Knot Probability for Self-Avoiding Loops on a Cubic Lattice

Yacov Kantor

*School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of
Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel*

Mehran Kardar

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

(Received 11 April 2004)

We investigate the probability for appearance of knots in self–avoiding loops (SALs) on a cubic lattice. A set of N–step loops is generated by attempting to combine pairs of $N/2$–step self–avoiding walks constructed by a dimerization method. We demonstrate that our method produces unbiased samples of SALs, and study the knot formation probability as a function of loop size. Our method produces knot probabilities slightly higher than those obtained by Yao *et al.*[1] using a Monte Carlo method to generate loops.

Keywords: Polymer rings, topology, knots, self-avoiding walks

1. Introduction

Knots and links naturally appear in long polymers [2], and play a prominent role in biological systems and processes[3]. Examples include chromosomes during cell division [4], knots in bacterial DNA [5], or knots in the native states of proteins[6]. It can be shown rigorously that very long self–avoiding loops are always knotted[7]. However, the theoretical proofs do not provide the frequency of the knots, or the functional dependence of their frequency on loop size. Quantitative insight into this question was first provided by a numerical study of random walks on a lattice [8]. Excluded volume effects, or self–avoiding (SA) interactions, are certainly crucial for correct description of polymers[9]; however, their incorporation into numerical studies is not simple. Earlier studies considered continuum models of self–avoiding loop (SAL) polymers with varying degrees of self–repulsion [10, 11], and demonstrated that with increasing number of monomers N, the fraction of unknotted loops decreases as e^{-N/N_o}. The characteristic size at which knots appear is surprisingly large: It increases from several hundred steps in the absence of self–avoidance, to hundreds of thousands for strongly SA polymers. Since the value of N used in typical simulations does not exceed several thousands, for SAL one can assume that the probability of an unknotted configuration simply decays exponentially. The value of N_o can then be ex-

tracted by noting that for $N \ll N_o$ the probability of the knotted configurations is $P_N \approx N/N_o$. Janse van Rensburg and Wittington [12] found $N_o \approx 1.3 \times 10^5$ on a face-centered cubic lattice in a study of SALs with $N \leq 1600$. A recent study of SALs by Yao *et al.*[1] on a cubic lattice (on which self-avoiding effects are more pronounced) with $N \leq 3000$ found $N_o \approx 2.5 \times 10^5$. Our results (obtained using a different approach for generating SALs) are similar to the latter study, although we obtain slightly larger probabilities of finding knots.

2. Loop-generating algorithms

Generating sufficiently large numbers of SALs has been the main obstacle to the study of statistics of knots in polymers. *Static* methods for creating SA polymers one at a time have the advantage of producing configurations that are independent of each other. If we do not require the two ends of a polymer to meet, several methods generate samples of "properly weighted" configurations: For lattice or for continuum models with "hard" potentials, this means that every configuration has the same weight. In the dimerization algorithm[13], an N-step SA walk (SAW) is created by generating two $(N/2)$-step SAWs and attempting to concatenate them. (If the concatenated walk is self–intersecting, it is discarded and the pair creation process is restarted.) The resulting SAWs are properly weighted, i.e. each
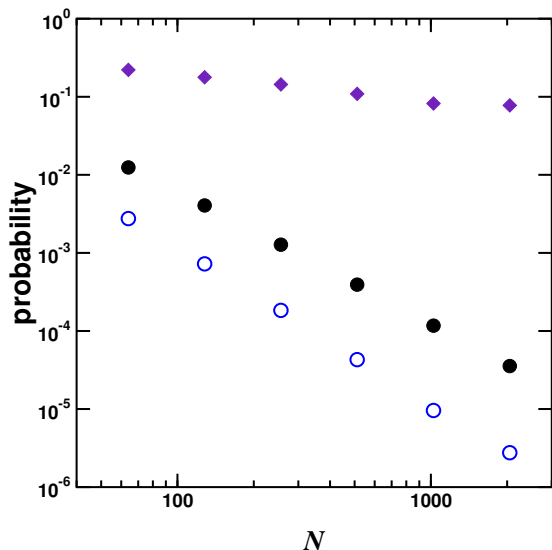
Figure 1. The probability that two $N/2$–step SAWs with the same origin terminate at the same point (full circles), and the probability of their forming a proper SAL (open circles). (These results were obtained in our simulations by including a 48–fold symmetry factor enhancement, as explained in the text.) The ratio between the latter and the former, i.e. the probability that the loop formed by the $N/2$-step pairs is self–avoiding (diamonds) as a function of N.
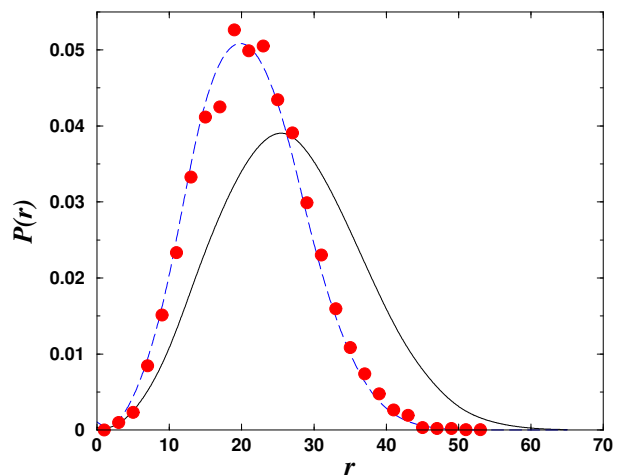

Figure 2. The solid line depicts the probability distribution for the end to end distance of a single self-avoiding walk of length 256. The dashed line corresponds to the "correct ensemble" in which two such walks accidentally meet at the same position. In Ref.[18], the former ensemble is used to generate instances of the latter, weighting the resulting configurations with the appropriate bias. The full circles are the distribution of the end-to-midpoint distance of actual, successfully generated, 512–step SALs in our dimerization procedure.

has the same weight. While this method is very efficient[15], it is not well suited for generating SALs: To produce a properly weighted $2N$–step SAL, we could first generate two N–step SAWs, assume that they both start at the origin, and check whether they do not intersect and end at the same point creating a loop, *discarding the pair if they do not form a proper SAL*. However, the probability of two SAWs in d dimensions accidentally terminating at the same point is proportional to $R^{-d} \sim N^{-d\nu}$, where R^2 is the mean squared end-to-end distance of N-step SAWs, and in $d = 3$ the swelling exponent is $\nu = 0.588$ [16]. Full circles in the Fig. 1 depict this decay with N. In addition, the probability that two such walks do not intersect each other decays as $N^{2(1-\gamma)}$, where $\gamma = 1.158$ in $d = 3$[17], as demonstrated by the diamonds in Fig. 1. Thus, successful formation of a three-dimensional SAL using this *direct* method is proportional to $N^{-2.08}$, as depicted by

open circles in Fig. 1.

To circumvent the problem of the rarity of cases where two SAWs starting from the origin terminate at the same point, Chen suggested [18] accumulating lists of SAWs. Whenever two SAWs in the list have the same end-to-end distance, they are rotated to make their endpoints coincide, thus producing a loop. Keeping the information on large numbers of SAWs is computationally memory–intensive, and this factor prevents application of the method for very long polymers. A more serious concern is that the ensemble of generated loops is *biased*: The solid line in Fig. 2 depicts the probability distribution of the end-to-end distance r for 256-step SAWs; the dashed line corresponds to the subset of cases where a pair of SAWs has the same end–point. The latter is the correct ensemble for loops, while the former is the ensemble produced when accumulated SAWs are rotated and linked. This distinction in weight

(bias) is well-known, and is properly corrected in Chen's algorithm. However, as in all bias-corrected methods, while the expectation value of a desired quantity is correctly reproduced, the variance of this quantity increases with system size. Essentially, bias–corrected methods attempt to reconstruct one distribution from a tail of another distribution; the distributions moving further apart with size. Fortunately, as can be seen from Fig. 2, for moderate values of N the distributions are not very different, and good results were obtained in Refs. [10, 11] by this method. Large memory requirements limit the use of the method to chains of moderate length.

Dynamic algorithms rely on Monte Carlo (MC) methods in which the shape of a polymer chain changes with time as a result of a possible set of deformations. In the *pivot* algorithm, a segment of the chain is chosen randomly, and one of several symmetry operations (such as rotations around pivot point(s)) is performed on it. This method was originally developed for linear polymers[14], and has proven to be extremely efficient in this context[15]. Since a randomly chosen segment is typically a significant portion of the entire chain, there is fast equilibration of large-scale properties such as the radius of gyration. The algorithm was successfully adapted to loops [19], and used to study knots [12]. Even higher efficiency was achieved by allowing fluctuations of the length of the polymer [20]. As in all MC methods, correct weights of configurations are assured in sufficiently long simulations. In practice, however, the configurations may be correlated. In particular, the correlation times for the persistence of a specific knot may significantly differ from the correlation times of large-scale geometrical features. Cognizant of this difficulty, Yao *et al.* [1] sampled the system at time intervals significantly exceeding the decay time of geometrical correlations, and verified that the times of appearance of knots behave like in a Poisson process. Our study provides an independent confirmation of the results obtained by the pivot method.

In this work we employ a *direct unbiased* approach to generating $(2N)$–step SALs. We first use successive dimerizations to generate two N– step SAWs, and then check whether (if starting from the same origin) they form a SAL. As discussed before, the probability of successful pairing of two SAWs decays as $\sim N^{-2.08}$; this includes the rejection probability that two segments

of a formed loop intersect. We find that the latter probability has only a weak dependence on the size of the loop. Indeed, the distribution of the distance between the origin and the 128th monomer in 256–step SALs, depicted by full circles in Fig. 2, practically coincides with the probability distribution (dashed line) of generating loops from two SAW segments, irrespective of their mutual intersections.

The small probability of two SAWs forming a loop can be enhanced 48-fold, by taking advantage of symmetries of the lattice to consider only SAWs whose end–point coordinates $\vec{r} = (x, y, z)$ satisfy the relations $x \geq y \geq z \geq 0$. This is achieved by generating a regular N-step SAW, and then performing the following transformations: If the coordinate x of the end–point is negative, the walk is reflected with respect to $y - z$ plane; and similarly for the other end–point coordinates. If the end–point has $x < y$ then x and y coordinates are interchanged along the whole walk; with corresponding interchanges for other pairs of axes. As a result of these transformations, we reduce the space in which the end–point can be located by a factor of 48, and thus increase the probability of forming a loop. In continuum, such reduction of space does not introduce any bias, since each SAW in the allowed subspace corresponds to 48 SAWs in the original space. This is, however, not true for SAWs ending on the boundaries of the allowed subspace; e.g. there are only 24 SAWs with $x = y > z > 0$, and consequently loops created by SAWs ending at such points are under–represented by a factor of 2. In continuum, configurations with $x = y$ have zero measure, but on discrete lattices a finite fraction of loops have this property, thus, creating a slightly biased sample. Since this is a boundary effect, it decreases with the inverse linear size of the space considered, i.e. it will be proportional to $1/R$.

3. Results and discussion

The direct method is quite efficient for generating small loops. The minimal (trefoil) knot on a cubic lattice consists of 24 steps[21], and the probability of its occurrence is extremely small. For $N = 64$, we had to examine 510 million pairs of 32–step SAWs in order to find 1.4 million SALs, out of which 27 formed a trefoil knot of the type depicted in Fig. 3. At this point the probability to have a knot is 0.000022 ± 0.000004. We used the
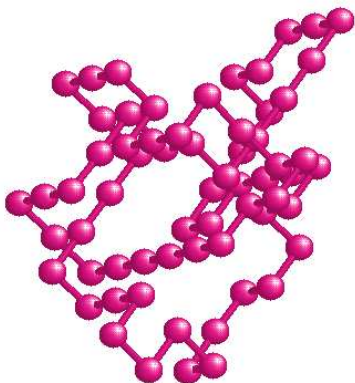
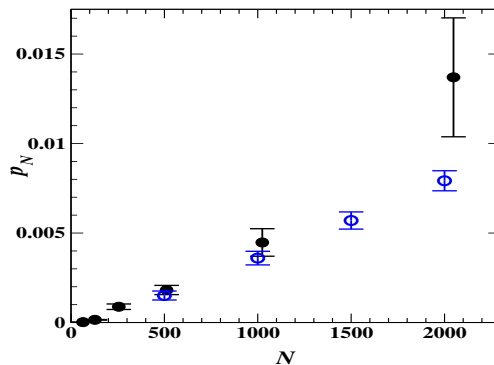Figure 3. A trefoil knot in a 64–step SAL on a cubic lattice.

Figure 4. Full circles depict the probability that an N-step SAL forms a knot. Error bars indicate one standard deviation. For comparison, open circles show the results of Ref. [1].

Alexander polynomial[22, 23] (at the value of its argument equal to -1) to determine the presence and type of a knot. Since almost all observed knots were trefoils, and very few more complicated knots were encountered, this invariant sufficed for our purposes. The limiting factor in our simulations was computer time, and we could not go beyond $N = 2048$, for which several months of CPU time were necessary to reach sufficient accuracy. Already at $N = 1024$, the probability to form SALs dropped to 10^{-5}, even after the 48–fold enhancement of the sampling, and it is obvious that this algorithm is significantly slower than pivoting or biased sampling methods. While the correction to bias in sampling symmetric configurations was important for $N = 64$, it became negligible for $N \sim 1000$.

Fig. 4 depicts our results for the probability p_N of having a knot in a SAL of length N. The curvature in the results for this range of values of N prevents a reliable extraction of N_o by a linear fit. Our last three ponts are slightly higher that the similar points in the results of Yao *et al.*[1]. In particular, result of Yao *et al.* is by 1.7 standard deviations lower than our last point. Statistically, this can occur with probability 0.1. If we take into account also discrepancies between lower points the assumption that these are accidental (statistical) differences becomes even less likely. We also note that out of the 59 knots detected for $N = 512$, only two were not trefoils, consistent with the results of Ref. [1].

In conclusion, we used the most direct method for generating an unbiased ensemble of SALs on a cubic lattice (up to symmetry factors). The method does not require large memory, but is very time consuming. While quite efficient for small and moderate sized SALs, it is limited to loops of few thosand steps, at which the probability of forming a knot is around one percent. While the efficiency of the method is significantly lower than pivot algorithms, for small and moderate Ns it can be used verify the results of other methods. Our results slightly deviate from those of Ref. [1]. In order to understand the origin of the differences larger samples, and larger N should be used, when faster computers become available.

4. Acknowledgements

References

[1] A. Yao, H. Marsuda, H. Tsukahara, M. K. Shimamura, and T. Deguchi, J. Phys. A **34**, 7563 (2001).

[2]   K. Mislow, *Introduction to Stereochemistry*, Benjamin, New York (1965); G. Schill, *Catenanes, Rotaxanes, and Knots*, Academic, New York (1971); D. M. Walba, Tetrahedron **41**, 3161 (1985); *Catenanes, Rotaxanes and Knots*, edited by J.-P. Sauvage and C. Dietrich-Buchecker, VCH, Weilheim (1999).

[3]   H. L. Frisch and E. Wassermann, J. Am. Chem. Soc. **83**, 3789 (1961); M. Delbrück, in *Mathematical Problems in Biological Sciences* (Proc. Symp. Appl. Math. **14**, 55 (1962)), edited by R. E. Bellman.

[4]   B. Alberts, K. Roberts, D. Bray, J. Lewis, M. Raff and J. D. Watson, *The Molecular Biology of the Cell* (Garland, New York, 1994).

[5]   V. V. Rybenkov N. R. Cozazarelli, A. V. Vologodskii, Proc. Natl. Acad. Sci. USA **90**, 5307 (1993); S. Y. Shaw and J. C. Wang, Science **260**, 533 (1993).

[6]   W. R. Taylor, Nature **406**, 916 (2000); R. Takusagawa and K. Kamitori, J. Am. Chem. Soc. **118**, 8945 (1996).

[7]   D. W. Sumners and S. G. Whittington, J. Phys. A **21**, 1689 (1988); N. Pippenger, Discrete Appl. Math. **25**, 273 (1989); Y. Diao, N. Pippenger and D. W. Sumners, J. Knot Theory Ramifications **3**, 419 (1994).

[8]   A. V. Vologodskiǐ, A. V. Lukashin, M. D. Frank-Kamenetskiǐ, and V. V. Anshelevich, Zh. Eksp. Teor. Fiz. **66**, 2153 (1974) [Sov. Phys. JETP **39**, 1059 (1975)]; M. D. Frank-Kamenetskii, A. V. Lukashin, and A. V. Vologodskii, Nature (London) **258**, 398 (1975).

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

[10] K. Koniaris and M. Muthukumar, Phys. Rev. Lett. **66**, 2211 (1991), and J. Chem. Phys. **95**, 2873 (1991).

[11] M. K. Shimamura and T. Deguchi, J. Phys. Soc. Japan **70**, 1523 (2001), and Phys. Lett. A **274**, 184 (2000); T. Deguchi and K. Tsurusaki, Phys. Rev. E **55**, 6245 (1997).

[12] E. J. Janse van Rensburg and S. C. Wittington, J. Phys. A **23**, 3573 (1990).

[13] K. Suzuki, Bull. Chem. Soc. Japan **41**, 538 (1968); Z. Alexandrowicz, J. Chem. Phys. **51**, 561 (1969).

[14] M. Lal, Molec. Phys. **17**, 57 (1969); B. MacDonal, D. L. Hunter, and M. O. Steinitz, J. Phys. A **18**, 2627 (1985).

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

[16] N. Madras and G. Slade, *The Self–Avoiding Walk*, Birkhäuser, Boston (1993).

[17] S. Caracciolo, M. S. Causo, and A. Pelissetto, Phys. Rev. E **57**, R1215 (1998).

[18] Y. D. Chen, J. Chem. Phys. **74**, 2034 (1981); J. Chem. Phys. **75**, 2447 and 5160 (1981).

[19] N. Madras, A. Orlitsky and L. A. Shepp, J. Stat. Phys. **58**, 159 (1990).

[20] E. J. Janse van Rensburg and S. C. Wittington, J. Phys. A **24**, 3935 (1991).

[21] Y. Diao, J. Knot Theory Ramifications **2**, 413 (1993).

[22] J. W. Alexander, Trans. Am. Math. Soc. **30**, 275 (1928).

[23] D. Rolfsen, *Knots and Links* (Publish or Perish, Berkeley, 1979); R. H. Crowell and R. H. Fox, *Introduction to Knot Theory* (Springer-Verlag, Heidelberg, 1977).