Conformations of randomly linked polymers

Yacov Kantor

School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69 978, Israel

Mehran Kardar

Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 6 February 1996)

We consider polymers in which $M$ randomly selected pairs of monomers are restricted to be in contact. Analytical arguments and numerical simulations show that an ideal (Gaussian) chain of $N$ monomers remains expanded as long as $M \ll N$, its mean squared end to end distance growing as $r^2 \propto M/N$. A possible collapse transition (to a region of order unity) is related to percolation in a one-dimensional model with long-ranged connections. A directed version of the model is also solved exactly. Based on these results, we conjecture that the typical size of a self-avoiding polymer is reduced by the links to $R \approx (N/M)^\phi$. The number of links needed to collapse a polymer in three dimensions thus scales as $N^\phi$, with $\phi \approx 0.43$. [S1063-651X(96)04009-3]

PACS number(s): 36.20.$-$r, 64.60.$-$i, 87.15.By

I. INTRODUCTION

Polymers subject to both repulsive and attractive self-interactions may have different equilibrium states depending on temperature. For example, a homogeneous self-avoiding polymer with short-range attractions between its monomers undergoes a collapse transition [1] from an expanded to a compact state as the temperature is lowered through a ‘‘θ point.’’ In the expanded state, the radius of gyration (root-mean-squared size) $R_g$ of the polymer scales as $N^\nu$, where $N$ is the number of monomers, and $\nu$ (approximately 0.588 in $d = 3$ dimensions [2]) is the swelling exponent. The compact state has a finite density, and hence $\nu \approx 1/3$. Upon collapse, contacts in the polymer (pairs of monomers located adjacent to each other in the embedding space) both increase drastically, and change qualitatively. The description of such changes constitutes another important characteristic of the collapse transition.

Statistics of such contacts is even more important in the characterization of the equilibrium states of heterogeneous polymers [3] such as biomolecules. Dynamics of protein folding is also influenced by contacts between amino acids (see, e.g., [4]). While in problems of this type the contacts between monomers are temporarily generated in the process of thermodynamic equilibrium, it is interesting to ask the inverse question of whether the configuration of a polymer can be described by specifying permanent contacts between its monomers. The issue of permanent contacts has also been extensively addressed in the context of rubber elasticity and the vulcanization process [5–8], where the typical situation is a polymer melt with permanent cross links between the polymers.

Gutin and Shakhnovich (GS) [9] analyzed the problem of a single polymer chain in which pairs of monomers are forced to remain adjacent to each other, i.e., are permanently linked. They note that in order to establish a meaningful relation between the distribution of such contacting pairs, and the behavior of real self-interacting polymers without permanent links, the choice of the ensemble of contacts must be very nonrandom: The number of random selections of contacting pairs of monomers is of the order of $N^N$, while the number of spatial configurations increases only exponentially with $N$. It is thus not possible to find a simple correspondence between the two random ensembles. GS suggest that a correlated selection of constrains is needed to generate appropriate spatial configurations.

Recently, Bryngelson and Thirumalai (BT) [10] considered a related problem in which links are randomly established between pairs of monomers on an ideal (i.e., non-self-interacting) polymer. The unconstrained ideal chain is expanded with $R_g \propto N^{1/2}$. By comparing variational estimates to the free energies of expanded and collapsed states, BT argue that increasing the number of (uncorrelated) links causes the polymer to collapse into a state in which $R_g$ is independent of $N$. In particular, they conclude that for a generic set of constraints, in which the typical distance $\ell$ (measured along the backbone) between linked monomers is of the order $N$, it suffices to have a negligible density of the constraints ($\sim 1/\ln N$) for such a collapse to occur.

In this work we consider several models of polymers with randomly linked monomers. An essential feature of all these models is that the links along the polymer are selected in an uncorrelated fashion. Unlike the previous work of BT, which concentrated on estimates of the free energy, we directly measure the spatial extent of the polymer. For ideal chains, we derive exact lower bounds which prove that uncorrelated links cannot cause the polymer to collapse. This is confirmed by extensive numerical simulations. Based on these results, we conjecture that, quite generally, the presence of $M$ random links reduces the typical size of a swollen polymer to $R \approx (N/M)^\phi$. For ideal chains, a collapse [to $R \approx O(1)$] occurs only when the number of links is of the order of $N$. However, it should be easier to collapse self-avoiding chains to a compact globular state with $R \propto N^{1/3}$.

The remainder of the paper is organized as follows. The simplest model of an ideal chain with permanent links is introduced in Sec. II. We show that calculation of the squared end to end distance is equivalent to determining the resistance of a related resistor network. This equivalence provides a powerful numerical tool that is exploited in the
The spatial extent of the polymer can be characterized by the thermal average of the squared end distance $r^2 = \langle (\vec{r}_0 - \vec{r}_N)^2 \rangle$, or by the squared radius of gyration $R_g^2$ which is the average of $r_{in}^2 = \langle (\vec{r}_{in} - \vec{r}_m)^2 \rangle$ over all pairs of monomers $\{l, m\}$. We note that the fluctuations of the polymer in each of the $d$ perpendicular space directions is independent of the others. Thus $r^2$ is simply the sum of $d$ contributions which can be calculated separately by considering one-dimensional chains. Therefore, without loss of generality we restrict ourselves to a one-dimensional Gaussian chain.

It is convenient to recast the problem in terms of a resistor network: Mathematically, calculation of $r^2$ for a Gaussian polymer with constraints is identical to finding the resistance of a chain built of elementary resistors with conductances $\kappa$, in which the pairs of sites $k_j^{(1)}, k_j^{(2)}$ are connected by zero resistance links (shorts). Thus, Fig. 1 may also represent an electrical circuit, where the dashed lines indicate the shorts. This exact correspondence [11] holds for any arbitrary Gaussian network: The dissipation on a link $ab$ is given by $\kappa_{ab}(\phi_a - \phi_b)^2$, where $\kappa_{ab}$ is the link conductance and $\phi_a$ is the potential on a node. This is analogous to the appropriate elastic term in the Gaussian Hamiltonian. The resistance between any two nodes $j$ and $k$ is calculated by minimizing the overall dissipation under applied external voltage. For a quadratic form this minimization is equivalent to calculating averages of $(\phi_j - \phi_k)^2$ with the Gaussian weight built using this form. Therefore, our task is reduced to calculating the resistance of networks such as the one in Fig. 1(b). In the remainder of the paper we shall use the terminology of the Gaussian polymer and of the resistor network interchangeably.

Note that $\kappa$, which is the force constant for the Gaussian chain or the conductance of an elementary resistor, appears only as an overall prefactor ($1/\kappa$) in calculations of $r^2$ or resistance. Thus without loss of generality we set $\kappa = 1$, making these quantities dimensionless. Calculating the resistance of a chain with a specified set of links is now accomplished using elementary methods: The configuration is first recast in the form of a simple electrical network, as in the process leading from Fig. 1(a) to Fig. 1(b). At this point each resistor is assigned a unit resistance. Pairs of resistors which are in series or parallel are replaced by effective resistors. Repeated application of this process leads to a network of not more than $M$ nodes connected by effective resistors. [For low densities of shorts, the number of nodes is much less than $M$. For example, reduction by series and resistor rules is sufficient to completely eliminate all internal nodes in Fig. 1(b).] Finally, the resistance of the reduced network is calculated by solving a system of linear equations. The number of unknowns (and equations) is of order of (or smaller than) $M$, and thus much smaller than $N$. Therefore, for each $N$ and $M$, we could easily average our numerical results over large numbers (up to 1600) of configurations with randomly distributed links.

## III. NEARBY LINKS

We start by considering links that join monomers that are close by along the chain. As a simple example consider a very long chain ($N \gg 1$) with $M$ contact pairs scattered randomly along the chain: The position $k_j^{(1)}$ of the first monomer depicted by a single circle.
mer of the \( j \)th pair is chosen with uniform probability anywhere along the chain, while the second member of the pair is located at \( k_j^{(2)} = k_j^{(1)} + \gamma \), i.e., the distances between any two members of a pair (measured along the backbone) are fixed at \( \gamma \). (We assume that \( \gamma \ll N \).)

For the corresponding electrical circuit, it is obvious that the total resistance is proportional to \( N \) (because the shorts are local). The problem is characterized by the density \( n = M/N \), and “coverage” \( c = n/\gamma \) of the links. As long as the coverage is small (\( c \ll 1 \)), the different pairs do not bridge over overlapping strands of the chain; the total resistance is obtained simply by removing the part of the chain that is shortened as \( r^2 = N - M/\gamma = (1 - c)N \approx e^{-c}N \).

When \( c \) is comparable to or larger than unity, the resistance of the chain drops significantly. However, it can still be bounded from below by the total resistance of the resistors that are not bridged by the shorts. An “unbridged resistor” is such that there are no shorts which begin to its left and end to its right. Since the probability of such a condition for each resistor is \( (1 - n)^r \), a lower bound on the resistance is given by \( (1 - n)^r N \). The continuum limit (where the discreteness of the chain can be disregarded) is reached when \( n \ll 1 \) and \( 1 \ll \gamma \ll N \). In this limit the bound becomes \( e^{-c}N \). (Note that \( c \) does not have to be small.) This lower bound shows that the chain is not collapsed for any coverage, i.e., \( r^2 \sim N \), although the prefactor may be very small when the links are dense.

The results of Monte Carlo simulations on this model are depicted in Fig. 2. In the continuum limit, the resistance is expected to have the form \( g(c)N \), where \( g \) is some unknown function. The collapse of the data for different values of \( \gamma = 8, 32, \) and 128 confirms this expectation. There are slight systematic deviations for \( \gamma = 8 \) which are due to the discreteness of the chain. The lower bound of \( e^{-c} \) is indicated by the solid line in this figure, and is quite a good estimate for small values of coverage.

There are close analogies between random resistor networks and percolation. In particular, the collapse of the Gaussian chain can be related to the presence of an infinite cluster of shorts. In the corresponding percolation problem, long-range shorts are added to a one-dimensional system [12] in an uncorrelated fashion. The results of this section were obtained for the simple case where the shorts connect points at a fixed distance \( \gamma \). However, the conclusions can be generalized to the case when \( \gamma \) is randomly distributed with a probability \( p(\gamma) \).

An important characteristic of long-range percolation is the coverage \( c = \sum_{\gamma=1}^{\infty} p(\gamma) \). As long as \( c \) is finite, the shorts do not form an infinite cluster [12]. The resistance (or \( r^2 \)) is then proportional to \( N \), with a lower bound of \( e^{-c}N \). Power-law distributions \( p(\gamma) \approx B\gamma^\alpha \) (for \( \gamma \to \infty \)) are frequently used to describe long-range percolation. For \( \alpha > 2 \) there is a finite \( c \), leading to the situation described above. For \( \alpha < 1 \), the sum \( n = \sum_{\gamma=1}^{\infty} p(\gamma) \) diverges, i.e., the number of contacts per monomer is infinite. In this case, an infinite cluster of shorts always exists, although the situation does not correspond to a realistic physical model, since \( M = N \to \infty \). A somewhat more realistic situation occurs for \( 1 < \alpha < 2 \), where it can be shown [13] that for \( n < 1/2 \) no infinite cluster is formed. For \( n > 1/2 \), depending on the details of \( p(\gamma) \) (e.g., the value of the constant \( B \)), percolation may or may not occur [14]. This again corresponds to \( M = N \).

**IV. DISTANT LINKS**

The distributions with \( \alpha < 2 \) cannot be directly used to describe the behavior of a finite polymer, because the divergence of \( c \) implies the presence of strong finite size effects. We shall, therefore, consider the extreme case of a broad distribution of \( \gamma \) by assuming that for finite \( N \) the typical \( \gamma \) is of order of \( N \). The simplest possible situation is obtained when \( k_j^{(1)} \) and \( k_j^{(2)} \) of the \( j \)th link are selected independently and uniformly among all monomers. We are interested in calculating \( r^2 \) for such a chain with \( M \) links.

We immediately notice a simple scaling argument: Consider a chain of length \( N \) with \( M \) links located at \( \{ k_j^{(1)}, k_j^{(2)} \} \), and compare it with another chain of \( \lambda N \) monomers with links at \( \{ \lambda k_j^{(1)}, \lambda k_j^{(2)} \} \). Clearly, \( r^2 \) of the latter sequence is exactly \( \lambda \) times larger than the former. (We consider the limit \( N \gg 1 \) where the discreteness effects can be disregarded.) The corresponding probabilities of finding such randomly linked chains are identical. [More precisely, the probability for a link in the first ensemble to be located between \( x \) and \( x + dx \) is equal to the probability for a link in the second ensemble to be located between \( \lambda x \) and \( \lambda(x + dx) \).] From the relation \( r_{MN}^2 = \lambda r_{MN}^2 \) it immediately follows that \( r^2 = f(M)N \). To obtain a collapsed state of the chain we must have \( f(M) \sim 1/N \). We shall show that such a small value of \( f \) is reached only when \( N \sim M \), when essentially every monomer is paired with another.

We first obtain a lower bound for \( r^2 \) in this case. The \( M \) shorts break the chain backbone into \( 2M + 1 \) segments. The resistance of the chain is certainly larger than that of the two extremal segments at its two ends. In Fig. 1, this corresponds to the distance between 0 and 1, plus the distance between 7 and 9. In the limit of large \( M \), each segment is independently taken from an exponential probability distri-
dinate of the ith monomer. The M randomly chosen monomers \( \{k_j\} \), for \( j = 1, 2, \ldots, M \) are linked to the origin, resulting in the statistical weight

\[
P(\{r_i\}) \propto \exp^{-\beta H} \prod_{j=1}^{M} \delta^d(r_k).
\]

The constraints break the chain into \( M+1 \) independent segments of lengths \( \{s_a\} \), with \( \alpha = 1, 2, \ldots, M+1 \). This model is easily solvable in the limit of large \( M \), where it is sufficient to apply the constraint \( \sum s_a = N \) only on average. Subject to this constraint, the joint probability of segment lengths is maximized for a product of independent exponential distributions

\[
p(s) = \frac{M+1}{N} \exp\left( -\frac{M+1}{N} s \right).
\]

For each segment, we thus have \( \langle s^\alpha \rangle = m! (N/M)^m \).

The end to end distance in this case is given simply by the contribution of the two end segments, and \( \langle r^2 \rangle = 2N/M \). We can also calculate an average of the squared distance from the origin

\[
R^2 = \frac{1}{N} \sum_{i=0}^{N} r_i^2 = \frac{1}{N} \sum_{a=1}^{M+1} r(s_a)^2,
\]

where we have taken advantage of the independence of segments. It is easy to show that each segment contributes \( r(s)^2 = s^2/6 \), resulting in

\[
\langle R^2 \rangle = \frac{1}{3} \frac{N}{M}.
\]

(We note that \( R^2 \) does not coincide with \( R_{\perp}^2 \) but differs from it only by a term of order \( N/M^2 \), which becomes negligible for large \( M \).) We can also consider cases where the chains are nonideal, such as directed polymers in a random medium [15]. In such cases, each segment wanders away from the origin by an amount \( s^\nu \) with \( \nu > 1/2 \). The overall radii of the randomly linked polymer are then characterized by the scale \( (N/M)^\nu \).

VI. DISCUSSIONS

Most of the results described so far apply to ideal chains. However, we may argue that some of the conclusions are expected to hold for self-avoiding polymers. Consider, for example, the lower bound obtained in Sec. IV. Even for a self-avoiding walk, it is reasonable to expect that the overall polymer size is larger than the contributions of its two end segments. Although the interactions among the segments make an exact argument difficult, it is probably reasonable to assume that each end segment has a characteristic size proportional to \( s^\nu \) with \( s \approx N/M \), where \( \nu \) is the swelling exponent. The numerical results of Sec. IV, and the analytical forms of Sec. V, suggest that this bound is generally satisfied. We thus conjecture that the typical size of a self-avoiding chain with \( M \) random links is given by
Assuming the validity of the above conjecture, we can ask how many links are necessary to compactify a chain. An ideal chain in its compact state is localized to a region of size \( R \sim O(1) \). This is achieved only with an extensive number of links \( M \approx N \). For a self-avoiding chain, however, the compact state has finite density, and hence \( R \approx N^{1/d} \). Comparing with Eq. (7) suggests that such compactification is achieved if the number of links scales as

\[
M \approx N^{\phi} \quad \text{with} \quad \phi = 1 - \frac{1}{d \nu}.
\]

For a self-avoiding polymer in \( d = 3 \), \( \phi \approx 0.43 \), and \( \phi = 1/3 \) in \( d = 2 \). (While this certainly gives the minimum number of bonds necessary for collapse, its sufficiency remains to be established.)

The above result suggests that it is much easier to compactify a self-avoiding polymer. However, it says very little about the final structure of the compact state. From the perspective of protein folding studies [9], the resulting state is most likely a compact globule; a liquidlike state with extensive entropy. Additional links would then be needed to freeze this compact object into a unique configuration. The radius of gyration is then not a good discriminator of the state of the macromolecule.

At the completion of this work, we became aware of a recent paper by Solf and Vilgis [16]. Their starting point is a randomly cross-linked Gaussian network, motivated by the Deam and Edwards [5] model of polymeric gels. Although they consider more general networks, they also perform simulations on the model of Sec. IV. They measure the radius of gyration, finding \( R_g^2 \approx 0.26N/M \), consistent with our results. Interestingly, the ratio of \( R_g^2 \) to \( r^2 \) (calculated in our paper) is approximately 1/6, as in the case of an unconstrained ideal chain.

ACKNOWLEDGMENTS

This work was supported by the Israel Science Foundation Grant No. 246/96, by the NSF through Grant No. DMR-94-00334, and the PYI program (M.K.).