TEL AVIV UNIVERSITY
RAYMOND AND BEVERLY SACKLER
FACULTY OF EXACT SCIENCES
SCHOOL OF PHYSICS & ASTRONOMY

אוניברסיטת תל-אביב
הפקולטה למדעים מדוייקים
ע"ש ריימונד ובברלי סאקלר
בית הספר לפיסיקה ואסטרונומיה

A Study of a Two-Dimensional Lattice Model of Polymers with Short-Range Interactions

Thesis submitted as part of the requirements for the degree of
Master of Science (M.Sc.) in Physics at Tel Aviv University
School of Physics and Astronomy

by

Ido Golding

The work was carried out under the supervision of
Professor Yacov Kantor

December 1996

Acknowledgments

I wish to thank my supervisor, Professor Yacov Kantor, for his guidance.
I am grateful to my fellow student, Eilon Brenner, for his moral support and for fruitful discussions during the course of this work.
Special thanks are due to all of my army colleagues, for giving me a hand whenever needed.
I dedicate this work to my parents.

Contents

Abstract

We study a two-dimensional lattice model of polymers, subject to a random short-range interaction between two monomer types ("charges"). The study is performed over a "canonical" ensemble (i.e. having a constant number of monomers) of polymers, with quenched randomness. For short polymer chains (up to 16 monomers), data is obtained by enumerating all spatial conformations and averaging over all possible quenches. For long chains (up to 100 monomers), we use a Monte Carlo simulation process.

The collapse of such *heteropolymers* (polymers made of more than one monomer type) may be related to the subject of protein folding, which is one of the fundamental problems in biophysical science. We therefore start our work with a short review of protein folding, including a reference to the issue of simple, low resolution models — such as ours.

Before going into our specific model, we review (in chapter 2) some basic ideas and results regarding a single polymer chain. This includes the ideal chain model, excluded volume effects, the θ-transition (a term referring to the collapse of a polymer when temperature is lowered or solvent is made poorer), the effects of heteropolymer interactions and the subject of two-dimensional polymers.

We then (in chapter 3) describe our model system, discuss the issues to be tested, and briefly describe the numerical methods we apply.

Chapter 4 details our findings regarding the θ-transition. We find that, like a homopolymer (a polymer composed of a single species of monomers), a neutral polymer on a square lattice undergoes a θ-transition, which is a tricritical phase transition. The transition occurs at a temperature $T_\theta = 0.83 \pm 0.02$ (in units of the interaction strength), with a critical exponent $\nu_\theta = 0.60 \pm 0.02$, which seems to be different from the value for homopolymer collapse. For non-neutral polymers, we observe a decrease in the θ-temperature with increasing excess charge, up-to a critical value of the net-charge fraction ($X_{\rm cr}$) where the collapse vanishes. To conclude this chapter, we draw the phase diagram of the polymer in the plane of temperature and excess charge.

Next, we explore (in chapter 5) the ground state and the energy landscape of a neutral polymer, in an attempt to find evidence for the existence of a glass-like freezing transition for such a polymer. We do not find much evidence for this transition, a result which is in accordance with theoretical predictions.

The last chapter presents our conclusions and discusses some future prospects, and the appendix contains additional details of our numerical procedures.

1 Prologue - Protein Folding

In this work, we study a two-dimensional lattice model of polymers, subject to a random short-rage interaction between two monomer types. One of the reasons for interest in this problem, is the possible connection between the collapse of such *heteropolymers* (polymers made of more than one monomer type) and the subject of protein folding, which is one of the fundamental problems in biophysical science. We therefore start with a short review of protein folding. For a more thorough review of the subject, we refer the reader to the paper by Chan and Dill [1] and to the book by Creighton [2].

1.1 What Are Proteins?

A protein is a linear polymer molecule, whose monomers are the 20 naturally occurring amino acids. Different proteins have different sequences of the amino acids. The amino acid sequence is known as the *primary structure* of the protein. Proteins may be classified into three types: *Fibrous* proteins serve mainly structural roles - in hair, skin and bones. *Membrane* proteins reside in cellular membranes, where they mediate the exchange of molecules and information across cellular boundaries. Enzymes, which are the catalysts for biochemical reactions in living cells, are *globular* proteins (on which we will focus from now on).

The most important state of a globular protein, known as its *native* or folded state, is extremely compact and unique. That is, a given protein folds to only one native state. The so called *secondary structure* of a globular protein includes hydrogen-bonded α-helices and β-sheets. The large-scale architecture of a protein — how the helices, sheets and other secondary structures fit together — is called its *tertiary structure*.

Proteins are in their native states in aqueous solvents near neutral pH at $20 - 40^{\circ}$C; this is the typical cellular environment. Under some nonphysiological conditions, such as high temperature, acidic or basic pH, or in some nonaqueous solvents, the unique folded structure of a protein unfolds or denaturates, often reversibly, through a sharp transition to an ensemble of more expanded conformations.

1.2 The Folding Problem

Under physiological conditions the native state is marginally more stable than the ensemble of denaturated conformations. Marginal stability may be necessary for biological

function, since catalysis and binding properties of proteins must be responsive to the environment and to regulatory molecules. Nevertheless, marginal stability poses the problem of understanding the net effect of large and diverse driving forces, and deciding which of them constitutes the main contribution to the folding, to the formation of the secondary structure and to the uniqueness and thermodynamic stability of the native structure:

- Some of the amino acids have net charge, which leads to Coulomb interactions, though screened.
- All amino acids can form hydrogen bonds, which may lead to the formation of the secondary structure.
- Hydrophobic interactions are viewed as a strong force for folding proteins, but there is no certainty as to whether they contribute also to the uniqueness and internal architecture of globular proteins.
- A large contribution to the balance of forces comes from the decrease in conformational entropy upon folding.

The balance of forces that folds a protein into its unique, compact native structure is encoded within its amino acid sequence. Solving the folding problem means understanding and predicting the native conformation of a protein from its amino acid sequence.

Why is this important? First, because we wish to know how such remarkable states of matter arise from the underlying laws of chemistry and physics. Second, solving the folding problem would unleash considerable new power on biotechnology, in principle permitting the design of new proteins.

To find the stable native state of a protein, ideally we should compute, for every possible conformation of the chain, the sum of the free energies of the atomic interactions within the protein and with the solvent, and find the conformation with the lowest free energy. But this is not feasible, because the number of conformations n of a chain molecule – each of them with a unique energy value – grows exponentially with the chain length: $n \sim \overline{z}^N$ where N is the number of monomers and $\overline{z} \simeq 2 - 6$ is the number of rotational isomers, determined by the types of monomers that make up the polymer. An exhaustive search is not a practical solution to the folding problem for a computer algorithm. Molecular dynamics techniques, in which we numerically solve the equations of motion using interaction energies, are also quite useless when trying to find the global energy minimum

for a protein: Supercomputers can currently simulate nanoseconds of real-time protein dynamics, a scale that doesn't approach the $10^{-1} - 10^{3}$ seconds typically required to fold real proteins.

Heuristic approaches also exist, in which we learn about the relationship between sequence and structure by observing patterns in the database of known protein structure [3].

1.3 Simple, Low Resolution Models

The only way by which we can explore the whole conformation and sequence spaces of proteins is by introducing models which are very simple and of low resolution: Proteins are modeled as self-avoiding walks on lattices. Specific sequences of monomers are studied in chains short enough that the full conformational space can be enumerated exhaustively.

The simplest model describing compactization of a polymer considers *homopolymer collapse*: Homopolymers are polymers composed of a single species of monomers. If these monomers possess a repulsive interaction with the solvent, leading to an effective attraction between monomers, then two competing terms determine the free energy: The internal energy, favoring compactization, and the entropy, favoring the expanded states (since there are fewer compact than expanded conformations). Thus, increasing monomer-monomer attraction – which depends on solvent conditions as well as on temperature – leads to a collapse from open to compact configurations.

Theories on the collapse of *heteropolymers*, such as proteins, are in a more primitive state than theories of homopolymers. It remains a challenge to learn how heteropolymers can collapse into unique states and how heteropolymer sequences produce the thousands of unique structures that are native proteins.

An example of a heteropolymer model [4] is the one which describes a protein as a linear chain of amino acids, each of which can be either of two types: H (nonpolar, i.e. hydrophobic) or P (polar, i.e. hydrophilic). The monomers are subject to excluded volume and an HH attraction free energy ϵ.

Another approach is based on spin-glass models, a term describing a magnetic system where ferromagnetic and anti-ferromagnetic bonds are randomly distributed. The first application of spin glass ideas to the coil-to-globule folding of proteins was done in 1987 by Bryngelson and Wolynes [5], who applied Derrida's random energy model [6] to the problem. In this model, each of a set of discrete states is given an energy chosen randomly from a Gaussian distribution. The statistical independence of different nearby states leads

to an extraordinarily rough energy landscape, in which minima can be surrounded by high barriers. In addition, the states that "just happen" to have the lowest energy need not resemble one another by any measure. This description is thought to be a very good model for the folding landscape in the conformational space of a heteropolymer with a random sequence [7]. The model exhibits a phase transition similar to glass transitions observed in the laboratory. The transition is basically an entropy crisis: The number of thermally available states decreases so fast at the transition temperature that the system freezes into one of a small number of states — exactly which ones depends on the randomness of the landscape and the thermal history. Based on this model, Bryngelson and Wolynes predict different folding and "freezing" transitions of a heteropolymer: A chain may *fold* into a given native structure specified in advance or *freeze* into a collection of misfolded (non-native) structures that have extremely slow dynamics of interconversion. Other model studies suggest that the kinetic accessibility of the native structure is strongly sequence dependent. Shakhnovich and Gutin developed a heteropolymer model in which the distribution of monomer-pair interaction strengths B_{ij} is assumed to be Gaussian [8] [9]. The width B of the heterogeneity distribution plays the crucial role of determining the number of lowest-energy states of the model. If the sequences are sufficiently heterogeneous (B large), they find that only a few states dominate the low temperature phase. Thus, they conclude that unique protein folds can arise simply from sequence heterogeneity.

2 Single Polymer Chain - a Theoretical Review

Before going into our specific model, we give a review of results for a single polymer chain. A clear discussion of the subject can be found in the books by de Gennes [10] and by Plischke and Bergerson [11].

2.1 The Ideal Chain

One of the simplest idealizations of a flexible polymer chain is to replace it by a random walk on a periodic lattice, as shown in Fig. 1. The walk is a succession of N steps,

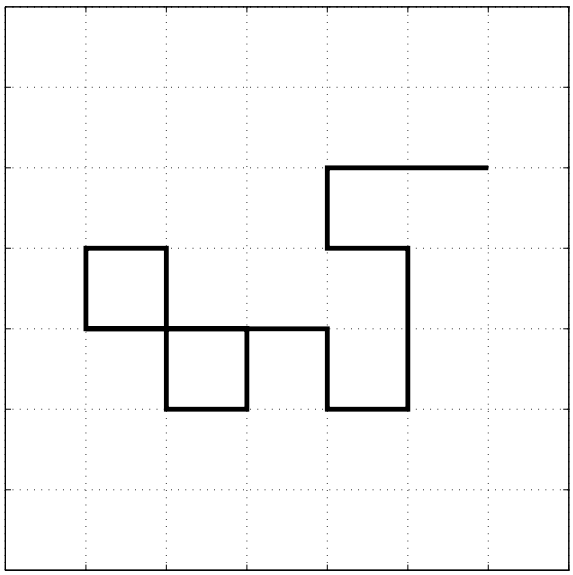

Figure 1: A random walk on a square lattice.

starting from one end and reaching an arbitrary end point. At each step, the next jump may proceed to any of the nearest-neighbor sites. All these possibilities have the same statistical weight. The length of one step will be denoted a.

The total number of such N steps walks is easily seen to be

$$n_N = z^N \ , \tag{2.1}$$

where z is the number of neighbors of each lattice site, also called the *coordination number*.

The end-to-end vector $\mathbf{R}_{ee}$ is the sum of N "jump vectors"

$$\mathbf{R}_{ee} = \mathbf{a}_1 + \mathbf{a}_1 + \ldots \mathbf{a}_N = \sum_n \mathbf{a}_n \ , \tag{2.2}$$

where each of the **a** terms is a vector of length a with z possible orientations. Based on the fact that different **a** vectors have completely independent orientations, we get the following results:

1. The average square end-to-end distance is linear in N,

$$\langle \mathbf{R}_{ee}^2 \rangle = Na^2. \tag{2.3}$$

2. For long walks, the distribution function for $\mathbf{R}_{ee}$ is Gaussian,

$$p(\mathbf{R}_{ee}) \sim e^{-d\mathbf{R}_{ee}^2/2Na^2} , \tag{2.4}$$

where d is the dimensionality of the lattice.

Eq. 2.4 gives a formula for the entropy of the chain at a certain elongation[1]:

$$S(\mathbf{R}_{ee}) = S(\mathbf{R}_{ee} = 0) - \frac{d\mathbf{R}_{ee}^2}{2Na^2} . \tag{2.5}$$

It is sometimes convenient to rewrite Eq. 2.5 in terms of the free energy,

$$F(\mathbf{R}_{ee}) = E - TS = F(0) + \frac{dT\mathbf{R}_{ee}^2}{2Na^2} , \tag{2.6}$$

giving the "spring constant" of an ideal chain.

Another ideal model, not restricted to a discrete lattice, is the one known as the *freely jointed chain*, consisting of $N + 1$ point particles, separated by bonds of length a that are free to take any orientation in the d dimensional space. For this model we get, again, that the average square end-to-end distance is linear in N. We also find that the *radius of gyration*, defined

$$R_g^2 = \frac{1}{N+1} \sum_{i=0}^{N} (\mathbf{R}_i - \overline{\mathbf{R}})^2 \tag{2.7}$$

(where $\mathbf{R}_i$ are the locations of the particles and $\overline{\mathbf{R}} = \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{R}_i$) behaves similarly for long chains:

$$\langle R_g^2 \rangle \approx \frac{Na^2}{6} . \tag{2.8}$$

Furthermore, it is found that restricting the orientation between successive bonds *still leads to ideal behavior*, with only a change of the constant, i.e. a is replaced by an effective bond length $\overline{a}$.

[1] Throughout this work, in all equations involving the temperature we will use units in which $k_B = 1$, i.e. temperature has the units of energy.

This last statement exemplifies the *global* (as opposed to local) point of view of polymer chains — where we omit the details of chain structure as much as possible and extract simple, universal features which will remain true for a large class of polymers.

The Gaussian chain is a simple model that can be shown to be formally equivalent to the freely jointed model in the thermodynamic limit (i.e. large N). We assume that the probability distribution $p(\mathbf{r}_i)$ for the vector $\mathbf{r}_i \equiv \mathbf{R}_i - \mathbf{R}_{i-1}$ connecting particles i and $i-1$ is given by

$$p(\mathbf{r}_i) \sim e^{-d\mathbf{r}_i^2/2a^2} \tag{2.9}$$

for all i. Then, writing formally the joint probability density for the locations of the $N+1$ particles in configuration space as a Boltzmann weight (plus a normalization constant) we can define an effective Hamiltonian

$$H = \frac{dT}{2a^2} \sum_{i=1}^{N} (\mathbf{R}_i - \mathbf{R}_{i-1})^2 \ , \tag{2.10}$$

which we will also write in a continuum version [11]:

$$H = \frac{dT}{2a^2} \int_0^N dx \, (\frac{d\mathbf{R}}{dx})^2 \ , \tag{2.11}$$

We will use the notation of Eq. 2.11 in the following sections.

2.2 Excluded Volume Effects

One of the important and subtle aspects of polymer statistics is the fact that real chains cannot cross or, equivalently, that no two particles can come closer than a minimum hard core distance. The effect of this on polymer conformations is easiest to see in the context of lattice models. One of the most thoroughly studied models of polymers is the *self-avoiding walk* (SAW) on a lattice — a random walk that can never intersect itself (Fig. 2).

For the unrestricted random walk, we have seen above that $R_g^2 \sim R_{ee}^2 \sim N$. Conversely, by enumerating SAWs on various two- and three-dimensional lattices, it can be seen that for large N

$$\langle \mathbf{R}_{ee}^2 \rangle \ \sim \langle R_g^2 \rangle \sim N^{2\nu(d)} \ , \tag{2.12}$$

where d is the dimensionality of the lattice, $\nu(2) = 0.75, \nu(3) \approx 0.6$. Therefore, the swelling of the random walk due to excluded volume is significant: It changes the exponent,

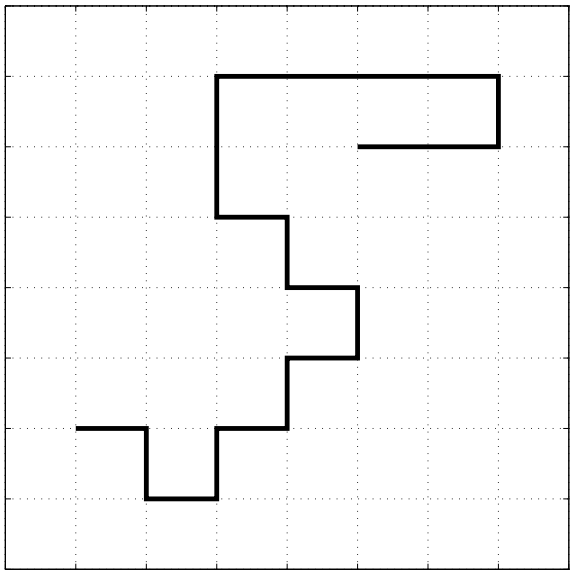

Figure 2: A self-avoiding walk on a square lattice.

not merely the amplitude, and has the universal characteristic of critical behavior — ν depends only on d, not on the type of the lattice.

It can also be shown that the total number of N-step SAWs has the property

$$n_N \sim \overline{z}^N N^{\gamma-1} \ , \tag{2.13}$$

where $\overline{z} < z$ is an "effective coordination number" that depends on the type of lattice, and γ is another universal exponent.

Flory [12] has devised a simple scheme for computing the exponent ν, which gives excellent values for all dimensionalities. It is based on writing two terms of the free energy: a repulsive energy (due to monomer-monomer excluded volume interaction) and an elastic (entropic) energy. By minimizing with respect to the polymer size R, one gets:

$$\nu = \frac{3}{d+2} \ . \tag{2.14}$$

This result, which actually "benefits from a remarkable cancellation of two errors" [10], is amazingly good.

The same result can be achieved using a continuum model. To the expression 2.11 for the entropic elasticity we add a simple two-body term, to get

$$\beta H = K \int_0^N dx \, (\frac{d\mathbf{R}}{dx})^2 + w \int_0^N dx \int_0^N dx' \delta^d(\mathbf{R}(x) - \mathbf{R}(x')) \tag{2.15}$$

(where $\beta = 1/T$). Then, using scaling arguments, we reach Flory's result (Eq. 2.14).

Here is the place to make a note about the dependence of ν on dimensionality. From 2.14 we see that ν is higher for smaller d, that is, the effect of excluded volume is more pronounced in lower dimensionalities. Moreover, it can be shown [10] that a self-avoiding chain is ideal for $d \geq 4$ — in this regime, repulsion between monomers represents only a weak perturbation.

2.3 The θ-Transition

A single polymer in a solvent is subject to both the hard core repulsion between monomers and a long range (e.g. van der Waals) attractive interaction. At high temperatures, the repulsive interactions are dominant, and the polymer size scales like a SAW. As temperature is lowered, a point is reached (called the θ-point), where the repulsive and attractive interactions effectively cancel and the polymer behaves essentially as an ideal random walk. For $T < \theta$, the polymer collapses into a compact object with $\nu = 1/d$. The same effect can be achieved by changing the solvent composition: In a "good" solvent, monomer-solvent interactions are attractive (as compared to monomer-monomer interactions), and the polymer swells. In a "poor" solvent, the polymer collapses in order to minimize monomer-solvent contacts.

This so called "θ-transition" has attracted a great deal of attention, both theoretical [13]–[15] and experimental [16]. De Gennes [17] has shown the analogy between the collapse of a polymer chain and a magnetic system near a *tricritical point* (a term used to classify a critical phenomenon according to the functional dependence of the free energy on the system's order parameter)[2]. Based on this analogy, use has been made of the critical exponents and the expected scaling equations calculated for tricritical phenomena, when investigating polymers near the θ-point [18] [19].

We will just mention that this analogy by de Gennes complements the known connection between self-avoiding walks and critical phenomena: It has been shown [10] that all properties of one self-avoiding walk on a lattice can be related to the spin correlation of a ferromagnet with an n-component magnetization vector, when we formally set $n = 0$.

[2] For an enlightening explanation see [11], and also [19].

2.4 Random Heteropolymers

The next degree of complexity we introduce into our system is *disorder* [20]. This disorder can be *external*, as in the case where we insert randomness into the lattice on which the polymer lies [21], or it can be *intrinsic*, as is the case for random heteropolymers. Below we will deal with intrinsic disorder only. We must also distinguish between *quenched* and *annealed* disorder, the first referring to a sequence of "charges" which is fixed in each realization of the system, whereas the latter describes charges which are free to move along the chain. A more precise definition (after [20]) would state that in an annealed (quenched) ensemble, the structure of the disorder, i.e. the charge sequence, relaxes on a much faster (slower) time-scale than the chain.

To be more specific, we describe a short-range interaction Hamiltonian [22]:

$$H_I = \frac{1}{2}\int_0^N dx \int_0^N dx' Q(x)Q(x')\delta^d(\mathbf{R}(x) - \mathbf{R}(x')) \tag{2.16}$$

(where $Q(x)$ is the quenched charge density along the polymer chain). Do these interactions modify the collapse transition of the polymer? That is, does the randomness change the critical phenomenon? This question will be of interest to us when we examine our model. The answer, however, is not clear. Kantor and Kardar [22] note that the relevance of the interactions is controlled by the sign of a scaling exponent $y_I = 1 - d\nu$, which means that for $\nu > 1/2$, the interactions are irrelevant at $d = 2$. In [23] it is claimed that quenched randomness results in repulsion, and the swelling exponent for $d < 2$ is computed.

The most remarkable feature of heteropolymer chains is the possible existence of a freezing transition at some finite temperature T_f [8] [7], below which the configurational entropy per monomer vanishes, and a few conformations with low energy dominate. Applying a kinetic description, we say that the heteropolymer can fold into its ground state or, instead, "freeze" into one of a few misfolded states. When describing the phase diagram, we may say that the compact state has split into several new phases: a liquid-like "molten globule", the native ground state, and a possible collection of misfolded "frozen" states.

An analytical investigation of random heteropolymers has shown an analogy between their physics and models of spin glasses. The freezing transition of heteropolymers is assumed to be analogous to the glass transition in Derrida's random energy model (see

Section 1.3), although this analogy is not always valid [24], as will be discussed when referring to our specific model.

2.5 Two-Dimensional Polymers

As will be discussed in the next chapter, our model deals with two-dimensional polymers. Although they are not, of-course, the common case, the treatment of 2-d polymers can be justified in a number of ways:

- Having mentioned the relation between polymer collapse and critical phenomena, there is a point in examining the problem of dimensionalities other than three, because the critical exponents are defined in each dimension, and depend only on d. Moreover, the two-dimensional case is thought favorable for examining the θ-point, since $d = 3$ is a marginal case and a logarithmic correction term is involved in the $R_g(N)$ dependence [25] [26].

- There is a big computational advantage in using a two-dimensional model over a 3-d one: We recall that for a given chain length N, the total number of SAWs is proportional (up to power law corrections) to $\overline{z}^N$ where $\overline{z} \approx 2.64$ for the square lattice and $\overline{z} \approx 4.68$ for the simple-cubic lattice [27]. When we later perform a complete enumeration of all chain conformations, our computer run-time will be proportional to the number of possible configurations. Therefore, for a specific chain length, the computational requirements in 2-d are much lower. On the other hand, when we use a Monte Carlo simulation to examine the polymers, we find that two-dimensional chains are much more prone to get stuck in conformations which the simulation fails to get out-of.

- Another argument [4] is that the surface-to-volume ratio, which primarily determines the physical behavior, is the same for shorter chains in two dimensions and for longer chains in three dimensions. Therefore, it is favorable to perform a numerical study — which is limited in the available chain-length — in two dimensions.

- Finally, Vilanove and Rondelez [28] give *experimental* evidence for the existence of two-dimensional polymers in Langmuir monolayers [29]. In particular, from the pressure isotherms they derive the scaling exponent ν for both a "good" solvent and a θ-solvent.

3 The Model System

In this section we describe the model under study, discuss the questions that will be of interest, and briefly describe the numerical methods by which we obtain the results. For an elaborate discussion of numerical aspects, we refer the reader to the appendix.

3.1 Description of the Model

Following Kantor and Kardar [30], we study a "canonical" ensemble (i.e. having a constant number of monomers) of polymers with quenched heterogeneity. The polymers are modeled as self-avoiding chains on a square lattice. Each polymer chain is formed from two types of monomers, labeled by $q_i = \pm 1$[1] , and subject to a short range interaction,

$$H_I = \sum_{i=0}^{N} \sum_{j \neq i} q_i q_j \Delta(\mathbf{R}_i - \mathbf{R}_j) \; , \tag{3.1}$$

where $\Delta(\mathbf{R}_i - \mathbf{R}_j) = 1$ if $\mathbf{R}_i$ and $\mathbf{R}_j$ are adjoining lattice sites (but i and j are not adjacent in position along the sequence) and $\Delta(\mathbf{R}_i - \mathbf{R}_j) = 0$ otherwise[2]. The only homogeneous interaction is the repulsion caused by the constraint of self-avoidance. Our interaction Hamiltonian, in which like charges repel and opposite ones attract, corresponds to strongly screened Coulomb interactions. This is in contrast to most two-letter models, i.e. models consisting on two monomer types [31]–[34], which try to simulate hydrophobic effects. The screened Coulomb case may be relevant to proteins: Three of the 20 natural amino-acids are positively charged, two are negatively charged, and the rest are neutral [35]. In addition, the screening length in biological solvents is often quite small.

3.2 Issues to be Tested

How do these heterogeneous interactions affect the θ-collapse of a polymer? This is the basic question that we would like to investigate. In particular, we will address the following issues:

[1] The thermodynamical properties of the model system depend only on the ratio q_i^2/T. We therefore choose arbitrary units in which $q_i^2 = 1$, and measure the temperature relative to these units.

[2] The interaction between monomers adjacent along the chain adds a constant term to the energy for all conformations of a certain quench. Discarding this term enables us to use a common energy scale for all quenches.

1. For "neutral" polymers (i.e. with equal number of positive and negative charges), we shall find that the polymer undergoes a "standard" θ-transition from a self-avoiding walk (with $\nu_{\rm SAW} = 0.75$) to a compact state ($\nu_{\rm compact} = 0.5$), and explore its properties — the θ-temperature T_θ and the critical exponent ν_θ. Our main finding is that ν_θ seems to be different from its value for homopolymer collapse. In other words, we find that the randomness may be a relevant perturbation at the θ-point.

2. For "charged" polymers, we quantify the asymmetry in the amount of positive and negative charges by $X \equiv \mid N_+ - N_- \mid /N$ (where N is here the number of monomers, equal to the number of steps plus one). For moderate values of X we shall see that the θ-transition still exists, with the transition temperature T_θ decreasing with X. However, there is a critical value $X_{\rm cr}$ above which there is no collapse — the chains are SAWs at both high and low temperatures. This is in accordance with results for the 3-d case [30].

3. We will try to explore the energy landscape of the polymers, in search of a possible freezing transition. Among the observables measured for this purpose are the $x(T)$ parameter (defined below), which estimates the number of relevant configurations at a given temperature, and $Q_{\alpha\beta}$ (likewise), a measure of the similarity between different configurations. We *do not* obtain conclusive evidence for the existence of a freezing transition.

3.3 Numerical Methods

The calculation of thermodynamic quantities is achieved as follows:

- For chain lengths of up to 15 steps (16 monomers), we use complete enumeration of all possible spatial conformations, and average over all possible charge sequences (quenches) of a given net charge.

- For chain lengths of up to 23 steps, we use complete enumeration of all possible conformations, and average over a limited number (20-100) of quenches of a given net charge.

- For chain lengths of up to 99 steps (100 monomers), we use Monte Carlo simulation and average the results over a limited number of quenches.

3.3.1 Complete Enumeration

In the exact enumeration process, we examine all possible spatial conformations of a chain of a given length. The number of these conformations, as seen from Eq. 2.13, grows exponentially with the chain length N. We take advantage of lattice symmetries to reduce the number of independent configurations. Thus, for a chain of 15 steps (16 monomers) we have to enumerate over 802,075 different configurations (unrelated by symmetry). The number of quenches of a certain net charge, for a chain of N monomers, is given by $\begin{pmatrix} N \\ N_+ \end{pmatrix} = \begin{pmatrix} N \\ N_- \end{pmatrix}$, giving, for example, 12870 neutral quenches for 16 monomers. For this case we get a total of roughly 10^{10} possibilities. To the best of our knowledge, previous authors have not performed a complete enumeration of configurations and quenches for this chain length. Some [4] have reached shorter chain lengths, while others have enumerated a limited number of quenches or a limited part of the conformational space [8] [31] [34] , restricting their study to compact configurations only.

The conformational enumeration process uses a depth-first algorithm [4], which seeks the longest branch of the self-avoiding walk. The algorithm backtracks either when the full chain of a given length has been generated or there is a dead-end due to an excluded-volume violation. At each step, the measured quantities (for example, the energy) are calculated by differences, that is, by calculating the change caused by the addition of a new monomer or the subtraction of one. The values are accumulated in a histogram, which is used later to calculate thermal averages at any given temperature. These averages are averaged again — this time over the possible quenches. We refer the reader to the appendix for further details of the numerical procedure.

3.3.2 Monte Carlo Simulation

In our simulation, we use a dynamic Monte Carlo (MC) method known as the "pivot" algorithm [27] [37] [38], described schematically in Fig. 3: Starting from some configuration of a SAW, we randomly select one site of the walk. A symmetry operation of the lattice (rotation or reflection) is applied to the part of the walk subsequent to the selected site, using this site as the origin. The choice of this symmetry operation is random. If the resulting walk is self-avoiding, it is accepted with a probability given (according to the

Metropolis method) by

$$p = \begin{cases} 1 & \text{for } E_{\text{new}} \leq E_{\text{old}} \\ e^{(E_{\text{old}} - E_{\text{new}})/T} & \text{for } E_{\text{new}} > E_{\text{old}} \ . \end{cases} \tag{3.2}$$

If the new walk is rejected, the old walk is counted once again. In each such iteration,

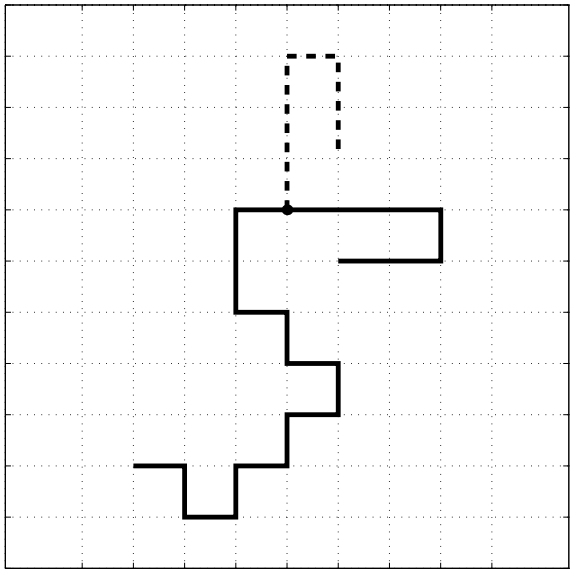

Figure 3: A "pivot" move on a self-avoiding chain. In this case, the move is a rotation of 90 degrees.

all observables are calculated and their thermodynamic mean value is obtained by averaging over all configurations counted (and later over different quenches). This process is performed for various temperatures, with the number of attempted moves in the range 10,000–300,000 (depending on temperature and chain length), and with 1 to 100 independent "runs" (series of such moves starting from an initial "rod" configuration) — needed because the algorithm gets stuck in local minima, at times. In each run, a predefined number of first samples is discarded, in order to remove the bias of initial conditions.

In order to verify the validity of the results and to obtain error estimates, we measure the correlation time — the time (in MC steps) needed for the decay of auto-correlation functions of the observables. For further details of the numerical procedure, we again refer the reader to the appendix.

4 Study of the θ-Transition: Results and Discussion

In this section we find evidence for a θ-transition of a neutral heteropolymer, which is a tricritical phase transition, and estimate the θ-temperature and the critical exponent ν_θ, which seems to have a different value from the homopolymeric case. We then turn to polymers having some net charge, and examine their collapse. We observe that the θ-transition persists up to some value $X_{\rm cr}$ of the excess charge, with the transition temperature decreasing with X. Above $X_{\rm cr}$ there is no longer a θ-transition.

4.1 Neutral Chains - Evidence of a Collapse

We first seek evidence of a collapse. We expect the polymer to be swollen (SAW-like, with $\nu_{\rm SAW} = 0.75$) at high temperature, and collapse to a compact shape ($\nu_{\rm compact} = 1/d = 0.5$) when we lower the temperature. Such evidence is not hard to find. Figs. 4–5 are an example: They show, respectively, the squared radius of gyration and end-to-end distance, both divided by the chain length, vs. the temperature, for various chain lengths. When the polymer is compact, we expect $R_g^2/N \sim R_{ee}^2/N \sim$ const, i.e. the value is independent of chain length N. On the other hand, when the polymer is swollen we should get $R_g^2/N \sim R_{ee}^2/N \sim N^{0.5}$, and the curves for different chain lengths should depart, with a higher value the higher the chain length. These estimates are evident quite nicely in Fig. 4, describing R_g^2/N. Fig. 5 includes a disturbing slight modification: In the collapsed region, R_{ee}^2/N decreases with N, a result that would imply $\nu < 0.5$ if we assume an exponential behavior $R_{ee} \sim N^\nu$. This may be a short-chain feature of the end-to-end distance, as contemplated in [25], where the same phenomenon was encountered (See also [47]).

We further show the energy per step E/N and the heat capacity per step C/N [1]. This is done in Figs. 6–7. The compactization is accompanied by a peak in the heat capacity, which shifts towards higher T and becomes higher for longer chains. This peak is another indication of the phase transition.

When comparing our results for R_g^2 , E and C with those for two-dimensional *homopolymers*, as seen, for example, in [13], we must note the qualitative similarity in the behavior of all observables. This further strengthens our expectation to observe in our

[1] For the exact enumeration results, C is calculated by numerically differentiating the energy: $C = \Delta E/\Delta T$, while in the MC simulation C is calculated from the fluctuations of energy: $C = (\langle E^2 \rangle - \langle E \rangle^2)/T^2$.

model, too, a tricritical phase transition as seen in the homopolymer case [13].

A visualization of the collapse transition is given by viewing two arbitrary conformations of a 50-monomer neutral polymer, sampled by the MC process, at $T = 2$ ("high temperature") and at $T = 0.6$ ("low temperature"). In Fig. 8 we see that the high temperature conformation is just a regular self-avoiding walk, whereas Fig. 9 depicts a conformation at a low temperature, where the polymer is very compact, with almost every monomer finding neighbors of the opposite charge.

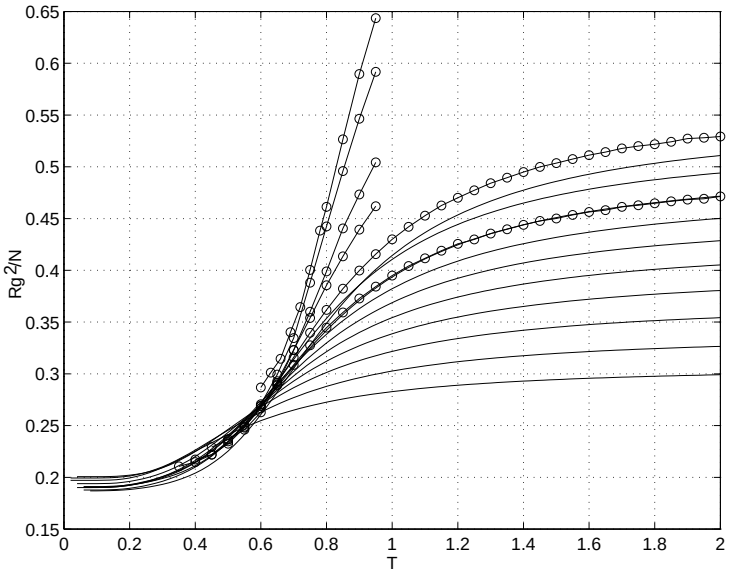

Figure 4: Squared radius of gyration, divided by chain length, for neutral polymers. Distances are measured in lattice constants, temperature is normalized by monomer interaction energy. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 6 , 8 , 10 , 12 , 14 , 16 , 18 (100 quenches) , 20 (100 quenches enumeration, 50 quenches MC) , 22 (100 quenches) , 24 (20 quenches) , 26 (50 quenches), 36 (25 quenches) , 50 (10 quenches) , 80 (10 quenches) , 100 (10 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

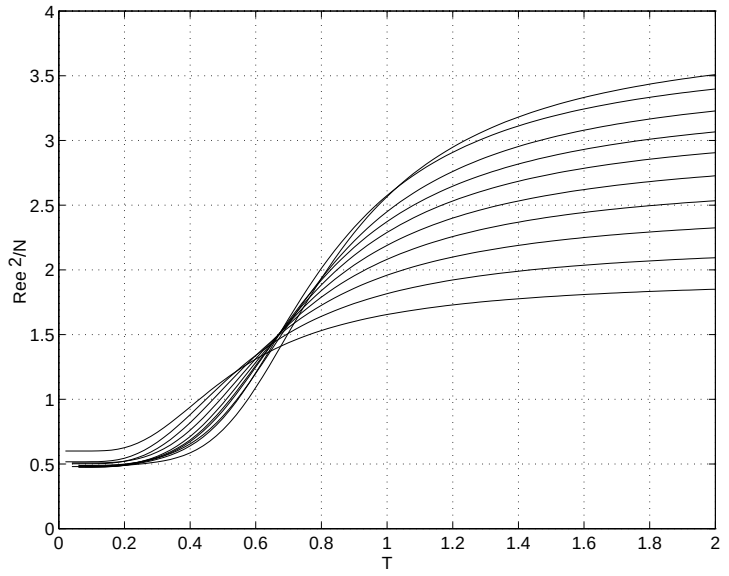

Figure 5: Squared end-to-end distance, divided by chain length, for neutral polymers. Data is from enumeration only. Distances are measured in lattice constants, temperature is normalized by monomer interaction energy. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 6 , 8 , 10 , 12 , 14 , 16 , 18 (100 quenches) , 20 (100 quenches) , 22 (100 quenches) , 24 (20 quenches).

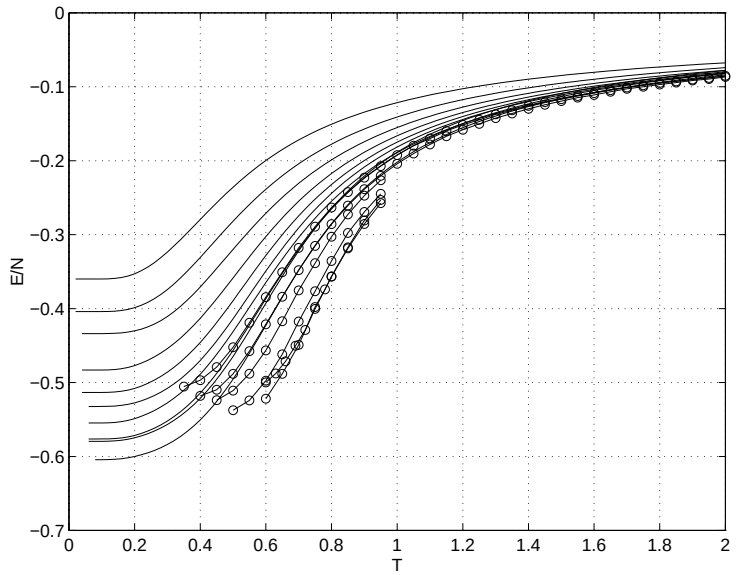

Figure 6: Energy divided by chain length, for neutral polymers. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from top left) : 6 , 8 , 10 , 12 , 14 , 16 , 18 (100 sequences) , 20 (100 quenches enumeration, 50 quenches MC) , 22 (100 quenches) , 24 (20 quenches) , 26 (50 quenches), 36 (25 quenches) , 50 (10 quenches) , 80 (10 quenches) , 100 (10 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

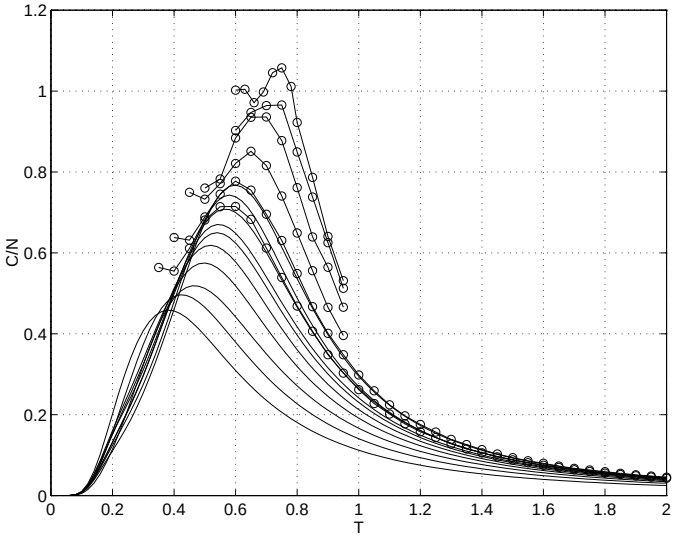

Figure 7: Heat capacity divided by chain length, for neutral polymers. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 6 , 8 , 10 , 12 , 14 , 16 , 18 (100 quenches) , 20 (100 quenches enumeration, 50 quenches MC) , 22 (100 quenches) , 24 (20 quenches) , 26 (50 quenches), 36 (25 quenches) , 50 (10 quenches) , 80 (10 quenches) , 100 (10 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

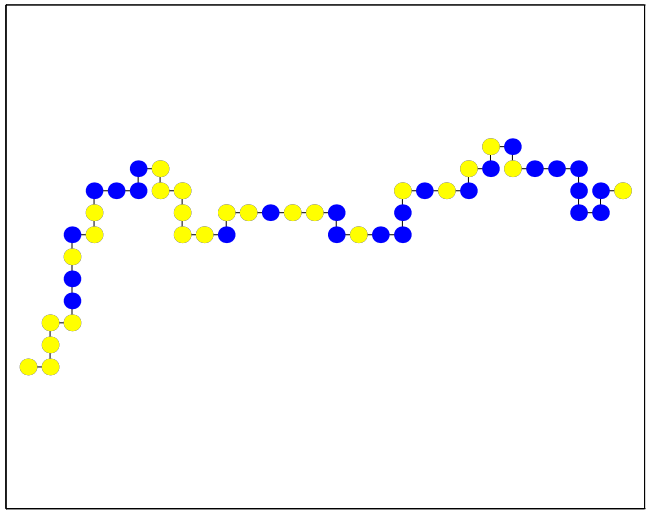

Figure 8: Conformation of a neutral 50-monomer polymer at $T = 2$, obtained by MC simulation. Oppositely charged monomers are denoted by dark and light filled circles.

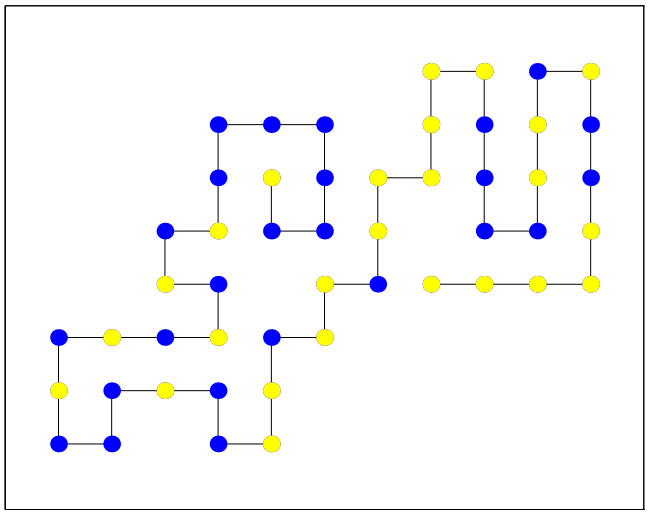

Figure 9: Conformation of a neutral 50-monomer polymer at $T = 0.6$, obtained by MC simulation. Oppositely charged monomers are denoted by dark and light filled circles.

4.2 Parameters of the θ-Transition

We now wish to validate the assumption that the collapse observed is indeed a θ-transition. If so, it should behave as a *tricritical* phase transition, as shown both theoretically [17] and numerically [13] (See also Section 2.3). We also wish to find the parameters of this transition, in particular the θ-temperature T_θ and the critical exponent ν_θ which describes the scaling of the radius of gyration at the θ-point.

Our first step is to try and estimate the transition temperature. We note that we are unable here to use the simple method applied in the 3-d case [30], which is to observe the intersection of the graphs of R_g^2/N vs. T for different lengths (using the facts that $\nu_{\text{ideal}} = \nu_\theta = 0.5$ and $\nu_{\text{SAW}} > \nu_\theta > \nu_{\text{compact}}$). This is because in the 2-d case, $\nu_{\text{compact}} = \nu_{\text{ideal}} = 0.5$. Moreover, unlike in 3-d, it is believed that for 2-d, $\nu_\theta \neq \nu_{\text{ideal}}$, and the 2-d value of ν_θ is not known a-priori.

Thus, we are forced to use other methods, the first of which is to identify the transition temperature T_θ with the position of the heat-capacity maxima (Fig. 7), and look for the limit of infinite chain length, which is expected to be the critical temperature. As depicted in Fig. 10, the $N \to \infty$ limit seems to exist, but a precise determination of $T_\theta(\infty)$ is quite impossible. This is possibly due to two reasons:

- The self-averaging of $C(T)$ over different quenches is quite slow for $T \approx T_\theta$, leading to variations in the position of the peak.
- The precision of the MC simulation deteriorates for low temperature and long chains (see appendix), giving large errors in the estimation of the peak position.

Thus, we give up the attempt to fit our results to the theoretical curve [13]:

$$T_\theta(\infty) - T_\theta(N) \sim \frac{1}{N^{\phi_\theta}} , \tag{4.1}$$

where ϕ_θ is a crossover exponent. We settle for observing that T_θ probably lies in the range $0.8 - 0.9$ and seek another method of estimation.

Such a method for accurately estimating T_θ and ν_θ, which at the same time helps validate the assumption of a tricritical transition, is by using finite-size scaling analysis of the observables, e.g. of R_g^2. In this technique, the expected behavior of a finite system is described in terms of a scaling theory involving the critical exponents of the corresponding infinite system. Based on the general theory for tricritical phase transitions, it has been

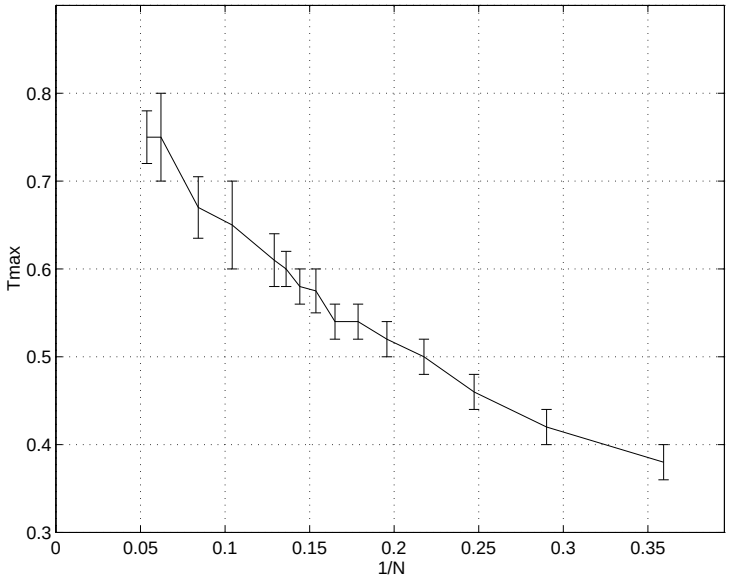

Figure 10: Location of the heat-capacity peak vs. inverse chain length, for a neutral polymer. Data is taken from Fig. 7.

shown theoretically [17] [18] and verified numerically for homopolymers [13] [42] [44] that the radius of gyration for different temperatures and chain lengths can be described using a single function as follows:

$$R_g^2/N^{2\nu_\theta} = f(N^{\phi_\theta}\tau) \ , \tag{4.2}$$

for $N \to \infty, \tau \to 0$, where $\tau \equiv |T - T_\theta|/T_\theta$, ν_θ is the correlation length exponent at the transition temperature T_θ , and ϕ_θ is the crossover exponent. The scaling function $f(x)$ should have the following limits:

$$f(x) = \begin{cases} \text{const} & \text{for } x \to 0 \\ x^{\mu_+} & \text{for } x \to \infty \ , \ T > \theta \ , \ \mu_+ = 2(\nu_{\rm SAW} - \nu_\theta)/\phi_\theta \\ x^{\mu_-} & \text{for } x \to \infty \ , \ T < \theta \ , \ \mu_- = 2(\nu_{\rm compact} - \nu_\theta)/\phi_\theta \ . \end{cases} \tag{4.3}$$

It is easily seen that the asymptotes of $f(x)$ at high and low temperatures reconstitute the behavior of a SAW and of a compact chain, respectively.

Fig. 11 depicts the scaling function $R_g^2/N^{2\nu_\theta}$ vs. the scaling variable $N^{\phi_\theta}\tau$, using data of all chain lengths (7–99) for temperatures satisfying $\tau \leq 0.25$, and choosing the parameters T_θ , ν_θ and ϕ_θ so that the points fall on two converging lines as required. As seen in the figure, a very good data collapse is achieved for $T_\theta = 0.83$, $\nu_\theta = 0.60$ and $\phi_\theta = 0.64$, and the resulting lines also approach the slopes of the theoretical asymptotes.

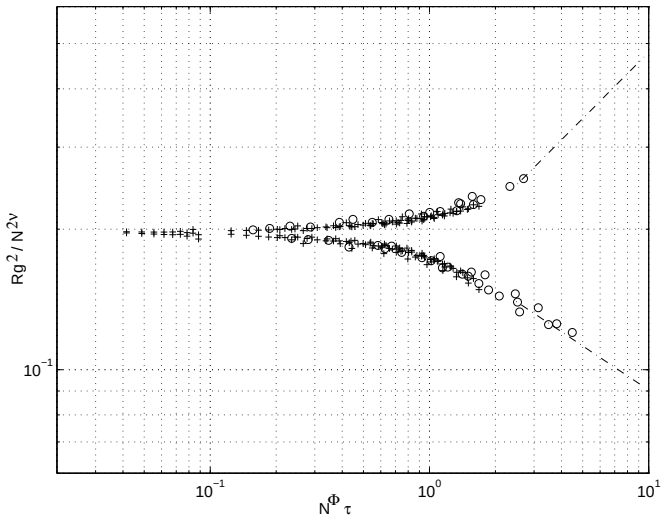

Figure 11: Log-log plot of the scaling function $R_g^2/N^{2\nu_\theta}$ vs. the scaling variable $N^{\phi_\theta}\tau$, close to the tricritical temperature. Data from chain lengths 7–99 is used, with values of τ up to 0.25. Parameters used: $\nu_\theta = 0.60$, $T_\theta = 0.83$, $\phi_\theta = 0.64$. Plus signs denote enumeration data, circles denote MC data. Dashed lines show theoretical asymptotes (Eq. 4.3), with amplitude fitted to data.

It should be further pointed out that the scaling behavior is found to be very sensitive to the value of ν_θ, slightly less to the value of T_θ, and quite insensitive to the choice of ϕ_θ. This is demonstrated in Fig. 12, where we have chosen $\nu_\theta = 0.57$. It is seen that the points, which previously lay on two converging lines, now diverge. More generally, we may state that the scaling behavior worsens as ν_θ departs from 0.60 and T_θ departs from 0.83. After examining various figures of the like of Fig. 12, we evaluate the transition parameters to be:

$T_\theta = 0.83 \pm 0.02$

$\nu_\theta = 0.60 \pm 0.02$.

where the error bars were set according to the parameter values where data did not collapse anymore (according to our subjective judgment). We did not evaluate ϕ_θ, due to the aforementioned insensitivity of the results to its value.

We verify these values by plotting $R_g^2/N^{2\nu_\theta}$ vs. T for various chain lengths (Fig. 13). The curves should intersect at $T = T_\theta$ (because $\nu_{\rm SAW} > \nu_\theta > \nu_{\rm compact}$). This is indeed what happens, within the estimated errors of the parameters and of the simulated data.

Another verification for the estimation of T_θ and ν_θ is achieved by examining the graph of $\log R_g$ vs. $\log N$ at a given temperature. At $T = T_\theta$ we expect a straight line, whose slope is equal to ν_θ, whereas for T slightly higher (lower) than T_θ, the slope should

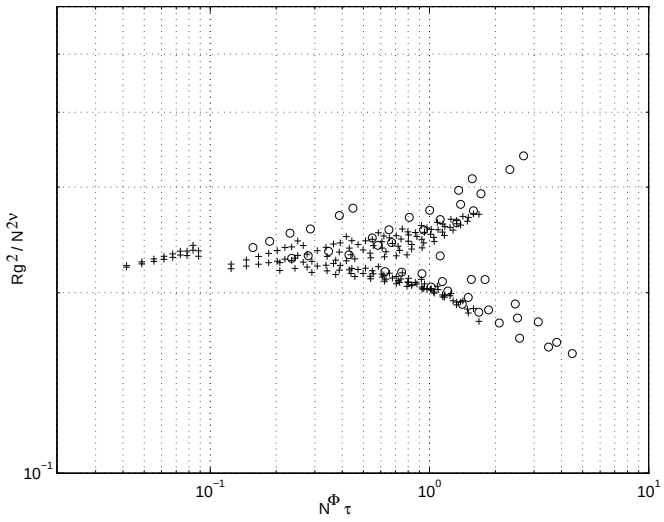

Figure 12: Log-log plot of the scaling function $R_g^2/N^{2\nu_\theta}$ vs. the scaling variable $N^{\phi_\theta}\tau$, close to the tricritical temperature. Data from chain lengths 7–99 is used, with values of τ up to 0.25. Parameters used: $\nu_\theta = 0.57$ $T_\theta = 0.83$ $\phi_\theta = 0.64$. Plus signs denote enumeration data, circles denote MC data.

increase (decrease) with N — recall Fig. 13. Fig. 14 confirms our assumption of linearity, with a least-mean-square fit giving $\nu = 0.60$ at $T = 0.83$. In Table 1 (after [42]) we examine the quality of the linear fitting at the temperature range 0.77–0.85. This is done by estimating ν separately for the lengths 9–25 and 35–99, and also calculating the sum of squared deviations between data and linear fitting (denoted by Σ), for the various temperatures. At T_θ, the estimates of ν from the two ranges of chain length should converge, and deviation from linearity should be minimal. Based on these requirements, we can estimate T_θ to be in the range 0.79–0.83, and ν_θ in the range 0.58–0.61. This agrees well with our scaling-based estimation, which we take to be our most precise one, mainly because it is obtained using many data samples at once (this decision is somewhat tentative, since one might claim that the estimation of parameters using a scaling analysis depends on our subjective choice of a specific plot as giving the "right" behavior, while others are judged as "wrong").

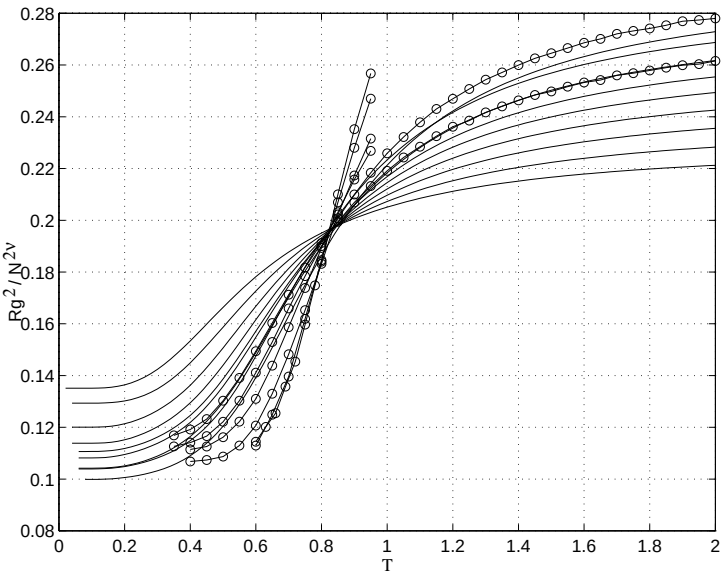

Figure 13: $R_g^2/N^{2\nu_\theta}$ vs. temperature, for various chain lengths (neutral polymers). The graphs intersect at the θ-point. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 8 , 10 , 12 , 14 , 16 , 18 (100 sequences) , 20 (100 quenches enumeration, 50 quenches MC) , 22 (100 quenches) , 24 (20 quenches) , 26 (50 quenches), 36 (25 quenches) , 50 (10 quenches) , 80 (10 quenches) , 100 (10 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

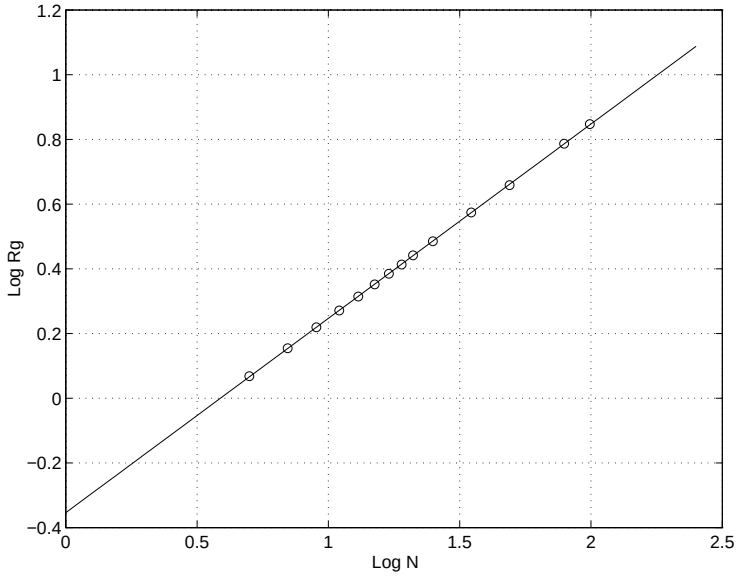

Figure 14: Log-log plot of the radius of gyration vs. chain length (neutral polymers) at $T = T_\theta$. The slope of the graph gives ν_θ.

T	0.77	0.79	0.81	0.83	0.85
$\nu(L = 9 - 25)$	0.579	0.588	0.593	0.602	0.608
$\nu(L = 35 - 99)$	0.574	0.585	0.596	0.607	0.617
Σ	4.0	2.59	2.71	2.44	4.69

Table 1: Results of the exponent ν for the radius of gyration, based on a linear fitting $R_g \sim N^{\nu}$, near the suspected critical temperature. Given are the estimates for two different ranges of chain-length, and also the sum of data deviations (squared) from the linear curve.

4.3 The Effect of Excess Charge

What happens to the collapse transition if the charges on the chain are not exactly balanced ($N_+ \neq N_-$) ? For such "charged" polymers, we measure the asymmetry in the amount of positive and negative charges by $X \equiv |N_+ - N_-|/N$ (where N is the number of monomers). Following the results for 3-d [30], we expect that for moderate values of X , a θ-transition still exists, with the transition temperature T_θ decreasing with X. However, there will exist a critical value $X_{\rm cr}$ above which no collapse will be seen — the chains will be SAWs at both high and low temperatures.

We have repeated the procedures of the previous section for the following values of X: 0 (neutral), $\frac{1}{4}$, $\frac{1}{3}$, $\frac{2}{5}$, $\frac{1}{2}$, $\frac{3}{5}$, $\frac{2}{3}$, $\frac{4}{5}$. Figs. 15–18 show $R_g^2/N^{2\nu_\theta}$ vs. temperature (various chain lengths), for four increasing values of X. It can be seen that for $X = \frac{1}{4}$ and $X = \frac{1}{3}$, we get the familiar intersection of curves, as observed for the neutral chains (Fig. 13). This indicates a θ-transition, with the same exponent $\nu_\theta = 0.6$, but with T_θ decreasing with X, as predicted. For $X = \frac{1}{2}$ the results are rather ambiguous, with their interpretation relying on how much we trust the Monte Carlo simulation at low temperatures. At $X = \frac{2}{3}$, however, there can be no doubt: The curves do not intersect. This means that the chain does not display θ-behavior no matter how much we lower the temperature. It follows, that the chain does not collapse ($\nu = 0.5$) either[2].

We have estimated the θ-temperature for the various X values using the aforementioned methods — the peak of $C(T)$, scaling analysis and the intersection of $R_g^2/N^{2\nu_\theta}$ curves. An approximate phase diagram in the (X, T) plane is depicted in Fig. 19. Note that the estimated error increases with X. This is due to the decrease in T_θ, which affects the quality of the MC simulation, and also possibly due to the transition itself which changes its nature and "deteriorates" for highly charged chains, before disappearing at $X_{\rm cr} \approx 0.6$.

[2]For X values that did not yield compactization, we have attempted also to obtain a different value for ν_θ using scaling analysis — but in vain.

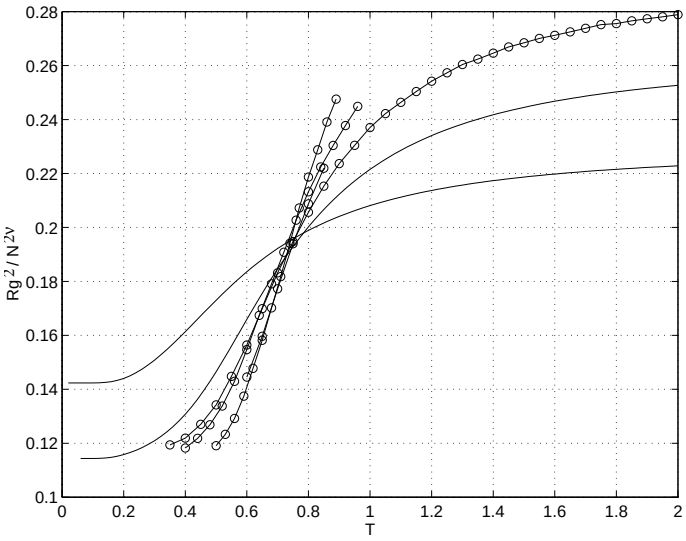

Figure 15: $R_g^2/N^{2\nu_\theta}$ vs. temperature for $X = \frac{1}{4}$ and various chain lengths. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 8 , 16 , 24 (50 quenches) , 32 (25 quenches) , 40 (25 quenches), 56 (10 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

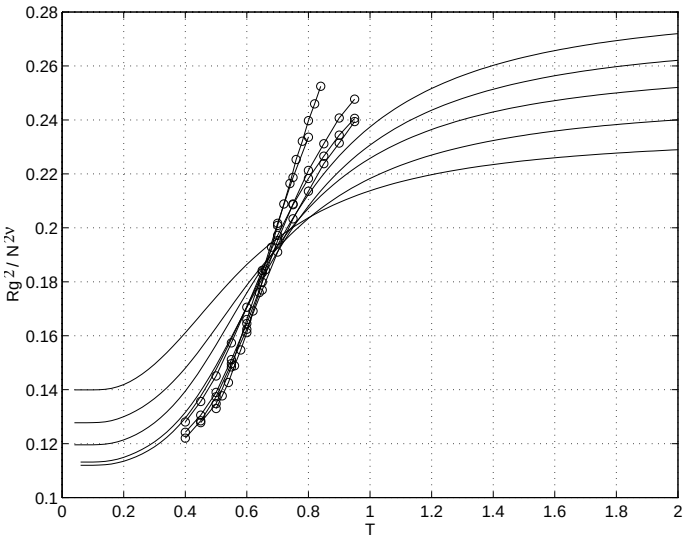

Figure 16: $R_g^2/N^{2\nu_\theta}$ vs. temperature for $X = \frac{1}{3}$ and various chain lengths. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 9 , 12 , 15 , 18 (100 quenches) , 21 (100 quenches) , 24 (25 quenches) , 27 (25 quenches), 30 (25 quenches) , 45 (25 quenches) , 60 (25 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

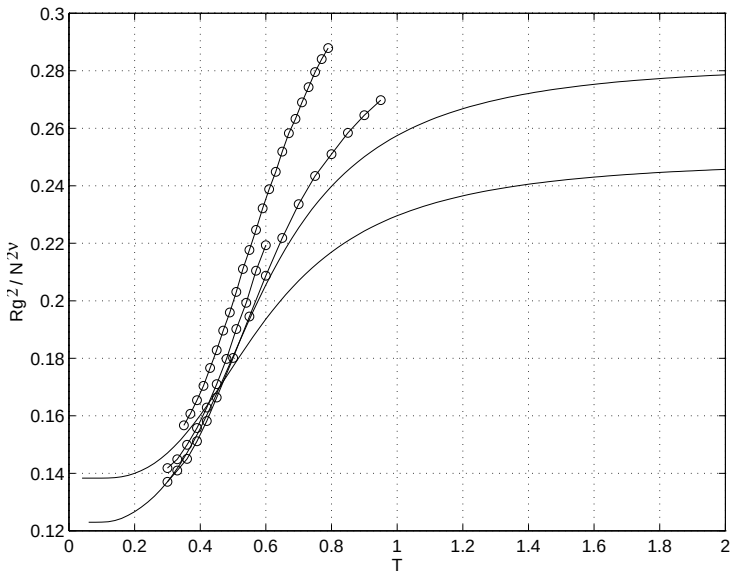

Figure 17: $R_g^2/N^{2\nu_\theta}$ vs. temperature for $X = \frac{1}{2}$ and various chain lengths. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 12 , 20 (100 quenches) , 28 (25 quenches) , 36 (25 quenches) , 48 (25 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

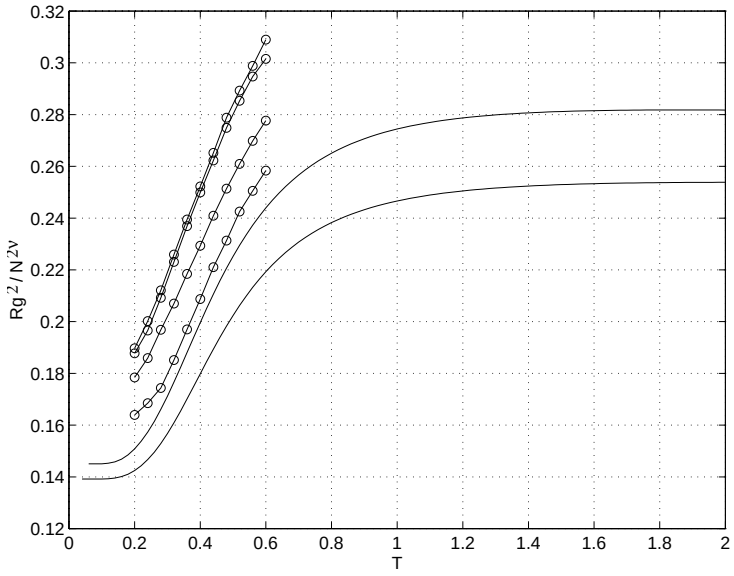

Figure 18: $R_g^2/N^{2\nu_\theta}$ vs. temperature for $X = \frac{2}{3}$ and various chain lengths. Curves are for all quenches (unless stated otherwise) and the following number of monomers (from bottom right) : 12 , 18 , 24 (15 quenches) , 30 (15 quenches) , 36 (15 quenches) , 42 (15 quenches). Solid lines represent results of enumeration, connected circles represent results of Monte Carlo simulation.

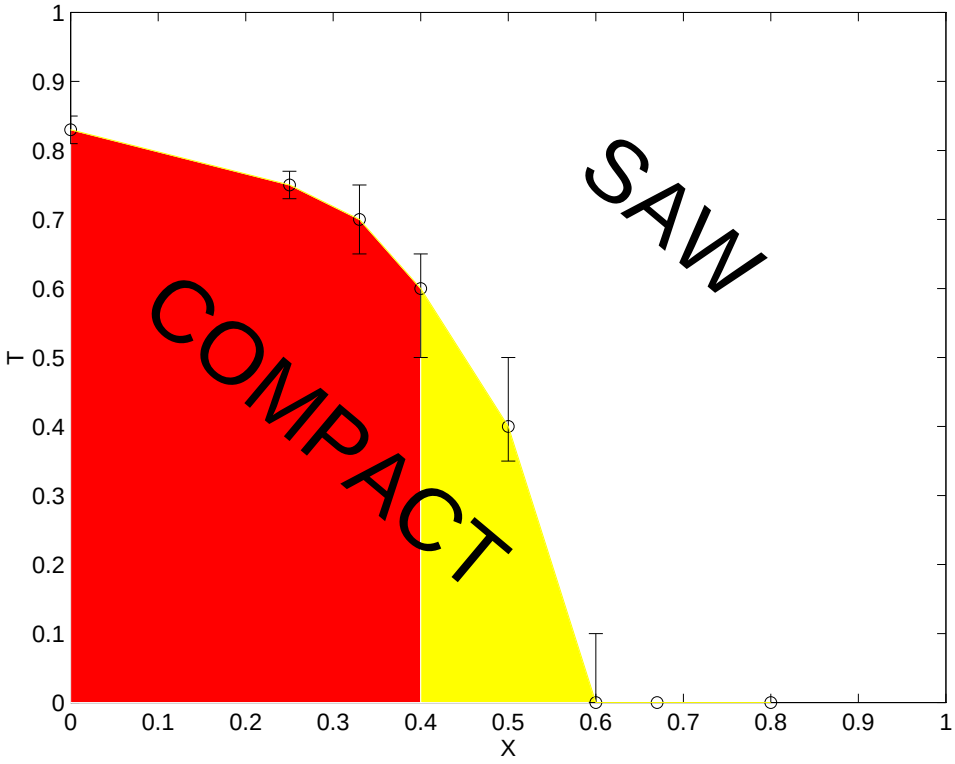

Figure 19: Phase diagram of a random heteropolymer in the plane of temperature (T) and excess charge (X). Vertical bars indicate estimated uncertainties in θ-temperature. The lighter shaded area indicates X values where the results are rather ambiguous, see for example Fig. 17.

4.4 Discussion

We have seen that a neutral polymer on a square lattice undergoes a tricritical θ-transition at a temperature $T_\theta = 0.83 \pm 0.02$, with a critical exponent $\nu_\theta = 0.60 \pm 0.02$. We take this value to be different from the homopolymer value. A glance at the various results for $\nu_\theta^{\rm homo}$ (see Table 2) may render this observation rather suspicious — the estimates span a vast range, with a value of 0.6 not unreasonable to be obtained. However, two reasons make us believe that our value is indeed different: First, of the results quoted in Table 2, the most recent and reliable simulations (to our judgment), e.g. [42], give estimates in the range 0.55-0.58. Second, we have used our own enumeration and MC procedures (properly altered, of-course) to simulate homopolymers, and obtained the result that $\nu_\theta^{\rm homo} = 0.55 \pm 0.01$. Therefore, our conclusion is that $\nu_\theta^{\rm hetero} \neq \nu_\theta^{\rm homo}$.

How does this observation withstand theoretical predictions? As discussed in section 2.4, there is no clear answer concerning the relevance of random interactions to the collapse transition. Kantor and Kardar [22] note that the relevance of the interactions is controlled by the sign of a scaling exponent $y_I = 1 - d\nu$, which means that for $\nu > 1/2$, the interactions are irrelevant at $d = 2$. Stepanow [23] states that quenched randomness results in repulsion, but computes the swelling exponent for $d < 2$ only.

While the analytical attempts at the problem are sporadic, the numerical ones are practically non-existing. A partially related work was done by Victor and Imbert [50]: They perform a MC simulation of 2-d polymers with an alternating sequence of charges, interacting through a logarithmic (long range) potential. The alternation of charges leads to an effective short-range ("dipole") interaction, which is similar to our model, though lacking the randomness. They obtain $\nu_\theta = 0.59 \pm 0.01$, a value very close to ours.

For non-neutral polymers, we have observed the decrease in θ-temperature with increasing excess charge, up-to a value $X_{\rm cr}$ where the collapse vanishes and SAW behavior prevails at all temperatures. We have drawn the phase diagram in (X, T) plane and found it similar to that obtained for 3-d by Kantor and Kardar [30]. A qualitatively similar phase diagram is obtained by Camacho and Schanke [33] for a 2-letter H-P model.

source	*result for* ν_θ
Vilanove & Rondelez [28]	0.56 ± 0.01
Ishinabe [25]	0.503 ± 0.01
de Gennes [17]	$0.505\ldots$
Baumgartner [13]	0.51
Meirovitch & Lim [42]	0.5795 ± 0.0030
	0.574 ± 0.006
Roy et al. [43]	$\frac{7}{12} = 0.583\ldots$
Barat et al. [44]	0.6 ± 0.2
Derrida & Saleur [45]	0.55 ± 0.01
Kholodenko & Freed [46]	0.55 ± 0.01
Privman [47]	0.535 ± 0.025
Seno & Stella [48]	0.57 ± 0.015
Marqusee & Deutch [19]	$\frac{2}{3} = 0.666\ldots$

Table 2: Various estimates of ν_θ for homopolymers

5 Examination of the Energy Landscape: Results and Discussion

We would now like to investigate in more detail the energy landscape of the (neutral) polymers, and in particular try to find possible evidence of a freezing transition.

5.1 The Ground State

First of all, we examine the ground state of neutral polymers. The mean ground state energy E_g (averaged over the different quenches) is shown vs. chain length in Fig. 20. As would be expected, it decreases monotonically with the increase in number of monomers. The error bars depict the standard deviations ΔE_g , which are of the order of 1 — i.e. E_g is highly self-averaging. If we are to give a phenomenological prediction for the N dependence of E_g (at least for large N), we would write

$$E_g \approx aN + bN^{1/2} \ , \tag{5.1}$$

where the first term represents "bulk" effects and the second one corresponds to a surface energy. To validate this estimate, we plot E_g/N vs. $N^{-1/2}$, expecting a straight line with slope b, intersecting the y axis at a ($E_g/N \approx a + bN^{-1/2}$). The results (Fig. 21) are quite satisfactory: For large N (small $N^{-1/2}$) linearity is observed, and we obtain $a \approx -0.87$, $b \approx 1.3$. The value of a is quite a reasonable one, as the homopolymer case gives $a = -1$, while in the random case, the connectivity of the polymer chain does not allow all monomers to choose optimal (i.e. of opposite sign) neighbors, thus giving rise to the phenomenon of *frustration* and yielding $a > -1$.

Another subject of interest is the *degeneracy* of the ground state. In other words — How many native states does a sequence have? Fig. 22 shows the distribution of degeneracy for all neutral 16-monomer chains. The main feature of the histogram is that the occurrence of a certain degeneracy tends to decrease when the degeneracy increases. That is, there are more sequences that have only one or two native structures than have 20 or 50. It would be correct to say that a typical sequence has "a few" native configurations. Similar results were found for the 2-letter H-P model [4] [31] . The degeneracy distribution is similar for other chain lengths. The average ground state degeneracy seems to increase slowly with chain length. However, even this observation should be regarded with caution, due to the large fluctuations of the values.

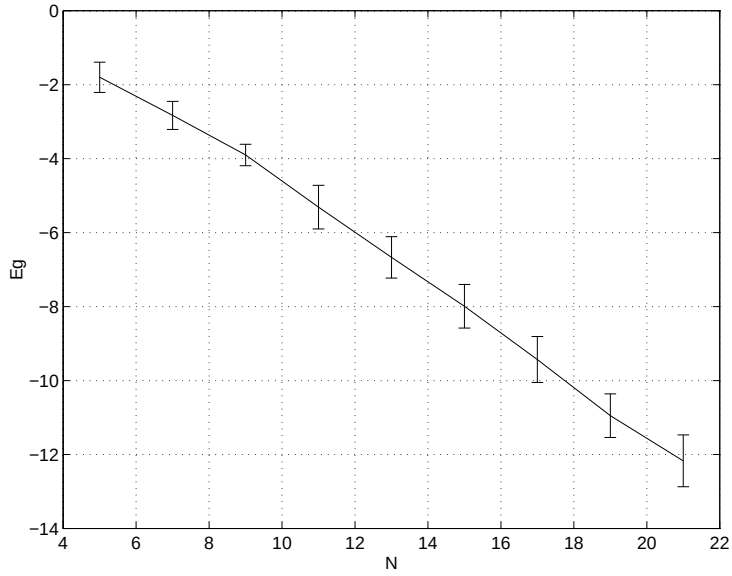

Figure 20: Average ground state energy of a neutral polymer vs. polymer length. Standard deviations of quenched averages are shown as error bars. For $N < 17$, averaging is over all possible neutral quenches. For $N \geq 17$ averaging is over 100 neutral quenches.

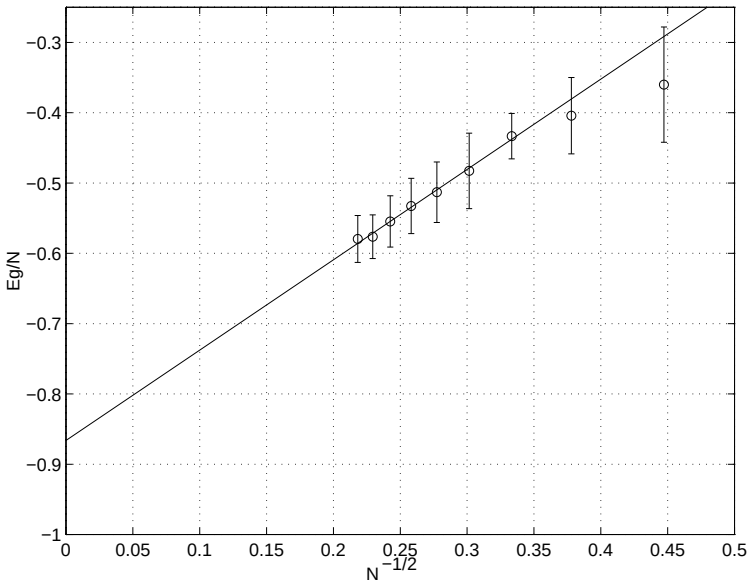

Figure 21: Average ground state energy of a neutral polymer (per monomer) vs. inverse square root of polymer length. Straight line shows linear fitting, when discarding small values of N.

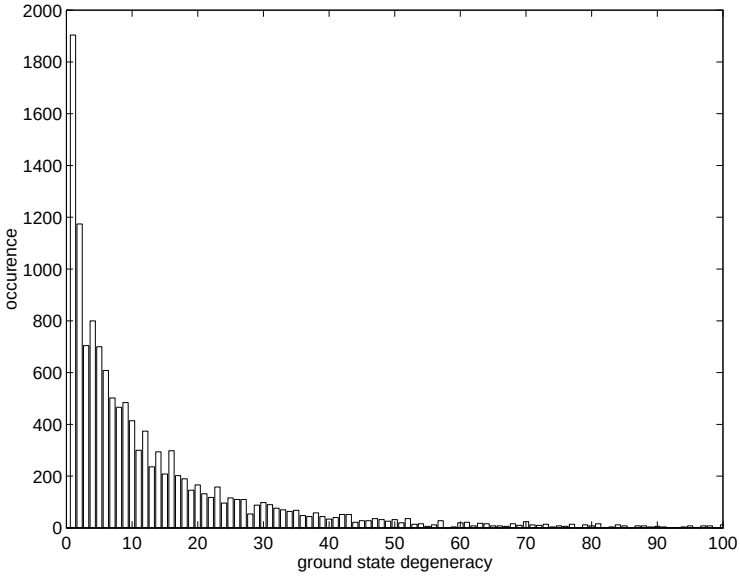

Figure 22: Distribution of ground state degeneracies for all neutral 16-monomer chains. Non-zero values are found for degeneracies as high as 720, but only the distribution up to 100 is shown here.

5.2 The Shape of the Energy Landscape, Search for Evidence of Freezing

5.2.1 Basic Definitions

As mentioned in section 2.4, one of the most remarkable features of heteropolymer chains is the possible existence of a freezing transition at some finite temperature T_f [8] [7], below which the configurational entropy per monomer vanishes, and a few conformations with low energy dominate. Applying a kinetic description, we say that the heteropolymer can fold into its ground state or, instead, "freeze" into one of a few misfolded states. When describing the phase diagram, we may say that the compact state has split into several new phases: a liquid-like "molten globule", the native ground state, and a possible collection of misfolded "frozen" states. Analytical investigation of random heteropolymers has shown analogy between their physics and models of spin glasses. The freezing transition of heteropolymers is assumed to be analogous to the glass transition in Derrida's random energy model. In this model, each of a set of discrete states is given an energy chosen randomly from a Gaussian distribution. The statistical independence of different nearby states leads to an extraordinarily rough energy landscape, in which minima can be surrounded by high barriers. In addition, the states that "just happen" to have the lowest energy need not resemble one another by any measure. This description is considered a very good model for the folding landscape in the conformational space of a heteropolymer with a random sequence [7]. The model exhibits a phase transition similar to glass transitions observed in the laboratory. The transition is basically an entropy crisis: The number of thermally available states decreases so fast at the transition temperature that the system freezes into one of a small number of states — exactly which ones depends on the randomness of the landscape and the thermal history.

To find evidence of freezing in a numerical model, we introduce two parameters [8]:

1. To characterize the number of conformations which are thermodynamically relevant at a given temperature, we use the x parameter, defined

$$x(T) = 1 - \sum_k p_k^2 \ , \tag{5.2}$$

where $p_k = e^{-E_k/T}/Z(T)$ is the Boltzmann weighted probability for a given conformation k. It can easily be seen that at high temperature when many conformations

have probabilities of the same order of magnitude, $x \approx 1$. In the opposite case, where only one conformation dominates, we get $x \approx 0$. In the case of a freezing transition at some temperature T_f, then according to [8], when averaging over all quenches we should encounter the following behavior:

$$\langle x(T) \rangle = \begin{cases} 1 & \text{for } T > T_f \\ \frac{T}{T_f} & \text{for } T < T_f \end{cases} . \tag{5.3}$$

2. An important feature of the frozen state of heteropolymers (when it exists) is that for a given sequence, the states with the lowest energies (and only these states are relevant in the frozen phase) are structurally different. To test this feature we measure similarity overlap between two conformations using the parameter

$$Q_{\alpha\beta} = \frac{N_{\text{common}}(\alpha, \beta)}{N_{\text{max}}} , \tag{5.4}$$

where N_{common} is the number of "contacts" which are the same in conformations α and β (i.e. monomers i and j are nearest neighbors in both conformations), and $N_{\text{max}} = \max[N_{\text{common}}(\alpha, \alpha), N_{\text{common}}(\beta, \beta)]$. When examining the low-energy conformations in a frozen phase, we expect to get low similarity between them, i.e. $Q \sim 0$. A value $Q \approx 1$ would indicate that the low energy conformations are close in structure to each other, meaning that the random energy model is not applicable to the system.

We will just mention, that when an analytical investigation of heteropolymers is made, in analogy with spin glasses, using the so-called *replica* technique, then parameters analogous to both x and Q appear in the calculations. For more details, see e.g. [9].

5.2.2 Should We Expect a Freezing Transition for Our Model ?

Although evidence of freezing, both analytical and numerical, has been found for various heteropolymer models, it has been shown [24] that the analogy between the energy landscape of a heteropolymer and the random energy model is not universal — it depends on the details of the heteropolymer model studied. In particular, there are a few features of our specific model which cast doubts on our prospects to observe freezing:

1. **Lack of strong homopolymer attraction.** The presence of an attractive homopolymer term in the Hamiltonian is assumed in all models which achieve a freezing transition. This attraction leads to the existence of a "molten globule" – a

compact phase with a large number of relevant conformations. Lowering the temperature then gives rise to heteropolymeric effects, and may create freezing [51] [52]. This assumption also allows Shakhnovich [8] to enumerate only *compact* configurations.

2. **Lack of sufficient heterogeneity.** A 2-letter model may not be heterogeneous enough to facilitate freezing. It is stated in [51] that the quenched disorder in the polymeric sequence introduces frustrations, which can lead the system to a frozen state. However, in the absence of sufficient types of charges, flexibility may enable the polymer to avoid frustration, and thus can suppress the freezing transition significantly. A similar claim is made by others [52]–[54]. The problem of heterogeneity can also be described from another point of view: To obtain freezing, we must have an energy spectrum with a continuous part, below which lie a few discrete states, sufficiently far apart — such is the spectrum resulting from the random energy model. To obtain such a spectrum (or as similar as possible), most models use more charge-types (e.g. 20 — the number of amino acids in a protein, as compared to our 2-charges model) or use randomly distributed interactions between monomers.

3. **Working on a two-dimensional lattice.** It appears [2] [55] that formation of a unique structure in heteropolymers in very sensitive to space dimensionality, with $d = 2$ being a marginal and nonuniversal case that strongly depends upon the type of lattice, type of sequence alphabet and so on. This is due to the fact that for $d > 2$ the majority of contacts are nonlocal, while for $d < 2$ the majority of contacts are local.

5.2.3 Results

We first examine the x parameter, which gives us indication of the number of thermodynamically relevant states (configurations) at each temperature. The value of $\langle x(T) \rangle$ averaged over all neutral quenches of a 16-monomer chain, is depicted in Fig. 23. Also shown are the average radius of gyration and heat capacity. It is seen that $x(T)$ does not reach a zero value — this is of-course due to the multiple ground state degeneracy of many sequences, as noted above. Another feature is that the temperature range where x decreases overlaps in general the range where the polymer collapse occurs — R_g^2 decreases and C has a peak. Therefore we cannot establish an estimate for a possible freezing tem-

perature. A relevant claim is made in [4], that when dealing with the H-P model, folding is the dominant factor in the decrease in number of relevant states.

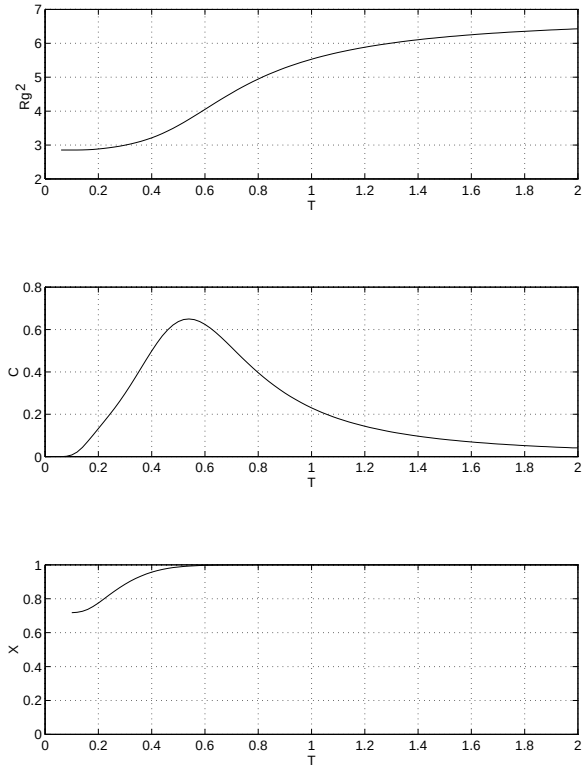

Figure 23: Squared radius of gyration (top), Heat capacity per monomer (middle) and x parameter (bottom) vs. temperature. Values are averaged over all quenches for neutral 16-monomer chains.

Next, we look at the energy spectrum of one neutral polymer (16-monomers, sequence chosen at random). Fig. 24 shows the energy levels and their occupation (logarithmic scale). The spectrum is seen to be very smooth and "continuous" (although it has, ofcourse, the inherent discreteness of the lattice model) — We do not obtain a number of singly-occupied states and a gap between them and the continuous regime, and we do not expect this energy spectrum to support freezing.

We also examine the similarity between the different ground-state ("native") conformations of a given sequence. Fig. 25 depicts the distribution of the average similarity between native states of a neutral 16-monomer chain. The averages ($\langle Q \rangle$) were calculated for 1000 quenches, and put in the histogram shown. It is seen that the different native states of a given chain are, on the average, structurally quite different from each other (Q values centered around $Q \sim 0.5$). A similar result has been reported for the 2-letter H-P model [4] (using a different measure for the distance between configurations). A small Q

value means that the minimum-energy states of the polymer are scattered randomly in its conformation space, in accordance with the random energy model. However, we find it hard to determine to what degree should a value of $Q \sim 0.5$ be regarded as "low".

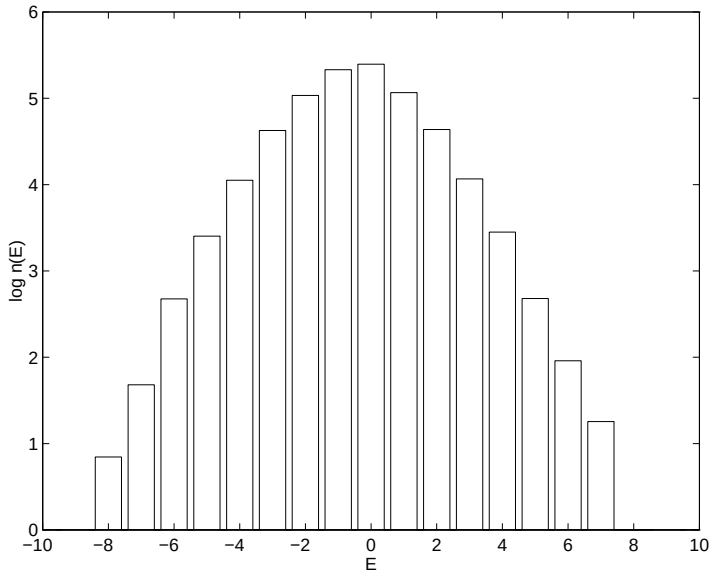

Figure 24: Logarithm of occupation of energy levels, for a 16-monomer neutral polymer (sequence chosen at random).

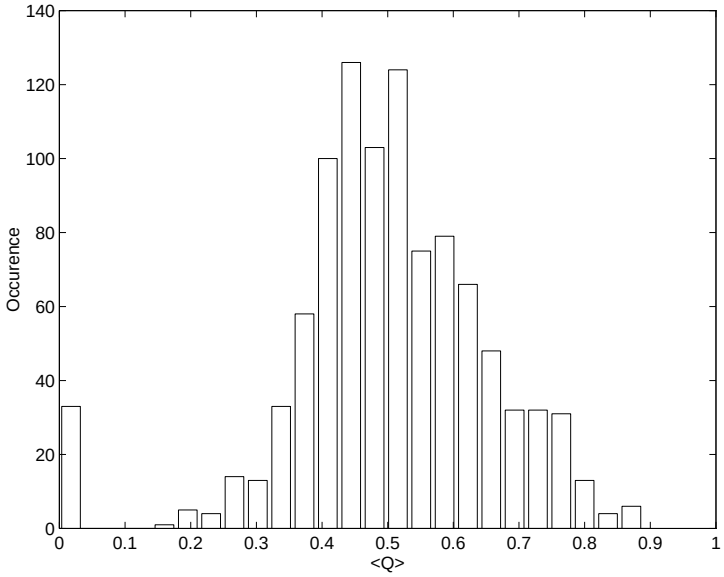

Figure 25: Distribution of average similarity between native states of a neutral 16-monomer chain. The averages ($\langle Q \rangle$) were calculated for 1000 quenches, and put in the histogram shown.

5.2.4 Discussion

We have not found much evidence for the existence of a freezing transition. Instead, results are quite ambiguous:

- On one hand, we have seen that the histogram of ground-state degeneracies is decreasing, meaning that many sequences have only a few native states. The average overlap between these native conformations, described by the Q parameter, is not high — $\langle Q \rangle \sim 0.5$, a value that, if we interpret it as "low", would lead us to believe that low-energy conformations are structurally different, as is the situation in the random energy model.

- On the other hand, we have seen that many sequences have a large number of degenerate ground states. Also, the typical energy spectrum of a neutral polymer is far from similar to that obtained in the random energy model, and seems most unprobable to support a freezing transition. Finally, we have not succeeded (using the x parameter) to find a possible location of the freezing temperature T_f, distinguishable from the location of the collapse temperature T_θ. These obstacles in finding conclusive evidence of freezing are in accordance with theoretical predictions regarding some particular aspects of our model, as discussed above.

Thus, we are unable to ascertain that our model exhibits a freezing transition. We find it more likely that it does not posses this feature.

6 Conclusions and Future Prospects

We have investigated a two-dimensional lattice model of polymers, subject to a quenched random short-range interaction. Our study follows the one done by Kantor and Kardar [30] in three dimensions. The lower dimension allows us to enumerate longer polymer chains: We explore all spatial conformations and charge quenches for a neutral polymer 16 monomers long, a procedure we have not seen in the literature to have been previously performed. We also extend the study to issues they did not investigate — the estimation of the critical exponent ν_θ, and the possible existence of a glass-like freezing transition.

We have seen that a neutral polymer undergoes a tricritical θ-transition at a temperature $T_\theta = 0.83 \pm 0.02$, with a critical exponent $\nu_\theta = 0.60 \pm 0.02$. The value of ν_θ seems to be different from its value for homopolymer collapse (for which there is a wide range of estimates). This conclusion is supported by a simulation we performed of homopolymers. If indeed $\nu_\theta^{\rm homo} \neq \nu_\theta^{\rm hetero}$, this would mean that heterogeneity is a significant perturbation in this model. We have reviewed several theoretical references of possible relation to this question, but they do not seem to point at a definite answer. This could be a subject for future analytic study, using, perhaps, renormalization-group methods.

For non-neutral polymers, we have observed a decrease in the θ-temperature with increasing excess charge, up-to a value $X_{\rm cr}$ where the collapse vanishes and SAW behavior prevails at all temperatures. We have drawn the phase diagram in the (X, T) plane and found it similar to that obtained for 3-d [30].

We have investigated the ground state and the energy landscape of a neutral polymer, in attempt to find evidence for the existence of a glass-like freezing transition for such a polymer. We do not find much evidence for this transition, a result which is in accordance with theoretical predictions for this model. Any attempt to observe a freezing phenomenon would require altering the model. A recommended change to be tested is the addition of a homopolymer-like attracting term in the Hamiltonian. This would raise the collapse temperature T_θ and would possibly enable us to observe the freezing transition, should it occur at some lower temperature (as is contemplated for models incorporating the H-P interaction [32]).

A Numerical Aspects of the Simulation Process

We will now discuss in more detail the various aspects of the numerical methods used in our work: the complete enumeration and the Monte Carlo simulation. The basic procedures are described in Section 3.3.

A.1 Enumeration of Spatial Configurations

In the exact enumeration process, we examine all possible spatial conformations for a chain of a given length. The conformational enumeration process uses a depth-first algorithm [4], which seeks the longest branch of the self-avoiding walk. The algorithm backtracks either when the full chain of a given length has been generated or there is a dead-end due to an excluded-volume violation. In order to reduce the number of mathematical operations required for each walk, the calculation of the various quantities (E, R_g^2 etc.) is done by calculating, at each step, certain differences, which are the changes caused by the addition of a new monomer or the subtraction of one. For example, we keep variables containing the sums of the x and y coordinates of the current walk, as well as their squares (i.e. $\sum_i x_i, \sum_i y_i, \sum_i {x_i}^2, \sum_i {y_i}^2$). This enables us to calculate, for each full-length walk, the radius of gyration thus:

$$R_g^2 = \frac{1}{N+1}\sum_{i=0}^{N}(x_i^2+y_i^2) - \frac{1}{(N+1)^2}[(\sum_{i=0}^{N}x_i)^2 + (\sum_{i=0}^{N}y_i)^2] \; . \tag{A.1}$$

Similarly, we calculate at each step the difference in energy caused by the added new monomer i

$$dE_i = \sum_j q_i q_j \tag{A.2}$$

for all monomers j nearest neighbors to i, but not adjacent along the chain. This enables us to calculate the energy of each completed walk simply by using the current sum of differences dE_i.

For each walk of the desired length, we put the measured variables in a multi-rowed histogram, whose columns are the discrete ($\Delta E = 1$) energy levels. The first row of the histogram contains the number of walks having each energy level (i.e. $n(E_i)$), whereas the other rows contain the summing of the various measured quantities for each energy level: For example, column i in the second row of the histogram is equal to $\sum_j R_g^2(E_j = E_i)$. Such histograms are created for each quench of charges, containing data of all possible

conformations, and than put in a file. This data is then used to calculate thermal averages for the desired observables, at any given temperature. For example, to calculate the average square radius of gyration at a temperature T we use

$$\langle R_g^2 \rangle_T = \frac{1}{Z_T} \sum_{E_i} e^{-E_i/T} (\sum_j R_g^2(E_j = E_i)) = \frac{1}{Z_T} \sum_{E_i} e^{-E_i/T} \langle R_g^2 \rangle_{E_i} n(E_i), \tag{A.3}$$

where $Z_T = \sum_{E_i} e^{-E_i/T} n(E_i)$, and average further over the different quenches.

A.2 Monte Carlo Simulation

A.2.1 Basic Definitions

Dynamic Monte Carlo methods [27] are based on some stochastic Markov process, where subsequent configurations $\overline{\mathbf{X}}_n$ of the system are generated from the previous configuration $\{\overline{\mathbf{X}} \to \overline{\mathbf{X}}' \to \overline{\mathbf{X}}''\}$ with some transition probability $W(\overline{\mathbf{X}} \to \overline{\mathbf{X}}')$. Various methods differ by the choice of the elementary step. For the choice of transition probability we require the principle of *detailed balance* with the equilibrium distribution $P_{\rm eq}(\overline{\mathbf{X}})$:

$$P_{\rm eq}(\overline{\mathbf{X}}) W(\overline{\mathbf{X}} \to \overline{\mathbf{X}}') = P_{\rm eq}(\overline{\mathbf{X}}') W(\overline{\mathbf{X}}' \to \overline{\mathbf{X}}) \; . \tag{A.4}$$

In the athermal case (pure SAW problem) each configuration has exactly the same weight. Then Eq. A.4 implies that the probability to select a motion $\overline{\mathbf{X}} \to \overline{\mathbf{X}}'$ must be the same as the probability for the inverse transition $\overline{\mathbf{X}}' \to \overline{\mathbf{X}}$. If there is an additional energy $E(\overline{\mathbf{X}})$ depending on the configuration $\overline{\mathbf{X}}$, the equilibrium distribution is $P_{\rm eq} \sim e^{-E(\overline{\mathbf{X}})/T}$ and hence Eq. A.4 leads to the requirement

$$\frac{W(\overline{\mathbf{X}} \to \overline{\mathbf{X}}')}{W(\overline{\mathbf{X}}' \to \overline{\mathbf{X}})} = e^{-[E(\overline{\mathbf{X}}') - E(\overline{\mathbf{X}})]/T}. \tag{A.5}$$

Following Metropolis et al. [39], it is common practice to take the transition probability as for the athermal case (which also results for the model in the limit $T \to \infty$) but multiply it with a factor $e^{-[E(\overline{\mathbf{X}}') - E(\overline{\mathbf{X}})]/T}$ if $E(\overline{\mathbf{X}}') - E(\overline{\mathbf{X}}) > 0$ and leave it unchanged if $E(\overline{\mathbf{X}}') - E(\overline{\mathbf{X}}) \leq 0$.

In practice, at every step of the algorithm one performs a trial move $\overline{\mathbf{X}} \to \overline{\mathbf{X}}'$. If $W(\overline{\mathbf{X}} \to \overline{\mathbf{X}}')$ is zero (SAW condition being violated), the old configuration is counted once more in the averaging, and the procedure is iterated. If $W(\overline{\mathbf{X}} \to \overline{\mathbf{X}}')$ is unity, the new configuration is accepted and counted in the averaging. If $0 < W < 1$, we need a

(pseudo-) random number ζ uniformly distributed between zero and one. We compare ζ with W: if $W \geq \zeta$ we accept the new configuration and count it, while if $W \leq \zeta$ we reject the trial configuration and again count once more the old configuration.

In the limit where the number of configurations M generated tends to infinity, the states $\overline{\mathbf{X}}$ generated with this procedure are distributed proportional to the equilibrium distribution $P_{\rm eq}(\overline{\mathbf{X}})$, provided there is no problem with the *ergodicity* of the algorithm. The canonical average of any observable $A(\overline{\mathbf{X}})$ is then approximated by the simple arithmetic average

$$\langle A \rangle \approx \overline{A} = \frac{1}{M - M_0} \sum_{t=M_0+1}^{M} A_t \ , \tag{A.6}$$

which can be interpreted as a time average if we associate a (pseudo-) time variable t with the label of successively generated configurations. Note, that in a dynamic MC simulation, the starting configuration is not representative of the equilibrium distribution (e.g. one may start – as indeed we do – with a completely stretched-out chain). So the system at the beginning of the simulation needs to be "equilibrated": This is why the first M_0 configurations in Eq. A.6 are omitted from the averaging.

The sample mean $\overline{A}$ has a variance given by [37]:

$$\mathrm{Var}(\overline{A}) \approx 2C_{AA}(0)\frac{\tau}{M - M_0} \ , \tag{A.7}$$

where τ is the relaxation time of the autocorrelation function

$$C_{AA}(t) \equiv \langle A_s A_{t+s} \rangle - \langle A \rangle^2 \ , \tag{A.8}$$

which is typically assumed to decay exponentially.

A.2.2 Simulation of the Polymer Chain

After rather unsuccessful attempts to use two other algorithms, the "Slithering Snake" method [27] and the relatively new method of Grassberger and Hagger [40]–[41], we have arrived at using the "pivot" algorithm [27] [37] [38], described schematically in Fig. 3: Starting from some configuration of a SAW, we randomly select one site of the walk. A symmetry operation of the lattice (rotation or reflection) is applied to the part of the walk subsequent to the selected site, using this site as the origin. The choice of this symmetry operation is random.

Various parameters used in the algorithm must be set correctly in order to overcome a few specific problems:

1. A sufficiently large number of first configurations must be omitted from averaging, to allow "equilibration".

2. At low temperatures, chains tend to get "stuck" in local energy minima and cul-de-sacs (dead-ends). So, a number of independent runs must be performed at each temperature to decrease the effect of this problem.

3. At each run, the number of iterations (attempted moves) must be sufficient for the number of independent samples to be large enough, taking into account the correlation between consecutive samples.

Using the various quality indicators described below, and noting that the severeness of the problems above depends strongly on the chain length N and temperature T, we have converged to the following values:

- Number of Runs $= \frac{A}{T}$, with A taken between 10 and 30.
- Number of iterations $= B\frac{N^2}{T}$, with $B = 30$ (For $T > 1$, the value taken is BN^2).
- Number of (first) samples discarded $= 0.2$ times the number of iterations.

A.2.3 Quality Indicators

We will now discuss the quality indicators used in conjunction with our MC simulation. The example used will be of a neutral 30-monomer chain, simulated at two extreme temperatures: $T = 10$ and $T = 0.4$. The high temperature will enable us to compare our results with the results of Madras and Sokal [37], who deal with the athermal (pure SAW problem) case. The low temperature, on the other hand, will enable us to demonstrate some of the problems encountered during the simulation — problems which are more pronounced the lower the temperature.

The first indication we can get about the quality of the simulation is from simply examining the raw samples obtained in the MC process. These samples are shown in Figs. 26–29. The samples of energy and radius of gyration at high temperature (Figs. 26–27) are characterized by rapid changes, and seem to cover nicely their range of values: For example, samples of E as high as $+9$ and as low as -8 can be observed.

The situation is drastically different at low temperature (Figs. 28–29). Here it is seen, that after an initial collapse to some value, this value not only changes just slightly, but changes occur just once in a few hundred or a few thousand iterations! This may imply, that the limited number of values observed is not necessarily a reliable representation of the limited configuration space at low temperature, but may imply that the simulated chain gets stuck at local minima, and rarely gets out to explore other subspaces of configuration space.

Another point arising from examining the samples, at both high and low temperatures, is that the "equilibration" value taken, i.e. discarding of the first 20 percent of the samples, is very much on the safe side, and no problems should occur regarding that issue.

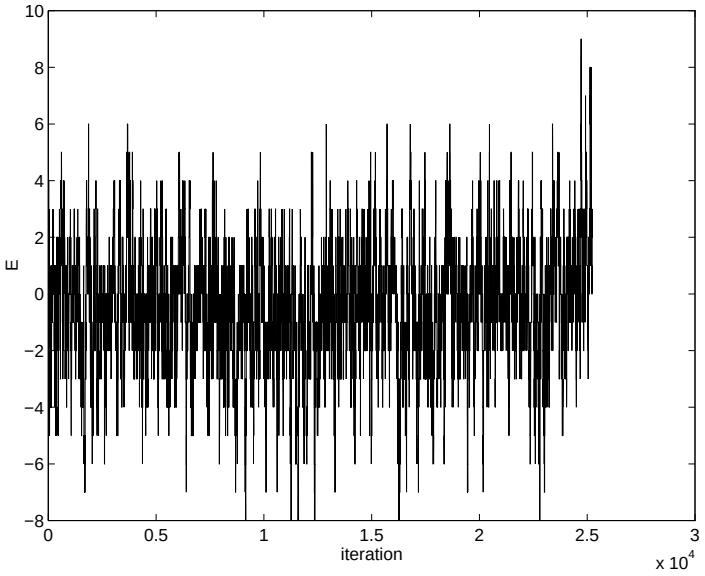

Figure 26: MC raw samples of energy, for a 30-monomer neutral chain (one quench), at $T = 10$.

A more quantitative indication for the quality of our simulation is achieved by examining the auto-correlation of the various sampled observables. Actually, we examine the *normalized autocorrelation function*

$$\rho_{AA}(t) = \frac{C_{AA}(t)}{C_{AA}(0)} \ , \tag{A.9}$$

where $C_{AA}(t)$ is defined in Eq. A.8. This is shown for R_g^2 in Figs. 30–31. The figures also present the forecast made by Madras and Sokal, assuming an exponential decay of the correlation function and assigning τ the value quoted by them [37]:

$$\tau \approx AN^p \ , \ A \approx 8 \ , \ p \approx 0.2 \tag{A.10}$$

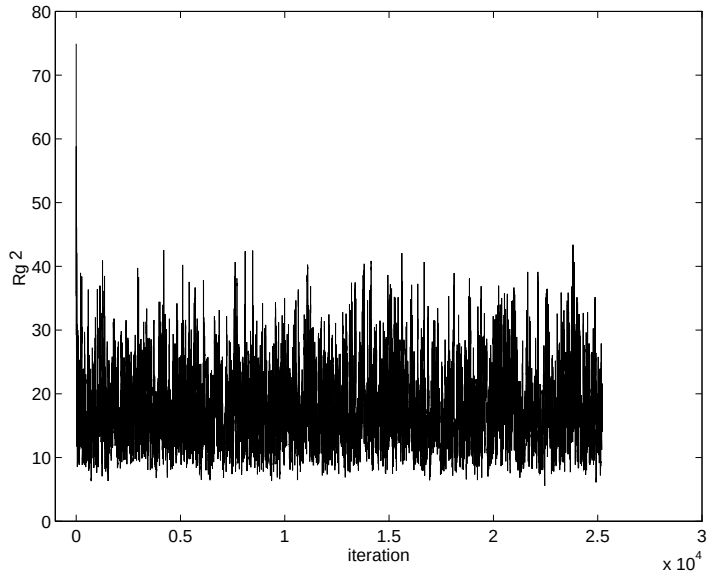

Figure 27: MC raw samples of squared radius of gyration, for a 30-monomer neutral chain (one quench), at $T = 10$.

(giving $\tau \approx 15$ for $L = 30$). It is seen that at high temperature (Fig. 30) the measured autocorrelation is "in the right ball-park" of the given prediction, with the actual function decaying faster than the prediction at first, and then fluctuating around zero value. At low temperature (Fig. 31), on the other hand, correlation decays much slower than predicted for the athermal case: It reaches $1/e$ value at a distance of more than 700 samples, and 0.1 at more than 2700. The meaning of this is obvious: At low temperature the samples are highly correlated, and a much bigger number of samples has to be taken in order to achieve a desired number of independent configurations.

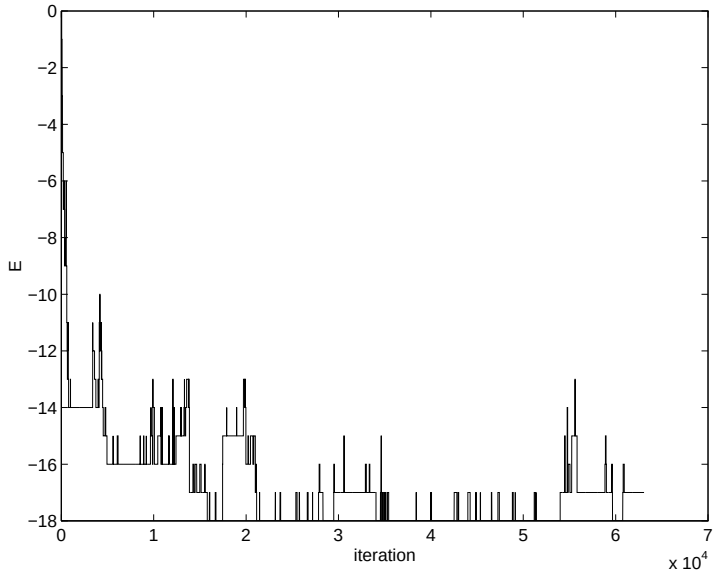

Figure 28: MC raw samples of energy, for a 30-monomer neutral chain (one quench), at $T = 0.4$.

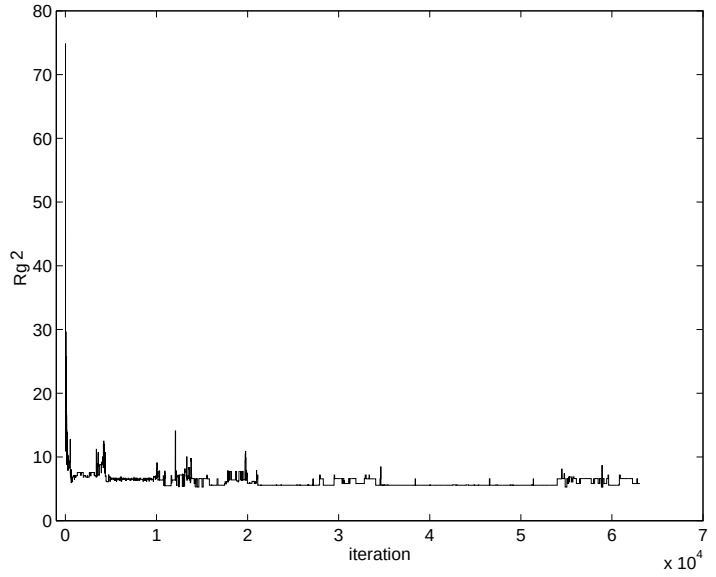

Figure 29: MC raw samples of squared radius of gyration, for a 30-monomer neutral chain (one quench), at $T = 0.4$.

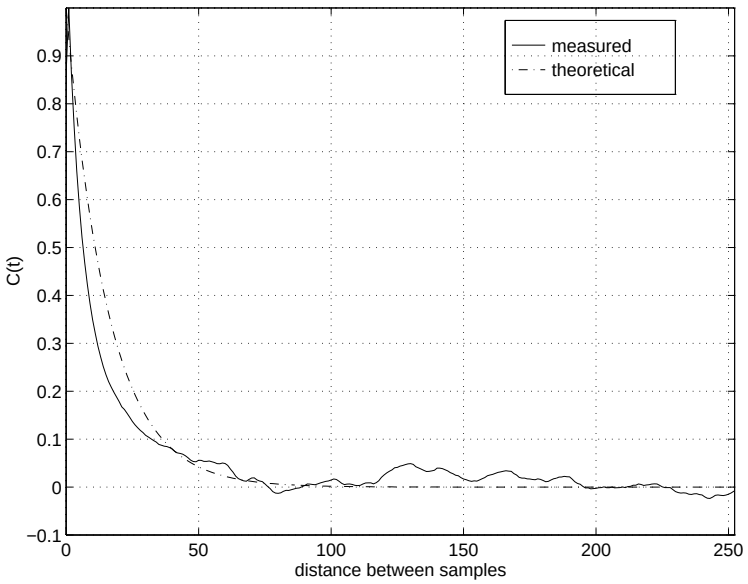

Figure 30: Normalized auto-correlation of R_g^2, for a 30-monomer neutral chain (one quench), at $T = 10$. The theoretical curve follows after Eq. A.10.

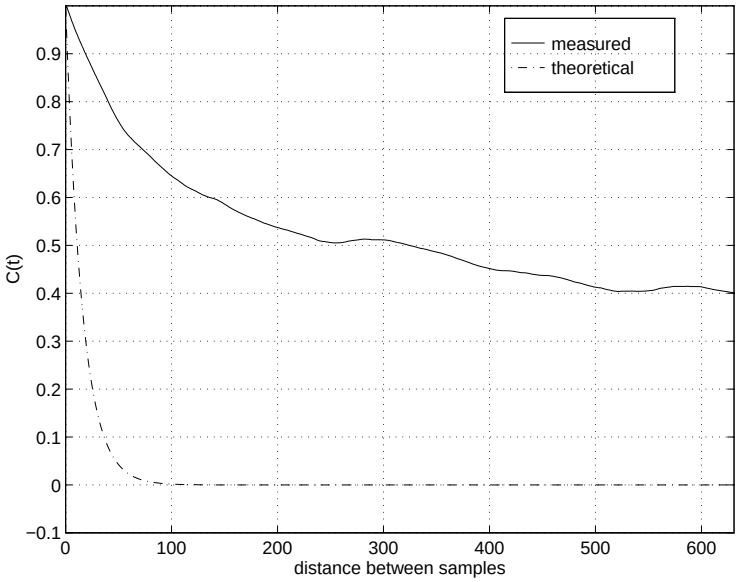

Figure 31: Normalized auto-correlation of R_g^2, for a 30-monomer neutral chain (one quench), at $T = 0.4$. The theoretical curve follows Eq. A.10.

Additional criterions used are the fraction of accepted moves f and distribution of this acceptance over the various chain sites $f(k)$ (k being the site along the chain). The values given in [37] are:

$$f \approx N^{-p} \ , \ p \approx 0.2 \ , \tag{A.11}$$

$$f(k) \sim \frac{n_N}{n_k n_{N-k}} \ , \tag{A.12}$$

where n_M is the number of M-step SAWs (see Eq. 2.13). For $N = 30$, the expected (athermal) acceptance fraction is $\approx 30^{-0.2} \approx 50\%$. At $T = 10$, the total number of accepted moves (after the "equilibration period") was indeed within 5 percent error of this value. At $T = 0.4$, however, the number was more than ten times lower — i.e. ten times as much samples must be taken to reach the same number of different configurations.

The severity of the problem is made clearer by examining how the accepted moves (again, only those after the "equilibration period") were distributed over the different chain sites. This is done in Figs. 32–34. The first of them shows the theoretical estimation of the distribution (Eq. A.12). The main feature of this figure is the preference for sites closer to the chain edge — this is quite obvious, since acting on small chain segments will less probably lead to violation of self avoidance. This feature is preserved (up to statistical errors) in the results for $T = 10$ (Fig. 33). The results at low temperature (Fig. 34), however, are drastically different: The "dislike" of inner-chain moves has turned into an almost total annihilation of such moves. That is, almost all of the accepted moves after equilibration are of the outermost chain parts. This further strengthens our suspicion that at low temperature the process gets stuck at local energy minima and loses ergodicity.

To conclude, when dealing with low temperatures, we must both significantly increase the number of attempted moves, and perform a considerable number of independent runs, if we wish to maintain the ergodicity and statistical meaningfulness of the algorithm. These requirements, after tested for various temperatures and chain lengths, lead to the values chosen and listed in subsection A.2.2. Still, one has to be careful when examining results at low temperatures.

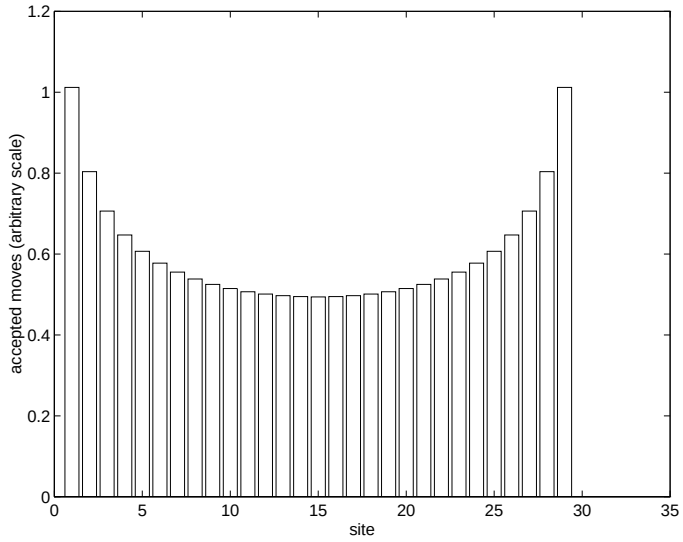

Figure 32: Theoretical acceptance distribution for a 30-monomer chain, after Eq. A.12.

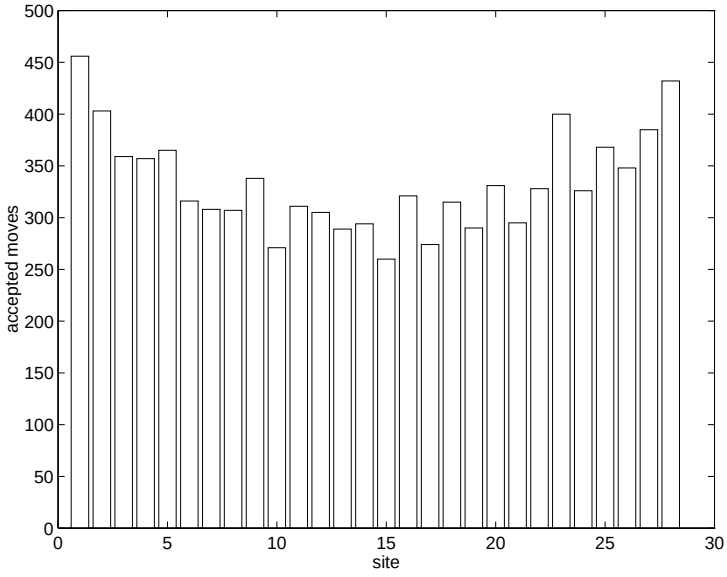

Figure 33: Acceptance per chain site, for a 30-monomer neutral chain (one quench), at $T = 10$.

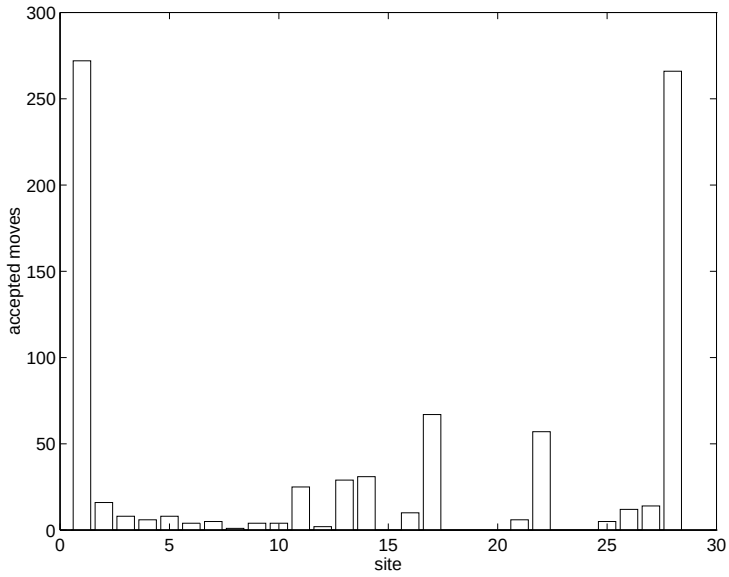

Figure 34: Acceptance per chain site, for a 30-monomer neutral chain (one quench), at $T = 0.4$.

A.3 Comparison of Enumeration and MC Results

For not-too-long chains, we can validate the results of the MC simulation by comparing the results (for a specific quench) with those of the exact enumeration[1]. This is done in Figs. 35–36, showing the radius of gyration and heat capacity of a 20-monomer neutral polymer (one quench). The parameters of the MC simulation were, in accordance with subsection A.2.2:

- Number of independent runs: $\frac{30}{T}$.
- Number of iterations at each run: $30\frac{N^2}{T}$ for $T \leq 1$, $30N^2$ for $T > 1$.
- Samples discarded: First 20% of each run.

The results are quite satisfactory: The MC results fall close to the exact ones. Averaging over more than a single quench (as done in the actual simulation) further improves the situation. However, we can observe that even for a relatively short chain (20), results become poorer at low temperature: The results for both R_g^2 and C/N depart from the exact line below $T \approx 0.5 - 0.6$.

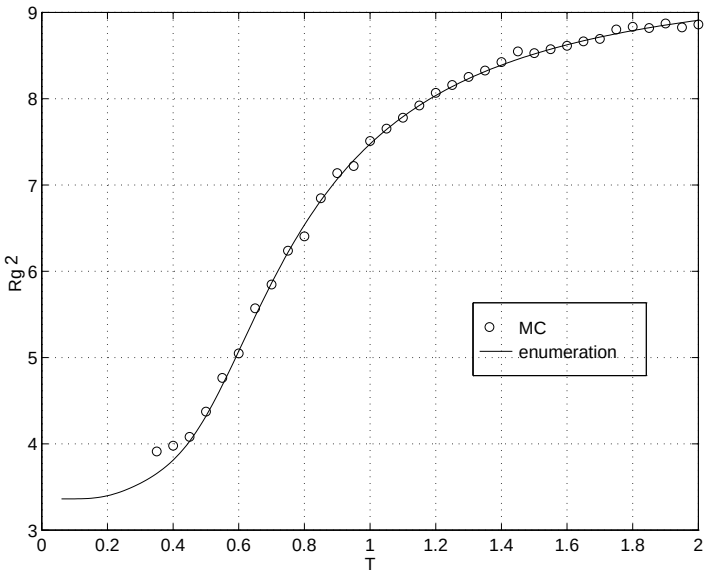

Figure 35: Squared radius of gyration vs. temperature for a 20-monomer neutral chain. Results of exact enumeration and Monte Carlo simulation (single quench) are shown.

[1]This is also a useful way to find "bugs" in the enumeration program ...

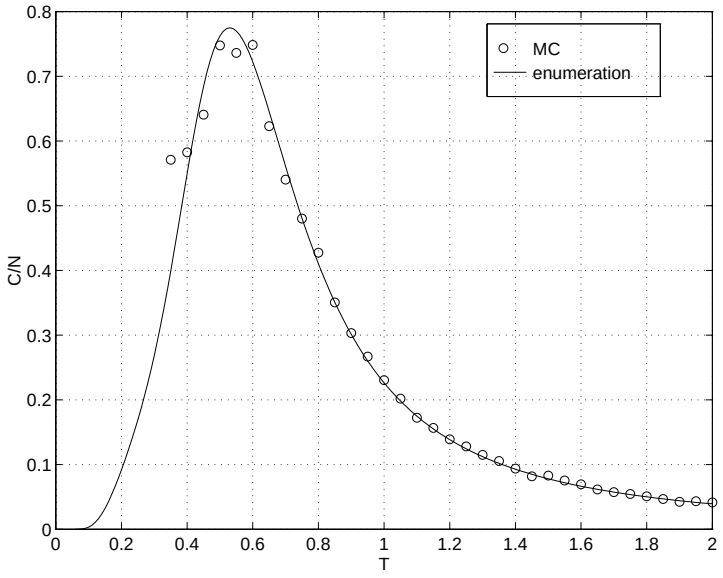

Figure 36: Heat capacity per monomer vs. temperature for a 20-monomer neutral chain. Results of exact enumeration and Monte Carlo simulation (single quench) are shown.

A.4 Hardware and Software Details

The simulations were performed using a Sun SPARC-20 workstation. Programs were written in C language and compiled using GNU C++ compiler (version 2.7.2). Additional data analysis and graphics were performed with MATLAB software (version 4.2a).

Maximal running time for a single simulation was 24 hours, for the complete enumeration of the 15 step (16 monomers) neutral polymer — all conformations and quenches.

References

[1] H. S. Chan and K. A. Dill, *Physics Today* p. 24, February 1993.

[2] T. E. Creighton (Ed.), *Protein Folding* (Freeman, New York 1992).

[3] W. R. Taylor (Ed.), *Patterns in Protein Sequence and Structure* (Springer-Verlag, New York 1992).

[4] K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).

[5] J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **84**, 7524 (1987).

[6] B. Derrida, *Phys. Rev. B* **24**, 2613 (1981).

[7] H. Frauenfelder and P. G. Wolynes, *Physics Today* p. 58, February 1994.

[8] E. Shakhnovich and A. Gutin, *J. Chem. Phys.* **93**, 5967 (1990).

[9] E. I. Shakhnovich and A. M. Gutin, *Biophys. Chem.* **34**, 187 (1989).

[10] P. G. De Gennes, *Scaling Concepts in Polymer Physics*, (Cornell University Press, Itacha, 1979).

[11] M. Plischke and B. Bergerson, *Equilibrium Statistical Physics*, (World Scientific, Singapore, 1994).

[12] P. Flory, *Principles of Polymer Chemistry*, (Cornell University Press, Itacha, 1953).

[13] A. Baumgartner, *J. Physique* **43**, 1407 (1982).

[14] C. Domb, *Polymer* **15**, 259 (1974).

[15] K. Kremer, A. Baumgartner and K. Binder, *J. Phys. A* **15**, 2879 (1981).

[16] I. Nishio, S. T. Sun, G. Swislow and T. Tanaka, *Nature* **281**, 208 (1979).

[17] P. G. de Gennes, *J. Physique Lett.* **36**, L-55 (1975).

[18] M. Daoud and G. Jannink, *J. Physique* **37**, 973 (1976).

[19] J. A. Marqusee and J. M. Deutch, *J. Chem. Phys.* **75**, 5179 (1981).

[20] A. Baumgartner and M. Muthumukar, In *Advances in Chemical Physics, Vol. 94 – Polymeric Systems*, Eds. I. Prigogine and S. A. Rice, (John Wiley and Sons, New York 1996).

[21] A. B. Harris, *Z. Phys. B* **49**, 347 (1983).

[22] Y. Kantor and M. Kardar, *Europhys. Lett.* **14**, 421 (1991).

[23] S. Stepanow, M. Schulz and J. U. Sommer, *Europhys. Lett.* **19**, 273 (1992).

[24] V. S. Pande, A. Y. Grosberg, C. Joerg and T. Tanaka, *Phys. Rev. Lett.* **76**, 3987 (1996).

[25] T. Ishinabe, *J. Phys. A* **18**, 3181 (1985).

[26] P. G. de Gennes, *J. Physique Lett.* **39**, L-299 (1978).

[27] K. Kremer and K. Binder, *Comp. Phys. R.* **7**, 259 (1988).

[28] R. Vilanove and F. Rondelez, *Phys. Rev. Lett.* **45**, 1502 (1980).

[29] S. P. Parker (Ed.), *McGraw Hill Dictionary of Scientific and Technical Terms*, (McGraw Hill, New York 1989).

[30] Y. Kantor and M. Kardar, *Europhys. Lett.* **28**, 169 (1994).

[31] H. S. Chan and K. A. Dill, *J. Chem. Phys.* **95**, 3775 (1991).

[32] A. Dinner, A. Sali, M. Karplus and E. Shakhnovich, *J. Chem. Phys.* **101**, 1444 (1994).

[33] C. J. Camacho and T. Schanke, preprint, COND-MAT / 9604174.

[34] H. Li, R. Helling, C. Tang and N. Wingreen, *Science* **273**, 666 (1996).

[35] D. Dressler and H. Potter, *Discovering Enzymes*, (Scientific American Library, New York 1990).

[36] V. S. Pande, A. Y. Grosberg, C. Joerg, M. Kardar and T. Tanaka, preprint, COND-MAT / 9609062.

[37] N. Madras and A. D. Sokal, *J. Stat. Phys.* **50**, 109 (1988).

[38] M. Lal, *Molec. Phys.* **17**, 57 (1969).

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

[40] P. Grassberger and R. Hagger, *J. Chem. Phys.* **102**, 6681 (1994).

[41] P. Grassberger and R. Hagger, *Europhys. Lett.* **31**, 351 (1995).

[42] H. Meirovitch and H. A. Lim, *J. Chem. Phys.* **91**, 2544 (1989).

[43] A. K. Roy, B. K. Chakrabarti and A. Blumen, *J. Stat. Phys.* **61**, 903 (1990).

[44] K. Barat, S. N. Karmakar and B. K. Chakrabarti, *J. Phys. A: Math. Gen.* **25**, 2745 (1992).

[45] B. Derrida and H. Saleur, *J. Phys. A: Math. Gen.* **18**, L-1075 (1985).

[46] L. Kholodenko and K. F. Freed, *J. Chem. Phys.* **80**, 900 (1984).

[47] V. Privman, *J. Phys. A: Math. Gen.* **19**, 3287 (1986).

[48] F. Seno and A. L. Stella, *J. Phys. (France)* **49**, 739 (1988).

[49] M. Kardar, private communication.

[50] J. M. Victor and J. B. Imbert, *Europhys. Lett.* **24**, 189 (1993).

[51] C. D. Stafos, A. M. Gutin and E. I. Shakhnovich, *Phys. Rev. E* **50**, 2898 (1994).

[52] V. S. Pande, A. Y. Grosberg and T. Tanaka, *Phys. Rev. E* **51**, 3381 (1995).

[53] E. I. Shakhnovich, *Phy. Rev. Lett.* **72**, 3907 (1994).

[54] C. D. Stafos, A. M. Gutin and E. I. Shakhnovich, *Phys. Rev. E* **48**, 465 (1993).

[55] E. I. Shakhnovich and A. M. Gutin, *J. Phys. A* **22**, 1647 (1989).

[56] H. S. Chan and K. A. Dill, *J. Chem. Phys.* **99**, 2116 (1993).

[57] A. Baumgartner, *J. Chem. Phys.* **72**, 871 (1980).