RECEIVED BY TIC MAY 8 1973

COO
CO - 2271 - 5

ENDPOINT DISTRIBUTION OF SELF-AVOIDING WALKS

Ralph Grishman

Department of Physics
Columbia University, New York, N. Y. 10027

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

MASTER

This research was supported in part by the U. S. Atomic Energy Commission.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

leg

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ABSTRACT

The dimerization algorithm for the generation of self-avoiding walks was modified to generate a large number of walks with a relatively large endpoint separation. Walks of up to 512 steps were generated on the square and cubic lattices using this modified algorithm. From these walks the distribution of endpoint x-coordinates of self-avoiding walks was accurately determined over a much larger range than would be possible using earlier techniques. The distributions obtained were found to be in general agreement with those proposed earlier for shorter walks by Domb, Gillis, and Wilmers.

Endpoint Distribution of Self-Avoiding Walks

Ralph Grishman

Department of Physics

Columbia University

Before 1960 it was generally assumed that the end-to-end distance of long polymer molecules had a Gaussian distribution, in analogy to a random walk.¹ In order to investigate this and other distributions, extensive numerical studies have been made since the 1950's using self-avoiding walks (SAW) as a simple model for polymers.² This model retains the crucial property of having interactions between chain elements which are nearby in space but separated by a long distance along the chain. In the early 1960's, several investigators reported measurements of the moments of the endpoints of randomly generated samples of SAW which were very different from the values predictable assuming a Gaussian distribution of the endpoints.^{3,4,5,6}

In a paper published in 1965,⁷ Domb, Gillis, and Wilmers reported that the distribution of endpoint x-coordinates of walks starting at the origin could be adequately described by

$$B_2 e^{-(x/\sigma)^4}$$

for a square lattice and

$$B_3 e^{-(x/\sigma)^{2.5}}$$

for a simple cubic lattice. The latter is equivalent to a radial endpoint distribution of

$$B_3 (r/\sigma)^{2.5} e^{-(r/\sigma)^{2.5}}$$

(the distribution on the square lattice could not be converted to a radial distribution in closed form). An analysis of the problem by Fisher in 1966⁸ provided strong theoretical support for the values

of the exponent found by Domb et al. Fisher's analysis was based primarily on the assumption that the endpoint distribution in d dimensions, for large n , "approaches a limiting shape

$$R_n^{-d} F(r/R_n)$$

where the scaling length R_n , which measures the end-to-end distance, varies as $r_0 n^{\nu}$, where n is the number of steps in the walk. He was then able to show that, under certain additional rather general assumptions about the distribution, the "shape factor has the form

$$F(y) = A(y) \exp(-y^{\delta})$$

where the $\delta = 1/(1 - \nu)$ and where $A(y)$ does not vary exponentially fast for large y ." Numerical studies have shown that, to a high degree of accuracy, the root-mean-square endpoint separation of SAW varies as $n^{0.75}$ and $n^{0.60}$ in 2 and 3 dimensions respectively,⁹ so Fisher's theory would predict values for δ of 4 and 2.5.

The data used by Domb et al. was limited to 18-step walks on the square lattice and 13-step walks on the cubic. Another study of the distribution was made in the same year (1965) by Mazur using 3-dimensional SAW of up to 60 steps, with nearest neighbors also excluded.¹⁰ Mazur conjectured that the radial distribution was

$$B_3 r^2 e^{-(r/\sigma)^t}$$

He measured several of the reduced moments of the endpoint distribution and compared them with the values which would be obtained from his conjectured distribution for various values of t ; in this way he arrived at a value for t of 3.2. More recently, Alexandrowicz and Accad¹¹ have generated walks of up to 8192 steps and tried to fit the endpoint distribution to a displaced Gaussian of the form

$$B_3 r^2 e^{-C_2 (r^2/n) + C_1 (r/\sqrt{n}) + C_0}$$

To the author's knowledge, this is the only paper which has attempted a fit of the distribution of long SAW; in particular, it appears that no test has been made of the distribution proposed by Domb et al. for longer walks.

The reason for this lies in the two ways in which SAW are generated: by the complete enumeration of short SAW and the Monte Carlo generation of longer SAW. Domb, Gillis, and Wilmers obtained their distributions by enumeration. Because of the exponential growth in the number of SAW as the number of steps in the walk increases, enumerative techniques have succeeded in extending the length of walks generated by only a few steps in the intervening years. In contrast, Monte Carlo programs, using the enrichment algorithm of Wall and Erpenbeck¹² or the dimerization algorithm of Alexandrowicz,¹³ have generated walks of several thousand steps. The accuracy of distributions computed by Monte Carlo methods, however, depends on the number of walks generated. This led Domb et al. to remark that "the Monte Carlo approach is not particularly suited to obtaining an explicit expression for the distribution of length since very few samples occur in the tail of the distribution. It seems better to use the Monte Carlo results to test the significance in standard statistical manner of the estimated distributions. . . . "

This paper will describe a modification to the Monte Carlo method, as suggested by Friedberg,¹⁴ which will overcome the limitation just described. Basically, this modification involves the generation of walks biased in the direction of the positive x-axis. This procedure will produce a large number of long walks, from which the tail of the original distribution may be quite accurately computed.

This modification is similar to a biasing technique described earlier by Wall, Windwer, and Gans.¹⁵ These authors suggested that,

in generating SAW on the tetrahedral lattice, trans and gauche steps be assigned different probabilities, so that more or less tightly coiled chains would be preferentially generated. This technique was used in the study of SAW with finite nearest neighbor interactions, in order to generate energetically important configurations which would occur rarely in an unbiased generation process.⁴

The following section will describe this modification in detail, and the section thereafter will present the results of a few measurements made on the generated distributions.

METHOD

The fundamental operation in the Monte Carlo generation of SAW involves selecting a single step at random (out of 4 on the square lattice and 6 on the cubic) and adding it to the walk currently being constructed. If the new vertex reached is already a part of the walk, the current walk is discarded and a new one is started; if not, the process is repeated until a walk of sufficient length has been generated. Friedberg's suggestion was to modify the procedure for choosing a single step at random so that the probability of taking a step in the $+x$ direction is greater than that of taking a step in one of the orthogonal directions by a factor e^{α} , and the probability of taking a step in the $-x$ direction is smaller by a factor e^{α} . The relative probability of walks with endpoint x_2 is thereby increased over the probability of walks with endpoint x_1 by a factor $\exp(\alpha (x_2 - x_1))$. Conversely, if the distribution with this bias is denoted $w_{\alpha}(x)$, the original distribution, $w_0(x)$, is proportional to $w_{\alpha}(x) \div \exp(\alpha x)$.

This modification can be incorporated into any of the methods for generating SAW: the simple Monte Carlo, Wall-Erpenbeck enrichment, or Alexandrowicz's dimerization. Dimerization appears to be the most efficient method of generating long walks, so it was chosen

for the current investigation. The generation of walks of a single bias was performed in three phases:

1. All walks of s steps were enumerated (where $s=8$ in 2 dimensions and $s=4$ in 3 dimensions).
2. A set of N_{2s} walks of $2s$ steps was generated by dimerization with biasing. First, each walk of s steps was assigned a weight of $\exp(\alpha x)$, where x is the end-to-end x displacement of the walk. A pair of s -step walks was then selected at random with this weighting and the two walks of a pair were joined end-to-end; if the new walk of $2s$ steps was non-self-intersecting, it was kept. This process was repeated until N_{2s} walks had been accumulated.
3. A set of N_{4s} walks of length $4s$ was generated by dimerization without biasing from walks of $2s$ steps. The procedure is identical to step 2, except that all walks of $2s$ steps are equally weighted.

Step 3 was repeated to generate walks of $8s$ steps out of $4s$ -step walks, walks of $16s$ steps out of $8s$ -step walks, and so forth until walks of l steps had been generated.

In order to check for self-intersection quickly, each lattice point is mapped into an integer between 0 and 511 by a "hashing function." To check whether a pair of walks can be joined without self-intersection, first each vertex of one walk is placed on one of 512 lists, corresponding to the value of the hashing function applied to that vertex. Each vertex of the other walk is then offset by the position of the endpoint of the first walk and the hashing function of the resultant lattice point is computed; only those points on the list corresponding to this value of the hashing function need to be checked for a match (i.e., self-intersection). If

the walks to be joined are less than 512 steps long, the list to be searched will on the average contain less than one entry; checking for self-intersection will then be very fast and will require (for the entire walk) a time approximately proportional to the length of the walk.

The program for generating walks was coded in FORTRAN IV, except for the routine for checking for intersections, which was coded in assembly language for greatest efficiency. All runs were made on an IBM 360 model 91, which executes approximately 5 million instructions per second. With no bias, the program required on one run 19803 trial pairings and 22.3 seconds to generate 3072 512-step walks out of a set of 256-step walks in two dimensions; in three dimensions 8393 trials and 22.2 seconds were required. When a bias is applied, the walks stretch out and the attrition is consequently greatly reduced. Thus with a bias of $\alpha = 0.3$ some corresponding figures are 5148 trials and 9.9 seconds in two dimensions and 4496 trials and 16.0 seconds in three dimensions.

Once some walks have been generated, the principal remaining problem lies in combining the distributions generated with different biases. If we let $w_{0,m}(x)$ be the number of m -step SAW, starting at the origin, with endpoint x -coordinate x , then the distribution of a sample of N_m walks of m steps with bias α is (neglecting random variation)

$$W_{\alpha,m}(x) = \frac{N_m}{\sum_{x=-m}^m w_{0,m}(x) \exp(\alpha x)} w_{0,m}(x) \exp(\alpha x) \quad (1)$$

Consequently, to combine distributions generated with different biases, one must compute the ratios of

$$C_{\alpha,m} = \frac{\sum_{x=-m}^m w_{0,m}(x) \exp(\alpha x)}{\sum_{x=-m}^m w_{0,m}(x) \exp(\alpha x)} \quad (2)$$

for the different biases. Two methods were tried for computing C :

1. If the difference of successive biases α and α' is small, the peaks of the biased distributions will partially overlap; a comparison of the distributions in the overlapping region can then be used to compute the ratio of $C_{\alpha'}$ to C_{α} . That is $C_{\alpha'}/C_{\alpha}$ can be computed as a weighted average of

$$\frac{W_{\alpha'}(x)}{W_{\alpha}(x)} = \frac{\exp(\alpha x)}{\exp(\alpha' x)}$$

over this region.

2. The attrition during the generation of the walks can be used to compute C_{α} directly (this method is due to Friedberg¹⁴). Since the set of shortest walks is completely enumerated, the value of $C_{\alpha,s}$ can be computed exactly. Successive values of $C_{\alpha,m}$, $m=2s, 4s, \dots$, can then be calculated using the formula

$$C_{\alpha,2m} = C_{\alpha,m}^2 S_{\alpha,2m} \quad (3)$$

where $S_{\alpha,2m}$ is the survival rate (1-attrition rate) in constructing walks of $2m$ steps out of walks of m steps, i.e., the fraction of pairs of walks of length m which are self-avoiding when joined together.

To compare these two methods, five sets of walks on the cubic lattice were generated; each set consisted of 10^4 walks of each length up to 512 steps, for each of the four biases $\alpha = 0.0, 0.1, 0.2$, and 0.3 . Values of $C_{\alpha,m}$ were calculated separately for each of the five sets of data using both methods 1 and 2. The average values and standard deviations of C obtained with the two methods were generally similar, but for some of the longer walks (where the overlap between distributions with different biases was not as great) the deviations were much larger with method 2. Method 1 was therefore selected for subsequent use.

For both methods it was observed that, very roughly, as the

length of the walk increased the fractional deviations in $C_{\alpha,m}$ increased proportionately. Thus even with method 2 the deviations were very large for the longest walks; the largest fractional standard deviation, for $C_{0.3,512}$, was 39%. The source of these large errors is evident from the first-order error propagation formula for equation (3):

$$\frac{SD(C_{\alpha,2m})}{C_{\alpha,m}} = \sqrt{4 \left[\frac{SD(C_{\alpha,m})}{C_{\alpha,m}} \right]^2 + \left[\frac{SD(S_{\alpha,2m})}{S_{\alpha,2m}} \right]^2} \quad (4)$$

Since $C_{\alpha,s}$ is computed exactly, the fractional standard deviation of

$$\frac{SD(C_{\alpha,m})}{C_{\alpha,m}} = \sqrt{\left(\frac{m}{2s}\right)^2 \left[\frac{SD(S_{\alpha,2s})}{S_{\alpha,2s}} \right]^2 + \left(\frac{m}{4s}\right)^2 \left[\frac{SD(S_{\alpha,4s})}{S_{\alpha,4s}} \right]^2 + \dots + \left[\frac{SD(S_{\alpha,m})}{S_{\alpha,m}} \right]^2} \quad (5)$$

For the trial runs, 10^4 walks of each length were generated, so the fractional standard deviation of each $S_{\alpha,m}$ would be expected to be around 1%. Equation (5) would then predict a fractional standard deviation for $C_{\alpha,512}$ of 74% (although the first-order formula is no longer accurate for such large errors).

Errors of this size in C would dominate errors in the biased distributions when computing the unbiased distribution. It is therefore important to adjust the relative number of walks of each length to be generated so as to reduce the error in C obtained with a fixed expenditure of computing time. To a first approximation, the time required to generate a SAW is proportional to its length, so the total time required to generate N_i walks of i steps, $i=2s, 4s, \dots, \ell$ is

$$T \propto 2s N_{2s} + 4s N_{4s} + \dots + \ell N_{\ell} \quad (6)$$

The typical fractional standard deviation of S_m is $N_m^{-1/2}$, so the fractional standard deviation of $C_{\alpha,m}$ is

$$\frac{SD(C_{\alpha,m})}{C_{\alpha,m}} = \sqrt{\left(\frac{m}{2s}\right)^2 \frac{1}{N_{2s}} + \left(\frac{m}{4s}\right)^2 \frac{1}{N_{4s}} + \dots + \frac{1}{N_m}} \quad (7)$$

With this approximation, $SD(C_{\alpha,m})/C_{\alpha,m}$ is minimum for fixed T when $N_i = 2\sqrt{2} N_{2i}$. For very short walks, however, the fixed computational overhead per walk becomes a major factor; in 3 dimensions, walks of 16 steps take only about 10% longer to generate than walks of 8 steps. In the limit that the time to generate a walk is independent of its length, the optimal ratio approaches $N_i = 2 N_{2i}$. The most efficient program would therefore use a ratio which gradually increased from near 2 for short walks to $2\sqrt{2}$ for long walks; however, for programming simplicity a uniform ratio of 2 was adopted for subsequent runs. This was estimated to produce an error in C only a few percent larger than the optimal program.

RESULTS

To generate the final data, five runs were made on the square and cubic lattices; each run generated 3072 (3×2^{10}) walks of 512 steps for each of the biases $\alpha = 0, 0.1, 0.2$, and 0.3 . Using the ratio $N_i = 2 N_{2i}$ just described required starting with sets of 196608 (3×2^{16}) walks of 8 steps in three dimensions and 98304 (3×2^{15}) walks of 16 steps in two dimensions. The maximum length and number of walks of that length were dictated by limitations of computer time; the walks were divided into five batches to permit some rough error estimates of the final results. The choice of biases was somewhat arbitrary, but was designed to generate walks with a large range of endpoint x -coordinates from a few different biases while providing some overlap of the peaks of the distributions of successive biases. For 512-step walks, this range of biases generated a significant number of walks through $x=260$ in two dimensions and through $x=150$ in three dimensions. This corresponds to a range of about 16 orders of magnitude in the unbiased distribution in two dimensions,

and 13 orders in three dimensions.

Each of these runs on the cubic lattice required approximately the same time as the earlier trial runs, which generated 10^4 walks of each length. In the final runs, however, the largest fractional standard deviation in $C_{\alpha,512}$ (computed using equation (3)) was 18%, less than half the earlier figure. While this may still seem quite large, it should be kept in mind that on a logarithmic scale an 18% error represents only about 1/2% of the total range of the distribution.

The one unbiased and three biased distributions generated by each run were combined into a single unbiased distribution $\bar{w}_0(x)$:

$$\bar{w}_0(x) = \frac{\sum_{\alpha} (W_{\alpha}(x) + W_{\alpha}(-x))}{\sum_{\alpha} ((\exp(\alpha x) + \exp(-\alpha x))/C_{\alpha})} \quad x \geq 0 \quad (8)$$

(data for $\pm x$ were combined because the unbiased distribution should be symmetric about $x=0$). A plot of $\bar{w}_0(x)$ was made for each lattice and length of walk. To permit a comparison with the distributions proposed by Domb, Gillis, and Wilmers,⁷ the axes chosen were $\ln \bar{w}_0(x)$ versus x^4 for the two dimensional SAW and $\ln \bar{w}_0(x)$ versus $x^{2.5}$ for the three dimensional walks. The five sets of values of $\bar{w}_0(x)$ were individually plotted in order to indicate the extent of the random variations in the data. The plots for walks of length 512 are shown in Figures 1 and 2. All of the graphs were very close to straight lines, thus lending strong support to the distributions proposed by Domb et al.

To obtain some quantitative measure of the agreement between the proposed distributions and the current data, a least-squares fit of $\ln \bar{w}_0(x)$ was made to the functional form $a + bx^{\gamma}$. The data from each run were fitted separately, so that five values of a, b , and γ were obtained for each lattice and length of walk. In making the fit, each point x was weighted by the number of walks actually

generated with that endpoint x-coordinate or one of its four nearest neighbors:¹⁶

$$\text{weight}(x) = \sum_{\alpha} \sum_{y=x-2}^{x+2} W_{\alpha}(y) + W_{\alpha}(-y)$$

The five values of γ for each lattice and length were combined to yield an average γ and 90% confidence limits;¹⁷ these values are given in Table I. In three dimensions the agreement with the hypothesis that the distribution of long walks is $B_3 e^{-(x/\sigma)^{2.5}}$ is excellent, although the exponent for the shorter walks is somewhat higher. In two dimensions, a value for γ of 4 is just outside the 90% confidence limits for the longest walks, so a distribution of precisely $B_2 e^{-(x/\sigma)^4}$ is neither completely confirmed nor rejected by the present data.

The small arrows which appear on Figures 1 and 2 indicate the importance of using biased distributions. These arrows mark the approximate range of endpoint x-coordinates over which a significant number of walks were generated by unbiased distribution.¹⁸ In one sense the arrows might be considered deceptively small, since the scales of the horizontal axes are not linear in x; the $\alpha = 0$ distribution alone extends about half as far as the combined distribution (in terms of endpoint x-coordinates) on the square lattice and about one-third as far on the cubic lattice.¹⁹ However, from the point of view of finding an analytic form for the distribution, the arrows properly indicate the very small range of values obtainable without the use of biased distributions.

FOOTNOTES

1. Barber and Ninham, Random and Restricted Walks (Gordon and Breach, 1971), p. 66.
2. C. Domb in Adv. Chem. Phys. 15 (1969); Barber and Ninham, Random and Restricted Walks (Gordon and Breach, 1971).
3. P. Verdier and W. H. Stockmayer, J. Chem. Phys. 36, 227 (1962).
4. F. Wall, S. Windwer, and P. Gans, J. Chem. Phys. 38, 2220 (1963).
5. T. F. Schatzki, J. Polymer Sci. 57, 337 (1962).
6. M. Fisher and B. Hiley, J. Chem. Phys. 34, 1253 (1961).
7. C. Domb, J. Gillis, and C. Wilmers, Proc. Phys. Soc. (London) 85, 625 (1965).
8. M. Fisher, J. Chem. Phys. 44, 616 (1966).
9. P. Gans, J. Chem. Phys. 42, 4159 (1965) and references cited therein.
Also R. Grishman, J. Chem. Phys. 58, no. 1 (1973).
10. J. Mazur, J. Chem. Phys. 43, 4354 (1965).
11. Z. Alexandrowicz and Y. Accad, J. Chem. Phys. 54, 5338 (1971).
12. F. Wall and J. Erpenbeck, J. Chem. Phys. 30, 634 (1959).
13. Z. Alexandrowicz, J. Chem. Phys. 51, 561 (1969).
14. Richard Friedberg, personal communication.
15. on next page →
16. The weight assigned to the data at x should be proportional to the reciprocal of the estimated variance of $\ln \bar{w}_0(x)$. Using the data at x alone, the best weight is therefore $\sum_{\alpha} W_{\alpha}(x) + W_{\alpha}(-x)$. Such a weighting scheme, however, assigns a greater weight to points with positive errors than those with negative errors; to minimize this effect a weighting smoothed over five points was used. The choice of weighting was in any case not critical; the several different weightings which were tried caused changes in the final value of γ of about 0.01, which is much smaller than the random errors.

17. The 90% confidence limits for a parameter are computed by taking the root-mean-square deviation of the 5 data points, dividing by $\sqrt{5}$ to get an estimate of the standard deviation of the average, then multiplying by 90% confidence limit for Student's t distribution with 4 degrees of freedom, $t_{.10}=2.132$. See for example P. Hoel, Introduction to Mathematical Statistics (Wiley, New York).
18. The limit of this range was taken to be the largest value of the x-coordinate for which at least ten walks were generated by at least one of the five runs.
19. The range, as defined in the previous footnote, extended to $x=124$ for the unbiased distribution in two dimensions, as compared with $x=266$ for the combined distribution (through $\alpha =0.3$); in three dimensions the unbiased distribution reached to $x=55$, while the combined distribution reached to $x=149$.
15. F. Wall, S. Windwer, and P. Gans, J. Chem. Phys. 37, 1461(1962).

Length of walk	γ on Square Lattice	γ on Cubic Lattice
8	---	2.72 ± 0.02
16	4.52 ± 0.03	2.64 ± 0.02
32	4.34 ± 0.04	2.54 ± 0.03
64	4.19 ± 0.05	2.52 ± 0.04
128	4.15 ± 0.04	2.50 ± 0.05
256	4.11 ± 0.09	2.49 ± 0.05
512	4.12 ± 0.11	2.48 ± 0.07

Table I. Values of γ in $\ln w_0(x) = a + bx^\gamma$, together with 90% confidence limits.

Figure Captions

Figure 1. Endpoint x-coordinate distribution of 512-step SAW on the square lattice. Arrow indicates range over which a significant number of walks were generated by the unbiased distribution (see footnote 18).

Figure 2. Endpoint x-coordinate distribution of 512-step SAW on the cubic lattice. Arrow indicates range over which a significant number of walks were generated by the unbiased distribution (see footnote 18).

FIGURE 1

FIGURE 2