical case. To visualize the densities, I used a logarithmic as well as a normal color scale. For $p > p_c$, the results were produced with PERM using fairly high statistics (10^{10} effective tours). The linear plots could probably not be distinguished from exact results. (Note that I did not plot the clusters here.) From these pictures, it seems that the conformation density does clearly disperse above p_c , which is a qualitative distinction to the critical case. However, the differences to the full-lattice case are also remarkable, even at $p = 0.9$: Although the clusters themselves are fairly symmetrical and homogeneous, the behavior of the SAWs is quite anisotropic.

5. Variations of the model

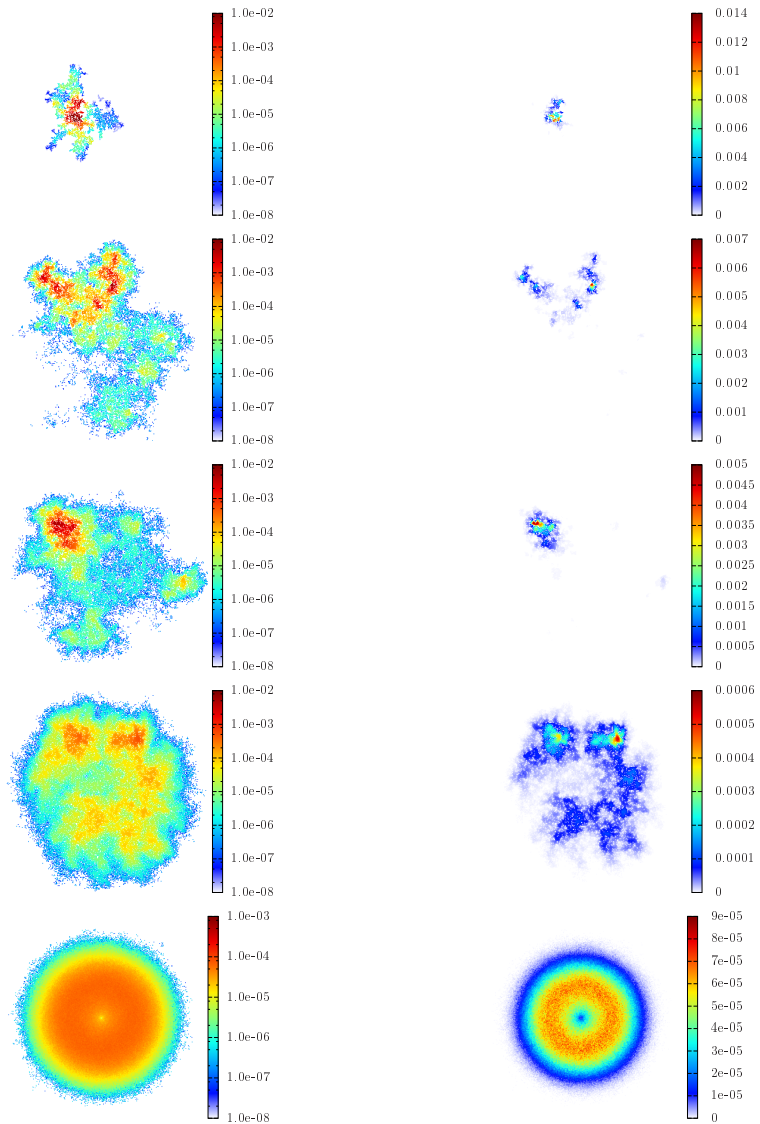


Figure 5.3.: Density distribution of end-point locations for a SAW of $N = 400$ steps on a $2D$ cluster at concentrations $p = p_c, 0.7, 0.8, 0.9, 1.0$, from top to bottom. A logarithmic color scale was used for the left plots, a linear one for those on the right. (Note that the scaling of the color axis varies.)

5.1. Away from criticality

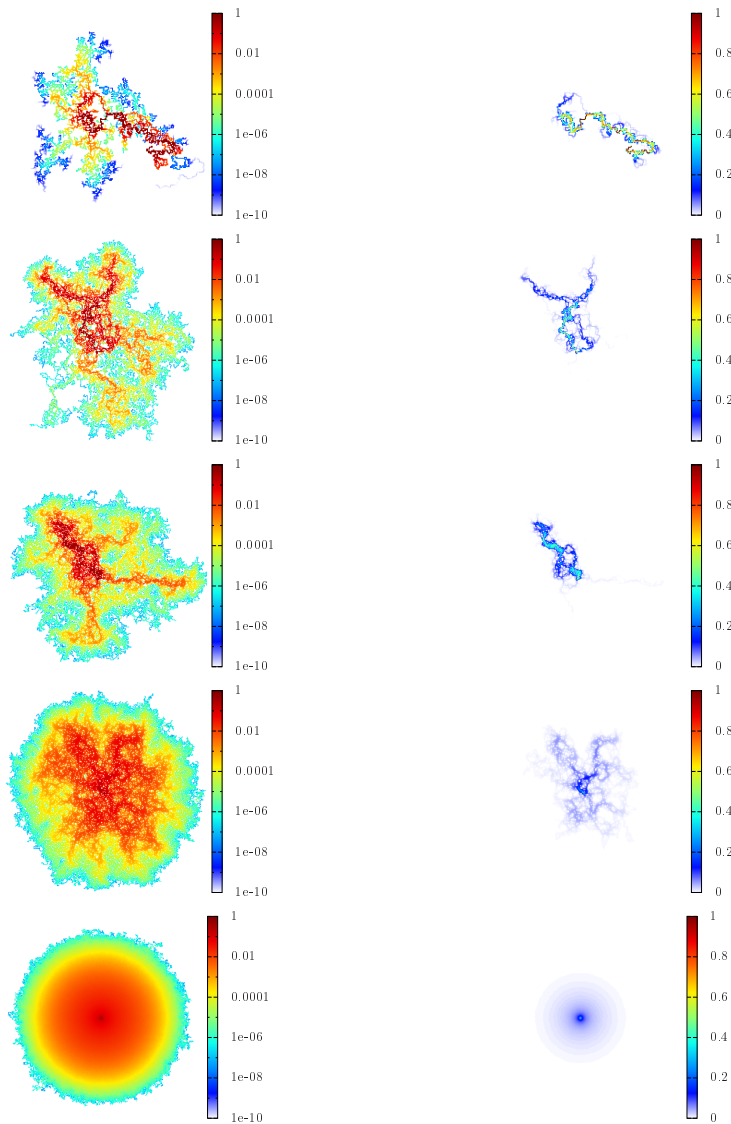


Figure 5.4.: Density of possible conformations through each site for a SAW of $N = 400$ steps on a $2D$ cluster at concentrations $p = p_c, 0.7, 0.8, 0.9, 1.0$, from top to bottom. A logarithmic color scale was used for the left plots, a linear one for those on the right.

5.2. Self-attracting self-avoiding walks

The SAW represents the limiting case for a long polymer in good-solvent condition and at high temperatures. In that regime, short-range monomer-monomer interactions are asymptotically irrelevant, and the polymer is in a stretched, *coiled* state. At lower temperatures (or in poorer solvents), however, these attractions will cause the polymer to assume a collapsed, *globular* conformation. The transition from the *coiled* to the *globular* phase happens at a specific temperature, called the Θ -point of the system. It is of second order and characterized by a set of universal critical exponents, describing the divergence of thermodynamic variables such as the specific heat. Directly at the Θ -point, the scaling of the end-to-end distance is random walk-like, while the space is filled (statistically) homogeneously below Θ [124, 125]:

$$\nu_T = \begin{cases} \nu_{\text{SAW}}, & \text{if } T > \Theta \\ 4/7 \text{ (2D) or } 1/2 & \text{if } T = \Theta \\ 1/D, & \text{if } T < \Theta. \end{cases} \quad (5.3)$$

At very low temperatures, there will also be crystalline phases, which depend on the specifics of the system. This behavior is captured by the simple model of a SAW with nearest-neighbor attraction, the so-called self-attracting self-avoiding walk (SASAW), which was already introduced in Section 3.2.5. Its Hamiltonian is given as a constant ϵ times the number m_ω of pairs of neighboring, non-consecutive monomers (“*contacts*”) of the conformation ω , and the thermal conformational average of an observable O of the system is hence defined as

$$\langle O \rangle_T = \frac{\sum_\omega O_\omega e^{-m_\omega \epsilon / (k_B T)}}{\sum_\omega e^{-m_\omega \epsilon / (k_B T)}}. \quad (5.4)$$

For convenience, I use the convention $\epsilon/k_B := 1$ here. The natural continuation to disordered media such as CPCs is to look at the quenched average:

$$[\langle O \rangle] = \frac{\sum_C \langle O \rangle_T}{\sum_C 1}, \quad (5.5)$$

where the sum is over all disorder conformations (incipient clusters). Naively, one would expect that SASAWs on CPCs should also exhibit a transition from entropy-dominated coils to energy-dominated globules. However, a couple of questions come up immediately:

- Will there be a fixed transition temperature that is the same for every cluster?
- How can the behavior at Θ be random walkish, i.e., $\nu_\Theta = \nu_{RW} < 1/2$ if that is already below the space-filling limit ($1/2 < 1/d_f$ in 2D and $1/2 < 1/d_{\text{BB}}$ in all dimensions)?
- Will $\nu_{T < \Theta}$ be $1/d_f$ or $1/d_{\text{BB}}$? Can a SAW actually be space-filling on a critical cluster?

Table 5.3.: Estimates for the Θ -temperatures and critical scaling exponents ν^Θ of the end-to-end distance for SASAWs on incipient critical percolation clusters in $2D$ and $3D$. They were obtained by Flory-like arguments (FL), exact enumeration (EE) and chain-growth Monte Carlo (MC).

study	Θ_{2D}	Θ_{3D}	ν_{2D,p_c}^Θ	ν_{3D,p_c}^Θ
FL [126]	-	-	0.578	0.564
FL [127]	-	-	0.678	0.611
FL [128]	-	-	0.72	0.62
EE [129]	0.67	-	-	-
EE [130]	0.71(8)	-	0.74(2)	-
EE [131]	(0.71)	0.5(2)	0.73(1)	0.60
EE [74]	0.62(6)	0.43(6)	0.80	0.60
MC [120, 132]	0.92(2)	0.71(2)	-	-

The issue of a Θ transition for SASAWs on CPCs has first been approached via Flory approximations [126–128], and several numerical investigations have subsequently been carried out [74, 120, 129–132]. However, the resulting estimates are rather far-flung, as can be seen in Table 5.3. This is perhaps not surprising given the controversies generated already by the “simple” case of normal SAWs on CPCs.

All these previous results were obtained presupposing the existence of a Θ -transition. The question of the existence itself has not been properly investigated, nor those questions raised above. A substantially more profound analysis would in fact have been difficult with the methods (and hardware) previously available, given that they already struggle with normal SAW on CPCs (see Section 3). As described in Section 3.2.5, the SFE method can, albeit with some effort, accommodate for nearest-neighbor interactions, allowing for a more refined look at the system.

5.2.1. Full temperature range

The more general but less efficient implementation of the SFE method for SASAWs (see Section 3.2.5) keeps track of the complete contact histograms, so that average values of observables at arbitrary temperatures can be calculated exactly via Eq. (3.15). This approach is particularly useful to gain an overview of the temperature-dependent behavior and single out interesting temperature locations for closer investigation.

I enumerated all conformations of SASAWs of length $N = 50, 100, \dots, 800$ in $2D$ and up to $N = 1600$ in $3D$, with each sample consisting of 10^4 clusters. For the largest systems, $N = 800$ in $2D$ and $N = 1600$ in $3D$, the method did not succeed for a small number of clusters, which had to be discarded (3 in both cases). This is unlikely to have biased the findings in a significant way.

Figure 5.5 shows the results for the (quenched) average internal energy per monomer,

$$[u]_T = -[\langle m \rangle]_T / N, \quad (5.6)$$

5. Variations of the model

m being the number of nearest-neighbor contacts. Its magnitude turned out to be surprisingly small: even at $T = 0$ the average number of contacts is only around 0.38 in $2D$ and 0.29 in $3D$. This is less than the average effective coordination numbers of the clusters minus two (which I determined as ≈ 0.52 in $2D$ and ≈ 0.31 in $3D$) showing how little of the area can actually be filled. Another striking aspect is the similarity of the curves; one can hardly discern any significant trend. In particular, they do not seem to become steeper with increasing N . This means there should be no divergence or discontinuity in the average specific heat per monomer, defined as

$$C_V = \frac{[\sigma_E^2]}{NT^2} = \frac{[\langle m^2 \rangle - \langle m \rangle^2]}{NT^2}. \quad (5.7)$$

As can be seen in Fig. 5.6, $C_V(T)$ has a clearly pronounced peak at roughly $T \approx 0.6$ in $2D$ and $T \approx 0.55$ in $3D$. These locations are broadly consistent with the Θ -temperatures reported in previous studies (see Table 5.3). However, with increasing N the height of the curves diminishes slightly instead of diverging. Interestingly, in both $3D$ and $2D$, a shoulder seems to develop at around $T = 0.1$, which might suggest some kind of freezing to the ground state. The lack of divergence of C_V (which was observed but not discussed in Ref. [120]) does not necessarily imply that there is no phase transition, but in any case, the behavior is clearly different from that of free SASAWs, where C_V^Θ diverges with some power of $\ln N$ [102].

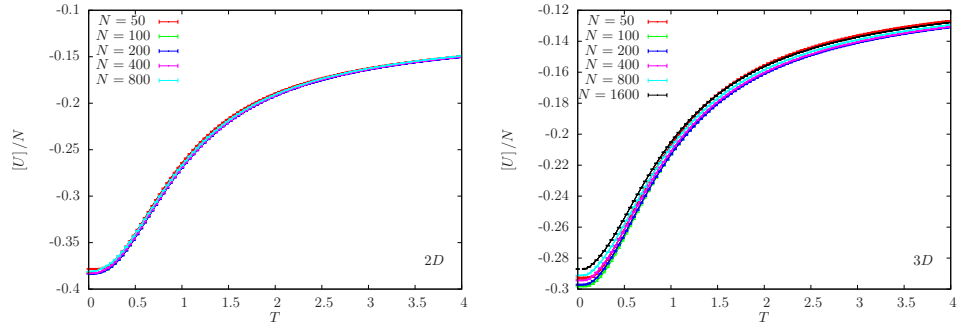


Figure 5.5.: Average internal energy per monomer vs. temperature for SASAWs of varying length on CPCs in $D = 2$ and $D = 3$.

5.2. Self-attracting self-avoiding walks

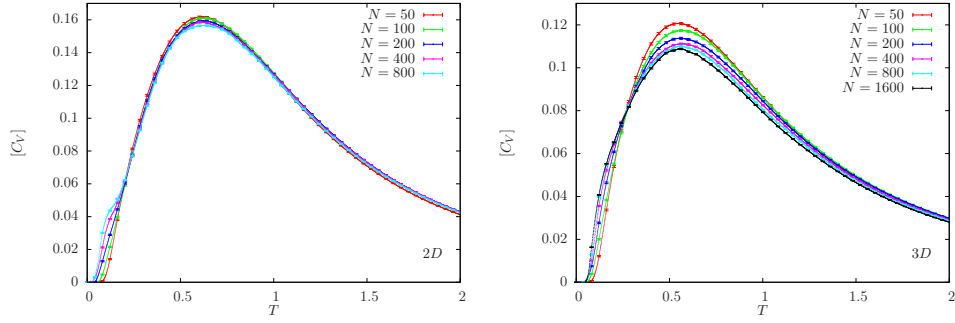


Figure 5.6.: Average specific heat as a function of temperature for SASAWs of varying length on CPCs in $D = 2$ (left) and $D = 3$ (right). The height of the peak diminishes with N .

For the end-to-end distances, the situation is qualitatively similar: Figure 5.7 shows the ratios of $[\langle R^2 \rangle]_T$ and the normal SAW averages (“ $[\langle R^2 \rangle]_\infty$ ”). This plot already indicates that the exponent ν cannot be very sensitive to changes in temperature: If there is a value Θ such that $\nu_T < \nu_\infty$ holds for $T < \Theta$, the ratio would have to vanish with $\sim N^{\nu_T - \nu_\infty}$. Instead, even for T close to zero, the value does not fall below 0.6, it even slightly increases with N . Moreover, there does not seem to be a distinctive temperature at which a drastic change takes place, but the end-to-end distance seems to change smoothly with T . This can be seen more clearly looking at the (scaled) thermal derivative of $[\langle R^2 \rangle]_T$ shown in Fig. 5.8: As for the specific heat, there is a peak around $T \approx 0.6$ / $T \approx 0.5$ but no indication of a divergence with N .

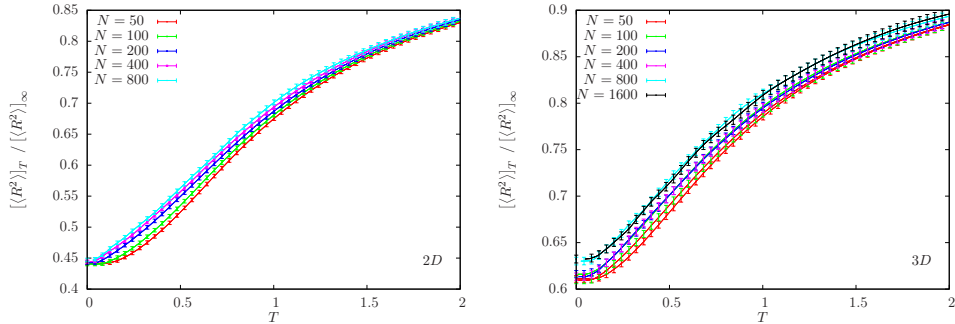


Figure 5.7.: Mean squared end-to-end distances of SASAWs on CPCs divided by the infinite temperature (SAW) limit.

5. Variations of the model

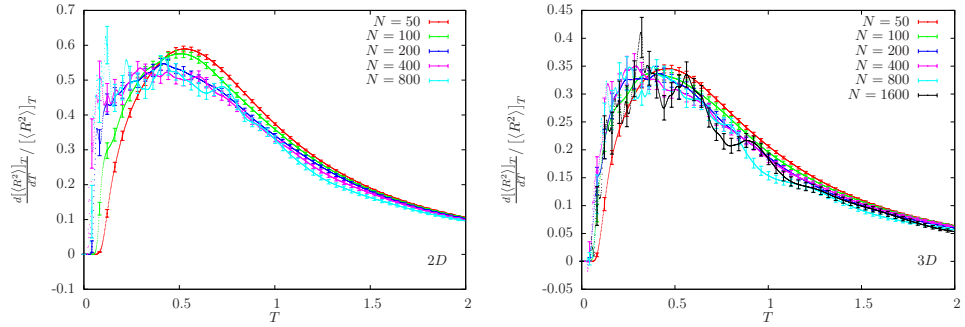


Figure 5.8.: Relative thermal derivatives of the mean squared end-to-end distances for SASAWs on CPCs in 2D (left) and 3D (right).

To understand what is going on, it is again helpful to look at individual clusters. As can be seen in Fig. 5.9, the specific heat displays very sharp spikes. These often correspond to sharp rises in the end-to-end distances but sometimes also to drops; see Fig. 5.10. However, the temperature locations of these “transitions” vary strongly, and on some clusters they are absent altogether. Fig. 5.11 shows the distributions of the temperature locations of where C_V has a global maximum for different chain lengths. Their sharpness does not significantly increase with N , and it is hence clear that the “smeared out” maxima of the quenched averages (Figs. 5.6) cannot diverge. Interestingly, the distributions seem to converge towards a universal shape.

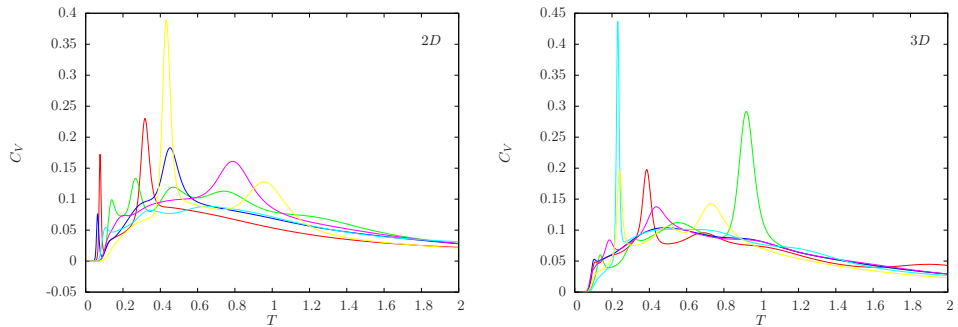


Figure 5.9.: Temperature dependence of the specific heat C_V on a couple of arbitrarily selected CPCs in 2D (left) and 3D (right).

5.2. Self-attracting self-avoiding walks

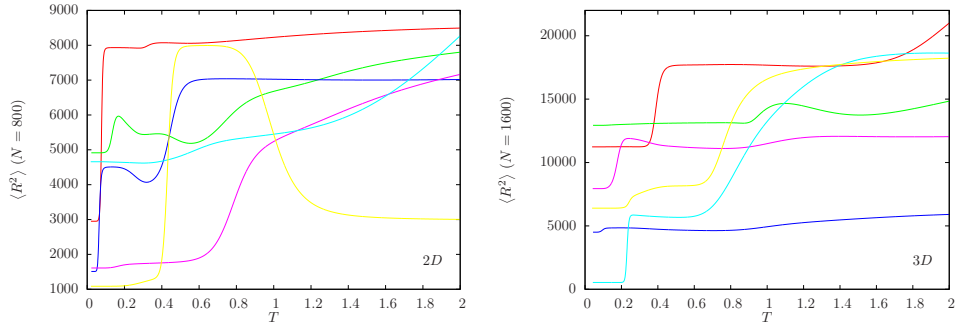


Figure 5.10.: Temperature dependence the end-to-end distance for the same set of clusters as in Fig. 5.9. Note how the jumps often correspond to peaks in the specific heat.

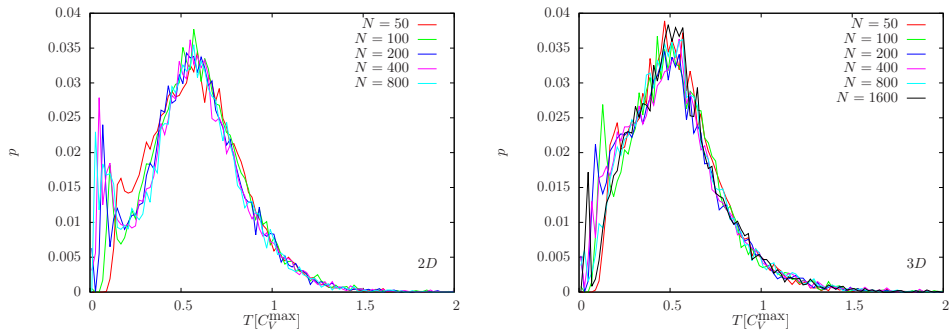


Figure 5.11.: Distribution of temperatures where the specific heat C_V has a global maximum for critical clusters in $D = 2$ (left) and $D = 3$ (right). Colors indicate different maximum lengths of the SASAWs.

5.2.2. Fixed temperatures

At a fixed temperature, the SFE method works much more efficiently as the averaging can be done on the fly; see Section 3.2.5. This approach does not yield as much information about the system, but it suffices as long as we only want to find out whether there is a change in the asymptotic scaling behavior due to temperature. To this end, the quenched averages of the end-to-end distances as a function of the number of steps (up to $N = 6400$) are compared at three different temperatures in $D = 2$ and $D = 3$, respectively:

- “ $T = \infty$ ”, i.e., without interactions (normal SAWs);
- $T = 0.55$, i.e., near (slightly below) the peak of $[C_V]$;

5. Variations of the model

- $T = 0$, i.e., only lowest-energy states contribute.

The results are plotted in Fig. 5.12 on double logarithmic scale. The curves differ by a constant factor, but the slopes are clearly very similar. Indeed, if one rescales the x -axis by a temperature dependent factor $a(T)$, the curves fall almost on top of each other; see Fig. 5.13. Note that the y -axis is rescaled by $\approx N^{2\nu}$ to magnify the differences, which are actually quite small.

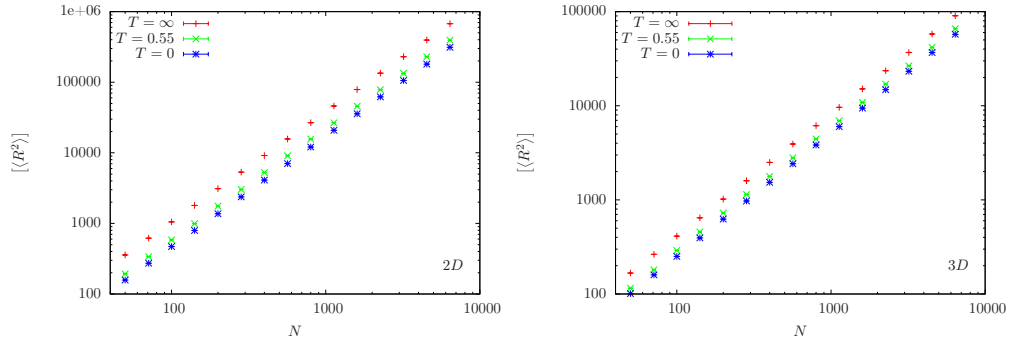


Figure 5.12.: Mean squared end-to-end distances of SASAWs at different temperatures on critical clusters in $2D$ (left) and $3D$ (right). The curves, from bottom to top, correspond to: Ground states, i.e., conformations with maximal number of contacts ($T = 0$); averages slightly the alleged transition temperature ($T = 0.55$); and standard SAWs ($T = \infty$).

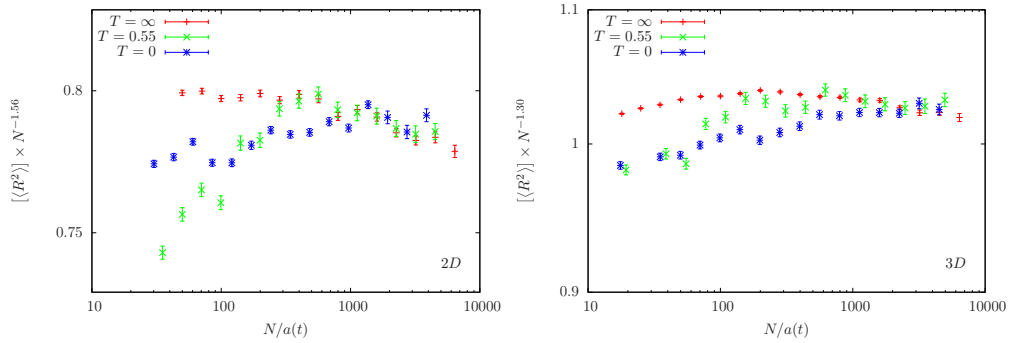


Figure 5.13.: Rescaled mean squared end-to-end distances for SASAWs at different temperatures. By dividing the values on the abscissa by a factor $a(T)$, the results were made to collapse onto one curve (for larger N at least); the scaling of the second axis serves to magnify the details.

5.2. Self-attracting self-avoiding walks

Table 5.4.: Estimates for the exponent ν for SASAWs at different temperatures on CPCs in $2D$ and $3D$.

$2D$				
T	χ^2	A	ν	a
∞	1.4	0.84(1)	0.0.776(9)	1
0.55	0.35	0.488(5)	0.776(6)	1.42
0	1.0	0.351(8)	0.782(2)	1.75
$3D$				
T	χ^2	A	ν	a
∞	0.94	1.082(7)	0.6466(4)	1
0.55	0.54	0.77(2)	0.648(1)	1.30
0	0.62	0.632(6)	0.651(3)	1.52

Estimates for the scaling exponents were again obtained by least-squares fits, and they turned out to agree within the statistical errors; see Table 5.4. This strongly indicates that the temperature has no effect whatsoever on the asymptotic scaling behavior, i.e., that there is no temperature-driven collapse transition. The exponent even resulted slightly larger at lower temperatures, which fits to the observed finite-size behavior with a slightly larger initial slope (Section 4.1) if we assume that the effective number at lower temperatures is smaller ($N \rightarrow N/a(t)$). The scaling factors $a(T)$ are also listed in Table 5.4. They were obtained from the ratios of amplitudes:

$$a(T) = \left(\frac{A(T)}{A(\infty)} \right)^{\frac{1}{2\nu}}. \quad (5.8)$$

To make the results for different temperatures collapse onto a single curve, one might also have rescaled the y -axis, but rescaling the x -axis captures the situation more accurately. This can be seen by looking at the end-to-end distances for SASAWs at different temperatures on single clusters as shown in Fig. 5.14. On top, where the values are not rescaled, the curves appear to have roughly the same average slope, but they look rather different otherwise, especially in $2D$. The rescaled curves shown on the bottom, however, do not only look similar on average, those from the same clusters even display almost the same fluctuation patterns. Hence, the thermal averages as functions of the number of steps N behave very similarly at different temperatures, the only significant difference being that the walks at lower temperatures are “slower” to realize the same ups and downs. The collapsing is less significant in $3D$, where the original curves are very similar already. This is intuitively clear as much fewer contacts can typically be realized in that case, so that their influence is smaller. It would be very interesting to look at the end-point and conformation densities (Section 4.3) for SASAWs in order to visualize how they tend to visit the same cluster areas as athermal SAWs. This is in principle feasible with the SFE method, at least for a couple of hundred steps.

5. Variations of the model

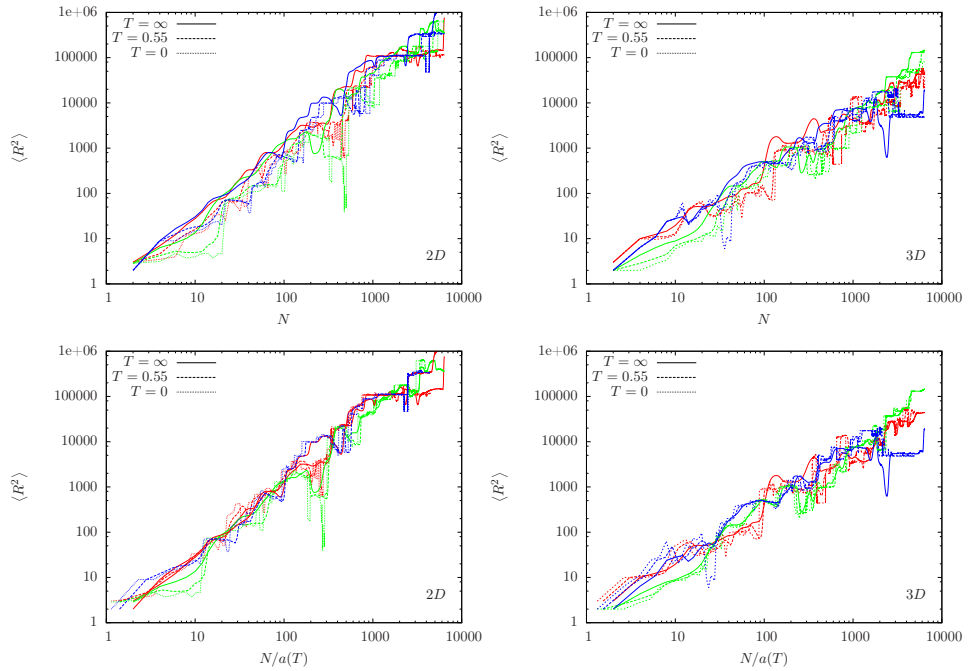


Figure 5.14.: Mean squared end-to-end distances as function of the number of steps for SAWSAWs at different temperatures on individual critical clusters in 2D (left) and 3D (right). The colors represent the clusters while the line types denote the temperature: solid for $T = 0$ (ground states), dashed for $T = 0.55$ and dotted for $T = \infty$ (normal SAWs). The abscissas in the bottom plots where rescaled by a temperature-dependent factor $a(T)$; see Table 5.4.

How can there be no transition?: The apparent complete absence of not only the Θ -point but of any kind of temperature-driven phase transition for SASAWs on CPCs is baffling but not inexplicable. Given the thin, fractal geometry of the critical clusters, it is intuitively clear that a SAW cannot efficiently fill the volume, not even that of the backbone. Even less can we expect the SAW to crystallize into some kind of ordered state in such a disordered medium. Still, with lowering temperature the energy must at some point become relevant and eventually dominant. But unlike for free SASAWs, an energetically favorable conformation is not collapsed around the origin: in order to realize as many contacts as possible, it has to seek out cluster regions that are most densely connected. As it happens, these are also entropically favorable. Hence, unlike for free SASAWs, there is simply no competition between energy and entropy on a global scale. Lowering temperature will therefore cause the chains to fill out the *blobs* (see Section 4.3) more densely, but since they are of a typical size, this only slows the scaling behavior by a constant factor. This is comparable to the behavior of a SASAWs

in a slab, where lowering the temperature would also not affect scaling exponent ν as it cannot fall below one even if the volume is filled out completely.

One might still ask whether there is some kind of collapse transition corresponding to a change from coil-like to globular conformations within the blobs, without effect on the scaling behavior of the end-to-end distance. However, this does not seem to occur at a sharply defined temperature as could be seen from the temperature dependence of the specific heat and the end-to-end distances in the previous subsection.

5.3. Kinetic-growth walks

The kinetic-growth (self-avoiding) walk (KGW) is a dynamic version of the SAW. The relation between these two models is similar to that between the diffusive random walk and the ideal chain [133]: While the standard SAW is conveniently described in terms of equilibrium statistical physics, i.e., using a partition function and averages over all conformations, defining it as a dynamic random process seems rather artificial and requires the use of Rosenbluth weights. The KGW, by contrast, is defined most conveniently as a diffusive random walk where the walker is not allowed to revisit a site. It was originally proposed as a model for polymerization [134], hence the “kinetic”. On CPCs, it can also be imagined as a more evolved version of de Gennes’ “ant in a labyrinth” [135] that uses a pheromone trail to avoid treading the same path twice, or more generally as a random search strategy. The only slight oddity of the KGW it can get trapped if it runs into a dead end, in which case the chain is canceled.

Based on MC results and Flory arguments, the KGW on the full lattice had initially been claimed to be in a different universality class than the standard SAW [134], but this has later been clearly refuted [98–100, 136]. On the honeycomb lattice, the KGW can furthermore be shown to be equivalent to a SASAW at a specific temperature [137]. This is because one can change from the SAW ensemble to the KGW ensemble by weighting conformations with the product of the number of blocked sites in each steps (cf. Section 3.1.1), which can also be written as a product of Boltzmann factors from a nearest-neighbor attraction. This equivalence does not hold for other lattice types [138], where one would need a more complicated multi-body interaction. Still, on regular lattices, the KGW can be regarded as an SAW with nearest neighbor interactions and hence as closely related to the SASAW.

KGWs on CPCs have also been investigated in the past, and various estimates for the scaling exponent ν have been reported: $\nu_{2D} \approx 0.68$ (from RSRG [90]), $\nu_{2D} = 0.760(5)$, $\nu_{3D} = 0.645(5)$ (from MC [73]), $\nu_{2D} = 0.725(7) / 0.768$ (from MC/ Flory arguments [139]), and Eq. (2.38) [81]. So far, it has not been clearly established whether ν^{KGW} compared to ν^{SAW} is smaller (as found in Ref. [73]), or larger (as claimed in Ref. [139]); or whether they are equal as for the free case.

Enumerating KGWs with the SFE approach would be difficult, though not impossible. It can be formulated as a SASAW with a multi-body nearest-neighbor interaction, but the problem is that in order to determine the weight for a KGW *segment* (see Section 3.2.5) one needs to know its direction, which is not accounted for in my implementation of the SFE method. At the same time, the KGW can be simulated

5. Variations of the model

with chain-growth (PERM) more easily and more efficiently than the equilibrium SAW (see Section 3.1.1 and Fig. 3.24), so the effort to adapt the SFE method did not seem worthwhile.

Using the KGW version of PERM with the “trap avoidance” improvement, I simulated walks of up to 1600 steps on $2D$ clusters and up to 800 steps on $3D$ clusters. As before, the length was increased by factors of $\sqrt{2}$. I used samples of at least 10^4 randomly generated clusters for each individual length. The *effective number of tours* on each cluster was chosen between $\tilde{t} = 2^{13}$ (for $N \leq 400$) and $\tilde{t} = 2^{16}$ (for $N \geq 1131$). I carefully verified that this is sufficient to avoid any significant bias effects. Figure 5.15 shows the measured mean squared end-to-end distances together with the results for normal SAWs. As can be seen, there is no significant difference in the slopes. Indeed, my best estimates for the exponents ν^{KGW} from least-squares fits, obtained using the ranges $N \geq 200$ for $2D$ and $N \geq 141$ for $3D$, are fully compatible with those for the standard SAW:

$$\nu_{2D}^{\text{KGW}} = 0.778(2) \quad \nu_{3D}^{\text{KGW}} = 0.644(2). \quad (5.9)$$

The equality of the exponents for the SAW and the KGW may not be too surprising considering that the KGW can be regarded as a SAW with a multi-body nearest-neighbor attraction and given that the same asymptotic behavior was also found for the SASAW regardless of temperature. Note, however, that the effective attraction for the KGW applies also to defect sites, which is a qualitative difference to the SASAW. While the finding $\nu^{\text{KGW}} = \nu^{\text{SAW}}$ is in line with the behavior observed for SASAWs, it is therefore not a trivial consequence.

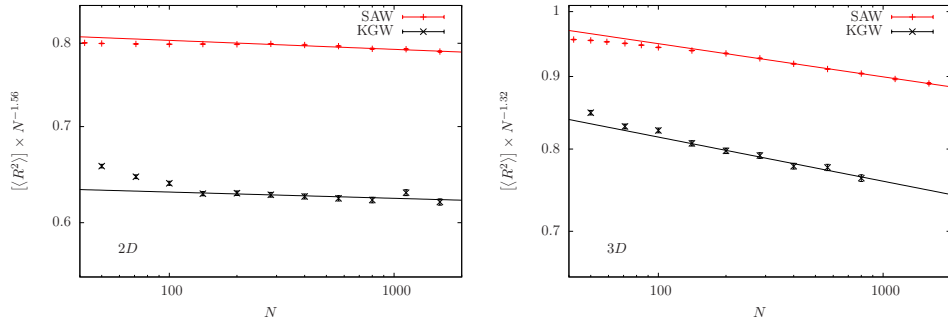


Figure 5.15.: Scaled mean squared end-to-end distance as function of the number of steps for KGWs (black data points) on $2D$ and $3D$ CPCs shown together with the results for standard SAWs from Section 4.1 (red points). The lines represent least squares fits.

5.4. Further ideas

Finally, I want to mention a couple of related systems which could also be investigated by way of SFE, in the hope to inspire future research.

5.4.1. Modifying the walks

Adhesive SAWs: In the previous section, the defects did not affect the walks other than by acting as a steric constraint. However, the method can just as well also include a short-range attraction between walks and defects. The strength of this interaction could in principle be chosen different than between two monomers, and exploring that phase space might be interesting. My interest in these cases is several-fold: For one, it is reasonable that if a polymer interacts with itself via some kind of short-range, van-der-Waals force it should also adhere to surrounding structures. A second motivation comes from the main conclusion of the previous section, namely that there is no thermal phase transition for SASAWs on CPCs because energy and entropy are not competing. This realization gives rise to the question if and how this union can be divided. Can the Θ -transition from free SASAWs be recovered through additional attraction to defects? After all, the interpretation had been that dense cluster regions dominate both entropy and energy by allowing for most conformations and for most contacts at the same time. This latter aspect might now change as contacts with defects can easily be established over the whole length of the chain. On the other hand, the KGW is effectively also attracted to defects and did not show a different behavior either. Maybe there are more profound reasons for such universal behavior still to be discovered.

SAWs under stretching force: A stretching force can very easily be realized from the spatial distribution of the end points by assigning energies to conformations as a function of the distance to the origin. This was done in Ref [113] using EE and in Refs. [120, 132] using PERM. These studies also included nearest neighbor attractions to model force induced unfolding. Keeping track of energy and position histograms will be demanding using the SFE, but one could probably still go beyond the lengths previously accessible with PERM, and the exactness of the results would give unrestricted access to the whole phase space.

SAWs with bending energy: As mentioned in Section 3.2.5, bending energy (stiffness) could be implemented very easily without causing much overhead. “Stiff SAWs” on CPCs would therefore surely be worth a look, even though it is not quite clear (to me) what questions to put to this model. Perhaps one could tie up to studies of SAWs with bending energy on regular fractals [140, 141], or perhaps the model will come up with something interesting of its own. Of course, stiffness might also be combined with attractions and/or stretching forces.

5. Variations of the model

5.4.2. Changing the medium

Close relatives: As already mentioned, subcritical clusters (*lattice animals*) in the AC ensemble still deserve some more attention. This would be a very easy task for the standard SFE method. Without much new effort, one could also use other lattice types or bond percolation. Such quasi-equivalent systems do not promise exciting new findings, but studying them could be valuable to verify and consolidate existing results. In particular, it would help separating universal aspects from unwanted finite-size effects.

Other finitely-ramified stochastic fractals: Studying SAWs in media similar to “standard” CPCs but from different universality classes could bring substantial qualitative understanding, e.g., concerning the roles of randomness and fractal structure of the medium. Easy targets would be *invasion percolation clusters* or clusters from *diffusion-limited aggregation* (DLA). More challenging would be systems with correlated defects such as critical Ising or Potts clusters or, more generally, instances the *random cluster model*. The bulky nature correlated clusters might be a problem for the SFE method, but it could presumably cope to some extent. Besides, there is also the option to use some PERM-like incomplete enumeration for very massive cells.

Complex fractal networks: As described in Refs. [142, 143], complex networks such as the WWW often have self-similar, fractal structures, allowing investigations by renormalization techniques. It is quite likely that the SFE method could hence also be adapted to some of these systems. However, I do not yet know how SAWs on complex fractal networks can be motivated. Perhaps, though, one could focus on the key principles of the method (see beginning of Section 3.2) which could be carried over to other discrete, decomposable problems such as particle transport or condensation.

6. Conclusion

This thesis presented a numerical treatment of self-avoiding walks on dilute lattices. I had carried out large-scale computer simulations to assess the asymptotic scaling behavior of SAWs on percolation clusters and their backbones with special focus was on the case of critical site concentration. This setting is particularly interesting because of the fractal and highly inhomogeneous nature of the system at that point. I developed a “scale-free” exact enumeration method (SFE) that makes use of this fractal structure and is much better-suited to investigate this setting than other numerical tools. Alongside this novel method, I employed a chain-growth Monte Carlo algorithm (PERM) to simulate SAWs on lattices with higher levels of site concentration. Both methods were tested and analyzed in detail. I also studied kinetic growth walks and self-attracting self-avoiding walks. For these latter, I closely looked at the influence of temperature on the conformational statistics.

Scale-free enumeration: The thinly connected, self-similar structure of critical percolation clusters poses a challenge for numerical methods: Markov-chain Monte Carlo is inefficient as global moves are almost always rejected, chain-growth methods often fail to find the way into relevant cluster regions, and all the exact enumeration tricks for regular systems cannot be played in presence of disorder. Fortunately, however, the structural properties of critical clusters can be exploited to factorize the problem, and a major part of this work (Chapter 3) was concerned with the development and testing of a novel exact enumeration method that does this. It operates on critical clusters that are partitioned into a tree hierarchy of nested cells on all length scales. SAW segments are independently enumerated within (small) cells, which are then decimated: they are treated essentially as point-like when conformations through larger cells are enumerated. By dint of this numerical renormalization scheme with irregular cells, it is possible to circumvent the problem of exponential complexity that is usually inherent to exact enumeration methods. I numerically demonstrated that the dependence of the runtime on the number of SAW steps is described by a power-law with an exponent around 2.4 for all dimensions considered instead of the exponential complexity of standard enumeration methods. The reduction translates to a massive increase of the accessible system sizes, from about 50 to over 10^4 steps. Since the runtime for the standard enumeration method increases with the number of conformations, it would need over 10^{1500} ages of the universe for this. I furthermore showed that even PERM, which is probably the best Monte Carlo method to treat the problem at hand, cannot compete in the vicinity of the percolation threshold. It is efficient for SAWs of a few hundred steps, but the results for longer chains turned out to be unreliable. Fortunately, PERM performs much better at higher concentrations, which is where

6. Conclusion

the SFE method fails. The two approaches hence complement each other perfectly.

To be efficient, however, the SFE method requires a hierarchical decomposition of the cluster that reflects its self-similar structure, and the algorithm used to setup this cell hierarchy constitutes a significant part of the program. Indeed, developing this part was less straightforward than the actual enumeration since it was not a priori clear how a hierarchy should ideally be organized. However, using heuristic guidelines and some amount of empirical tinkering, I succeeded in devising an “amalgamation algorithm” that does the job near perfectly.

Scaling behavior of SAWs on critical percolation clusters: By exactly enumerating SAWs of several thousand steps (up to 12800 in $2D$ and $3D$) on large samples of critical percolation clusters, I could assess their asymptotic scaling behavior in unprecedented detail. The exponent ν describing the increase of the end-to-end distance with chain length resulted slightly but significantly smaller than previously reported in the literature for both $2D$ and $3D$. Accurate values for higher dimensions were also obtained, though these measurements were impaired by the fact that generating the large clusters became very costly. The scaling exponents on cluster backbones turned out to be very similar to those on full clusters, though the values are not perfectly consistent in all dimensions. A clear relationship to the fractal exponents of the clusters could unfortunately not be established, and though some predictions from field theory and Flory approximations are not far from the numerical estimates, the agreement is not overly impressive.

The number of SAW conformations fluctuates very strongly for different clusters, and these large deviation impeded direct investigation of the quenched averages. However, since the distribution is very close to log-normal in all dimensions on full incipient clusters and backbones, the value can be approximately calculated from the average entropy. Instead of the expected power-law factor, the scaling law for the number of conformations on incipient CPCs (and backbones) seems to feature a stretched exponential. This finding is very surprising as power-law “enhancement” is correct for the “all cluster average” (AC), i.e., when all clusters that can support at least one SAW conformation of the respective length are included in the statistics. The growth of the exponential of the average entropy was also found to involve a stretched-exponential correction factor. This is not as surprising but nonetheless unexpected: previous studies had hinted that the scaling behavior of this quantity might be unusual, but the particular form observed was not among the many suggestions.

Future attempts to explain the behavior of SAWs on CPCs might be aided by the qualitative observations presented for distributions of end-points an *conformation densities*. These are strongly concentrated in entropically favorable cluster regions and seem and not to disperse with increasing chain length. The regions where the majority of conformations end and the channels by which they get there tend to remain unchanged over wide ranges of lengths before switching very abruptly. It may well be that the scaling behavior of the average end-to-end distances and numbers of conformations reflect the distribution of entropically favorable cluster regions and how these are explored by the SAWs.

SAWs on supercritical clusters: To investigate SAWs on clusters above the percolation threshold I employed either SFE or PERM, depending on concentration and chain length. For this case, the well-established Meir-Harris model predicts the same asymptotic scaling behavior of the end-to-end distance as on regular lattices. The results of my diploma thesis for the $2D$ system had already put this into question to some extent, and the more profound analysis presented here confirmed those findings: In both $2D$ and $3D$, the scaling exponent ν appears to be significantly larger above the percolation threshold than directly at criticality. Indeed, the measured slopes in the log-log plots increase with growing chain length rather than crossing over to the (smaller) regular-lattice value. Intriguingly, the values of the slopes for different concentrations show a tendency of convergence, hinting at the existence of a stable fixed point for supercritical systems. This issue clearly deserves to be looked at more closely in the future.

Self-attracting SAWs and KGWs on CPCs: Nearest-neighbor attractions turned out to have no influence on the asymptotic behavior of SAWs on CPCs. The average specific heat was found to have at cusp a certain temperature, but it does not diverge. Moreover, the swelling of the end-to-end distance is always described by the same exponent as for SAWs without attraction — even at zero temperature. This behavior is different from that of SASAWs on the full lattice, which collapse and eventually freeze when the temperature is lowered. The complete absence of a transition is highly surprising and goes against all expectations and claims in the literature. However, in light of the observations concerning the distributions of end-points and conformation densities described at the end of Chapter 4, it can be understood qualitatively: The entropy on a cluster is typically dominated by one or two channels, while the remaining area does not contribute significantly. These entropically favorable channels are concatenations of cluster regions with above-average connectivity, where the growth of conformational possibilities is strongest. Incidentally, this means that they allow for most nearest-neighbor contacts, so that they are also energetically favorable at lower temperatures. Hence, unlike on full clusters, there is no competition between energy and entropy that would warrant a phase transition.

The average end-to-end distance of kinetic growth walks on CPCs appears to be characterized by the exponent for normal SAWs as well. This is possibly due to the same effect I proposed for SASAWs since KGWs can be interpreted as equilibrium SAWs with a multi-body nearest-neighbor attraction. This would mean that field-theoretical analyses of KGWs [81, 97] might be used to understand the behavior of equilibrium SAWs.

Outlook: While my work answered some questions, others have remained open and more have emerged. Most important, in my opinion, are the following two: whether (and if so, why) the quenched average of the number of conformations on incipient critical clusters is truly governed by such an unusual scaling law, and whether there is a universal fixed-point for the swelling exponent for SAWs on supercritical clusters. To further look into the first issue numerically, one would need to overcome the

6. Conclusion

problem of large deviations. This might be realized via importance sampling of the clusters as sketched in Section 4.2. The second one is easier since one should be able to simulate significantly longer SAWs on weakly diluted lattices with PERM, possibly in conjunction with multicanonical Markov-chain MC methods. Perhaps more urgent than further numerical work, however, is the need for a new theory able to explain my findings. For instance, it would be desirable to properly understand why there is no phase transition for SASAWs (including KGWs), and how the scaling of the average entropy relates to the large deviations in the distribution of the number of conformations. A possible starting point here might be to investigate the spatial fluctuations of density and connectivity on critical clusters, which is responsible for the non-dispersion of the conformation density.

As for the SFE method, there is still much scope for further application, as described at the end of Chapter 5. There is also possibility to use the underlying strategy of numerical “adaptive-cell” renormalization to problems that are also embedded in a disordered fractal medium but completely different from SAWs.

Appendix A.

Generating and analyzing critical percolation clusters

A.1. Setup

I exclusively used site percolation clusters on hypercubic lattices with periodic boundary conditions. As criterion for percolation, I required that a cluster must wrap around the lattice in at least one dimension. More precisely, a cluster percolates if a self-avoiding polygon fits on it that crosses at least one boundary an odd number of times. I adjusted the sizes lattice L to avoid that SAWs could “feel” the boundaries, while trying not to waste too much time and memory (in higher dimensions). The precise values that I used are listed in Table A.1 below.

Table A.1.: Lattice extensions L used for different dimensions and numbers of steps.

N	$2D$	$3D$	$4D$	$5D$	$6D$	$7D$
50	100	52	33	25	20	14
71	142	67	41	31	24	17
100	200	86	51	37	28	20
141	282	111	64	45	34	24
200	400	143	80	54	41	29
283	566	184	100	66	48	35
400	800	237	125	79	58	41
566	1132	305	156	96	69	49
800	1600	392	195	117	83	59
1131	2262	504	243	141	99	70
1600	3200	649	303	170	119	84
2263	4526	836	379	206	142	100
3200	6400	1076	473	252	170	
4525	9050	1384	590	301		
6400	12800	1782	736			
9051	18102	2293	920			
12800	25600	2951				

A.1.1. Creating the clusters

I generated the critical clusters using a depth-first implementation of a burning algorithm known as the Leath method [144]: Each lattice site can either be occupied (1), empty (-1), or unchecked (0). At the beginning, all sites are 0 except for the starting site (“seed”), which is always 1. A random number is then drawn for each unchecked neighbor in turn. If the number is larger than the occupation probability p , the neighbor is set to -1; else it set to 1 and the routine recursively calls itself from there. The program finishes once there are no unchecked neighbors left, i.e., when all neighbors of the cluster are -1. To check for percolation, I assign a number to each occupied site storing which boundaries have been crossed to reach it. Each dimension is represented by one bit which is flipped at each boundary crossing. Whenever a neighbor is found occupied, I compare its number to that of the current site (with one bit flipped if a boundary is between the two site). If the two numbers differ, percolation has occurred. This approach is faster than randomly occupying the sites on the whole lattice and then identifying the clusters using the Hoshen-Kopelman (HK) algorithm [145]. I should note that the two methods are not exactly identical: the burning method samples more massive clusters with increased probability (which is sensible in my view). However, this does not result in any measurable difference in practice.

The easiest and fastest implementation of the burning algorithm uses a D -dimensional array to store the status of the sites, which I did for $D = 2, 3$. However, the required memory for a lattice of extension L increases with L^D , which a problem in higher dimensions. Instead, one can use a hash table where only the visited sites (“1”) are stored, and the memory needed to store a (percolating) cluster hence goes $O(L^{d_f})$.

A.2. Estimating the fractal dimensions d_f and d_{\min}

I measured the average masses $[M]$ for percolating clusters on lattices of varying size L , see Fig. A.1, and estimated the fractal dimension d_f by fitting

$$M = aL^{d_f}. \tag{A.1}$$

The results are given in Table A.2. In $D = 2, 3$, they are fully consistent with the literature. In higher dimensions, finite-corrections were found to be more significant, which can be seen when the y -axis is rescaled so that the curves become flat. To adjusted the lower cutoff for the fits to get a χ^2 close to one, but I also used optical judgment to decide when the slopes had stabilized. The results for $4D$ and $5D$ are more precise than previous numerical estimates although the fit errors may be slightly optimistic as the approach seems to overestimate d_f for small systems. For $D > 6$, periodic boundary conditions are known to raise the fractal dimension of the incipient cluster to $d_f = 2D/3$ [146–148], which is consistent with my findings.

A.2. Estimating the fractal dimensions d_f and d_{\min}

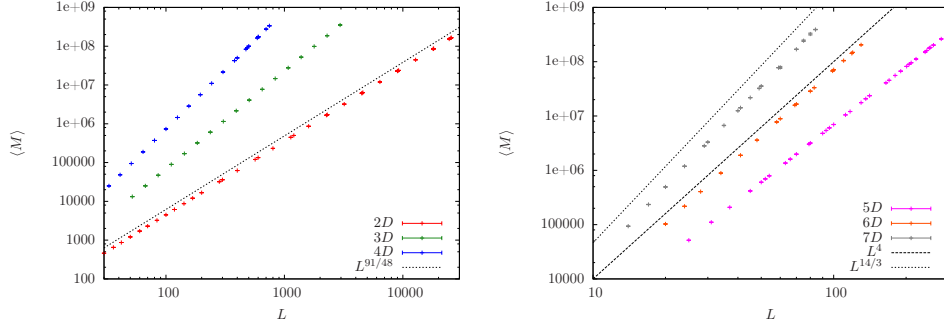


Figure A.1.: Masses vs extensions L and $\langle R_0 \rangle$ for CPCs in different dimensions. Slopes are a measure of d_f ; solid lines represent exactly known values.

Table A.2.: Estimates for the fractal dimension d_f for CPCs in different Euclidean dimensions.

D	fit range (L)	χ^2	d_f	(previous studies)
2	200 - 25800	0.67	1.89584(2)	$91/48 = 1.8958\bar{3}$ [23]
3	111 - 2951	0.41	2.52294(10)	2.52295(12) [26]
4	100 - 747	1.49	3.044(2)	3.05(5) [149]
5	130 - 280	1.49	3.517(7)	3.69(2) [150]
6	100 - 130	0.45	4.02(3)	4
7	30 - 84	1.08	4.65(1)	4 ($14/3 = 4.\bar{6}$ [146–148])

To estimate the shortest-path dimension d_{\min} , I measured the average chemical distance $\langle l \rangle$ between cluster sites as a function of L as can be seen in Fig. A.2. The results of fitting

$$[\langle l \rangle] = aL^{d_{\min}} \quad (\text{A.2})$$

are given in Table A.3 ¹ Here the $2D$ estimate is slightly but significantly different than the one reported previously [25]. That value is probably more reliable, since the analysis carried out in that study was more thorough than mine and since I found the slope ($\frac{d \ln[\langle l \rangle]}{d \ln L}$) to converge relatively slowly in $2D$ (systematically decreasing with L). By contrast, estimates for $D = 3-5$ are perfectly consistent with those from previous studies. The $6D$ estimate for d_{\min} is slightly above the theoretical value, while the $7D$ estimate is again in line with the prediction for systems with periodic boundary conditions ($D/3$).

¹I also measured the average cluster mass within a chemical distance of N steps to estimate d_l ($=d_f/d_{\min}$), which yielded similar but less accurate results for d_{\min} .

Appendix A. Generating and analyzing critical percolation clusters

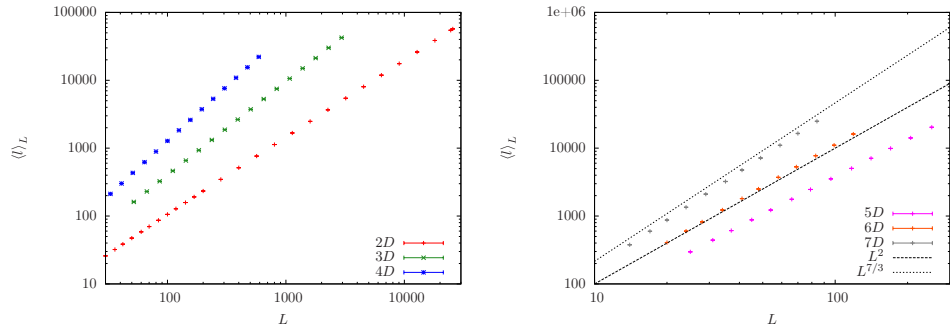


Figure A.2.: Average chemical distance on CPCs vs. lattice extension. The dashed and dotted lines correspond to the theoretical predictions for $D = 6$ and $D = 7$, respectively

Table A.3.: Estimates for the shortest-path dimension d_f for CPCs in different Euclidean dimensions.

D	fit range (L)	χ^2	d_{\min}	(previous studies)
2	1131-25800	0.82	1.1312(2)	1.13077(2) [25]
3	200-2951	1.72	1.3760(3)	1.3756(3) [26]
4	195-590	1.08	1.604(3)	1.607(5) [151]
5	79-252	0.901	1.818(3)	1.812(6) [151]
6	48-119	1.15	2.057(4)	2
7	29-84	1.31	2.331(9)	2 ($7/3 = 2.\bar{3}$)

A.3. Extracting the backbone and measuring d_{BB}

I defined the backbone as the largest bi-connected component wrapping around the lattice. To identify the bi-connected components, I used an adaptation of Tarjan's bridge-finding algorithm [152], which identifies all cutting edges of a graph in linear time. The average backbone masses, $[M_{\text{BB}}]$, are shown in Fig. A.3. The backbone dimension was then obtained by fitting

$$[M_{\text{BB}}] = aL^{d_{\text{BB}}}; \quad (\text{A.3})$$

the results are listed in Table A.4. For $D = 2-4$, the estimates are very accurate and in agreement with previous findings. For $D \geq 5$, however, the results are clearly too large given that the upper critical value is expected to be $d_{\text{BB}} = 2$. This discrepancy is probably not due to finite-size effects, which appeared to be rather insignificant. I therefore suspect that my definition of the backbone leads to a larger fractal dimension for $D \geq 5$ than the original one as the set of sites connecting to distant seeds [27]. However, since

A.3. Extracting the backbone and measuring d_{BB}

the difference must be due to the boundary conditions it should not substantially affect the scaling behavior of SAWs.

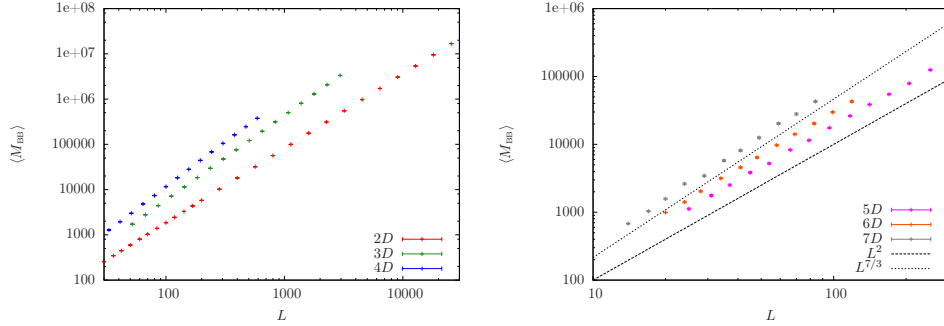


Figure A.3.: Backbone mass vs. lattice size for percolating clusters in dimensions 2-4 (left) and 5-7 (right). The dashed and dotted lines correspond to the theoretical predictions for $D = 6$ and $D = 7$, respectively.

Table A.4.: Estimates for the backbone dimension d_{BB} in different Euclidean dimensions.

D	fit range (L)	χ^2	d_{BB}	(previous studies)
2	200-25600	1.00	1.6433(2)	1.64336(10) [19]
3	111-2951	0.481	1.8736(5)	1.87(3) [38]
4	200-590	0.711	1.932(8)	1.9(2) [28]
5	37-252	0.969	2.03(2)	1.93(16) [28]
6	20-119	0.668	2.13(1)	2
7	14-84	2.41	2.32(2)	2 ($7/3 = 2.\bar{3}$)

Bibliography

- [1] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1976).
- [2] C. Vanderzande, *Lattice Models in Polymer Physics* (Cambridge University Press, Cambridge, 1998).
- [3] M. Rubinstein and R. Colby, *Polymer Physics* (Oxford University Press, Oxford, 2003).
- [4] N. Madras and G. Slade, *The Self-Avoiding Walk* (Birkhäuser, Boston, 1993).
- [5] R. D. Schram, G. T. Barkema, and R. H. Bisseling, *J. Stat. Mech.* **6**, P06019 (2011).
- [6] N. Clisby, *Phys. Rev. Lett.* **104**, 055702 (2010).
- [7] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969).
- [8] P. de Gennes, *Physics Letters A* **38**, 339 (1972).
- [9] H. E. Stanley, *Phys. Rev.* **176**, 718 (1968).
- [10] J. des Cloizeaux and G. Jannink, *Polymers in Solution* (Claderon, Oxford, 1990).
- [11] L. Schäfer, *Universal Properties of Polymer Solutions as Explained by the Renormalization Group* (Springer, Berlin, 1999).
- [12] S. Redner and P. J. Reynolds, *J. Phys. A: Math. Gen.* **14**, 2679 (1981).
- [13] H. Saleur and B. Derrida, *J. Stat. Phys.* **44**, 225 (1986).
- [14] B. K. Chakrabarti and S. S. Manna, *J. Phys. A: Math. Gen.* **16**, L113 (1983).
- [15] H. Kesten, *Percolation Theory for Mathematicians* (Birkhäuser, Basel, 1984).
- [16] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1992).
- [17] D. Ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, 2000).
- [18] M. E. J. Newman and R. M. Ziff, *Phys. Rev. E* **64**, 016706 (2001).

Bibliography

- [19] X. Xu, J. Wang, Z. Zhou, T. M. Garoni, and Y. Deng, Phys. Rev. E **89**, 012120 (2014).
- [20] P. Grassberger, Phys. Rev. E **67**, 036101 (2003).
- [21] B. Mandelbrot, *Les Objets Fractals: Forme, Hasard et Dimension* (Flammarion, Paris, 1975).
- [22] M. P. M. den Nijs, J. Phys. A: Math. Gen. **12**, 1857 (1979).
- [23] B. Nienhuis, J. Stat. Phys. **34**, 731 (1984).
- [24] S. Smirnov and W. Werner, Math. Res. Lett. **8**, 729 (2001).
- [25] Z. Zhou, J. Yang, Y. Deng, and R. M. Ziff, Phys. Rev. E **86**, 061101 (2012).
- [26] J. Wang, Z. Zhou, W. Zhang, T. M. Garoni, and Y. Deng, Phys. Rev. E **87**, 052107 (2013).
- [27] G. Shlifer, W. Klein, P. J. Reynolds, and H. E. Stanley, J. Phys. A: Math. Gen. **12**, L169 (1979).
- [28] D. C. Hong and H. E. Stanley, J. Phys. A: Math. Gen. **16**, L475 (1983).
- [29] M. F. Gyure, M. V. Ferer, B. F. Edwards, and G. Huber, Phys. Rev. E **51**, 2632 (1995).
- [30] M. Porto, A. Bunde, and S. Havlin, Physica A **266**, 96 (1999).
- [31] A. Coniglio, J. Phys. A: Math. Gen. **15**, 3829 (1982).
- [32] Y. Gefen, A. Aharony, B. B. Mandelbrot, and S. Kirkpatrick, Phys. Rev. Lett. **47**, 1771 (1981).
- [33] A. Bunde and S. Havlin, *Fractals and Disordered Systems* (Springer, Berlin, 1991).
- [34] S. Alexander and R. Orbach, J. Physique Lett. **43**, 625 (1982).
- [35] D. C. Hong, S. Havlin, H. J. Herrmann, and H. E. Stanley, Phys. Rev. B **30**, 4083 (1984).
- [36] P. Grassberger, Physica A **262**, 251 (1999).
- [37] J. Adler, Y. Meir, A. Aharony, A. Harris, and L. Klein, J. Stat. Phys. **58**, 511 (1990).
- [38] M. Porto, A. Bunde, S. Havlin, and H. E. Roman, Phys. Rev. E **56**, 1667 (1997).
- [39] B. K. Chakrabarti and J. Kertész, Z. Phys. B **44**, 221 (1981).
- [40] K. Kremer, Z. Phys. B **45**, 149 (1981).

- [41] B. Derrida, *J. Phys. A: Math. Gen.* **15**, L119 (1982).
- [42] A. Roy and B. Chakrabarti, *Physics Letters A* **91**, 393 (1982).
- [43] Y. Kim, *J. Phys. C: Solid State Phys.* **16**, 1345 (1983).
- [44] A. B. Harris, *Z. Phys. B* **49**, 347 (1983).
- [45] J. Rexakis and P. Argyrakis, *Phys. Rev. B* **28**, 5323 (1983).
- [46] B. Derrida, *Phys. Rep.* **103**, 29 (1984).
- [47] P. Lam and Z. Zhang, *Z. Phys. B* **56**, 155 (1984).
- [48] R. Rammal, G. Toulouse, and J. Vannimenus, *J. Phys. (Paris)* **45**, 389 (1984).
- [49] M. Sahimi, *J. Phys. A: Math. Gen.* **17**, L379 (1984).
- [50] J. W. Lyklema and K. Kremer, *Z. Phys. B* **55**, 41 (1984).
- [51] J. Nadal and J. Vannimenus, *J. Phys. France* **46**, 17 (1985).
- [52] A. K. Roy and S. S. Manna, *Z. Phys. B* **61**, 205 (1985).
- [53] D. Marković, S. Milošević, and H. Stanley, *Physica A* **144**, 1 (1987).
- [54] S. B. Lee and H. Nakanishi, *Phys. Rev. Lett.* **61**, 2022 (1988).
- [55] A. Aharony and A. B. Harris, *J. Stat. Phys.* **54**, 1091 (1989).
- [56] J. Machta, *Phys. Rev. A* **40**, 1720 (1989).
- [57] J. Machta and R. A. Guyer, *J. Phys. A: Math. Gen.* **22**, 2539 (1989).
- [58] S. B. Lee, H. Nakanishi, and Y. Kim, *Phys. Rev. B* **39**, 9561 (1989).
- [59] Y. Meir and B. Harris, *Phys. Rev. Lett.* **63**, 2819 (1989).
- [60] Y. Kim, *Phys. Rev. A* **41**, 4554 (1990).
- [61] P. M. Lam, *J. Phys. A: Math. Gen.* **23**, L831 (1990).
- [62] A. Roy and A. Blumen, *J. Stat. Phys.* **59**, 1581 (1990).
- [63] S. P. Obukhov, *Phys. Rev. A* **42**, 2015 (1990).
- [64] J. Machta and T. R. Kirkpatrick, *Phys. Rev. A* **41**, 5345 (1990).
- [65] P. Le Doussal and J. Machta, *J. Stat. Phys.* **64**, 541 (1991).
- [66] H. Nakanishi and S. B. Lee, *J. Phys. A: Math. Gen.* **24**, 1355 (1991).
- [67] K. Y. Woo and S. B. Lee, *Phys. Rev. A* **44**, 999 (1991).

Bibliography

- [68] C. Vanderzande and A. Komoda, *Europhys. Lett.* **14**, 677 (1991).
- [69] Y. Kim, *Phys. Rev. A* **45**, 6103 (1992).
- [70] H. Nakanishi and M. J, *Physica A* **191**, 309 (1992).
- [71] C. Vanderzande and A. Komoda, *Phys. Rev. A* **45**, R5335 (1992).
- [72] P. Grassberger, *J. Phys. A: Math. Gen.* **26**, 1023 (1993).
- [73] M. Rintoul, J. Moon, and H. Nakanishi, *Phys. Rev. E* **49**, 2790 (1994).
- [74] K. Barat and B. K. Chakrabarti, *Phys. Rep.* **258**, 377 (1995).
- [75] S. B. Lee, *J. Korean Phys. Soc.* **29**, 1 (1996).
- [76] J. Hovi and A. Aharony, *J. Stat. Phys.* **86**, 1163 (1997).
- [77] H. E. Roman, M. Porto, A. Ordemann, A. Bunde, and S. Havlin, *Philos. Mag. B* **77**, 1357 (1998).
- [78] A. Ordemann, M. Porto, H. E. Roman, S. Havlin, and A. Bunde, *Phys. Rev. E* **61**, 6858 (2000).
- [79] A. Ordemann, M. Porto, H. Eduardo Roman, and S. Havlin, *Phys. Rev. E* **63**, 020104 (2001).
- [80] C. v. Ferber, V. Blavatska, R. Folk, and Y. Holovatch, *Phys. Rev. E* **70**, 035104 (2004).
- [81] H. K. Janssen and O. Stenull, *Phys. Rev. E* **75**, 020801 (2007).
- [82] V. Blavatska and W. Janke, *Europhys. Lett.* **82**, 66006 (2008).
- [83] V. Blavatska and W. Janke, *Phys. Rev. Lett.* **101**, 125701 (2008).
- [84] V. Blavatska and W. Janke, *J. Phys. A: Math. Theor.* **42**, 015001 (2009).
- [85] V. Blavatska and W. Janke, *Physics Procedia* **3**, 1431 (2010).
- [86] H. K. Janssen and O. Stenull, *Phys. Rev. E* **85**, 011123 (2012).
- [87] V. Dotsenko, *Replica Theory of Disordered Statistical Systems* (Cambridge University Press, Cambridge, 2001).
- [88] A. B. Harris, *J. Phys. C: Solid State Phys.* **7**, 1671 (1974).
- [89] J.-P. Bouchaud and A. Georges, *Phys. Rev. B* **39**, 2846 (1989).
- [90] P. M. Lam and Z. Q. Zhang, *Z. Phys. B* **57**, 65 (1984).
- [91] D. Dhar, *J. Math. Phys.* **19**, 5 (1978).

- [92] D. Ben-Avraham and S. Havlin, Phys. Rev. A **29**, 2309 (1984).
- [93] Y. Liu and Z. Q. Zhang, J. Phys. A: Math. Gen. **18**, 1027 (1985).
- [94] F. D. A. A. Reis and R. Riera, J. Stat. Phys. **71**, 453 (1993).
- [95] Y. Hotta, Phys. Rev. E **90**, 052821 (2014).
- [96] V. Blavatska, C. von Ferber, and Y. Holovatch, in *Statistics of Linear Polymers in Disordered Media*, edited by B. Chakrabarti (Elsevier, Amsterdam, 2005).
- [97] H.-K. Janssen and O. Stenull, Phys. Rev. E **85**, 051126 (2012).
- [98] L. Peliti, J. Physique Lett. **45**, 925 (1984).
- [99] L. Pietronero, Phys. Rev. Lett. **55**, 2025 (1985).
- [100] A. L. Stella, Phys. Rev. Lett. **56**, 2430 (1986).
- [101] N. Fricke, *Self-avoiding walks on disordered lattices*, Diploma thesis, Universität Leipzig (2010).
- [102] P. Grassberger, Phys. Rev. E **56**, 3682 (1997).
- [103] N. Fricke and W. Janke, Physics Procedia **34**, 39 (2012).
- [104] N. Fricke and W. Janke, Europhys. Lett. **99**, 56005 (2012).
- [105] N. Fricke and W. Janke, Eur. Phys. J. Special Topics **216**, 175 (2013).
- [106] N. Fricke and W. Janke, Phys. Rev. Lett. **113**, 255701 (2014).
- [107] I. Jensen, J. Phys. A: Math. Gen. **37**, 5503 (2004).
- [108] R. D. Schram, G. T. Barkema, and R. H. Bisseling, Comput. Phys. Commun. **184**, 891 (2013).
- [109] N. Clisby, J. Stat. Phys. **140**, 349 (2010).
- [110] M. N. Rosenbluth and A. W. Rosenbluth, J. Chem. Phys. **23**, 356 (1955).
- [111] H.-P. Hsu, V. Mehra, W. Nadler, and P. Grassberger, J. Chem. Phys. **118**, 444 (2003).
- [112] H.-P. Hsu and P. Grassberger, J. Stat. Phys. **144**, 597 (2011).
- [113] A. R. Singh, D. Giri, and S. Kumar, Phys. Rev. E **79**, 051801 (2009).
- [114] D. E. Knuth, *The Art of Computer Programming, Vol. 4, Fascicle 4* (Addison-Wesley, Upper Saddle River, NJ, 2006).
- [115] N. Clisby, R. Liang, and G. Slade, J. Phys. A: Math. Theor. **40**, 10973 (2007).

Bibliography

- [116] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing* (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975).
- [117] S. Kumar, I. Jensen, J. L. Jacobsen, and A. J. Guttmann, *Phys. Rev. Lett.* **98**, 128101 (2007).
- [118] B. Efron, *Ann. Stat.* **7**, 1 (1979).
- [119] A. J. Guttmann, in *Phase Transitions and Critical Phenomena*, Vol. 13, edited by C. Domb and J. L. Lebowitz (Academic Press, New York, 1989) p. 1.
- [120] V. Blavatska and W. Janke, *Phys. Rev. E* **80**, 051805 (2009).
- [121] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [122] A. Giacometti and A. Maritan, *Phys. Rev. E* **49**, 227 (1994).
- [123] A. J. Guttmann and A. R. Conway, *Ann. Comb.* **5**, 319345 (2001).
- [124] P. De Gennes, *J. Physique Lett.* **36**, 55 (1975).
- [125] B. Duplantier and H. Saleur, *Phys. Rev. Lett.* **59**, 539 (1987).
- [126] B. K. Chakrabarti and S. M. Bhattacharjee, *J. Stat. Phys.* **58**, 383 (1990).
- [127] A. K. Roy, B. K. Chakrabarti, and A. Blumen, *J. Stat. Phys.* **61**, 903 (1990).
- [128] I. Chang and A. Aharony, *J. Phys. I France* **1**, 313 (1991).
- [129] K. Barat, S. N. Karmakar, and B. K. Chakrabarti, *J. Phys. A: Math. Gen.* **24**, 851 (1991).
- [130] K. Barat, S. N. Karmakar, and B. K. Chakrabarti, *J. Phys. A: Math. Gen.* **25**, 2745 (1992).
- [131] K. Barat, S. N. Karmakar, and B. K. Chakrabarti, *J. Phys. I France* **3**, 2007 (1993).
- [132] V. Blavatska and W. Janke, *Comput. Phys. Comm.* **182**, 1966 (2011).
- [133] N. Fricke, J. Bock, and W. Janke, *diffusion-fundamentals.org* **20**, 111 (2013).
- [134] I. Majid, N. Jan, A. Coniglio, and H. E. Stanley, *Phys. Rev. Lett.* **52**, 1257 (1984).
- [135] P. G. De Gennes, *La Recherche* **7**, 919 (1976).
- [136] K. Kremer and J. W. Lyklema, *Phys. Rev. Lett.* **55**, 2091 (1985).
- [137] P. H. Poole, A. Coniglio, N. Jan, and H. E. Stanley, *Phys. Rev. B* **39**, 495 (1989).

- [138] M. Ponmurugan, S. Narasimhan, and K. Murthy, *Physica A* **371**, 171 (2006).
- [139] S. L. Narasimhan, *Phys. Rev. E* **53**, 1986 (1996).
- [140] A. Giacometti and A. Maritan, *J. Phys. A: Math. Gen.* **25**, 2753 (1992).
- [141] D. Lekić and S. Elezović-Hadžić, *Physica A* **390**, 1941 (2011).
- [142] C. Song, H. Makse, and S. Havlin, **433**, 392 (2005).
- [143] C. Song, S. Havlin, and H. A. Makse, *Nature Phys.* **2**, 275 (2006).
- [144] P. L. Leath, *Phys. Rev. B* **14**, 5046 (1976).
- [145] J. Hoshen and R. Kopelman, *Phys. Rev. B* **14**, 3438 (1976).
- [146] M. Aizenman, *Nucl. Phys. B* **485**, 551 (1997).
- [147] M. Heydenreich and R. van der Hofstad, *Commun. in Math. Phys.* **270**, 335 (2007).
- [148] M. Heydenreich and R. van der Hofstad, *Probab. Theory Rel.* **149**, 397 (2011).
- [149] P. Grassberger, *J. Phys. A: Math. Gen.* **19**, 1681 (1986).
- [150] N. Jan, D. C. Hong, and H. E. Stanley, *J. Phys. A: Math. Gen.* **18**, L935 (1985).
- [151] G. Paul, R. M. Ziff, and H. E. Stanley, *Phys. Rev. E* **64**, 026115 (2001).
- [152] R. Tarjan, *Inform. Process. Lett.* **2**, 160 (1974).

Acknowledgments

My work was funded for the most part by the Deutsche Forschungsgemeinschaft (DFG) via Grant No. JA 483/29-1 of the Sächsische Forschergruppe FOR 877. Further funding was provided by the ESF. I also received considerable support from the Leipzig Graduate School of Excellence GSC185 “BuildMoNa” in the form of travel funding, access to scientific training, and work equipment. Travel funding was also supplied by the Deutsch-Französische Hochschule (DFH-UFA) under grant No. CDFA-02-07 and by the Alexander von Humboldt Foundation (AvH). This has allowed me to spend interesting and exciting weeks with scientists in Nancy, Coventry, and Lviv.

I gratefully acknowledge the generous support given by these friendly institutions, but most of my thanks go to the people who supported me during these last years. First of all I want to thank my supervisor Wolfhard Janke for constant support and advice; the numerous endless discussions have helped shaping not only this thesis but also my development as a scientist. I furthermore want to thank all my dear colleagues and friends at the Institute of Theoretical Physics for the nice atmosphere. Special thanks go to Johannes Zierenberg and Martin Marenz for many helpful conversations and for proofreading parts of this thesis. As all members of the ITP, I am forever indebted to Martin and to Hannes Nagel for keeping the Computers alive and for patiently helping me with countless technical problems.

Last but not least, I want to thank all the people who kept me going: my friends (in particular Kelvin Autenrieth and Johannes Stelzer, who helped me getting this done), my family, and my girlfriend Carina.

Academic Curriculum Vitae

Personal Details:

Name	Niklas Fricke
Date of birth	January 09, 1984
Nationality	German

Education:

1990 - 1997	Freie Waldorfschule Kassel Grundschule
1998 - 2000	Albert-Schweitzer-Schule Kassel
2000 - 2003	Jakob-Grimm-Schule Kassel, <i>Abitur</i>
2003 - 2005	Studies of Physics at the Universität Hamburg, <i>Vordiplom</i>
2005 - 2010	Studies of Physics at the Universität Leipzig, <i>Diplom</i>
since 2011	Ph.D. thesis in Theoretical Physics at the Universität Leipzig within the Graduate School “BuildMoNa”

Publications:

- N. Fricke, S. Sturm, M. Lämmel, S. Schöbl, K. Kroy, and W. Janke, *Polymers in Disordered Environments*, diffusion-fundamentals.org **23**, 7 (2015).
- N. Fricke, and W. Janke, *Asymptotic Scaling Behavior of Self-Avoiding Walks on Critical Percolation Clusters*, Phys. Rev. Lett. **113**, 255701 (2014).
- V. Blavatska, and N. Fricke, and W. Janke, *Polymers in Disordered Environments*, Condens. Matter Phys. **17**, 33604 (2014).
- N. Fricke, and J. Bock, and W. Janke, *Diffusion and Polymers in Fractal, Disordered Environments*, diffusion-fundamentals.org **20**, 111 (2013).
- N. Fricke, and W. Janke, *Self-avoiding walks on strongly diluted lattices: Chain-growth simulations vs. exact enumeration*, Eur. Phys. J. - Special Topics **216**, 175 (2013).
- N. Fricke, and W. Janke, *Scale-free enumeration of self-avoiding walks on critical percolation clusters*, Europhys. Lett. **99**, 56005 (2012).
- N. Fricke, and W. Janke, *Exact Enumeration of Self-Avoiding Walks on Percolation Clusters*, Physics Procedia **34**, 39 (2012).

Bibliographische Beschreibung:

Fricke, Niklas

Polymers in Fractal Disorder

Universität Leipzig, Dissertation

159 S., 152 Lit., 79 Abb., 30 Tab.

Referat:

Die vorliegende Arbeit präsentiert eine numerische Studie von selbstvermeidenden Zufallswegen (SAWs) auf Perkolationsclustern, ein kanonisches Modell für Polymere in stark ungeordneten Medien. Hierfür wurde ein neuer Algorithmus entwickelt, welcher es ermöglicht SAWs von mehr als zehntausend Schritten exakt auszuzählen. Dies bedeutet eine Steigerung von mehreren Größenordnungen gegenüber der zuvor existierenden Methode, welche kaum mehr als vierzig Schritte zulässt. Solch eine Steigerung wird erreicht, indem die fraktale Struktur der Perkolationscluster gezielt ausgenutzt wird: Die Cluster werden hierarchisch in lose verbundene Gebiete unterteilt, innerhalb welcher Wegstücke separat ausgezählt werden können. Nach dem Auszählen wird ein Gebiet "dezimiert" und verhält sich während der Behandlung größerer Gebiete effektiv wie ein Gitterpunkt. Da diese neue Methode nur nahe der Perkolationsschwelle funktioniert, wurde zum Erzielen der Ergebnisse zudem ein Kettenwachstums-Monte-Carlo-Algorithmus (PERM) eingesetzt.

Untersucht wurde zunächst das asymptotische Skalenverhalten des Abstands der beiden Kettenenden als Funktion der Schrittzahl auf kritischen Clustern in verschiedenen Dimensionen. Dank der neuen hochperformanten Methode konnten die bisherigen Schätzer für den dies beschreibenden Exponenten signifikant verbessert werden. Neben dem Abstand wurde zudem die Anzahl der möglichen Konformationen und die mittlere Entropie angeschaut, für welche ein ungewöhnliches Skalenverhalten gefunden wurde. Für Konzentrationen oberhalb der Perkolationsschwelle wurde festgestellt, dass der Exponent, welcher das Wachstum des Endabstands beschreibt, nicht dem für freie SAWs entspricht, was nach gängiger Lehrmeinung der Fall sein sollte. Schlussendlich wurden SAWs mit Anziehung zwischen benachbarten Monomeren untersucht. Hier zeigte sich, dass es auf kritischen Perkolationsclustern keinen Phasenübergang zu geben scheint, an welchem die Ketten kollabieren, sondern dass das Skalenverhalten des Endabstands selbst am absoluten Nullpunkt der Temperatur unverändert ist.