UNIVERSITÀ DEGLI STUDI DI MILANO
Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di laurea in Fisica

COMPORTAMENTO CRITICO DEL MODELLO DI DOMB–JOYCE

Relatore: Prof. Sergio Caracciolo

Correlatore: Prof. Bruno Bassetti

PACS: 05.40.Fb, 05.70.Jk, 61.25.Hq, 64.60.Fr

Tesi di Laurea di
Marco GHERARDI
matr. 604032

Anno Accademico 2003/2004

"Anche se la via dell'acqua è sconosciuta all'acqua, l'acqua funziona; anche se la via dell'acqua non è sconosciuta all'acqua, l'acqua effettivamente funziona."

(Eihei Dōgen Zenji, Shōbōgenzō-Sansui-Kyō)

Ai miei genitori

Contents

Chapter 1

Introduction

The *Domb–Joyce model* is a model in statistical mechanics, which has been introduced to describe polymers in good solutions through N-step lattice walks. The interaction Hamiltonian is defined in such a way as to discourage self-intersections of the walks, and the strength of the self-excluding constraint is regulated by an *excluded volume* parameter v. For $v = 0$ the model reduces to the Random Walk (non-constrained free walk), while for $v = \infty$ it reduces to the Self-Avoiding walk (completely self-excluding walk). As a consequence, the Domb–Joyce model appears as an interpolating model between the two limiting cases of RW and SAW.

The most interesting properties are observed in the large-N regime, where the static behavior ceases to be dependent on the microscopic particulars of the model, and *universality* is displayed. In general, the study of systems with infinite degrees of freedom is interesting exactly because some properties are independent of the underlying micro-structure. This happens in the *critical limit*, that is, the limit where an appropriate correlation length — which measure the scale length of likely fluctuations of the system — diverges, and some quantities show non-continuous behavior. The critical limit corresponds to the limit $N \to \infty$ in walk models; clearly the walk has only a finite number of degrees of freedom for N finite, so they can not sum up but to smooth functions. It is quantities defined in the critical limit, that are universal; that is, they do not depend on the particular form of the interaction — as long as it remains short-range in space — nor on the underlying lattice. Universality is the reason why extremely simple models — like the Domb–Joyce model — can be used to describe in some measure real physical systems — like polymers in good solutions, which are governed by manifold types of interactions, and of which there exists a wide variety of chemically and structurally different realizations.

Above 4 dimensions the critical behavior of polymer solutions is expected

to be Gaussian — that is, 4 is the *upper critical dimension*. The three-dimensional Domb–Joyce model has been deeply studied, and many results are found in literature, also for other model in the same universality class. Here, we wish to study the two-dimensional case on the square lattice by means of a very efficient Monte Carlo technique, and to compare numerical results with theoretical predictions. We focus on the observable *mean square end-to-end distance*, which is a measure of the swelling of the chain.

Two very different (critical) limits will be investigated. The first is the usual scaling limit, where N is sent to infinity, while v remains fixed. In this limit, the Domb–Joyce model is thought to belong to the SAW universality class for every $v > 0$. One then expects the transition $v \to 0$ to be discontinuous, because in the critical limit the model jumps from one kind of behavior (SAW-like) to another (RW-like). This is true at the critical point $N = \infty$, but far from the critical region this transition is smooth. As a consequence, one is interested in the characteristics of this transition, which is called *crossover*. Unfortunately, crossover behavior is not universal, and such a greatly non-critical regime can be described only through fenomenological models. Yet, there happens to exist a limit — called *critical crossover limit* — in which the system shows universality: it is the limit where $N \to \infty$ and $v \to 0$ while their product is kept fixed. This limit corresponds to going to the critical limit ($N \to \infty$) while still observing the crossover region. The second limit we are interested in is exactly the critical crossover limit.

In chapter 2, a brief account on polymers solutions and their modelization is given, with particular stress on good solutions. Chapter 3 contains a thorough description of Monte Carlo methods applied to walk models, with particular attention on the statistical analysis of data, and on the *Pivot algorithm*, which is the algorithm we use to simulate the model on a computer. In chapter 4, we give a review of the theoretical tools and the results that can be found in literature, and we obtain — with a standard cluster expansion — the one-loop form of the crossover function, which describes the behavior of the observable in the critical crossover limit. Chapter 5 is devoted to the presentation of numerical result. Finally, chapter 6 draws the conclusions.

Chapter 2

Polymers in Solution

2.1 Generalities

A *polymer* [1, 2, 3, 4, 5, 6] is a long, repeating chain of atoms, formed through the linkage of many molecules, called *monomers*. When the number of bonds becomes large, the overall polymer dimensions greatly exceed those of the constituent monomer units; because of its large dimensions the polymer is also called a *macromolecule*; real polymers can consist of more than 10^5 monomers.

$$X - X - X - \cdots - X$$

If the monomers (X) are identical then the chain is called a *homopolymer*, otherwise it is called a *heteropolymer*. Here, we will be interested mainly in the former type of chains. By appending different long polymers after each other one obtains a *copolymer*. The number of monomers in a single chain (usually denoted by N) is called *degree of polymerization*, and is a very important quantity, because it is responsible — as will be clear in the following — of the approach to criticality. When one considers ensembles of polymers, two different choices are possible with respect to the degree of polymerization: one considers either fixed-N or variable-N ensembles. The former case is referred to as *monodisperse*, while the latter is the *polydisperse* ensemble.

Adjacent monomers along the chain are bound together by covalent bonds. This kind of linkage does not in general allow complete independence between successive monomers: each monomer is given only a certain freedom of rotation about the axis of the preceding unit. It is the characteristics of this very freedom that account for the different degrees of *flexibility* that are observed in real polymers. A polymer is called flexible if the correlation between the orientations of successive bonds persists only over the range of a few monomer

units, otherwise it is called stiff. The actual range of this correlation (that is, the flexibility of the polymer) is not expected to affect the behavior of very long chains, as long as it remains finite so that many important quantities one is interested in are not sensitive on the details of this property.

The monomer units do not in general have a *functionality* of only two, that is they need not form only two bonds with two other monomers. When the monomers are bonded one after another as in the figure on page 5 they form a *linear polymer*. But in general one can obtain more complex branched structures or lattice-like structures, where the elements are (linear) polymer segments. Two very popular structures of this kind are the so called *star polymers*, in which many different polymer segments share the same endpoint, and *comb polymers*, chains constituted of a polymeric backbone with several branches attached along it.

When a large number of polymers is put together, the different chains interact with each other. The bulk properties of this ensemble depend on the type of polymers and on the temperature: the phase can be solid, liquid, or glassy. The liquid phase is characterized by its viscosity, which raises rapidly with the molecular mass, but its behavior under external solicitation is viscoelastic, that is, the liquid reacts at first in an elastic way, and the viscosity manifest itself only after a time which raises with the molecular mass. In general, the bulk properties of mass polymers are hard to describe mathematically, because several kinds of interactions are involved, and because the polymers all interact with each other. An example of the complex behavior of these ensembles is the glassy transition, that is the transition between the glassy and the solid phase: the transition temperature depends on the speed at which the temperature is lowered. This complexity is smoothed when one considers polymers in dilute solutions [see section 2.2].

A complete theory of polymers — be it solutions, melts, or solid state — would be enormously complicated and almost certainly the mathematical and computational difficulties in solving or extracting valuable information from a thorough model would be severe. Modelization of polymers in the framework of statistical mechanics and the renormalization group [see chapter 4] helps in both simplifying the models one has to study and providing new techniques and new perspectives that are valuable tools in coping with the problems encountered when investigating such a vast and faceted subject.

In general, a polymer model on a configuration space Ω (be it on a lattice, on continuum space, with discrete segments or with continuous chains) can be described by means of an ensemble of instances $\{\omega\}$, specified — with respect to an underlying measure $d\mu$ on the state space — by a Hamiltonian $H[\omega]$ which specifies the energy of a configuration. The Hamiltonian gives information about the system statistics through the postulation of the Gibbs

distribution, by which the mean value of an observable $O[\omega]$ over the ensemble is given by

$$\langle O \rangle_H = \frac{1}{Z} \int_\Omega e^{-\beta H[\omega]} O[\omega] d\mu[\omega] \tag{2.1}$$

where

$$Z = \int_\Omega e^{-\beta H[\omega]} d\mu[\omega] \tag{2.2}$$

is called *partition function*, and β is the inverse temperature.

2.2 The Role of Solution

When two monomers become close in space, they repel each other through a force F whose strength depends on the separation between them, as well as on the chemistry of each bead and on temperature and pressure conditions. When a liquid solvent is introduced into the system, and a dilute solution is generated, other kinds of interactions are present. Now each bead has a chance to interact with solvent molecules as well as other beads. As a result, the effective force $F + F'$ that acts between a pair of beads is no longer equal to the vacuum value F, because it gets mediated by the solvent, which contributes a force F'. The type of this mediation — which modifies the effective interaction that drives the chain — depends on the particular choice of the two constituents of the solution, and on the temperature.

Generally speaking, the monomer–monomer interactions can be divided into two categories:

- Short-range interaction. This indicates those interactions that have a short range along the chain; they are responsible for the local connectivity of the polymer — that is, its chain-like structure — and for the kind of constraints that successive monomer units have to comply.

- Long-range interactions. These interactions are long-range along the chain, but are usually short-range in space — an important exception is the case where charged species are present, but treating long-range interacting chains is hard from the theoretical point of view. These forces can be both attractive and repulsive, and are responsible for the deviation from ideal-chain (Gaussian) behavior.

The configuration statistics of short-range interaction polymer systems can be treated with standard mathematical tools, such as the Markov methods; hence, models involving only this type of interactions are often exactly solvable. When it comes to studying the long-range case, things get much more complicated.

Effective — solvent-mediated — interactions of the second type can be either repulsive or attractive. This is an effect of the presence of the solvent. If the bead-solvent interaction favors bead-solvent contact over the bead-bead one, the solvent is said to be a *good solvent* (for the polymer and temperature considered), while if the reverse is the case, it is called a *bad solvent*. In a good solvent, then, the segment-solvent interaction tends to pull a pair of segments apart, so that the effective force F' should be repulsive as is F. On the other hand F' will be attractive in poor solvents. A solution of two fixed constituents can be made poorer or better by changing the temperature. There happen to exist a certain temperature — called *theta temperature* — for which on average $F+F'=0$. Roughly speaking, at the theta temperature the attractive forces exactly balance the repulsive ones; the chain is then expected to behave ideally. The theta point is then seen to be similar to the Boyle temperature for real gases, which is the temperature for which the (effective) behavior is ideal.

In the good solvent regime, the chain is expected to be in a swelled state. This means that the average extension of the polymer has to be larger than in the non-interacting case, because now the interactions are mainly repulsive. On the contrary, for a poor solvent under the theta temperature, the polymer is expected to be dominated by the attractive interactions between its monomers, so that it will be found in a globular self-collapsed state. When attractive forces prevail, the monomers all tend to occupy the same volume. Yet, this is physically impossible, because short-range steric (hard-core) repulsion — which is due to atom–atom forces between the monomers, and no longer to the solvent-mediated van der Waals interactions — forbids the coincidence of different monomers. Thus, below the theta temperature the chain is found in a low-entropy globular phase, and the statistics is energy-dominated, while above the theta point the chain is found in a swelled, high-entropy phase.

A natural observable which measures the overall swelling of the chain is the *mean square end-to-end distance* (sometimes it will be called simply *end-to-end distance*)

$$\left\langle \mathbf{R}_e^2 \right\rangle := \left\langle \left(\sum_i \mathbf{r}_i \right)^2 \right\rangle \tag{2.3}$$

where the $\mathbf{r}_i$ are the vectors specifying the monomer locations. The end-to-end distance is observed to follow a power-law behavior for N large:

$$\left\langle \mathbf{R}_e^2 \right\rangle \sim N^{2\nu} \tag{2.4}$$

Thus one expects ν to be $\sim 1/d$ in the collapsed phase — where the chain densely occupies a volume — and $\nu \geq \nu_{ideal}$ in the swelled phase, where

ν_{ideal} is the corresponding value for the ideal chain. A direct measure of the swelling of the chain is the *swelling* or *expansion factor*

$$\alpha^2 := \frac{\langle \mathbf{R}_e^2 \rangle}{\langle \mathbf{R}_e^2 \rangle_0} \tag{2.5}$$

that is, the ratio of the mean square end-to-end distance of the excluded-volume chain to that of the ideal chain.

When describing a polymer solution within the framework of statistical mechanics, the Hamiltonian is given by the total interaction energy W:

$$H[\omega] = W\left(\{\mathbf{R}_i\}\right) \tag{2.6}$$

where $\{\mathbf{R}_i\}$ stands for the set of all monomer displacement vectors. For the treatment of the theory it is necessary to make some approximation. The fundamental approximation which is done in the good-solvent regime is the *binary cluster approximation* or *superposition approximation*

$$W\left(\{\mathbf{R}_i\}\right) = \sum_{0 \le i < j \le N} w(\mathbf{R}_{ij}) \tag{2.7}$$

where $w(\mathbf{R}_{ij}$ is the pair potential between the i-th and the j-th bonds as a function of their separation. Equation (2.7) expresses the fact that only *two-body interactions* are present, which seems to be a reasonable assumption, since one does not expect that three-monomer interactions even exist.

Yet, the binary cluster approximation is valid only in the good-solvent regime. Under the theta temperature, attractive forces begin to win over repulsive ones. If one assumes only purely attractive interactions, a simple argument [see section XIV-2.1 in [5]] shows the triviality of the model. Pick for example a purely attractive walk on a square lattice. Indeed, it is easy to see that entropy is always proportional to the chain-length N, while the energy of the fundamental state is proportional to N^2 — the fundamental state being the chain $\{0, \mu, 0, \mu, \ldots\}$, where μ is a neighbor of the origin 0. Hence, the attractive phase is energy-dominated. Moreover, the energy of the first excited state is seen to be orders of N greater than the fundamental state, so that in the $N \to \infty$ limit the chain will essentially stay in the latter. This implies that the mean square end-to-end radius is equal to the lattice constant. The situation is even worse when one takes a continuum limit of the model, by removing the lattice, because the chain is seen to collapse onto a single point of the space, and its mean metric dimension is null. It is then clear that two-body interactions are not enough, and many-body terms have to be introduced in order to obtain a non-trivial theory of poor and theta solvents.

Another approximation — which is implicit when one postulates a canonical measure like the one in (2.1) — is that the theory actually describes single chains: there are no interaction terms coming from the forces between *different* chains. Physically this can be achieved by considering only *dilute* solutions, where the mean distance between different chains is much larger than the typical chain size.

2.3 Models

Several models have been proposed to describe polymers in solution. Each has its own advantages and drawbacks: some are easier to treat theoretically, because they can be handled using the standard tools of theoretical physics, such as perturbative expansions and renormalization; others are easier to simulate and treat numerically; still others provide more realistic results in the non-critical region.

The first step in developing theories of large length scale polymer properties involves the introduction of a minimal model that adequately describes these properties while using the smallest possible number of phenomenological parameters. The simplest possible models are the continuum *random flight* and the lattice *random walk*. These are thought to correctly describe some properties of polymer solutions at the theta temperature (in dimension $d \geq 3$).

Away from the theta point, the effective interactions between distant polymer units become an important aspect for the statistic description of the system, even regarding those properties that are thought not to depend on the strength or short-scale details of these interactions. So more sophisticated models are to be considered; these need to have a larger number of tunable parameters, thus increasing the difficulties in theoretical and numerical investigation of these models.

2.3.1 Chain Models

Chain models describe real polymers through discrete straight segments attached to each other in a chain-like fashion. Let N denote the number of such segments, and let $\mathbf{R}_i$ denote the position in $\mathbb{R}^d$ of the i-th monomer along the chain ($i = 0 \dots N$), so that the vector describing j-th segment is $\mathbf{r}_j = \mathbf{R}_j - \mathbf{R}_{j-1}$ ($j = 1 \dots N$). The distribution functions τ_j for the bond vectors $\mathbf{r}_j$ is defined as

$$\tau_j(\mathbf{r}_j) = e^{-\beta u_j(\mathbf{r}_j)} \tag{2.8}$$

where u_j specifies the bond potential, and its zero is chosen in such a way that the probabilities $\tau_j(\mathbf{r}_j)$ are normalized. In the *random-flight model* the probability associated to a chain specified by the monomer vectors $\{\mathbf{R}_j\}_{j=0\cdots N}$ is chosen as to depend only on the τ_j's, so that every segment is independent of the rest of the chain. It is easy to guess that — by the central limit theorem — the particular choice of the bond potential does not radically change the long-chain behavior — which will be Gaussian — in accordance with universality considerations. An important scale specified by u_j is the mean square bond length

$$\left\langle |\mathbf{r}_j|^2 \right\rangle = \int d\mathbf{r} |\mathbf{r}|^2 u_j(\mathbf{r}) = a_j^2 \tag{2.9}$$

If the bond distribution is the same for each bond, then the mean square end-to-end distance is easily calculated in terms of the unique bond length a:

$$\left\langle R_e^2 \right\rangle = \left\langle \left(\sum_i \mathbf{r}_i \right)^2 \right\rangle = \sum_{i=1}^{N} \left\langle |\mathbf{r}_i|^2 \right\rangle + \sum_{i \neq j} \left\langle \mathbf{r}_i \cdot \mathbf{r}_j \right\rangle = a^2 N \tag{2.10}$$

In some references the random-flight chain is taken as a chain with the Markov property of independence between different bond distributions and with all rod-like fixed-length segments. This model is not convenient for theoretical formulation of global properties, because of the constraint that the length of each segment be fixed.

A particular form of the random-flight model is the *spring-bead* model, where the potential between two successive monomers is taken to be elastic, so that the chain can be thought as being composed of beads joined together by springs. A nice feature of this model is that it can be demonstrated that a spring-bead chain — which is 'locally Gaussian' by definition — behaves as a Gaussian regardless of the value of N, so that the behavior of its observables can be calculated exactly and the large-N expansion can be actually written in a closed form without corrections to scaling.

Other different chain models are possible and are found in literature. The *freely rotating chain* is a polymer made of fixed-length segments that are not completely free to rotate, but are subject to the constraint that the angle between neighboring segments is fixed. A continuous chain generated from the freely rotating chain by an appropriate limiting process has been proposed, and is called *Kratky-Porod chain*, or *worm-like chain*.

All these models present a local granular structure, that corresponds to the microscopic granularity of real polymers. This is not a fundamental requirement of a polymer model, and *continuous chain models* have been studied, which are described in continuum space (the chain is parametrized

by means of a continuous index) so that they do not have local stiffness [see section 2.3.3 for an example]. These models are usually easier to treat with the standard tools of theoretical physics, so that they have been widely used to study polymer properties with renormalization group techniques.

2.3.2 Walk Models

Walk models are defined on a lattice, so they are naturally regularized with respect to ultraviolet divergences. These turn out to be very friendly models when trying to investigate polymer properties with numerical methods such as Monte Carlo or series extrapolation [see chapter 3]. On the other hand they are usually harder to treat theoretically, and require additional care in understanding how non-universal properties depend on the choice of the underlying lattice. The role of the lattice is a non-trivial one, not only on a local level. Actually, in certain circumstances careful choice of the lattice even permits exact solution of some special models.

On the other hand, the lattice is often responsible for artifacts due to the special local and global properties it has. For example, on a triangular lattice a walk has a non-zero probability of returning to the origin after $N \geq 2$ steps, while on a square lattice this probability is null for every odd N. For the self-avoiding walk this parity breaking is responsible for additional corrections to scaling with different sign for N and $N+1$; these are called *antiferromagnetic corrections*, and are strictly linked to the appearance of an additional singularity of the generating functions (for a negative real value in the complex temperature plane) which is called *antiferromagnetic singularity*.

The simplest walk model on a lattice is the *random walk* (RW)[7, 8], which describes polymers in a condition where all forces acting on the monomers are neglected. A RW is a path on the lattice which is completely unrestricted; the ensemble of RWs is defined as the set of all RW where each walk is given the same statistical weight. This model is a true Markov process[1], since the probability distribution for the N-th point depends only on that for the $(N-1)$-th point. Many properties of the RW can be calculated exactly. Let $P(\mathbf{x}_1, t_1; \mathbf{x}_0, t_0)$ be the conditional probability for the walk to be on the site $\mathbf{x}_1$ at time t_1, knowing that its initial position is $\mathbf{x}_0$ at time t_0. The obvious observation that a walker can reach a certain point only if one unit of time before it is on a neighboring site leads to a recurrence relation for P:

$$P(\mathbf{x}, t+1; \mathbf{x}_0, t_0) = \frac{1}{2d} \sum_{\hat{\mu}} P(\mathbf{x} + \hat{\mu}, t; \mathbf{x}_0, t_0) \tag{2.11}$$

[1] A brief introduction to Markov processes is given in chapter 3, but the goal is very different than the description of the RW.

where the sum is over all unit length vectors that generate the lattice. The seed of the recursion is

$$P(\mathbf{x}_1, t_0; \mathbf{x}_0, t_0) = \delta_{\mathbf{x}_1, \mathbf{x}_0} \tag{2.12}$$

The two equations (2.11) and (2.12) represent a generalization to arbitrary dimension of Pascal's construction of the binomial coefficient, which corresponds to $d = 1$.

These recursion relations can be easily solved by working in momentum space, and give

$$P(\mathbf{x}_1, t_1; \mathbf{x}_0, t_0) = \int_{-\pi}^{\pi} \frac{d\mathbf{k}}{(2\pi)^d} e^{i\mathbf{k}\cdot(\mathbf{x}_1 - \mathbf{x}_0)} \left(\frac{1}{d} \sum_{\mu} \cos k_\mu \right)^{t_1 - t_0} \tag{2.13}$$

A non-trivial limit of this propagator can be obtained by letting the time-scale τ and the space-scale a go to zero in such a way that the ratio τ/a^2 be kept constant. One gets

$$P(\mathbf{x}_1 - \mathbf{x}_0, t_1 - t_0) = \frac{1}{(4\pi(t_1 - t_0))^{d/2}} \exp\left(-\frac{(\mathbf{x}_1 - \mathbf{x}_0)^2}{4(t_1 - t_0)} \right) \tag{2.14}$$

that is the well-known kernel of the diffusion equation in continuous space. From this distribution it is possible to compute the different observables of interest; for example it can be seen that the squared end-to-end distance $\langle |\mathbf{x}_1 - \mathbf{x}_0|^2 \rangle$ is proportional to $t_1 - t_0$.

The random walk is a very simple model which does not take into account any of the interactions that take place in a polymer solution. A somewhat more refined model — which is still very simple in its definition — is the *Self-Avoiding Walk*. The goal of this model is to describe the excluded volume interactions that dominate the good solvent regime.

The self-avoiding walk (SAW)[2] is a path on a lattice that does not visit the same site more than once. This simple and straightforward definition must not lead to an underestimation of the problem: the SAW turns out to be a most fascinating and rich subject, arising manifold problems and requiring deep investigation with tools borrowed from many different fields of mathematical and theoretical physics. In a strict mathematical sense we still do not know exactly even how many SAWs reach a certain point after a certain number of steps (that is, the two-point function). On the other hand, a variety of non-rigorous methods have been applied to the SAW, which have shed light on some important aspects of this model.

[2] An excellent and quite complete review of results and methods for the self-avoiding walk can be found in the book by Madras and Slade [9].

The basic questions about SAWs regard the mean distance the walk is from the starting point after N steps. Answers to this questions are known for very small values of N (from exact enumeration) or asymptotically for N large (from non-rigorous and rigorous methods, which give estimates and bounds on the quantities of interest).

In general, a SAW lives on an (undirected) graph, which is a collection of points, together with a collections of edges (pairs of points). The most common — and maybe the most simple and natural — graph is the *d-dimensional lattice* $\mathbb{Z}^d$, the points of which are the $x \in \mathbb{R}^d$ with integer coordinates, and the edges of which are all the unit vectors connecting the points. In dimension one, everything is known exactly for the SAW: once an initial direction is chosen, the walk is completely stretched in that direction, since it is never allowed to turn back. This is quite a simple and uninteresting behavior. In $d < 1$ the structure becomes vastly richer.

One defines c_N as the number of all N-step SAWs starting at the origin. Computation of $c_N(d)$ is possible for small values of N, but the combinatorics are seen to become extremely complicated as N increases. Several observations allow the computation of bounds on c_N. The simplest of such observations is that walks for which each step is in one of the d positive coordinate directions are SAW, so that d^N is a lower bound for c_N. On the other hand, the number of N-step SAWs is not greater than the number of random walks for which immediate reversal is prohibited, so that this number provides an upper bound on c_N. Summing up

$$d^N \leq c_N \leq 2d(2d-1)^{N-1} \tag{2.15}$$

When it comes to computing mean values, the flat ensemble is considered, in which all N-step SAWs get the same weight, so that for an observable $O(\omega)$

$$\langle O(\omega) \rangle_N := \frac{1}{c_N} \sum_{\omega\colon\, |\omega|=N} O(\omega) \tag{2.16}$$

where the sum is extended to all N-step SAWs.

For the simple random walk one clearly has $c_N^{RW} = (2d)^N = \mu^N$, where μ is the coordination number of the lattice. The number of random walks thus follows a simple exponential growth for every N. The behavior for the SAW is conjectured to follow a similar exponential law, but only asymptotically for large N, with a different value of μ (accounting for the effective action of the self-avoidance constraint) and with power-law corrections; that is, for large N

$$c_N \sim A\mu^N N^{\gamma-1} \tag{2.17}$$

where μ is called *connective constant* and γ is an example of a *critical exponent*[3] (that is, an exponent describing power-law behavior of observables in the vicinity of a second-order phase transition). While no prove of the finiteness of γ is yet known, the existence of the connective constant can be demonstrated rigorously as follows.

Given an N-step SAW ω_1 and an M-step SAW ω_2, their *concatenation* $\omega_1 \circ \omega_2$ is defined as the $(N+M)$-step walk obtained by appending ω_2 after ω_1. The number of such walks is $c_N c_M$, which represents the number of random walk that are self-avoiding for the initial N steps and for the final M steps, but for which other kind of intersections are allowed. Clearly such a number is an upper bound for the number of $(N+M)$-step SAWs, so that

$$c_{N+M} \leq c_N c_M \tag{2.18}$$

$$\log c_{N+M} \leq \log c_N + \log c_M \tag{2.19}$$

Equation (2.19) expresses a property of sequences of numbers called *subadditivity*. One can then apply a standard theorem and obtain that the following limit exists

$$\mu := \lim_{N\to\infty} c_N^{1/N} \tag{2.20}$$

It then follows immediately from the simple bound (2.15) that

$$d \leq \mu \leq 2d-1 \tag{2.21}$$

Much more refined bounds for c_N are available. In dimensions $d = 2$ the strictest bound is provided by the Hammersley–Welsh theorem [10] (the proof of which is based on an 'unfolding' procedure that brings a SAW into a self-avoiding bridge, which is superadditive), while for $d = 3, 4$ the best result is given by Kesten's bound [11, 9]. Both these bounds (that are the best bounds available) are in the form $c_N/\mu^N \leq \exp(O(N^p))$ for some constant $0 < p < 1$; it is a major open problem to replace this bound by a polynomial in N.

Moreover, it is conjectured that the following behaviors hold in the large-N regime

$$\begin{array}{lcll} \langle|\omega(N)|^2\rangle & \sim & BN^{2\nu} & \\ G(x;z_c) & \sim & A/|x|^{d-2+\eta} & (\text{as } |x|\to\infty) \\ c_N(x) & \sim & C\mu^N N^{\alpha-2} & \end{array} \tag{2.22}$$

where $c_N(x)$ is the number of SAWs with end-point at site x, $G(x;z_c)$ is the generating function of $c_N(x)$ evaluated at the singularity, and ν, η and α

[3]The particular expression $1-\gamma$ at the exponent is chosen to match the definition of a corresponding exponent γ for spin systems.

are critical exponents, between which hyperscaling relations hold (see [9] for details)[4].

Several modified SAW and RW models turn out to be of great interest, both in deepening the understanding of the 'full' SAW and in providing a workbench to test theoretical and numerical tools. Examples are the loop-erased RW, self avoiding bridges, self avoiding polygons, the SAW and RW in constrained geometries, the SAW with medium range jumps, the myopic SAW (sometimes called 'true' SAW, because it is defined as a true stochastic process, while the SAW is not), and the *weakly self avoiding walk* or *Domb–Joyce model*.

Loosely speaking, the Domb–Joyce model [12] is an ensemble of ordinary random walks, where each walk ω pays an energy for each self-intersection, so that self-intersecting walks are actually possible, but more or less probable depending on the interaction coupling. The measure can be defined as

$$d\mu_{DJ}^{(v)}[\omega] := \left(\prod_i d\omega_i\right) \exp\left[-v \sum_{i\in\Lambda} n_i(n_i-1)\right] \tag{2.23}$$

where n_i is the occupation number of lattice site i, and $\prod_i d\omega_i$ is the underlying flat measure on lattice random walks. The partition function Z_v is the mass of this measure:

$$Z_v = \int d\mu_{DJ}^{(v)} \tag{2.24}$$

For $v = 0$ the measure collapses onto the flat RW measure, while for $v \to \infty$ only occupation numbers equal to 0 or 1 contribute to the energy, so that one gets the usual SAW measure.

$$\begin{array}{rcl} d\mu_{DJ}^{(v\to\infty)}[\omega] & = & d\mu_{SAW}[\omega] \\ d\mu_{DJ}^{(v=0)}[\omega] & = & \prod_i d\omega_i \end{array} \tag{2.25}$$

Definition (2.23) can be rewritten by counting self-intersections along the walk itself — instead of counting them on the whole lattice — and by making the dependence on N explicit:

$$d\mu_{DJ}^{(v,N)}[\omega] = \left(\prod_{i=1}^{N} d\omega_i\right) \exp\left[-v \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \delta_{\omega_i,\omega_j}\right] \tag{2.26}$$

[4]See chapter 4 for an exact result on ν that is relevant to the model studied here.

Another equivalent form is obtained by remembering that $\delta_{ij}^n = \delta ij$ for every n and resumming the exponential:

$$d\mu_{DJ}^{(N)}[\omega] = \left(\prod_i d\omega_i\right) \prod_{i\neq j} \left[1 + \tilde{v}\delta_{\omega_i,\omega_j}\right] \tag{2.27}$$

where the new parameter is related to the old one through the following relation

$$\tilde{v} = e^{-v} - 1 \tag{2.28}$$

Both these latter forms are found in the literature[5]. We will be employing definition (2.26), where the parameter v will be called the *excluded volume*, *interaction strength*, or *coupling* of the theory, and N will be called *number of bonds*, or *walk length*.

By comparing the foregoing description with the notions explained in the previous two sections, the Domb–Joyce model appears then to describe (universal properties of) monodisperse ensembles of linear flexible homopolymers in good solutions at high dilution.

Sometimes a generalization of this model has been proposed and studied [13], where the weight of a self-intersection depends on the length of the loop the intersection produces. Such a model is called *forgetful weakly self-avoiding walk* (FWSAW) — because of its memory-loss property.

When it comes to describing polymers near or below the theta temperature, all the foregoing models fail, because they are completely neglecting the attractive interactions between monomers that are distant along the chain. A model which is refined in this direction is an improved version of the SAW, to which attractive interactions are added. Such an ensemble is defined as the set of all SAWs with Hamiltonian proportional to the number of nearest-neighbor (on the lattice) contacts. Such an interaction is called *bridge interaction* and is a model of the short range attractive forces between the monomers.

2.3.3 Two-Parameter Theory

The so-called *two parameter theory* is a continuum theory of polymer solutions above the theta temperature. It relies on the *Edwards model* [14], which is a continuum measure on paths ω with the same characteristics of the Domb–Joyce model — Gaussian backbone, point-wise range in space,

[5]Sometimes the sum at the exponent is restricted to $i < j$ so that the parameter v takes an additional $1/2$ factor.

long range on proper time. It is described by the partition function over walks $\omega(\tau)$ with $\tau \in [0, L]$

$$Z(L, v) = \int \mathscr{D}[\omega] \exp\left[-\frac{1}{2}\int_0^L d\tau\, \dot{\omega}^2(\tau) - \frac{v}{2}\iint_{\tau_1>\tau_2} d\tau_1 d\tau_2\, \delta^d\left(\omega(\tau_1) - \omega(\tau_2)\right)\right]$$

The two parameters of the theory are L and v, the former being the chain-length and the latter being the excluded volume parameter. The first integral in the exponential is the well-known geodesic weight that appears in the Wiener measure on random paths.

This is the modern version of the two-parameter theory, which was first formulated by Flory [15] in a different form. The original method of Flory approximately views the polymer as a spherical cloud of segments whose density decreases according to a (α-dependent) Gaussian distribution in the distance from the center of mass. Such a model is often referred to as the *smoothed-density model*. In the Flory theory, the equilibrium value of the expansion factor α is calculated from the balance between the osmotic force which tends to swell the molecule, and the elastic force arising from the resulting molecular expansion to a less probable configuration. The resulting equation (in three dimensions) for α is called *Flory equation*

$$\alpha^5 - \alpha^3 = 2.6z \tag{2.29}$$

where z is defined as

$$z := \left(\frac{3}{2\pi l^2}\right)^{\frac{3}{2}} v N^{\frac{1}{2}} \tag{2.30}$$

This equation appears to be correct at zeroth order, but is slightly different from more precise theoretical calculations (for example by the Wang-Uhlenbeck method) already at first order in z, where $4/3$ is expected, instead of 2.6.

Many variants of the original Flory approach have been proposed (see for instance chapter VIII in [5]), and they all share the same features of the two-parameter theory.

Chapter 3

Monte Carlo Integration

The name *Monte Carlo* is used to denote a class of techniques and algorithms that are used to perform approximate calculations in cases where one is not able to treat them analytically. The *leit motif* of these methods is the fundamental role played by random numbers. These are used to probe the space in a 'uniform' fashion, in such a way that integrals or infinite series be substituted by finite sums that approximate them. Monte Carlo integration is — in general — a *bad* method, because of its rate of convergence: the statistical errors it produces are almost universally proportional to the inverse square root of the computational budget; this is essentially a consequence of the central limit theorem. Almost all other numerical methods converge at far faster rates in low dimensions, but in high dimensional systems (such as those in statistical mechanics and quantum field theory) Monte Carlo is almost always the best choice. Also, Monte Carlo (also denoted *MC* from now on) methods are easier to use in cases where the domain of integration is irregular. Moreover the accuracy of results may be improved by adding a single point (i.e. a single MC iteration), while for other deterministic methods one usually has to go to higher order rules.

3.1 The Monte Carlo Method

The term 'Monte Carlo' can be traced back to a popular game played in Monaco[16]. Strangely enough, this has little to do with random numbers and luck; it is a street game where children toss pebbles[1] blindfolded (that is, randomly) on a circle inscribed in a square. The ratio between the two areas is $\pi/4$. If — after a lot of tossing — the stones uniformly span the

[1]It is worth noting that incidentally (or maybe not) the word 'pebble' is 'calculus' in Latin.

whole square one is able to compute an estimate of π simply by counting the pebbles inside and outside the circle.

This game perfectly describes the whole philosophy behind a class of MC methods that are called *static*, or *equilibrium* Monte Carlo. In a more abstract way, one wants to compute integrals

$$\int_\Omega d\mu(x)\, f(x) \tag{3.1}$$

of a function f w.r.t. a measure $d\mu$ on a domain Ω. Using static MC means generating a (large) set of *samples* $x \in \Omega$ in such a way that they are

- *uniform* (w.r.t. the measure $d\mu$): this means that for a subset $S \subset \Omega$ the probability of generating a sample in S is proportional to the volume of S, i.e. to $\int_S d\mu$;
- *independent*: that is, the probability of generating a given sample must be independent of the previously generated samples.

Sampling a measure space with an algorithm that follows these two properties is called *direct sampling*.

As often as unfortunately, there are cases in which one is not able to direct sample the configuration space. For instance, one of these cases is represented by the famous coins-in-a-shoe-box problem, treated in the original paper by Metropolis *et al.*. Here, the problem is to sample the distribution of a fixed number of coins packed inside a box, so that they do not overlap. A naive simple sampling algorithm would be to choose randomly the position of the first coin, and then to insert the other coins with a uniform distribution in the allowed (non-overlapping) region. Unfortunately, this procedure does not yield the correct distribution; in particular, the maximum density of random sequential deposition is much smaller than the close packing density of the coins.

For problems for which a static MC is unfeasible, one has to change strategy and drop at least one of the two requirements above. *Dynamic* MC does not produce independent samples anymore. Now successive MC iterations give correlated samples: this means that the points in the configuration space are not randomly dispersed anymore, but follow a 'path'. But this dynamics is carefully chosen in such a way that the walk ends up covering 'uniformly' the space of configurations. The drawback of this approach is that one has to be more careful when analyzing the data obtained from such a sampling scheme, because of correlation (the dependence of successive values) and relaxation (the initial 'time' it takes for the path to cover 'uniformly enough'

the space). [For a more precise discussion of this aspects see sections 3.2 and 3.2.3.]

In the pebble game metaphor dynamic MC corresponds to throwing a pebble and then moving to the place it lands, before throwing another one. This procedure brings up an important question: what should be done when a stone is thrown outside the square. A correct answer is given by the *Metropolis–Rosenbluth–Teller* algorithm [see section 3.1.2], which states that when a pebble has fallen out of reach the player should remain where she is and pile a new pebble on top of the old one, before continuing.

To state this more precisely, let Σ denote the configuration (or *state*) space, and let the integral be expressed in terms of a probability density π. Then the idea of dynamic MC methods is to invent a *stochastic process* with state space Σ having π as its unique equilibrium distribution. One is then assured that time averages over this process will eventually converge to π-averages, irrespectively of how the process is initialized. It is important to note (with the author in [17]) that this time evolution need not correspond to any *real* dynamics, but is to be chosen on efficiency grounds.

3.1.1 Markov Chains

Stochastic processes for dynamic MC are *Markov chains*[2]. A Markov chain with state space S is a sequence of S-valued random variables $X_0, X_1, \ldots$ such that successive transitions $X_t \to X_{t+1}$ are statistically independent. Obviously the fact that Markov chains have a one-timestep memory[3] does not imply that correlation between distant elements on the chain be null. On the contrary, often several thousand iterations are required for the system to lose memory of initialization.

Once an initial distribution α is chosen, a Markov chain is completely specified by its *transition matrix* (also called *transition kernel*, in general state space). The elements in the transition matrix P represent the probabilities associated with the transitions between elements of the state space: $P = \{p_{xy}\}_{x,y\in S} = \{p(x \to y)\}_{x,y\in S}$. In order to let P be a *probability* transition matrix (or a *stochastic matrix*) it is required that

- $p_{xy} \geq 0$ for all $x, y \in S$
- $\sum_y p_{xy} = 1$ for all $x \in S$

[2]There are lots of references on the theory of Markov chains. Among them the books by Iosifescu[18] and Chung[19] are worth a look; they focus on Markov chains with finite state space and general state space respectively.

[3]Alternatively one says that the *order* of the chain is one.

Now if the chain is initialized so that $\text{Prob}(X_0 = x) = \alpha(x)$ then the stochastic process is specified by the joint probability

$$\text{Prob}(X_0 = x_0,\, X_1 = x_1,\, \ldots) := \alpha(x_0)\, p_{x_0 x_1}\, p_{x_1 x_2} \cdots \tag{3.2}$$

The goal is to construct a chain which will (in some sense to be made more precise) converge — no matter what the initial distribution be — to the probability distribution that MC is intended to simulate. Clearly not all Markov chains have the property that they converge to the same distribution for all initializations. For example the chain with identity transition matrix always converges (in any sense) to the starting distribution. But that of convergence is just the first difficulty. Actually one wants to be able to *construct* a Markov chain *given* the equilibrium distribution, so asymptotic distributions should be asked to have some caracterizing property that is more easily checked. One is then brought to introduce stricter classes of chains and distributions.

A Markov chain is said to be *irreducible* if for each pair $x, y \in S$ there exists an $n \geq 0$ for which

$$(P^n)_{xy} > 0 \tag{3.3}$$

Stated differently an irreducible chain is one for which every state has a finite probability of being reached starting from any other state (including itself). This definition is suggested from the necessity of having a *unique* equilibrium distribution, because if the state space were divided into mutually 'non-interacting' domains (in the sense specified by failure of equation 3.3), one could end up having different equilibrium probability measures when starting from distributions supported in different domains. The *period of* $x \in S$ (denoted d_x) is defined as the greatest common divisor of the numbers $n > 0$ for which $(P^n)_{xx} > 0$. It can be shown that for an irreducible chin all states have the same period, so one can speak of the *period of the chain* d. Moreover the state space can be partitioned into subsets $S_1, S_2 \ldots, S_d$ around which the chain moves cyclically, i.e. $(P^n)_{xy} > 0$ for every couple $x \in S_i$ $y \in S_j$ for which $j - i \neq n (\text{mod } d)$. If $d = 1$ the (irreducible) chain is called *aperiodic*. Aperiodicity assures that the chain does not cycle periodically through some subsets of the state space, so that — as stated by the following theorem — convergence to the equilibrium distribution can be intended in a 'strong' sense. Finally, a probability measure $\pi = \{\pi_x\}_{x \in S}$ is called *stationary* if it is a (left) eigenvector of the trasfer matrix P, that is if

$$\sum_x \pi_x (P)_{xy} = \pi_y \quad \text{for all } \; y \in S \tag{3.4}$$

With these definitions a theorem about convergence can be proved:

Theorem. *Let P be the transition probability matrix of an irreducible Markov chain of period d. If a stationary probability measure exists, then it is unique and*

$$\lim_{n\to\infty} \left(P^{nd+r}\right)_{xy} = \begin{cases} d\pi_y & \text{if } x \in S_i,\ y \in S_j,\ j - i = r(\mathrm{mod}\ d) \\ 0 & \text{if } x \in S_i,\ y \in S_j,\ j - i \neq r(\mathrm{mod}\ d) \end{cases} \tag{3.5}$$

In particular for an aperiodic chain $d = 1$, so that

$$\lim_{n\to\infty} \left(P^{n}\right)_{xy} = \pi_y \tag{3.6}$$

In the general case a stationary probability distribution need not exist, but if it does, then the theorem assures that it is the distribution reached in the long run, irrespectively of the initial distribution α.

Other theorems can be proved under the conditions of this theorem, such as a central limit theorem and a law of the iterated logarithm (for statements and proofs see [19]).

3.1.2 Dynamic Monte Carlo and the Metropolis Method

It is now clear how to set up a MC dynamics to calculate expectation values w.r.t. a given probability measure π. It suffices to invent a Markov process with a transfer matrix P such that

- P be *irreducible*
- π be *stationary* for P.

Then the theorem in section 3.1.1 implies that expectation values calculated along the process (*time* averages) converge — in the 'weak' sense specified by that theorem — to expectation values w.r.t. the measure π (*equilibrium* averages)[4].

A convenient way of satisfying the condition of stationarity without explicitly checking it is by requiring a stricter condition called *detailed balance*. While stationarity means that the net 'current' coming out of an arbitrary state x is null (so it is a rather global condition, because it concerns all other

[4]Notice that this equivalence between time-averages and π-averages is resemblant of a fundamental aspect in *ergodicity* theory. Actually, the term *ergodic* is found in the literature about Markov chains, where it denotes a chain which is irreducible and (in general non-finite state space) also aperiodic and positive Harris recurrent (for a definition see [20]). Elsewhere (for instance in [21]) the term is used to denote an irreducible chain for which a stationary distribution exists.

states), detailed balance requires that the net current between any *couple* of states be null, i.e.

$$\pi_x(P)_{xy} = \pi_y(P)_{yx} \quad \text{for each } x, y \in S \tag{3.7}$$

Detailed balance is then a more local condition. It is easily seen that equation 3.7 implies global stationarity: summing over x one obtains the product πP in the left hand side, and π in the right hand side, since summing over an index of $(P)_{xy}$ gives unity (P is a *probability* transition matrix). A Markov chain satisfying detailed balance is called *reversible*. It is worth noting that if P_1, P_2, ..., P_n are transition probabilities satisfying stationarity for π, then so does any convex combination (that is, any matrix of the form $\sum_{i=1}^{n} \lambda_i P_i$ with $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$). The same holds for the product $P = P_1 P_2 \cdots P_n$ [5].

Now, given a probability measure π on the state space S the goal is to construct a transition matrix P satisfying detailed balance. The fact is, that usually one already has some sort of idea about how the MC algorithm should work, i.e. one already knows some features of the process, such as locality or non-locality [see section 3.2.3], symmetries, and even the kind of atomic moves the process should be driven by. So, in order to keep these rough features in the final algorithm one is concerned about finding a way of minimally modifing an existing probability transition matrix so that it satisfies detailed balance with respect to π.

A general method of performing this task was first introduced by Metropolis *et al.*[22] and was later extended and generalized by Hastings[23]. Let P_0 (which will be called *proposal matrix*) be an irreducible probability transition matrix that includes the fundamental features of the algorithm one wants to develop. One could then use this matrix to generate the atomic moves of the process; these are called *proposed moves*. The core of the *Metropolis–Hastings* method is then a way of determining whether a proposed move is to be accepted or rejected. A most important point here is that in the case of a rejected move the process should not simply wait for another move to be processed, because following this way the rejected move would bring no information at all to the chain. On the contrary, if a proposed move is rejected, then a *null transition* $x \to x$ is made, so that 'memory' of the rejection is kept in the dynamics. In the pebble game metaphor [see the introduction to

[5] These correspondences are useful when constructing new MC algorithms, because one can build a new dynamics using other known dynamics — for which stationarity is demonstrated — as bricks. In fact, the convex combination means choosing at each step with probability λ_i a move among the atomic moves specified by P_i, while the product amounts to perform the moves sequentially.

section 3.1] this rule equates to pile a new pebble on top of the old one at x when the new pebble has been tossed out of the game square.

If the probability of accepting a proposed move $x \to y$ is denoted by α_{xy}, the transition matrix of the modified algorithm is then

$$\begin{array}{rcll} (P)_{xx} & = & (P_0)_{xx} + \sum_{y \neq x} (P_0)_{xy} (1 - \alpha_{xy}) & \\ (P)_{xy} & = & (P_0)_{xy} \alpha_{xy} & \text{for } x \neq y \end{array} \tag{3.8}$$

Now for P to satisfy the detailed balance condition w.r.t. π, the following equation for α must be satisfied for all pairs $x \neq y$:

$$\frac{\alpha_{xy}}{\alpha_{yx}} = \frac{(P_0)_{yx} \pi_y}{(P_0)_{xy} \pi_x} \tag{3.9}$$

Setting

$$\alpha_{xy} = f\left(\frac{(P_0)_{yx} \pi_y}{(P_0)_{xy} \pi_x} \right) \tag{3.10}$$

where $f : [0, +\infty] \to [0, 1]$, one sees that solutions to equation 3.9 correspond to those functions f which satisfy

$$\frac{f(z)}{f(1/z)} = z \quad \text{for all } z \tag{3.11}$$

The original procedure proposed by Metropolis *et al.* corresponds to the choice

$$f(z) = \min(z, 1) \tag{3.12}$$

but other choices are possible, for instance

$$f(z) = \frac{z}{1 + z} \tag{3.13}$$

which is sometimes used. The Metropolis choice 3.12 is the *maximal* function satisfying equation 3.11, meaning that for all other possible choices of f the following inequality must be satisfied[6].

$$f(z) \leq \min(z, 1) \tag{3.14}$$

Also notice that with the Metropolis choice a transition kernel already satisfying detailed balance is not modified at all.

[6] In fact, $f(z) \leq 1$ because it represents a probability. For $z > 1$ this suffices to conclude that $f(z) \leq \min(z, 1)$. On the other hand, if there exists a $z^* < 1$ for which $f(z^*) > z^*$, then equation 3.11 implies $f(1/z^*) > 1$, absurd.

The special case in which the transition matrix is symmetric and the target distribution is of the Gibbs form is worth analyzing directly. In this case one has

$$\alpha_{xy} = f\left(\frac{\pi_y}{\pi_x}\right) \tag{3.15}$$

with a distribution of the form

$$\pi_x = \frac{1}{Z} e^{-\beta E_x} \tag{3.16}$$

so that

$$\alpha_{xy} = f\left(e^{-\beta(E_y - E_x)}\right) \tag{3.17}$$

The most important aspect here is that the partition function Z has disappeared from this expression: this is crucial, because in general one is not able to compute it exactly. Using the Metropolis acceptance probability one finally obtains the following update rules:

- If $\Delta E \leq 0$, then just accept the proposal.
- If $\Delta E > 0$, then accept the proposal with probability $e^{-\beta \Delta E}$.

The second case is treated with the help of pseudo-random number generators, as in the prior procedure of generating a (random) proposed move. One generates a random number uniformly distributed in $[0, 1]$ and then accept the proposal if this number is $\leq e^{-\beta \Delta E}$. This scheme for actually generating the MC dynamics is the one we followed for the simulations in this work.

3.2 Statistical Analysis of Data

A most important aspect of Monte Carlo studies is the analysis of obtained data. One usually gets long series of values for every observable, actually one value for each MC iteration. The number of iterations achieved depends on many factors — such as CPU power, time available, complexity of the actual algorithm used — but it often gets as high as several million cycles. This means several million values to analyze for *each* observable and for *each* choice of the model parameters (if it has any). It is then clear that analysis of data is a computationally expensive task, and its cost in terms of CPU time — though not as high as that of actual MC computation — must not be underestimated. Moreover analysis of data constitutes the 'interface' between the model being simulated and the scientist, because it provides output in human readable form. So it must be treated with special care,

not to completely misunderstand the behavior of the system that is being simulated, and not to misjudge properties of the MC dynamics that is used.

MC work shares many of the features of ordinary experimental work. As noted long ago in [24], this requires the inclusion in MC papers of estimates of statistical error, descriptions of experimental conditions (i.e. parameters used in the calculation), details about the experimental apparatus (i.e. about the algorithms and the program), careful discussion of systematic errors, and so on. Berretti and Sokal in [25] emphasize two points in particular:

- The importance of a correct statistical treatment of *autocorrelations* and their effects, including a theoretical and an empirical study, leading to valid error bars.

- The distinction between *systematic errors*, due to misspecification or approximation of the mathematical model on wich the analysis is based (these errors often result from unincluded corrections to scaling or from not taking finite-size effects into account), and *statistical errors*, due to the unavoidable random fluctuation inherent in any probabilistic experiment.

The first point will be treated in this section.

3.2.1 Autocorrelation

In *dynamic* MC successive samples X_t, X_{t+1} are correlated, perhaps even very strongly. So *autocorrelation* in obtained data arises. This phenomenon may cause the variance of estimates produced from the simulation to be much higher than in static Monte Carlo, where successive samples are independent. So a rigorous analysis of autocorrelation is needed to produce valid error bars. Moreover the Markov chain is not in general initialized with the stationary distribution (actually, since the simulation starts from one *sample* x taken from the initial distribution, this should be thought as always being a delta distribution $\delta(x)$). Thus the dynamics must be allowed some time in order to reach the stationary regime; this is again an effect of autocorrelation. So also a treatment of *thermalization* (that is, convergence to the stationary distribution) is needed to correctly analyze the data.

Consider the *stationary* Markov chain. This means observing the system after a time long enough for it to have reached equilibrium (or else it suffices to start the chain with the stationary — and unique — distribution π). Let $f : S \to \mathbb{R}$ be a real-valued function defined on the state space S; it represents a real-valued *observable*. $\{f_t\} = \{f(X_t)\}$ is a (stationary) stochastic process.

Its *mean* is

$$\mu_f := \langle f_t \rangle = \sum_{x \in S} \pi_x f(x) \tag{3.18}$$

and its *unnormalized autocorrelation function* — sometimes also called *autocovariance* — is

$$C_{ff}(t) := \langle f_s f_{s+t} \rangle - \mu_f^2 = \sum_{x,y \in S} f(x) \pi_x \left(\left(P^{|t|} \right)_{xy} - \pi_y \right) f(y) \tag{3.19}$$

Note that even if the above definition contains the Monte Carlo *times* s and $s+t$ the autocorrelation function actually depends, by stationarity, only on their difference t. Finally, the *normalized autocorrelation function* is defined as

$$\rho_{ff}(t) := \frac{C_{ff}(t)}{C_{ff}(0)} \tag{3.20}$$

Typically one expects the autocorrelation time to decay exponentially for large t, so one is led to define — as in the case of *spatial* equilibrium correlation functions — the *exponential autocorrelation time* as

$$\tau_{exp}[f] := \limsup_{t \to \infty} \left(-\frac{t}{\log |\rho_{ff}(t)|} \right) \tag{3.21}$$

This definition highlights the fact that autocorrelation could manifest itself at different time scales for different observables. But one is interested in the relaxation of the Markov chain itself, which constitutes the dynamics behind *every* observable that is calculated. So it is natural to slightly change this definition into the following

$$\tau_{exp} = \sup_f \tau_{exp}[f] \tag{3.22}$$

Thus, τ_{exp} represents the relaxation time of the *slowest* mode in the system. Yet, this latter definition — though being important from a theoretical point of view — is modified in actual MC work, because calculation of the sup in (3.22) is unfeasible. In fact one only considers the observables one has collected data for; but clearly the actual autocorrelation of the chain (in the sense of equation (3.22)) is in general different from that calculated this way.

Another way of looking at this issue is by considering the transition probability matrix P as an operator on the Hilbert space $l^2(\pi)$. Then it is possible to derive a more mathematically precise result about the deviation from equilibrium (see [17] for details). Indeed the distance from equilibrium (in the l^2 sense) is found to satisfy the following inequality asymptotically as $t \to \infty$:

$$d_2 \left(\alpha P^t, \pi \right) \leq d_2(\alpha, \pi) \, e^{-\frac{t}{\tau_{exp}}} \tag{3.23}$$

where α is the initial distribution, and $d_2(\nu, \pi)$ is the l^2-distance from ν to the equilibrium distribution. Notice that a sup over all observables is hidden here, too. It appears in the expression for the distance d_2.

So, the exponential autocorrelation time defined above focuses on the exponential decay of the autocorrelation function for large times. On the other hand, for a *given* observable f another τ can be defined. It is called the *integrated autocorrelation time*

$$\tau_{int}[f] := \frac{1}{2} \sum_{t=-\infty}^{\infty} \rho_{ff}(t) \tag{3.24}$$

So this new time-like quantity measures the magnitude of the 'overall' autocorrelation[7]. One then expects it to be a measure of the statistical errors in Monte Carlo measurements of averages $\langle f \rangle$, because the variance of sample means is expected to be more and more important as the autocorrelation of successive samples increases.

A very general bound on the autocorrelation times has been established in [26] for algorithms of the Metropolis type. It is proved that for a suitable observable H the autocorrelation function at lag 1 (from which it is possible to obtain bounds on the autocorrelation times) is bounded below by an expression involving var(H).

3.2.2 Estimators

All characteristic times defined this way share the feature of containing limits for $t \to \infty$. When analizing data from a MC simulation, one never has infinitely many values for each observable. Thus, a feasible way of estimating the autocorrelation times — together with valid statistical error bars for averages of observables — is needed.

There is a branch of statistics which is of great help in the task of finding and studying estimators for the quantities one is interested in; it is called *time series analysis*. Systematic expositions can be found for instance in the books by Priestley [27] and Anderson [28].

Let us focus on a given observable f in the stationary regime (i.e. let us start the chain in the stationary distribution, or equivalently let it evolve for a long time). Then $\{f_t\}$ will be a real-valued stationary stochastic process with mean $\mu = \langle f_t \rangle$ and unnormalized autocorrelation function

[7] Notice that in general $\tau_{exp} \neq \tau_{int}$, and this is an important issue to remember when analizing Monte Carlo data. With the previous definitions the two characteristic times coincide in the special case where $\rho_{ff}(t)$ behaves as $e^{-\frac{|t|}{\tau}}$. This also explains the factor of 1/2 inserted in definition (3.24).

$C(t) = \langle f_s f_{s+t} \rangle - \mu^2$. The goal is to obtain valid estimates of μ, $C(t)$, $\rho(t)$, τ_{int} and τ_{exp} from a finite — but large enough — set of samples $f_1, \ldots, f_n$. The natural estimator of μ is the sample mean

$$\overline{f} := \frac{1}{n} \sum_{i=1}^{n} f_i \tag{3.25}$$

that is, the arithmetic mean of all the values one has collected for the observable. This estimator is *unbiased*. This means that its average is equal to the quantity it is an estimate of. In this case one can easily check that indeed $\langle \overline{f} \rangle = \mu$. Its variance — as calculated from the definition — is

$$\begin{aligned} \mathrm{var}\left(\overline{f}\right) &= \frac{1}{n^2} \sum_{r,s=1}^{n} C(r-s) && (3.26) \\ &= \frac{1}{n} \sum_{t=-(n-1)}^{n-1} \left(1 - \frac{|t|}{n}\right) C(t) && (3.27) \\ &\approx \frac{1}{n} \left(2\tau_{int}\right) C(0) && (3.28) \end{aligned}$$

The latter relation is an approximate equality valid in the regime $n \gg \tau$[8]. Equation (3.28) gives a precise meaning to the integrated autocorrelation time. The variance of an observable is a factor $2\tau_{int}$ larger than it would be if the samples were statistically independent (for instance in static MC). One can see here how static and dynamic quantities mix together in calculating estimates. Even if one is interested only in the static quantity μ, it is necessary to compute a dynamic quantity (viz. τ_{int}), in order to determine the error bar. This is due to the peculiar way one tries to observe properties of a *static* model — such as one in equilibrium statistical mechanics — by simulating a *dynamic* process which approximates it.

It is then clear that in order to get an estimate of the statistical error on μ an estimator for $C(t)$ has to be found. Here (as clearly explained in [25]) two different situations can occur, depending on whether the mean μ is known or unknown. In the former case the natural estimator of $C(t)$ is

$$\widetilde{C}(t) := \frac{1}{n - |t|} \sum_{i=1}^{n-|t|} \left(f_i - \mu\right) \left(f_{i+|t|} - \mu\right) \tag{3.29}$$

[8] The requirement that n be much greater than τ is important in the whole discussion about MC analysis, not only because of the intuitive meaning it gives to equation (3.28). [See the discussion about self-consistency in section 3.2.3].

while in the latter case one has to use the estimator of μ defined above, instead of μ:

$$\widehat{C}(t) := \frac{1}{n-|t|} \sum_{i=1}^{n-|t|} \left(f_i - \overline{f}\right)\left(f_{i+|t|} - \overline{f}\right) \tag{3.30}$$

The latter is almost always the case, because one usually wants to compute both the mean μ and its error from a MC simulation. $\widetilde{C}(t)$ is unbiased, that is $\langle \tilde{C} \rangle(t) = C(t)$. On the other hand, the samples f_i in the definition of $\widehat{C}(t)$ are entangled in such a way that it is not unbiased[9]; anyway the bias is of order $1/n$. Notice that as t increases these estimates become rougher, because the number of couples $(i, i+t)$ available in the sampled data decrease[10].

Likewise one defines two estimators for the normalized autocorrelation function $\rho(t)$ as

$$\widetilde{\rho}(t) := \frac{\widetilde{C}(t)}{\widetilde{C}(0)} \tag{3.31}$$

and

$$\widehat{\rho}(t) := \frac{\widehat{C}(t)}{\widehat{C}(0)} \tag{3.32}$$

Formulas for the variances and covariances of $\widetilde{C}(t)$, $\widehat{C}(t)$, $\widetilde{\rho}(t)$ and $\widehat{\rho}(t)$ can be computed (see for instance [28]).

All the natural estimators described so far turned out to be good estimators, in the sense that their variances vanish as the sample size n goes to infinity, so the MC worker may get an arbitrarily precise estimate of these quantities by adjusting the run-length n. Unluckily, this happy property does not hold anymore when one turns to consider other estimators, such as that for the integral autocorrelation time. In fact, a naive estimator for τ_{int} is directly constructed from the definition, substituting the autocorrelation function with its estimator, and truncating the sum

$$\bar{\tau}_{int} := \frac{1}{2} \sum_{t=-(n-1)}^{n-1} \widehat{\rho}(t) \tag{3.33}$$

but this estimator turns out to have a variance that approaches a nonzero constant as n goes to infinity. This happens because the sample autocorrelations $\widehat{\rho}(t)$ carry most of the information in the region where t is less then or of order τ. For $t \gg \tau$ the information is 'hidden' inside the noise, and the

[9] This is essentially due to the fact that $\left\langle \overline{f}^2 \right\rangle \neq \mu^2$.

[10] Again, one needs a number of samples much greater than tipical decay times involved in the objects that have to be computed, in order to have valid estimates. See footnote 8.

noise happens to sum up to a total variance of order 1. The solution is to give more importance to the region $t \lesssim \tau$ in the definition of the estimator for τ_{int}. This is achieved by insertion of a *window function* $\lambda(t)$ which is of order unity for $|t| \lesssim \tau$ and ≈ 0 for $|t| \gg \tau$:

$$\widehat{\tau}_{int} := \frac{1}{2} \sum_{t=-(n-1)}^{n-1} \lambda(t)\widehat{\rho}(t) \tag{3.34}$$

(the definition for $\widetilde{\tau}_{int}$ — in the case where the mean μ is known — is the same with $\widetilde{\rho}(t)$). A good and natural choiche for $\lambda(t)$ is the rectangular window

$$\lambda = \chi\,[-M, M] \tag{3.35}$$

where χ is the characteristic function of an interval, and M is a suitably chosen cutoff. This new estimator is biased, and the bias does not go to 0 as $n \to \infty$, but it does as the cutoff M goes to infinity:

$$\mathrm{bias}(\widehat{\tau}_{int}) \equiv \langle \widehat{\tau}_{int} \rangle - \tau_{int} = -\frac{1}{2} \sum_{|t|>M} \rho(t) + o\left(\frac{1}{n}\right) \tag{3.36}$$

On the other hand, the variance is seen to be — under some approximation

$$\mathrm{var}(\widehat{\tau}_{int}) \approx \frac{2(2M+1)}{n} \tau_{int}^2 \tag{3.37}$$

so that for M fixed it can be made arbitrarily small for n large enough. One then understands that the choice of M is a tradeoff between bias and variance. Indeed, one can make the bias small by taking M large enough so that the sum in (3.36) includes a negligible range of $\rho(t)$. On the other hand one also wants to keep M small, in order to have a small variance at fixed n, or — equivalently — in order to obtain a fixed variance from a shorter MC run.

Again, some self-consistency is required, since the choice of M is closely related to the magnitude of the typical decay time τ_{int}, but this quantity is estimated using M itself as a cutoff. The following *automatic windowing* [29, 17] procedure seems to be a most convenient one in assuring self-consistency. Choose M as the smallest integer such that $M \geq c\widehat{\tau}_{int}(M)$ for some *windowing factor* c. The choice of the factor is a very euristic one. If $\rho(t)$ follows a pure exponential decay then $c \approx 4$ should be enough, because $e^{-4} < 2\%$ and therefore the bias would not change the significance of the estimated integral autocorrelation time. In a more realistic situation the decay of the autocorrelation function can be slower than exponential (in the pre-asymptotic and/or asymptotic region), so a more careful choice is to take the windowing factor between, say, 6 and 10.

3.2.3 Relaxation Problems

Summing up, convergence to equilibrium and autocorrelation at equilibrium are very important subjects in Monte Carlo work. The questions arised by these topics are answered by the tools and techniques described in the previous two sections.

First, there is the problem of initialization bias. The Markov chain has to be started in some configuration x_i. For example, in simulating an Ising model x_i could be that with all spins pointing in one direction; this is called a *cold start*. Another choice is a configuration having spin directions chosen at random from some simple probability distribution, for instance the flat distribution; this is called a *hot start*. A simulation usually works 'pointwise' in state space, meaning that it deals with *single* configurations $x \in S$, and not with *ensembles* $\{x\}$. This means that the initialization distribution of a MC run will always be a delta distribution $\delta(x_i)$, that is, the chain is initialized in the state x_i. However, if one considers a *single* MC run, some initialization bias will always be present, since the initial distribution will always differ from the equilibrium distribution.

In order to treat this aspect more quantitatively, one introduces the exponential autocorrelation time τ_{exp}, which serves as a measure of the rate of convergence to equilibrium. Actually, the formal definition of τ_{exp} (eq. (3.22) provides an *upper bound* on the amount of time one has to wait before equilibrium is attained 'for all practical purposes'. But one is never able to compute an estimate of this quantity, because it implies a sup over all possible observables. Is is usually impossible to know τ_{exp} (or an upper bound for it) even theoretically. For these reasons one computes an estimate of the exponential autocorrelation time only for the observables one is interested in and has collected data for, and the value of τ_{exp} is taken as the maximum among these values and multiplied by some factor k to get the discard time t_d. Then, the first t_d data in the run are discarded, and the means and variances of the observables are computed only from the data left. The choice of k depends on the degree of accuracy one wants to achieve, but usually a value around 20 is more than enough, for in this case the deviation from equilibrium will be at most e^{-20} ($\approx 10^{-9}$) times the initial deviation from equilibrium (in the sense specified by equation (3.23)).

Notice that the problem of initialization bias is not a most important one, because the bias on the sample means goes to zero as $1/n$, while the statistical errors are of order $1/n^{\frac{1}{2}}$. However, it is also true that in practice the coefficient of $1/n$ (which of course depends on τ_{exp} too) may be large, depending on how far (in the l^2 sense) the initial delta distribution is from equilibrium. Rejecting the data from the initial transient is not very expensive, and grants

the avoidance of a potentially large systematic error. Moreover — assuming that $\tau_{int} \approx \tau_{exp}$[11] — in order to practically forget about initialization it suffices to throw away $\sim 20\tau$ of the data, while the run-legths needed to obtain reasonably small statistical errors are $\sim 1000\tau$ (for $\sim 1\%$ accuracy), so that the loss due to rejection is totally negligible. From this discussion one can also get an idea of how uncritical the choice of the factor k (and then of t_d) is.

Autocorrelation in equilibrium, of which τ_{int} is a measure, controls the variance of the sample means in a dynamic MC simulation. As stated by equation (3.28) the variance on the sample mean $\bar{f}$ is a factor $2\tau_{int}[f]$ higher than it would be in independent sampling; roughly speaking, a run of length n only contains $n/2\tau_{int}[f]$ effectively independent (non-correlated) points as far as the observable f is concerned. The integrated autocorrelation time therefore represents a close measure of the computational efficiency of a MC algorithm. Nonetheless, when comparing different algorithms another important aspect must be taken into account, that is the computational complexity *per iteration*, or the time it takes to complete a single MC step. This is because it is time in the definition of τ_{int} is measured in units of iterations, but every iteration could take a lot of machine time to complete. Actually, in the most common scenario one has to find a tradeoff between good autocorrelation times and speed of the algorithm.

One of the most important things to keep in mind when estimating τ_{int} from a MC run is *self-consistence*. The estimate of τ_{int} depends on the number of iterations n, but this number is to be fixed upon knowledge of τ_{int} itself. So one must require that the run length n be much greater than the estimates of τ_{int} obtained from the same run. Unfortunately self-consistence is only a necessary condition for the trustworthiness of MC data: in general it is not sufficient, because the analyzed data could be poisoned by *metastability*.

Metastability is one of the wildest and less controllable problems that may arise in MC simulation. It is a danger common to all empirical and numerical methods of determining when equilibrium has been achieved, for example in optimization algorithms, and in non-linear fitting methods. Metastability means that equilibrium only *appears* to be settled, while the algorithm is stuck in a metastable region of the state space. Metastable regions are those regions of the state space in which the dynamics is 'trapped' and spends a lot of time (iterations) to come out. One way to try and reduce the consequences of this problem is to exploit the arbitrariness of the initial configuration, ob-

[11] It is important to note that this is not always the case, since the two time-scales need not be equal; however this is a working assumption used to roughly compare the importance of thermalization versus correlations at equilibrium.

serving how the results change as the initial state is changed, and maybe trying to avoid the metastable region. For example, near a first-order phase transition most Monte Carlo methods present metastability, with metastable regions associated with the distinct pure phases. One then starts the dynamics with a hot and a cold start, and then tests consistency between results. Even if this procedure does not guarantee that metastability is absent, it does increase the confidence.

Yet, the worst problems arise when the autocorrelation times diverge. This is a universal situation, studied in the theory of critical dynamics, and is called *critical slowing down*. It happens that near a critical point (that is, when the equilibrium measure is that of a model which is near a critical point) the (exponential or integrated) autocorrelation time τ diverges, typically as

$$\tau \sim \min(L, \xi)^z \tag{3.38}$$

Here L is the lattice size, ξ is the correlation length (in an infinite-volume system with the same parameters), and z is a *dynamic critical exponent*. This kind of behavior is typical of second-order phase transitions. Near a *first*-order phase transition, things get even worse, with τ usually growing exponentially fast with a power of L. This is essentially because the algorithm needs to tunnel through very improbable configurations involving interfaces, in order to move from one phase to another, so it takes a long time to get really independent configurations.

Since there are two different definitions of a characteristic time for the dynamics, there could be two different dynamic critical exponents z_{exp} and z_{int}. This is in fact the case, and in general $z_{exp} \neq z_{int}$, once again highlighting the fact that the two autocorrelation times actually correspond to different aspects of the dynamics. In fact, near a critical point one expects the autocorrelation function to obey a dynamic scaling law [30] of the form

$$\rho_{ff}(t; \beta) \sim |t|^{-\alpha} F\left((\beta - \beta_c)\,|t|^b\right) \tag{3.39}$$

for $|t| \gg 1$, $\beta \approx \beta_c$, with $|\beta - \beta_c||t|^b$ bounded, where a and b are dynamic critical exponents, β is a relevant parameter and β_c is its value at criticality. If $F(x)$ happens to have an exponential decay for large $|x|$, then letting $|t|$ scale as $(\beta - \beta_c)^{-1/b}$ in order to keep x bounded, one gets

$$\tau_{exp}[f] \sim |\beta - \beta_c|^{-\frac{1}{b}} \tag{3.40}$$

$$\tau_{int}[f] \sim |\beta - \beta_c|^{-\frac{1-a}{b}} \tag{3.41}$$

so it is clear that in general (unless $a = 0$) $z_{exp} \neq z_{int}$.

It is important — when designing new MC algorithms or when improving existing ones — to understand the underlying physical reasons of critical slowing down, in order to devise ways of making the exponent z smaller. One property of conventional algorithms that is most responsible of critical slowing down is *locality*. A local dynamic MC algorithm is one in which the atomic moves change the configuration only locally (examples are the single spin flip in an Ising model, or the one-bead flip in a SAW). Roughly speaking one could say that a local algorithm changes only a little part of the configuration at a time, so that 'information' travels only at a finite speed, spreading from a site only to its neighbors at each MC step. One might guess — very crudely indeed — that information executes a random walk, so that after t^2 steps it has traveled on average a distance t. One can also guess — now less crudely — that information is to be allowed a journey of a (static) correlation length, for the system to have reached an essentially new configuration. But the correlation length diverges at the critical point (in the thermodynamic limit), so that one expects the correlation time to diverge roughly as ξ^2, that is, the dynamic critical exponent z equals 2.

From this picture one sees that for a local algorithm the slow modes are the long-wavelength modes. The natural solution to critical slowing down is then to take care of these modes by some kind of global updating[12].

Unfortunately, working with collective-mode algorithms is not easy as it might seem. While they help — once the proper modes are identified and exploited to generate non-local moves — in reducing critical slowing down, on the other hands they usually require much more computational time per cycle. So one needs to keep an eye on computational efficiency, too. The bad news is, that usually the computational cost outweighs the advantages[13].

It is then clear that implementing these ideas on the construction of new algorithms is a highly model-dependent task. Nonetheless, some trends in the struggle for 'globalization' may be categorized:

- Multi-grid Monte Carlo [31]

[12]Non-local moves may also be combined with local ones (*hybrid* algorithms), in order to reduce the value of z while still retaining ergodicity. Oftentimes a clever hybrid algorithm is more efficient than either of its 'pure' constituents; these are the cases when the slow modes of the local moves are speeded up by the non-local moves and vice-versa. An extremely happy situation is when a (hybrid) ergodic algorithm can be constructed from non-ergodic algorithms.

[13]For example, in simulating self-avoiding walks the computational weight is given by the self-avoidance check [actually this is not the whole story, since there are other sources of slowing down; see the discussion on non-local moves in section 3.3.1]. When only local moves are employed this check can be made faster, exploiting the very locality of the changes; instead, when a global change is applied one has to check the whole walk again. [But see the discussion about the Pivot algorithm in sections 3.3.2 and 3.3.3]

- Fourier acceleration [32]
- Swendsen–Wang type algorithms [33]

See for instance [17, 34] and references therein for a description of these methods.

3.3 Algorithms for Polymer Measures

The numerical study of the critical properties of self-avoiding walks makes use of essentially two classes of methods. The first is *exact enumeration*. With exact enumeration one is able to collect very fine-grained and precise (actually, exact) data, but the computational weight is high, so one restricts to the simulation of very short SAWs (usually $N \approx 30$). Then this information is used to extrapolate the behavior for large N, with techniques such as Padé or differential approximants. The second class is — of course — Monte Carlo simulation[14]. In MC study, by contrast, one aims to probe directly the regime of fairly long SAWs (up to $N \approx 10^5$). Both these methods require some assumptions about the large-N behavior of the system, in order to correctly extrapolate the quantities of interest in the critical ($N \to \infty$) limit. But since the region probed by MC simulations is much deeper into the critical regime, one can expect that in general *systematic* errors will be less important than they are in results obtained from exact enumerations (which, on the other hand, do not suffer from statistical errors).

SAWs have two big advantages over spin systems, when it comes to MC simulations. Firstly, there are *no* finite-size effects due to the finiteness of the lattice, that is, one can work with walks directly on an infinite lattice[15]. This eliminates the need for finite-size scaling, and the risk of systematic errors caused thereby. Secondly, there is no L^d factor in the computational weight, because the walk is always 'one-dimensional' from a computational point of view. This means one can go closer to criticality, especially for high dimensions.

Since the definition of self-avoiding walks is only geometrical, many different ensembles can be constructed from SAWs. Since there are two parameters, one can freely choose among these the ones to be fixed, that is, the

[14] See [35] (and references therein) for a thorough explanation of the different algorithms (both static and non-static) that have been proposed and used, together with an in-depth analysis of the pros and cons of each algorithm.

[15] Of course this requires some clever programming, and some a priori knowledge of the walk lengths that are to be probed; but in practice it is possible to completely avoid border and finite-size effects.

ones that will not be changed by the MC dynamics. The two parameters are the walk length N and the end-point x[16], so it is natural to define three different ensembles[17]:

- Microcanonical — Fixed N, variable x
- Fixed-endpoint canonical — Variable N, fixed x
- Free-endpoint canonical — Variable N, variable x

MC studies of the self-avoiding walk go back to the early 1950's [36]. The first algorithms used were the most obvious and natural ones: static MC algorithms. The most naive of them is *simple sampling*. In simple sampling one just generates a random walk of length N and checks if it is self-avoiding. If it is, then it is accepted, otherwise one goes back to the origin and starts building a totally new walk. It is immediately clear that such a way of proceeding is not at all efficient (even if one checks for self-avoidance after every step, rejecting the walk as soon as an intersection is found) because for large N the probability of getting a SAW decreases exponentially fast. In fact this probability is given by the ratio between the number of N-step self-avoiding walks and the number of ordinary random walk of length N, that is

$$\frac{c_N}{(2d)^N} \underset{N\to\infty}{\sim} \left(\frac{\mu}{2d}\right)^N N^{\gamma-1} = e^{-\lambda N} N^{\gamma-1} \tag{3.42}$$

where $\lambda = \log(2d/\mu)$ is called *attrition constant*, and is a measure of the exponential decay of the probability of finding a SAW among ordinary random walks. (λ is approximately 0.42 on the square lattice).

A slight improvement that alleviates the problem and lowers the attrition constant a little is achieved by considering only *non-reversal random walks*. This ensemble is constituted by those random walks which never return to a site they visited two steps before. This could be viewed as a *forgetful* SAW, with an extremely short memory (namely, a 2-step memory). In practice, at every step the walk chooses among all neighboring sites but the one it comes from. With this modification, the factor $(2d)^N$ in equation (3.42) is replaced

[16] Actually, when simulating a polymer with excluded-volume interaction, there is another parameter, that regulates the strength of the interaction. This is always a constant of the MC dynamics, and is fixed before each simulation, because one wants to know how properties of the model are related to changes of this parameter. As far as we know, no 'grand canonical' ensemble with a variable excluded-volume has ever been used in MC simulations.

[17] The terminology is not a standard one. Some authors use very different nomenclatures. Essentially, the differences rise from considering monomer or polymer ensembles. The terms used here allow for the use of 'grand canonical' for ensembles of many SAWs.

by $2d(2d-1)^{N-1}$ and the algorithm does work faster, since the attrition constant is now

$$\lambda^{(2)} = \log \frac{2d-1}{\mu} \tag{3.43}$$

(For comparison, on the square lattice $\lambda^{(2)} \approx 0.13$). The next logical step in improving the algorithm within this scheme is to consider forgetful SAWs with r-step memory, that is random walks with only 'long' (viz. with length $> r$) loops [37]. This is implemented by enumerating (before the actual simulation starts) all r-steps SAWs, and building the walk using these as bricks[18]. Now the probability of getting a SAW is

$$\frac{c_N}{(c_r)^{\frac{N}{r}}}$$

At first sight, the exponential behavior seems to have been beaten, so that the attrition constant is zero. But this is not true, because in the regime where r is not $\gg 1$, c_r is still much higher than μ^N, thus leading to a non zero value for $\lambda^{(r)}$. Of course, attrition can in principle be made arbitrarily small by taking r large, but this brings a serious drawback in terms of used memory (for storing all r-step SAWs) and pre-simulation computing time. The trouble is that $c_r^{1/r}$ converges to μ rather slowly.

One can further improve the r-step sampling algorithm by choosing at each step only among those walks which do not go back to the previous walk's initial site; this is called *r-step non-reversal sampling* (for comparison, with 10-step SAWs one gets an attrition constant $\lambda_{NR}^{(10)} \approx 0.071$ on the square lattice).

Another very natural way of generating SAWs is through *restricted sampling*. Since simple sampling algorithms seem to waste a lot of time discarding non-self-avoiding walks, one is tempted to have the algorithm choose at each step only those neighboring sites that are still available. Unfortunately, this method does not generate SAWs with the correct (flat) weights. Rather, a given walk ω is generated with probability

$$P(\omega) = \text{const} \prod_i \frac{1}{k_i(\omega)} \tag{3.44}$$

where k_i is the number of choices available at the i-th step. For instance, a walk that has been forced to follow a particular direction many times

[18] Actually, this implementation does not eliminate *all* k-step loops with $k < r$, because there could still be loops generated by the coincidence of monomers belonging to different sub-chains.

during its path (such as a walk with lots of 'passageways' between two walls) weighs more in this ensemble than a walk that has been more 'free'. So mean values of observable must be estimated using a weighted sum (divided by the sum of the weights) where each weight is given by equation (3.44). Restricted sampling has some hidden disadvantages. Firstly, it does not avoid exponential attrition, despite the fact that self-intersections are now completely avoided. Indeed, the avoidance of self-intersections does eliminate one cause of attrition, but silently introduces another one, that of *trapping*. A trapped SAWs of length N is a walk for which $k_N = 0$, that is, the walk could not be extended by a single step without violating self-avoidance. As one could expect, this phenomenon is much more severe in two dimensions. Secondly, the variance of estimates obtained with this method is usually very high. This is due to the fact that as N gets large the greater part of the data — namely those data whose corresponding weight is negligible — gets practically ignored, so in a sense it is wasted data.

Yet another way of producing independent SAWs is through *dimerization* [38], which is a recursive algorithm. One can build an N-step SAW by concatenating two $N/2$-step SAWs and checking if the resulting walk is self-avoiding. The two constituent walks may themselves be generated in a similar way. The 'seeds' of the recursion, that is the first two (shorter than a fixed cutoff length) SAWs, are to be generated using some other simple algorithm, such as non-reversal simple sampling. This algorithm happens to be very efficient in high dimensions ($d \geq 5$), where generating an N-step SAW this way takes a time that grows almost linearly in N.

A class of non-static methods is that of the so called *quasi-static* algorithms. Here, one drops the strict requirement that the generated walks not be correlated, and generate (non correlated) batches of (correlated) samples. Every static method can be made quasi-static by working with walks at different values of N at the same time, and exploiting the fact that a walk that self-intersects at the n-th step gives an n-step SAW, if cut on the intersection, and m-step SAWs as well (with $m < n$), if cut before the intersection. Other methods are intrinsically quasi-static, such as *enrichment* and *incomplete enumeration*[19]. Enrichment is another recursive algorithm, where if an s-step walk is a SAW, then t copies of it are made, and used as starting points on top of which longer walks are constructed and tested (and then again copied, if they reach lengths $2s$, $3s$, and so on). This algorithm needs clever adjusting of the parameters s and t. Incomplete enumeration is a modification of any exact enumeration algorithm in which the tree is not completely explored. Rather, every branch b is followed only with probability

[19] See [35] for details and references on these methods.

$P(b) < 1$.

3.3.1 Dynamic Methods for SAWs

Here only MC methods for simulating SAWs will be reviewed. It must be stressed that once some attractive interaction is inserted in the model (for describing theta-point behavior and polymers in bad solvents) all known method behave very badly in the critical limit, because near the collapse transition nearly all moves will get rejected, since the walk is no longer sparse enough. On the contrary, most SAW algorithms can be easily implemented for simulations of weakly SAWs and other purely *repulsive* polymer measures.

The atomic moves used by a dynamic MC algorithm for the SAW are naturally categorized by means of three properties, regarding the way the moves change a given walk. According to this categorization, moves can be

- N-conserving (e.g. k-bead, pivot) or N-changing (e.g. kink insertion and deletion)
- endpoint-conserving (all *internal* local moves, both N-changing and N-conserving) or endpoint-changing (e.g. slithering-snake, pivot)
- local (e.g. k-bead), bilocal (e.g. slithering-snake, kink-end reptation) or non-local (e.g. pivot, cut-and-paste)

Here only a brief review of the most important moves will be given. For a more thorough description see [35].

The most general local N-conserving move is the *k-bead* move (with $k \geq 1$). A move of this kind is defined to be one that changes the walk at most on k consecutive steps, that is it changes ω into ω' such that $\omega(i) = \omega'(i)$ for all i except possibly for k consecutive values of i[20]. If one of the changed steps is the end-point or the origin, then the move is said to be *end-group*, otherwise it is called *internal*. Writing all the possible k-bead moves reduces to enumerating all maps between k-step SAWs[21].

A bilocal move is generated by consecutive application of a local (in general N-changing) move in two different parts of the walk. These places may (ad usually will) be far from each other on the chain. Also, usually the two

[20] One also asks that the modified walk be different from the original one at some time i_{min} *and* at time $i_{min} + k$, so that a k-bead move is not also a $(k+l)$-bead move for all l.

[21] This is because it is clearly not clever to use moves that surely introduce self-intersections, because they will be surely rejected. The situation is quite different when considering algorithms for weakly self-avoiding walks, where a self-intersection is not sure to be discarded.

constituent local moves are highly correlated, so that they can be thought of as a single bilocal move. The most important bilocal moves are the *slithering-snake* (a bond is deleted at one end of the walk, and another one is appended at the other end), the *kink-transport* (a kink — that is, a figure formed by three sides of a square — is deleted at one location and inserted at another one), the *kink-end* and *end-kink reptation* (a kink is deleted from the walk and two new bonds are appended at one end, and vice-versa). These moves are all N-conserving. The simplest N-changing moves are kink insertion and deletion (which are internal), and end-bond addition and deletion (which are end-group moves).

As far as non-local moves are concerned, not many of them are known and used. This is because it is much harder to find effective and fast non-local moves. Firstly, they generate new configurations that are 'far' (that is, very different) from the original ones, so they are more likely to get discarded, slowing the dynamics. For instance in a SAW, a thorough change of the walk will be more intersection-prone because in general the move is not exploiting the self-avoidance of the walk it is attempting to change[22]. For an interacting walk, on the other hand, non-local moves generate new configurations with large energy differences and thus they will be rejected often. Secondly, while a local move can be carried out in a (machine) time of order 1[23], a non-local move often requires a time $\sim N$ (or at least $\sim N^p$, with $p > 0$) because the number of sites to be updated at each MC step diverges for $N \to \infty$[24].

Several algorithms can be constructed for different ensembles using a combination of these moves. The first algorithms used for SAWs (in the 1960's) were local and N-conserving [39] (k-bead). These are also very easy to implement and use. So it came as a great surprise the demonstration that *all* such algorithms are in fact non-ergodic [40] in dimensions $d = 2, 3$ for sufficiently large N. The proof is constructive and is based on showing that for each k there exists a class of SAW configurations which are *frozen* with respect to any k-bead local move (here 'frozen' means that any such move generates an intersecting walk).

The critical dynamics of local N-conserving algorithms may be crudely estimated using a heuristic argument. Of course, the autocorrelation times of

[22] Again, the pivot algorithm is an excellent compromise between non-locality and acceptance probability [see the discussion in section 3.3.2].

[23] For instance, when using k-bead moves, k does not depend on N, so the time to algorithmically perform the move is constant.

[24] The pivot algorithm excels also when it comes to walk updating, because the information one has to store on each step is of order 1 even if the walk itself is changed in a non-local fashion. This involves clever programming, and a tricky use of data structures [see section 3.3.3].

a non-ergodic dynamics are strictly infinite, but one could restrict the analysis to an *ergodic class*, that is, the ensemble of all configurations connected (through successive iterations of the Markov chain) to some particular initial state. Within this argument one observes the mean-square radius of gyration $\langle R_m^2 \rangle$, pretending it represents the typical dynamics of the system. For every update of the walk, a few (order 1) steps move a distance of order 1, so that the change in $\langle R_m^2 \rangle$ is of order $N^{\nu-1}$ (ν is the static critical exponent of the correlation length). One must allow the mode to walk an average distance of order $N^{2\nu}$ in order to have it lose memory of past configurations. Guessing that the mean-square radius of gyration performs a random walk around the lattice one concludes that an average $\left(N^{2\nu}/N^{\nu-1}\right)^2$ moves are needed. So $\tau \sim N^{2+2\nu}$: this dynamics is very slow.

Bilocal (N-conserving) algorithms are a little faster. The oldest one of this kind was the *reptation* algorithm, whose moves are those of the slithering-snake kind. This algorithm is clearly non ergodic, all configurations with both ends trapped being frozen. It is easily seen by a heuristic argument that the correlation time behaves like N^2 (it is the time it takes for the random-walking information to cover the entire N-step walk). Many other moves can be added to this set, in order to recover ergodicity, thus obtaining a plethora of different algorithms (for recent work on new extended reptation dynamics see [41], where it is showed how some cleverly designed algorithms show very mild slowing down).

In the N-changing ensembles two algorithms are worth citing, which have $\pi_\beta(\omega) = \text{const}\ \beta^{|\omega|}$ as invariant probability distribution. One is the *Berretti–Sokal* (BC) algorithm [25][25] (also called *slithering-tortoise*). The elementary moves are as follows: either one attempts to append a new step (equally distributed in each direction) to the walk, or else one deletes the last step from the walk. The probabilities of appending and deleting a step are chosen so that their ratio is $2d\beta$. Clearly this dynamics varies the end-point $\omega(N)$. It is easy to see that this algorithm is ergodic and satisfies detailed balance with respect to π_β. With an argument similar to that used for estimating the critical slowing down of the reptation algorithm one can heuristically argue that $\tau \sim \langle N \rangle^2$ for the BC dynamics.

The other is the *BFACF* algorithm [42, 43], which leaves the end-point fixed instead. It uses the one-bead flip (the only possible one-bead internal move), which has $\Delta N = 0$, and kink insertion and deletion, which have $\Delta N = \pm 2$. Ergodicity and dynamical behavior of this algorithm are most

[25]It is also worth noting that the authors of this work devise a way of computing *a priori* (that is, prior to the actual performing of the simulation) error estimates, based on the particular scaling law of the number of N-step SAWs.

subtle problems (see [35] for a review).

3.3.2 The Pivot Algorithm

The *pivot* algorithm is was invented in 1969 by Lal [44], and then reinvented in 1985 by MacDonald *et al.* [45][26]. It had not been studied comprehensively until the work of Madras and Sokal [29], where it is shown to be highly efficient, and its dynamical properties are investigated.

The pivot is a dynamics for walks in the fixed-N, variable end-point ensemble (the other end-point is fixed at $\omega_0 = 0$). The invariant probability measure is the standard equal-weight SAW distribution $\pi_\omega = 1/c_N$. The elementary moves are as follows: for every MC step one chooses a proper time k along the walk and an element of the symmetry group of the lattice; one then applies the symmetry-group element to the sub-walk constituted by $\omega_{k+1} \dots \omega_N$ using ω_k as a pivot, that is as a temporary 'origin' around which the move is performed. Then the proposed walk is checked for self-avoidance and rejected if it is self intersecting (in which case a null transition is made). Obviously, the pivot time is chosen so that $k < N$ because a symmetry around ω_N would always be an identity of the whole walk.

A specific variant of the pivot algorithm is specified by fixing the distributions of the object that are to be chosen 'at random' (the pivot location and the symmetry-group element). The pivot time k can be chosen according to any preset probability $(P_i)_{i=0,\dots,N-1}$, provided that $P_i > 0$. Strict positivity must be satisfied in order to ensure ergodicity. In fact, if some proper time k along the chain is never hit, then the angle between the two bonds that tie together at k will never change. The symmetry operation $g \in G$, where G is the group of orthogonal transformations that leave the lattice invariant, may be chosen according to any probability distribution that satisfies $p_g = p_{g^{-1}}$ for all g and has enough nonzero entries to ensure ergodicity (the question about ergodicity is a rather subtle one for the pivot algorithm[27]). It is easy to see that the equality of the probabilities for an element of the group and its inverse is a necessary and sufficient condition for the dynamics to satisfy detailed balance with respect to the uniform distribution[28].

Some variations of the pivot algorithm are possible, though not always advantageous. Firstly it is historically worth noting that the original work of Lal [44] uses the *bonds* of the walk as pivot locations, and the symmetry

[26] In this latter work the dynamics was called 'wiggling'.

[27] For instance, in two dimensions one would guess that $\pi/2$ rotations are enough for all the configurations to reach each other. This is false! See below.

[28] Because it is a necessary and sufficient condition for the equality of $p_{g_1}(k_1)\, p_{g_2}(k_2) \cdots p_{g_l}(k_l)$ and $p_{g_l^{-1}}(k_l)\, p_{g_{l-1}^{-1}}(k_{l-1}) \cdots p_{g_1^{-1}}(k_1)$.

operation is performed with respect to axis specified by the pivot bond. For the simple hyper-cubic lattice this algorithm is clearly not ergodic, since a rod-like configuration is frozen.

Another variant is one in which the choice of the transformation depends on the configuration in the vicinity of the pivot location. For instance a useful such variant could be one that excludes immediate reversals of the walks, which would get rejected in a SAW[29].

One obtains yet another variant — which is not strictly a variant, but rather an optimization — by applying the symmetry-group element to the shorter 'half' of the walk, not always to the part of the walk subsequent to the pivot point. The initial point of the walk would no longer stay at the origin, but one can get a slight (~ 2) improvement in the computational work (the average number of points the algorithm needs to move lowers from $N/2$ to $N/4$)[30].

At first sight, the pivot seems to carry a great drawback: as N gets large, nearly all proposed walks will get rejected, because of the very non-locality of the moves. Indeed, the acceptance fraction does go to zero as $N \to \infty$, but it does so like N^{-p} with a very low exponent p (in two dimensions it is found to be around 0.19). This is perhaps because the moves exploit the self-avoidance of the two segments that are moved, by transforming them in a rigid fashion so that they remain self-avoiding: the only intersections may be those between points belonging to different segments[31]. On the other hand, the non-locality of the moves carries the benefit of producing a totally different configuration — an independent one, for all practical purposes — once every few moves. So the pivot algorithm sums up to a very good one, actually the best known algorithm for simulating polymers *above* the theta temperature.

A crude estimate helps in understanding the behavior of the acceptance fraction f. The two sub-walks into which the pivot time k divides the whole walk are both self-avoiding. The estimate is based on the assumption that these two walks are *typical* k-step and $(N-k)$-step SAWs. Under this hypothesis the acceptance fraction is simply the ratio between the number of N-step truly self-avoiding walks and the number of possible (in general in-

[29] Clearly, this kind of variant makes no sense in a simulation of an interacting walk for which self-intersections are not forbidden (such as the Domb–Joyce model), because the dynamics would live in the wrong ensemble.

[30] This optimization has no effect when one uses the clever data structures accounted for in section 3.3.3. [See that section for a much greater improvement of the computational weight].

[31] This property is exploited also for speeding up the self-avoidance check in the optimized implementation of section 3.3.3.

tersecting) walks built with the two sub-SAWs, that is

$$f \approx \frac{c_N}{c_k \; c_{N-k}} \tag{3.45}$$

Which, for N large, gives the estimate

$$f \sim N^{-(\gamma-1)} \tag{3.46}$$

Hence $p = \gamma - 1 \approx 0.344$ in two dimensions. Of course, the assumption made here is not realistic: the two sub-walks are far from being typical. First of all, they are appended after each other to form a restricted walk, so they are expected to be more stretched than a typical SAW, in order to favor mutual self-avoidance. Moreover, their relative orientation is not random, because they are more likely to be found 'pointing in the same direction', that is, with a small angle between their end-to-end vectors. Again, this is because this configuration is more likely not to cause self-intersections.

It should be noted, also, that *local* observables evolve a factor N slower than global ones, because for the latter a few accepted moves are enough to radically change their value, while the former need the pivot point to be inside the part of the walk they belong to (for example, the angle between the first two bonds is changed only when the pivot point is exactly the second step of the walk).

The ergodicity of the pivot dynamics depends on which transformations are given nonzero probabilities. It turns out that in general $\pi/2$ rotations alone are not sufficient to ensure ergodicity; as a matter of fact, there exists a 223-step SAW in $\mathbb{Z}^2$ which is frozen with respect to all possible such rotations[32]. Happily, it suffices to add all axis reflections in order to have an ergodic algorithm. More precisely, the following theorems can be proved (the versions provided here apply to $d = 2$, but similar theorems hold for the general hyper-lattice $\mathbb{Z}^d$).

Theorem. *The pivot dynamics is ergodic, provided all axis reflections, and either all $\pi/2$ rotations or all diagonal reflections are given nonzero probability.*

Theorem. *The pivot dynamics is ergodic, provided all π rotations, and either all $\pi/2$ rotations or all diagonal reflections are given nonzero probability.*

The proofs of these theorems — the detailed versions of which can be found in [29] — are based on proving that any N-SAW may be transformed

[32] This is a stunning display of the complexity of self-avoiding walks. See [29] for a picture of such a walk.

into a straight rod by some sequence of at most $2N-1$ moves of the kind considered. A sketch of the proofs is as follows. Let $B(\omega)$ denote the smallest rectangular box containing the walk ω; let $D(\omega)$ be the sum of the sides of $B(\omega)$ (so that D is the l^1 diameter of B); let $A(\omega)$ be the number of straight internal angles in ω. A walk ω having its end-points ω_0 and ω_N in opposite corners of its $B(\omega)$ will be called *diagonal*. Then the set of all N-step SAWs can be partitioned into the two subsets of diagonal and non-diagonal walks.. Any walk belongs in exactly one subset. Then it is proved that for any non-diagonal SAW ω there exists a pivot move that transforms it into a new SAW ω' with $D(\omega') > D(\omega)$ (the box containing the walk is 'bigger') and $A(\omega') = A(\omega)$ (the new walk has the same number of turns as the old one). Likewise, it is proved that for any diagonal SAW ω which is not a straight rod there exists a pivot move that transforms it into a new SAW ω' with $D(\omega') \geq D(\omega)$ and $A(\omega') = A(\omega) + 1$. Then every N-step SAW which is not a straight rod can be transformed by a single pivot move into another SAW with strictly larger $A + D$, but since $A \leq N-1$ and $D \leq N$, and $A + D = 2N - 1$ if and only if the walk is a rod, it follows that any N-step SAW can be transformed into a rod by a sequence of at most $2N-1$ pivots.

Notice that when applied to weakly self-avoiding walks (such as the Domb–Joyce model), the pivot dynamics is always ergodic, at least for finite interaction coupling. This is because the actual ensemble for the Domb–Joyce model contains the set of all random walk, and they all have non-zero probabilities — at least in the coupling regime where the model does not strictly become a SAW — so there can be no frozen configurations. So, for the Domb–Joyce model, even taking the sole $\pi/2$ rotations suffices. But this does not mean that there can not exist bottlenecks, that is very unlikely configurations through which the system is forced to pass, in order to go from one subset of the state space to another. These are usually responsible for making the dynamics slower, or even cause severe metastability, when the system is quasi-trapped in one of the subset and will take long to jump to another. For instance, the 223-step SAW which is frozen with respect to $\pi/2$ rotations lies behind a bottleneck, because if the repulsive force is strong it must cross a very unlike (i.e. whose probability is low) configuration involving (on average) a lot of self-intersections.

3.3.3 Optimized Implementation of the Pivot Algorithm

Naive implementations of the pivot algorithm have performances limited to $O(N)$, meaning that the time to carry out a move (accepted or not) in an

N-step walk is of order N. This includes the time needed to scan the walk for self-intersections, and the time needed to write the new walk to memory when the proposed move is accepted.

Checking self-avoidance is the most delicate part. Using a brute-force routine that checks non-coincidence for *each* pair of walk points is clearly not feasible, because it would take a time $O(N^2)$, which is far too expensive. A more clever way of checking self avoidance is by using a so-called *bit table*. A bit table is a large block of memory in which each lattice site is assigned a bit; this bit is set to 1 if the walker passed on that site and 0 if that site does not belong to the walk. One stores information about the walk in two (redundant) ways: one is the usual structure that contains the walk points ω_i (be it a vector, a linked list, or whatever), and the other is a bit table. Now checking whether a site is empty is a matter of reading one bit in the table, an operation that takes a time $O(1)$. Thus the self-avoidance check using bit tables takes a time $O(N)$. The drawback of using such a data structure is the heavy memory requirements: the memory needed to store the table raises fast with N, especially in high dimensions. A more feasible solution is the use of *hash tables* (see for instance [46]), a structure in which several lattice sites share the same memory address, and collisions are treated only when they happen; hash tables are almost as fast as bit tables, but use much less memory. Whatever method is chosen to store the data and to test the self-avoidance, it is clearly useless to compare walk points that lie on the same sub-walk with respect to the pivot point, so any implementation of this check should consider only couples of points lying on opposite sides of the point around which the pivot is performed.

The specific optimized implementation of the pivot algorithm that will be used here is due to the recent work of Kennedy [47]. The two aspects that are treated differently than in older implementations are the test for self-intersections and the routine that actually carries out the pivot move, that is the piece of code that stores the new walk into memory. Both these steps used to reduce the performance to $O(N)$, because the bit table was the best method available for the self-avoidance check, and because actually writing the whole new walk to memory clearly takes a time of order N.

The method for speeding-up the test for self-intersections takes advantage of the fact that the walk only takes nearest neighbor steps, that is, $\omega_{i+1} - \omega_i = e_\mu$, where e_μ is one of the vectors that generate the lattice. When comparing the walk at times i and j one does not only check whether $\omega_i = \omega_j$, but one computes the distance $d(\omega_i, \omega_j) = ||\omega_i - \omega_j||_1$ instead. This norm $|| s ||_1$ (with s belonging to the lattice) is to be defined as the minimum num-

ber of steps needed to get to site s (from the origin)[33]. For the hyper-cubic lattice this distance is simply the l^1 norm. Clearly, if the distance between ω_i and ω_j happens to be null, then $\omega_i = \omega_j$. The information carried by this quantity is valuable in speeding-up the test, because of the connectedness of the walk; in fact if d is non zero one can conclude not only that $\omega_i \neq \omega_j$, but also that

$$\omega_{i'} \neq \omega_{j'} \quad \text{for all } i', j' \leq N \text{ such that } |i - i'| + |j - j'| < d(\omega_i, \omega_j) \tag{3.47}$$

One now needs a clever way of choosing what values of i and j to check. The most obvious strategy would be to let i run through all the times between $p+1$ and N (where p is the pivot time), and update j (in the range $[0, p-1]$) through 'jumps' of length d, that is $j_{k+1} = j_k - d(\omega_i, \omega_{j_k})$; the procedure should stop when an intersection is found, that is when $d = 0$, or when both i and j are out of their ranges. Unfortunately, such an algorithm would take a time $O(N)$, because i is still taking *all* the values in the range $[p+1, N]$, that is an average $N/2$ different values if the pivot location is chosen with uniform probability distribution.

It is then clear that in order to effectively improve the behavior of the self-avoidance check one must find a way of *simultaneously* let both i and j make jumps of more than 1 step. The improved strategy is as follows. Throughout the algorithm i and j will be times on opposite sides of the pivot time p, so that $j < l < i$. The algorithm is initialized with $i = p + 1$ and $j = p - 1$. Then at each step one either decreases j or increases i in such a way that

$$\omega_{i'} \neq \omega_{j'} \quad \text{for all } i', j' \text{ such that } j < j' < p < i' < i \tag{3.48}$$

until a self-intersection is found, or until both the indexes i, j have reached the end of the walk. This property is clearly satisfied in the initial state, so it will hold during the whole procedure. The algorithms for increasing i and decreasing j are analogous, so only the procedure for increasing i will be explained. Let m_i be the distance of ω_i from the set $\{\omega_k : j < k < p\}$, that is the minimum of the distances from ω_i to ω_k with k running on the values from j to p. If $m_i > 0$ then $\omega_{i'} \neq \omega_{j'}$ for all i', j' such that $i \leq i' < i + m_i$ and $j < j' < p$. Thus, one can increase i by m_i, and property (3.48) will still hold. The value of m_i need not be computed exactly (moreover, such a computation would take way too long, and would completely destroy the advantages of this implementation): a lower bound b_i will suffice, then one can increase i by b_i.

[33] In this definition the actual walk that sits on the lattice does not matter: the minimum number of steps is intended on a void lattice, so that self-avoiding constraints do not enter in the definition.

In order to compute such a lower bound for m_i one uses a loop on j' running from $p-1$ down to j and proceed as follows, again exploiting the connectedness of the chain. At the start of the loop b_i is set equal to N, so that it can not grow higher. At each step one computes the distance $d = d(\omega_i, \omega_{j'})$. Then an integer $s < d$ is picked, so that the distance of ω_i from the set $\{\omega_k \colon j' - s \leq k \leq j'\}$ is surely at least $d - s$. Now the estimate b_i can be replaced by $\min(b_i, d-s)$, and the updated b_i will be a lower bound on the distance of ω_i from the set $\{\omega_k \colon j' - s \leq k < p\}$, so one reduces j' by $s+1$. When j' reaches j, b_i will be a lower bound on m_i.

The possible choices of s are several, the most natural of which would be to take s as the integer part of $d/2$. However there are significantly better choices, such as the following, which exploits the knowledge of the previous value of b_i in order to find a trade-off between a fast decrease in j' and a lower bound b_i which is not too far from the real m_i.

$$\begin{cases} s = d/2 & \text{if} \quad d < b_i \\ s = d - b_i & \text{if} \quad d \geq b_i \end{cases} \tag{3.49}$$

If the algorithm reaches the point in which both i and j exceed their ranges, then the walk is self avoiding and should be accepted. As long as both indexes stay in their ranges, one is free to choose whether to increase i or decrease j. A reasonable choice is to increase i if it is closer to p and decrease j otherwise, so that the both indexes travel away from p roughly at the same rate. This kind of updating searches for intersections in the vicinity of the pivot point first, so that immediate returns and short loops are detected fast.

The second performance bottleneck one has to speed up is the writing down of the new walk, when it is accepted. The key idea here is not to write down the *whole* walk at each MC step, but to solely keep track of the transformations and the pivot locations. One then has a list of operations that only after a certain number of iterations are to be actually carried out. Of course, the routine that checks for self intersections needs to fetch coordinates of the walk steps, so some of them must be computed at each step. But the important point here is, that the optimized version of the self-avoidance test does not need *all* walk sites. Indeed, it turns out to be more effective to store the information about the walk in some 'encrypted' form that takes a time $O(1)$ to write to memory, and find a way of effectively fetch the actual walk locations by reading from such a structure.

This structure could be for instance an array of the pivot times p and the symmetry-group elements g. But it turns out that the time required to compute the position of a location on the walk from such a data structure would be significant. Instead, it is found to be more efficient to use a different

data structure that contains information about the segments of walk that lie between two successive[34] pivot times. The following data structure is motivated by the observation that one can think of the segment of ω' that lies between times p_i and p_{i+1} as being rigid, and when a pivot move is performed on a pivot time p the transformations needed to get ω_j from ω'_j change only for $j > p$. Thus for each accepted pivot one stores

- The old walk ω'. This is the last walk that has been completely enumerated.
- The number of rigid segments (as defined above) which constitute the walk. Note that this needs note be the 'age' of ω' — that is, the number of accepted pivot not yet carried out — because two different pivot moves could have the same pivot time, so that the lists defined in the following do not grow.
- The pivot times $p_1 < p_2 < \cdots < p_n$. These are stored in increasing order, not in the original order they were picked.
- The lattice symmetries $g_1, g_2, \cdots, g_n$. These are not the symmetry-group elements of the various pivot moves, but rather the symmetries that are to be applied to the rigid segments.
- The lattice sites $x_1, x_2, \cdots, x_n$ that specify the translations that are to be applied to the rigid segments.

With this data structure any walk location ω_j belonging to the i-th segment ($p_i < j \leq p_{i+1}$) can be calculated using the equation

$$\omega_j = g_i \omega'_j + x_i \tag{3.50}$$

Then, for every accepted pivot move, the elements of the data structure must be updated so that equation (3.50) holds for each element of the chain. This is done in two steps. In the first step one simply adds the pivot time p to the list. In order to do this, the right position in the list has to be found for p. This means that a k must be found, such that $p_k < p < p_{k+1}$; there are efficient ways of performing this task, but this does not seem to be a critical aspect of the whole algorithm, so that even a naive searching procedure should suffice. If p is already present in the structure then nothing is done, and one skips to the second step. Otherwise, the changes in the data structure are

[34] Here 'successive' is to be intended with respect to proper time on the walk, not MC time.

- $n \rightsquigarrow n+1$ (one of the formerly rigid segments has been cut in two by the new pivot point, so their number increases by 1).
- $p_1, \cdots, p_n \rightsquigarrow p_1, \cdots, p_k, p, p_{k+1}, \cdots, p_n$ (the new pivot time is inserted in the list).
- $g_1, \cdots, g_n \rightsquigarrow g_1, \cdots, g_{k-1}, g_k, g_k, g_{k+1}, \cdots, g_n$ (g_k is inserted in the list as a placeholder; the only symmetries that will be recalculated during step two are those from this g_k to g_n).
- $x_1, \cdots, x_n \rightsquigarrow x_1, \cdots, x_{k-1}, x_k, x_k, x_{k+1}, \cdots, g_n$ (same as above).

The second step is to update the elements of the two lists containing the symmetry operations g and the translations x. Under a pivot move specified by time p and symmetry g (and then with pivot point $x := \omega_p$, and with k such that $p_k < p < p_{k+1}$), a walk site ω_j belonging to the i-th segment with $i > k$ (so that ω_j does get affected by the move) transforms according to the following equation

$$\omega_j \rightsquigarrow g(\omega_j - x) + x = g(g_i\omega'_j + x_i - x) + x = gg_i\omega'_j + gx_i - gx + x \quad (3.51)$$

Hence, the right update rule for the structure is the following

- $g_i \rightsquigarrow gg_i$ for all $i > k$ (nothing changes for $i \leq k$).
- $x_i \rightsquigarrow gx_i - gx + x$ for all $i > k$ (nothing changes for $i \leq k$).

So one keeps on updating the structure at each successful MC step; then, after a given number N_p of such 'implicit' updates, the walk is completely calculated and substituted to ω'.

Summing up, there are three steps the time required for completion of which depends on the length of the walk:

- The self-avoidance test.
- The updating of the data structure at each accepted pivot.
- Once in a while, the complete enumeration of the current walk.

The first of these bottlenecks is very hard to analyze a priori, and needs some empirical study. This step involves repeated computations of the distance $d(\omega_i, \omega_j)$; one assumes that the average number of such computations grows with N^σ. Moreover, every time the procedure that computes d is called, some evaluations of walk points are to be made by applying equation (3.50). In order to use this equation for ω_j one must first find i such that $p_i \leq j \leq p_{i+1}$.

This searching task can be accomplished in a time of order $\log n$ with a bisection algorithm, and this translates in an average $O(\log N_p)$ behavior. So the first bottleneck of the list is believed to take $O\left(N^\sigma \log N_p\right)$.

The second bottleneck involves only operations made on the data structure. This structure is stored in a linear fashion, and its size is specified by n, so that one expects this step to take an average time of order N_p.

The third task in the list above is very time-consuming (many of the intricacies of this whole algorithm share the goal of reducing the times this very task has to be accomplished): since it must travel the whole walk, it takes a time $O(N)$. Happily, now one needs not carrying this step out for every accepted pivot, but only once every N_p, so that the overall behavior is expected to be $O(N/N_p)$.

Finally, the total time per accepted pivot is

$$O\left(N^\sigma \log N_p\right) + O(N_p) + O\left(\frac{N}{N_p}\right) \tag{3.52}$$

It is then clear that the best choice for N_p is to take it proportional to $N^{1/2}$, in order to balance the second and the third terms in the expression above. As far as the exponent σ is concerned, Kennedy [47] finds 0.57 as an upper bound for σ in two dimensions, but the real value could be significantly less than this estimate.

Now, we want to apply (a modified version of) the pivot algorithm to the simulation of the Domb–Joyce model. Unfortunately, one of the great advantages of the test for acceptance seems to be broken in this case, since a self-intersecting walk should not be discarded with probability 1. Recalling the Metropolis criterion, one should compare the energy E_2 of the current walk to that of the previous one E_1; if the difference is non-positive — that is, if the proposed walk has a lower energy than the old one — then the proposed walk is accepted with probability 1. Instead, if the difference is positive, then the walk is accepted with a probability proportional to $\exp(-v\Delta E)$, where v is the excluded volume. It would seem that the routine for checking self-avoidance should cross the *whole* walk every time, in order to compute the current energy.

Happily, there is a way of doing things that prevents this routine to reach the end of the walk all the times, thus reducing the time needed to complete the task. When one has to accept a walk with probability proportional to $p = \exp(-v\Delta E)$, in fact one compares this value with a uniformly distributed random number generated in $[0, 1]$, and accepts the walk if the random number is not greater than the statistical weight. The trick is to generate this random number *before* the algorithm starts scanning the chain for self-intersections. Then, for each self-intersection that is found ΔE increases,

so that p decreases, and one can be sure that the walk will be rejected as soon as the value of p reaches the previously generated random number.

Clearly, the algorithm will be faster and faster as one increases the value of v (that is, as one reaches the true self-avoiding walk), because p increases faster for higher values of v, and the rejection point will be reached sooner. Inversely, the algorithm is seen to become much slower in the vicinity of the random walk ($v \to 0$).

Chapter 4

Theoretical Results

4.1 Non-Rigorous Results

Most non-rigorous results are obtained by considering models equivalent to the Edwards model. Notation in this section will be changed a bit sometimes, in order to keep it consistent with the symbols and definitions used in the original works cited. The chain length will be denoted by s, L or N, the chain itself by c, ω or r and the proper time by τ or by latin indexes (depending on whether it is continuous or discrete), moreover the excluded volume parameter will be denoted by g, v or v_2.

4.1.1 Flory Formula and Mean Field

The most simple approximation one can apply to the estimation of critical exponents is based on an energy balance argument. Following [48] one considers the following measure

$$\mathscr{D}[\mathbf{c}] \exp\left[-\frac{1}{4}\int_0^s \dot{\mathbf{c}}^2(\tau)\,d\tau - \frac{g}{6}\int_0^s d\tau_1\,d\tau_2\,\delta^d\left(\mathbf{c}(\tau_1)-\mathbf{c}(\tau_2)\right)\right] \tag{4.1}$$

The two terms that appear at the exponent in (4.1) represent respectively the free action and the interaction. The energy balance argument consists in asking that these two terms satisfy the same power scale law for s large. By straightforward dimensional analysis [see also section 4.1.2] and by requiring that $|\mathbf{c}|$ scale as s^ν one has

$$\int_0^s \dot{\mathbf{c}}^2(\tau)\,d\tau \sim s^{2\nu-1} \tag{4.2}$$

$$\iint_0^s d\tau_1\,d\tau_2\,\delta^d\left(\mathbf{c}(\tau_1)-\mathbf{c}(\tau_2)\right) \sim s^{2-d\nu} \tag{4.3}$$

which can be easily obtained by remembering that

$$\delta^d(a\mathbf{x}) = \frac{1}{a^d}\delta^d(\mathbf{x}) \tag{4.4}$$

The requirement that the power scaling behavior be the same for the two terms can then be written as $2\nu - 1 = 2 - d\nu$, so that for the swelling exponent ν one gets

$$\nu = \frac{3}{d+2} \tag{4.5}$$

which is known as the *Flory formula*.

The original argument due to Flory was rather different, but the main features (and the result) are the same. His method was as follows. The starting point is an N-chain in d-dimensional space, with a certain unknown radius (end-to-end distance) R, and an internal mean monomer concentration given by

$$c_m \approx \frac{N}{R^d} \tag{4.6}$$

The repulsive energy density due to monomer-monomer interactions is written in terms of the *local* monomer concentration c. When only two-body interactions are taken into account[1], the local energy density will be proportional to c^2:

$$E_{rep}^{(loc)} = vc^2 \tag{4.7}$$

where v is a positive excluded volume parameter[2]. The whole Flory argument relies on two approximations. The first consists in replacing the average of the local c^2 with the square of the average

$$\left\langle c^2 \right\rangle \rightsquigarrow \left\langle c \right\rangle^2 = c_m^2 \tag{4.8}$$

so that the mean repulsion energy density is

$$E_{rep} = v\frac{N^2}{R^d} \tag{4.9}$$

This is a typical *mean field* approximation, where all correlations among monomers are ignored. This tends to swell the chain indefinitely, because the more far away the monomers are, the lower the repulsion is. To obtain a non-trivial result, some attractive interaction between the monomers is to

[1] When considering the approach described in the previous paragraph, this assumption is implicit in the choice of the action as in (4.1), where the interaction term is in the form of a single delta function.

[2] In the original Flory notation $v = a^d(1 - 2\chi)$, where a is the monomer radius and $\chi < 1/2$ is an interaction parameter.

be included. The second approximation regards the particular choice of this interaction. Flory included a term describing the attraction between the first and the last monomers, which are a distance R apart. Following the result for ideal (Gaussian) chains , an elastic form of the attractive energy density was picked:

$$E_{el} = \frac{R^2}{a^2 N} \tag{4.10}$$

The total energy density then presents a minimum for some radius R_f, called the *Flory radius*

$$R_f^{d+2} = \frac{d}{2} a^2 v N^3 \tag{4.11}$$

from which one gets the Flory formula (4.5).

Since one expects ν to be bounded from below by the Gaussian value of $1/2$, equation (4.5) is expected to hold at most for $1 \leq d \leq 4$. This appears as another clue that the upper critical dimension of the excluded volume problem be 4; in the picture given by (4.1) this means that for $d \geq 4$ the interaction term can not be strong enough to balance the free action. The Flory formula gives stunning results in $d = 1, 2$, where it predicts the values $\nu = 1, 3/4$, which are confirmed by other methods. It is tempting to conjecture that the formula is valid also near dimension 4, but the (purportedly more precise) renormalization group results deviate from it even at first order in $\epsilon = 4 - d$.

De Gennes strongly criticized Flory's argument, but his more refined version of the derivation actually produced worse results. As a matter of fact, the whole Flory's argument seems to benefit from the cancellation of two errors: the overestimation of the repulsive energy by the mean field approximation, and the overestimation of the elastic attraction between the ends of the chain, which is truly elastic only in the Gaussian approximation. Indeed, as de Gennes himself underlines [3], "As often happens in self-consistent field calculations, two errors cancel each other to a large extent. Many post-Flory attempts, which tried to improve on one term leaving the other unaltered [either the repulsive or the elastic energy], led to results that were poorer than the Flory formula".

A somewhat more accurate and modern mean field approach to the model of weakly self-avoiding walks is that developed in [13]. In this work a more general model than the Domb–Joyce is considered, in which the repulsion between monomers decreases as a power of the distance along the chain $|i - j|^{-\lambda}$; this model is referred to as the forgetful weakly self-avoiding walk (FWSAW). This model is constructed on continuum space, but the 'local stiffness' of the polymer is preserved by considering chains built with discrete

segments. The measure is

$$\mathscr{D}[\omega]\,\frac{1}{Z}\exp(-H[\omega]) \tag{4.12}$$

where the Hamiltonian is given by

$$H[\omega]=\frac{1}{2}\sum_{i=0}^{N}\left(\vec{\omega}_i-\vec{\omega}_{i-1}\right)^2+\frac{1}{2}\,g\sum_{i\neq j}\frac{V\left[\left(\vec{\omega}_i-\vec{\omega}_j\right)^2\right]}{|i-j|^{\lambda}} \tag{4.13}$$

in which V stands for any short range potential, used to regularize a delta function. The mean field approximation is based on a variational specialization of the general Gaussian trial measure

$$H_0[\omega]=\frac{1}{2}\sum_{i,j}G_{ij}^{-1}\,\vec{\omega}_i\cdot\vec{\omega}_j \tag{4.14}$$

The function G_{ij} is determined by minimization of the following functional

$$F[G]=\langle H-H_0\rangle_0-\log Z_0 \tag{4.15}$$

One then introduces the Fourier transform $\widetilde{G}(p)$ of G. In the large-N limit p has infinite range, and the exponent ν can be extracted from the small-p behavior of $\widetilde{G}(p)$. The results are

$$\nu_{MF}(\lambda,d)=\begin{cases}1/2 & \text{for } \lambda>(4-d)/2\\ (2-\lambda)/d & \text{for } \lambda<(4-d)/2\end{cases} \tag{4.16}$$

for $2\leq d\leq 4$ (with logarithmic corrections to $\langle R^2\rangle$ for $\lambda=(4-d)/2$), and

$$\nu_{MF}=1/2 \quad \text{for } d>4 \tag{4.17}$$

This result again enforces the hypothesis on the upper critical dimension.

These mean field results are clearly not correct for $\lambda=0$ — that is for the infinite memory case, the Domb–Joyce model — for dimension less that the critical dimension. For this reason the authors in [13] propose the following conjecture (as supported by numerical results)

$$\nu(\lambda,d)=\min\left(\nu_{SAW}(d),\nu_{MF}(\lambda,d)\right) \tag{4.18}$$

4.1.2 Scaling Relations

It is interesting to study the behavior of the Edwards model under general scale transformations. By means of this approach [49] — which will be called *naive*, as opposed to more sensible formulations based on the renormalization group — it is possible to obtain general scale relations without perturbative ϵ expansions.

The partition function considered is of the form

$$Z = \int \mathscr{D}[r] \, \exp\left(-\frac{3}{2l} \int_0^L \dot{\mathbf{r}}^2(\tau) \, d\tau - \frac{v}{2l^2} \iint_0^L d\tau_1 \, d\tau_2 \, \delta^d(\mathbf{r}(\tau_1) - \mathbf{r}(\tau_2)) \right) \tag{4.19}$$

where l is the Kuhn length — which represents the typical bond length — and v is the usual excluded volume parameter. The degree of polymerization is naturally defined as $N = L/l$. The scale transformations under which the behavior of the model will be studied are of the kind

$$\begin{aligned} \mathbf{r}'(\tau') &= a_0 \, \mathbf{r}(\tau) \\ \tau' &= b_0 \, \tau \end{aligned} \tag{4.20}$$

This scaling transforms the partition function (4.19) into the following

$$\begin{aligned} Z = \int \mathscr{D}[r'] \, \exp\Bigg(& -\frac{b_0}{a_0^2} \frac{3}{2l} \int_0^{b_0 L} \dot{\mathbf{r}}'^2(\tau') \, d\tau' \\ & - \frac{a_0^d}{b_0^2} \frac{v}{2l^2} \iint_0^{b_0 L} d\tau_1' \, d\tau_2' \, \delta^d(\mathbf{r}'(\tau_1') - \mathbf{r}'(\tau_2')) \Bigg) \end{aligned} \tag{4.21}$$

apart from an infinite volumetric factor. The partition function (4.21) can be seen to be equivalent to (4.19) through the following positions

$$L' = b_0 L, \;\; l' = a_0^2 l / b_0, \;\; v'/l'^2 = (a_0^d / b_0^2)(v/l^2) \tag{4.22}$$

This makes a connection between a partition function and infinitely many polymer theories; the model is then really a *two-parameter* model, in spite of the dependence of (4.19) from three parameters. By the further position

$$b := b_0 l, \;\; a := a_0 l, \;\; b = a^2 \tag{4.23}$$

one gets for the partition function Z the functional form

$$Z = Z(bN, \, b^{(d-4)/2} v l^{-d}) \tag{4.24}$$

By considering (4.21) and choosing

$$b_0 = L^{-1} \tag{4.25}$$

$$a_0^2 = b_0 \tag{4.26}$$

so that the rescaled coupling multiplying the free term and the rescaled contour length $b_0 L$ are of order unity (that is, do not depend on L or v), one gets that the two-body interaction coupling $\tilde{v} = v/2l^2$ takes a factor $L^{2-d/2}$. This is of particular importance in the critical crossover regime, where one considers the limit in which $v \to 0$ and $L \to \infty$ [see section 4.4]. Indeed, by the foregoing scaling argument one expects the perturbative expansion (for small v and large $N = L/l$) of observables to depend only on the combination

$$x = vN^{2-\frac{d}{2}} \tag{4.27}$$

which makes sense for $d < 4$.

A similar argument can be applied to the propagator

$$G(\mathbf{R}, L) = \int_{\mathbf{r}(0)=0}^{\mathbf{r}(L)=\mathbf{R}} \mathscr{D}[r]\, Q_{L,v}[r] = \int \mathscr{D}[r]\, \delta^d(\mathbf{R} - \mathbf{r}(L))\, Q_{L,v}[r] \tag{4.28}$$

where $Q_{L,v}[r]$ is the exponential weight in the partition function (4.19). The following scaling behavior is obtained

$$G = (bl^{-2})^{d/2}\, G(b^{1/2} R l^{-1},\, bN,\, b^{(d-4)/2} v l^{-d}) \tag{4.29}$$

From this relation it is found for the end-to-end distance

$$\left\langle \mathbf{R}^2 \right\rangle = l^2 b^{-1}\, f(bN,\, b^{(d-4)/2} v l^{-d}) \tag{4.30}$$

which can be rewritten in a specialized form by exploiting the arbitrariness of the scale factor b and choosing $b = N^{-1}$:

$$\left\langle \mathbf{R}^2 \right\rangle = lL\, f(N^{(4-d)/2} v l^{-d}) \tag{4.31}$$

An immediate consequence of this relation is that in dimension $d > 4$ the behavior in the long chain limit is Gaussian, that is $\langle \mathbf{R}^2 \rangle \sim lL$. This result once again highlights the special role played by $d = 4$ for the relevance of the excluded-volume interaction.

An analogous analysis can be carried out for the general two-body interaction specified by a potential $W(\mathbf{r}(\tau_1) - \mathbf{r}(\tau_2))$. The treatment of the general interaction term represents an interesting subject, because it helps in understanding and closely observing what the meaning and consequences of universality are. One assumes that $W(\mathbf{r})$ solely depends on $|\mathbf{r}|$ — for translational and rotational invariance — and that it has a finite total interaction strength (binary cluster integral) $\int d\mathbf{r} W(\mathbf{r}) < \infty$. It is useful to study the

dependence of the Fourier transform $\widetilde{W}(k)$ on the scale factor. One then observes that in the large-N limit only small arguments of $\widetilde{W}$ contribute to the integration in the transform. Then an expansion around $k = 0$ is used, and one finally obtains

$$\left\langle \mathbf{R}^2 \right\rangle_W = l^2 N \, f\left(N^{(4-d)/2} v_0 l^{-d}, \left\{ N^{-n+(4-d)/2} l^{-(d+2n)} v_{2n} \right\}_{n\in\mathbb{N}} \right) \tag{4.32}$$

where the coefficients v_{2n} are the $2n$-th derivatives of $\widetilde{W}(k = 0)$. Since $N^{-n+(4-d)/2} \to 0$ for every n in the $N \to \infty$ limit, the dependence of $\langle \mathbf{R}^2 \rangle_W$ on the terms that contain derivatives of $\widetilde{W}$ is dropped. This argument then explains in a simple way — at least in three dimensions — why introduction of potentials different from the δ is not important (in the approximate and asymptotic sense of equation (4.32)): only the binary cluster integral $v_0 = \widetilde{W}(0)$ matters.

4.1.3 Renormalization Group

The first attempts towards applying the methods of the renormalization group to polymers were based on transformations *à la* Kadanoff. This kind of approach was introduced in [50] and developed in [51]. The first k monomers on the chain are grouped into the first renormalized monomer, the next k monomers in the second, and so on, until the chain is described by a renormalized chain of N/k units. Unfortunately — as it is claimed in [52] — it is not possible to obtain results at higher orders than ϵ in this scenario. The need for a different formulation of a renormalization group transformation for polymers has been taken into consideration in [52].

The 'naive' approach permits an easy analysis of excluded-volume behavior under rescaling of the model. Yet, it introduces an extreme simplification which is not justifiable: one ignores the dependence of physical observables on the cutoff Λ. This cutoff is introduced in order to reproduce the discrete nature of polymer chains, and to give mathematical sense to the integral of the delta function. Obviously, Λ will scale under a rescaling of the model, too. This non-trivial dependence of the cutoff on the scale is the heart of this renormalization group analysis. One introduces transformations — which will be called $\mathscr{S}_s$ —

$$\begin{aligned} \mathbf{c}'(\tau') &= \mathbf{c}(\tau)/s^{\nu} \\ \tau' &= \tau/s^{(2-\eta)\nu} \end{aligned} \tag{4.33}$$

where $s > 0$ is the scaling factor. This transformation will affect Λ too, transforming it into Λ'. Actually, it will be useful to keep the cutoff fixed

— by means of another transformation, which will be defined later — so that it will be possible to ignore the dependence on Λ, or absorb it into the functional relations. In the 'naive' approach this goal was reached by trivial substitution of Λ' with Λ. Here, another transformation will be introduced, which will take care of transforming Λ' back into Λ: it is the *coarse-graining* or *Kadanoff* transformation $\mathscr{K}_s$, chosen in such a way that

$$\mathscr{K}_s \mathscr{S}_s(\Lambda) \equiv \mathscr{R}_s(\Lambda) = \Lambda. \tag{4.34}$$

One focuses on the regularized propagator

$$\begin{aligned} G(\mathbf{R}, N) = \int_{\mathbf{c}(0)=0}^{\mathbf{c}(N)=\mathbf{R}} \mathscr{D}[c] \, \exp\Bigg[-\frac{d}{2l^2} \int_0^N \dot{\mathbf{c}}^2(\tau) \, d\tau \\ - \frac{1}{2} v_2 \iint_{|\tau_1 - \tau_2| > \Lambda} d\tau_1 \, d\tau_2 \, \delta^d(\mathbf{c}(\tau_1) - \mathbf{c}(\tau_2)) \Bigg] \end{aligned} \tag{4.35}$$

The Kadanoff transformation is constructed as follows.

- (I) Interactions whose loop[3] is smaller than Λ in the chain rescaled with $\mathscr{S}_s$ are absorbed into a redefinition of l^2 and N.

- (II) When two interactions — one then considers two pairs of interaction points, say a and b, and c and d — have their ends separated by proper times less than Λ — that is, a is 'close' to c, and b to d —, these interactions are combined into a redefinition of the interaction parameter v_2.

The overall effect of $\mathscr{K}_s$ is then to integrate out interactions whose scale is between Λ' and Λ. Loosely speaking, $\mathscr{R}_s$ corresponds to looking at the chain from far away: it appears smaller and the fine grain is no longer apparent.

[3]The loop of an interaction is defined as the piece of chain which lies between the two interaction points that contribute to the δ.

The perturbative calculation is based on the *loop-wise* expansion

$$
\begin{aligned}
&G(\mathbf{R},N) = G_0(\mathbf{R},N) + \\
&- v_2 \int_0^N d\tau_0 \int_{\tau_0+\Lambda}^N d\tau_1 \int d\mathbf{r}_0\, G_0(\mathbf{r}_0,\tau_0)\, G_0(\mathbf{0},\tau_1-\tau_0)\, G_0(\mathbf{R}-\mathbf{r}_0,N-\tau_1) \\
&+ v_2^2 \int_0^N d\tau_0 \int_{\tau_0+\Lambda}^N d\tau_1 \int_{\tau_1+\Lambda}^N d\tau_2 \int_{\tau_2+\Lambda}^N d\tau_3 \int d\mathbf{r}_0 \int d\mathbf{r}_1\, G_0(\mathbf{r}_0,\tau_0)\cdot \\
&\Big[G_0(\mathbf{0},\tau_1-\tau_0)\, G_0(\mathbf{r}_1-\mathbf{r}_0,\tau_2-\tau_1)\, G_0(\mathbf{0},\tau_3-\tau_2) \\
&+ G_0(\mathbf{r}_1-\mathbf{r}_0,\tau_1-\tau_0)\, G_0(\mathbf{0},\tau_2-\tau_1)\, G_0(\mathbf{r}_0-\mathbf{r}_1,\tau_3-\tau_2) \\
&+ G_0(\mathbf{r}_1-\mathbf{r}_0,\tau_1-\tau_0)\, G_0(\mathbf{r}_0-\mathbf{r}_1,\tau_2-\tau_1)\, G_0(\mathbf{r}_1-\mathbf{r}_0,\tau_3-\tau_2) \Big] \cdot \\
&\cdot G_0(\mathbf{R}-\mathbf{r}_1,N-\tau_3) \;+\; \mathcal{O}(v_2^3)
\end{aligned}
\tag{4.36}
$$

where the free (Gaussian) propagator was used, which is

$$
G_0(\mathbf{r},n) = \int_{\mathbf{c}(0)=\mathbf{0}}^{\mathbf{c}(n)=\mathbf{r}} \mathscr{D}[c]\, \exp\left[-\frac{d}{2l^2} \int_0^n \dot{\mathbf{c}}^2(\tau)\, d\tau \right] \tag{4.37}
$$

Equation (4.36) is easily obtained from an expansion of the exponential statistical weight

$$
\int \mathscr{D}[c]\, e^{-H[c]} = \int \mathscr{D}[c]\, e^{-H_0[c]}\, [\, 1 \;-\; H_I[c] \;+\; \cdots] \tag{4.38}
$$

The Kadanoff transformation is implemented by subdividing the domain of $d\tau$ integrals into contributions due to interactions with scale larger or smaller than Λ. One then defines the renormalized free propagator as

$$
\begin{aligned}
\tilde{G}_0(\mathbf{R}',\mathscr{R}_s(N);\mathscr{R}_s(l)) &:= G_0(\mathbf{R}',N';l') - v_2' \int_0^{N'} d\tau_0' \int_{\Lambda' \le \tau_1' - \tau_0' \le \Lambda} d\tau_1' \cdot \\
&\cdot \int d\mathbf{r}_0'\, G_0(\mathbf{r}_0',\tau_0')\, G_0(\mathbf{0},\tau_1'-\tau_0')\, G_0(\mathbf{R}'-\mathbf{r}_0',N'-\tau_1') \;+\; \mathcal{O}({v_2'}^2)
\end{aligned}
\tag{4.39}
$$

which includes contributions of all 'small' loops. Expression (4.39) can be calculated using some approximation, namely first order in ϵ and large N. It is then possible to obtain $\mathscr{R}_s(N^{-1})$, while including also two-loop diagrams one finally gets

$$
\mathscr{R}_s(N^{-1}) = s^{2\nu} N^{-1} \left(1 - v_2 \left(\frac{2}{\pi l^2} \right)^2 2\nu\, \log s \right) \tag{4.40}
$$

$$\mathscr{R}_s(v_2) = s^{\epsilon\nu} v_2 \left(1 - 4v_2 \left(\frac{2}{\pi l^2}\right)^2 2\nu \log s\right) \tag{4.41}$$

The fixed points are

$$\begin{aligned} {v_2}^* &= \pi^2 l^4 \epsilon/32 \\ N^* &= 0, \infty \end{aligned} \tag{4.42}$$

The solution with $N^* = 0$ is not interesting since it is not physical, and does not carry information about the critical behavior of polymers. By linearizing around $(N^{-1})^*$ one obtains

$$\mathscr{R}_s(N^{-1}) = s^{2\nu - \epsilon\nu/4} N^{-1} \tag{4.43}$$

In order to extract information about the critical exponent ν one considers the observable $\langle \mathbf{R}^2 \rangle$, for which the following scaling relation holds

$$\left\langle \mathbf{R}^2 \right\rangle = s^{2\nu} f(\mathscr{R}_s(N^{-1}), \mathscr{R}_s(v_2), \mathscr{R}_s(\Lambda) = \Lambda) \tag{4.44}$$

By considering this scaling form at the fixed point $v_2 = {v_2}^*$, the dependence of f from the scale factor will be solely due to the term $\mathscr{R}_s(N^{-1})$. One then chooses s so that

$$s^{2\nu - \epsilon\nu/4} \sim N \tag{4.45}$$

Hence, the dependence of $\langle \mathbf{R}^2 \rangle$ on s is only in the scaling factor which multiplies f. Finally, one obtains

$$\left\langle \mathbf{R}^2 \right\rangle \sim N^{\frac{2}{2-\epsilon/4}} = N^{1 + \frac{\epsilon}{8} + \cdots} \tag{4.46}$$

It is interesting to compare this RG result with the Flory formula for ν; as one can see, the two expressions differ already at first order in ϵ.

$$\begin{aligned} \nu_{Flory} &= \tfrac{1}{2}\left(1 + \tfrac{\epsilon}{6} + \cdots\right) \\ \nu_{RG} &= \tfrac{1}{2}\left(1 + \tfrac{\epsilon}{8} + \cdots\right) \end{aligned}$$

By observing the renormalization group flow for $N = \infty$ (critical theories), one sees that the fixed point divides the parameter manifold into two *branches*, depending on whether $v_2 \gtrless {v_2}^*$. One then expects to observe leading corrections to scaling with amplitudes with different signs for the two regimes, which are called *weak* and *strong coupling branch*. This observation is of fundamental importance in polymer physics, because it clearly highlights the non-universality of certain aspects of the systems one investigates, even near criticality. The amplitudes of the corrections to scaling are highly model-dependent: even their sign depends on the microscopic parameters of the model. Thus one should not be surprised if experimentally observed properties of real polymers in solution approach their asymptotic values from, say, above, while a particular model that is used to describe them (say, the two-parameter theory in the weak coupling regime) predicts the opposite behavior.

4.2 Rigorous Results

Not many exact results are known for the Domb–Joyce model. Moreover, most of them are derived in dimensions above the critical dimension ($d \geq 5$), or for the particular (and rather uninteresting) case $d = 1$. Even for the SAW model — which is relatively simpler — only few exact results are known for every dimension. The only rigorous SAW critical exponent that is known through a rigorous method in dimension $d = 2$ is the swelling exponent $\nu = 3/4$, which is the same value obtained with a variety of other non-rigorous methods, so that one is tempted to guess that in two dimensions non-rigorous methods for the excluded volume problem actually provide precise results. It is interesting to note that the demonstration of the fifty-year-old conjecture $\nu_{d=2} = 3/4$ is based on the accurate study of a process that is nearly resemblant of the Domb–Joyce walk [53].

4.2.1 Bounds on the Number of Self-Intersections

Both an upper and a lower bound on the number of self-intersections of a weakly self-repelling walk are derived in [54]. For generality, the walk is constructed on a graph $\mathscr{G}$ which is infinite (in order to consider the long chain limit $N \to \infty$), connected and transitive (in order to exploit a classical argument in the study of self-excluding objects [see below]).

One considers the set Γ_n of strictly self-avoiding walks of length n defined on $\mathscr{G}$. The transitivity of the graph permits the definition of the *concatenation* of two walks $\gamma_1 + \gamma_2$, with $\gamma_1 \in \Gamma_n$ and $\gamma_2 \in \Gamma_m$ for every couple of integers (m, n). Clearly, the general walk obtained by concatenation of two self-avoiding walk is not itself self-avoiding, so that it does not in general belong to Γ_{n+m}. One then has

$$|\Gamma_{n+m}| \leq |\Gamma_n| + |\Gamma_m| \tag{4.47}$$

that is, the sets Γ_n are *subadditive* in size. A classical theorem on subadditivity then assures that

$$\lim_{n\to\infty} |\Gamma_n|^{1/n} = \mu \tag{4.48}$$

$$|\Gamma_n| \geq \mu^n \tag{4.49}$$

for some positive μ, which is called *connective constant*, and depends only on the choice of the graph $\mathscr{G}$.

It is very interesting to note that an argument of this kind does not rely on the particular sub-object whose appearance is forbidden on a walk, so that a completely similar reasoning can be applied to other classes of constrained

walks. Indeed, every time one defines a subset of random walks as the set of those walks that do not present a certain type of sub-object — be it a rectangle, nested loops, or any other geometrically-constrained loop — it will happen that the general concatenation of two such walks does not obey the same restriction, and the subadditivity property will be satisfied. It is then natural to define *generalized weakly self-avoiding walks* (GWSAW), for which an energy price is payed whenever some given event occurs on a walk. Manifold interesting questions can then be considered about such a class of walks. What classes of GWSAW belong to the same universality class of the SAW? What classes — if any — behave as the random walk? Is there a connection between the dimensionality of the excluded sub-object and the behavior of the model?

Turning back to the Domb–Joyce model, one defines the number of self-intersections, or *self-intersection local time*

$$J_n := \sum_{i<j\leq n} \delta\{\omega(i) = \omega(j)\} \tag{4.50}$$

The estimates for the bounds are based on the following argument. Let Q_n^β be the Domb–Joyce measure with excluded volume β and walk-length n. Let A_1 and A_2 be two subsets of $[0, n^2] \subset \mathbb{N}$. Moreover, let $\langle\cdot\rangle_0$ denote the expectation value with respect to the free measure Q_n^0. Then it is easily verified that

$$\begin{aligned} Q_n^\beta(J_n \in A_1) &= \frac{\langle \exp(-\beta J_n)\, \delta\{J_n \in A_1\}\rangle_0}{\langle \exp(-\beta J_n)\rangle_0} \\ &\leq \frac{\langle \exp(-\beta J_n)\, \delta\{J_n \in A_1\}\rangle_0}{\langle \exp(-\beta J_n)\, \delta\{J_n \in A_2\}\rangle_0} \end{aligned} \tag{4.51}$$

Now, if one is able to demonstrate the existence an A_2 and a number $\tau > 0$ — both independent of n — such that for every n and every A_1 the following property be satisfied

$$\langle \exp(-\beta J_n)\, \delta\{J_n \in A_1\}\rangle_0 < e^{-\tau n}\ \langle \exp(-\beta J_n)\, \delta\{J_n \in A_2\}\rangle_0 \tag{4.52}$$

then one can conclude by relation (4.51) that

$$Q_n^\beta(J_n \in A_1) < e^{-\tau n} \tag{4.53}$$

that is, the subset A_1 contributes 'little' on the number of self-intersections (that is, its contribution on the energy is little). Using arguments of the kind outlined above, the following propositions can be proved (see propositions 1 and 2 in [54])

- (*Upper Bound*) For fixed $\beta > 0$ and $\epsilon > 0$, if $\tau = \epsilon(\log(\deg(\mathscr{G})) - \omega(\mathscr{G})) > 0$ and $\tilde{b} = (1+\epsilon)(\log(\deg(\mathscr{G})) - \omega(\mathscr{G}))/\beta > 0$ then for every $b \geq \tilde{b}$ and every $n > 0$,

$$Q_n^\beta(J_n > bn) < e^{-\tau n} \tag{4.54}$$

- (*Lower Bound*) For fixed $\beta > 0$ there exist $\tilde{b}, \tilde{n}$ and τ, independent of n but dependent on β and $\mathscr{G}$, such that for every $n > \tilde{n}$ and $0 < b < \tilde{B}$,

$$Q_n^\beta(J_n < bn) < e^{-\tau n} \tag{4.55}$$

The proof of the upper bound is quite simple, and is based on (4.51) together with the choice $A_2 = \{0\}$. The proof of the lower bound, on the contrary, is much more intricate, though still elementary.

4.2.2 High Dimensions and the Lace Expansion

In high dimensions ($d \geq 5$) the excluded volume interaction is no longer influent for the asymptotic behavior of the model. The most powerful tool to derive asymptotic relations for the SAW in high dimensions is the *lace expansion*, introduced by Brydges and Spencer two decades ago [55]. This expansion is based on a procedure which is similar to the *cluster expansion*, and provides a renewal equation for the propagator in $\mathbb{Z}^d$

$$C_n(x) := \sum_{\substack{\omega:0\to x \\ |\omega|=n}} \prod_{0\leq s<t\leq n} (1 - \lambda U_{st}(\omega)) \tag{4.56}$$

where $U_{st}(\omega) = -\delta_{\omega(s),\,\omega(t)}$. One defines

$$K[a,b](\omega) := \prod_{a\leq s<t\leq b} (1 - \lambda U_{st}(\omega)) \tag{4.57}$$

with which the propagator can be written as

$$C_n(x) = \sum_{\substack{\omega:0\to x \\ |\omega|=n}} K[0,n](\omega) \tag{4.58}$$

Expansion of the product in (4.57) gives a sum over all possible graphs Γ. A *graph* Γ is defined as the set of all couples $\{s,t\}$ with $s,t \in [a,b] \subset \mathbb{Z}$ and $s < t$. The set of all graphs is called $\mathcal{B}[a,b]$. For brevity one refers to the couple $\{s,t\}$ as st.

$$K[a,b](\omega) = \sum_{\Gamma\in\mathcal{B}[a,b]} \prod_{st\in\Gamma} (-\lambda U_{st}(\omega)) \tag{4.59}$$

It is now natural — when trying to obtain an expression for C_n involving two-point functions in some sense 'irreducible' — to define new objects by an expression closely similar to (4.59) but restricting the sum only to connected graphs. A graph in $\Gamma \in \mathcal{B}[a,b]$ containing at least two elements ax and yb is said to be *connected* if for every $c \in [a,b]$ there exists $st \in \Gamma$ such that $s < c < t$. The set of such graphs is denoted by $\mathcal{G}[a,b]$.

$$J[a,b](\omega) := \sum_{\Gamma \in \mathcal{G}[a,b]} \prod_{st \in \Gamma} (-\lambda U_{st}(\omega)) \tag{4.60}$$

One then defines the *lace functions*

$$\Pi_m(x) := \sum_{\substack{\omega: 0 \to x \\ |\omega| = n}} J[0,m](\omega) \tag{4.61}$$

With these definitions it is simple to prove the following recursion relation. It suffices to observe that the contribution to $K[0,n]$ from graphs that do not contain couples of type $0x$ is exactly $K[1,n]$. The other contributions can be resummed by decomposing every Γ into its connected components. One then substitutes the expression obtained for $K[0,n]$ into (4.58):

$$C_n = (\chi(\Omega) * C_{n-1}) + \sum_{m=2}^{n} \Pi_m * C_{n-m} \tag{4.62}$$

where Ω is the set of nearest neighbors of the origin, and $*$ stands for the convolution between two discrete measures. By briefly rearranging expression (4.62) and calculating its Fourier transform one obtains an equation that gives an important meaning to the lace functions:

$$\tilde{C}(k,z) = \frac{1}{1 - z\tilde{D}(k) - \tilde{\Pi}(k,z)} \tag{4.63}$$

where

$$\tilde{C}(k,z) = \sum_T z^T \sum_x C(x,T)\, e^{ik\cdot x} \tag{4.64}$$

is the generating functional of the two-point functions, and a similar relations holds for $\tilde{\Pi}(k,z)$ as functions of $\Pi(x,T)$), and where

$$\tilde{D}(k) = \frac{1}{2d} \sum_{\mu \in \Omega} \cos k_\mu \tag{4.65}$$

is the well known function that appears in the propagator of the free walk. Equation (4.63) hence expresses the fact that in some sense the lace functions measure the deviation from Gaussian behavior.

So far the objects formally called laces have not yet appeared. A *lace* is defined as a minimally connected graph, that is a graph such that it becomes disconnected if any couple st is removed. These objects are involved in the demonstration of bonds on the lace functions. These estimates allow the proof — starting from (4.63) — that for $d \geq 5$ there exists a $\lambda_0 > 0$ such that for every excluded volume $\lambda < \lambda_0$

$$\left\langle \omega^2(T) \right\rangle_\lambda = DT(1 + \mathcal{O}(T^{-1/16})) \tag{4.66}$$

where D is a *diffusive* constant.

While this analysis is carried out in momentum space, the authors in [56] examine equation (4.62) with a fixed point argument directly in $\mathbb{Z}^d$. With a somewhat involved argument a point-wise estimate is obtained of the error one makes when approximating $C_n(x)$ with a Gaussian distribution $\varphi_\sigma(x)$ (see [56] for details).

In [57] the lace expansion is investigated and abstracted from an axiomatic point of view. Let P be a set (of properties). A map l which associates a subset $l(S) \subset P$ to every subset $S \subset P$ is called *lace map* if it satisfies the following properties for every $S, S' \subset P$:

$$\begin{array}{l} (a)\ l(S) \subset S \\ (b)\ l(S) \subset S' \subset S \Rightarrow l(S') = l(S) \\ (c)\ l(S) = l(S') \Rightarrow l(S \cup S') = l(S) \end{array} \tag{4.67}$$

A set L for which $l(L) = L$ is called *lace*. The set $C(L)$ of properties *compatible* with a certain lace L is defined as the set of the properties $p \in P$ such that $l(L \cup \{p\}) = L$. With these definitions, the following general result can be obtained, which is called *abstract lace expansion*. Let X be a set of elements, where each element is associated to ('has') a subset of the properties of P. For every L, let $N(L)$ be the number of elements of X that have all the properties in L e none in $C(L)$. Then the number of elements in X which have no properties in P is given by

$$N_0(X) = \sum_{L \in \mathcal{L}} (-1)^{|L|}\, N(L) \tag{4.68}$$

where $\mathcal{L}$ is the set of all laces. The author in [57] guesses that the introduction of a map l — different and perhaps more complicated than that formulated by Brydges and Spencer — could allow a treatment of the asymptotic behavior of the Domb–Joyce model even for $d = 2, 3, 4$.

4.2.3 Ballistic Behavior in $d = 1$

The case $d = 1$ is relatively simple — but not completely trivial —, and in this case it is possible to obtain theorems of type 'law of large numbers' and

'central limit'. In [58] it is proved that for every $\beta \in (0,1)$ there exists a $\theta(\beta)$ such that for every ϵ

$$\lim_{n\to\infty} Q_n^\beta \left(\left| \frac{1}{n} S_n - \theta(\beta) \right| \leq \epsilon \right) = 1 \tag{4.69}$$

where $\theta(\beta)$ is called *speed* of the polymer. This result — which is obtained with a *large deviation* technique — highlights the *ballistic* behavior of the model in one dimension. A more accurate analysis is carried out in [59], and gives the result that, for every $C \in \mathbb{R}$,

$$\lim_{n\to\infty} Q_n^\beta \left(\frac{S_n - n\theta(\beta)}{\sigma(\beta) n^{1/2}} \leq C \right) = \mathcal{N}((-\infty, C]) \tag{4.70}$$

where $\mathcal{N}$ is the normal distribution, and $\sigma(\beta)$ is called the *spread* of the polymer.

The speed of the polymer is expected to go to zero in the limit $\beta \to 0$, because for $\beta = 0$ the behavior is no longer ballistic, but diffusive. More precisely, in [60] it is proved that

$$\lim_{\beta\to 0} \beta^{\frac{1}{3}}\, \theta(\beta) = \alpha > 0 \tag{4.71}$$

Other results — for example in the limit $\beta \to 0$ with $n^{3/2}\beta \to \infty$ — are obtained in [61]. Notice that the particular choice of this limit is not casual, and the exponent 3/2 is not chosen for ease of proof, or for technical reasons: the combination $n^{3/2}\beta$ exactly corresponds to the critical combination $vN^{2-d/2}$ (in the special case $d = 1$) which was derived with a scaling argument in section 4.1.2.

4.3 Universality

The interest in the study of polymer models is focused on the long-chain limit. This is not only because real polymers have high degrees of polymerization. First of all — as it is clear from the other sections of this chapter — some calculations are possible only in the $N \to \infty$ approximation, or at least they are much more feasible in that limit. But the most important aspect of this particular regime is related to a fundamental aspect of statistical (infinite degrees of freedom) systems: *universality*.

4.3.1 Generalities

General statistical system exhibit a vast variety of behaviors, which are specified by the particular choices of a multitude of model parameters, for instance

the microscopic characteristics of the model, the approximations introduced, the type of lattice on which the system is placed. But there exists a particular situation in which much more symmetry and simplicity seems to emerge from this apparently wild plethora of behaviors. When observed in the true thermodynamic limit (infinite degrees of freedom), statistical systems happen to have a point — that is, a particular choice of their parameters — for which they behave in some non-continuous fashion, and many of their measurable quantities diverge. This point is called *critical point* (or *second order phase transition*), and is defined as that particular set of model parameters for which an appropriate *correlation length* diverges. The correlation length is a quantity that measures the typical size of likely fluctuations, that is, the distance along which the system exhibits coherence. More precisely, a *correlation function* is defined, which measures the correlation between points that are distant on the lattice. Away from the critical point, this correlation function is believed to damp exponentially for large distances, so that the correlation length is defined as the inverse factor multiplying the distance at the exponent. At the critical point the divergence of the coherence length says that fluctuations can travel along the 'whole' (infinite!) space.

The correlation length defines a typical scale of the system. It is easy to guess that, when this scale becomes infinite, the small-scale characteristics of the model become irrelevant. For example, if one considers a spin model with medium range interactions — every spin interacts only with spins at most a distance R away — one expects that in the critical limit some global properties of the system will not be dependent on R, because any *fixed* R becomes 'negligible' in the limit where the correlation length becomes infinite. This is indeed a general property of statistical systems at a critical point: they exhibit universality; that is, some properties do not depend on the short-scale specifics of the models, and are therefore called *universal*. The *critical exponents*, which describe the rate of divergence of some observables near the critical point, are examples — and perhaps the most important ones — of universal quantities.

Most qualitative and quantitative features of critical phenomena (that is, statistical systems at the critical point) can be understood in the framework of the *renormalization group* (RG) [62, 63]. This fertile point of view was stimulated by the pioneering work of Kadanoff [64]. The RG idea is to study the behavior of the system under some sort of transformation which neglects the small scale particulars of the model, so that truly universal features become apparent. This transformation induces a *renormalization flow* on the space of theories. The fixed points of this flow correspond to those theories which exhibit scale invariance. For a walk model, this means that the walk is self-similar, and — for example — small loops are as likely to be found as

large ones, at *all* scales. Critical theories flow towards the fixed points: their universal quantities are the fixed point universal quantities[4]. In this picture, the parameter space is divided into 'attraction basins', each corresponding to a fixed point. Critical theories in the same basin flow towards the same point. As a consequence they will share the same universal properties, so that they are said to be in the same *universality class*.

4.3.2 Correspondence with Field Theory

With universality in mind, it is natural to ask whether there exist a field theory whose universal quantities are those of polymer models. For polymers in good solutions with only short-range (in space) interactions the answer is affirmative. This answer is due to the original intuition of de Gennes [66], an subsequent work of others [see [43] and reference therein].

Following [48], one considers the Edwards model. The delta function is linearized by introducing auxiliary imaginary fields. The Laplace transform of the two-point function can then be calculated. Then, a variant of the replica trick is applied, so that one introduces additional fields ϕ, but at the same time the limit $n \to 0$ — where n is the number of components of ϕ — must be taken, in order to delete a det factor. Finally the integration in the auxiliary field can be carried out, so that one end up with the usual two-point function of $g\phi^4$ theory. The remarkable conclusion is that the statistical properties of polymers are related to the properties of the $g\phi^4$ field theory in the 'unphysical' limit $N = 0$. More precisely, universal critical properties of polymer chains are related to the singularities of correlation functions of the $N = 0$ Φ^4-theory in the massless (critical) limit [see chapter 28 in [48] for details][5].

As a consequence, all critical exponents for polymer models are those obtained for the zero-component field theory. In two dimensions, a very accurate (five-loop) calculation of perturbative series for critical quantities can be found in [68]. Also, in two dimensions there are other methods which are of great help in extracting critical quantities, the most important of which are conformal invariance methods [see [67] for an introduction]. The study of conformal invariance properties of polymers, whose central charge — in the terminology of conformal field theory — is zero, has been done in [69]. The two exponents ν and γ are known exactly from $N = 0$ field theory in

[4]For a comprehensive review of critical phenomena treated with renormalization group techniques, see [65].

[5]Other derivations of this connection can be found in literature, see for instance [67] chapter 9, where the correspondence between SAWs and a lattice model of interacting spins is showed in a 'geometric' fashion.

two dimension [70]:

$$\nu = \frac{3}{4} \qquad \gamma = \frac{43}{32} \tag{4.72}$$

On the other hand, the leading non-analytic correction to scaling exponent Δ_1 is still the subject of controversy and debate, since its value as derived with conformal-invariance methods (11/16, [69]) differs from that obtained with Coulomb-gas arguments (3/2, [71]). This uncertainty does not even tell whether the leading non-analytic correction is more important than the first analytic correction.

Another thing that is exactly known from field theory is the upper critical dimension, which is 4, as conjectured for polymers by other methods. Dimension $d = 4$ is the dimension above which excluded-volume polymers behave asymptotically exactly as free (Gaussian) chains. Indeed, this is very easy to explain heuristically, since Brownian motion paths have dimension 2, and one expects that two such paths almost never interact in $d > 2 + 2$: a 4-dimensional walker has a 'large' volume to explore, and will rarely come back to a place it already visited.

So the critical universal properties of polymers in good solutions are described by the critical $N = 0$ field theory. But what happens if one considers the properties of poor- or θ-solvent polymers? As sketched in chapter 2, in these regimes two-body interactions are no longer enough, and one has to include at least three-body interactions. At a field-theoretic level, this translates into the introduction of a ϕ^6 term. The qualitative picture of the critical behavior then changes substantially, because now a tricritical point appears. The upper critical dimension for the ϕ^6 operator is 3, so one concludes that while over $d = 3$ the behavior at the θ-point is strictly Gaussian (with logarithmic corrections in three dimensions), in low dimensions θ-point polymers do not behave like random walks.

From the foregoing discussion it should be clear that it is very important to distinguish what properties of the system are universal. It is exactly by means of universality, that very simple models — such as the Domb–Joyce model — can claim to describe real physical systems — such as real polymers, that are governed by manifold kinds of interactions, and of which there exist a variety of chemically and structurally different realizations. One should not be tempted to try and match the behavior of a real finite-length (that is, non-critical) polymeric chain by carefully adjusting the interaction strength in the Edwards model (for instance by scaling it like $L^{d/2-2}$, see [5], because only universal quantities can be obtained from a simplified model. Actually, there exist various quantities — for instance the interpenetration ratio and some effective exponents — that are seen to approach their asymptotic value from above or from below depending on the particular model chosen. This

is a sign of the non-universality of correction-to-scaling amplitudes, even the sign of which is non-universal [see [72] and reference therein].

4.4 Crossover Behavior

As it is clear from the discussion in the previous sections, the Domb–Joyce model is expected to have a non-trivial SAW-like behavior in the critical limit for every $v \neq 0$. But when $v = 0$ the behavior suddenly becomes exactly that of the RW, because then not only is the model in the Gaussian universality class, but it becomes exactly the same model as the RW. Within this picture, the interest is naturally turned to the study of the *crossover* between the two types of behavior. This means that one is interested in how the observables change their behavior from that of one fixed point to that of the other, by letting v run from 0 to non-zero values. The parameter v is therefore called *running parameter*.

4.4.1 Generalities

It is clear that in the strict critical limit $N \to \infty$ the crossover must be abrupt — that is, discontinuous —, because universal quantities have values that do not depend on v when $v > 0$, but they assume different values for $v = 0$. On the contrary, in the non critical region $N < \infty$ the crossover is expected to be smooth[6]. Loosely speaking, this smoothness is obvious for walk models, because the non-critical walk is of finite size, then it does not have but a finite number of degrees of freedom, which can not sum up but to a smooth function. In general, a (small) distance from the critical point contributes to the critical asymptotic scaling laws through correction to scaling terms. It is exactly these correction terms that are responsible of the non-critical smoothness, because they sum up to give a gradual — that is, continuous — transition between the two regimes.

Manifold quantities can be defined in the crossover region, which are of help in matching the non-critical behavior of polymer models with that of real polymers — here we refer to chain models, but similar quantities are introduced in the study of non-critical crossover aspects of other models. For example, the so-called *effective exponents* measure the effective 'asymptotic' behavior of the chain, which for finite v and N behaves 'halfway' between the RW and the SAW. For instance, one can define an effective swelling exponent

[6] See, for instance, [73].

by

$$\nu_{eff}(N, v) := \frac{1}{2 \log 2} \log \frac{\langle R_e^2(2N, v) \rangle}{\langle R_e^2(N, v) \rangle} \tag{4.73}$$

From the point of view of the Wilson renormalization group, crossover phenomena are explained by the competition between two fixed points: the Gaussian (classical, mean-field) fixed point and the non-classical (Wilson-Fisher) fixed point. The most important issue in the study of crossover phenomena is to determine whether one is able to define crossover scaling functions that are universal. The 'range' at which the crossover takes place (the crossover region) is identified by the *Ginzburg criterion* [74]. If the reduced temperature $|t|$ is greater than the Ginzburg number G — which of course depends on the model, through its running parameter — the system shows an approximate mean-field behavior, while non-classical critical behavior begins to manifest itself when $|t| < G$. The Ginzburg number G is then a measure of the relevance of space fluctuations, which drive the system to criticality. When varying t while keeping G fixed, the type of crossover observed is non-universal, since it depends on the details of the model. The description of such a crossover is usually performed through effective quantities by phenomenological models. As was pointed out by Bagnuls and Bervillier [75], a universal behavior can be obtained only in a properly defined critical limit. In fact, in order to observe universality, one has to let the reduced temperature go to zero — that is, observe the system at criticality. Moreover, if one wants to observe a non-trivial crossover limit, the Ginzburg number has to go to zero, too, so that one keeps the model in the crossover region. A proper formulation of the problem is due to Luijten, Blöte and Binder [76], who argue that a universal behavior can be observed by letting $t \to 0$ and $G \to 0$ while keeping their ratio fixed. This particular limit will be called *critical crossover limit* [77].

4.4.2 Crossover Functions in the Domb–Joyce Model

The first task in studying the crossover behavior of the Domb–Joyce model is to find the function of the running parameter v and the 'inverse temperature' N which corresponds to the Ginzburg combination. By the scaling analysis performed on the Edwards model [see equation (4.27)], the crossover scaling variable is expected to be $x = vN$ in two dimensions. Actually, from the rigorous expansion performed directly on the lattice for the Domb–Joyce model [see section 4.4.3], the expansion variable turns out to be exactly that combination of the bare parameters. Notice, in passing, that the criticality of $d = 4$ has a clear effect on the crossover variable, which is $vN^{2-d/2}$ in

arbitrary dimension. In fact, the critical crossover limit with respect to this expression is meaningless for $d \geq 4$, because it would be impossible to keep x fixed in the large-N limit.

Hence, with the foregoing definition of x, the critical crossover limit is defined as

$$\text{CCL:} \quad \begin{cases} N \to \infty \\ v \to 0 \\ vN \ \text{ fixed} \end{cases} \tag{4.74}$$

In this limit the scaling argument of equation (4.31) predicts

$$\frac{\langle R_e^2 \rangle_N}{N} = f(x) \tag{4.75}$$

It is this very function $f(x)$ which exhibits universality. Here, universality does not mean the values of $f(x)$ are universal. The function $f(x)$ is universal apart from an overall rescaling of the two axes, that is of x and of $\langle R_e^2 \rangle$. This is actually quite natural, since x is a combination of *bare* parameters. Moreover, in the critical crossover limit effective exponents like the one defined in (4.73) become universal. The scaling form (4.75), which has been derived by a scaling argument, is verified to first order in x in an expansion around $x = 0$ in section 4.4.3.

If one compares the crossover scaling functions f_1 and f_2 of two different models in the same universality class, they will differ by a rescaling of the kind described above.

$$f_1(x) = a f_2(bx) \tag{4.76}$$

This means that once the two non-universal scaling factors a and b are known, and the axes are rescaled appropriately, the two functions collapse into the same. For instance, let f_1 belong to the Domb–Joyce model, and let f_2 be the crossover function as computed from field theory. If one knows, say, f_2 to second order, and f_1 from first order, then the second order coefficient of f_1 can be extracted, because the two non-universal factors are fixed by comparison of the two functions to zeroth and first order. f_1 to first order for the Domb–Joyce model will be calculated in the following section. f_2 can be calculated from field theory by using the results in [68] and the method described in [78].

4.4.3 Cluster Expansion

In order to study the crossover functions in the vicinity of the Gaussian fixed point — that is, near the random walk — a *cluster expansion* is used. This expansion naturally turns out to be an expansion in the crossover parameter

$x = vN^{2-d/2}$. Here we will restrict ourselves to the two-dimensional case, so that $x = vN$. A cluster expansion for the Domb–Joyce model on the cubic lattice in $d = 3$ has been carried out by Domb and Joyce — in the original paper [12] where the model was first proposed — and by Barrett and Domb in [79], using the results for random walks of reference [80]. Very precise results — up to three loops — are obtained in these works for the three-dimensional Domb–Joyce model.

Here we wish to compute the first order (one loop) diagram of the expansion for the case of interest, that is for $d = 2$. The tool of choice in performing a cluster expansion for a walk model on the lattice is the use of *lattice generating functions*. Following the notation of [12], we shall denote by q the coordination number of the lattice ($q = 4$ for the square lattice), by $P_n(\mathbf{l})$ the probability for a random walk of length n to end at lattice site $\mathbf{l}$, and by u_n the number of n-step loops (returns to the origin). Moreover, let c_n be the number of n-step random walks, and let $c_n(\mathbf{l})$ be the number of such walks that end at lattice site $\mathbf{l}$. One can then define the following generating functions, where $x \equiv qt$

$$\sum_{n=0}^{\infty} c_n t^n = \sum_{n=0}^{\infty} q^n t^n = (1-qt)^{-1} = (1-x)^{-1} =: P(x) \tag{4.77}$$

$$\sum_{n=2}^{\infty} u_n t^n = \sum_{n=2}^{\infty} P_n(\mathbf{0}) x^n =: R(x) \tag{4.78}$$

$$\sum_{n=0}^{\infty} c_n(\mathbf{l}) t^n = \sum_{n=0}^{\infty} P_n(\mathbf{l}) x^n =: P(\mathbf{l}, x) \tag{4.79}$$

so that $P(\mathbf{0}, x) = 1 + R(x)$. It is useful to introduce the *generating polynomial* of the lattice, which for the 2-dimensional square lattice is

$$\Psi(\mathbf{x}) \equiv \Psi(x_1, x_2) := \frac{1}{4}\left(x_1 + x_1^{-1} + x_2 + x_2^{-1}\right) \tag{4.80}$$

It is easy to see that

$$\sum_{\mathbf{l}} P_n(\mathbf{l}) \mathbf{x}^{\mathbf{l}} = \Psi(\mathbf{x})^n \tag{4.81}$$

where $\mathbf{l}$ is a 2-dimensional multi-index, and $\mathbf{x}^{\mathbf{l}} \equiv x_1^{l_1} x_2^{l_2}$. Hence

$$\sum_{\mathbf{l}} P(\mathbf{l}, x) \mathbf{x}^{\mathbf{l}} = \left(1 - x\Psi(\mathbf{x})\right)^{-1} =: Q(\mathbf{x}, x) \tag{4.82}$$

Notice that all the foregoing quantities are completely analogous to those usually defined for the random walk (see for instance [8]) in terms of β, β_c, $\hat{p}$,

etc. in momentum space description. We shall denote the partition function of the Domb–Joyce model in the v-expansion by

$$c_n(v) := \sum_{\omega} \left[1 - v \sum_{i,j} \delta_{\omega_i,\omega_j} + \cdots \right] \tag{4.83}$$

This appears as a natural generalization of the number c_n of random walks to the interacting case. Consequently, the interacting generating function of $c_n(v)$ will be defined as

$$P(x,v) := \sum_{n=0}^{\infty} c_n(v) t^n \tag{4.84}$$

The zeroth order contribution to this object is the free function $P(x)$.

The term that contributes to first order in v is obtained by considering the one-loop diagram One must sum over all possible i and j such that the i-th and j-th points of the walk are coincident on the lattice. For each chain this gives

$$\sum_{n_1+n_2=n} (n_1+1) c_{n_1} u_{n_2} \tag{4.85}$$

and it is easy to see that this is the n-th coefficient in the t-expansion of

$$R(x) \frac{d}{dx} \left[x P(x) \right] = P(x)^2 R(x) \tag{4.86}$$

This is the term that accounts for the first order of the partition function. Since we are interested in the end-to-end distance, also a slightly different object must be evaluated, which keeps account of the position of the end point of each chain. By multiplying $c_n(\mathbf{l})$ by $\mathbf{x}^l$ and summing over $\mathbf{l}$ it is readily found that the generating one-loop term is

$$Q(\mathbf{x},x)^2 Q(x) \tag{4.87}$$

From such a generating function $G(\mathbf{x},x)$, the mean square end-to-end distance is extracted by the operator

$$\left[\left(x_1 \frac{\partial}{\partial x_1} \right)^2 + \left(x_2 \frac{\partial}{\partial x_2} \right)^2 \right] G(\mathbf{x},x) \Bigg|_{x_1=1,\, x_2=1} \tag{4.88}$$

After some algebra one obtains, to first order in v,

$$\sum_{n=0}^{\infty} c_n t^n = P + v P^2 R + \cdots \tag{4.89}$$

$$\sum_{n=1}^{\infty} c_n(v) \left\langle R_e^2 \right\rangle_n t^n = xP^2 + 2vxP^3R + \cdots \tag{4.90}$$

By expanding the expression for the end-to-end distance to first order in v one gets

$$\begin{aligned} \left\langle R_e^2 \right\rangle_n &= \frac{\sum_\omega |\omega(n)|^2 \left[1 - v \sum_{ij} \delta_{\omega_i,\omega_j} + \cdots\right]}{\sum_\omega \left[1 - v \sum_{ij} \delta_{\omega_i,\omega_j} + \cdots\right]} \\ &= \left\langle R_e^2 \right\rangle_n^{(0)} - v \left[\left\langle |\omega(n)|^2 \sum_{ij} \delta_{\omega_i,\omega_j} \right\rangle_n^{(0)} + \left\langle |\omega(n)|^2 \right\rangle_n^{(0)} \left\langle \sum_{ij} \delta_{\omega_i,\omega_j} \right\rangle_n^{(0)} \right] \end{aligned} \tag{4.91}$$

where $\langle \cdot \rangle_n^{(0)}$ means expectation with respect to the flat RW measure. The zeroth order gives the random walk value n, which can be obtained in this framework by studying the n-th derivative of the free term $xP(x)^2$ in the generating function. The terms which contribute to first order in v are respectively the connected and disconnected components. Since the expression in (4.90) only gives the numerator of (4.91), it corresponds only to the connected component. By equation (4.89), the disconnected component is readily evaluated as nP^2R, so that the first order contribution to the end-to-end distance is given by the n-th term in the expression for $2xP^3R$ minus n times the n-th term in the expression for P^2R. Let L_n denote the operator that extracts the n-th term:

$$L_n \left[F(x)\right] := \frac{1}{n!} \frac{d^n}{dx^n} F(x) \bigg|_{x=0} \tag{4.92}$$

Then the mean square end-to-end distance to first order in v is

$$\left\langle R_e^2 \right\rangle_n = n - v \left(L_n \left[2xP^3R\right] - nL_n \left[P^2R\right] \right) + \cdots \tag{4.93}$$

In two dimensions [80], the loop generating function near the singularity[7] $x = 1$ is given by

$$R(x) = -\frac{1}{\pi} \log(1 - x) - 1 \tag{4.94}$$

With this result, we are able to compute the connected and disconnected components

$$A = L_n \left[2xP^3R\right] = L_n \left[\frac{2x}{(1-x)^3} \left(\frac{1}{\pi} \log(1 - x) - 1 \right) \right] \tag{4.95}$$

[7]Since we are interested in the behavior for n large, only the behavior near the singularity is needed.

$$B = L_n\left[P^2 R\right] = L_n\left[\frac{1}{(1-x)^2}\left(\frac{1}{\pi}\log(1-x) - 1\right)\right] \tag{4.96}$$

In order to do so, we have to compute the n-th derivative of the objects in squared parentheses. They are in the form $h(x) = f(x)g(x)$, so we shall make use of Leibniz formula

$$\frac{d^n}{dx^n}\left[f(x)g(x)\right] = \sum_{j=0}^{n}\binom{n}{j}\frac{d^j}{dx^j}f(x)\frac{d^{n-j}}{dx^{n-j}}g(x) \tag{4.97}$$

The term 1 which appears together with the log in the parentheses does not happen to contribute to leading order[8]; so we will neglect it. So, for A we have

$$\left.\frac{d^k}{dx^k}\left[\frac{x}{(1-x)^3}\right]\right|_{x=0} = \frac{(k+2)!}{2} - (k+1)! \tag{4.98}$$

$$\left.\frac{d^k}{dx^k}\left[\log(1-x)\right]\right|_{x=0} = -(k-1)! \quad \text{for } k > 0 \tag{4.99}$$

so that

$$\begin{aligned} L_n\left[\frac{x}{(1-x)^3}\log(1-x)\right] &\sim \frac{1}{n!}\sum_{j=0}^{n-1}\binom{n}{j}\left[\frac{(j+2)!}{2} - (j+1)!\right]\left[-(n-j-1)!\right] \\ &= -\sum_{j=0}^{n-1}\frac{(j+1)!}{(n-j)!j!}\left(\frac{j+2}{2} - 1\right)(n-j-1) \\ &= -\sum_{j=0}^{n-1}\frac{j+1}{n-j}\frac{j}{2} \\ &= -\sum_{j=0}^{n-1}\left[\frac{n(n+1)}{2}\frac{1}{n-j} - \frac{j+1}{2} - \frac{n}{2}\right] \end{aligned} \tag{4.100}$$

where $\sim$ has been used to denote that the two expressions are equal at leading order in n. The last two terms in the last sum of equation (4.100) can be evaluated exactly, and respectively give $n^2/2$ and $n(n+1)/4$. The remaining term has a logarithmic divergence, and can be evaluated by transforming the

[8]The fact is, that use of Leibniz formula gives the k-th derivatives of the object, so that the constant term 1 gives a null contribution, apart from the 0-th term. But when this term is evaluated at $x = 0$, what remains is only the n-th derivative of the other factor (that is, $f(x)$), whose behavior is subleading — as it is clear from the following discussion.

sum into an integral; it gives a term $-\log n$. Finally, for the connected term we have the following leading behavior

$$A \sim \frac{n^2}{\pi}\left(\log n - \frac{3}{2}\right) \tag{4.101}$$

The logarithmic divergence must not appear in the final expression, so we expect it to be canceled by the disconnected term.

By a completely analogous procedure, the disconnected part B can be evaluated. It turns out to be

$$nB \sim \frac{n^2}{\pi}\left(\log n - 1\right) \tag{4.102}$$

As expected, the logarithmic term is canceled, and finally one is left with the following one-loop expression for the expansion factor:

$$\frac{\langle R_e^2\rangle_n}{n} = 1 + \frac{1}{2\pi} vn + \cdots \tag{4.103}$$

Chapter 5

Numerical Results

We simulated the two-dimensional Domb–Joyce model by means of the modified optimized Pivot algorithm which is described in chapter 3. The two free parameters of the simulated lattice model that are to be chosen for every Monte Carlo run are the excluded volume parameter v and the walk length N. The parameter space to be probed is then the two-dimensional plane $v - N$. Two different classes of paths on this plane have been followed in the simulations, each corresponding to a different non-trivial limiting regime of the model.

Firstly, the critical scaling limit has been investigated, that is the limit in which N goes to infinity, while v is kept fixed. This corresponds to the critical limit which is accounted for in section 5.1. It is important, when making assertions by looking at plots in this regime, to remember that the two limits $N \to \infty$ and $v \to 0$ do not commute in this regime. Indeed, one expects the scaling limit to be of the Wilson-Fisher type for every fixed v when N goes to infinity, so that $\lim_{v\to 0} \lim_{N\to\infty}$ is expected to show the same non-classical scaling. On the other hand, when letting v go to zero one gets the Gaussian behavior for any fixed N, so that the limit $\lim_{N\to\infty} \lim_{v\to 0}$ corresponds to the mean-field regime.

Secondly, we studied the critical crossover limit of the model. This regime corresponds to taking $N \to \infty$ and $v \to 0$ strictly *at the same time*, while keeping the combination $x = vN$ fixed. This regime is described in section 5.2.

For a correct analysis of the data obtained from the simulations, we relied on the time-series analysis results described in chapter 3. The number of MC iterations executed for every coordinate in the parameter plane (v, N) is 500,000. These iterations are to be intended as consecutive, that is they are the run length of a single Markov chain, starting from a single initial configuration.

Not every sample was recorded, in order to reduce the correlation time, and to save disk space; one sample out of 100 was saved and used in the analysis. The integrated autocorrelation time has been measured with the automatic windowing procedure, where a windowing factor $c = 8$ was taken. This procedure turned out to produce nearly independent samples, that is, the resulting integrated autocorrelation time was slightly above 0.5.

In order to reduce the bias due to the initial transient, the first few samples of every run were discarded. In order to be assured about the consistency of the method, the exponential autocorrelation time was computed both for data including and not including the samples that were chosen to be discarded. It turned out that for every observable and every point chosen in the parameter plane[1] (the estimator of) the exponential autocorrelation time was $\tau_{exp} < 10$, so that a few tens discarded samples should suffice. Indeed, after discarding the first 100 samples (that is, the first 10,000 MC iterations) the exponential autocorrelation time became 1 for every observable[2].

All fits in this chapter were obtained by the *Levenberg–Marquardt method* [81, 82], which seems to be the best tradeoff between speed and accuracy. It must be noticed that non-linear fitting algorithms are not *exact*, since they are not guaranteed to give the exact solution to the residual minimization problem. Moreover, they may suffer from slow convergence and — even worse — the dreaded metastability. In order to rule out, as far as it was possible, metastability, we performed the fits several times, each time starting from a different initialization of the parameters. No metastable attractive regions have ever emerged from the fits.

5.1 Scaling Behavior

The hope in studying the fixed-v behavior of the model for N large was to verify that criticality was SAW-like for every $v > 0$. The difficulty here is, that the estimation of critical behavior via extrapolation of MC data (as well as exact enumeration data) is largely subjective, and heavily relies on assumptions of smoothness, and on the expressions chosen for describing the corrections to the critical behavior. According to general renormalization group arguments [83], observables in the SAW (that is, in the extreme critical

[1] The exponential autocorrelation time is expected to depend mainly on the choice of N, being higher for N large, because more pivot moves are required to 'completely bend' a long chain. Yet, quite surprisingly, the exponential autocorrelation time measured on the whole set of samples does not seem to depend strongly on N.

[2] The exponential autocorrelation time is probably now much less than unity, but the estimator operatively computes it as the least integer greater than the 'definition value'.

Domb–Joyce model) near criticality are expected to follow a scaling law of the following kind

$$\langle \mathcal{O} \rangle_N = A N^p \left[1 + \frac{a_1}{N} + \frac{a_2}{N^2} + \cdots + \frac{b_0}{N^{\Delta_1}} + \frac{b_1}{N^{\Delta_1+1}} + \cdots \right] \tag{5.1}$$

where the Δ exponents are called *correction to scaling exponents*. They describe the *non-analytic* corrections which — in the RG picture — are due to the approach to the fixed point, that is, to irrelevant operators. Scaling form (5.1) also contains *analytic* corrections to scaling, which are due to the approach to criticality. The first non-analytic correction exponent Δ_1 is expected to be universal. Notice that knowledge of this exponent is of great importance when fitting MC data, because it is the most important ingredient in a systematic-error free extrapolation. Actually, systematic errors in such extrapolations are *always* present, but fitting versus an expression like (5.1) truncated at order Δ_1 is a great improvement over a naive power scaling ansatz.

In most three-dimensional systems, $\Delta_1 \approx 0.5$, so that non-analytic corrections are far more relevant than analytic ones. The problem is much more subtle in two-dimensions, and both theoretical and numerical analyses have failed in clarifying whether $\Delta_1 \gtrless 1$ [see chapter 4]. From the numerical point of view, measuring correction to scaling exponents is difficult. This is because one is actually measuring an *effective* exponent — as is always the case in fitting numerical data — which incorporates, in an averaging fashion, all corrections to scaling that lie further than the truncation point in the fitting function. The fact is, that this could lead to largely underestimated exponents, because — for example — two corrections to scaling with exponents of the same order — as is the case for the first analytic and the first non-analytic corrections —, but with amplitudes of different signs, could combine to give a phenomenological exponent near 0.

To complicate things, in the Domb–Joyce model another parameter is present beside N. This means that every amplitude in (5.1) will depend on v. Studying a one-parameter family of Hamiltonians, like the Domb–Joyce model, can be interesting, because comparison of exponent estimates for different values of v might give an idea of the extent to which measured effective exponents depend on the non-universal amplitudes, and on the choice of the fitting function (as observed in [84]). We tried such an approach, at least to make this subtle cancellation of behaviors apparent. We chose two different values of the excluded volume parameter (1 and 0.02), and measured the effective correction to scaling exponent Δ_{eff} by fitting the mean square

end-to-end distance versus[3]

$$f(N) = AN^p \left(1 + \frac{B}{N^{\Delta_{eff}}}\right) \tag{5.2}$$

where p was fixed to the theoretical value $3/2$. The two values we obtained did differ from each other (0.8 for $v = 0.02$, and 1.5 for $v = 1$), but the second value had an extremely large error (over 100%). These estimates do not help in understanding more deeply the problem.

In [85] for the first time strong numerical evidence in favor of one of the two theoretical values of Δ_1 was given, by careful study of SAWs on the lattice. The authors conclude that $\Delta_1 = 3/2$ in two dimensions, and this is the value we will be using for fits in this section, also because it is the value that gives best fits.

We simulated the model for three fixed values of v, namely $v = 0.01,\ 1,\ 10$. The values of N were taken in the range $[10, 1500]$ covering all even numbers. A dense covering of the range surprisingly appeared to be more important — for the measure of the leading exponent ν — than the distance from criticality — that is, the cutoff N_{max}. A representative plot of the observable is reported in figure 5.1. The fitting function was chosen to be

$$f(N) = AN^p \left(1 + BN^{-1} + CN^{-3/2}\right) \tag{5.3}$$

and the fit was performed via variation of A, B, C and p. The resulting values of the exponent 2ν are reported in table 5.1.

v	2ν
0.02	1.506(24)
1	1.515(21)
10	1.499(25)

Table 5.1: The swelling exponent 2ν for different values of the excluded volume. Its theoretical value is 1.5.

These values are in excellent agreement with the theoretical prediction: the behavior is the critical SAW-like behavior even for relatively small values of v. No trends can be observed in the measured value as the interaction strength is varied, meaning that the fitting function accurately describes corrections to scaling, so that we are measuring a quantity that is very close to the asymptotic one. As a consequence, one expects that this exponent does indeed exhibit a discontinuity for $v = 0$.

[3]In order to get better results, one could try and include also higher order terms in the fitting function, such as the first analytic correction. Nonetheless, such an improved fit turned out to give much wilder statistical errors, because another parameter is introduced.

5.2 Crossover Behavior

We wish now to qualitatively characterize the critical crossover function in two dimensions, and to check quantitatively their behavior for small values of the crossover variable. As described in the chapter 4, crossover functions are the scaling functions that describe observables in the critical crossover limit. Here, we will focus on the crossover function for the end-to-end distance:

$$f(x) \equiv f(vN) := \frac{\langle R_e^2(v,N)\rangle}{N} \quad \text{for } N \to \infty,\ v \to 0 \text{ and } x \equiv vN \text{ fixed} \tag{5.4}$$

First of all, in order to get a global picture of the crossover function, we simulated the Domb–Joyce model for increasing values of N in the range [800,1200] and for values of v distributed in a logarithmic fashion in the range [0.001,100]. The crossover function is expected to converge to 1 in the small-x limit — this corresponds to the random walk regime — and to diverge in the large-x limit — the self-avoiding walk regime. A plot of this function in the ranges described above is in figure 5.2.

The convergence to a single function for fixed x in the critical limit is very evident even for such relatively small values of N. The accordance is better for lower values of x. The asymptotic regimes are clearly satisfied: the function is seen to diverge for large values of the scaling variable, while it is nearly equal to 1 for $x < 0.1$.

To move on to a more quantitative level, we focus on the small-x regime. We simulated the model for values of N up to 16000, and for x in the range [0.1,2]. The results are plotted in figure 5.3, where they are presented together with the theoretical one-loop prediction. In this range the function is seen to be linear with good approximation. Nonetheless, we performed both linear and quadratic fits, and the latter gave much better results, the former displaying slight systematic underestimation of the linear coefficient. The quadratic fits have been done with the following fitting function

$$f(x) = a + bx + cx^2 \tag{5.5}$$

where all three parameters were varied in the fit. Notice that the correlation between the linear and the quadratic terms in such a fit turns out to be rather high, around 0.9. The results for the linear coefficient b are reported in table 5.2, and they are to be compared with the theoretical prediction of equation (4.103), which is

$$b = \frac{1}{2\pi} \approx 0.159 \tag{5.6}$$

N	b
250	0.150(20)
500	0.168(18)
1000	0.157(26)
2000	0.167(21)
4000	0.158(17)
8000	0.187(24)
16000	0.162(24)

Table 5.2: The linear coefficient of the crossover function, as obtained from numerical data for different values of N. Its theoretical value is ≈ 0.159

The numerical values all agree with the theoretical prediction, and no trends for increasing N are apparent. This last circumstance strengthens the result that the crossover functions only depend on the combination vN, and corrections to crossover behavior are not detectable in this range. Notice that the theoretical curve in figure 5.3 does not have free parameters, so the accordance between theory and simulations is not mediated by any numerically-fitted parameter.

Finally, we turn to the observation of the effective exponent

$$\nu_{eff}(N, v) := \frac{1}{2 \log 2} \log \frac{\langle R_e^2(2N, v) \rangle}{\langle R_e^2(N, v) \rangle} \tag{5.7}$$

As explained in chapter 4, this quantity is expected to become universal in the critical crossover limit. Then, we expect — as for the crossover function — to observe that the curves for different values of N converge to the same scaling curve for N large. We simulated the model for $N = 500 \cdot 2^k$ with $k = 0, 1, 2, 3, 4$, and for values of v logarithmically distributed in the range [0.0001,2]. Then, these values of N and v were combined to plot the quantity versus x. A plot of the effective swelling exponent is reported in figure 5.4. The curves corresponding to different values of N nicely converge onto the same curve, as expected. The overall behavior is qualitatively the one expected: the curves for different values of N reduce to the same curve in the critical limit. Actually, the coincidence between the functions is very close even for low values of N, in the middle range. The curve also displays monotonicity to a good extent, as expected, since the Domb–Joyce model is always in the non-broken phase [86].

When one considers the curve quantitatively, the accordance with theory is less precise. In the small x regime, the exponent shows the correct limiting RW value 1/2. In the large x regime, the effective exponent is expected to

approach the critical SAW value 3/4, which does not seem to be the limiting value of the curve in figure. In that region the values of v are of order unity, and perhaps the disagreement is due to corrections in v, which sum up to important deviations from universal behavior.

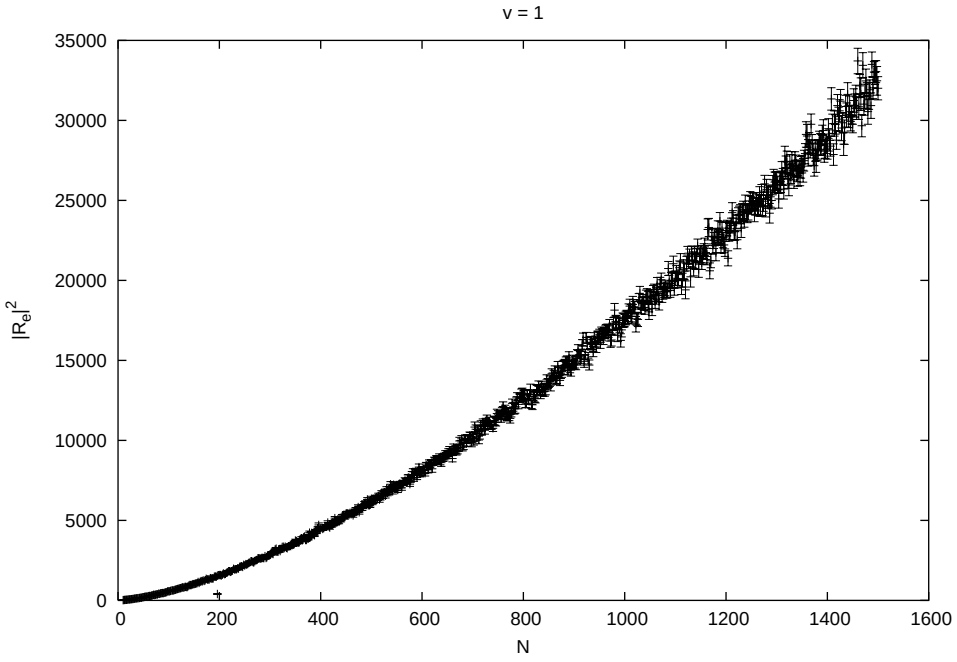

Figure 5.1: A plot of the mean square end-to-end distance for $v = 1$.

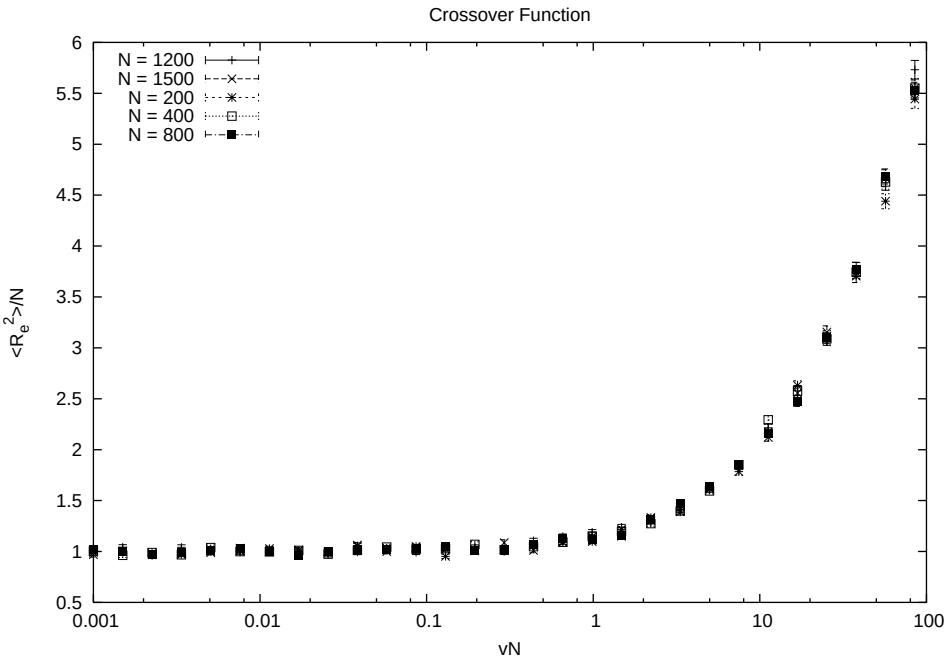

Figure 5.2: The crossover function. The scale on the x axis is logarithmic.

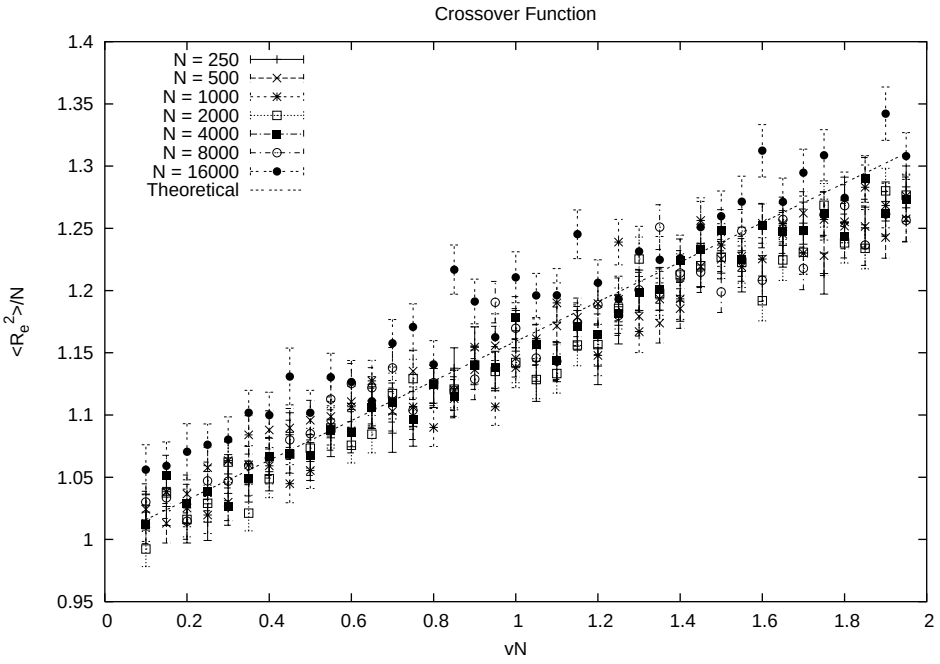

Figure 5.3: The crossover function for small values of the variable x. The line represents the one-loop (linear) theoretical prediction, which has no fitted parameters.

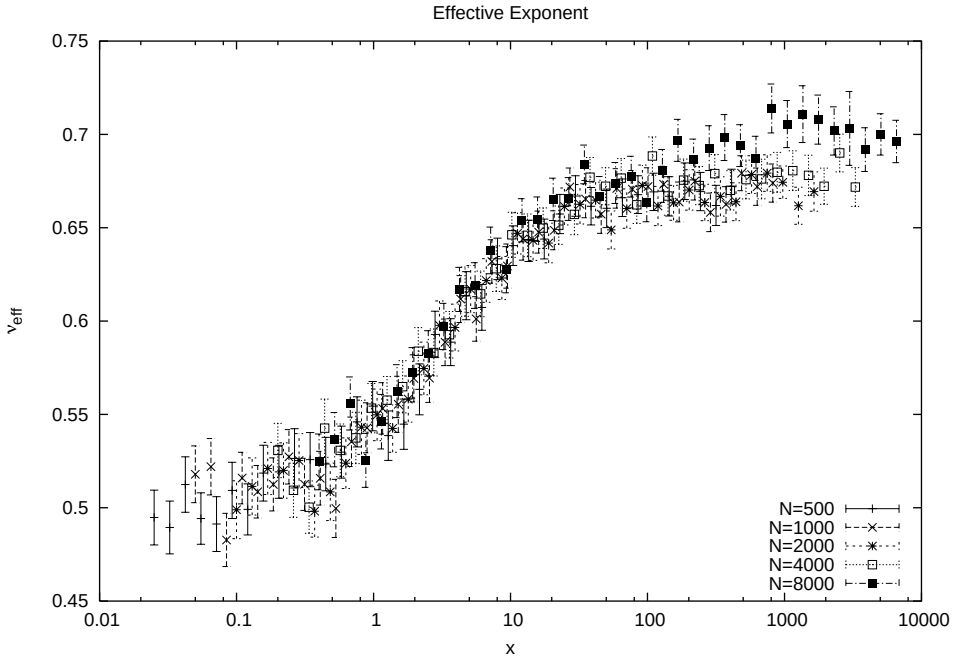

Figure 5.4: A plot of the effective swelling exponent versus the crossover variable.

Chapter 6

Conclusions

In this work we analyzed the critical behavior of the Domb–Joyce model, as a model of polymers in good dilute solutions. We reviewed previous results — which mainly refer to other models belonging to the same universality class, or they are obtained for the *three-dimensional* Domb–Joyce model —, calculated the one-loop crosover function directly for the Domb–Joyce model on the square lattice, and checked theoretical predictions through Monte Carlo simulations. We focused on the end-to-end distance, and analyzed two different limits: the scaling limit and the critical crossover limit. The results for the scaling limit are in perfect accordance with the expected behavior: the model belongs to the SAW universality class for every strictly positive value of the interaction strength. In the critical crossover limit we checked that the non-critical curves actually converge to a universal function, and displayed such function in a wide range of the crossover variable. Moreover, we obtained estimates for linear coefficient of the crossover function in the random walk regime, that is, for small values of the variable. The check between numerical data and theoretical calculations is good.

Future work on the Domb–Joyce model should perhaps be directed towards the study of other observables, first of all the second virial coefficient, in order to study critical amplitude rations, which are expected to be universal [87]. From the theoretical point of view, the calculation we performed in two dimensions gives the non-universal amplitude that is needed to match the critical crossover behavior of the lattice model with field-theoretical predictions. Thus, future theoretical work should be devoted to obtaining high order perturbation series from field theory, and matching them against numerical Monte Carlo data. Also, another interesting extension of the present work is towards the study of branched chains and polymers with different topological structures.

Bibliography

[1] P.J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, Ithaca, NY, 1953.

[2] H. Yamakawa. *Modern Theory of Polymer Solutions*. Harper & Row, New York, 1971.

[3] P.-G. de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, Ithaca, NY, 1979.

[4] K.F. Freed. *Renormalization Group Theory of Macromolecules*. John Wiley & Sons, New York, 1987.

[5] J. des Cloizeaux, G. Jannink. *Les Polymères en Solution*. Les Editions de Physique, Les Ulis, 1990.

[6] H. Fujita. *Polymer Solutions*. Elsevier, New York, 1990.

[7] G.F. Lawler. *Intersections of Random Walks*. Birkhäuser, Boston, 1991.

[8] C. Itzykson, J.-M. Drouffe. *Statistical Field Theory Vol.1: from Brownian Motion to Renormalization and Lattice Gauge Theory*. Cambridge University Press, 1989.

[9] N. Madras, G. Slade. *The Self-Avoiding Walk*. Birkhäuser, 1993.

[10] J.M. Hammersley, D.J.A. Welsh. Further results on the rate of convergence to the connective constant of the hypercubical lattice. *Quart. J. Math. Oxford*, 13:108, 1962.

[11] H. Kesten. On the number of self-avoiding walks. II. *J. Math. Phys.*, 5:1128, 1964.

[12] C.Domb, G.S.Joyce. Cluster expansion for a polymer chain. *J. Phys. C: Solid State Phys.*, 5:956, 1972.

[13] S.Caracciolo, G.Parisi, A.Pelissetto. Random walks with long-range self-repulsion on proper time. *J. Stat. Phys.*, 77:519, 1994.

[14] S.F.Edwards. *Proc. Phys. Soc.*, 85:613, 1965.

[15] P.J. Flory. *J. Chem. Phys*, 17:179, 1949.

[16] W. Krauth. Introduction to monte carlo algorithms. *e-print cond–mat/9612186*, 2003.

[17] A. Sokal. Monte carlo methods in statistical mechanics: Foundations and new algorithms.

[18] M. Iosifescu. *Finite Markov Processes and Their Applications*. Wiley, Chichester, 1980.

[19] K. L. Chung. *Markov Chains with Stationary Transition Probabilities*. Springer, New York, 1967.

[20] D. Revuz. *Markov Chains*. North–Holland, Amsterdam, 1975.

[21] A. Pelissetto. Introduction to the monte carlo method. In *Elementary Particles, Quantum Fields and Statistical Mechanics*. Università di Parma, 1994.

[22] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller. Equation of state calculation by fast computing machines. *J. Chem. Phys.*, 21:1087, 1953.

[23] W.K. Hastings. Monte carlo sampling methods using markov chains and their application. *Biometrika*, 57:97, 1970.

[24] W.W. Wood, J.J. Erpenbeck. Molecular dynamics and monte carlo calculations in statistical mechanics. *Ann. Rev. Phys. Chem.*, 27:319, 1976.

[25] A. Berretti, A.D. Sokal. New monte carlo method for the self-avoiding walk. *J. Stat. Phys.*, 40:483, 1985.

[26] S. Caracciolo, A. Pelissetto. A general limitation on monte carlo algorithms of metropolis type. *e.print hep–lat/9307021*, 1993.

[27] M.B. Priestley. *Spectral Analysis and Time Series*. Academic, London, 1981.

[28] T.W. Anderson. *The Statistical Analysis of Time Series*. Wiley, New York, 1971.

[29] N.Madras, A.D. Sokal. The pivot algorithm: A highly efficient monte carlo method for the self-avoiding walk. *J. Stat. Phys.*, 50:109, 1988.

[30] P.C. Hohenberg, B.I. Halperin. Theory of dynamic critical phenomena. *Rev. Mod. Phys.*, 49:435, 1977.

[31] J. Goodman, A.D. Sokal. Multigrid monte carlo method for lattice field theories. *Phys. Rev. Lett.*, 56:1015, 1986.

[32] C.G. Batrouni, G.R. Katz, A.S. Kronfeld, G.P. Lepage, B. Svetitsky, K.G. Wilson. Langevin simulations of lattice field theories. *Phys. Rev.*, D32:2736, 1985.

[33] R.H. Swendsen, J.S. Wang. Nonuniversal critical dynamics in monte carlo simulations. *Phys. Rev. Lett.*, 58:86, 1987.

[34] A.D. Sokal. New monte carlo algorithms for quantum field theory and critical phenomena or how to beat critical slowing down. In *Monte Carlo Methods in Theoretical Physics*. Eds. Sergio Caracciolo, Adelchi Fabrocini, ETS Editrice, Pisa, 1991.

[35] A.D. Sokal. Monte carlo methods for the self-avoiding walk. *e-print hep–lat/9405016*, 1994.

[36] F.T. Wall, L.A. Hiller Jr., D.J. Wheeler. Statistical computation of mean dimensions of macromolecules. i. *J. Chem. Phys.*, 22:1036, 1954.

[37] J.M. Hammersley, D.C. Handscomb. *Monte Carlo Methods*. Methuen, London, 1964.

[38] K. Suzuki. Excluded volume effect of very-long-chain molecules. *Bull. Chem. Soc. Japan*, 41:538, 1968.

[39] M. Delbrük. In *Mathematical Problems in the Biological Sciences*. Ed. R.E. Bellman, American Mathematical Society, Providence, 1962.

[40] N. Madras, A.D. Sokal. Nonergodicity of local, length-conserving monte carlo algorithms for the self-avoiding walk. *J. Stat. Phys.*, 47:573, 1987.

[41] S. Caracciolo, M. Papinutto, A. Pelissetto. Dynamic critical behavior of an extended reptation dynamics for self-avoiding walks. *e-print cond–mat/0110455v1*, 2004.

[42] B. Berg, D. Foerster. Random paths and random surfaces on a digital computer. *Phys. Lett. B*, 106:323, 1981.

[43] C. Aragão De Carvalho, S. Caracciolo. Polymers and $g|\phi|^4$ theory in four dimensions. *Nucl. Phys. B*, 215:209, 1983.

[44] M. Lal. Monte carlo computer simulation of chain molecules i. *Molec. Phys.*, 17:57, 1969.

[45] B. MacDonald, N. Jan, D.L. Hunter, M.O. Steinitz. Polymer conformations through 'wiggling'. *J. Phys. A*, 18:2627, 1985.

[46] D.E. Knuth. *The Art of Computer Programming*, volume 3. Addison-Wesley, Massachusetts, 1973.

[47] T. Kennedy. A faster implementation of the pivot algorithm for self-avoiding walks. *e-print cond–mat/0109308v1*, 2001.

[48] J. Zinn-Justin. *Quantum Field Theory and Critical Phenomena*. Oxford Science Publications, 1990.

[49] M.K.Kosmas, K.F.Freed. On scaling theories of polymer solutions. *J. Chem. Phys.*, 69:3647, 1978.

[50] P.G.de Gennes. *Riv. Nuovo Cimento*, 7:363, 1977.

[51] M.Gabay, T.Garel. *J. Phys. (Paris) Lett.*, 39:123, 1978.

[52] Y.Oono, K.F.Freed. Conformation space renormalization of polymers I; single chain equilibrium properties using wilson-type renormalization. *J. Chem. Phys.*, 75:993, 1981.

[53] I.Hüter. Proof of the conjecture that the planar self-avoiding walk has root mean square displacement exponent 3/4. *e-print math.PR/0108077*, 2001.

[54] I.Hüter. Weakly self avoiding walks on graphs and self-intersection events. *Preprint*, 2003.

[55] D.Brydges, T.Spencer. Self-avoiding walk in 5 or more dimensions. *Commun. Math. Phys.*, 97:125, 1985.

[56] E.Bolthausen, C.Ritzmann. Strong pointwise estimates for the weakly self-avoiding walk. *e-print math-PR/0103218*, 2001.

[57] D.Zeilberger. The abstract lace expansion. *e-print math.CO/9808076*, 1998.

[58] A.Greven, F.den Hollander. A variational characterization of the speed of a one-dimensional self-repellent random walk. *Ann. Appl. Probab.*, 3:1067, 1993.

[59] W.König. A central limit theorem for a one-dimensional polymer measure. *Ann. Probab.*, 24:1012, 1996.

[60] R.van der Hofstad, F.den Hollander. Scaling for a random polymer. *Commun. Math. Phys*, 169:397, 1995.

[61] R.van der Hofstad, F.den Hollander, W.König. Central limit theorem for a weakly interacting random polymer. *Markov Processes and Related Fields*, page 1, 1997.

[62] K.G. Wilson. Renormalization group and critical phenomena. *Phys. Rev. B*, 4:3174, 1971.

[63] K.G. Wilson, J. Kogut. The renormalization group and the ϵ-expansion. *Phys. Rep. C*, 12:75, 1974.

[64] L.P. Kadanoff *et al.* Static phenomena near critical points: Theory and experiment. *Rev. Mod. Phys.*, 39:395, 1967.

[65] A. Pelissetto, E. Vicari. Critical phenomena and renormalization-group theory. *Phys. Rep.*, 368:549, 2002.

[66] P.G. de Gennes. Exponents for the excluded volume problem as derived by the wilson method. *Phys. Lett.*, 38A:339, 1972.

[67] J. Cardy. *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, 2000.

[68] E.V. Orlov, A.I. Sokolov. Five-loop renormalization-group expansion for two-dimensional euclidean $\lambda\phi^4$ theory. *e-print hep–th/0301077*, 2003.

[69] H. Saleur. Conformal invariance for polymers and percolation. *J. Phys. A: Math. Gen.*, 20:455, 1987.

[70] B. Nienhuis. Exact critical point and exponents of the $o(n)$ model in two dimensions. *Phys. Rev. Lett.*, 49:1062, 1982.

[71] B. Nienhuis. Critical behavior in two dimensions and charge asymmetry of the coulomb gas. *J. Stat. Phys.*, 34:731, 1984.

[72] A. Sokal. Static scaling behavior of high-molecular-weight polymers in dilute solutions: a reexamination. *e-print hep–lat/9305009*, 1993.

[73] A. Pelissetto, J.-P Hansen. Corrections to scaling and crossover from good- to θ-solvent regimes of interacting polymers. 2004.

[74] V.L. Ginzburg. *Fiz. Tverd. Tela*, 2:2031, 1960.

[75] C. Bagnuls, C. Bervillier. *Phys. Rev. Lett.*, 76:4094, 1996.

[76] E. Luijten, H.W.J. Blöte, K. Binder. Medium-range interactions and crossover to classical critical behavior. *Phys. Rev. E*, 54:4626, 1996.

[77] A. Pelissetto, P. Rossi, E. Vicari. Crossover scaling from classical to nonclassical critical behavior. *e-print cond–mat/9804246*, 1998.

[78] S. Caracciolo, M.S. Causo, A. Pelissetto, P. Rossi, E. Vicari. Crossover phenomena in spin models with medium-range interactions and self-avoiding walks with medium-range jumps. *e-print cond–mat/0105160*, 2001.

[79] A.J. Barrett, C. Domb. Virial expansion for a polymer chain: the two parameter approximation. *Proc. Roy. Soc. London A*, 367, 1979.

[80] E.W. Montroll, G.H. Weiss. Random walks on lattices. II. *J. Math. Phys.*, 6:167, 1965.

[81] K. Levenberg. A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, 2:164, 1944.

[82] D. Marquardt. An algorithm for least-square estimation of nonlinear parameters. *J. Appl. Math.*, 11:431, 1963.

[83] F.J. Wegner. Corrections to scaling laws. *Phys. Rev. B*, 5:4529, 1972.

[84] P. Belohorec, B.G. Nickel. Accurate universal and two-parameter model results from a monte-carlo renormalization group study. *Unpublished*, 1997.

[85] S. Caracciolo, A.J. Guttmann, I. Jensen, A. Pelissetto, A. Rogers, A. Sokal. Correction-to-scaling exponents for two-dimensional self-avoiding walks. *e-print cond–mat/0409335*, 2004.

[86] E. Luijten, H.W.J. Blöte, K. Binder. Nonmonotonical crossover of the effective susceptibility exponent. *e-print cond–mat/9706195*, 1997.

[87] V. Privman, P.C. Hohenberg, A. Aharony. In *Phase Transitions and Critical Phenomena. ed. C. Domb and J.L. Lebowitz.* Academic Press, London-San Diego, 1991.

Ringraziamenti

Vorrei ringraziare alcune persone, ognuna delle quali è ed è stata fondamentale per me. Grazie di cuore: senza di voi non sarei dove sono.

Grazie ai miei genitori, che hanno avuto fiducia in ciò per cui la persi io. Grazie: vi devo molto.

Grazie a Olivia, che mi ha fatto scoprire che il cammino della vita, per quanto casuale possa essere, è davvero autoescludente: non si passa mai due volte sulla stessa strada. Grazie Oli, anche e soprattutto per altre infinite infinitesime briciole di pane.

Grazie a Francesco, che mi è stato vicino anche da cosí lontano.

Grazie a Bortolo, per aver ragionato con me - che sono irragionevole - durante tutto il tempo che mi ha dedicato.

Grazie al Prof. Caracciolo, alla cui pazienza e comprensione devo piú che a ció - ed è molto - che mi ha insegnato.

Grazie a tutti i miei maestri, da mio padre a Guccini.

Grazie ai miei amici, che per molto tempo sono stati tutto, e per tutto il tempo saranno moltissimo.

Grazie a Marco.