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

Randomly Charged Polymers as One Dimensional Random Walks

by

Shay Wolfling

Thesis submitted in partial
fulfillment of the requirements
for the M.Sc. degree
at Tel-Aviv University
School of Physics and Astronomy

The research work for this thesis has been
carried out under the supervision of
Professor Yacov Kantor

December 1997

Acknowledgments

I wish to thank my supervisor, Professor Yacov Kantor, for his guidance. His knowledge and endless patience made this work possible.
Special thanks are due to my colleagues at work, who carried the burden without me during the course of this work.
I am grateful to my family and friends for their encouragement to accomplish this work.

Contents

Abstract

The ground state of randomly charged polyampholytes (polymers with positively and negatively charged groups along their backbone) is conjectured to have a structure similar to a necklace, made of weakly charged parts of the chain compacting into globules, connected by highly charged stretched 'strings'. We suggest a specific structure, within the necklace model, where all the neutral parts of the chain compact into globules: The longest neutral segment compacts into a globule; in the remaining part of the chain, the longest neutral segment (the 2nd longest neutral segment in the entire chain) compacts into a globule, then the 3rd, and so on. We show that the length of the nth longest neutral segment in a sequence of N monomers is proportional to N/n^2, while the mean number of neutral segments increases as $\sqrt{N}$. The polyampholyte in the ground state within our model is found to have an average linear size proportional to $\sqrt{N}$, and an average surface area proportional to $N^{2/3}$.

We map the charge sequence of the polyampholyte into a one-dimensional random walk, and investigate the size distribution of the longest neutral segments in the chain through their analogy to the longest loops inside the random walk. We generalize the results obtained for a specific class of random walks, in which a unit displacement appears at each step (i.e. each monomer in the chain is charged ± 1), by defining the problem of longest loops for Gaussian random walks. In such walks the probability density of the longest loop depends on the chosen distance ϵ defining a loop. (Two different steps of the walk should be located closer than ϵ from each other so that the segment between them will be called a closed loop). Using Monte-Carlo simulations, along with analytical methods, we obtain and analyze the dependence of the probability density on the length of the chain and on ϵ, for different values of ϵ. We show that, independent of ϵ, the probability density of the longest loop converges with increasing number of steps in the walk to the probability density of the longest loop in random walks with steps of fixed length.

We use a scaling process to obtain a probability density of the longest loop, which (for long walks) is independent of the number of steps and of the nature of the single step of the random walk. This probability density is identical for random walks with steps of fixed length and for Gaussian random walks. We prove that this probability density is universal for large classes of random walks, and we obtain some of its analytical properties.

1 Polymers and Random Walks – Review of Previous Studies

Polymers are multi-linked chains, forming a long and flexible macromolecule. The units of a polymer are called monomers, and are assumed to be unbreakable from each other (in real polymers the bonds between the monomers are unbreakable under normal physiological conditions). Polymers in which all the monomers are identical are called *homopolymers*, and polymers which contain several types of monomers are known as *heteropolymers* or *copolymers*. The desire to understand long chain biological macromolecules, and especially proteins, stimulates extensive studies of polymers [1, 2, 3]. Proteins are polymer macromolecules, in which the monomers are amino acids. Different proteins have different sequences of amino acids. The structure of the protein is determined by the interactions between its constituent amino acids. According to these interactions, the protein folds into a specific shape, which is an energetically stable ground state, that is responsible for its activity and properties.

In this section we review previous studies, and quote known results that are needed to understand our study, or that are relevant to the raised questions, suggested model and obtained results. We begin by a short review of simple polymer models, followed by a discussion of randomly charged polyampholytes, focusing on their ground state energy and structure. In the work we map the problem of size distributions of neutral segments in a randomly charged polymer into similar problems of one dimensional random walks (1-d RW's), and we therefore review related RW problems. We devote a special section to the review of models of randomly broken objects, which are mathematical models related to the problem of dividing the polymer into neutral segments.

1.1 Simple Polymer Models

Given the large number of monomers and the complexity of the interactions in real molecules, the primary method for exploring the structure and properties of real polymers is by introducing very simple models.

One of the simplest models of a polymer is a chain of monomers, separated by bonds of length a, that are free to take any orientation in a d dimensional space. This model is known as the *ideal chain model* or the *freely jointed model* [2], and is actually a $d-$dimensional random walk. The vector between the origin and the end point of the

chain is called the end-to-end vector, and is denoted by $\mathbf{R}_{ee}$.

$$\mathbf{R}_{ee} = \sum_{i=1}^{N} \mathbf{a}_i \ , \tag{1.1}$$

where $\mathbf{a}_i$ are the displacement vectors between two connected monomers (satisfying $|\mathbf{a}_i| = a$ for all i), and the polymer has $N+1$ monomers, labeled $0, \ldots, N$. The square root of the average of $\mathbf{R}_{ee}^2$ over all possible conformations, which is a measure for the average size of the polymer, is proportional to $\sqrt{N}$. A more robust measure of the polymer size is the *radius of gyration*, defined as:

$$R_g^2 = \frac{1}{N+1} \sum_{i=0}^{N} \left(\mathbf{R}_i - \overline{\mathbf{R}} \right)^2 \ , \tag{1.2}$$

where $\mathbf{R}_i$ are the positions of the monomers ($\mathbf{R}_i = \sum_{n=1}^{i} \mathbf{a}_n$), and $\overline{\mathbf{R}} \equiv \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{R}_i$. The square root of the average of R_g^2 is also proportional to $\sqrt{N}$. Stated generally we see that:

$$\sqrt{\langle \mathbf{R}_{ee}^2 \rangle} \sim \sqrt{\langle R_g^2 \rangle} \sim N^{\nu} \ , \tag{1.3}$$

where ν is the scaling exponent ($\nu = \frac{1}{2}$ in the ideal chain). Furthermore, it is was shown [3] that restricting the walk to a lattice or to limited angle rotations, does not change the scaling exponent. This is an example of a universal feature in polymer chains – omitting the details of the chain's structure and extracting a universal scaling exponent, which is valid for many classes of polymers.

As in the case of a simple random walk, the thermodynamic limit (large N) of the freely jointed model can also be reached when the displacement vectors $\mathbf{a}_i$ are no longer of constant length a, but assigned a Gaussian probability density:

$$p(\mathbf{a}_i) \sim e^{-d\mathbf{a}_i^2/2a^2} \ . \tag{1.4}$$

Writing formally the joint probability density of a given configuration $p(\{\mathbf{a}_i\})$ as a Bolzman weight ($e^{-\frac{\mathcal{H}}{k_B T}}$), we can define an effective Hamiltonian:

$$\mathcal{H} = \frac{dk_B T}{2a^2} \sum_{i=1}^{N} \left(\mathbf{R}_i - \mathbf{R}_{i-1} \right)^2 \ , \tag{1.5}$$

where k_B is the Bolzman constant, and T is the temperature. Adopting a continuous description of the chain (rather than of discrete monomers), in which monomer configurations are described by $\mathbf{R}(x)$, where x is the internal label for the position of a monomer

along the chain, the effective Hamiltonian (1.5) becomes [4]:

$$\mathcal{H} = \frac{dk_BT}{2a^2}\int_{x=0}^{N}\left(\frac{d\mathbf{R}}{dx}\right)^2 dx \ . \tag{1.6}$$

One of the important aspects of real polymers is that they cannot self intersect – two monomers cannot come closer than a minimum distance. This is modeled by a *self-avoiding walk* (SAW) – a random walk that can never intersect itself (see, e.g., [5]). For a SAW Eq. (1.3) still holds, but with a value of ν different than $\frac{1}{2}$. The Gaussian model Hamiltonian (Eq. 1.6) can be generalized to describe this excluded volume effect:

$$\frac{\mathcal{H}}{k_BT} = K\int_{x=0}^{N}\left(\frac{d\mathbf{R}}{dx}\right)^2 dx + \omega\int_0^N dx\int_0^N dx'\delta^d\left(\mathbf{R}(x)-\mathbf{R}(x')\right) \ , \tag{1.7}$$

where K and ω are some constants. Using scaling arguments, requiring that the interaction energy be the same on all length scales, a result of $\nu = 3/(d+2)$ is obtained. This result was obtained by Flory [6], based on the approximation of taking the free energy as a sum of an elastic (entropic) energy of a regular RW and a mean field estimate of the repulsive energy, and minimizing the sum in respect to the polymer size. This result for ν is only approximate in 3 dimensions. A current estimate of ν for a 3 dimensional SAW is $\nu = 0.57875 \pm 0.0003$ [7], but there are many other Monte Carlo, exact-enumeration and renormalization-group predictions of ν (see references in [7]). Since the scaling exponent is different than for ideal random walks (leading to a significant swelling of the chain), the SAW is said to be in a different universality class than the ideal chain.

Another important aspect of several real polymers is the presence of electric charge along their backbone. A polymer which contains charged monomers (of only one type) is known as a *polyelectrolyte*. This work deals mainly with *polyampholytes* (PA's), which are polymers with positive and negative charges along their backbone [8]. Models of PA's are important to the study of proteins, since under normal physiological conditions, 5 of the 20 naturally occurring amino acids have an excess charge (three are positively charged and two are negatively charged), each appearing in about 2%-7% of the proteins [1].

1.2 Polyampholytes

We consider a polymer of charged monomers, interacting via unscreened Coulomb interactions. At high temperatures the effect of the electrostatic interaction is small (the

thermal energy dominates), and the structure is similar to that of an uncharged polymer (specifically, R_g is proportional to $N^{0.58}$, as in a SAW). We are interested in the ground state structure of PA's, where the structure becomes sensitive to the charge sequence of the chain, and especially to its total (excess) charge. Throughout this work, we discuss PA's that consist of a random mixture of positive and negative charges, which cannot move along the chain, i.e. they are quenched.

1.2.1 Debye-Hückel Arguments

It was first suggested by Edwards *et al.* [9] that a neutral PA confined to a sphere smaller than its natural radius, so as to have a uniform density, behaves as a regular (micro)electrolyte. Elaborating these arguments, Higgs and Joanny [10] assumed that there exists a collapsed state of the chain, with a volume significantly smaller than the volume of ideal Gaussian chain $(a\sqrt{N})^3$, but significantly larger than the volume of the completely close-packed monomers Na^3. They assumed, that taking advantage of the presence of two types of charges, the electrostatic interactions create correlations between the positions in space of the positive and negative charges, so that every charge is predominantly surrounded by charges of an opposite sign. The Coulomb interactions are thus screened at long distances. Higgs and Joanny treated this effect for charges on a polymer in the same way that free charges in an ionic solution are treated, and calculated the electrostatic energy of the PA as if it were a small volume of a Debye-Hückel electrolyte solution [11]. The calculation resulted in a structure of a single globule of close-packed blobs, where each blob has a typical radius of the Debye-Hückel screening length. This screening length, which is the distance determining the dimensions of the ion cloud due to a given ion, is the length where the electrostatic energy, due to charge fluctuations in the polymer, is equal to the thermal energy. Within the blobs the chain remains a SAW, since the thermal energy dominates, while between the blobs there is a screened electrostatical attraction, packing them closely together (see Fig. 1). This is a collapsed structure, in which $R_g \sim N^{1/3}$, and hence the solution is self-consistent.

Wittmer *et al.* [12] discussed the influence of the distribution of charges on the conformational properties of neutral PA's. They calculated the fluctuations of charge density inside the globule (which lead to the collapse) within the framework of the random phase approximation. By introducing the charge correlations, they obtained an expression which enabled calculation of both completely random neutral sequence (recovering the results

of Higgs and Joanny) and of correlated sequences (recovering results obtained by Victor and Imbert [13] for alternating sequences).

1.2.2 Scaling Arguments

A different approach to the study of the ground state is by scaling arguments, requiring that the interaction energy be the same on all length scales. This approach was taken by Kantor and Kardar [14] for PA's. Adding Coulomb interactions to the Hamiltonian of an excluded volume polymer (Eq. 1.7), we get:

$$\frac{\mathcal{H}}{k_BT} = K\int_0^N \left(\frac{d\mathbf{R}}{dx}\right)^2 dx + \int_0^N dx \int_0^N dx' \left[\omega\delta^d\left(\mathbf{R}(x)-\mathbf{R}(x')\right) + \frac{q(x)q(x')}{k_BT|\mathbf{R}(x)-\mathbf{R}(x')|^{d-2}}\right] \tag{1.8}$$

where $q(x)$ is the charge density, interacting through a d-dimensional Coulomb interaction decreasing with range as $1/r^{d-2}$.

In order to explain the nature of the scaling process, we will first derive it (after [15]) for homogeneously charged polymers (polyelectrolytes), where $q(x)$ is constant and equal some q_0: Under rescaling of the internal coordinate $x \to \lambda x$, and the external coordinate $\mathbf{R} \to \lambda^\nu \mathbf{R}$, the coefficients of the first (entropic), second (SAW) and third (electrostatic) terms in Eq. (1.8) scale respectively as $\lambda^{2\nu-1}$, $\lambda^{2-d\nu}$ and $\lambda^{2-(d-2)\nu}$. For $d > 6$, choosing $\nu = \frac{1}{2}$ leaves the entropic term of Eq. (1.8) unchanged under rescaling (since the exponent equals zero), while the remaining terms decay to zero (since their exponents become negative). Thus, excluded volume and Coulomb interactions are irrelevant above six space dimensions. For $d < 6$, the electrostatic term is most relevant, and is set to be unchanged under rescaling by choosing $\nu = 2/(d-2)$. This value of ν is unity for $d = 4$, and below this dimension the polymer is fully stretched. Through a full renormalization-group treatment, Pfeuty *et al.* [15] proved that the above result is exact, while numerical results in 4 and 5 dimensions [16] support this picture.

In applying similar scaling arguments to PA's (after [14]), one should note that due to the averaging of independent charges, the random product $q(x)q(x')$ in Eq. (1.8) contributes an additional factor of $1/\lambda$ to the rescaling of the electrostatic term. (The sum of λ random charges increases as $\sqrt{\lambda}$, as opposed to the λ increase for polyelectrolytes.) The coefficient of the electrostatic term now scales as $\lambda^{1-(d-2)\nu}$, and we find that $\nu = 1/(d-2)$, for $3 \le d \le 4$, and a stretched polymer ($\nu = 1$) is obtained for $d \le 3$.

Qualitative views of the PA chain structure in the Debye-Hückel and in the scaling approaches are depicted in Fig. 1 (figure from [21]). According to the Debye-Hückel motivated arguments the structure is a collapsed single globule of blobs, while according to scaling arguments the polymer is stretched.

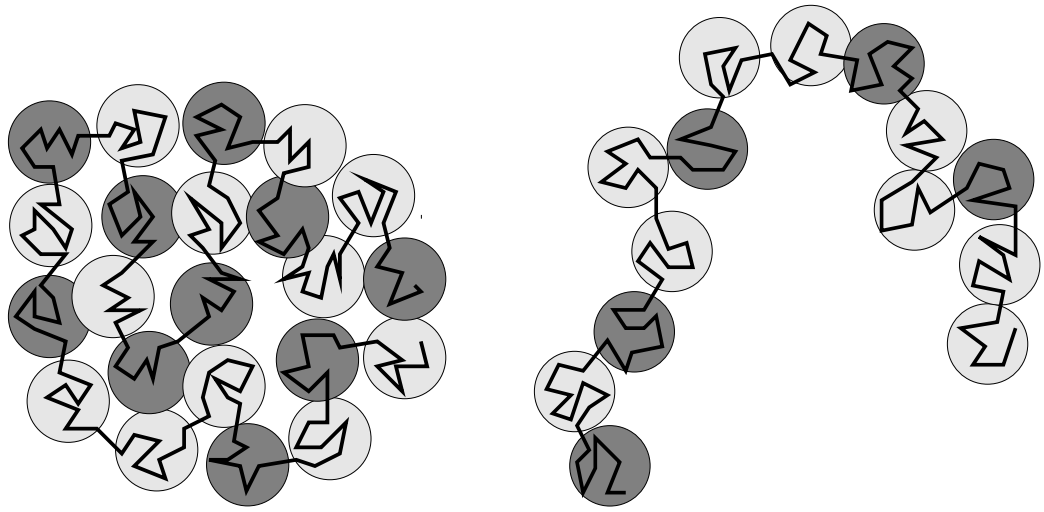

Figure 1: Qualitative views of the spatial arrangement of blobs in a PA, according to Debye-Hückel arguments, where the chain collapses (left), and according to scaling arguments, where the chain stretches (right). Lighter and darker blobs are positively or negatively charged blobs. Electrostatic interaction within each blob is dominated by thermal energy.

The apparent contradiction between the scaling and the Debye-Hückel motivated arguments, was resolved by Kantor, Li and Kardar [17, 18], noting that the low-T ground state of the PA is extremely sensitive to its overall (excess) charge Q. The overall neutrality of a PA is an essential condition for the validity of the Debye-Hückel approximation. (The formal derivation of the approximation [11] begins with the free energy of the uncharged gas, attempting to evaluate the corrections due to Coulomb energy, under the assumption that introduction of charges without change of particle positions does not change the energy. This is true only for completely neutral systems.) Under the neutrality constraint, the PA collapses at low temperatures. In the absence of this constraint (i.e. the complete ensemble of all quenches), where the neutral configurations have negligible probability, and the typical excess charge is $Q \sim \sqrt{N}$, the dimensional analysis, predicting $\nu = 1$ is valid. These statements are supported by Monte Carlo (MC) simulations [17, 18] and by variational mean field calculations [19].

1.2.3 Energy and Geometry of Polyampholytes

Further studies of the size and structure of randomly charged PA's, and especially the dependence on the excess charge, were performed by Kantor and Kardar [20, 21, 22]. Numerical results show that the radius of gyration of a PA strongly depends on its charge Q, and is very weakly influenced by other details of the sequence. (Srivastava and Muthukumar [23] claim, however, after studying different randomly charged neutral PA's by MC simulations, that the locations of charges do play a crucial role in the conformational behavior of neutral PA's). At high temperatures, the typical electrostatic energy of a configuration, composed of a random sequence of N charges of $\pm q_0$, is shown [21] to change sign at an excess charge of:

$$Q = Q_c \equiv q_0\sqrt{N} \ . \tag{1.9}$$

As the temperature is lowered, chains with excess charge larger than Q_c expand, while those with excess charge lower than Q_c shrink. At low temperatures, R_g^2 is obtained numerically to be monotonically increasing with Q. Furthermore, for small Q the increase is small, while for $Q > Q_c$ an extremely fast increase begins. The dependence of R_g^2 on Q suggests that the transition from compact to stretched configurations at low temperatures occurs for $Q \simeq Q_c$.

To explain the above low-temperature results, Kantor and Kardar [20, 21] started with the empirical observation that a neutral PA compacts to a spherical globule. This suggested that the quench-averaged energy can be presented as a sum of condensation and surface energies:[1]

$$E(Q=0) = -\frac{q_0^2}{a}N + \gamma S \ , \tag{1.10}$$

where q_0^2/a is (approximately) the condensation energy gain per particle, since every monomer is predominantly surrounded by monomers of opposite sign, and $\gamma \sim q_0^2/a^3$ is the surface tension, which expresses the fact that the monomers on the surface have no neighbors outside the globule. Uniformly adding charge Q to each configuration, the energy increases by Q^2/R, becoming:

$$E(Q) = -\frac{q_0^2}{a}N + \gamma S + \frac{Q^2}{R} \ . \tag{1.11}$$

The low temperatures critical excess charge was calculated by Kantor and Kardar [21] through an analogy of charged PA's to charged drops. The optimal shape of a PA is

[1] In all the equations of this section we omit dimensionless prefactors of order unity.

obtained by minimizing the overall energy of Eq. (1.11): For small values of Q the surface tension keeps the PA in an approximately spherical shape. However, a spherical drop charged beyond a certain charge, called the Rayleigh charge Q_R, which depends on the surface tension and the volume of the drop, becomes locally unstable to elongation, since the internal pressure (of Coulomb repulsion) exceeds the surface tension. Even before the total charge reaches Q_R, the drop becomes unstable to splitting into two equal drops, separated by an infinite distance. Additional splittings of the drop occur for larger charges. Similar behavior is expected in PA's charged to $Q_R \sim \sqrt{N}$. Although a PA cannot split, the analogy to charged drops can still be exploited: Constraining the structure to maintain its connectivity by attaching droplets with narrow tubes, results in a necklace-type structure of droplets connected by strings.

Dobrynin Rubinstein and Obukhov [24] studied the case of polyelectrolytes, and found that there is a range of temperatures and charge densities, for which a polyelectrolyte in a poor solvent (corresponding to an effective attraction between monomers) has a necklace-type shape, with compact beads joined by narrow strings. By changing the charge or the temperature, the polymer undergoes a cascade of transitions between necklaces with different number of beads. Using the charged drops analogy, Dobrynin *el al.* characterized the structure completely (including the number of beads and strings, their sizes, and the number of monomers in them, for a given temperature and charge density), and concluded that the necklace structure provides a good picture for polyelectrolytes in a poor solvent.

Trying to apply the necklace model to quenched PA's having random charges [20, 21], several difficulties occur due to the randomness. It was noted, for instance, that a situation occurs, in which most spherical shapes are unstable, while there is on average no energetic gain in splitting the sphere into two equal parts. A consistent theoretical picture for random PA's beyond the instability threshold was not found, but a typical PA is conjectured to be composed of compact globules connected by long strings. In order to reduce the electrostatical energy, the globules consist of segments of the chain that are almost neutral (collapsing according to the Debye-Hückel picture), while the strings are formed by highly charged segments. MC simulations and exact enumeration suggest [21, 22] that the linear size of such a PA (the average of R_g over the unrestricted ensemble) grows with the number of monomers faster than the linear size of a SAW, i.e. $\nu > 0.6$.

Motivated by the necklace structure of randomly charged PA's, Kantor and Ertaş [25, 26, 27] discussed the size distribution of the longest segments with total charge Q

("Q-segments") in such PA's. They mapped the charge sequence of the PA to a 1-d RW, and investigated the probability that the longest Q-segment in such a N step RW has length L. A probability density was defined and investigated in the limit when $N, L, Q \to \infty$, while the reduced length $l \equiv L/N$ and reduced charge $q \equiv Q/\sqrt{N}$ are fixed. The probability density of the longest loop in a RW was obtained numerically and investigated analytically, but a complete analytical solution to the problem of longest loop was not found. Within the class of RW's, in which a unit displacement occurs at each step, numerical evidence was presented for a continuum limit, where the properly scaled functions become independent of N. However, the numerical 'proof' of such 'universality' was limited to this particular class of RW's.

The elongation of a PA at low temperatures and at excess charge of $Q \gg q_0\sqrt{N}$ suggested by the necklace model, was also described by Gutin and Shakhnovich [28], using scaling arguments. However, they claimed that the structure of the polymer is an elongated compact globule, composed of blobs. Dobrynin and Rubinstein [29] analyzed the behavior of PA chains in the framework of a two-parameter Flory theory. They found three different regimes for PA's with excess charge $Q > q_0\sqrt{N}$, and also claimed that at low temperatures, the charge density fluctuation causes a collapse into an elongated globule. The aspect ratio of the globule was determined by the excess charge, and was independent of temperature. The chain's length was found, however, to be linear with N (for excess charge $\sim \sqrt{N}$), and it was noted that it was impossible to find the difference between the elongated cylindrical shape, and a necklace configuration.

We note that there are other related systems in which strings and globules coexist: Schiessel and Blumen [30] claim that a neutral PA under external electrical field exhibit a transition similar to that of a charged PA with excess charge below and above $q_0\sqrt{N}$. (This is reasonable since the transition in a charged PA is also induced by an electrical field, due to the charge density fluctuation). Under a moderate field the PA elongates, and as the field increases the PA stretches into an extended form, consisting of a series of blobs, whose sizes increase towards both ends. A similar structure is taken by polymers in strong flows of the solvent [31].

1.3 One-Dimensional Random Walks

We saw in the previous section, that although the analysis of the structure of a randomly charged PA within the necklace model did not result in a consistent theoretical picture, a

key role was played by the neutral segments in each specific chain (forming the beads in the necklace). In section 2, we map the problem of size distribution of neutral segments in a random sequence of charges into a 1-d RW (following [25, 26]). From this analogy, we will see that the distribution of sizes of the longest neutral segments is equivalent to the distribution of longest loops (i.e. segments that return to their origin) in the corresponding RW. We therefore review in this section RW problems and models (relying greatly on [32, 33, 34]), which will help us to analyze problems of longest loops.

1.3.1 Classical Random Walk Problems

The model of 1-d RW describes a particle taking a series of steps of equal length, each step is taken either forward or backward, with equal probability. The probability $W(k,N)$, that the particle arrives at the point k after N steps (when starting from the origin: $k=0$), can be written in terms of the binomial coefficients C_r^n:

$$W(k,N) = C^N_{(N+k)/2} \cdot \left(\frac{1}{2}\right)^N \ . \tag{1.12}$$

The large N limit ($N \gg k$) of this probability is obtained by using the Stirling formula. A continuum limit of the problem is obtained by defining a net displacement $x = ka$ (where a is the length of each step), and a continuous time $t = Nt_0$ (where the particle takes a step every t_0 unit times). The probability of a particle to be at a position between x and $x + \Delta x$ (where we substitute $\Delta k = \Delta x/2a$, since k can take only even or odd values, depending on N) at time t is given by [33]:[2]

$$p(x,t)\Delta x = \frac{1}{a\sqrt{2\pi(t/t_0)}} e^{-\frac{x^2}{2a^2(t/t_0)}} \Delta x \ . \tag{1.13}$$

The average of x^2, regarding this probability density, satisfies:

$$\langle x^2 \rangle = \frac{a^2 t}{t_0} \ . \tag{1.14}$$

It will become evident from our work, that the problem of longest loop is related to several classical RW problems. Among these related issues are the probability density of the maximum and minimum of a RW, the probability of first passage through a given position, and the probability of last return to the origin. A very useful tool in calculating

[2] In the following sections we substitute $t_0 \equiv 1$, and therefore the number of steps N is equivalent to the time t.

these properties is known as the *method of reflections*, in which one reflects the path of a RW about a given point. We will illustrate this method by considering another problem, which is related to the problem of longest loop – a RW in the presence of an absorbing wall (after [33]). We are interested in the probability $W(k_0, N; k_1)$ of a particle, making a RW, to reach the position $k = k_0$ after N steps, at the presence of an absorbing wall at $k = k_1$. (When the particle arrives at k_1 it is 'absorbed' by the wall and stops its movement). We count the number of sequences which lead to k_0, excluding all sequences which arrive at k_1. It is shown, using the reflection method, that every excluded sequence uniquely defines another sequence, leading to the position $(2k_1 - k_0)$: For each sequence

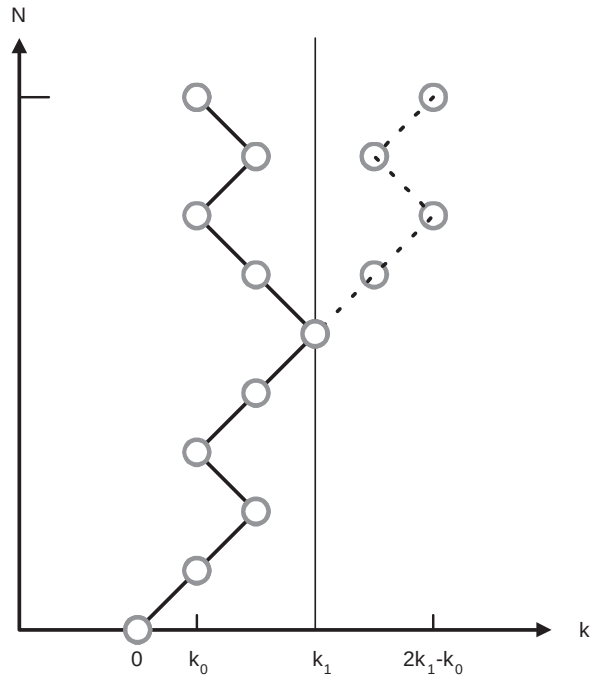

Figure 2: Example of reflection of a path about the line $k = k_1$. Every trajectory arriving at k_0 after passing through k_1 (solid) is reflected by a trajectory which arrives at $2k_1 - k_0$ (dashed), and vise versa.

leading to k_0 and passing through k_1, an 'image' sequence is defined, by reflecting about the line $k = k_1$ the part of the trajectory after its last contact with k_1, before reaching k_0 (see Fig. 2). This sequence is unique, and it reaches the position $k = 2k_1 - k_0$. Similarly, for each trajectory leading to $2k_1 - k_0$ we obtain by reflection a 'forbidden' trajectory leading to k_0. Therefore:

$$W(k_0, N; k_1) = W(k_0, N) - W(2k_1 - k_0, N) \ . \tag{1.15}$$

For the continuum limit (defining $x_1 = k_1 a$ and $x = k_0 a$) this leads to:

$$p(x,t;x_1) = \frac{1}{a\sqrt{2\pi t}}\left[e^{-\frac{x^2}{2a^2 t}} - e^{-\frac{(2x_1-x)^2}{2a^2 t}}\right] . \tag{1.16}$$

The probability $Q(k_1, N)$, that a particle making a RW will arrive at k_1 after N steps, without ever having touched the line $k = k_1$ at any earlier step, (problem known as the first passage problem) can be calculated using $W(k, N; k_1)$. The event of first passage through k_1 in the Nth step occurs if and only if the particle reaches the position $k_1 - 1$ after $N-1$ steps, in the presence of an absorbing wall at k_1, and then takes a step forward (which has probability $\frac{1}{2}$). Hence:

$$Q(k_1, N) = \frac{1}{2}W(k_1 - 1, N-1; k_1) = \frac{k_1}{N}W(k_1, N) . \tag{1.17}$$

For the continuum limit, this probability leads to the probability density $q(x_1, t)$ of a particle to arrive at x_1 for the first time after t steps:

$$q(x_1, t) = \frac{x_1}{t}p(x_1, t) . \tag{1.18}$$

The distribution of T, the first return time to the origin of a RW with long range correlations (i.e. $\langle x^2 \rangle \sim t^{2H}$ for $0 < H < 1$), is found by Ding and Yang [35] to be $P(T) \sim T^{H-2}$, dependent only on the mean square displacements at large times, and not on the distribution of the RW steps. In section 4 we shall attempt to prove that the probability density of the longest loop in a RW is also independent on the distribution of single steps, and depends only on the displacement at large scales.

Let us consider the probability density $M(r, t)$ of the maximum of a RW r after time t. We consider the collection of paths, in which $X(0) = 0$ and $X(t) > r > 0$ (where $X(t)$ is the position at time t), and reflect them about the line in which for the first time the position is r. The reflected paths satisfy $X(t) < r$, and for both the original and the reflected paths the maximum is greater than or equal to r. Therefore, for every path with $X(t) > r$ the reflection argument displays two paths with the same probability, having maximum greater than r. Elaborating this argument, it can be shown [32] that the probability of a $t-$step RW to have a maximum greater than r is twice the probability of the RW to have a position greater than r exactly at the last step:

$$\int_r^\infty M(x,t)dx = 2\int_r^\infty p(x,t)dx , \tag{1.19}$$

which leads for $r \geq 0$ to:

$$M(r,t) = 2p(r,t) \ . \tag{1.20}$$

Similarly, it can be shown [34] that the probability of a $N-$step RW never to return to its origin is equal to the probability that it reaches its starting position exactly at the Nth step. Using this equality, the probability $\alpha(\xi, N)$, that the last return to the origin of a $N-$step RW occurs at the ξth step, is obtained:

$$\alpha(\xi, N) = W(k=0,\xi)W(k=0, N-\xi) \ . \tag{1.21}$$

The first term is the probability to return to the origin at the ξth step, and the second term is the probability to never return to the origin after the ξth step.

Most of the results mentioned in this section can be related to the investigation of longest loops, which start from one specified step. The search for the longest loop of the RW, among *all* possible starting steps, however, creates a more complicated problem, which is actually more related to SAW's, than to regular RW's.

1.3.2 Self-Avoiding Walks

The relation between the problems of longest loops and SAW's lies in the fact that having a longest loop of some length L, means that for lengths greater than L the walk is a SAW. Chains in which $L = 0$ are SAW's, and chains in which L is small are similar to a SAW (for each step, all the steps in the chain, apart from the closest L steps along the chain, obey the SAW statistics). As L increases the similarity to SAW vanishes.

A somewhat similar notion, but for shortest loops rather than for longest loops, was introduced by Fisher and Sykes [36], while trying to find upper bounds to the limit:

$$\mu \equiv \lim_{N\to\infty} \frac{C_{N+1}}{C_N} \ , \tag{1.22}$$

where C_N is the number of $N-$step SAW's on a given lattice. The bounds were found by considering restricted random walks, which are only allowed to intersect themselves after k or more steps (for $k = 2, 3, 4, \ldots$). The problems of finding the $N \to \infty$ limit of $C_{(k)N}/C_{(k)N-1}$ were solved (for small values of k, up to 12) by direct construction of a recurrence relation for $C_{(k)N}$, the total number of kth order restricted walks. As k (which is the length of the shortest allowed loop) increases to infinity, the bounds become closer to the required limit μ.

A different model, related to the problem of longest loops, is called the loop erased self-avoiding walk (LESAW) [37]. A LESAW is constructed from a classical RW, in which any loop generated by a self intersection is erased, resulting in a SAW. The distribution of the LESAW differs from ordinary SAW in that the probability given to a particular $N-$step path depends on the probability that the walk will remain self-avoiding for times $k \geq N$: The LESAW distribution gives zero probability to all 'trapped' paths (i.e. all paths $[x_0, \ldots, x_N]$ for which there exists a $k \geq N$ such that there are no self-avoiding paths $[y_0, \ldots, y_k]$ with $y_i = x_i$ for all $i \leq N$). Numerical results for the critical exponent ν in two and three dimensions obtained by Guttmann an Bursill [38] and by Bradley and Windwer [39] showed that LESAW and SAW are in different universality classes (SAW being closer to an ideal RW). It was shown by Lawler [40] that for dimensions $d > 4$ the LESAW behaves as a simple RW (as does ordinary SAW), and that in $d \leq 4$ the fraction of steps remaining unerased in a LESAW, vanishes with increasing number of total steps. For two and three dimensions it was shown [41], that the scaling exponent ν for LESAW is limited from below by the Flory value for SAW, $\nu = 3/(d+2)$. Within this proof, the probability for the length of the loop originating in the nth step, if the $(n-1)$st step was not erased, was found. Dhar and Dhar [42] studied the distribution of sizes of the erased loops, in the LESAW model, and obtained the probability $P(L)$ of having a loop of length L, for large L's:

$$P(L) \sim L^{-\sigma} \ , \quad \text{where } \sigma = 1 + \frac{2}{\zeta} \ , \tag{1.23}$$

and ζ is the fractal dimension of the loop erased walks in the graph. This result is valid for dimensions $1 < d < 4$ (with corrections when $d < 2$).

We see that none of the results mentioned above for the LESAW model consider the *longest* erased loops, but rather the length of loops in general, which is of less interest to us. Moreover, all the results proven are for space dimensions higher than 1, where as we are interested in the probability of longest loops in a 1-d RW. (It may be possible to apply some of the properties obtained for the probability of loops in the LESAW model to the problem of longest loop in higher dimensions. This problem is addressed at [26], but is beyond the scope of our work.)

1.4 Models of Randomly Broken Objects

In section 2 of the work we will divide a RW into loops and we will therefore investigate not only the longest loop in a given RW, but also the second longest loop (which does not overlap the longest loop - i.e. none of its steps belong to the longest loop), third longest loop and so on. The problem of dividing (almost) the entire RW into non-overlapping loops, is related to the problems of randomly broken objects, introduced by Derrida and Flyvbjerg [43, 44]. In one of these models (the random breaking model), a segment of unit length is divided into mutually exclusive parts by a self-similar random process. The analytical expressions for the probabilities $P_i(W)$ of the ith longest segment (at the end of the cutting process) are found to be different on each interval $1/(k+1) < W < 1/k$ (integer k), and therefore, one expects to see singularities of $P_i(W)$ at all the values $W = 1/k$. These expected singularities were obtained by Derrida *et al.* [43, 44], while investigating numerically the random breaking model and several other models, in which the phase space is broken into non-overlapping unit-sum valleys. Similar singularities, resulting from similar reasons, were obtained by Frachebourg *et al.* [45], while studying the probability distributions of the longest time interval between successive departures and arrivals to the origin of a 1-d RW.

The basic reasoning for the existence of singularities in all these models seems to be applicable to the problem of longest loops in a RW. The longest segment in the random breaking model is analogous to the longest loop, the 2nd longest segment is analogous to the longest loop in the remaining part of the chain (the 2nd longest non-overlapping loop in the chain), and so on (the process is detailed in section 2). Continuing the analogy, we see that $P_n(W)$ is equivalent to the probability of the nth longest loop having length W (when the entire chain is of unit length). The division of a segment into parts in the random breaking model resembles the 'breaking' of a RW into loops: The probability densities of the longest segments in some models of randomly broken objects [43, 44] resemble the probability density of the longest loop in [26].

One main difference between the models described in [43, 44, 45] and the longest loop problem, is that the probabilities of the longest loops are not self-similar. The probability of a certain fraction of the chain to form the longest loop and the probability of the same fraction of the remaining chain to form a second longest loop are different. The original chain is a classical RW, while the remaining part of the chain, after the longest loop is

erased, is not a RW (on average the remaining part of the chain is more 'stretched' than a RW having the same length). We will return to this absence of self-similarity in section 2, where we will discuss the absence of evidence for singularities in the probability densities of the nth longest loop in a RW, as opposed to the singularities in $P_n(W)$.

2 Possible Structure of Polyampholyte in a Ground State

In this chapter we explore a possible ground state of a randomly charged PA: a necklace of neutral compact blobs connected by highly charged stretched strings. In order to characterize this state, we determine the typical sizes of neutral segments in a random sequence of N charges.

2.1 A 'Typical' Necklace-Type Structure

As detailed is section 1.2, numerical studies suggest [20, 21, 22] that a PA forms a *necklace* of weakly charged globules, connected by highly charged strings. This structure is a compromise between the tendency to reduce the surface area (i.e. to form globules) due to surface tension, and the tendency to expand, in order to reduce the Coulomb interaction caused by the excess charge.

Kantor and Ertaş [25, 26, 27] attempted to quantify the qualitative necklace model, by postulating that the ground state of a PA will consist of a single globule, formed by the longest neutral segment of the PA, while the remaining part will form a tail. We follow [25], and investigate the problem of the size distribution of neutral segments in randomly charged PA's by mapping the charge sequence into a 1-d RW. The charge sequence $\omega = \{q_i\}$ $(i = 1, ..., N;\ q_i = \pm 1)$ is mapped into a sequence of positions $S_i(\omega) = \sum_{j=1}^{i} q_j$ $(S_0 = 0)$ of a random walker. (From now we will measure charges in units of the basic charge q_0, and therefore q_j will be dimensionless.) The random sequence of charges is thus equivalent to a RW, a chain segment with an excess charge Q corresponds to a RW segment with total displacement of Q steps, and a neutral segment is equivalent to a loop inside the RW. Throughout the work we will use the terminologies of randomly charged PA's and of RW's interchangeably.

When the longest neutral segment forms a globule (as assumed in [20, 21, 22]), the remaining part of the chain is very large ($\sim N$). It is natural to assume that neutral segments on that tail will further reduce the total energy by folding into globules. Eventually, the necklace will consist of many neutral globules. However, there are many ways in which the chain can be divided into neutral segments, and we are interested in a simple unique structure, which is physically reasonable. We therefore suggest a *specific* necklace-type structure, and construct the ground state for a randomly charged PA in the following

way, depicted in Fig. 3: The longest neutral segment contains L_1 monomers; it compacts into a globule of linear size proportional to $L_1^{1/3}$; in the remaining part of the chain the longest neutral segment (the 2nd longest neutral segment in the entire chain) of size L_2 also compacts into a globule, then the 3rd and so on, until the segments become very small (of only a few monomers). Generally, L_n denotes the number of monomers in the nth longest neutral segment, which compacts into a globule of radius $L_n^{1/3}$. Eventually, all the neutral segments are exhausted and we are left only with strings which carry the PA's excess charge Q, and connect the globules. The total number of monomers in the chain, is the number of monomers in all the neutral segments, plus the absolute value of the excess charge. (From now on, we will denote by Q the *absolute value* of the excess charge). This process of generating globules out of neutral segments, is similar to the loop

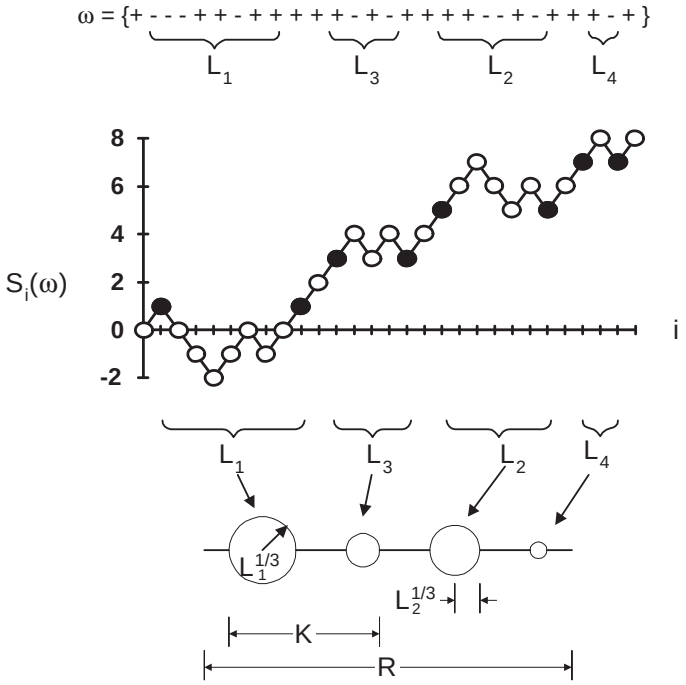

Figure 3: An example of a charge sequence ω, mapped into a 1-d RW $S_i(\omega)$, and a typical loops structure. Filled circles indicate the starting and ending points of loops. The longest loop in the RW has 8 steps ($L_1 = 8$), the 2nd longest loop has 6 steps ($L_2 = 6$), $L_3 = 4$ and $L_4 = 2$. The excess charge (which is equivalent to the total displacement of the RW) is $Q = +8$ and the total length is $N = 28$. See text for explanation of other labels.

erasing process in the LESAW model [37], where loops are erased rather than compact. The main difference between the processes is that in the LESAW model, loops in the RW are erased in chronological order, as they appear in the walk, and in our process the neutral segments are compactified according to their lengths, starting from the longest

one. Division of a charged sequence into neutral or almost neutral segments played an important role in determination of the ground state properties of a model self-interacting random polymer, represented by a directed 1-d RW [46, 47]. Both discrete and continuous charge distribution have been considered, and the resulting ground state resembles a necklace. In that model however, all possible divisions into neutral segments play a role, while we concentrate on a particular division of the charge sequence into neutral segments.

The 'ground state', generated by this process of longest neutral segments compacting into globules, does not necessarily minimize the total energy of the PA. First of all, the process does not consider the possibility of weakly charged globules, which can include much more monomers than the neutral globules, thus compensating by surface energy for the additional electrostatic energy. (For instance, two long neutral segments separated by one or two charges, would probably have lower total energy when they compact into a large single charged globule. See the segment K in Fig. 3). In addition, even when considering only neutral globules, it is not necessary that the procedure of compacting the longest neutral segment at each step generates the lowest energy state: It is possible that two 'medium sized' globules will have lower surface energy than a long one and a very short one (which remains after the first long segment was already chosen).

Using this 'longest neutral globules' picture we construct the structure of the ground state, within our model. We find the number of steps in the nth longest neutral segment, and calculate its dependence on N and n. We also find the total number of neutral segments in a chain, and we show that the total number of monomers not in any neutral segment (the number of monomers in the 'strings' connecting the globules) is exactly Q, the excess charge of the PA. We define R, the linear size of the chain, according to the picture of Fig. 3: The neutral segments in the chain compact into globules (each with a linear size of $R_{\rm segment} \sim L_{\rm segment}^{1/3}$). If all the globules are linearly packed then the total linear size is the sum of the linear sizes of all the globules ($R = \sum R_{\rm segment}$). The linear size of the chain must include the monomers not in any neutral segment (the 'strings' in the necklace, which are the absolute value of the total excess charge Q). We therefore get a means to describe the chain's size:[3]

$$R \equiv \sum_n L_n^{1/3} + Q \ . \tag{2.1}$$

[3]Here, and throughout this section, we omit prefactors of order unity.

It is evident that the generated state captures some essential features of the ground state suggested by the necklace model: The necklace type structure is compact (i.e. $R \sim N^{1/3}$) when the PA is neutral (the longest neutral segment is the entire chain) or has very small excess charge, and begins to stretch as the excess charge increases (the charged strings become longer). Finally, the PA becomes completely stretched (i.e. $R \sim N$) for the fully charged polymer.

Similarly to the definition of R, we can define the surface area S of the chain. Each neutral segment compacts into a globule of surface area $\sim L_{\rm segment}^{2/3}$, all the globules are linearly packed, and the surface area of the 'strings' is proportional to the number of monomers not in any neutral segment, which is the absolute value of the excess charge. We can therefore define:

$$S \equiv \sum_n L_n^{2/3} + Q \ . \tag{2.2}$$

The absolute value Q of the excess charge in the definition (2.2) can be viewed either as the surface area of the "necks" of the necklace (in the language of continuum drop model), since their diameter is of order unity (one monomer in diameter) or as the loss of condensation energy due to removal of the "necks" from the globules. S can have values ranging from $N^{2/3}$, for a neutral chain, to N, for a completely charged chain.

So far in this section, only the properties of a specific chain were mentioned. However, each of the mentioned properties has a distribution of values for different chains. In the following section we will analyze these distributions.

2.2 Numerical Results

2.2.1 Size Distribution of Longest Loops

We examine the statistics of the loops in a 1-d RW of N steps, using MC method for several N's up to $N = 10^4$. For each N we randomly select 10^6 sequences, and for each sequence we find the lengths L_n of all the (non-overlapping) loops. We then calculate distributions and averages of the parameters mentioned in the previous section.

We denote by $P_{N,n}(L_n)$ the probability of the nth longest loop in a RW of N steps to be of length L_n. The average length of the longest loop was found [25, 26] to be proportional to N. Expecting the same behavior for the length of the nth longest loop,

we define the probability density of the nth longest loop:

$$p_n(l_n) \equiv \frac{N}{2}\left[P_{N,n}(L_n) + P_{N,n}(L_n+1)\right] \ , \tag{2.3}$$

where $l_n \equiv L_n/N$. (Note that at least one term in Eq. (2.3) vanishes, since loops can be only of even length. Therefore, definition (2.3) includes average of probabilities for L_n and $L_n + 1$ as in the definitions used in continuum limits for discrete RW's, in order to prevent even-odd oscillations.)

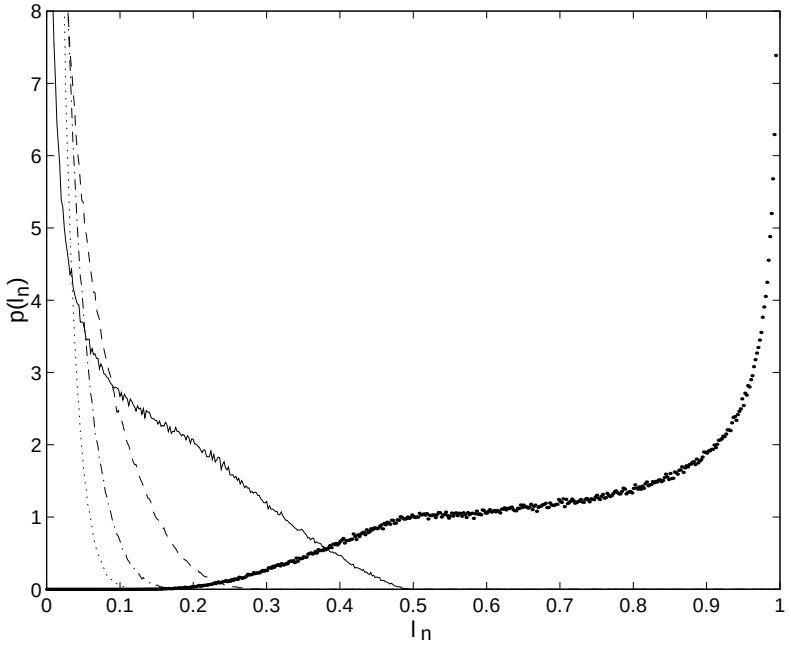

Figure 4: Probability densities of 5 longest loops from right to left: $p(l_1)$-thick points, $p(l_2)$-solid line, $p(l_3)$-dashed line, $p(l_4)$-dot-dashed line, $p(l_5)$-dotted line, as a function of $l = L/N$ from MC results of 10^6 random sequences of length $N = 1000$.

The probability densities for the five longest loops in a chain are depicted in Fig. 4 for $N = 1000$. Several properties of $p(l_n)$ [4] are evident from this figure. The probability density of the longest loop $p(l_1)$ was shown by Kantor and Ertaş [26] to have a square root divergence at $l_1 = 1$, and a discontinuous derivative at $l_1 = \frac{1}{2}$. When the length of the first loop L_1 in a specific chain is long, then the lengths of the other loops L_n (for $n > 1$) in that chain must be short, since the total length of all the loops cannot exceed N. Therefore, since in many chains the length of the longest loop is almost equal to N (as

[4] Throughout this section we denote the probability density of the nth longest loop by $p(l_n)$, in order to avoid the double indexation, i.e. $p(l_n)$ is a shorthand notation for $p_n(l_n)$. The different functions p_n will be therefore identified by their argument l_n.

indicated by the divergence of $p(l_1)$ when $l_1 \to 1$), the lengths of the other loops approach zero. This is evident from the divergence of $p(l_n)$ when $l_n \to 0$ for all $n > 1$. Because in any specific chain the length of the nth longest loop is shorter than the length of the kth longest loop, for $n > k$, the divergence of $p(l_n)$ near zero is stronger for large n. Since the probability densities $p(l_n)$ are normalized separately for each n, then any two of them must intersect (i.e. $p(l_n)$ always intersects $p(l_{n'})$ for $n \neq n'$). The length of the second longest loop never exceeds the length of the first longest loop, and the sum of their lengths never exceeds N. Therefore, the length of the second longest loop cannot exceed half the length of the chain. Consequently, $l_2 \leq \frac{1}{2}$ for all the chains, and $p(l_2)$ vanishes identically for $l_2 > \frac{1}{2}$. Similarly we can show that $l_n \leq 1/n$ for all n and $p(l_n) = 0$ for $l_n > 1/n$.

Kantor and Ertaş [26] presented numerical evidence for, but not a mathematical proof of, the $N-$independence of $p(l_1)$, when $N \to \infty$. In the same way, we expect the probability density of the nth longest loop to approach an $N-$independent limit, when $N, L_n \to \infty$, while l_n is fixed. Fig. 5 depicts the results of MC calculations of sev-

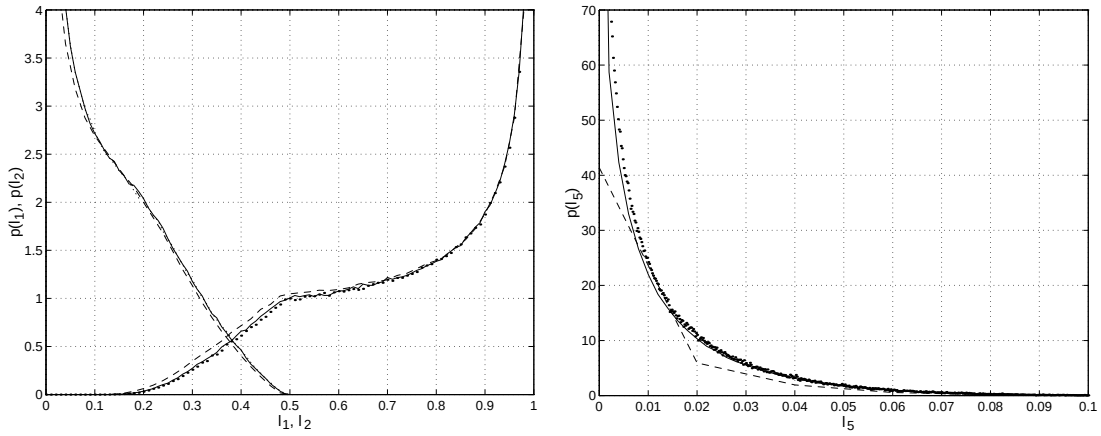

Figure 5: Probability density of the longest loop, which diverges at $l_1 \to 1$ and of the 2nd longest loop, which diverges at $l_2 \to 0$ (left) and of the 5th longest loop (right) as a function of $l_n = L_n/N$ ($n = 1, 2, 5$). MC results from 10^6 random sequences of length $N = 10000$ (dots), 1000 (solid), 100 (dashed). Note the different scales in both figures.

eral $p(l_n)$ for several values of N. We note that the behavior of the probability densities is virtually independent of N. However, we note that for larger n the probability densities are more sensitive to N for small values of N, and higher N's are needed to reach the 'continuum limit': On the right of Fig. 5 we see that $p(l_5)$ for $N = 1000$ and $N = 10000$ are similar but distinguishable. However, for $N = 100$ the probability density is different. The difference is understood, since for large n the loop length L_n is short, and continuum

limit is expected only when $L_n \gg 1$, i.e. for sufficiently large N.

We note that $p(l_2)$ is qualitatively similar to the probability densities of the length of the second segment in different models of randomly broken objects [43, 44] (see Figs. 1(b) and 3(b) in [43]). However, the probability densities of the length of the second segment in [43] have strong singularities (of the first derivative) at $l_2 = 1/4$, and are shown to have singularities at $l_2 = 1/k$ for all integer $k \geq 2$. All the probability densities $p(l_n)$ of the longest loops have a singularity when they become identically zero ($p(l_n)$ vanishes for $l_n > 1/n$ and therefore is non-analytical at $l_n = 1/n$). It is possible that due to the singularity in $p(l_{n'})$ at $1/n'$, all the other probability densities $p(l_n)$ (for $n < n'$) also have singularities at $1/n'$, since all the probability densities are dependent. Apparently, these singularities do not cause discontinuity of the first derivative, and, therefore are not visible in the numeric data. In several models of randomly broken objects [43, 44], the probability densities of the length of the nth segment have singularities at $l = 1/n'$ (for $n \leq n'$). These singularities are due to a self-similar process, which leads to a different analytical expression for the probabilities on each interval $1/(n'+1) < l < 1/n'$. Since the random process of generating the longest loops in our process is not self-similar (as indicated in section 1.4), the reason for the singularities in the models of randomly broken objects does not hold in the case of the nth longest loop (as was suggested by Frachebourg *et al.* [45]). Therefore, the probability of the nth longest loop does not necessarily have singularities at values of $l = 1/n'$.

2.2.2 Conformational Properties of the Constructed Chain

In order to verify the $N-$independence of $p(l_n)$ (and therefore the generality of the properties of $p(l_n)$ obtained for a specific N), we show that as $N \to \infty$ the average $\langle L_n \rangle$ becomes linear with N. Fig. 6 depicts the dependence of the average length of the nth longest loop $\langle L_n \rangle$ (for $n = 1 - 5$) on N. From the linear dependence with unit slope of $\log\langle L_n \rangle$ [5] on $\log N$ for $n = 1, 2$ (two upper graphs in Fig. 6), we deduce that the average lengths of the longest and second longest loops are proportional to N. The slopes of the graphs of $\log\langle L_n \rangle$ *vs.* $\log N$ increase for higher values of n, and the fit of the data to a straight line becomes worse. All the slopes (for all n) include the value 1 within their error limits (the linear fit slopes are 1.04 ± 0.06 for $n = 3$, 1.08 ± 0.11 for $n = 4$, and 1.14 ± 0.19 for $n = 5$). The slight increase in the effective slope with n is due to effects of finite N:

[5] The base of the logarithm in all the following equations and figures is 10.

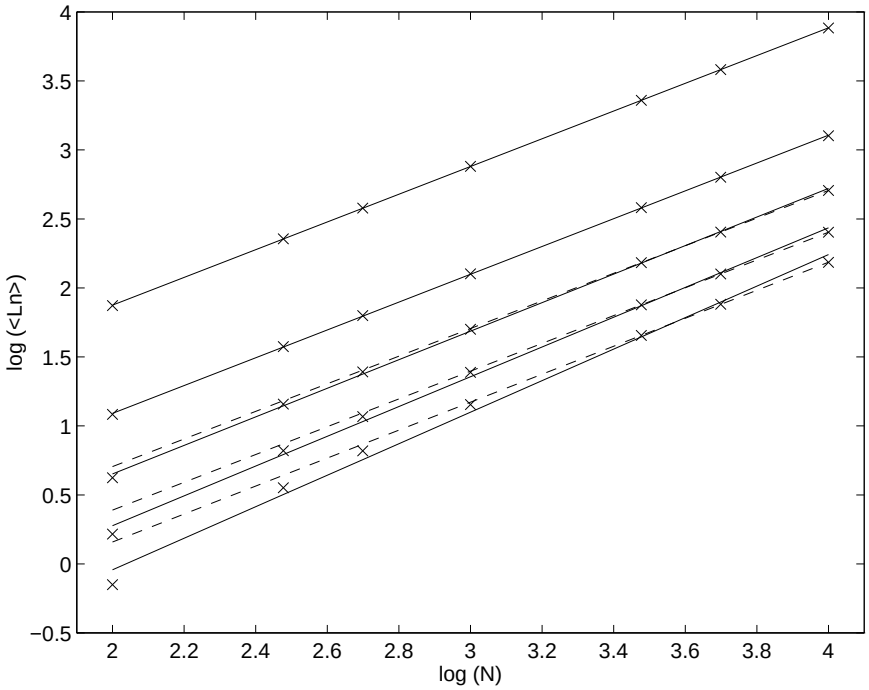

Figure 6: Logarithm of average length of 1st to 5th longest loop (from top to bottom), as a function of logarithm of chain length N. '$\times$' are averages of numeric data, solid lines are linear fits, and dashed lines are extrapolated linear fits. For the two longest loops (upper lines) the average length is proportional to N. For $n = 3, 4, 5$ the slopes of the linear fits increase with n, while the slopes of the extrapolated fits (see text) remain close to unity.

For any given dinite N, a considerable part of the sequences have no nth longest loop. (For instance, 83% of the random sequences of $N = 100$ have only 4 or less loops, and no 5th loop, while for $N = 10000$ only 11% do not have 5th longest loop). The absence of nth longest loop for small N means that the average length of the nth longest loop (for $n \gg 1$) should increase faster than N for small N. In order to overcome this systematic error, we determine the $N-$dependence of $\langle L_n \rangle$, through an extrapolation of the slopes of the linear fits of $\log\langle L_n \rangle$ *vs.* $\log N$ to $N \to \infty$. For a given n (3, 4 or 5) we calculate, through a local linear fit (i.e. linear fit for only the neighboring values), the slopes for several values of N, and estimate them at $N \to \infty$. These extrapolated slopes (dashed lines in Fig. 6) are much closer to unity than the slopes of the linear fits for all the data points: The extrapolated slopes are 1.00 ± 0.04 for $n = 3$, 1.01 ± 0.06 for $n = 4$, and 1.02 ± 0.12 for $n = 5$. We therefore expect that as $N \to \infty$, $\langle L_n \rangle$ becomes proportional to N for any finite n. Since $\langle L_n \rangle$ is proportional to N and therefore $\langle l_n \rangle$ is independent of N, we get that:

$$\langle L_n \rangle = N\langle l_n \rangle = N \int_0^1 l_n p(l_n) dl_n \ . \tag{2.4}$$

The error bars of the numeric data in Fig. 6 due to statistical errors increase with increasing N, since the number of random sequences averaged is the same for all N. The error bars are negligible for $N < 1000$, and are almost the size of the '$\times$ mark' for $N = 10^4$. The error estimates are the same for all the figures in this chapter with data plotted *vs.* $\log N$.

Since in any specific RW, $L_1 \geq L_2 \geq \cdots \geq L_n$, the average length of the nth longest loop $\langle L_n \rangle$ decreases (for fixed N) with increasing loop number n. There is no typical scale in the problem, and we may expect a power law dependence $\langle L_n \rangle \sim N n^{-\alpha}$, with $\alpha > 0$. The total number of steps in all the loops in any given RW cannot exceed N (i.e. $\sum_n L_n \leq N$), and therefore $\sum_n \langle L_n \rangle \leq N$. The convergence of the sum means that $\alpha > 1$.

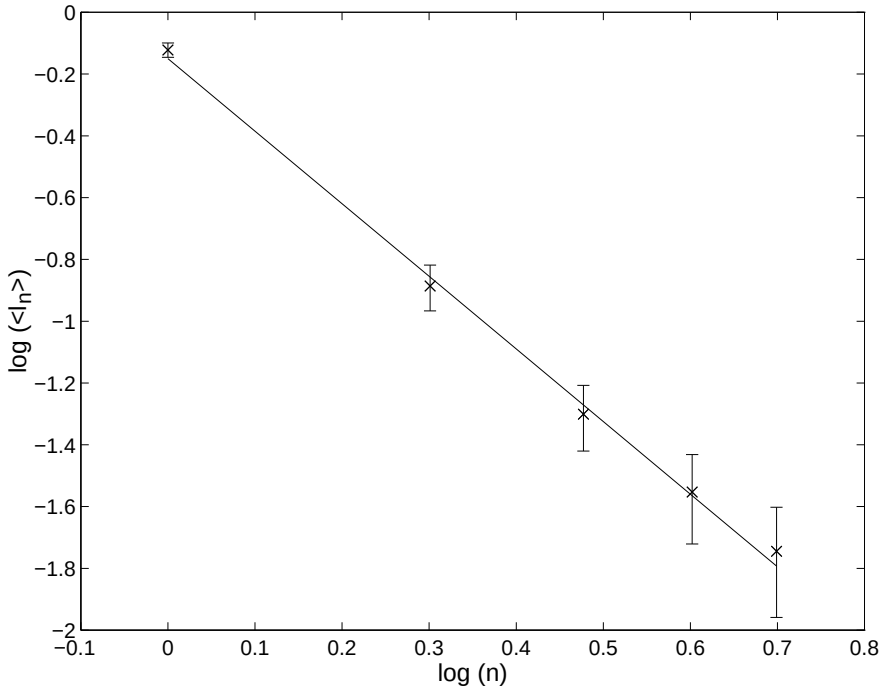

Figure 7: Logarithm of average reduced length of the nth longest loop as a function of logarithm of loop number n. The '$\times$' are averages of the numeric data, plotted with their error bars, and the line is the linear fit, having a slope of -2.3 ± 0.4.

We depict in Fig. 7 the average reduced length $\langle l_n \rangle$ *vs.* n on a logarithmic scale. To avoid systematic errors due to finite N, each value of $\langle l_n \rangle$ in the graph was determined through an extrapolation: For each n, we plotted $\langle l_n \rangle$ *vs.* $1/N$, and found the extrapolated value of $\langle l_n \rangle$ near $1/N = 0$. These values of $\langle l_n \rangle$ are depicted in Fig. 7. The linear fit to the data points has a slope of -2.3 ± 0.4. We therefore conclude that as $N \to \infty$:

$$\langle L_n \rangle \sim \frac{N}{n^\alpha} \quad , \quad \text{where} \quad \alpha = 2.3 \pm 0.4 \ . \tag{2.5}$$

At the end of this section we will argue that $\alpha = 2$, which is within the error limits of Eq. (2.5). The linear dependence of $\langle L_n \rangle$ on N and the 'no-scale' power law will enable us (in section 3) to treat the problem of the longest loops in the continuum limit.

2.2.3 Neutral Segments in Finite Chains

So far we have discussed only long chains and their properties in the $N \to \infty$ limit. There are several differences between the properties of infinitely long chains and finite size chains. First, in finite size chains the monomers that do not belong to any neutral segment constitute a finite part of the chain, as opposed to a vanishing part as $N \to \infty$. In addition, in infinitely long chains the number of neutral segments is infinite, while for finite N at some point there are no more neutral segments.

Let us consider the total number of monomers which do not belong to any neutral segment, for finite N. It is easily shown that the number of monomers left out from all the neutral segments in any specific random sequence is exactly the absolute value Q of the excess charge (defined in section 1.2) of the corresponding PA: The excess charge cannot exceed the number of monomers left out, because, by definition, the neutral segments do not have excess charge. On the other hand, all the monomers left out from the neutral segments must be part of the excess charge: If, for instance, a PA has excess positive charge, then no negative charge can be left out from all the neutral segments. (If there had been a group of negative charges, it would have joined a group of positive charges to make a neutral segment, or together with a group of positive charges, would have joined an existing neutral segment and make it longer). Since the r.m.s. excess charge of a randomly charged sequence of N charges of ± 1 is equal to $\sqrt{N}$, the r.m.s. number of monomers not in any neutral segment in the chain is also equal to $\sqrt{N}$. We note that this dependence satisfies the general claim shown by Lawler [40] for the LESAW model – the fraction of the steps remaining unerased, vanishes as $N \to \infty$ (for $d \leq 4$).

At some point the process of search for the next longest loop exhausts all the loops in the RW. We investigate this stage in the process, by analyzing the number n_f of loops in a RW. When N is infinite, the average length of the nth longest loop is given by $\langle L_n \rangle \sim N n^{-\alpha}$. Application of this equality for finite N's would predict, for large enough n's, $\langle L_n \rangle \leq 1$. Since this is not possible (the minimal length of a loop is two steps), we argue that Eq. (2.5) is valid, for finite N's, only to describe the average lengths of the longest n_f loops. There is no typical scale to the problem of the total number of

loops, and we therefore expect a power law dependence of $\langle n_f \rangle$ on N. The length of the last loop, for all chain lengths, is usually very short (consisting of only few positive and negative charges), having length independent of N, i.e. $\langle L_{n_f} \rangle \sim N^0$, which means that $\langle l_{n_f} \rangle \sim N^{-1}$. Since $\langle l_{n_f} \rangle \sim n_f^{-\alpha}$ (substituting $n = n_f$ in Eq. 2.5), we can expect $\langle n_f \rangle \sim N^y$, where

$$y = \frac{1}{\alpha} \ . \tag{2.6}$$

Substituting α from Eq. (2.5), leads to a value of $y = 0.43 \pm 0.09$.

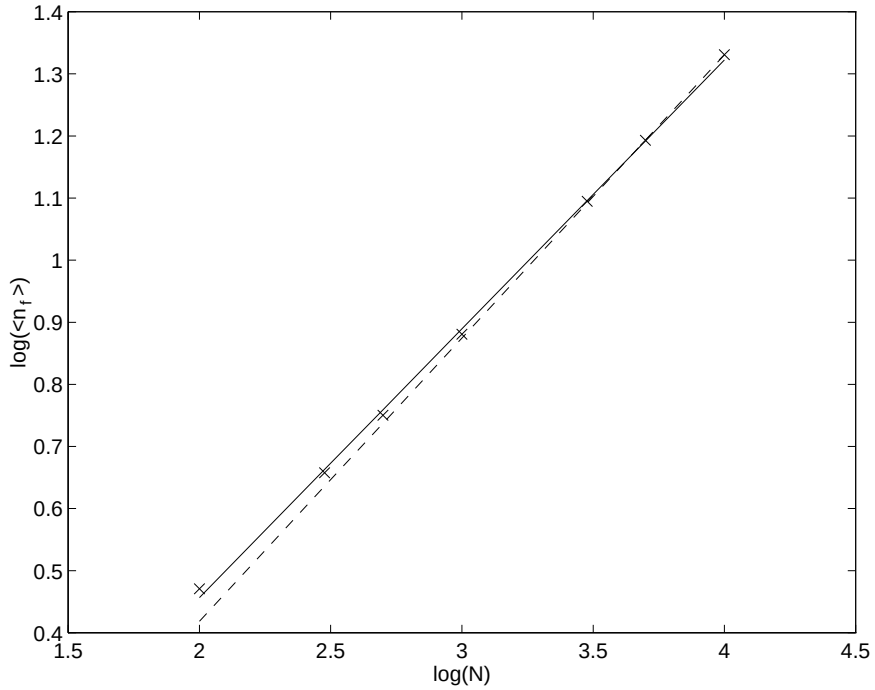

Figure 8: Logarithm of average number of loops in a chain $\langle n_f \rangle$ *vs.* $\log N$ for $N = 100 \div 10000$. '$\times$' are averages of numeric data, each of 10^6 random sequences, solid line is the linear fit for all the data, and dashed line is the extrapolated fit in the $N \to \infty$ limit. The extrapolated fit has a slope of 0.46 ± 0.06.

In Fig. 8 the average number of loops $\langle n_f \rangle$ is plotted as a function of N on a logarithmic scale. The data fit a straight line, confirming the power law dependence of $\langle n_f \rangle$ on N. The slope of the linear fit for all the data (solid line in Fig. 8) is 0.43 ± 0.02. As in Fig. 6, we estimate the power of the dependence of $\langle n_f \rangle$ on N through extrapolating the linear fit in the $N \to \infty$ limit (dashed line in Fig. 8), in order to overcome the systematic error for finite N. The slope of the fit in the $N \to \infty$ limit is 0.46 ± 0.06. We therefore get:

$$\langle n_f \rangle \sim N^y \ , \quad \text{where} \quad y = 0.46 \pm 0.06 \ . \tag{2.7}$$

This value of y is within the error limits of the value predicted by Eqs. (2.5) and (2.6).

At the end of this section we will argue that $\alpha = 2$ and $y = 0.5$, values which are within the error limits of those deduced from the numeric data. In order to confirm the $\langle n_f \rangle \sim \sqrt{N}$ relation, we show in Fig. 9 the probability density of n_f divided by $\sqrt{N}$ for several chain lengths ($N = 10^2 \div 10^4$). The division by $\sqrt{N}$ causes a reasonable collapse of the graphs for different values of N to a single function, which is (almost) $N-$independent. We see that the probability density has a maximum when $n_f/\sqrt{N}$ is close to zero (the most probable value of n_f is finite and independent of N), and it decreases to zero with increasing $n_f/\sqrt{N}$.

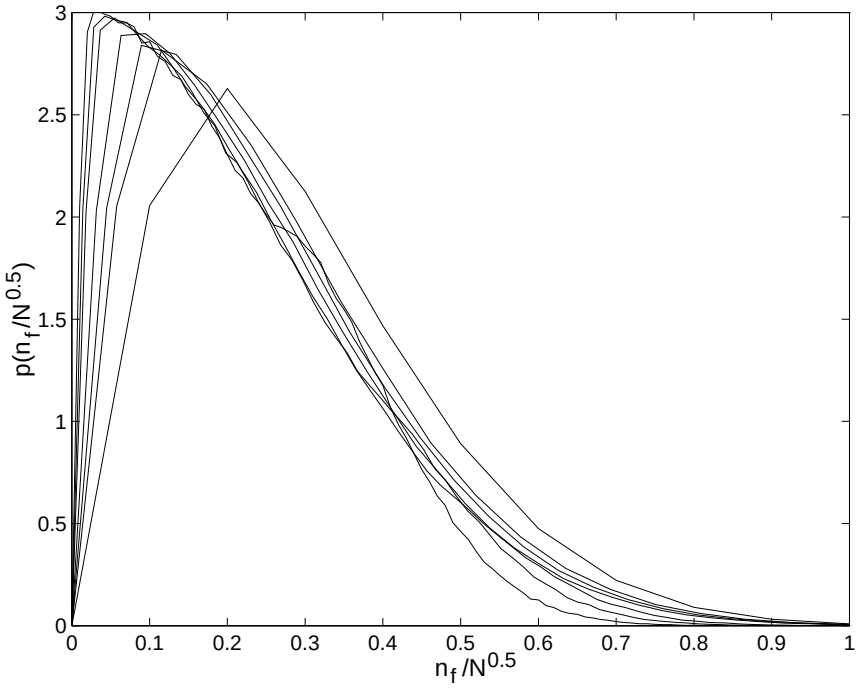

Figure 9: Probability density of n_f, the total number of loops in a chain, divided by $\sqrt{N}$, for $N = 100$, 300, 500, 1000, 3000, 5000, 10000 (from right to left).

Knowing the statistical properties of the chain, we try to construct a self-consistent complete picture, in which our numerical results fit together. We know that the total length of all the loops ($\sum_{n=1}^{n_f} L_n$) equals to the entire length of the chain minus the steps not in any loop (which are the excess charge, which are on the average $\sqrt{N}$). Dividing this equality by N, taking an average over the random sequences, we get:

$$1 - \langle \sum_{n=1}^{n_f} l_n \rangle \sim \frac{1}{N^{\frac{1}{2}}} \ . \tag{2.8}$$

On the other hand, from Eqs. (2.5) and (2.7) we know (omitting constants of order unity):

$$\left\langle \sum_{n=1}^{\langle n_f \rangle} l_n \right\rangle = \sum_{n=1}^{\langle n_f \rangle} \langle l_n \rangle \simeq \int_{n=1}^{N^y} n^{-\alpha} dn \sim 1 - N^{y(1-\alpha)} \ . \tag{2.9}$$

Comparison of the powers of N in Eqs. (2.8) and (2.9) leads to

$$y = \frac{1}{2(\alpha - 1)} \ , \tag{2.10}$$

which together with Eq. (2.6) is satisfied by:

$$\alpha = 2 \quad , \quad y = \frac{1}{2} \ . \tag{2.11}$$

These equalities are satisfied by the values of $y = 0.46 \pm 0.06$, and $\alpha = 2.3 \pm 0.4$ obtained numerically, and constitute a self consistent picture, in which the average conformational properties of the constructed ground state fit together.

2.3 Physical Properties of the Ground State

In this section we investigate some of the physical characteristics of the constructed ground state of randomly charged PA's. We focus on the linear size R and surface area S (defined in Eq. 2.1, 2.2) of the proposed ground state, trying to explain their dependence on N through the self consistent picture constructed in the previous section.

From the linear fit in Fig. 10, depicting the $N-$dependence of the average linear size of a RW of N steps, we deduce that $\langle R \rangle \sim N^{\nu}$, with $\nu = 0.50 \pm 0.01$:

$$\langle R \rangle \equiv \langle \sum_{n=1}^{n_f} L_n^{1/3} + Q \rangle \sim N^{0.50 \pm 0.01} \ . \tag{2.12}$$

Fig. 11 depicts the probability density of R divided by $\sqrt{N}$ for several values of N. From the data collapse we deduce that the $N-$dependence in Eq. (2.12) represents a scaling of the entire probability density. The $\nu \simeq \frac{1}{2}$ power in Eq. (2.12) means that the chain is not compact (although the distribution is 'peaked' near the lowest possible value of R), and is not completely stretched, but has a linear size as an ideal RW with N steps. We note that there is an oscillatory behavior near $R = R_0$ (not shown in the figure) due to discrete possible values of R for finite N. (The smallest value of R is $N^{1/3}$, the next value is $(N-2)^{1/3} + 2$, and so on, such that the values in between have zero probability).

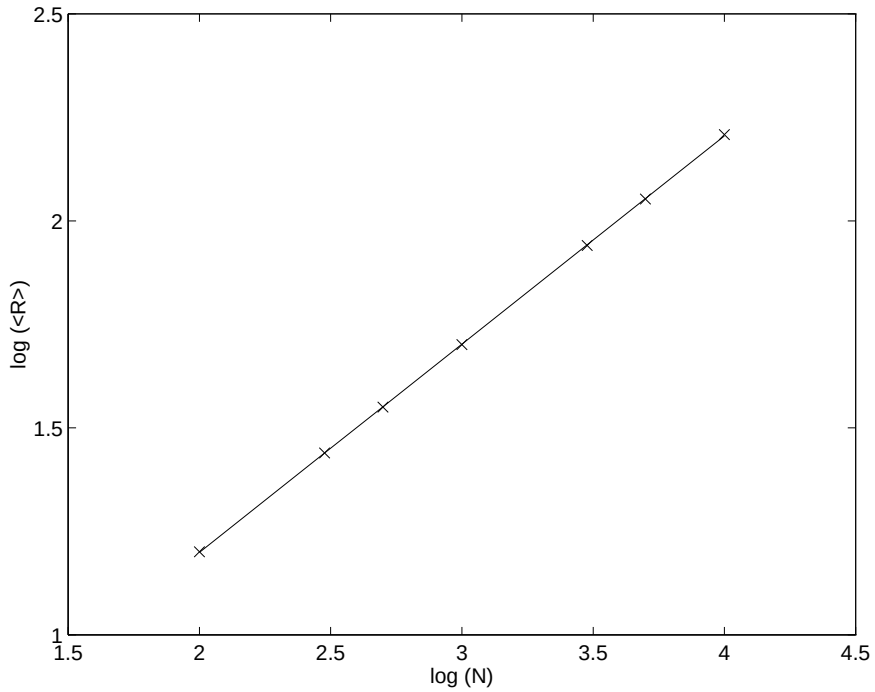

Figure 10: Logarithm of average size of chain $\langle R\rangle$ *vs.* $\log N$ for $N = 100 \div 10000$. '$\times$' are averages of numeric data, each of 10^6 random sequences, and the line is the linear fit, having slope 0.50 ± 0.01.

We can explain the dependence in Eq. (2.12), by assuming that $\langle \tilde{R}\rangle \equiv \sum_n \langle {L_n}^{1/3}\rangle$ and $\sum_n \langle L_n\rangle^{1/3}$ have the same $N-$dependence, and by using the power laws of $\langle n_f\rangle$ and $\langle L_n\rangle$ (Eqs. 2.5 and 2.7) with $\alpha = 2$ and $y = 0.5$ (Eq. 2.11):

$$\langle \tilde{R}\rangle \equiv \sum_{n=1}^{n_f} \langle {L_n}^{1/3}\rangle \sim \sum_{n=1}^{\langle n_f\rangle} \langle L_n\rangle^{1/3} \sim \int_{n=1}^{N^y} \left(\frac{N}{n^\alpha}\right)^{1/3} dn \sim N^{0.5} \ . \tag{2.13}$$

In order to confirm this dependence, Fig. 12 depicts $\langle \tilde{R}\rangle$, as a function of N (for $N = 100$ to 10000) on a logarithmic scale. The slope of the linear fit is 0.49 ± 0.02, in accordance with the prediction of Eq. (2.13). We see that the average linear chain size $\langle R\rangle$ is a sum of two terms, each proportional to $\sqrt{N}$ – the average linear size of all the loops, and the average excess charge. The $\sqrt{N}$ dependence of $\langle R\rangle$ is therefore understood.

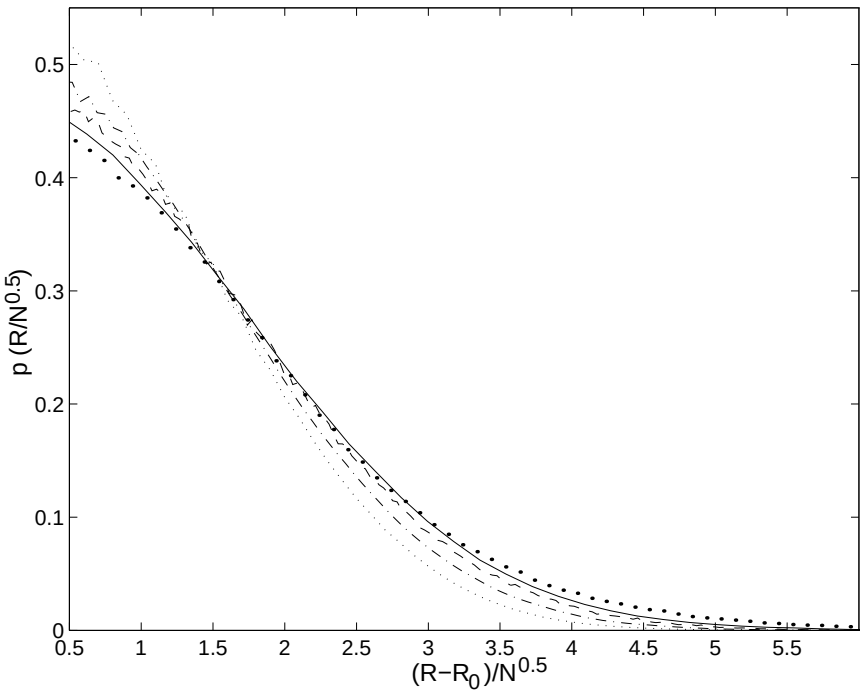

Figure 11: Probability density of R, the linear size of the entire chain, for several chain lengths: N =100 (dotted line), 300 (dot-dashed line), 1000 (dashed line), 3000 (solid line), 10000 (thick points). The minimal possible value of R ($R_0 = N^{1/3}$) is subtracted from R, and the result is divided by $\sqrt{N}$ to collapse the data.

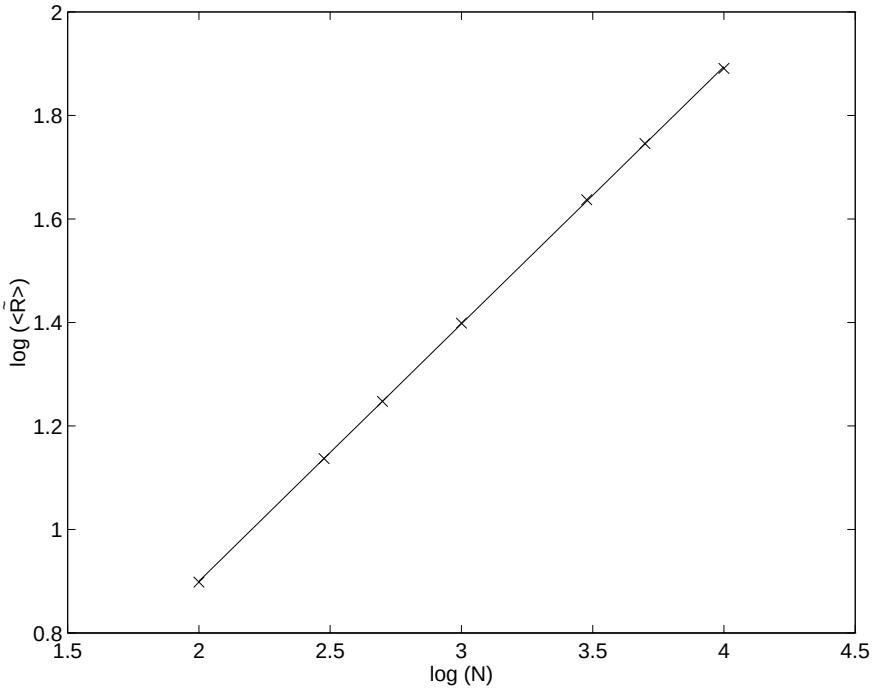

Figure 12: Logarithm of $\langle \tilde{R} \rangle$, the average linear size of all the loops *vs.* $\log N$. '$\times$' are averages of numeric data, each of 10^6 random sequences, and the line is the linear fit, having slope 0.49 ± 0.02.

When we defined S, the surface area of the chain (Eq. 2.2), we saw that it can have values ranging from $N^{2/3}$, for a neutral chain, to N, for a completely charged chain. These limits indeed correspond to the expected behavior of PA. Obviously $S \geq N^{2/3}$. We analyzed the N-dependence of S (Fig. 13), and found that:

$$\langle S \rangle \equiv \langle \sum_{n=1}^{n_f} {L_n}^{2/3} + Q \rangle \sim N^{0.67 \pm 0.01} \ . \tag{2.14}$$

When we subtract from S its minimal value, and divide the result by $N^{2/3}$, we get a distribution which is (almost) identical for all N (Fig. 14). This dependence means that the average surface energy of the generated structure has the same $N-$dependence as the surface energy of a single compact globule (or several compact globules, each containing a finite part of the chain). From the same arguments that led to Eq. (2.13), we get:

$$\langle S \rangle = \langle Q \rangle + \sum_{n=1}^{\langle n_f \rangle} \langle L_n \rangle^{2/3} \sim \sqrt{N} + \int_{n=1}^{N^y} \left(\frac{N}{n^\alpha} \right)^{2/3} dn \sim N^{2/3} \ . \tag{2.15}$$

This power of N is in accordance with Eq. (2.14), and indicates that the N-dependence of the average surface area is determined by the neutral segments (i.e. the 'beads' in the necklace), and is not affected by the excess charge (the 'strings' in the necklace).

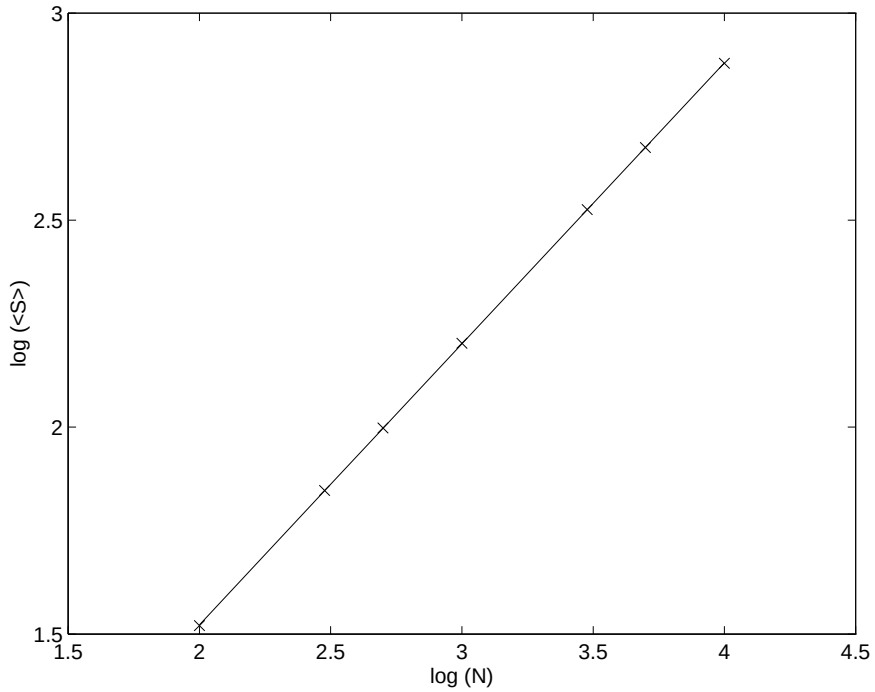

Figure 13: Logarithm of average surface area of chain $\langle S \rangle$ *vs.* $\log N$ for $N = 100 \div 10000$. '$\times$' are averages of numeric data, each of 10^6 random sequences, and the line is the linear fit, having slope 0.67 ± 0.01.

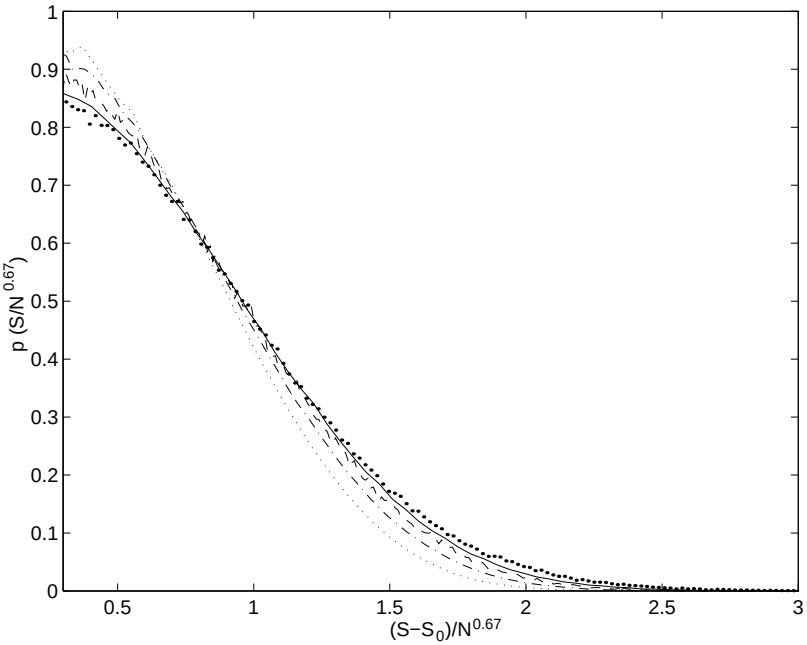

Figure 14: Probability density of S, the surface area of the entire chain, for several chain lengths: N =100 (dotted line), 300 (dot-dashed line), 1000 (dashed line), 3000 (solid line), 10000 (thick points). The minimal possible value of S ($S_0 = N^{2/3}$) is subtracted from S, and the result is divided by $N^{2/3}$ to collapse the data.

2.4 Discussion and Conclusions

We investigated the expression $S + \frac{Q^2}{R}$, which has the same N-dependence as the energy of the generated structure. (We considered the energy terms of Eq. (1.11), and omitted the condensation term, since it is the same for all structures of a given N). Fig. 15 depicts

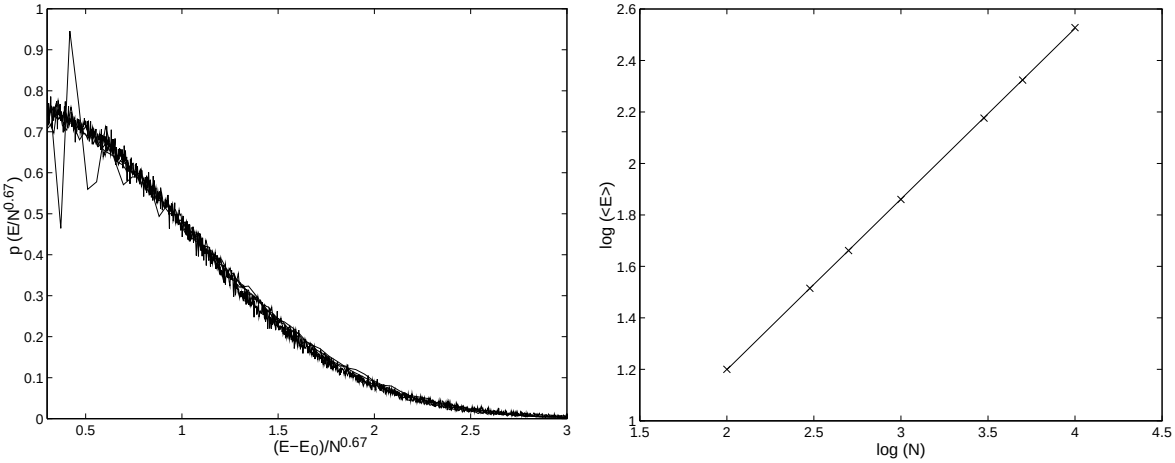

Figure 15: Total 'energy' of entire chain (omitting the condensation term). Probability density for several chain lengths ($N = 100 \div 10000$), while the minimal possible value of E (E_0) is subtracted from E, and the result is divided by $N^{2/3}$ to collapse the data (left). $\log\langle E\rangle$ *vs.* $\log N$ (right), where '$\times$' are averages of the numeric data, and the line is the linear fit, having slope 0.66 ± 0.01.

the probability density and average of the expression $S + \frac{Q^2}{R}$, denoted as E. From the value of the slope of the linear fit on the right of Fig. 15 we deduce:

$$\langle E\rangle \equiv \langle S + \frac{Q^2}{R}\rangle \sim N^{0.66\pm 0.01} \ . \tag{2.16}$$

When subtracting from E its minimal value and dividing the result by $N^{2/3}$, we get a probability density which is identical for all N. This dependence means that the total energy is very low: The surface energy term (S) gets its minimal value ($\sim N^{2/3}$), and $R \sim N^{1/2}$, thus bringing the electrostatic energy term to a point where it does not affect the N-dependence of the total energy. The total energy behaves very much like the surface energy, as if the chain is constructed of a *single compact neutral globule*.

In comparing our results with previous studies of the structure of the ground state in randomly charged PA's ([17, 18, 20, 21, 22] and discussed in section 1.2.3), we see that our suggested structure satisfies qualitative properties evident in other studies: The structure is compact when the chain is neutral or weakly charged, and it stretches as the excess charge increases. However, the result $\langle R\rangle \sim N^{0.5}$ ($\nu \simeq \frac{1}{2}$) is not in complete accordance with other findings. By an exact enumeration, Kantor and Kardar showed [22]

that the average PA (average over an unrestricted ensemble – not restricting the value of the excess charge) is not compact, i.e. $\nu > \frac{1}{3}$ (which agrees with our results). Investigating the $N-$dependence of short PA's with Coulomb interactions, Kantor and Kardar obtained a value of about $\nu \simeq \frac{1}{2}$ (although they claimed that the chains were too short to extract a meaningful value). Moreover, the qualitative behavior of our probability distribution of R is also quite similar to the one obtained in [22]: The distribution is peaked near its smallest possible value, and has a tail, which determines the asymptotic behavior of ν. However, the probability distribution of R obtained by Kantor and Kardar in [22] has a much broader (possibly power law) tail, so that as N increases, the large values of R of the stretched configurations are assumed to dominate the total average. They thus expect (though not completely prove) that the value of $\nu \simeq \frac{1}{2}$, extracted from the behavior of the short loops, underestimates the true asymptotic value. Kantor and Kardar therefore conclude that the $\langle R \rangle$ of the unrestricted ensemble increases with N at least as fast as a SAW (i.e. $\nu > 0.6$), as opposed to our $\nu \simeq \frac{1}{2}$. However, we note that if we allow for weakly charged globules (and not just neutral ones), we should expect the linear size of the chains in our model to be proportional to N^{ν} with $\nu \geq 0.5$: The excess charge will be partly included in the globules, and therefore may cause an increase in the power in which $\langle \tilde{R} \rangle$ (the average linear size of all the loops without the excess charge) depends on N. This model of weakly charged globules is worth further study, but is beyond the scope of our work.

3 Continuous Random Walks

In this section, we return to the probability density of the longest loop in a 1-d RW, which serves as a measure for the longest neutral segment in a randomly charged PA. According to the central limit theorem [34] any RW, in which each step is assigned a given distribution, approaches a Gaussian RW, when the number of steps increases. (By Gaussian RW we mean a RW, in which each step is assigned a Gaussian probability of zero average and a fixed standard deviation. Since the sum of Gaussians is a Gaussian, then the probability of a position x after t steps is a Gaussian.) Therefore, the Gaussian RW can serve as a general model for all RW's. Furthermore, every unbiased RW (i.e. the average displacement of each step is zero) can be characterized by a single parameter – the standard deviation of the Gaussian. We therefore explore the problem of the longest loop (introduced in section 1.2.3) for continuous Gaussian RW's. Such a RW is referred to as a continuous RW, since the position of a given step at a given time can take any value x, with a certain probability density (see Eq. 1.13), and is therefore a continuous variable, as opposed to the positions of steps in a 'discrete' RW (i.e. a RW with steps of fixed length). In continuous RW's the steps are still discrete, but the probability of the position is a continuous Gaussian. The problem of how to define a loop for Gaussian RW's is not trivial, and seems to determine the probability density of the longest loop.

3.1 Numerical Investigation of the Probability Density

3.1.1 Definition of the Problem

How can we define a loop for a continuous RW, where the positions are distributed in continuum, and the path never returns *exactly* to a position visited before? According to the definition of a loop in a RW with steps of fixed length, where a loop is formed when the 'random walker' returns exactly to a previously visited position, no loops are formed in a continuous RW. We therefore introduce a new parameter (denoted ϵ), and say that if two steps in a continuous RW are closer than ϵ from each other, the segment between them is called a closed loop. Therefore, the probability of having a longest loop of length L along a RW depends on N, the number of steps in the RW, on a, the standard deviation of each step in the walk, and on ϵ defined above.

We try to decrease the number of parameters in the problem. Let us consider a specific RW of given a and N, where a loop is defined by some ϵ, and scale it by some λ (i.e.

$a \to \lambda a$). Each step in the RW becomes λ times longer, and therefore the distance between any two steps is multiplied by λ. When we rescale the distance defining a loop by λ (i.e. $\epsilon \to \lambda\epsilon$) then the lengths of the loops in the rescaled walk are identical to the lengths in the original walk. This scaling argument shows that the probability of the longest loop depends on ϵ and a only through ϵ/a.

As in the discrete problem (see Eq. (2.3) and [26]), for large N's it is convenient to work with a probability density of the longest loop, and explore it as a function of the reduced length $l \equiv L/N$.[6] We denote the probability density of the longest loop by:

$$p(l; N, \epsilon/a) = N \cdot P(L; N, \epsilon/a) \ , \tag{3.1}$$

where $P(L; N, \epsilon/a)$ is the probability of having a longest loop of length L in a RW of given a and N, when the distance defining a loop is ϵ. We use MC simulations to investigate the probability density $p(l; N, \epsilon/a)$ of the longest loop, and especially its dependence on ϵ and N.

3.1.2 Probability Density in the ϵ, N Plane

For a RW with steps of fixed length, the probability density of the longest loop becomes universal (i.e. $N-$independent) for modest values of N (see Fig. 5 in section 2.2 and also [26]). However, the probability density of the longest loop for the continuous RW seems to be drastically dependent on the chosen value of ϵ. This dependence is depicted in Fig. 16, showing $p(l; N, \epsilon/a)$ for several values of ϵ/a from 10^{-5} to 10, and for a given $N = 300$ (this dependence is qualitatively similar for all values of N over two orders of magnitude, from 30 to 3000). When ϵ is extremely small (we shall see later that the requirement is $\epsilon \ll a/\sqrt{N}$) no loops are formed in all the sequences, the longest loop is thus of reduced length $l = 0$, and $p(l; N, \epsilon/a)$ becomes identically zero for all $l \neq 0$. In the other limit, when ϵ is extremely large (we shall see that the requirement is $\epsilon \gg a\sqrt{N}$), every step closes a loop which originates at almost all the other steps in the walk, and therefore the first step generates a loop with the last, resulting in $l = 1$ for all the sequences. We therefore get that by changing ϵ/a for fixed N, the probability density $p(l; N, \epsilon/a)$ changes gradually from a function similar to $\delta(l)$, when the value of ϵ/a is extremely small, to a

[6] Throughout this section we use the parameters L and l to denote length and reduced length *along the RW* (i.e. the internal coordinate).

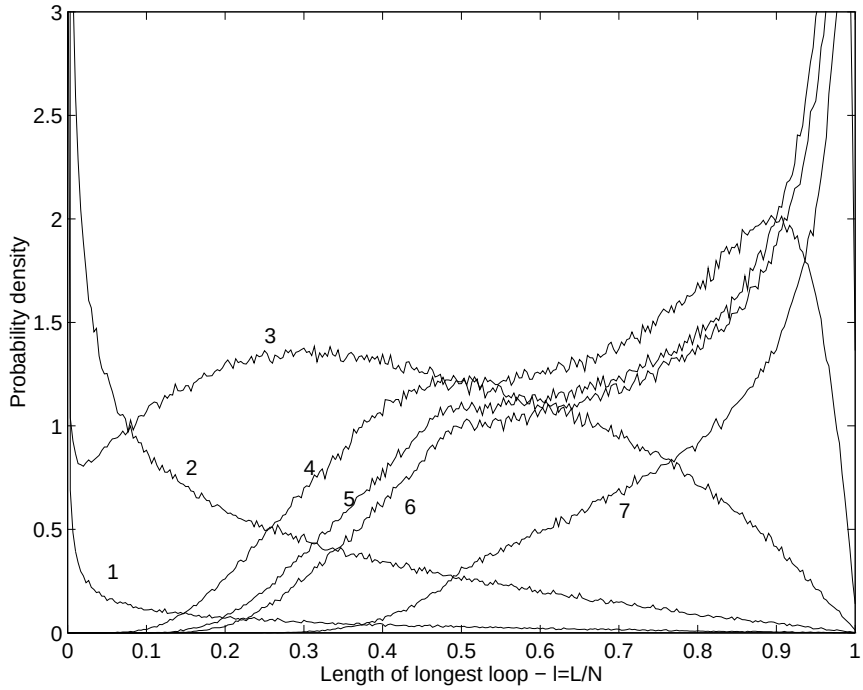

Figure 16: $p(l;N,\epsilon/a)$ *vs.* l for several values of ϵ/a: 10^{-5} (1), 10^{-4} (2), 10^{-3} (3), 10^{-2} (4), 0.1 (5), 1 (6), 10 (7). All graphs show MC results of 10^6 randomly selected sequences of length $N = 300$.

function similar to $\delta(l-1)$, when the value of ϵ/a is extremely large.[7] Somewhere along this transition from $\delta(l)$ to $\delta(l-1)$, approximately where $\epsilon \simeq a$, the probability density $p(l;N,\epsilon/a)$ becomes very similar (at least qualitatively) to $p(l)$, the probability density of the longest loop in RW's with steps of fixed length, in the large N limit. (We adopt the notation $p(l)$ rather than $p(l_1)$ of section 2, since in this section we are interested only in the probability density of the longest loop.) We note that all the numerical probability densities for continuous RW's in the following sections are based on MC results of 10^6 randomly selected sequences. Therefore, the statistical errors are of order $\sqrt{N/10^6}$ (when the probability density is of order of unity), varying from $\sim 5\cdot 10^{-3}$ to ~ 0.05 for the tested values of N from 30 to 3000.

Fixing ϵ/a to some value and investigating the $N-$dependence of $p(l;N,\epsilon/a)$, resulted in different qualitative behaviors for different values of ϵ/a. These behaviors are shown in Figs. 17, 18 and 19, for $\epsilon/a = 10^{-3}$, 10 and 1, respectively. In Fig. 17 ($\epsilon/a = 10^{-3}$) $p(l;N,\epsilon/a)$ changes gradually from $\delta(l)$ towards $p(l)$ as N increases. These functions are qualitatively similar to those in Fig. 16 (constant N and changing ϵ/a) in a way that

[7] The limit of $p(l;N,\epsilon/a) = \delta(l-1)$, i.e. $\epsilon \to \infty$, is not shown in Fig. 16, due to computational limitations: For $\epsilon/a \gg \sqrt{N}$ the total number of loops, and therefore the computation time, increases as N^2, since every step closes a loop which originates at almost all the other steps in the RW.

increasing N (for constant ϵ/a) is equivalent to increasing ϵ/a (for constant N). Similar results are obtained for all $\epsilon/a < 1$ (10^{-5} to 0.5). For all $\epsilon/a > 1$ (ϵ/a=10 in Fig. 18) $p(l; N, \epsilon/a)$ changes from $\delta(l-1)$ towards $p(l)$ as N increases, in a way that increasing N (for constant ϵ/a, as in Fig. 16) is equivalent to *decreasing* ϵ/a. At the 'transition point' ($\epsilon/a = 1$ in Fig. 19), the $N-$dependence of $p(l; N, \epsilon/a)$ is almost non-existing, and the probability density converges very quickly to $p(l)$. Even for very small values of N (N=10 – circles in Fig. 19), where the effects of discrete steps should be evident, the probability density is very close to its $N \to \infty$ limit.

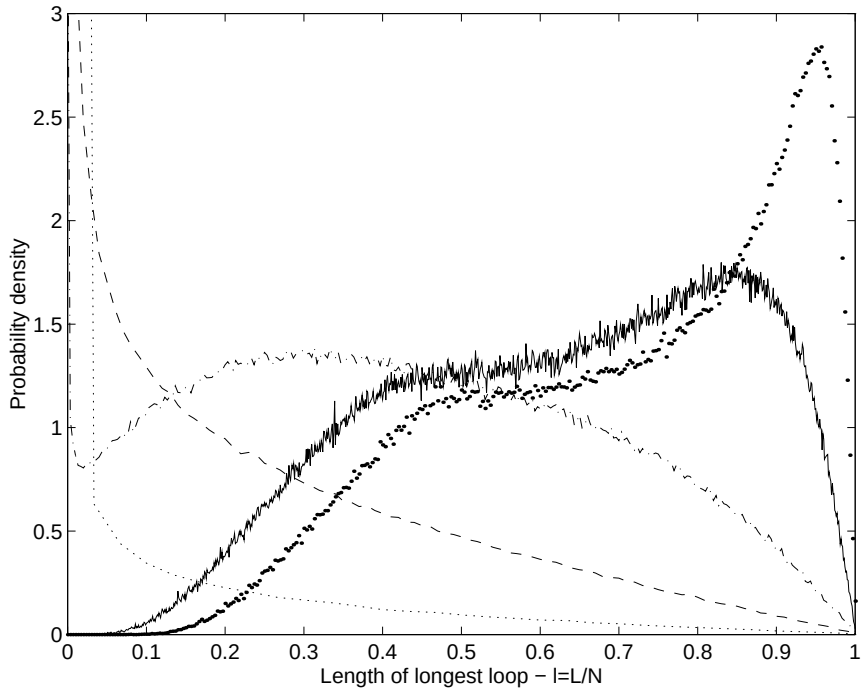

Figure 17: $p(l; N, \epsilon/a)$ *vs.* l for $\epsilon/a = 10^{-3}$ and several chain lengths: $N = 30$ (dotted line), 100 (dashed line), 300 (dot-dashed line), 1000 (solid line), 3000 (thick points). All graphs show MC results of 10^6 randomly selected sequences.

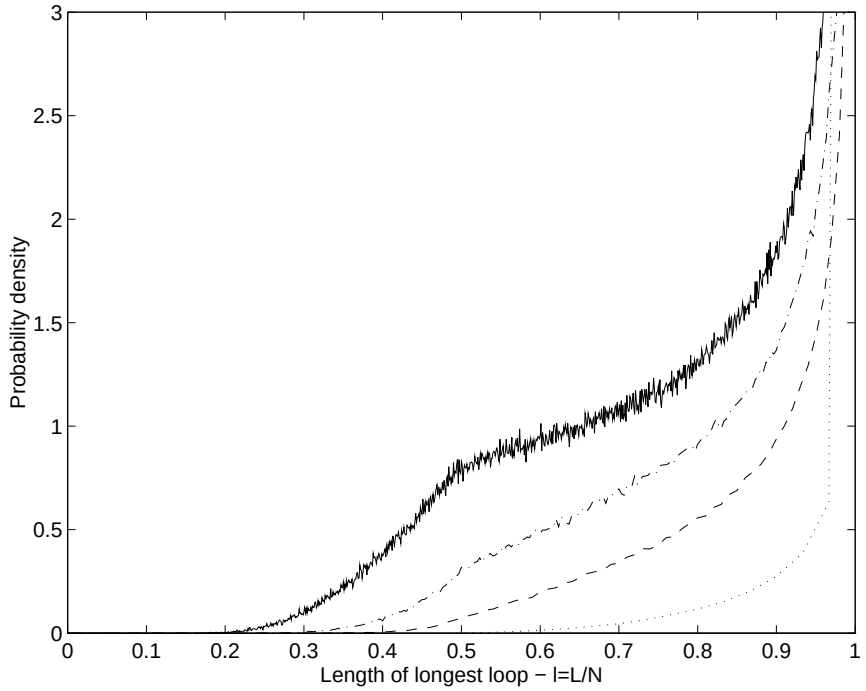

Figure 18: $p(l; N, \epsilon/a)$ *vs.* l for ϵ/a=10 and several chain lengths N (from bottom): $N = 30, 100, 300,$ 1000. All graphs show MC results of 10^6 randomly selected sequences.

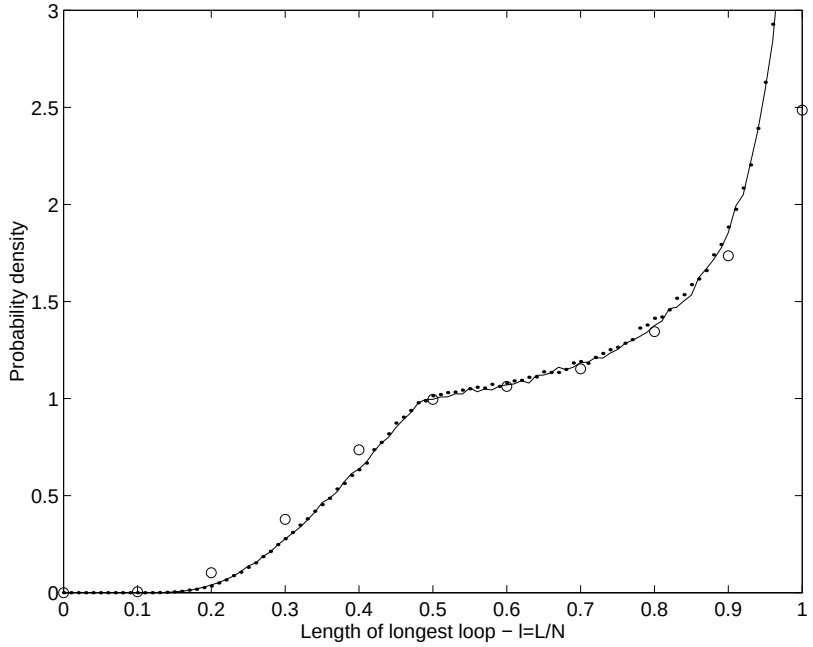

Figure 19: $p(l; N, \epsilon/a)$ *vs.* l for ϵ/a=1 and several chain lengths. $N = 1000$ (dots), 100 (solid line), 10 (circles). All graphs show MC results of 10^6 randomly selected sequences.

The numerical results can be summarized as follows:

- There is numerical evidence (although the limit was not reached for all ϵ/a) that as $N \to \infty$, for any value of ϵ/a we get:

$$p(l; N, \epsilon/a) \to p(l) \ . \tag{3.2}$$

  However, the rate of this convergence depends on the value of ϵ/a. The convergence of $p(l; N, \epsilon/a)$ to $p(l)$ with increasing N and fixed ϵ/a is faster when ϵ/a is closer to unity. This claim is evident from Fig. 20, depicting $p(l; N, \epsilon/a)$ for several values of ϵ/a (≤ 1) and N (ϵ/a=1, N=100; ϵ/a=0.1, N=1000 and ϵ/a=0.01, N=3000), compared with $p(l)$ (with N=1000). The fit of the numeric data to $p(l)$ is better for ϵ/a closer to unity (even when as the values of ϵ/a become closer to unity, we take lower values of N). For ϵ/a=1, 82% of the numeric values for different l's fit $p(l)$ within their error bars, for ϵ/a=0.1, only 65% of values fit $p(l)$ within their error bars, and for ϵ/a=0.01, 62% of the values fit $p(l)$ within their error bars (but 92% of the remaining values are above $p(l)$).

- The different functions $p(l; N, \epsilon/a)$, changing gradually from $\delta(l)$ through $p(l)$ to $\delta(l-1)$, seem to belong to one 'family' of functions. By one 'family' we mean that for any fixed value of (large enough) N we can get all the possible functions just by changing ϵ/a, and that for a given ϵ/a ($\neq 1$, $\neq 0$) we again can get, by changing N, all the possible functions (above or below $p(l)$, depending on ϵ/a). This observation may indicate that there is only one parameter (i.e. a combination of N and ϵ/a) that determines $p(l; N, \epsilon/a)$ for all values of N and $\epsilon/a < 1$, and another single parameter that determines $p(l; N, \epsilon/a)$ for $\epsilon/a > 1$.

- There is a qualitative different nature to the probability density $p(l; N, \epsilon/a)$ for values of $\epsilon/a < 1$ and for values of $\epsilon/a > 1$, and a very rapid convergence to $p(l)$ with increasing N is evident at the transition point $\epsilon \simeq a$.

In the following section we explore the group of probability densities $p(l; N, \epsilon/a)$ analytically, and give qualitative and quantitative arguments to explain the numerical observations.

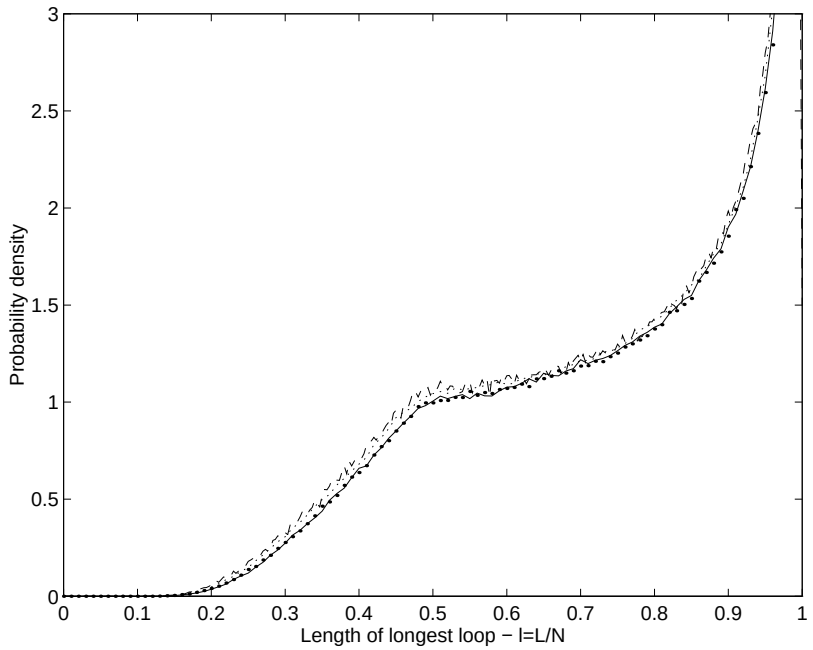

Figure 20: $p(l;N,\epsilon/a)$ *vs.* l for several values of N and ϵ/a: N=100, ϵ/a=1 (thick points); N=1000, ϵ/a=0.1 (dotted line); N=3000, ϵ/a=0.01 (dashed line), compared with $p(l)$ with N=1000 (solid line). All graphs show MC results of 10^6 randomly selected sequences. Although the fit seems rather good for all values of ϵ/a, quantitative measurements show that the fit to $p(l)$ is better as ϵ/a approaches unity (see text).

3.2 Arguments for the Behavior of $p(l;N,\epsilon/a)$

In this section we present qualitative explanations and partial analytical solutions for the behavior of the probability density of the longest loop $p(l;N,\epsilon/a)$ in its extreme ϵ/a values $\epsilon/a \ll 1$, and $\epsilon/a \gg 1$. Through this behavior we try to understand the behavior and $N-$dependence of $p(l;N,\epsilon/a)$ in the entire ϵ/a region, and especially at $\epsilon/a \simeq 1$.

3.2.1 Probability Density for Low ϵ/a

The numerical results indicate that $p(l;N,\epsilon/a)$ depends on N and ϵ/a only through a single parameter. We present qualitative arguments, which are rather intuitive, and lead to the parameter $N^{3/2}\cdot\epsilon/a$ in the $\epsilon/a \ll 1$ limit. A continuous RW of N steps and step size a is spread over an average length of $a\sqrt{N}$ (see Eq. 1.14). In order to make a loop, the position of one step in the RW should be closer than ϵ to the position of another given step. The average probability of a loop between two given steps is thus of order of this distance ϵ divided by the entire spread of the RW, i.e. $\frac{\epsilon}{a\sqrt{N}}$. Since there are $\sim N^2$ pairs of

steps, which can form a loop (each step can close a loop which originates at all the other steps), we conclude that the probability for a loop in the $\epsilon \to 0$ limit depends only on $\frac{\epsilon}{a\sqrt{N}}N^2$, denoted by B. Obtaining the probability of a loop by multiplying the probability of a loop for a single pair of steps in the number of pairs, is valid only when there is no overlap between the '$\epsilon-$coverage' (i.e. locations within distance ϵ) of any two steps (if there is an overlap, we should subtract the overlapping area from the multiplication). This no-overlap requirement means that $N \cdot \epsilon$, the total '$\epsilon-$coverage' of all the steps, should be much smaller than the length spread by the RW ($a\sqrt{N}$). The single parameter dependence of the probability density, and the no-overlap requirement can be formed mathematically:

$$p(l; N, \epsilon/a) = \tilde{p}\left(l; B \equiv N^{3/2}\frac{\epsilon}{a}\right) \quad \text{for } \frac{\epsilon}{a} \ll \frac{1}{\sqrt{N}} \ . \tag{3.3}$$

The qualitative arguments presented above are not valid when there is more than one loop in the RW, since we only discussed the probability of a loop, and not the probability of a longest loop. Numerical comparison of several probability densities, having different values of N and ϵ/a but the same value of B, confirms the dependence of Eq. (3.3). Fig. 21 depicts the probability density of the longest loop for several values of N and $\epsilon/a \ll 1/\sqrt{N}$. It is evident from the figure, that different functions $p(l; N, \epsilon/a)$ of different N and ϵ/a, having the same value of the parameter B, collapse to a single graph. The collapse was numerically evident for values of B up to 10 and for all values of N (up to 3000). Note that for B=10 and N=30 (circles in the upper graph of Fig. 21), where $\frac{\epsilon}{a} \simeq 0.06$ and $\frac{1}{\sqrt{N}} \simeq 0.18$, and the requirement of Eq. (3.3) is not fulfilled ($\epsilon/a < 1/\sqrt{N}$, but not $\ll$), the fit becomes quite poor.

In the $\epsilon/a \ll 1/\sqrt{N}$ limit it is possible to find $p(l; N, \epsilon/a)$ analytically. The inequality $\epsilon/a \ll 1/\sqrt{N}$ means that loops are very rear – in the extreme case there are no loops at all, and $p(l; N, \epsilon/a) = \delta(l)$. The next approximation is having no more than one loop per chain – in this case, *having* a loop of length L means that it is the *longest* loop. We therefore calculate the probability of having a loop of length L. A loop of length L in a RW occurs when the Mth step of the walk has a position x (this happens with probability density $p(x, M)$ as in Eq. 1.13), and the $(M + L)$th step has a position distant y from x, where $-\epsilon \leq y \leq \epsilon$. When we keep L constant, M can take any integer value from zero (i.e. making a loop between the origin ant the Lth step) to $N - L$ (i.e. making a loop between the $(N - L)$th step and the end-position). Therefore we get for the probability

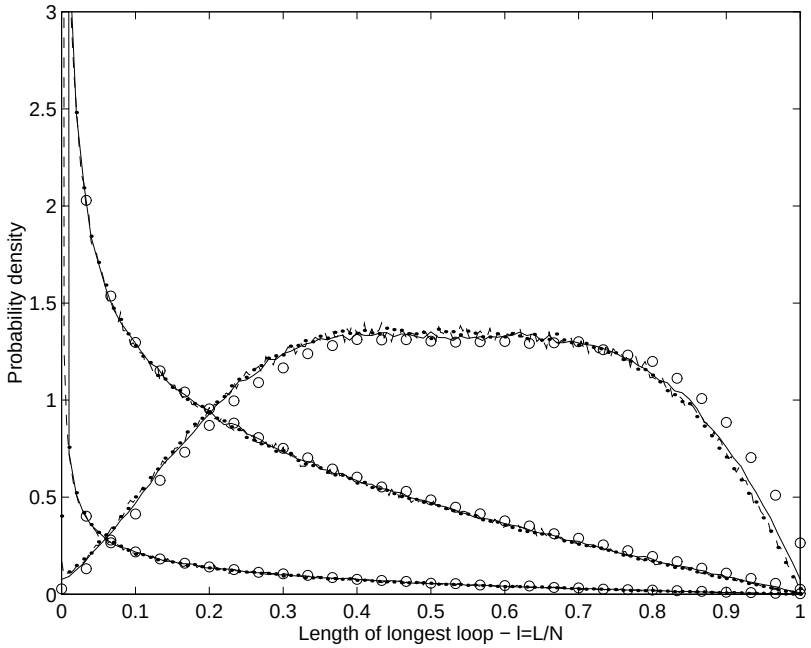

Figure 21: $p(l; N, \epsilon/a)$ *vs.* l for several values of $B = N^{3/2}\epsilon/a$ (from bottom): B= 0.1, 1, 10. For each B, several values of chain lengths N are represented (where ϵ/a is matched for the value of B): N = 1000 (dots), 300 (dashed line), 100 (solid lines), 30 (circles). Other values of B (up to B=10) and of N (up to $N = 3000$) were tested, and showed the same picture.

of having a loop of length L:

$$P(L; N, \epsilon/a) = \sum_{M=0}^{N-L} \int_{x=-\infty}^{\infty} \int_{y=-\epsilon}^{\epsilon} p(x, M) p(y, L) dy dx \ . \tag{3.4}$$

To get Eq. (3.4) we assumed that the events of all possible loops (i.e. a loop starting at the origin, a loop starting at the second step and so on) are mutually exclusive, which is in accordance with the assumption of having no more than one loop per chain. Without this assumption we could not simply sum over M, but should have subtracted the probability of having two or more loops. Integrating x out ($\int_{-\infty}^{\infty} p(x, M)dx = 1$), substituting $p(y, L)$, and performing the integral for y, while neglecting terms $\mathcal{O}\left[(\epsilon/a)^2\right]$, we get from Eq. (3.4):

$$P(L; N, \epsilon/a) = \sqrt{\frac{2}{\pi}} \frac{N - L}{\sqrt{L}} \frac{\epsilon}{a} \ , \tag{3.5}$$

which for probability density in this single-loop (s.l.) approximation and reduced length l becomes:

$$p_{\text{s.l.}}(l; N, \epsilon/a) = \sqrt{\frac{2}{\pi}} N^{3/2} \frac{\epsilon}{a} \frac{1 - l}{\sqrt{l}} \ . \tag{3.6}$$

We can take this approximation further by assuming that even if there is more than one loop in a specific RW, then all the loops are independent (i.e. having a loop of length L_1 does not change the probability of having a loop of length L_2). We denote the probability density of having a *longest* loop of length l under this independent-loops (i.l.) approximation by $p_{\text{i.l.}}(l; N, \epsilon/a)$, and get:

$$p_{\text{i.l.}}(l; N, \epsilon/a) = p_{\text{s.l.}}(l; N, \epsilon/a) \cdot \left[1 - \int_{k=l}^{1} p_{\text{i.l.}}(k; N, \epsilon/a) dk\right] \ , \tag{3.7}$$

where $p_{\text{s.l.}}(l; N, \epsilon/a)$ is the probability of having a loop of length l, and $[1 - \int p_{\text{i.l.}}]$ is the probability of not having a loop longer than l. (For dependent loops we should replace $1 - \int p_{\text{i.l.}}$ by the conditional probability of not having a loop longer than l, given a loop of length l). Solving Eq. (3.7) for $p_{\text{i.l.}}(l; N, \epsilon/a)$ leads to:

$$p_{\text{i.l.}}(l; N, \epsilon/a) = p_{\text{s.l.}}(l; N, \epsilon/a) \cdot \exp\left[-\frac{2}{\sqrt{\pi}} N^{3/2} \frac{\epsilon}{a} \left(\frac{4}{3} - 2\sqrt{l} + \frac{2}{3} l^{3/2}\right)\right] \ . \tag{3.8}$$

As indicated by the qualitative arguments (and although there is more than one loop), the dependence of $p_{\text{i.l.}}(l; N, \epsilon/a)$ on N and ϵ/a is only through the parameter $B = N^{3/2} \frac{\epsilon}{a}$. The comparison of this analytical probability density to the numerical results is depicted in Fig. 22 for values of B ranging up to $B = 10$. The analytical function agrees with the numerical results for $B < 1$ (where about 70% of the numerical values fit the analytic function within their error limits). For values of N and ϵ/a, which satisfy $B = 1$, almost two thirds of the sequences have one or more loops (the probability of having a longest loop of length $l = 0$ is 0.37). For these values, the assumption of independent loops is no longer valid, and the fit of the numeric data to the analytical probability density $p_{\text{i.l.}}$ becomes rather poor (less than 50% of the numerical values fit the analytic function within their error limits, and more than 45% of the values' error limits are above the analytical prediction).

When we look at the analytical function $p_{\text{i.l.}}(l; N, \epsilon/a)$ for larger values of B (Fig. 23), where the assumption of independent loops is no longer valid, we see that although $p_{\text{i.l.}}(l; N, \epsilon/a)$ has some of the qualitative properties of the numeric probability density (such as transition from $\delta(l)$ to $\delta(l-1)$), it has none of the singularities of the real probability density. This suggests that all the singularities and the special properties of $p(l; N, \epsilon/a)$ originate from the dependencies between the loops (these dependencies are absent from $p_{\text{i.l.}}(l; N, \epsilon/a)$ by definition).

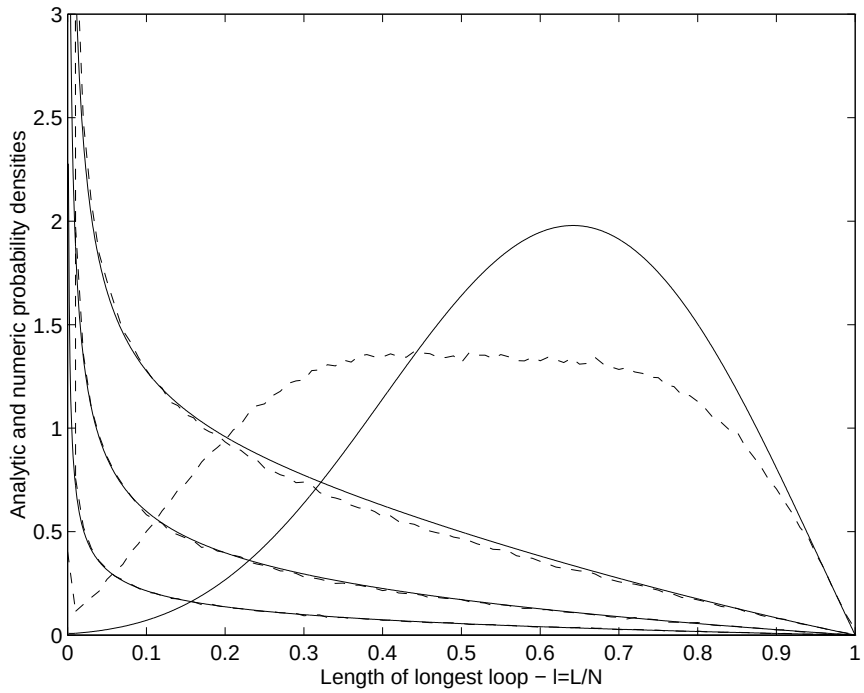

Figure 22: Comparing analytic $p_{\mathrm{i.l.}}(l; N, \epsilon/a)$ (solid lines) and numeric (dashed lines) $p(l; N, \epsilon/a)$ probability densities of longest loops *vs.* l ($N = 1000$) for several values of $B = N^{3/2}(\epsilon/a)$ (from bottom): B= 0.1, 0.3, 1, 10.

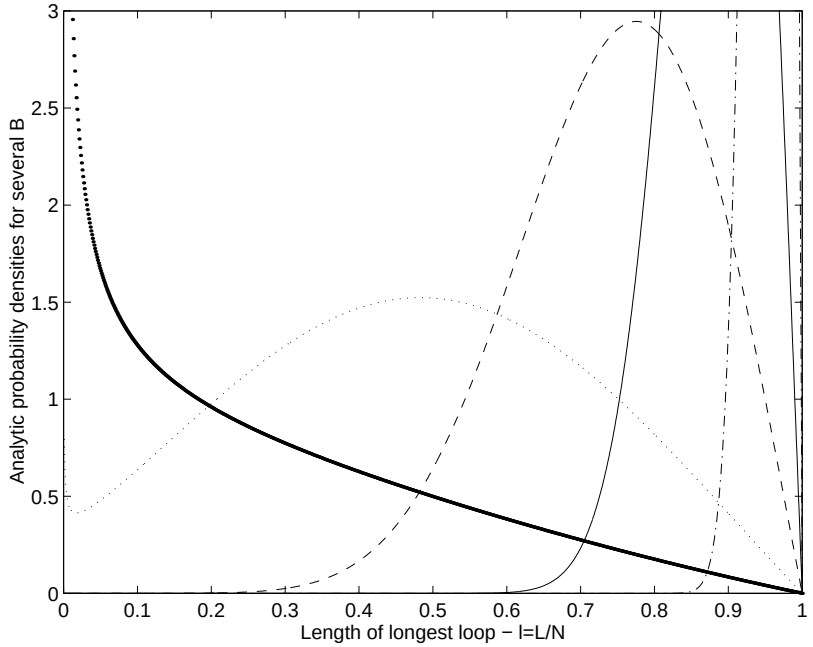

Figure 23: Analytical predictions of the probability density of longest loops $p_{\mathrm{i.l.}}(l; N, \epsilon/a)$ for several large values of B (from left: $B = 1, 5, 25, 125, 1000$). The function changes from $\delta(l)$ to $\delta(l-1)$, but without the singularities of the numeric results.

3.2.2 Probability Density for High ϵ/a

Through rescaling arguments we find a single parameter for the dependence of $p(l; N, \epsilon/a)$ on N and ϵ/a in the $\epsilon/a \gg 1$ limit. We want to rescale a chain with given N, ϵ and a to a chain with N^* 'effective steps' and $\epsilon^*/a^* = 1$. We therefore group n steps of the original RW to an 'effective step' of size $a^* = \epsilon$. In order to find n we note that a RW of n Gaussian steps, each of average size a, is spread over an average size of $a\sqrt{n}$ (defined as an effective size a^*). An 'effective step' (which includes n of the original steps) therefore spreads over an average size of ϵ, if it includes $n = (\epsilon/a)^2$ of the steps of the original chain. If we divide the original chain to $N^* = \frac{N}{n} = \frac{N}{(\epsilon/a)^2}$ 'effective steps', then each one will have an effective size of $a^* = a\sqrt{n} = a\sqrt{(\epsilon/a)^2} = \epsilon$ (where $\epsilon^* = \epsilon$). This is an exact rescaling, since the probability of having a loop of certain length (for loops longer than $(\epsilon/a)^2$ steps in the original chain) is the same in both chains. We see that any chain with N steps and $\epsilon/a > 1$ can be rescaled to a chain with $N^* = \frac{N}{(\epsilon/a)^2}$ steps and $\epsilon^*/a^* = 1$ (if ϵ/a is large enough, that a RW of $(\epsilon/a)^2$ steps becomes truly Gaussian, but lower than $\sqrt{N}$, where N^* becomes unity). We therefore get:

$$p(l; N, \epsilon/a) = \tilde{p}\left(l; C \equiv \frac{N}{(\epsilon/a)^2}\right) \quad \text{for } 1 \ll \frac{\epsilon}{a} < \sqrt{N} \ . \tag{3.9}$$

It should be noted that such rescaling arguments for $\epsilon/a < 1$ are problematic, because we do not know how many steps should be collected to make an 'effective step' (i.e. we do not know n). The number n, of steps in the original chain which make an 'effective step' in a chain with $\epsilon^*/a^* = 1$, should be obtained by the requirement that any added step (i.e. the $(n+1)$th step) will fall within distance ϵ of one of the n previous steps. This requirement means that the '$\epsilon-$coverage' (i.e. locations within distance ϵ) of all the n steps together, covers the entire size spread by the n steps, and no 'holes' (i.e. locations distant more than ϵ from all the steps) are left. This requirement for no 'holes' is fulfilled (on the average) when $\epsilon/a \geq 1$ even for a single step ($n = 1$).

Fig. 24 depicts probability densities $p(l; N, \epsilon/a)$ for several values of N and $\epsilon/a > 1$. The data collapse of probability densities having the same value of C to a single function is not so good, since in each probability density one of the limiting conditions of ϵ/a (Eq. 3.9) is not completely met: For $C = 10$ and $C = 5$ there is a probability density with a value of $\epsilon/a = 5$, which is not enough to produce a truly Gaussian RW, while for $C = 1$,

there is a value of $\epsilon/a = \sqrt{N}$.[8] Nevertheless, we deduce that the rescaling arguments are correct and the single parameter in this limit is indeed $C = \frac{N}{(\epsilon/a)^2}$.

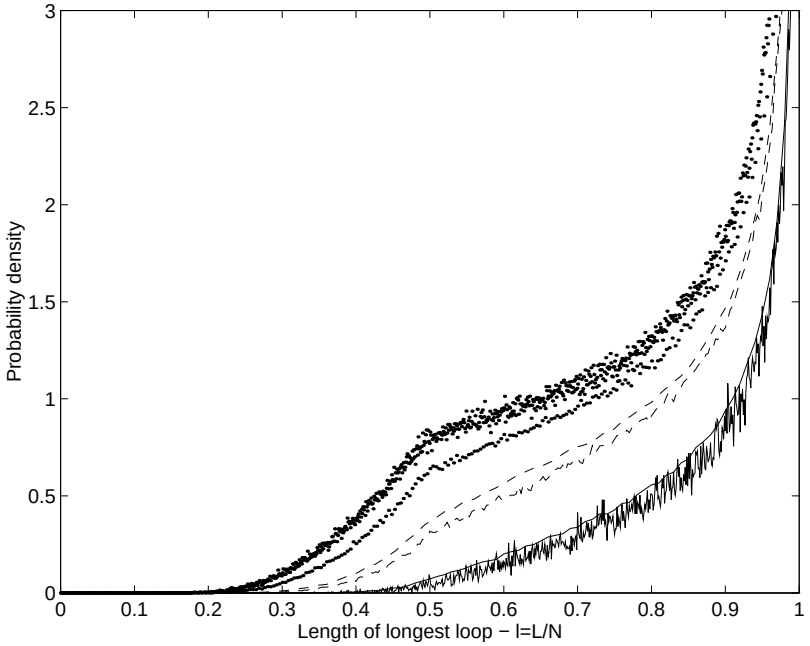

Figure 24: $p(l; N, \epsilon/a)$ *vs.* l for several values of $C = \frac{N}{(\epsilon/a)^2}$. For each C, two values of chain length N are represented, where ϵ/a is matched for the value of C. From top: $C = 10$ - dots ($N = 10^3$, ϵ/a=10; N=250, ϵ/a=5), $C = 5$ - dashed lines (N=75, ϵ/a=5 ; N=300, ϵ/a=10), $C = 1$ - solid lines (N=100, ϵ/a=10; N=900, ϵ/a=30).

3.2.3 Probability Density for $\epsilon/a \simeq 1$

The most prominent property of $p(l; N, \epsilon/a)$ for $\epsilon/a \simeq 1$ is the rapid convergence with increasing N to $p(l)$. The fact that $p(l; N, \epsilon/a)$ becomes $N-$independent means that increasing the number of steps, while keeping ϵ/a constant, does not change the statistics of the lengths of the loops. We demonstrate that for $\epsilon/a \simeq 1$, an increase in the number of steps N, for a fixed value of ϵ/a, should not change the statistics of the lengths of the loops: We compare the loops statistics of an 'original' RW of N steps of size a, and distance defining a loop $\epsilon = a$ to a scaled RW with $2N$ steps of size $a/\sqrt{2}$ and a distance defining a loop $\epsilon/\sqrt{2}$. The distribution of positions of every second step in the scaled RW is identical to that of the steps in the original RW, so we can look at the scaling as a

[8] Higher values of ϵ/a and N, required for the data collapse, were not tested due to the long computation times, proportional to N^2 for $\epsilon/a \gg 1$, where every step closes a loop which originates at almost all the other steps.

process of adding a new step in the RW between any two existing steps, while dividing ϵ, the distance defining a loop, by $\sqrt{2}$. We argue (statistically speaking) that all the loops of the original RW remain in the rescaled RW, and on the other hand, no new loops appear in the rescaled RW:

- The decrease in ϵ in the scaled RW (which causes that only about $1/\sqrt{2}$ of the original loops which close with each step remain in the scaled walk) is compensated by the increase in N (there are twice the possibilities for a loop closing with each step), and therefore, on average, all the loops in the original RW remain in the scaled RW. This argument is not valid for $\epsilon \gg a$, where the decrease in the number of loops with decreasing ϵ is much greater than can be compensated by the increase in N.

- The newly added steps are (statistically) closer than $\epsilon/\sqrt{2}$ to an existing step in the original RW, and any loop generated by the new steps already exists for one of the neighboring old steps in the original RW. If we look at each step as 'covering' range ϵ of the position axis, then for $\epsilon = a$ the entire RW path is covered, in a way that adding steps to the walk does not generate new loops. Therefore, no new loops appear in the rescaled RW. This argument is not valid for $\epsilon \ll a$, where not the entire position axis is covered by the RW.

The fact that in the rescaling process all the original loops remain and no new loops appear, means that the statistics of the loops lengths is unchanged by increasing N (when $\epsilon/a \simeq 1$), and a rapid convergence of $p(l; N, \epsilon/a \simeq 1)$ with increasing N is evident.

Another way to understand the fast convergence of $p(l; N, \epsilon/a \simeq 1)$ to $p(l)$ (faster than at any other value of ϵ/a) is to look more carefully at the parameters B and C (defined in the previous sections). For $\epsilon/a \gg 1$ we showed through rescaling arguments that the probability density depends only on $C = \frac{N}{(\epsilon/a)^2}$, which means that increasing N is equivalent to decreasing ϵ/a, i.e. making it closer to unity. For $\epsilon/a \ll 1$ however, the single parameter is $B = N^{3/2}(\epsilon/a)$, which means that increasing N is equivalent to increasing ϵ/a, again making it closer to unity. We see that making ϵ/a closer to unity is equivalent to increasing N. Therefore, when $\epsilon/a \simeq 1$ the convergence with N is the fastest.

3.2.4 Conclusions

Many of the properties mentioned in the previous sections, including the concluding remarks of section 3.1, are evident from the following figure. The curves in Fig. 25 are lines

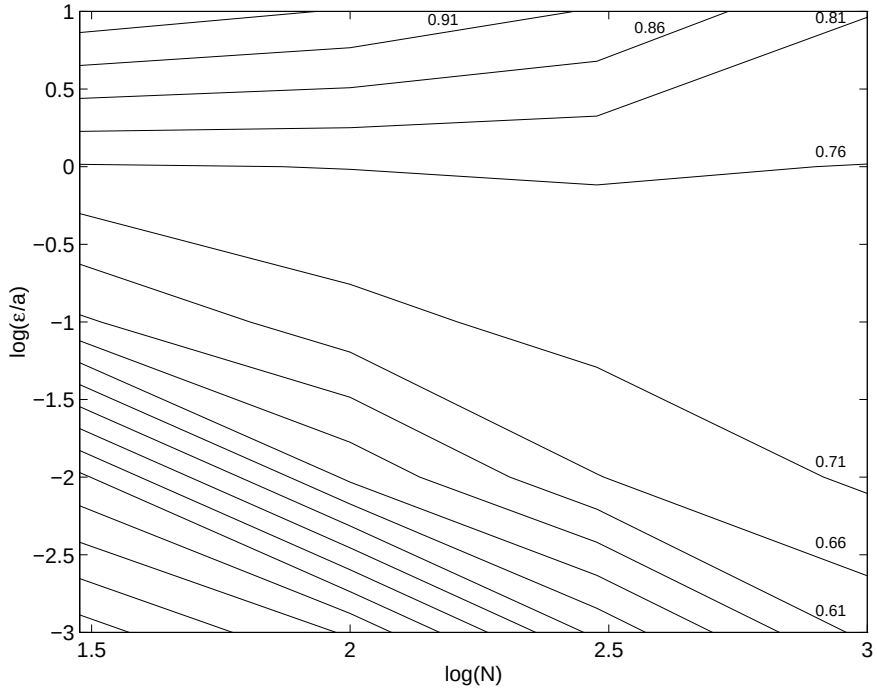

Figure 25: 'Equi-average' lines, along which $\langle l \rangle$ is constant (see text), as a function of N and of ϵ/a on a logarithmic scale ($N = 30$ to 1000 and $\epsilon/a = 0.001$ to 10). The labels above the lines are the average lengths $\langle l \rangle$ along them. The values of $\langle l \rangle$ along adjacent lines differ by 0.05.

in the '$\epsilon - N$' plane (on a logarithmic scale), along which the average reduced length of the longest loop $\langle l \rangle$, (calculated by $p(l; N, \epsilon/a)$ from 10^6 random sequences) has constant value ('Equi-average' lines). The labels above the lines are the values of these averages. The values of $\langle l \rangle$ along adjacent lines differ by 0.05. Therefore, the distance between the curves represents the rate of change of the average. Since $p(l; N, \epsilon/a)$ are normalized functions which change gradually from $\delta(l)$, where $\langle l \rangle = 0$ to $\delta(l-1)$, where $\langle l \rangle = 1$, equal averages indicate equal (or at least very similar) probability densities for all l. We review the properties evident from this figure:

- The uniqueness of $\epsilon/a \simeq 1$ emerges again. The value $\epsilon/a \simeq 1$ is the only value of ϵ/a for which the probability density is (almost) the same for all N (resulting in $\langle l \rangle \simeq 0.76$) – indicated by the (almost) horizontal line at $\log(\epsilon/a) \simeq 0$.

- The curves from both sides of the $\epsilon/a \simeq 1$ line are not identical, confirming that the dependence of $p(l; N, \epsilon/a)$ on ϵ/a is different in value (and not just in sign) between $\epsilon/a < 1$ (where $p(l; N, \epsilon/a)$ depends on $N^{3/2}\epsilon/a$) and $\epsilon/a > 1$ (where it depends on $N(\epsilon/a)^{-2}$).

- For all values of ϵ/a, when $N \to \infty$, we get $p(l; N, \epsilon/a) \to p(l)$: Moving parallel to the horizontal axis, (increasing N), the value of $\langle l \rangle$ gets to a region between the curves of $\langle l \rangle = 0.71$ and $\langle l \rangle = 0.81$ (above and below the $\log(\epsilon/a) = 0$ line). In this region the probability density is very similar to $p(l)$. In this sense the probability density $p(l; N, \epsilon/a)$ is universal: at the $N \to \infty$ limit it does not depend on the values of ϵ, a and N.

3.3 Alternative Definitions of a Loop

All the results obtained so far in this chapter were for a specific definition of what is called a loop (given at the beginning of the chapter). In order to demonstrate that the obtained results are not just an effect of this definition, and are typical for continuous Gaussian RW's, we repeat the numerical processes for several other definitions of a loop. For instance, we can define a loop between two steps in a continuous RW, if their positions are closer than a certain ρ, randomly generated (for each two steps) from a Gaussian distribution of zero average and standard deviation ϵ. The probability of having a loop between two steps distant Φ from each other along the position axis is thus:

$$P(\text{loop}) = 2\int_{\rho=\Phi}^{\infty} \frac{1}{\sqrt{2\pi}}\frac{1}{\epsilon}e^{-\frac{\rho^2}{2\epsilon^2}}d\rho = \frac{2}{\sqrt{2\pi}}\int_{y=\Phi/\epsilon}^{\infty} e^{-y^2/2}dy = P(\Phi/\epsilon)\ . \tag{3.10}$$

Repeating some of the procedures of section 3.1.2 for this definition, resulted in similar qualitative and quantitative behaviors. In Fig. 26 we compare the probability density of the longest loop for the two definitions, for several values of ϵ/a. We note that the same considerations of section 3.2 apply to this definition, and can be used to explain the behavior of the probability density of the longest loop. Results obtained for several other distributions of ϵ (so that the probability of a loop will become a Gaussian, for instance) were also very similar to those obtained in section 3.2.

The similar results for several definitions of a loop lead us to the notion that the probability density of the longest loop is universal, not just in the sense that it does not depend on the *values* of ϵ and a, but that it does not depend on the details of the *definition*

of what is called a loop. In the next section we characterize this universality, and try to prove it rigorously.

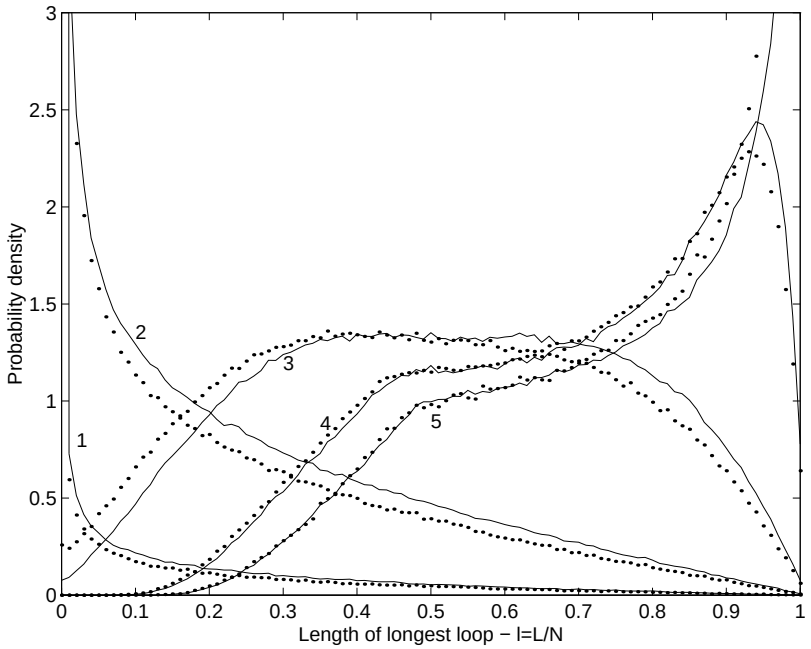

Figure 26: $p(l; N, \epsilon/a)$ *vs.* l for several values of ϵ/a. Comparing results of deterministic ϵ/a (line) with normally distributed ϵ/a (dots). The values of ϵ/a are: 10^{-4} (1), 10^{-3} (2), 10^{-2} (3), 0.1 (4), 1 (5). All graphs show MC results of 10^6 randomly selected sequences of length $N = 100$.

4 $N-$independence of the Probability Density

In this section we want to prove that the probability densities of the longest loop for both discrete and continuous RW's[9] become (for long walks) independent of N, the number of steps in the walk. Furthermore, we show that the probability densities in both cases converge with increasing N to the same 'universal' probability density. This probability density is universal in the sense that it does not depend on N or on the details of the probability of a single step of the RW. The main part of the section is devoted to prove the $N-$independence of the probability density for discrete RW's, while later we generalize the proof to continuous RW's.

4.1 Discrete Random Walks

We are interested to prove the existence and to find the $N-$independent probability density of the longest loop for discrete RW's. We therefore divide a long RW into several parts, where each such sub-RW is long enough to have a Gaussian statistics, and calculate the probability of having a loop between two sub-RW's. Based on this probability we rescale the long RW of discrete steps to a relatively short RW, where each 'effective step' is actually a sub-RW. The probability density of the longest loop calculated from the scaled RW is independent of the number of steps and of their discrete characterization in the original RW.

4.1.1 Probability of a Loop Between Random Walks

We want to calculate the probability of having a loop between two sub-RW's of a long discrete RW. By a loop between the sub-walks we mean a loop (in the long RW), which starts at a step in the first sub-walk and ends at a step in the second sub-walk. Such a loop is formed when the two sub-walks intersect (both of them reach the same position), and we therefore calculate the probability of intersection of two RW's. Fig. 27 depicts two such sub-RW's (solid lines), which are part of one long RW (dashed line). Each sub-walk is a discrete RW, having τ steps of size a. (We are interested in the limit where $\tau \to \infty$ and $a \to 0$ so that $a\sqrt{\tau}$ is finite.) The first sub-RW starts at the origin, and the position

[9] Throughout this section we use the terms discrete RW's and continuous RW's relating to a single step in the RW, according to the meaning defined in the previous section. (In a discrete RW the steps are of fixed length, and in a continuous RW each step is assigned a Gaussian probability).

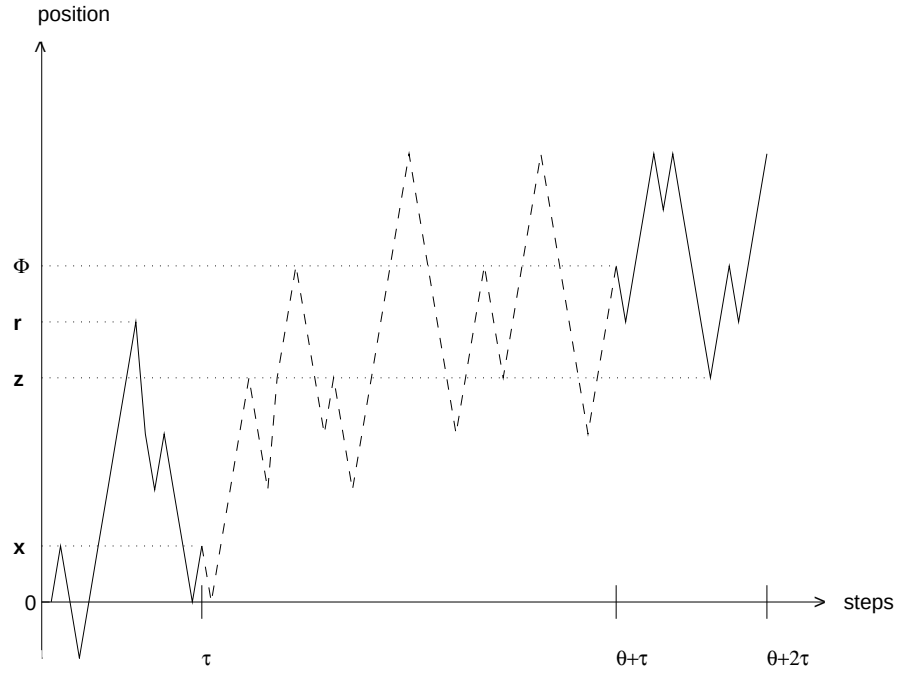

Figure 27: Two discrete RW's (solid lines), each of τ steps, which are sub-walks of a single RW (connecting dashed line). The origin of the second walk is shifted Φ along the position axis. See text for explanations of other labels.

of its last step is x. The second sub-RW begins at a position Φ relative to the origin, and Θ steps (along the original RW) after the first sub-RW ends. We denote the maximal coordinate reached by the first sub-walk by r, and the minimal coordinate reached by the second sub-walk by z. The two sub-walks intersect (forming a loop in the original RW) if and only if the maximum of one walk (the first walk in Fig. 27) is greater or equal the minimum of the other walk.

Two sub-walks, which are part of one long RW, are not independent RW's: Fixing the position Φ of the origin of the second sub-walk affects the probabilities of the possible states of the first sub-walk (and specifically the probability of the end-step position), thus making it not completely random. This 'dependence' between the sub-walks is stronger when they are closer along the original RW (i.e. Θ is small). In the extreme case, of neighboring sub-walks (i.e. $\Theta = 0$ and the second sub-walk begins where the first one ends), the end-position of the first sub-walk is fixed to be equal Φ. When the sub-walks are distant Θ (> 0) steps from each other along the original RW, the influence of fixing Φ on the possible states of the first sub-walk decreases. When the sub-walks are very far from each other along the original RW (i.e. $\Theta \gg \tau$), then fixing Φ has almost no effect on

the statistics of the first sub-walk, and its end-position x can take any value, according to a true RW statistics probability density. In this $\Theta \gg \tau$ limit we say that the sub-walks are independent RW's.

We first solve the simplest case, of the intersection probability of two independent RW's. In section 1.3, the probability density $M(r,\tau)$ of the maximal coordinate of a RW after τ steps was found to be (for $r > 0$) twice the probability density of the position of the RW after τ steps (see Eq. 1.20). For $r < 0$ the probability density $M(r,\tau)$ vanishes, since the maximum of a RW cannot be lower than its origin. By reflecting each RW about its origin (replacing $+a$ with $-a$ and vise versa), we see that for every RW having a maximum position of r, there is a reflected RW having a minimum position of $-r$. Therefore, the probability density $m(r,\tau)$ of the minimal coordinate of a RW after τ steps equals $M(-r,\tau)$. We thus get:

$$M(r,\tau) = m(-r,\tau) = \begin{cases} 2p(r,\tau) = \frac{2}{a\sqrt{2\pi\tau}} e^{-\frac{r^2}{2a^2\tau}} & ; \quad \text{for } r \geq 0 \\ 0 & ; \quad \text{for } r < 0 \end{cases} \tag{4.1}$$

where we substituted $p(r,\tau)$ from Eq. (1.13). The two independent RW's intersect if the maximum of the first walk r takes a higher value than the minimum of the second walk z (which is $z - \Phi$ relative to the origin of the second walk) for any value of z. Therefore, the intersection probability of two independent RW's, which is the probability of a loop between independent sub-walks in one long RW is:

$$P_{\text{ind}} = \int_{z=-\infty}^{\infty} \left(\int_{r=z}^{\infty} M(r,\tau) dr \right) \cdot m(z-\Phi,\tau) dz \ . \tag{4.2}$$

By definition, for $z > \Phi$, we get $m(z-\Phi,\tau) = 0$, and for $z \leq 0$ the inner integral becomes unity, since the maximum of a RW is always greater than its origin. The resulting probability (4.2) is thus:

$$P_{\text{ind}} = \int_{z=-\infty}^{0} 2p(z-\Phi,\tau) dz + \int_{z=0}^{\Phi} \left(\int_{r=z}^{\infty} 2p(r,\tau) dr \right) 2p(z-\Phi,\tau) dz \ . \tag{4.3}$$

Using the definitions (see [34]) of the *normal density function*

$$\Pi(y) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \ , \tag{4.4}$$

and of the *normal distribution function*

$$\eta(y) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{\xi^2}{2}} d\xi \ , \tag{4.5}$$

we rewrite Eq. (4.3) in the form (using the equality $\eta(-y) = 1 - \eta(y)$):

$$P_{\text{ind}}(\Delta) = 2\eta(\Delta) - 4\int_{y=-\Delta}^{0} \Pi(y)\eta(y+\Delta)dy \ , \tag{4.6}$$

where

$$\Delta \equiv \frac{\Phi}{a\sqrt{\tau}} \ . \tag{4.7}$$

This probability depends only on one parameter, which is the 'normalized' distance (along the position axis) between the origins of the walks. $P_{\text{ind}}(\Delta)$ is depicted by the solid line in Fig. 28.

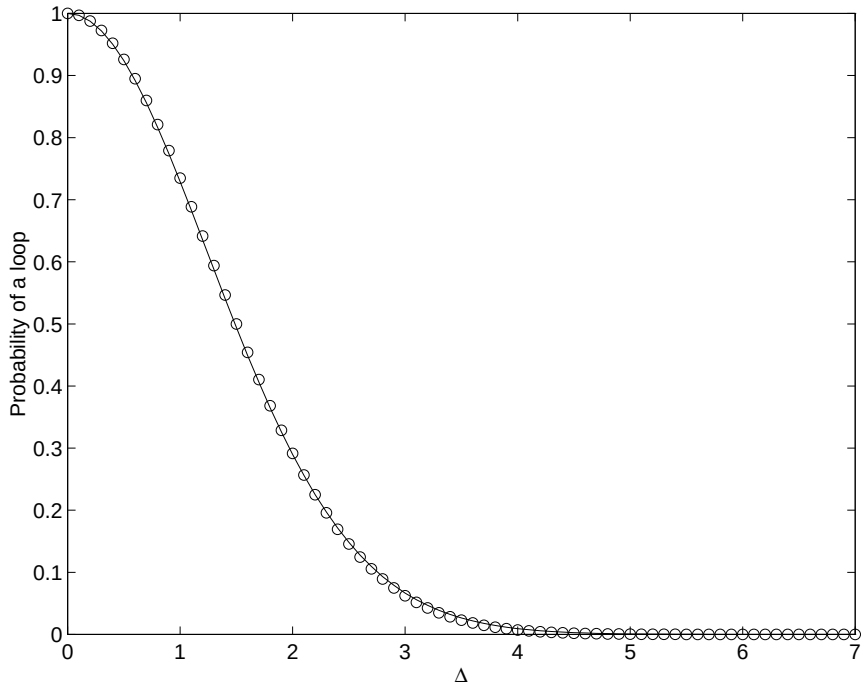

Figure 28: Probability of a loop between two independent RW's (which are sub-walks in one long RW) of τ steps of size a, distant Φ from each other along the position axis, as a function of $\Delta = \Phi/(a\sqrt{\tau})$. The analytical probability $P_{\text{ind}}(\Delta)$ (line), can be represented by a (non-normalized) Gaussian, having width 1.274 (circles).

For small Δ we get:

$$P_{\text{ind}}(\Delta) \simeq 1 - \frac{\Delta^2}{\pi} \ . \tag{4.8}$$

This result could be expected, since not having a loop when $\Delta \to 0$ means that one RW must be always below its origin, while the other always above it. Because the probability of a τ steps RW never to return to its origin is equal to the probability that it reaches its origin exactly at the τth step (see [34] and section 1.3.1), the probability of not having a

loop is proportional to $p(0,\tau)^2$, which grows like $1/(a^2\tau) \simeq \Delta^2$. For large Δ we get:

$$P_{\rm ind}(\Delta) \to \frac{4}{\sqrt{\pi}}\frac{1}{\Delta}e^{-\Delta^2/4} \ . \tag{4.9}$$

For practical purposes $P_{\rm ind}(\Delta)$ can be represented by a (non-normalized) Gaussian (circles in Fig. 28), having width of 1.274.

Let us consider an opposite extreme, where the RW's are neighboring (i.e. $\Theta = 0$). Such RW's do not intersect (apart from their meeting point) only when the first walk reaches its maximal position for the first time at its last step, and the second walk is always above its origin. The probability that a RW arrives at a position Φ for the first time at the τth step is given by $\frac{\Phi}{\tau}p(\Phi,\tau)$ (see Eq. (1.18), where $p(\Phi,\tau)$ is the probability density of the position Φ at time τ defined in Eq. (1.13)), while the probability of a RW to always be above its origin is $\frac{1}{2}p(x=0,\tau)$ (see section 1.3.1 and [34]). Thus, the probability of having no intersection (i.e. no loop) between neighboring RW's is:

$$P_{\rm no-loop} = \frac{\Phi}{\tau}p(\Phi,\tau)\frac{1}{2}p(x=0,\tau) = \frac{\Phi}{4\pi\tau^2 a^2}e^{-\frac{\Phi^2}{2a^2\tau}} \ . \tag{4.10}$$

In the limit where $\tau \to \infty$ and $a \to 0$ while $a\sqrt{\tau}$ is fixed, this probability vanishes, which means that in the continuum limit there is always an intersection between neighboring RW's. (Neighboring sub-walks in a chain always form a loop).

Let us consider the general case, of the probability of a loop between two (dependent) sub-walks, that are part of one RW. The dependence is represented by stating that the two sub-walks are separated by κ sub-walks in the original RW (i.e. $\Theta = \kappa\tau$ in Fig. 27). In order to have a loop, three independent events must occur:

(1) The maximal coordinate reached by the first sub-walk must be greater or equal to the minimal coordinate reached by the second sub-walk.

(2) There is a sub-RW of $\kappa\tau$ steps, that starts at the end position of the first sub-walk, and ends at the origin of the second sub-walk.

(3) The minimal coordinate of the second sub-walk equals to some value z.

Integrating the probabilities of these events for all possible end-positions of the first sub-walk and minimum positions of the second sub-walk, we get the probability of a loop between two sub-walks, denoted by $P_{\rm RW}(\kappa,\Delta)$. This derivation is detailed in appendix A. Fig. 29 depicts $P_{\rm RW}(\kappa,\Delta)$ as a function of Δ for several values of κ. These probabilities

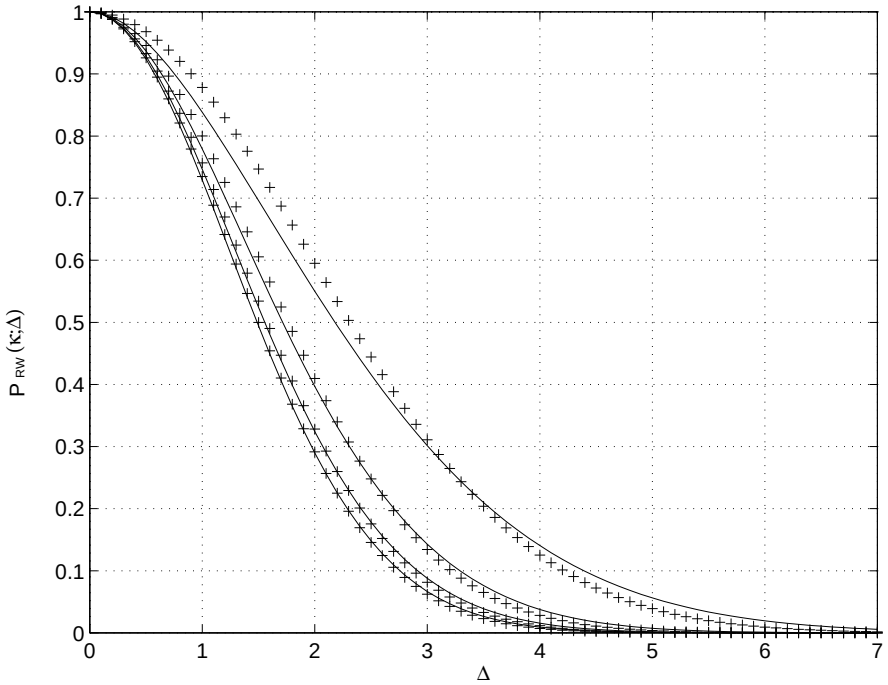

Figure 29: $P_{\rm RW}(\kappa,\Delta)$ *vs.* Δ. Different lines for different values of κ (from top): 1, 3, 10, $P_{\rm ind}(\Delta)$. The '+' are Gaussian fits (The widths of the fits are 1.96 for κ=1; 1.50 for κ=3; 1.34 for κ=10; 1.27 for $P_{\rm ind}(\Delta)$).

are qualitatively very similar to $P_{\rm ind}(\Delta)$, the independent-walks probability. We see that $P_{\rm RW}(\kappa,\Delta=0)=1$, and it decreases in a Gaussian-like manner to zero with increasing Δ. The probability of a loop between dependent sub-walks is always higher than the probability of a loop between independent sub-walks, and it increases as the walks become more dependent (i.e. smaller κ). This is to be expected, since sub-walks that are closer within the path of the original RW are also closer in their positions, and are more likely to 'intersect'.

It is evident from Fig. 29 that $P_{\rm RW}(\kappa,\Delta)$ approaches $P_{\rm ind}(\Delta)$ with increasing κ. This is reasonable because two remote sub-walks ($\Theta \gg \tau$ or equivalently $\kappa \gg 1$) become independent RW's. To first order in $1/\kappa$ we show in appendix A that:

$$P_{\rm RW}(\kappa,\Delta) = P_{\rm ind}(\Delta) + \frac{1}{\kappa}\cdot f(\Delta)\ , \tag{4.11}$$

where f is some function. Since the difference between $P_{\rm RW}(\kappa,\Delta)$ and $P_{\rm ind}(\Delta)$ decays only as $1/\kappa$, we *cannot* assume that for most values of κ, the function $P_{\rm RW}(\kappa,\Delta)$ is approximated by $P_{\rm ind}(\Delta)$.

The probability of a loop between two sub-walks on a RW is defined by $P_{\rm RW}(\kappa,\Delta)$, according to the distance between the sub-walks along the original RW (κ) and the nor-

malized distance between the origins along the position axis (Δ). The definition of a loop according to a probability (rather than a deterministic definition) is similar to the definition in section 3.3, where a loop between two steps in a RW was formed if the distance between them Φ was smaller than a randomly generated number (from a Gaussian distribution). The definition of a loop according to $P_{\mathrm{RW}}(\kappa, \Delta)$, and the fact that each 'effective step' (or sub-walk) is actually a RW by itself, set the ground for a rescaling process of the problem, which leads to a proof of the universality of the probability density of the longest loop.

We rescale a long RW of discrete steps into a relatively short RW of Gaussian steps in the following way: We divide the original RW into equal sub-walks, which are the 'effective steps' in the new scaled RW. Each such effective step includes τ steps of the original walk, and is by itself a RW already in the continuum limit (with step size $a \to 0$ and $\tau \to \infty$). Between any such two effective steps the probability of a loop is known (defined by $P_{\mathrm{RW}}(\kappa, \Delta)$), and the probability of having a loop of a given length can be calculated. This calculation results in probability densities which are independent of the number of steps N in the original RW and independent of the original step size a. However, this calculation, apart from being very cumbersome, does not enable us to find the probability density of the *longest* loop, since it disregards the dependencies between the probabilities of loops. (As evident in section 3.2, the probability of having a loop of length l and the probability of not having a loop longer than l are dependent. These dependencies are the source of the singularities in $p(l)$.) $P_{\mathrm{RW}}(\kappa, \Delta)$ is the probability of having a loop between two effective steps, when nothing is known about the walk. However, the fact that there is a loop (or that there are no loops) between two other steps in the walk, changes this probability. In the following section we alter the scaling process a little, in order to obtain the probability density of the longest loop.

4.1.2 Rescaling The Problem

In the previous section a loop between two sub-walks, which were the 'effective steps' of one long RW, was defined if the sub-walks intersected each other, i.e. if the maximal coordinate of one sub-walk was higher than the minimal coordinate of the other. Following this idea, we perform a rescaling process: We divide a given long RW with steps of fixed length into m equal sub-walks, which are the 'effective steps' of the rescaled walk. Each of the m segments, still long enough to be an ideal Gaussian RW (the steps in each segment

are of fixed length, but the probability of the positions is Gaussian), is assigned a minimum value and a maximum value. These values are assigned to each segment according to the mutual distribution of a minimum and a maximum of a RW, given its origin and end position (the distribution is independent of the positions of steps 'inside' the sub-walk and on distributions of other segments). These minima and maxima therefore depend only on the positions of the first and last steps of each sub-walk, and not on other details of the original RW. A loop between such two segments is defined if they 'intersect' each other (i.e. if either the assigned minimum or maximum of one segment has a value lower than the other segment's assigned maximum and higher than its assigned minimum). This rescaling process is statistically exact: The probability of having a loop between any two segments, when a loop is defined according to the randomly assigned minima and maxima, is exactly the probability of having a loop (in the original RW), which starts with a step in the first segment, and ends with a step in the second segment. When all the loops in a specific scaled RW of m segments are obtained, the longest loop can be found. The probability density of the longest loop is obtained by finding the longest loop for many independent RW's of m 'effective steps'.

We note that the rescaling process does not result in an analytical expression for the probability density of the longest loop, since the independent RW's of m 'effective steps', required in order to get a probability density, are generated numerically by MC simulation. However, since the rescaling process is statistically exact, the generated probability density of the longest loop is equal to the probability density of the longest loop obtained numerically for discrete RW's in the $N \to \infty$ limit [25, 26]. The equality of the probability density for a RW of m 'effective steps' to the $N \to \infty$ probability density of a 'discrete' RW is exact up to the number of effective steps into which the initial RW was divided: For instance, a loop between the 2nd segment and the 7th segment in a 20-segments RW includes loops of lengths ranging from 4/20 to 6/20 of the chain's length, depending on the exact location of the steps which generated the loop in these segments. (If the steps generating the loop in the rescaled chain were located at the beginning of the 2nd segment and at the end of the 7th segment in the original chain, then the loop length in the original chain is almost 6/20 of the chain's length. Similarly a loop of length just over 4/20 of the original chain's length can be generated between those segments.) In general, when there is a loop between two given segments it is impossible to know (in the scaled chain) where in the segments were the 'original' steps which generated the loop. Therefore, for

each loop in the rescaled chain there is an inherent inaccuracy of up to (plus or minus) one segment.

The essence of the described scaling process is the assigning of a mutually generated minimum and maximum to each sub-walk. We therefore explain in the remaining of this section how to determine the minimum and maximum for each segment. The mutual probability density of minimum and maximum of a RW, is closely related to the problem of diffusion between two absorbing walls. When a diffusive particle does not hit the walls, then its corresponding path is a RW, with a minimum greater than the position of one wall and maximum lower than the position of the second wall. We therefore solve the diffusion equation in the presence of absorbing walls, and relate the solution to the mutual probability density of minimum and maximum of a RW. We solve the diffusion equation in one dimension [32]:

$$\frac{a^2}{2}\frac{\partial^2 p(x,t)}{\partial x^2} = \frac{\partial p(x,t)}{\partial t} \ , \tag{4.12}$$

for the probability density $p(x,t)$ of a particle starting at the origin and taking steps of size a to be at a position x after t steps. The solutions are of the form [48]:

$$p(x,t) = \sum_k A_k \sin(kx + \varphi_k) e^{-\frac{1}{2}k^2 a^2 t} \ . \tag{4.13}$$

The boundary conditions of absorbing walls at $-w$ and at W are:

$$p(x = -w, t) = p(x = W, t) = 0 \ . \tag{4.14}$$

These conditions lead to the possible values of k and φ:

$$k_n x + \varphi_n = \frac{n\pi(w + x)}{w + W} \ , \tag{4.15}$$

where n is any integer number. A_k of Eq. (4.13) is obtained through the initial condition $p(x, t = 0) = \delta(x)$. The expansion of $\delta(x)$ in a Fourier sine series in the interval $(-w, W)$ is given by [48]:

$$\delta(x) = \frac{2}{w + W} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi w}{w + W}\right) \sin\left(\frac{n\pi(w + x)}{w + W}\right) \ . \tag{4.16}$$

Comparing Eq. (4.16) to Eq. (4.13), after substituting there $t = 0$ and k_n, φ_n from Eq. (4.15), leads to an expression for A_k. The resulting probability density (with respect to x) $p_{w,W}(x,t)$ of a particle to be at a position x after t steps, when there are absorbing

walls at $-w$ and W (which is the mutual probability density of $M(W,t)$ and $m(-w,t)$ of Eq. 4.1) is given by:

$$p_{w,W}(x,t) = \frac{2}{w+W}\sum_{n=1}^{\infty}\sin\left(\frac{n\pi w}{w+W}\right)\sin\left(\frac{n\pi(w+x)}{w+W}\right)e^{-\frac{1}{2}\left(\frac{n\pi}{w+W}\right)^2 a^2 t}\ . \tag{4.17}$$

We are interested in the conditional probability of a *given* RW not to hit absorbing walls at $-w$ and W (i.e. the probability of a given RW, of t steps of size a starting at the origin and ending at x, to have a maximum lower than W, and a minimum higher than $-w$). This conditional probability is equal to the derived probability density $p_{w,W}(x,t)$, divided by the probability density $p(x,t)$ of the RW to be at a position x after t steps. When measuring all distances (i.e. w, W and x) in units of $a\sqrt{t}$, we get for the conditional probability:

$$G(w,W;x) = \frac{2\sqrt{2\pi}}{w+W}e^{\frac{x^2}{2}}\sum_{n=1}^{\infty}\sin\left(\frac{n\pi w}{w+W}\right)\sin\left(\frac{n\pi(w+x)}{w+W}\right)e^{-\frac{1}{2}\left(\frac{n\pi}{w+W}\right)^2}\ . \tag{4.18}$$

This function is the conditional probability of a given RW to have a minimum higher than $-w$ and a maximum lower than W.

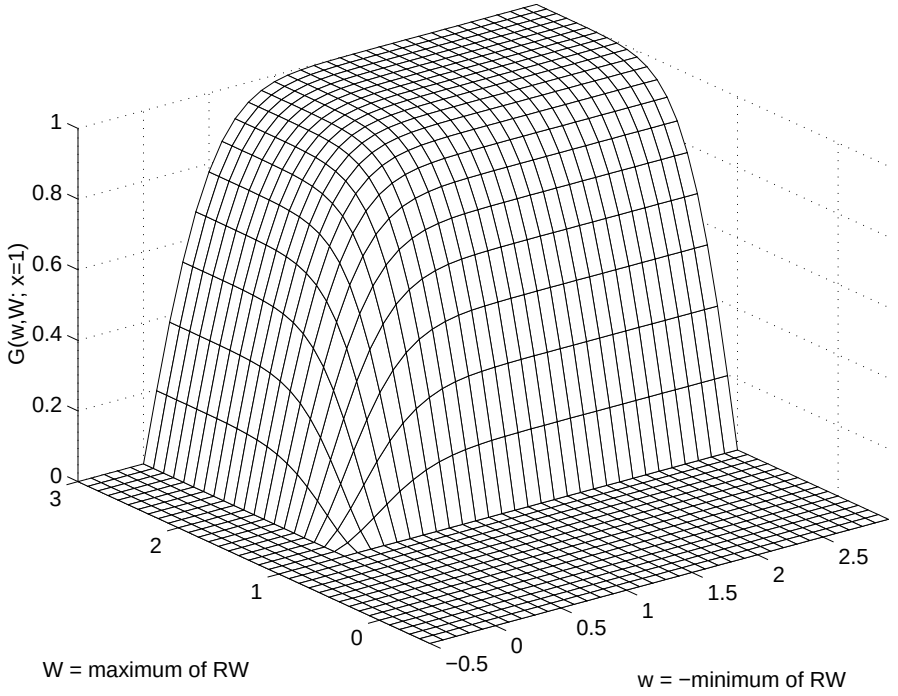

Figure 30: $G(w,W;x=1)$ *vs.* w and W. All lengths are measured in units of $a\sqrt{t}$. It is clear that for $W < \max(x,0)$ and for $w < \max(-x,0)$ the function vanishes. For $w, W \to \infty$, we get $G \to 1$.

We approximate the function $G(w, W; x)$ in order to facilitate the computational process of obtaining its values (which otherwise requires infinite summation): When $W \gg \max(x, 1)$ the probability $G(w, W; x)$ becomes independent of W, since the probability of a RW from 0 to x not to hit a very far absorbing wall, is unity (i.e. $G(w, W_1 \gg \max(x, 1); x) \simeq G(w, W_2 \gg \max(x, 1); x)$). Similarly, $G(w, W; x)$ becomes independent of w, when $w \gg \max(-x, 1)$. Therefore, it is sufficient to find $G(w, W; x)$ for values of w and W where the value of $(w + W)$ is less or equal to several times the value of $\max(x, 1)$. In these cases, due to the $-n^2$ numerator in the fraction in the exponent, only the first few terms of the infinite series of Eq. (4.18) are important. Fig. 30 depicts $G(w, W; x)$ as a function of w and W, for a specific value of $x = 1$. It is clear that for $W < x$ or for $W < 0$ (and in the same manner for $w < -x$ or $w < 0$) the probability is zero, since the origin or end-position of the RW are beyond one of the walls. It is also evident that $G \to 1$ when $W, w \to \infty$, since when the walls are 'far enough', the RW never hits them.

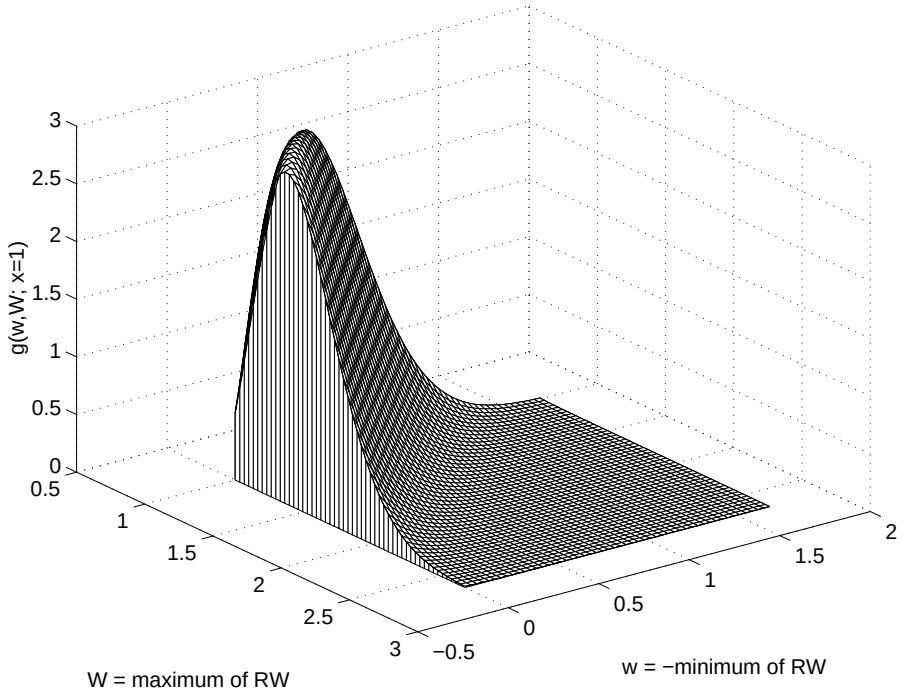

Figure 31: Mutual probability density of having a RW with a minimum $-w$ and a maximum W: $g(w, W; x = 1)$ *vs.* w and W. All lengths are measured in units of $a\sqrt{t}$. The function is defined only for $W > \max(x, 0)$ and for $-w < \min(x, 0)$. For $w, W \to \infty$ we get $g \to 0$. (Note the different view angle than Fig. 30).

The derivative of G in respect to w and W:

$$g(w,W;x) = \frac{\partial^2 G(w,W;x)}{\partial w \partial W} \ , \tag{4.19}$$

is the joint probability density (in respect to w and W) of a given RW, ending at a position x, to have a minimum $-w$ and a maximum W. Fig. 31 depicts this probability density. According to this probability density we randomly generate the minimum and maximum of a RW, ending at a position x (again, all distances are measured relative to the origin and are in $a\sqrt{t}$ units). According to the behavior of $G(w,W;x)$ we can see that the function $g(w,W;x)$ is defined only for $W > \max(x,0)$ and for $-w < \min(x,0)$, where the origin (0) and the end position (x) of the RW are between the minimum and the maximum. For $w, W \to \infty$, we get $g \to 0$, since the walls are far, and the probability to hit them vanishes. The process of randomly generating w and W from $g(w,W;x)$ is detailed in appendix B.

The rescaling process described above constitutes a proof for the $N-$independence (for large N's) of the probability density of the longest loop, for discrete RW's: We can perform this rescaling process to any given discrete (long) RW, and obtain the same probability density of longest loop, independent of the number of steps N, and of the step size a of the given RW. Since the scaling process is statistically exact, the probability density of the longest loop, calculated from the rescaled RW, is equal to the probability density calculated from the given long RW. We therefore conclude that the probability density of the longest loop for discrete RW's is universal, and for large N's it does not depend on the number of steps N, and on the step size a of the RW.

4.2 The Continuous Case

In this section we show that the probability density of the longest loop for Gaussian RW's for large N's is independent of N and of ϵ, the distance between steps defining a loop (see section 3.1.1). Furthermore, we show that this probability density is equal to the 'universal' probability density obtained in the previous section for discrete RW's.

We perform the rescaling process, discussed in the previous section, on a given (long enough) Gaussian RW. (We divide the Gaussian RW into m equal segments, randomly generate a minimum and a maximum for each segment according to the probability density $g(w,W;x)$ of Eq. (4.19), and define a loop if two such segments intersect each other.) The emerging probability density of the longest loop is the 'universal' probability density,

obtained for the discrete RW's, because when performing the rescaling in the previous section, we did not limit the RW to be discrete. However, we must show that the probability of a having a loop of certain (reduced) length between any two segments according to this process, is equal to the probability of having a loop of the same length in the original Gaussian RW (i.e. a loop which starts with a step in the first segment, and ends with a step in the second segment). In order to show that the probabilities of a loop are the same in the rescaled and original RW, we must prove two claims:

(1) The minimum and maximum, randomly assigned to each segment in the rescaled RW according to $g(w, W; x)$, are valid (i.e. statistically exact) for Gaussian RW's.

(2) When two segments in the rescaled RW intersect (i.e. the minimum of one being between the minimum and maximum of the other), a loop between them is formed in the original Gaussian RW.

The first claim is valid since $g(w, W; x)$, the mutual probability density of the minimum and maximum for discrete RW's, was calculated only based on the Gaussian statistics of each sub-walk, and is therefore also the mutual probability density of the minimum and maximum for Gaussian RW's.

We prove the second claim by considering a Gaussian RW of N steps of size a, where a loop between two steps is defined if their positions are closer than an arbitrary ϵ from each other. We scale the positions of steps of the RW by $a\sqrt{N}$, the average distance between the minimum and the maximum (the distance 'spread' by the RW), making the distance between the minimum and maximum of order unity. The N steps of the RW are spread along the position axis between the minimum and the maximum of the RW, according to some probability density $q(x)$ (where $\int_{\min}^{\max} q(x)dx = 1$). The average number of steps of the RW at a certain position x within the (rescaled) interval defining a loop of $\Delta x = \epsilon/(a\sqrt{N})$ is given by:

$$Nq(x)\Delta x = Nq(x)\frac{\epsilon}{a\sqrt{N}} \sim \sqrt{N} \ . \tag{4.20}$$

The $\sqrt{N}$ dependence means that the average number of RW steps within the Δx interval diverges with increasing N, for all the positions along the RW between the minimum and the maximum, and for all $\epsilon > 0$. We see that in the large N limit the entire range along the position axis between the minimum and the maximum is covered by the '$\epsilon-$ranges' (i.e.

positions closer than ϵ) of the steps in the Gaussian RW. Therefore, when the minimal coordinate reached by one Gaussian sub-RW is between the minimum and maximum of another Gaussian sub-RW, it is always closer than ϵ to a position of a certain step in the second sub-RW, and a loop is formed in the original Gaussian RW.

This second claim is also valid when a loop between two steps in a Gaussian RW is defined by a probability, proportional to the distance between the steps divided by ϵ (as in section 3.3): As N increases, there is an infinite number of steps closer than ϵ to a certain given step, each having a finite probability to make a loop with the given step, and therefore a loop is generated with unit probability.

We have thus proved the 'universality' of the probability density of the longest loop: We showed that the probability density of the longest loop (for large enough N's) is independent of N and is the same for both continuous RW's with a loop defined according to an arbitrary $\epsilon > 0$, and for discrete RW's. In the notations of section 3, we have shown that:

$$\lim_{N\to\infty} p(l; N, \epsilon/a) = p(l) \ . \tag{4.21}$$

4.3 The 'Universal' Probability Density

We have seen that the probability density of the longest loop is 'universal' for many classes of RW's, in which the probability of the position of steps after a large number of steps is Gaussian, and the entire region between the minimum and maximum of the RW is 'covered' by the distance defining a loop for each step. For all these classes of RW's the scaling process of section 4.1.2 can be repeated, resulting in the same universal probability density, independent of the details of the RW or of what is called a loop.

For continuous RW's we have lost through this scaling process the parameter ϵ which defined a loop. This was accomplished by assigning to each 'effective step' a certain width, randomly generated from a given probability, determined only by the position of the next step in the RW. We also note that by the definition of an 'effective step', which is actually a RW in the general continuum limit, the discretization of the problem and the distribution of a single step are lost (each sub-walk can represent a sequence of n steps of size a, or kn steps size $a/\sqrt{k}$ or a single step of size $a\sqrt{n}$ in a Gaussian RW).

4.3.1 Comparison to the Probability Density for Discrete Random Walks

In this section we explore the universal probability density emerging from the scaling process, and compare it to the probability density of the longest loop for discrete RW's. We show that as the number of sub-walks (to which the original RW was divided in order to get the rescaled walk) increases, the universal probability density converges to the $N \to \infty$ limit of the probability density for discrete RW's. We denote the probability density of having a longest loop of (reduced) length $l = L/m$, in a RW of m 'effective steps', by $f_m(l)$. We have numerically obtained $f_m(l)$ for m=4, 10, 20, 50 and 100 (each from 10^5 independent random sequences of length m), and compared them to $p(l)$, obtained by MC simulation of 10^6 random sequences of $N = 1000$ discrete steps. These functions are depicted in Fig. 32. For each value of m, the possible values of l are $\frac{1}{m}, \frac{2}{m}, \cdots, \frac{m-1}{m}$. The

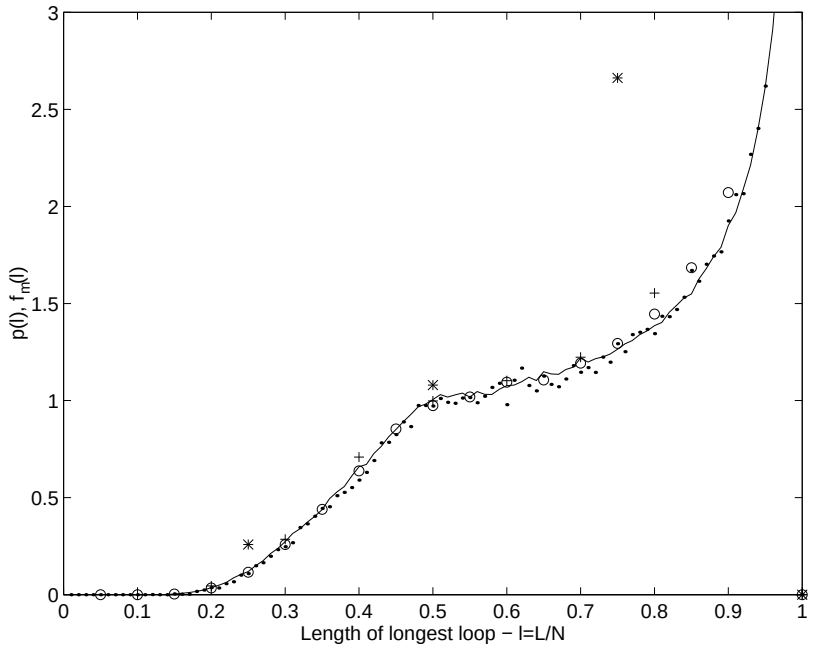

Figure 32: $f_m(l)$ *vs.* l for several values of m: 4 (star), 10 (plus), 20 (circle), 100 (dot) compared with $p(l)$ (line) as generated by MC simulation of 10^6 random sequences of length $N = 1000$. The rapid convergence is evident (see text).

length of the longest loop is never zero, since neighboring sub-walks always make a loop (as evident in section 4.1.1), and is never equal to m, since a loop from the first to the last (mth) step is $m - 1$ segments long. It is evident from Fig. 32 that $f_m(l)$ converges very quickly with increasing m to $p(l)$ (even for $m = 10$ the function $f_m(l)$ fits $p(l)$ very well for $l < 0.8$). We note that (for small values of m), there is no convergence in the

$l \to 1$ limit. We will come to this issue at the end of the section.

The convergence of $f_m(l)$ to $p(l)$ is analyzed quantitatively at table 1, where for each m we calculate the average (over all values of l) distance $|f_m(l) - p(l)|$, compare it with the average errors of $f_m(l)$ (errors resulting from the numeric process), and find the percentage of numeric data which fit the simulation value within their error bars.[10] Although

m	4	10	20	50	100
$\langle \lvert f_m(l) - p(l) \rvert \rangle$	0.12	0.038	0.029	0.032	0.031
$\frac{\langle \lvert f_m(l) - p(l) \rvert \rangle}{\langle \text{numeric error} \rangle}$	24	5.1	2.6	1.8	1.2
fit to simulation	0%	67%	79%	65%	78%

Table 1: Convergence measurements of $f_m(l)$ to $p(l)$ for several numbers of 'effective steps' m in the rescaled RW. The measurements include the average (over l) distance between the functions, the distance divided by the average error resulting from the numeric process, and the percentage of numeric values which fit the simulation within their error limits.

some of these convergence measurements depend on the size of the sample (which is 10^5 independent random sequences of length m), their relative sizes for different m's indicate the fast convergence of $f_m(l)$ to $p(l)$ with increasing m. We see that for $m \geq 10$, most of the data fit the simulation within their error limits, and for $m \geq 20$ the errors due to the numeric process become in order of the inaccuracy of the scaling process. As stated in section 4.1.2 the inherent inaccuracies of the rescaling process are proportional to the length of the segments (i.e. for $m = 10$, the inherent inaccuracy is of order of one tenth of the chain). These diminishing inaccuracies explain the convergence of $f_m(l)$ to $p(l)$ with increasing m.

The probability density of the second longest loop (or third longest loop and so on) can be calculated similarly to the probability density of the longest loop. In Fig. 33 the probability density (from 10^5 random sequences) of the second longest loop of a rescaled chain, having $m = 20$ segments, is compared to the same probability density of a discrete RW of $N = 1000$ steps, obtained by MC simulation of 10^6 random sequences. These probabilities are found to be (qualitatively) very similar. However, we cannot derive the universality of the probability density of the second longest loop from exactly the same considerations that led to the universality of the probability density of the longest

[10] When we measure distance or fit to the simulation, we do not compare the numeric value to the specific value of the simulation at a given point L, but rather to the average over $\frac{L-0.5}{m}$ to $\frac{L+0.5}{m}$, which is the most reasonable representation of the simulation with m steps.

loop. The remaining part of the chain, after the longest loop is 'erased', does not follow the statistics of a classical RW (on average, it is more stretched than a RW having the same length). Therefore, we cannot perform the scaling process of section 4.1.2 on the remaining part of the chain, and derive the probability density of the second longest loop in a universal way (the mutual probability density of the minimum and maximum $g(w, W; x)$ is valid only for unbiased RW's). To prove the universality of the second longest loop, the mutual probability of minimum and maximum should be generalized for the statistics of RW's with erased longest loop, and shown to be independent of the single step probabilities in this case. This proof is beyond the scope of our work.

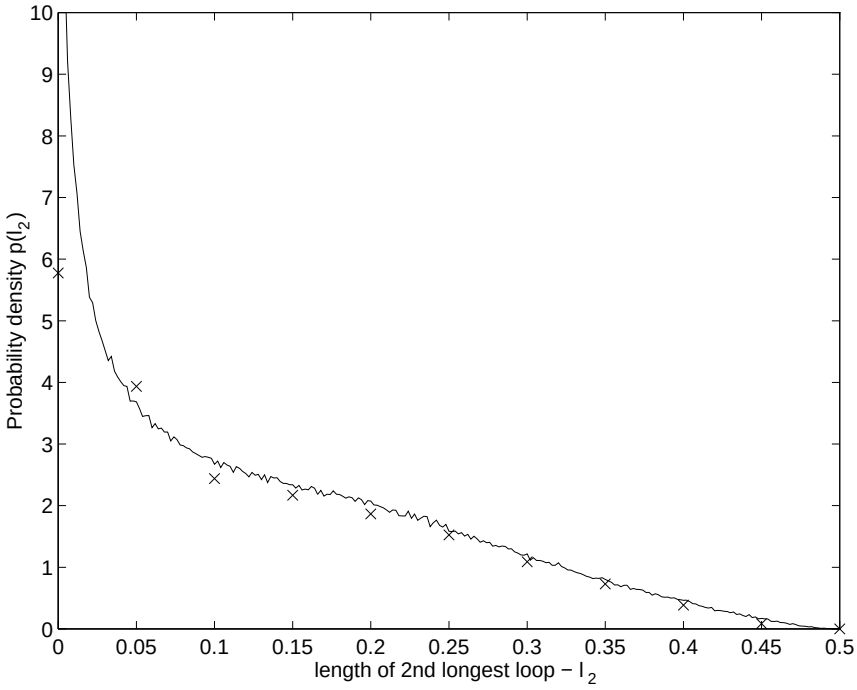

Figure 33: Probability density of the second longest loop of a rescaled chain of m=20 segments ($\times$), compared with the probability density of the second longest loop of a discrete RW with N=1000 steps (line).

4.3.2 Analytical Properties in the Limit of Long Loops

In this section we investigate $p(l)$ in the $l \to 1$ limit, using the rescaled probability density. An infinitely long RW with N steps of fixed length a, is divided into m segments of τ steps. We are interested in $f_m(\frac{m-1}{m})$, the probability of having a loop between the first and the last segments. It was found in section 4.1.1 that when the last segment's origin is shifted Φ from the origin of the RW, the probability of a loop between the first and

last segments is $P_{\rm RW}\left(\kappa = m-2,\ \Delta = \frac{\Phi}{a\sqrt{\tau}}\right)$. In order to obtain the probability of a loop between the first and last segments, $P_{\rm RW}$ is integrated over all possible values of Φ with the probability of the RW to reach Φ after $m-1$ segments:

$$f_m(\frac{m-1}{m}) = \int_{\Phi=-\infty}^{\infty} P_{\rm RW}\left(m-2, \frac{\Phi}{a\sqrt{\tau}}\right) p(\Phi,(m-1)\tau)d\Phi \ . \tag{4.22}$$

We therefore get:

$$f_m(\frac{m-1}{m}) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m-1}} \cdot I(m) \ , \tag{4.23}$$

where

$$I(m) \equiv \int_{\Delta=-\infty}^{\infty} e^{-\frac{1}{2}\frac{\Delta^2}{m-1}} P_{\rm RW}\left(m-2,\Delta\right) d\Delta \ . \tag{4.24}$$

In the $m \to \infty$ limit, Eqs. (4.23) and (4.24) become:

$$f_m(\frac{m-1}{m}) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}} \cdot I \ , \tag{4.25}$$

where

$$I = \int_{\Delta=-\infty}^{\infty} P_{\rm ind}(\Delta)d\Delta \simeq 3.2 \ . \tag{4.26}$$

Substituting $l_0 = \frac{m-1}{m}$ in Eq. (4.25), we get:

$$f_m(l_0) = \frac{1}{\sqrt{2\pi}}\sqrt{1-l_0} \cdot I \ . \tag{4.27}$$

The behavior of $p(l)$ in the $l \to 1$ limit was obtained by Kantor and Ertaş [26]:

$$p(l \to 1) = \frac{A}{\sqrt{\pi(1-l)}} \ , \tag{4.28}$$

where A is a numerically obtained constant ($A = 1.011 \pm 0.001$). This means that the probability P_{l_0} of a loop to be longer than l_0 is (in the $l_0 \to 1$ limit):

$$P_{l_0} = \int_{l=l_0}^{1} p(l \to 1)dl = \frac{2}{\sqrt{\pi}} A\sqrt{1-l_0} \ . \tag{4.29}$$

We see (by comparing Eq. (4.27) to Eq. (4.29)) that the $l-$dependence of $f_m(l)$ is in accordance with the behavior of $p(l)$ in the $l \to 1$ limit as derived by Kantor and Ertaş. As indicated before, the probability $f_m(\frac{m-1}{m})$ includes events of longest loops being both shorter and longer than exactly $m-1$ segments long. We can see that any loop longer than $\frac{m-1}{m}$ of the chain, begins at the first segment and ends at the last one (and therefore

is 'counted' by $f_m(\frac{m-1}{m})$), and that all the loops that are between the first and the last segments are longer than $\frac{m-2}{m}$ of the chain. We therefore get for all m:

$$P_{l_0=\frac{m-1}{m}} < f_m(\frac{m-1}{m}) < P_{l_0=\frac{m-2}{m}} \ . \tag{4.30}$$

In the $m \to \infty$, $l_0 \to 1$ limit we get by substituting the definitions of P_{l_0} and $f_m(\frac{m-1}{m})$ to Eq. (4.30):

$$\frac{2}{\sqrt{\pi}} A \frac{1}{\sqrt{m}} < \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{m}} I < \frac{2}{\sqrt{\pi}} A \sqrt{\frac{2}{m}} \ . \tag{4.31}$$

Eq. (4.31) allows us to find *analytical* upper and lower limits to A:

$$0.80 < A < 1.13 \ , \tag{4.32}$$

which are satisfied by the known numeric value of $A = 1.011$ obtained in [26].

The probability $f_m(\frac{m-2}{m})$ can be calculated similarly to the calculation of $f_m(\frac{m-1}{m})$. It is the probability of having a loop between the first and the one-before-last segments or second to last segments, while not having a loop between the first and last segments. The calculation of $f_m(\frac{m-2}{m})$ is very cumbersome, since these probabilities are dependent, but is still possible, because the dependencies are known (they effect the possible values of the minimum or maximum of the segments). In the same way, $f_m(\frac{m-k}{m})$ can be calculated (for every finite k) without the numeric process of mutually generating minimum and maximum to each segment. We attempted to find boundaries for A through the inequality:

$$f_m(\frac{m-k}{m}) < p\left(\frac{m-(k+1)}{m} < l < \frac{m-(k-1)}{m}\right) = \frac{2A}{\sqrt{m\pi}}(\sqrt{k+1} - \sqrt{k-1}) \ . \tag{4.33}$$

We substituted the numerical values of $f_m(\frac{m-k}{m})$, but the obtained boundaries for A were not tighter than Eq. (4.32).

From the analytical value of $f_m(\frac{m-1}{m})$ in the $m \to \infty$ limit (Eqs. 4.25 and 4.26) we can see that $f_m(l)$ does not converge to $p(l)$ in the $l \to 1$ limit:

$$f_m(\frac{m-1}{m}) \simeq 2.26\sqrt{\frac{m}{\pi}} \ , \quad \text{for } m \to \infty \ . \tag{4.34}$$

We compare this value to the average of $p(l)$ over values of l from $\frac{m-1-0.5}{m}$ to $\frac{m-1+0.5}{m}$:

$$\left\langle p(\frac{m-1}{m}) \right\rangle \equiv m \int_{l=\frac{m-1.5}{m}}^{\frac{m-0.5}{m}} p(l) dl = \sqrt{\frac{2m}{\pi}} A(\sqrt{3}-1) \simeq 1.04\sqrt{\frac{m}{\pi}} \ . \tag{4.35}$$

The first equality is derived by substituting $p(l)$ in the $l \to 1$ limit from Eq. (4.28) and performing the integral, while the second is derived by substituting the numeric value of A. We see that for the last step in the rescaled RW, $f_m(l)$ does not converge to $p(l)$, but rather to a value greater than twice the value of $p(l)$. This non-convergence for the last step, although clear from Fig. 32 for small values of m, is negligible for $m \to \infty$, since it concerns only $1/m$ of the chain.

5 Conclusions and Discussion

We have investigated the size distribution of neutral segments in randomly charged polymers with positive and negative charges (polyampholytes). According to the necklace model [20, 21], in the ground state of polyampholytes, such segments compact into globules. Following Kantor and Ertaş [25, 26], we have mapped the problem of longest neutral segments in polyampholytes to the problem of longest loops in one-dimensional random walks, and applied numeric methods along with analytic estimates to study the probability density of longest loops.

Since we believe that a polyampholyte's structure based on the necklace model has a very low energy, we suggested a specific detailed necklace-type structure for polyampholytes in the ground state, and numerically obtained its conformational and physical properties (the number and sizes of 'beads' and 'strings' in the necklace, the spatial extent, the surface area and energy). This structure is compact when the chain is neutral or weakly charged, and stretches as the chain becomes charged. We find that the ground state structure has a very low energy, which depends on the number of monomers as the energy of a single compact neutral globule. We find that the unrestricted average of the linear size of the polymer in the ground state depends on the number of monomers as the linear size of an ideal chain, with a critical exponent of $\nu = 0.50 \pm 0.01$. This finding is not in accordance with previous studies [21, 22], concluding that the average linear size increases with N at least as fast as a self-avoiding walk (i.e. $\nu > 0.6$). We believe that it is worthwhile to slightly alter the way in which compact globules are formed within our model (by allowing weakly charged segments to compact into globules or by not forcing all the neutral segments to completely compact), in order to try to reproduce this 'swelling' of the average chain.

We have defined the problem of the longest loop for continuous Gaussian random walks, investigated the resulting probability densities, and showed numerically that they converge with increasing number of steps in the random walk to the probability density of the longest loop in random walks with steps of fixed length. These results motivated a scaling process, which enabled us to obtain a probability density of the longest loop, which is independent of the number of steps and of the nature of the single step of the random walk. This probability density is identical for random walks with steps of fixed length and for Gaussian random walks. We have presented numerical and analytical evidence that

this probability density of the longest loop is universal for large classes of random walks. Investigating this universal probability density, we have obtained some of its analytical properties, in the limit of long loops. It may be possible to establish additional analytical properties of the problem, through further investigation of this universal function. However, a full renormalization-group treatment of the problem within the derived scaling process, in order to find a *complete* analytical solution to the problem, is expected to be quite complicated. As opposed to a standard renormalization-group treatment [2, 49], where a limiting point or exponent is searched, we are interested in the entire probability density, which is actually a 'limiting function'.

It may be possible to generalize the continuous definitions and the scaling process leading to a universal probability density of the longest loop, in order to find the probability density of the second longest loop, or other related probabilities, that are relevant to the understanding of the suggested ground state structure of polyampholytes. Finding the probability density of the longest segment of a given charge (largest "Q-segment") in the continuum limit, and proving its universality, is another possible generalization of this study.

A Probability of a Loop Between Random Walks

We present the derivation of the probability of an 'intersection' between two RW's, that are sub-walks of one RW with steps of fixed length. Such an intersection indicates that a loop is formed between the two sub-walks (see Fig. 27 in section 4). The probability of a loop is derived for an infinitely long RW, where each sub-walk can be treated within the Gaussian statistics. We use the notations of section 4 – the number of steps in each sub-walk is τ, the size of each step is a, there are $\kappa\tau$ steps of the RW between the two sub-walks, the end position of the first sub-walk is x, and the second sub-walk begins at a position Φ relative to the origin of the first sub-walk. Without loss of generality we can suppose that $\Phi \geq 0$. The two sub-walks form a loop when the maximal coordinate of the first walk is greater or equal to the minimal coordinate of the second walk (labeled z).

As stated in section 4, in order to have a loop, three independent events, with probabilities denoted by P_1, P_2 and P_3, must occur:

(1) The maximal coordinate of the first sub-walk must be greater or equal to the minimal coordinate of the second sub-walk. The probability of a RW to reach a position x and to have a maximum greater than z after τ steps is given by [34]:

$$P_1(x,z) = \begin{cases} p(2z-x,\tau) & ; \text{ for } z \geq \max(0,x) \\ p(x,\tau) & ; \text{ otherwise} \end{cases} , \tag{A.1}$$

where $p(x,\tau)$ is the probability of a RW to be at position x after τ steps (given by Eq. 1.13).

(2) There is a RW between the end position of the first sub-walk and the origin of the second sub-walk, i.e. a RW of $\kappa\tau$ steps and a total displacement of $\Phi - x$. However, the position Φ is fixed, and therefore this RW is restricted by the existence of a RW of $\kappa\tau + \tau$ steps from the origin to Φ. This probability is given by

$$P_2(\Phi,x) = \frac{p(\Phi - x, \kappa\tau)}{p(\Phi,(\kappa+1)\tau)} . \tag{A.2}$$

(3) The minimum of the second sub-walk is equal to z, i.e $\Phi - z$ relative to its origin. This probability is given by Eq. (4.1):

$$P_3(\Phi,z) = m(z-\Phi,\tau) = \begin{cases} 2p(\Phi - z,\tau) & ; \text{for } z \leq \Phi \\ 0 & ; \text{for } z > \Phi \end{cases} . \tag{A.3}$$

In order to get the probability of a loop, denoted by $P_{\rm RW}(\kappa,\Delta)$ (for given Φ, a, τ and κ, when $\Delta \equiv \Phi/a\sqrt{\tau}$), we integrate $P_1 \cdot P_2 \cdot P_3$ over all values of z and x:

$$P_{\rm RW}(\kappa,\Delta) = \int_{z=-\infty}^{\infty}\int_{x=-\infty}^{\infty} P_1 P_2 P_3 dz dx \ . \tag{A.4}$$

From the definitions of P_1 and P_3 we see that if $z<0$ then $P_1 = p(x,\tau)$, if $0 \le z \le \Phi$ then $P_1 = p(x,\tau)$ when $x > z$, and if $z > \Phi$ then P_3 vanishes. We therefore get:

$$\begin{aligned} P_{\rm RW}(\kappa,\Delta) \quad &= \int_{z=-\infty}^{0} P_3(\Phi,z)dz \int_{x=-\infty}^{\infty} p(x,\tau)P_2(\Phi,x)dx \\ &+ \int_{z=0}^{\Phi} P_3(\Phi,z) \int_{x=-\infty}^{z} P_1(x,z)P_2(\Phi,x)dxdz \\ &+ \int_{z=0}^{\Phi} P_3(\Phi,z) \int_{x=z}^{\infty} p(x,\tau)P_2(\Phi,x)dxdz \ . \end{aligned} \tag{A.5}$$

We denote the first term in Eq. (A.5) by $A(\kappa,\Delta)$, the second one $B(\kappa,\Delta)$ and the third $C(\kappa,\Delta)$, and calculate each of them, substituting the definitions of P_1, P_2, P_3 and using the definitions of the normal density function $\Pi(x)$ and the normal distribution function $\eta(x)$ of Eqs. (4.4) and (4.5):

$$\begin{aligned} A(\kappa,\Delta) \quad &= \int_{z=-\infty}^{0} P_3(\Phi,z)dz \int_{x=-\infty}^{\infty} p(x,\tau)\frac{p(\Phi-x,\kappa\tau)}{p(\Phi,(\kappa+1)\tau)}dx \\ &= \int_{z=-\infty}^{0} P_3(\Phi,z)dz = 2\eta\left(\frac{-\Phi}{a\sqrt{\tau}}\right) = 2\eta(-\Delta) \ . \end{aligned} \tag{A.6}$$

$$\begin{aligned} B(\kappa,\Delta) &= \int_{z=0}^{\Phi} P_3(\Phi,z)\sqrt{\frac{\kappa}{\kappa+1}}e^{\frac{\Phi^2}{2a^2(\kappa+1)\tau}}e^{-\frac{(2z-\Phi)^2}{2a^2(\kappa+1)\tau}}\frac{1}{a\sqrt{2\pi\tau}}\int_{x=-\infty}^{z} e^{-\frac{1}{2a^2\kappa\tau}\left(\sqrt{\kappa+1}x-\frac{2\kappa z+\Phi}{\sqrt{\kappa+1}}\right)^2}dxdz \\ &= \int_{z=0}^{\Phi} P_3(\Phi,z)e^{\frac{4z(\Phi-z)}{2a^2(\kappa+1)\tau}}\eta\left(-\frac{(\kappa-1)z+\Phi}{a\sqrt{\kappa(\kappa+1)\tau}}\right)dz \\ &= \sqrt{\frac{2(\kappa+1)}{\pi(\kappa+5)}}\frac{1}{\Pi\left(\frac{2\Delta}{\sqrt{(\kappa+1)(\kappa+5}}\right)}\int_{\frac{-(\kappa+3)\Delta}{\sqrt{(\kappa+1)(\kappa+5)}}}^{\frac{2\Delta}{\sqrt{(\kappa+1)(\kappa+5)}}}\Pi(y)\eta\left(\frac{(1-\kappa)y}{\sqrt{\kappa(\kappa+5)}}-\frac{\kappa+2}{\kappa+5}\Delta\sqrt{\frac{\kappa+1}{\kappa}}\right)dy \end{aligned} \tag{A.7}$$

$$\begin{aligned} C(\kappa,\Delta) \quad &= \int_{z=0}^{\Phi} P_3(\Phi,z)\eta\left(\frac{\Phi}{a\sqrt{\tau\kappa(\kappa+1)}}-\sqrt{\frac{\kappa+1}{\kappa}}\frac{z}{a\sqrt{\tau}}\right)dz \\ &= 2\int_{y=-\Delta}^{0}\Pi(y)\eta\left(-\sqrt{\frac{\kappa}{\kappa+1}}\Delta-\sqrt{\frac{\kappa+1}{\kappa}}y\right)dy \ . \end{aligned} \tag{A.8}$$

Adding these three terms we get the probability of a loop:

$$P_{\rm RW}(\kappa,\Delta) = 2\eta(-\Delta) + 2\int_{y=-\Delta}^{0} \Pi(y)\eta\left(-\sqrt{\frac{\kappa+1}{\kappa}}y - \sqrt{\frac{\kappa}{\kappa+1}}\Delta\right)dy + \tag{A.9}$$

$$\sqrt{\frac{2(\kappa+1)}{\pi(\kappa+5)}}\frac{1}{\Pi\left(\frac{2\Delta}{\sqrt{(\kappa+1)(\kappa+5}}\right)}\int_{y=-\frac{(\kappa+3)\Delta}{\sqrt{(\kappa+1)(\kappa+5)}}}^{y=\frac{2\Delta}{\sqrt{(\kappa+1)(\kappa+5)}}} \Pi(y)\eta\left(\frac{(1-\kappa)y}{\sqrt{\kappa(\kappa+5)}} - \frac{\kappa+2}{\kappa+5}\Delta\sqrt{\frac{\kappa+1}{\kappa}}\right)dy \ . \tag{A.10}$$

We are interested in exploring $P_{\rm RW}(\kappa,\Delta)$ in the $\kappa\to\infty$ limit, showing that it converges to the probability of a loop between independent sub-walks, and finding how fast is the convergence. We therefore expand $P_{\rm RW}(\kappa,\Delta)$ in powers of $1/\kappa \equiv \zeta$, neglecting terms of order ζ^2. We use the following approximations, which are valid when neglecting terms of order of $1/\kappa^2$:

$$\begin{array}{ll} \sqrt{\frac{\kappa+1}{\kappa+5}} = 1-\frac{2}{\kappa} & \frac{\kappa-1}{\sqrt{\kappa(\kappa+5)}} = 1-\frac{7}{2\kappa} \\ e^{\frac{2\Delta^2}{(\kappa+1)(\kappa+5)}} = 1 & \frac{\kappa+2}{\kappa+5} = 1-\frac{3}{\kappa} \\ \frac{1}{\sqrt{(\kappa+1)(\kappa+5)}} = \frac{1}{\kappa} & \sqrt{\frac{\kappa+1}{\kappa}} = 1+\frac{1}{2\kappa} \\ \frac{\kappa+3}{\sqrt{(\kappa+1)(\kappa+5)}} = 1 & \sqrt{\frac{\kappa}{\kappa+1}} = 1-\frac{1}{2\kappa} \ . \end{array} \tag{A.11}$$

We expand $B(\kappa,\Delta)$ and $C(\kappa,\Delta)$ in powers of ζ (the expression $A(\kappa,\Delta)$ is independent of κ) according to these approximations:

$$B(\kappa,\Delta) = 2(1-2\zeta)\int_{y=-\Delta}^{2\Delta\zeta} dy\Pi(y)\eta\left[-y(1-\frac{7\zeta}{2}) - \Delta(1-\frac{5\zeta}{2})\right] \ . \tag{A.12}$$

The $B(\kappa\to\infty,\Delta)$ limit is obtained by substituting $\zeta=0$ in Eq. (A.12), while for the first correction we take the derivative in respect to ζ, for $\zeta=0$.

$$\begin{aligned} \left.\frac{dB(\kappa,\Delta)}{d\zeta}\right|_{\zeta=0} &= 2\int_{y=-\Delta}^{0}\left[\left(\frac{7y}{2}+\frac{5\Delta}{2}\right)\Pi(y)\Pi(-y-\Delta)\right]dy + 4\Delta\cdot\Pi(0)\eta(-\Delta) \\ &-4\int_{y=-\Delta}^{0}\Pi(y)\eta(-y-\Delta)dy \\ &= \frac{3\Delta}{2\sqrt{2}}\Pi\left(\frac{\Delta}{\sqrt{2}}\right)\left[2\eta\left(\frac{\Delta}{\sqrt{2}}\right)-1\right] + \frac{4\Delta}{\sqrt{2\pi}}[1-\eta(\Delta)] - 4\eta(\Delta) + 2 \\ &+4\int_{y=-\Delta}^{0}\Pi(y)\eta(y+\Delta)dy \ . \end{aligned} \tag{A.13}$$

For $C(\kappa,\Delta)$ we get, when neglecting terms of order of ζ^2:

$$C(\kappa,\Delta) = 2\int_{y=-\Delta}^{0} \Pi(y)\eta\left(-y-\Delta+\frac{\zeta(\Delta-y)}{2}\right)dy \ , \tag{A.14}$$

and

$$\left.\frac{dC(\kappa,\Delta)}{d\zeta}\right|_{\zeta=0} = 2\int_{y=-\Delta}^{0}(\Delta-y)\Pi(y)\Pi(-y-\Delta)dy = \frac{3\Delta}{2\sqrt{2}}\Pi\left(\frac{\Delta}{\sqrt{2}}\right)\left[2\eta\left(\frac{\Delta}{\sqrt{2}}\right)-1\right] \ . \tag{A.15}$$

From substituting $\zeta = 0$ in Eqs. (A.12) and (A.14) we get:

$$\begin{aligned} P_{\mathrm{RW}}(\kappa\to\infty,\Delta) &= A(\kappa,\Delta)+B(\zeta=0,\Delta)+C(\zeta=0,\Delta) \\ &= 2\eta(-\Delta)+4\int_{y=-\Delta}^{0}\Pi(y)\eta(-y-\Delta)dy \\ &= 2\eta(\Delta)-4\int_{y=-\Delta}^{0}\Pi(y)\eta(y+\Delta)dy \ , \end{aligned} \tag{A.16}$$

which is exactly the expression for $P_{\mathrm{ind}}(\Delta)$, the probability of a loop between independent sub-walks (see Eq. 4.6). For the first correction to $P_{\mathrm{ind}}(\Delta)$ we get:

$$\begin{aligned} f(\Delta) \equiv \left.\frac{dP_{\mathrm{RW}}(\kappa\Delta)}{d\zeta}\right|_{\zeta=0} &= \left.\frac{dB(\kappa,\Delta)}{d\zeta}\right|_{\zeta=0}+\left.\frac{dC(\kappa,\Delta)}{d\zeta}\right|_{\zeta=0} \\ = 2+4\int_{-\Delta}^{0}\Pi(y)\eta(y+\Delta)dy &+ \frac{4\Delta\eta(-\Delta)}{\sqrt{2\pi}}-4\eta(\Delta)+\frac{3\Delta}{\sqrt{2}}\Pi\left(\frac{\Delta}{\sqrt{2}}\right)\left[2\eta\left(\frac{\Delta}{\sqrt{2}}\right)-1\right] \ , \end{aligned} \tag{A.17}$$

where to first order in $1/\kappa$:

$$P_{\mathrm{RW}}(\kappa,\Delta) = P_{\mathrm{ind}}(\Delta)+\frac{1}{\kappa}\cdot f(\Delta) \ . \tag{A.18}$$

The first order correction function $f(\Delta)$ is depicted in Fig. 34, and is compared with its numeric value for $\kappa = 50$ (i.e.: $[P_{\mathrm{RW}}(\kappa=50,\Delta)-P_{\mathrm{ind}}(\Delta)]\cdot 50$). We see that for typical values of Δ, the difference between $P_{\mathrm{RW}}(\kappa,\Delta)$ and $P_{\mathrm{ind}}(\Delta)$ falls off only as $1/\kappa$, and we therefore cannot assume that for most values of κ, the function $P_{\mathrm{RW}}(\kappa,\Delta)$ is approximated by $P_{\mathrm{ind}}(\Delta)$.

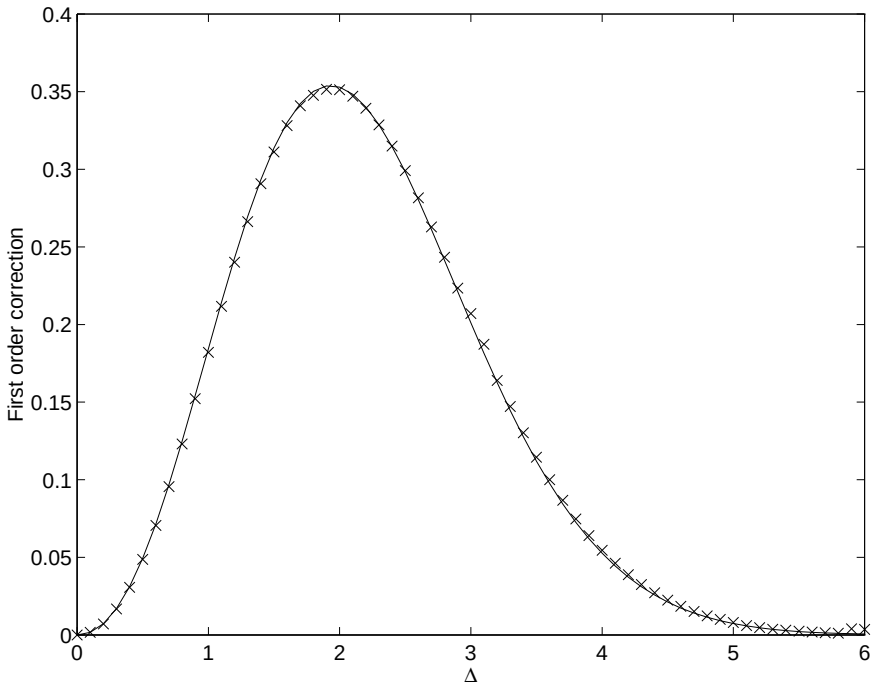

Figure 34: First order correction function to $P_{\mathrm{RW}}(\kappa, \Delta)$ (line), compared with its numeric value for κ=50 ('$\times$'). $P_{\mathrm{RW}}(\kappa \to \infty, \Delta) \simeq P_{\mathrm{ind}}(\Delta) + \frac{1}{\kappa} \cdot$ (Correction).

B Random Generation of Extrema of Random Walks

We present a numerical process of randomly generating a minimum $(-w)$ and a maximum (W) of a RW, ending at a position x.[11] The 2-dimensional (2-d) joint probability density of a given RW, ending at a position x, to have a minimum $-w$ and a maximum W, is given by $g(w, W; x)$, defined in Eq. (4.19). The problem of randomly generating w and W from their 2-d joint probability density has two parts: First of all we must deal with the general theoretical problem of generating random numbers according to a predefined 2-d distribution, using only uniform distributions of random numbers (generated by standard random number generators). Secondly, we must find methods to facilitate the numerical process of generating the specific probability density $g(w, W; x)$, since the general theoretical methods usually involve cumbersome inversions of the probability density.

One of the standard methods to generate a predefined (positive and normalized) probability density $f(y)$ from a uniform probability density, is the *transformation method* [50]: We first calculate F, the indefinite integral of f:

$$F(y) = \int_{y'=-\infty}^{y} f(y')dy' \ . \tag{B.1}$$

The inverse function $y(F)$ takes a uniform density into a one distributed as $f(y)$. In order to use this method for a 2-d probability density, we express $g(w, W; x)$ as a product of two 1-dimensional probability densities: $f_1(W)$, the probability density of the maximum of the RW, and $f_2(w|W_0)$, the conditional probability density of the minimum w, when the maximum is W_0. These functions are given by:

$$f_1(W) = \int_{w'=0}^{\infty} g(w', W; x)dw' \ , \tag{B.2}$$

$$f_2(w|W_0) = g(w, W; x|W = W_0) = \frac{g(w, W_0; x)}{\int_{w'=0}^{\infty} g(w', W_0; x)dw'} \ . \tag{B.3}$$

We calculate $I(W)$, the integral of $f_1(W)$, where W can take only positive values:

$$I(W) = \int_{W'=0}^{W} f_1(W')dW' = \int_{W'=0}^{W}\int_{w'=0}^{\infty} g(w', W'; x)dw'dW' \ . \tag{B.4}$$

Since $g(w, W; x)$ is the probability density of a given RW, ending at a position x, to have a minimum position of $-w$ and a maximum position of W, then $I(W)$ is the probability

[11] Throughout this appendix we discuss RW's whose origins are at 0, and all the positions are measured relative to the origin.

that a given RW, ending at a position x, has a maximal coordinate lower than W. This probability is equal to the probability density $p(x,t;W)$ of a RW after t steps never to reach an absorbing wall at W (see Eq. 1.16), divided by $p(x,t)$, the probability density of a RW to be at a position x after t steps. When we measure all the positions in units of $a\sqrt{t}$, where a is the step size of the RW and t is the number of steps of the walk (as was done in section 4), we get:

$$I(W) = \frac{p(x,t;W)}{p(x,t)} = \left[e^{-\frac{x^2}{2}} - e^{-\frac{(2W-x)^2}{2}}\right] e^{\frac{x^2}{2}} = 1 - e^{-2W(W-x)} \ . \tag{B.5}$$

Inverting $I(W)$ for $W \geq 0$ we get:

$$W(I) = \frac{x}{2} + \sqrt{\left(\frac{x}{2}\right)^2 - \frac{\ln(1-I)}{2}} \ . \tag{B.6}$$

When I is a random number, generated from a uniform probability density between 0 and 1, then W of Eq. (B.6) is distributed according to $f_1(W)$.

After W has been set to a given W_0, we randomly generate w according to $f_2(w|W_0)$. We calculate $J(w)$, the integral of f_2 for positive values of w:

$$J(w) = \int_{w'=0}^{w} f_2(w'|W_0)dw' = \frac{\int_{w'=0}^{w} g(w',W_0;x)dw'}{\int_{w'=0}^{\infty} g(w',W_0;x)dw'} = \frac{g_2(w,W_0;x)}{g_2(w=\infty,W_0;x)} \ , \tag{B.7}$$

where

$$g_2(w,W;x) \equiv \int_{w'=0}^{w} g(w',W;x)dw' = \frac{\partial G(w,W;x)}{\partial W} \ , \tag{B.8}$$

and $G(w,W;x)$ is the probability of a given RW ending at a position x to be always above $-w$ and below W (see Eq. 4.18). The expression $g_2(w,W;x)\Delta W$ is the probability of a given RW, ending at a position x, to have a maximum between W and $W+\Delta W$ and a minimum higher than $-w$. Fig. 35 depicts $g_2(w,W;x)$ as a function of w and W, for a specific value of $x = 1$. For $W < x$ or $w < 0$ the function vanishes, since the probability of a maximum lower than the end-position or a minimum higher than the origin is zero. The function $g_2(w,W;x)$ decreases to zero with increasing W, since a RW is not likely to reach a maximum value which is much greater than its given origin and end-position. The function $g_2(w,W;x)$ becomes independent of w for large values of w, because in this limit the requirement to have a minimum higher than $-w$ does not affect the probability. This limiting probability density $g_2(w=\infty,W;x)$ can be calculated from $I(W)$ (see Eqs. B.4, B.7 and B.8):

$$g_2(w=\infty,W;x) = \frac{\partial I(W)}{\partial W} = 2(2W-x)e^{-2W(W-x)} \ . \tag{B.9}$$

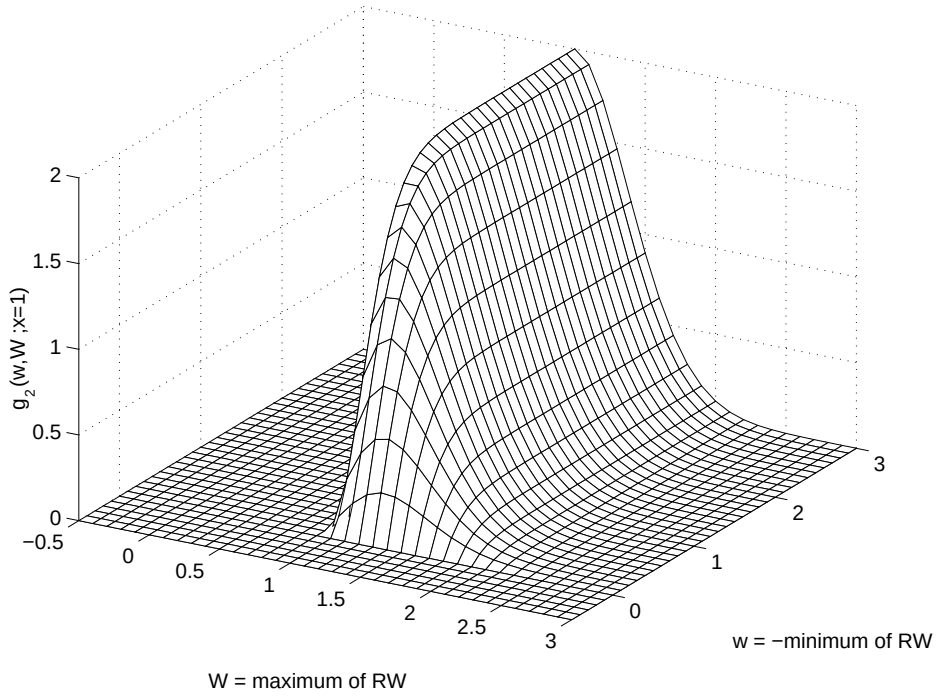

Figure 35: $g_2(w, W; x)$ *vs.* w and W. The function $g_2(w, W; x)$ is the probability density of a given RW ending at a position x to have a maximum value of W and a minimum higher than $-w$.

In order to randomly generate w according to $f_2(w|W_0)$, we numerically invert $g_2(w, W_0; x)$ for any given W_0 and x: Given W_0 and x, we calculate $g_2(w, W_0; x)$, divide it by $g_2(w = \infty, W_0; x)$ and find the value of w for which this division equals (or is closest to) J, a random number, generated from a uniform probability density between 0 and 1. The w's generated from this process are distributed according to $f_2(w|W_0)$.

Through the procedure detailed above, we can generate, for any given value of x, the minimum $(-w)$ and the maximum (W) of a RW ending at a position x, according to their mutual probability density $g(w, W; x)$.

References

[1] T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd ed. (W. H. Freeman and Company, New York, 1993).

[2] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, New York, 1979).

[3] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, New York, 1994).

[4] M. Plischke and B. Bergersen, *Equilibrium Statistical Physics*, 2nd ed. (World Scientific, Singapore, 1994).

[5] S. Windwer, In *Markov Chains and Monte Carlo Calculations in Polymer Science*, Ed. G. G. Lowry (Marcel Dekker, New York, 1970).

[6] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York, 1971).

[7] S. Caracciolo, M. S. Causo and A. Pelissetto, preprint cond-mat/9703250.

[8] C. Tanford, *Physical Chemistry of Macromolecules* (John Wiley and Sons, New York, 1961).

[9] S. F. Edwards, P. R. King and P. Pincus, Ferroelectrics **30,** 3 (1980).

[10] P. G. Higgs and J. F. Joanny, J. Chem. Phys. **94,** 1543 (1991).

[11] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon Press, New York, 1980).

[12] J. Wittmer, A. Johner and J. F. Joanny, Europhys. Lett. **24,** 263 (1993).

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

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

[15] P. Pfeuty, R. M. Velasco and P. G. de Gennes, Le J. de Phys. (Paris) **38,** L5 (1977).

[16] Y. Kantor, Phys. Rev. **A42,** 2486 (1990).

[17] Y. Kantor, H. Li and M. Kardar, Phys. Rev. Lett. **69,** 61 (1992).

[18] Y. Kantor, M. Kardar and H. Li, Phys. Rev. **E49,** 1383 (1994).

[19] D. Bratko and A. M. Chakraborty, J. Phys. Chem. **100,** 1164 (1996).

[20] Y. Kantor and M. Kardar, Europhys. Lett. **27,** 643 (1994).

[21] Y. Kantor and M. Kardar, Phys. Rev. **E51,** 1299 (1995).

[22] Y. Kantor and M. Kardar, Phys. Rev. **E52,** 835 (1995).

[23] D. Srivastava and M. Muthukumar, Macromolecules **29,** 2324 (1996).

[24] A. V. Dobrynin, M. Rubinstein and S. P. Obukhov, Macromolecules **29,** 2974 (1996).

[25] Y. Kantor and D. Ertaş, J. Phys. **A27,** L907 (1994).

[26] D. Ertaş and Y. Kantor, Phys. Rev. **E53,** 846 (1996).

[27] D. Ertaş and Y. Kantor, Phys. Rev. **E55,** 261 (1997).

[28] A. M. Gutin and E. I. Shakhnovich, Phys. Rev. **E50,** R3322 (1994).

[29] A. V. Dobrynin and M. Rubinstein, J. Phys. II France **5,** 677 (1995).

[30] H. Schiessel and A. Blumen, J. Chem. Phys. **105,** 4250 (1996).

[31] F. Brochard-Wyart, Europhys. Lett. **23,** 105 (1993).

[32] S. Karlin, *A First Course in Stochastic Processes* (Academic Press, New York, 1966).

[33] S. Chandrasekhar, Rev. Mod. Phys. **15,** 1 (1943).

[34] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. (John Wiley and Sons, New York, 1968), Vol. 1.

[35] M. Ding and W. Yang, Phys. Rev. **E52,** 207 (1995).

[36] M. E. Fisher and M. F. Sykes, Phys. Rev. **114,** 45 (1959).

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

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

[39] R. E. Bradley and S. Windwer, Phys. Rev. **E51,** 241 (1995).

[40] G. F. Lawler, Duke Math. J. **53,** 249 (1986).

[41] G. F. Lawler, J. Stat. Phys. **50,** 91 (1988).

[42] D. Dhar and A. Dhar, Phys. Rev. **E55,** 2093 (1997).

[43] B. Derrida and H. Flyvbjerg, J. Phys. **A20,** 5273 (1987).

[44] B. Derrida and H. Flyvbjerg, J. Phys. **A19,** L1003 (1986).

[45] L. Frachebourg, I. Ispolatov and P. L. Krapivsky, Phys. Rev. **E52,** R5727 (1995).

[46] B. Derrida, R. B. Griffiths and P. G. Higgs, Europhys. Lett. **18,** 361 (1992).

[47] B. Derrida and P. G. Higgs, J. Phys. **A27,** 5485 (1994).

[48] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed. (Academic Press, London, 1985).

[49] J. J. Binney, N. J. Dowrick, A. J. Fisher and M. E. J. Newman, *The Theory of Critical Phenomena - an Introduction to the Renormalization Group* (Clarendon Press, Oxford, 1992).

[50] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, 1992).

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

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

פולימרים טעונים אקראית כמהלכים אקראיים חד-מימדיים

חיבור זה מוגש כחלק מן הדרישות לקבלת תואר "מוסמך למדעים" M.Sc.

בית הספר לפיסיקה ואסטרונומיה

אוניברסיטת תל–אביב

על–ידי

שי וולפלינג

העבודה הוכנה בהדרכתו של פרופסור יעקב קנטור

דצמבר 1997