Linking of uniform random polygons in confined spaces

J Arsuaga[1], T Blackstone[2], Y Diao[3], E Karadayi[4] and M Saito[4]

[1] Department of Mathematics, San Francisco State University, 1600 Holloway Ave,
San Francisco, CA 94132, USA
[2] Department of Computer Science, San Francisco State University, 1600 Holloway Ave,
San Francisco, CA 94132, USA
[3] Department of Mathematics and Statistics University of North Carolina, Charlotte 9201,
University City Blvd Charlotte, NC 28223, USA
[4] Department of Mathematics, University of South Florida, 4202 E Fowler Avenue Tampa,
FL 33620, USA

Abstract

In this paper, we study the topological entanglement of uniform random polygons in a confined space. We derive the formula for the mean squared linking number of such polygons. For a fixed simple closed curve in the confined space, we rigorously show that the linking probability between this curve and a uniform random polygon of n vertices is at least $1 - O\left(\frac{1}{\sqrt{n}}\right)$. Our numerical study also indicates that the linking probability between two uniform random polygons (in a confined space), of m and n vertices respectively, is bounded below by $1 - O\left(\frac{1}{\sqrt{mn}}\right)$. In particular, the linking probability between two uniform random polygons, both of n vertices, is bounded below by $1 - O\left(\frac{1}{n}\right)$.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Synthetic polymers and biopolymers are long thin molecules that, under appropriate experimental conditions, can be intermingled with themselves or other molecules. Quenching processes of synthetic polymer melts trap the microscopic entangling of polymer strands, and it is believed that these entanglements affect the rheological properties of the material [8, 15]. For instance semicrystalline polyethylene consists of alternating amorphous and crystalline regions. Polyethylene chains extend in and out of the crystalline regions into the amorphous regions. These chains can either penetrate into the next crystalline region or return to the region where they originated. In the latter case the chains form loops that may intermingle

with other loops forming nontrivial links [5, 6]. In the case of biopolymers, entangling of DNA in the form of knots and links provides information about biological processes of recombination [20], segregation [7] and chromosome organization [1]. Similarly, knots and links in proteins and RNA are believed to reflect functional properties of the folded chain [14, 21]. Linking of DNA chains is particularly relevant in chromosome biology. The end products of replication in Escherichia coli are two linked circles. Resolution of this topological problem is essential to ensure cell division. An extreme case of linked replication products is found in the mitochondria of trypanosomes. This organism contains a large network of circles comprising several thousands of linked DNA molecules called kinetoplasts (reviewed in [18]).

These physical and biological systems motivate our study. The study of linking between ring polymer chains dates back more than two decades ago [10], and the linking of DNA chains in free solution has also been studied previously, see for example [3, 12] and the references therein. These studies have shown that the linking probability of two chains depends on the distance between the centre of masses between the two polymer chains and for DNA chains also on the volume exclusion of the DNA molecule and on the degree of supercoiling [22].

An important factor contributing to the formation of links in physical and biological systems is the spatial confinement of molecules. This consideration is particularly relevant in biological systems where genomes are condensed up to 10^4 times to fit in the cell nucleus (reviewed in [9]). Here we address the problem of linking of polymer chains in confined volumes.

Linking of fibres in confined volumes has been previously studied for chains in the simple cubic lattice [17] but remains mostly unexplored for chains in R^3. Furthermore, the actual physical parameters, such as chain flexibility, of synthetic polymers and biopolymers confined to small volumes are mostly unknown. Here we use the uniform random polygon (URP) model. The URP model was first introduced to analyse knot distributions in confined volumes [16]. Results using the URP model are qualitatively very similar to those using other polymer models [1, 13]. Therefore, although our aim in this work is not to provide a realistic model for polymer chains in confined volumes, we believe that the URP model will yield results that are qualitatively similar to those obtained using other models. Furthermore, the URP model provides a reference system that can be theoretically checked.

First, we rigorously show that the linking probability of two chains, one with fixed length and another with variable length n, confined to a box of fixed size increases to 1 with a rate at least of $1 - O\left(\frac{1}{\sqrt{n}}\right)$. Next, we carry out extensive numerical studies of the linking probability of two confined uniform random polygons of any lengths n and m and propose that the linking probability should increase with the rate at least of $1 - O\left(\frac{1}{\sqrt{nm}}\right)$. We finish by discussing open questions and possible extensions of this work.

2. Uniform random polygons in a confined space

For the sake of simplicity, throughout this paper, we will use the unit cube $[0, 1]^3$ as our confining space. However, our approach and results will apply to other confining spaces such as the unit ball or any symmetric convex set in $\mathbb{R}^3$, modulo some constant coefficient. For $i = 1, 2, \ldots, n$, let $U_i = (u_{i1}, u_{i2}, u_{i3})$ be a three-dimensional random point that is uniformly distributed in the unit cube C^3 (or in a unit ball) such that $U_1, U_2, \ldots, U_n$ are independent. Let e_i (called the ith edge) be the line segment joining U_i and U_{i+1}, then the edges $e_1, e_2, \ldots, e_n$ define a uniform random polygon R_n in the confined space, where e_n is the line segment joining U_n and U_1. Figure 1 shows two uniform random polygons in red and green. Volumes of the polygons are included only for illustrative purposes and are not considered in this study.
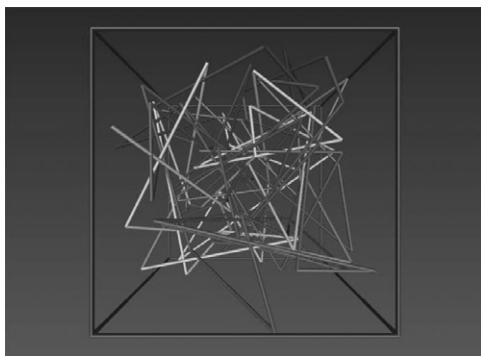
Figure 1. A uniform random link of two components confined in a cube generated by the URP model.
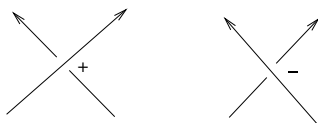

Figure 2. Assignment of ±1 at a crossing. The sign for any two edges whose relative position is as shown in the figure is determined by the right-hand rule.

Throughout this paper, uniform random polygons are always assumed to be confined in the unit cube.

Remark. The URP model is very different from the usually studied models such as the equilateral random polygons and Gaussian random polygons. Like the Gaussian random polygon model, the URP does not have a fixed bond length. However, for large n, one could think of the bond length of an R_n as the average distance d between two independent and uniform random points in the unit cube (or the confining space), since the length of R_n is close to nd (with a large probability) if n is large. The obvious advantage of the URP model is that it is very easy for simulation purposes as well as theoretical reasoning, so that simulation results can be checked using theoretical reasoning.

3. The mean squared linking number of two equal length uniform random polygons in a confined space

Use the xy-plane as the projection plane, we will need to study the projection diagram of R_n under this projection. It is not hard to show that with probability 1, this projection is a regular projection (that is, there are no crossing points of multiplicity more than 2). First, let us consider the case when there are only two (independent) oriented random edges ℓ_1 and ℓ_2. Since the end points of the edges are independent and are uniformly distributed in C^3, the probability that the projections of ℓ_1 and ℓ_2 intersect each other is a positive number, which we will call $2p$. Assume that ℓ_1 and ℓ_2 are oriented. Define a random variable ϵ in the following way: $\epsilon = 0$ if the projection of ℓ_1 and ℓ_2 have no intersection, $\epsilon = -1$ if the projection of ℓ_1 and ℓ_2 has a negative intersection and $\epsilon = 1$ if the projection of ℓ_1 and ℓ_2 has a positive intersection. For the definition of positive and negative intersections, see figure 2.

By symmetry, we then see that $P(\epsilon = 1) = P(\epsilon = -1) = p$. It follows that $E(\epsilon) = 0$ and $Var(\epsilon) = E(\epsilon^2) = 2p$.
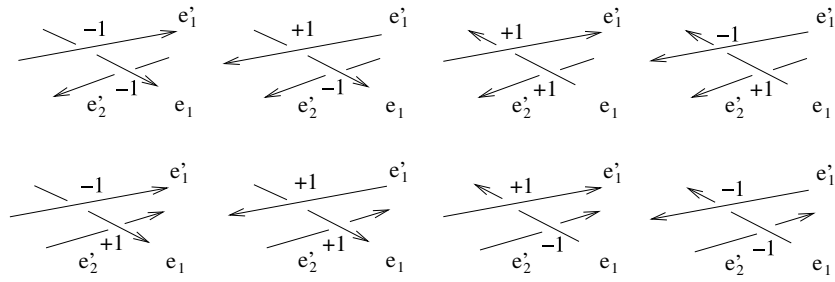
Figure 3. For each configuration of ℓ_1, ℓ'_1 and ℓ'_2 for which $\epsilon_1\epsilon_2 \neq 0$, there are eight symmetric ways of assigning the order of the vertices to them (i.e., the orientations of them). Four of them yield $\epsilon_1\epsilon_2 = -1$ and four of them lead to $\epsilon_1\epsilon_2 = 1$.

In this section, we consider the mean squared linking number of two equal length uniform random polygons in the confined space C^3. We will need the following lemma which concerns the case when there are four edges (some of them may be identical or have a common end point) involved: ℓ_1, ℓ_2, ℓ'_1 and ℓ'_2. Let ϵ_1 be the random number ϵ defined above between ℓ_1 and ℓ'_1 and let ϵ_2 be the random number defined between ℓ_2 and ℓ'_2.

Lemma 1. *(1) If the end points of ℓ_1, ℓ_2, ℓ'_1 and ℓ'_2 are distinct, then $E(\epsilon_1\epsilon_2) = 0$ (this is the case when there are eight independent random points involved);*

(2) If $\ell_1 = \ell_2$, and the end points of ℓ'_1 and ℓ'_2 are distinct (this reduces the case to where there are only three random edges, with six independent points, involved), then $E(\epsilon_1\epsilon_2) = 0$;

(3) In the case that $\ell_1 = \ell_2$ and ℓ'_1 and ℓ'_2 share a common point (so there are only five independent random points involved in this case), let $u = E(\epsilon_1\epsilon_2)$ and in the case that ℓ_1 and ℓ_2 share a common point, ℓ'_1 and ℓ'_2 also share a common point (so there are four edges defined by six independent random points involved in this case), let $E(\epsilon_1\epsilon_2) = v$. We have $q = p + 2(u + v) > 0$, where p is as defined before.

Proof. (1) This is obvious since ϵ_1 and ϵ_2 are independent random variables in this case.

(2) For each configuration in which the projections of ℓ'_1 and ℓ'_2 both intersect the projection of ℓ_1 (since otherwise $\epsilon_1\epsilon_2 = 0$), there are eight different ways to assign the orientations to the edges (see figure 3). Four of them yield $\epsilon_1\epsilon_2 = -1$ and four of them lead to $\epsilon_1\epsilon_2 = 1$. Since the joint density function of the vertices involved is simply $\frac{1}{V^6}$, where V is the volume of the confined space C^3, thus by a symmetry argument, we have $E(\epsilon_1\epsilon_2) = 0$.

(3) Consider the case when there are two random triangles involved. Orient them and name the edges of the first triangle ℓ_1, ℓ_2 and ℓ_3. Similarly, name the edges of the second triangle ℓ'_1, ℓ'_2 and ℓ'_3. Let ϵ_{ij} be the crossing sign number ϵ between the edges ℓ_i and ℓ'_j. Consider the variance of the summation $\sum_{i,j=1}^{3} \epsilon_{ij}$ (the summation indices are taken using mod (3)):

$$
V\left(\sum_{i,j=1}^{3} \epsilon_{ij}\right) = E\left(\left(\sum_{i,j=1}^{3} \epsilon_{ij}\right)^2\right)
$$

$$
= \sum_{i,j=1}^{3} E\left(\epsilon_{ij}^2\right) + 2\sum_{i=1}^{3}\left(E(\epsilon_{ij}\epsilon_{i(j-1)}) + E(\epsilon_{ij}\epsilon_{i(j+1)})\right)
$$

$$
+ 2\sum_{i=1}^{3}\left(E(\epsilon_{ij}\epsilon_{i+1,j+1}) + E(\epsilon_{ij}\epsilon_{i-1,j+1})\right).
$$

Since the ϵ_{ij}'s are identical random variables, i.e., they have the same distributions, each term in the first summation of the right-hand side in the above yields $2p$, each term in the second summation yields u (see lemma 1) and each term in the third summation yields v. There are 9 terms in the first summation and 18 terms in the second and third summation. This leads to $V\left(\sum_{i,j=1}^{3}\epsilon_{ij}\right) = 18p + 36(u+v) = 18(p+2(u+v))$. Since $V\left(\sum_{i,j=1}^{3}\epsilon_{ij}\right) > 0$, this implies that $p+2(u+v) > 0$, as claimed. □

Let us consider the case of two uniform random polygons R_1 and R_2, each of n vertices. Name the edges of R_1 and R_2 by $\ell_1, \ell_2, \ldots, \ell_n$, and $\ell_1', \ell_2', \ldots, \ell_n'$ such that this order matches the order inherited from the orientations of the polygons. As in the proof of lemma 1, let ϵ_{ij} be the crossing sign number between ℓ_i and ℓ_j'. Then the linking number between R_1 and R_2 is defined as $\frac{1}{2}\sum_{i,j=1}^{n}\epsilon_{ij}$. The linking number between two polygons is a well-known topological invariant. That is, the linking number computed from any regular projection (in which only double crossings occur) of two polygons topologically equivalent to the original ones stays the same. Applying lemma 1 to this case, we obtain the following theorem.

Theorem 1. *The mean squared linking number between two uniform random polygons R_1 and R_2 of n edges each (in the confined space C^3) is $\frac{1}{2}n^2q$ where $q = p + 2(u+v) > 0$ is as defined in lemma 1. Similar results hold if C^3 is replaced by a symmetric convex set in $\mathbb{R}^3$.*

Proof. Name the edges of R_1 and R_2 by $\ell_1, \ell_2, \ldots, \ell_n$, and $\ell_1', \ell_2', \ldots, \ell_n'$ such that this order matches the order inherited from the orientations of the polygons. Since the linking number between R_1 and R_2 is $\frac{1}{2}\sum_{i,j=1}^{n}\epsilon_{ij}$, it is easy to verify that

$$E\left(\left(\frac{1}{2}\sum_{i,j=1}^{n}\epsilon_{ij}\right)^2\right) = \frac{1}{2}n^2q.$$

□

Remark. Note that if we consider the mean squared crossing number instead, then we would get a number of the order $O(n^4)$ since one would expect that two typical uniform random polygons in a confined space to have an average of $O(n^2)$ crossings in a projection of it. In a sense, the mean squared linking number behaves more like the mean ACN of a random polygon in $\mathbb{R}^3$ without restriction. See [4].

Remark. In our numerical study we estimated that $p = 0.115 \pm 0.002$, $u = -0.0526 \pm 0.006$, $v = 0.012 \pm 0.005$ and $q = 0.0338 \pm 0.024$.

4. The linking probability of a uniform random polygon with a fixed simple closed curve in a confined space

In this section, we consider the topological entanglement of a uniform random polygon and a fixed simple closed curve in a confined space.

Before we move forward, we will state the following crucial theorem from probability theory due to Charles Stein [19].

Theorem 2. *Let $x_1, x_2, \ldots, x_n$ be a sequence of stationary and m-dependent random variables such that $E(x_i) = 0$, $E\left(x_i^2\right) < \infty$ for each i and*

$$0 < C = \lim_{n\to\infty}\frac{1}{n}E\left(\left(\sum_{i=1}^{n}x_i\right)^2\right) < \infty,$$

then $\frac{1}{\sqrt{nC}}\sum_{i=1}^{n} x_i$ *converges to the standard normal random variable. Furthermore, if we let* $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \mathrm{e}^{-\frac{x^2}{2}} \, \mathrm{d}x$ *be the distribution function of the standard normal random variable, then we have*

$$\left| P\left(\frac{1}{\sqrt{nC}} \sum_{i=1}^{n} x_i \leqslant a \right) - \Phi(a) \right| \leqslant \frac{A}{\sqrt{n}}$$

for some constant $A > 0$.

For simplicity, we will assume that the confined space is the cube given by the set $\left\{ (x, y, z) : -\frac{1}{2} \leqslant x, y, z \leqslant \frac{1}{2} \right\}$ and assume that the simple closed curve is the circle S on the xy-plane whose equation is $x^2 + y^2 = r^2$, where $r > 0$ is a constant that is less than $1/2$. As we did in the last section, we will let ϵ_j be the sum of the ± 1 assigned to the crossings between the projections of j th edge ℓ_j of R_n and S, we need to take the sum since in this case the projection of ℓ_j (onto the xy-plane) may have up to crossings with S. It is easy to see that $\epsilon_j = 0, \pm 1, \pm 2$ for each j, the ϵ_j s have the same distributions and by symmetry that we have $E(\epsilon_j) = 0$ for any j. It is obvious that if $|i - j| > 1 \mod(n)$, then ϵ_i and ϵ_j are independent, hence we have $E(\epsilon_i \epsilon_j) = 0$. By a similar argument as we did in the last section, we have $p' + 2u' > 0$, where $p' = E(\epsilon_1^2)$ and $u' = E(\epsilon_1 \epsilon_2)$. It follows that

$$0 < C = \frac{1}{n} E\left(\left(\sum_{j=1}^{n} \epsilon_j \right)^2 \right) = p' + u'$$

for any n. If we ignore the last term ϵ_n in the above, then we still have

$$0 < C = \lim_{n \to \infty} \frac{1}{n} E\left(\left(\sum_{j=1}^{n-1} \epsilon_j \right)^2 \right) = p' + u'.$$

Furthermore, it is obvious that the sequence $\epsilon_1, \epsilon_2, \ldots, \epsilon_{n-1}$ is a stationary and 2-dependent random number sequence since ϵ_j s have the same distributions, and what happens to $\epsilon_1, \ldots, \epsilon_j$ clearly do not have any affect to what happens to $\epsilon_{j+2}, \ldots, \epsilon_{n-1}$ (hence they are independent). By theorem 2, there exists a constant $A > 0$ such that

$$\left| P\left(\frac{1}{\sqrt{nC}} \sum_{i=1}^{n-1} \epsilon_i \leqslant a \right) - \Phi(a) \right| \leqslant \frac{A}{\sqrt{n}},$$

where $\Phi(a)$ is the standard normal distribution function. It follows that

$$P(\mathrm{Lk}(S, R_n) \neq 0) = P\left(\sum_{i=1}^{n} \epsilon_i \neq 0 \right)$$

$$\geqslant 1 - P\left(-a - \frac{2}{\sqrt{nC}} \leqslant \frac{1}{\sqrt{nC}} \sum_{i=1}^{n-1} \epsilon_i \leqslant a + \frac{2}{\sqrt{nC}} \right)$$

$$\geqslant 1 - \left(\Phi\left(a + \frac{2}{\sqrt{nC}} \right) - \Phi\left(-a - \frac{2}{\sqrt{nC}} \right) \right) - 2\frac{A}{\sqrt{n}},$$

where $\mathrm{Lk}(S, R_n)$ is the linking number between S and R_n and $a > 0$ is an arbitrary number (since $|\epsilon_n| \leqslant 2$). It follows that P (the linking number between S and $R_n \neq 0) \geqslant 1 - O\left(\frac{1}{\sqrt{n}} \right)$. So the linking probability between S and R_n approaches 1 at the rate of $1 - O\left(\frac{1}{\sqrt{n}} \right)$.

We list this result as the following theorem.

Theorem 3. *Let S and R_n be as defined in this section, then as n approaches infinity, the probability that S and R_n form an unsplittable link approaches 1 at the rate at least* $1 - O\left(\frac{1}{\sqrt{n}}\right)$.

Remark. Ideally, we would like to prove theorem 3 for the case of two uniform random polygons. In fact, our numerical results in section 6 strongly suggest that the linking probability between two uniform random polygons (with n and m vertices respectively) is at least of the rate $1 - O\left(\frac{1}{\sqrt{mn}}\right)$. Unfortunately, we are unable to prove this rigorously at this time. Although it seems that the convergence of $\frac{1}{n\sqrt{2q}} \sum_{i,j=1}^{n} \epsilon_{ij}$ can be determined by a simple application of the central limit theorem, it is actually not quite so simple. This is mainly because of the dependency among the random numbers ϵ_{ij} (defined in the last section), which is a major problem in proving any central limit theorem in probability theory. It is probably why theorem 2 was regarded as a big breakthrough at the time it was published. We are able to prove theorem 3 because the random variables ϵ_j turned out to be 2-dependent and stationary so we were able to apply theorem 2. But the random variables ϵ_{ij} are not m-dependent for any fixed m. One may try to separate the ϵ_{ij} s into disjoint groups and take the sum of each group in a hope that these sums turn out to be m-dependent for some fixed m so theorem 2 can be applied. Unfortunately, that does not work either. An interested reader may try this for him/herself to be convinced. Thus, a rigorous proof for theorem 3 in the case of two uniform random polygons has to wait for the development of a suitable central limit theorem that does not require m-dependency of the random variables. Proving such a theorem is a serious undertake in probability theory and is beyond the scope of this paper and the expertise of the authors.

The fixed curve S does not have to be symmetric. As long as it is placed in the cube such that the projection of it to the xy-plane is regular and that the intersection of any line segment with its projection can have at most a constant number of intersections, and that there is a positive probability that R_n can form an unsplittable link with S, the result of theorem 3 will hold. The setting in the theorem is to make the proof less complicated and tedious. We will state this as the following theorem.

Theorem 4. *If S' is a fixed simple closed curve and R'_n is a uniform random polygon of n vertices, both confined in (the interior of) a symmetric convex set of $\mathbb{R}^3$, then the probability that S' and R'_n form a unsplittable link approaches 1 at the rate at least* $1 - O\left(\frac{1}{\sqrt{n}}\right)$.

5. Numerical methods

The generation of uniform random polygons is straightforward. Each coordinate of a vertex of the uniform random polygon of n edges was drawn from a uniform distribution over [0, 1]. Although in this paper we only considered polygons confined in rectangular boxes, spheres or other convex confining volumes can be easily implemented and one should expect similar results.

We estimated the linking probability of two polygons by computing their linking number. It is known that if the linking number between two polygons is not zero, then the two polygons are non-trivially linked. In this case the two polygons are said to be homologically linked. Nevertheless there exist nontrivial links whose linking number is zero. In this case the polygons are topologically linked but are not homologically linked. It turns out that the linking number is very effective in detecting linking, as our numerical simulations later show that the probability of two uniform random polygons (of n and m vertices each) being homologically linked approaches 1 with the rate of $1 - O\left(\frac{1}{\sqrt{mn}}\right)$. This matches the observation in [17] for the case of two lattice polygons tightly confined in a box. It is quite conceivable that the probability of
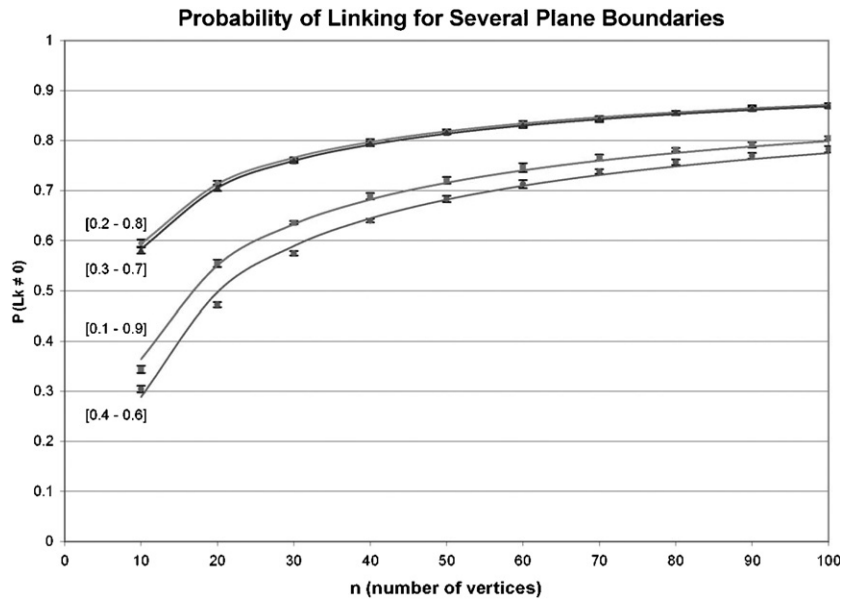
Figure 4. The plot of P(linking # $\neq$ 0). The *x*-axis represents the number of vertices of the uniform random polygon and the *y*-axis the probability P(linking # $\neq$ 0).

two uniform random polygons being topologically linked approaches 1 with a rate much faster than $1 - O\left(\frac{1}{\sqrt{mn}}\right)$, however we will not go further in that direction. For the linking number computation, we used the algorithm described in [11], which is based on the Gaussian integral form of the linking number. In our numerical studies in the following section, the sample size is set to achieve convergence. For most cases, it turns out that 50 000 is enough. For example, each of the linking number distributions shown in figure 5 for $n = 20, 40$ and 80 is based on a sample of size 50 000.

6. Numerical results

A. Our first numerical study concerns the linking between a uniform random polygon and a fixed equatorial flat curve. By theorem 3, this linking probability should grow at a rate as least as fast as $1 - O\left(\frac{1}{\sqrt{n}}\right)$. For comparison with this analytical result, we calculated the probability of having a non-zero linking number between a uniform random polygon of varying length and a fixed closed curve lying flat in the equatorial plane of the cube $E = \mathbb{R}^2 \times \left\{\frac{1}{2}\right\}$. To illustrate that the dependence of this probability on the dimensions of the curve does not affect the general trend (a claim of the theorem), we consider four squares on the plane E with different sizes, given below by their two diagonal vertices:

(a) (.1, .1, .5), (.9, .9, .5);   (b) (.2, .2, .5), (.8, .8, .5);
(c) (.3, .3, .5), (.7, .7, .5);   (d) (.4, .4, .5), (.6, .6, .5).

Results are shown in figure 4. Each curve in the figure is fitted to a curve of the form $1 - \frac{a}{\sqrt{n}}$, where a is estimated to be 2.010 ± 0.011, 1.284 ± 0.012, 1.316 ± 0.010 and 2.249 ± 0.011 corresponding to (a) through (d) respectively in that order. Clearly, while the position and size
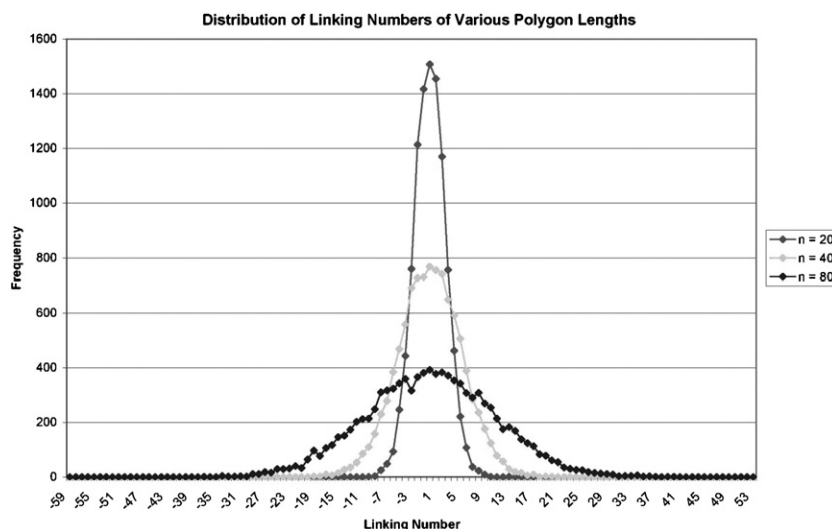
Figure 5. Distribution of linking numbers between two uniform random polygons with $n = 10, 20, 40$ and 80 vertices.

of the fixed close curve do affect the linking probability (visible for small values of n in the figure), the overall trend of the form $1 - O\left(\frac{1}{\sqrt{n}}\right)$ is strongly supported by the data.

B. Our second numerical study concerns the distribution of linking numbers between two uniform random polygons of equal length, as well as the probability that the linking number between two such uniform random polygons is not zero. Samples were drawn independently for both uniform random polygons. The plot of distribution of linking numbers is shown in figure 5 for three different values of n.

As expected, it is more likely for shorter polygons to have linking number zero. One would of course also expect that the actual linking probability for shorter polygons to be smaller as well. However, in order to determine whether two polygons are nontrivially linked when the linking number between them is zero, we would have to resolve to the use of knot polynomials, which is beyond the scope of this paper. Instead, we focused our investigation to $P(\text{linking} \# \neq 0)$, since $P(\text{linking} \# \neq 0)$ is a lower bound for the actual linking probability and we believe that $P(\text{linking} \# \neq 0)$ converges to 1 fairly fast so it does provide a good estimate for the topological linking probability. Results are shown in figure 6. We observe that $P(\text{linking} \# \neq 0)$ behaves as $1 - \frac{b}{n}$, where n is the number of vertices of the two uniform random polygons and b is estimated to be about $\sqrt{10}$.

C. The numerical result in part **B** suggests that $P(\text{linking} \# \neq 0)$ between two uniform random polygons of n vertices follows the rate $1 - \frac{b}{n}$. For two uniform random polygons with different number of vertices, this suggests that $P(\text{linking} \# \neq 0)$ should behave like $1 - \frac{b}{\sqrt{nm}}$, at least for large values of m and n, where n, m are the numbers of vertices of the two polygons respectively. We thus simulated $P(\text{linking} \# \neq 0)$ for $m, n = 10$ up to 100.

Figure 7 shows our numerical result on $P(\text{linking} \# \neq 0)$ between two uniform random polygons of different lengths (grey) as well as the 3D plot of the function $1 - \frac{b}{\sqrt{nm}}$ (red) with $b = \sqrt{10}$. The simulation result clearly suggests that $1 - \frac{b}{\sqrt{nm}}$ as a lower bound for the linking probability between two uniform random polygons, one with m vertices and the other with n
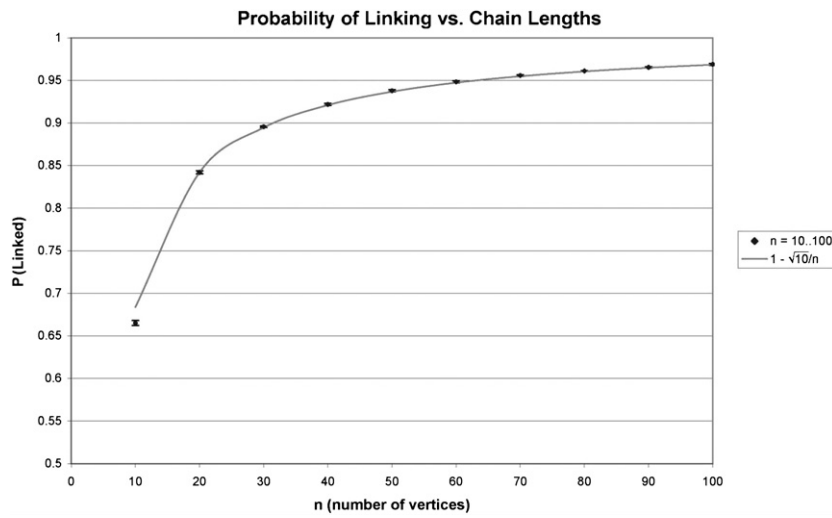
Figure 6. Probability of getting a zero linking number between two uniform random polygons. The number n of vertices considered in the simulations varied from $n = 10$ up to 100. The curve is the graph of $1 - \frac{\sqrt{10}}{n}$.
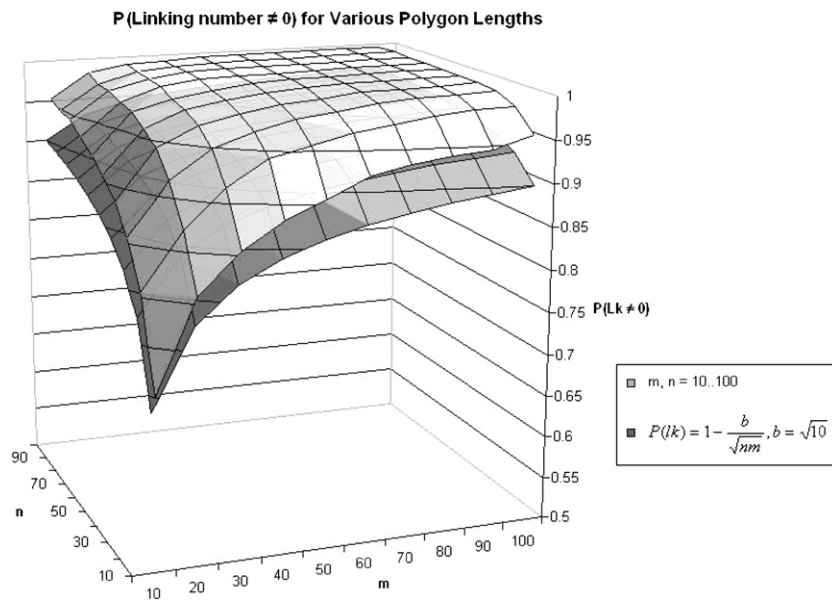

Figure 7. Three dimensional plot of the P(Linking # $\neq$ 0) between two uniform random polygons with different number of vertices. The x and y axes represent the number of vertices for the two polygons. The red surface is the lower bound estimated analytically. The grey curve was obtained by computer simulations.

vertices. Note that for larger n and m values, the difference between our numerical result and the function $1 - \frac{b}{\sqrt{nm}}$ is much smaller, suggesting that $1 - \frac{b}{\sqrt{nm}}$ is a better model for larger n, m values only. However, even when m is fixed, $1 - \frac{b}{\sqrt{nm}}$ still serves as a good lower bound
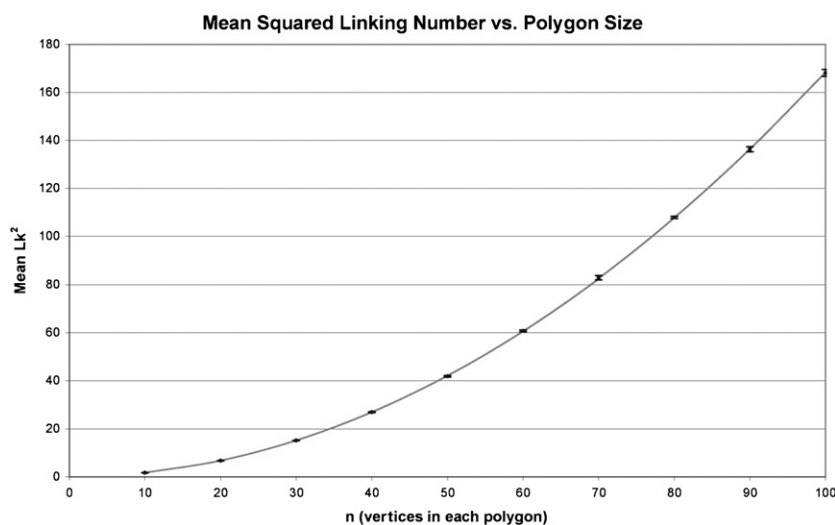
Figure 8. The mean squared linking number. Values of $\langle Lk^2 \rangle$ obtained by computer simulations are shown in blue. The continuous curve shows the theoretical value by $\frac{1}{2}qn^2$.

of the linking probability since the behaviour of $1 - \frac{b}{\sqrt{nm}}$ is the same as $1 - O(\frac{1}{\sqrt{n}})$ in this case, which coincide with our theoretical result obtained in section 4.

D. Our last numerical study confirms that the growth rate of the mean squared linking number follows the theoretically predicted growth rate $\frac{1}{2}qn^2$. We estimated that $q \approx 0.0338$. Results are shown in figure 8 where we show the $\langle Lk^2 \rangle$.

7. Conclusion

Little is known of the physical properties of polymers when confined to small volumes. In this paper, we have presented a detailed analysis of the linking probability of two curves in confined volumes under the URP model. We have given a rigorous proof of the asymptotic behaviour of the probability of linking between a curve of fixed length and a random curve. We have also investigated numerically the asymptotic behaviour of two random polygons and concluded that the rate is at least $1 - O(\frac{1}{\sqrt{nm}})$. Although we have restricted ourselves to cases where polygons are confined to boxes, our results remain true for other convex volumes. Our results can also be applied to study the particular case observed in synthetic polymers where polymer branches extend from the crystalline to the amorphous phase. This can be achieved by rooting the polygonal chains to the surfaces of the box. We intend to carry out more theoretical and numerical studies in this direction in the future.

Our results may be of special importance in biological systems where DNA molecules are confined to very small volumes. For instance, if we were to compare the volume occupied by the human genome and that of a random polygon of equal length we would find a reduction of the order of 10 000 times [9]. Furthermore, it has recently been proposed that the degree of intermingling between any chromosome with the rest of the genome during interphase may be higher than expected, reaching up to 40% of the volume of the chromosome [2]. These biological observations suggest that the linking of DNA chains may be relevant also in higher organisms. This is also an area for our future study.

Although intuition as well as our numerical results strongly suggests that the linking probability between two uniform random polygons tends to 1 as the number of the vertices of the polygons approaches infinity, a theoretical proof remains a challenge at this time and we intend to make progress towards it in our future research.

Acknowledgments

References

[1] Arsuaga J, Vazquez M, McGuirk P, Trigueros S, Sumners D and Roca J 2005 *Proc. Natl. Acad. Sci. USA*
      102 9165–9
[2] Branco M and Pombo A 2006 *Public Libr. Sci. Biol.* **4** 780–8
[3] Deguchi T and Tsurusaki K 1997 *Lectures at Knots 96* ed S Suzuki (Singapore: World Scientific) 95–122
[4] Dobay A, Dubochet J, Millett K, Sottas P E and Stasiak A 2003 *Proc. Natl. Acad. Sci. USA* **100** 5611–5
[5] Edwards S F 1967 *Proc. Phys. Soc.* **91** 513–9
[6] Edwards S F 1968 *J. Phys. A: Gen. Phys.* **1** 15–28
[7] Flammini A, Maritan A and Stasiak A 2004 *Biophys. J.* **87** 2968–75
[8] Flory J P 1953 *Principles of Polymer Chemistry* (Ithaca, NY: Cornell University Press)
[9] Holmes V and Cozzarelli N R 2000 *Proc. Natl. Acad. Sci. USA* **97** 1322–4
[10] Iwata K and Kimura T 1981 *J. Chem. Phys.* **74** 2039–48
[11] Klenin K and Langowski J 2000 *Biopolymers* **54** 307–17
[12] Klenin K V, Vologodskii A V, Anshelevich V V, Dykhne A M and Frank-Kamenetskii M D 1989 *J. Biomol.*
      Struct. Dyn. **6** 707–14
[13] Mansfield M L 1994 *Macromolecules* **27** 5924–6
[14] Mathews D H and Turner D H 2006 *Curr. Opin. Struct. Biol.* **16** 270–8
[15] Men Y, Rieger J and Strobl G 2003 *Phys. Rev. Lett.* **91** 095502–2
[16] Millett K 2000 *Knots in Hellas'98 (Delphi) (Series on Knots and Everything)* vol 24 (Singapore: World
      Scientific) 306–34
[17] Orlandini E, Janse van Rensburg E J, Tesi M C and Whittington S G 1994 *J. Phys. A: Math. Gen.* **27** 335–45
[18] Shlomai J 2004 *Curr. Mol. Med.* **4** 623–47
[19] Stein C 1972 *Proc. 6th Berkeley Symposium on Mathematical Statistics and Probability (Probability Theory)*
      vol 2 (Berkeley, CA: University California Press) 583–602
[20] Sumners D W, Ernst C, Spengler S J and Cozzarelli N R 1996 *Q. Rev. Biophys.* **28** 253–313
[21] Virnau P, Mirny A L and Kardar M 2006 *Public Libr. Sci. Comput. Biol.* **152**
[22] Vologodskii A V and Cozzarelli N R 1993 *J. Mol. Biol.* **232** 1130–40