Characteristics of shape and knotting in ideal rings

Characteristics of shape and knotting in ideal rings

Laura Zirbel and Kenneth C Millett

Department of Mathematics, University of California, Santa Barbara, CA 93016, USA

E-mail: lzirbel@math.ucsb.edu and millett@math.ucsb.edu

Abstract

We present two descriptions of the local scaling and shape of ideal rings, primarily featuring subsegments. Our focus will be the squared radius of gyration of subsegments and the squared internal end to end distance, defined to be the average squared distance between vertices k edges apart. We calculate the exact averages of these values over the space of all such ideal rings, not just a calculation of the order of these averages, and compare these to the equivalent values in open chains. This comparison will show that the structure of ideal rings is similar to that of ideal chains for only exceedingly short lengths. These results will be corroborated by numerical experiments. They will be used to analyze the convergence of our generation method and the effect of knotting on these characteristics of shape.

PACS number: 02.10.Kn

(Some figures may appear in colour only in the online journal)

1. Introduction

Long strings of connected molecules, called polymers, are central structures in life and physical sciences, as well as engineering. Prominent examples are DNA, proteins, polystyrene and silicone. With regard to DNA, Fiers and Sinsheimer first showed that the DNA of a specific virus is a single-stranded ring [8]. Because of this closure condition, DNA, like other closed polymers, can be knotted. In 1976, Liu *et al* discovered examples of knotted DNA, which was followed by the discovery of topoisomerases, enzymes which knot and unknot DNA [2, 13]. These discoveries suggest that knotting plays an important role in the behavior and shape of DNA and other polymers.

We will examine ideal rings, embedded equilateral polygons in $\mathbb{R}^3$, which provide a model for polymers under the θ-condition, where excluded volume can be ignored [3, 9, 22]. We will compare ideal rings with ideal chains, random walks in $\mathbb{R}^3$, to determine the effect of the closure constraint as well as the effect of knotting.

These models will be used to analyze the relationship between shape, scale, knotting and two physical characteristics specific to their local structure. We define the average squared

radius of gyration of subsegments, calculated by averaging the standard squared radius of gyration of each subsegment of length k, and the squared internal end to end distance, the average distance between vertices k edges apart. Rather than finding approximations of these averages, we will determine the exact theoretical averages for these descriptions of shape, as well as discuss numerical simulations and show how these characteristics are affected by knotting.

In numerical studies examining squared end to end distance, squared end to end distance is referred to internal end to end distance or two point correlation [21, 24, 26]. These characteristics are simplified in open chains because subsegments of length k in chains of length 100 are identical to subsegments of length k in chains of length 100 000: the ambient length has no effect on the behavior of subsegments. However, in the case of ideal rings, subsegments of length k in a ring of length n do not behave the same as a subsegment of length k in a ring of length $m \neq n$. Thus, subsegment behavior relies on both k and n. It will be shown that the structure of ideal rings is similar to that of ideal chains for only exceedingly short lengths, which may correspond with the difference in θ-temperature between open chain polymers and ring polymers [22]. Understanding in the case of closed chains may also allow us to use these characteristics to identify and describe the knotted portions of open chains [11, 21].

2. An introduction to ideal rings and their notation

Definition 1. *An ideal ring, P, is an n-edged equilateral polygon embedded in $\mathbb{R}^3$, with one vertex at the origin. Let $e_1, \ldots, e_n$ be unit vectors such that the kth vertex of P is $v_k = \sum_{i=1}^{k} e_i$. Let each e_i be called an edge vector of P. The fact that P is a polygon is equivalent to requiring closure, that is, that $v_n = \sum_{i=1}^{n} e_i = 0$. Let $\mathcal{P}_n$ denote the space of all such polygons.*

The careful reader may note that these ideal rings are more specific than usual. We require these polygons to be based at the origin, and be oriented. This definition makes notation concise, and does not affect the averages we will calculate, as they are independent of the base point and orientation.

For each $P \in \mathcal{P}_n$, the edge vectors e_i are all identically distributed. Therefore, when $i \neq j$, $(e_i \cdot e_j)$ is a random variable that does not depend on i and j.

We will compare these with ideal chains, a similar population but without the closure constraint.

Definition 2. *An open chain, W, also sometimes called an ideal chain or ideal open chain, is an n-edged random walk in $\mathbb{R}^3$, where each edge has unit length. Let $e_1, \ldots, e_n$ be unit vectors such that the kth vertex is given by $v_k = \sum_{i=1}^{k} e_i$. As with ideal rings, let e_i be called an edge vector. Let $\mathcal{W}_n$ denote the space of all such open chains.*

Again, the careful reader will see that these are also based at the origin for ease of notation.

3. Rigorous calculation of theoretical averages for ideal rings and ideal chains

3.1. Squared end to end distance in ideal chains

For comparison, we will first calculate the squared end to end distance in ideal chains, as in [6].

Lemma 1. *In $\mathcal{W}_n$, the average squared end to end distance is n.*

Proof. Let $W \in \mathcal{W}_n$. Calculating the end to end distance of W:

$$\left\| \sum_{i=1}^{n} e_i \right\|^2 = \left(\sum_{i=1}^{n} e_i \cdot \sum_{i=1}^{n} e_i \right)$$

$$= \sum_{i=1}^{n} (e_i \cdot e_i) + 2 \left(\sum_{i=1}^{n-1} \sum_{j=1}^{n} (e_i \cdot e_j) \right).$$

For random walks $W \in \mathcal{W}_n$, the direction of each edge vector is completely uncorrelated with the direction of the previous edge vector, and each edge vector e_i is a uniformly distributed, unit length, random vector. Thus, for $i = j$, $(e_i \cdot e_j) = 1$, and for $i \neq j$ the average value of $(e_i \cdot e_j)$, $\langle e_i \cdot e_j \rangle$, is 0. Thus we have that the average end to end distance, taken over all $W \in \mathcal{W}_n$, is

$$\left\langle \left\| \sum_{i=1}^{n} e_i \right\|^2 \right\rangle = n.$$

$\square$

3.2. Average edge product in rings

Now we will consider ideal rings. As noted before, $(e_i \cdot e_j)$ is independent of i and j. Here, we find an average value for $(e_i \cdot e_j)$, taken over all i, j for all $P \in \mathcal{P}_n$.

Definition 3. *Consider the space of ideal rings, $\mathcal{P}_n$, for some n. Let r_n denote the average of the set $\mathcal{R}_n$,*

$$\mathcal{R}_n = \{(e_i \cdot e_j) : e_i, e_j \text{ are edge vectors of some } P \in \mathcal{P}_n, i \neq j\}.$$

We will call r_n the average edge product.

The following lemma is generally known, and is foundational to the following proofs [10].

Lemma 2. *For all $n \in \mathbb{N}$, the average edge product, over all ideal rings of length n, is given by $r_n = \frac{-1}{n-1}$.*

Proof. For all $P \in \mathcal{P}_n$, $\sum_{i=1}^{n} e_i = 0$. Then squaring both sides we have

$$(0 \cdot 0) = 0 = \left(\sum_{i=1}^{n} e_i \right) \cdot \left(\sum_{i=1}^{n} e_i \right)$$

$$= \sum_{i=1}^{n} (e_i \cdot e_i) + 2 \sum_{1 \leqslant i < j \leqslant n} (e_i \cdot e_j).$$

Taking the average over all $P \in \mathcal{P}$, we replace $(e_i \cdot e_j)$ with $\langle e_i \cdot e_j \rangle = r_n$:

$$0 = n + 2 \frac{n(n-1)}{2} r_n.$$

Solving for r_n we have

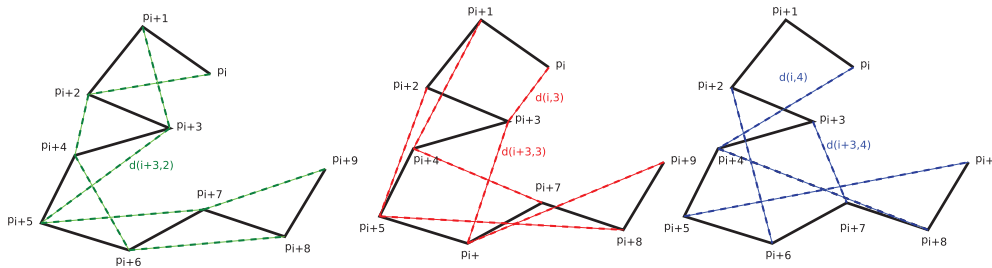$$r_n = \frac{-n}{n(n-1)} = \frac{-1}{n-1}.$$

$\square$

Figure 1. In the above image, we have end to end distances marked for $k = 2$, $k = 3$ and $k = 4$. We would take the average of their squares to find the average squared end to end distance for $k = 2, 3$ and 4.

3.3. Squared end to end distance

In open chains, the average squared end to end distance of a subsegments of length k in a chain of length n is identical to the average squared end to end distance of chains of length k. That is, the ambient chain has no effect on the shape of the subsegment. The same is not true in ideal rings, where the closure constraint plays a pivotal role in local scale, as we will see.

Definition 4. *For any $P \in \mathcal{P}_n$, define*

$$d_P(k, j) = \left\| \sum_{i=j}^{j+k} e_i \right\|^2 = \left(\sum_{i=j}^{j+k} e_i \right) \cdot \left(\sum_{i=j}^{j+k} e_i \right),$$

the squared distance between the jth and $(j + k)$th vertex. We call $d_P(k, j)$ the squared end-to-end distance of length k at j of P, as in figure 1. Define $A(k, n)$ to be the average value of

$$\{d_P(k, j) : 1 \leqslant j \leqslant n, P \in \mathcal{P}_n\}.$$

that is, let $A(k, n)$ denote the average squared end to end distance of a subsegment of length k in an ideal ring of length n. We will call $A(k, n)$ the average end to end distance of length k.

Theorem 3. $A(k, n) = \frac{k(n-k)}{n-1}$.

Proof. First we have that

$$d_P(k, j) = \left\| \sum_{i=j}^{j+k} e_i \right\|^2 = \left(\sum_{i=j}^{j+k} e_i \right) \cdot \left(\sum_{i=j}^{j+k} e_i \right)$$

$$= \sum_{i=j}^{j+k} (e_i \cdot e_i) + 2 \sum_{j \leqslant i < m \leqslant j+k} (e_i \cdot e_m)$$

$$= k + 2 \sum_{j \leqslant i < m \leqslant j+k} (e_i \cdot e_m).$$

We have shown that independent of P, k, i and j, the average value of $(e_i \cdot e_m) = \frac{-1}{n-1}$. Therefore the average value of $d_P(k, j)$ is, independent of P, i and j, $k + 2 \sum_{j \leqslant i < m \leqslant j+k} r_n$. This average
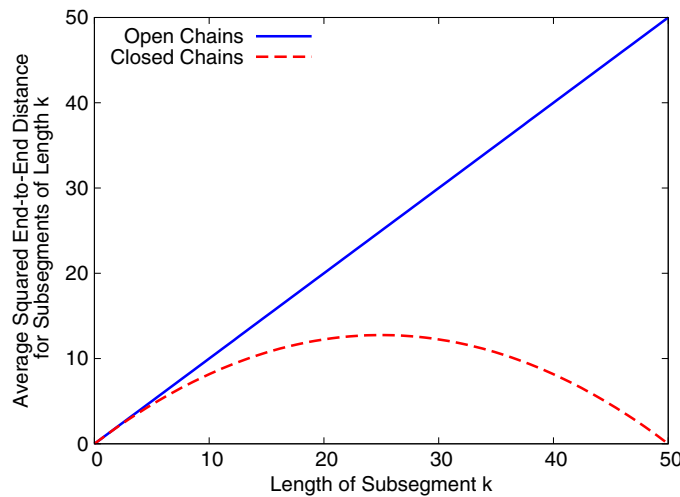
Figure 2. Here we compare the theoretical average squared end to end distance of subsegments of length k for open chains (blue) and ideal rings (red and dashed), both with total length $n = 50$.

is precisely $A(k, n)$. We conclude that

$$
\begin{aligned}
A(k, n) &= k + 2 \sum_{j \leqslant i < m \leqslant j+k} r_n \\
&= k + 2 \frac{(k-1)(k)}{2} \frac{-1}{n-1} \\
&= \frac{k(n-k)}{n-1}.
\end{aligned}
$$

Therefore the average value of $\| v_k \|^2$ is $A(k, n) = \frac{k(n-k)}{n-1}$. $\square$

We can verify some key features of this identity immediately. At $k = 1$ and $k = n - 1$, we have that the average squared end to end distance is 1. Vertices one edge apart must be distance 1 apart, as we would expect, so the theoretical average agrees with the physical reality. Also, the function is symmetric about $\frac{n}{2}$, as a segment of length k and its complement, a segment of length $n - k$, should have the same end to end distance. Lastly, the function achieves its maximum at $\frac{n}{2}$, as after that point, on average, the end to end distance must decrease due to the closure condition. Witz showed through a different argument that squared end to end distance scales like $\frac{k(n-k)}{n}$, which is close to our value [25]. This is an approximation though, as for $k = 1$, we have $\frac{n-1}{n} \neq 1$, though it is close for large values of k and n.

As in [6] and lemma 1, for open chains, the average squared end to end distance for ideal chains is n (or nb^2 where b is the length of the segments.) The squared end to end distance of subsegments of length k within an open chain of length n will be identical to the squared end to end distance of an open chain of length k. Therefore, in the open case, the average squared end to end distance of a subsegment of length k within a chain of length n is k. As we can see in figure 2, the average squared end to end distance of subsegments of an open chain and the average squared end to end distance subsegments of an ideal ring look very different for subsegments of very short lengths when $n = 50$. We observe that for the average squared end to end distance of segments of length k in open chains to be within $\frac{1}{100}$th of the average squared end to end distance of segments of length k in ideal rings, we would need $k \approx \frac{n}{100}$ or smaller. For length 50, as in figure 2, k must be length 1, which is a very strong restraint.

3.4. Average center of mass

For the following definition and lemma, it is helpful to recall that these ideal rings are based at the origin, $\sum_{i=1}^{n} e_i = 0$, as we stipulated originally.

Definition 5. *Let $n \in \mathbb{N}$. Let $P \in \mathcal{P}_n$. Define c_P, the center of mass of P, to be $c_P = \frac{1}{n} \sum_{k=1}^{n} v_k$, the average of the vertices.*

Lemma 4. *For any $P \in \mathcal{P}_n$,*

$$\| c_P \|^2 = \frac{1}{n^2} \left(\sum_{k=1}^{n} (n - k + 1)^2 + 2 \sum_{1 \leqslant i < j \leqslant n} (n - j + 1)(n - i + 1)(e_i \cdot e_j) \right).$$

Proof.

$$\| c_P \|^2 = \left(\frac{1}{n} \sum_{k=1}^{n} v_k \cdot \frac{1}{n} \sum_{k=1}^{n} v_k \right)$$

$$= \frac{1}{n^2} \left(\sum_{k=1}^{n} \sum_{i=1}^{k} e_i \cdot \sum_{k=1}^{n} \sum_{j=1}^{k} e_j \right)$$

$$= \frac{1}{n^2} \left(\sum_{k=1}^{n} (n - k + 1)^2 + 2 \sum_{1 \leqslant i < j \leqslant n} (n - j + 1)(n - i + 1)(e_i \cdot e_j) \right).$$

$\square$

Definition 6. *As above, let us denote the average of the set $\{ \| c_P \|^2 : P \in \mathcal{P}_n \}$ as $\| c_n \|^2$, the average center of mass of $\mathcal{P}_n$.*

Lemma 5. $\| c_n \|^2 = \frac{n+1}{12}$.

Proof. From lemma 4, we have that for any $P \in \mathcal{P}_n$,

$$\| c_P \|^2 = \frac{1}{n^2} \left(\sum_{k=1}^{n} (n - k + 1)^2 + 2 \sum_{1 \leqslant i < j \leqslant n} (n - j + 1)(n - i + 1)(e_i \cdot e_j) \right).$$

We will replace $(e_i \cdot e_j)$ with r_n to find $\| c_n \|^2$:

$$\| c_n \|^2 = \frac{1}{n^2} \left(\sum_{k=1}^{n} (n - k + 1)^2 + 2 \sum_{1 \leqslant i < j \leqslant n} (n - j + 1)(n - i + 1) r_n \right).$$

$$= \frac{1}{n^2} \left(\sum_{k=1}^{n} (n - k + 1)^2 \right) + \frac{2}{n^2} \sum_{j=2}^{n} \sum_{i=1}^{j-1} \left((n - j + 1)(n - i + 1) \left(\frac{-1}{n-1} \right) \right).$$

Simplifying the first term, we have

$$\frac{1}{n^2} \left(\sum_{k=1}^{n} (n - k + 1)^2 \right) = \frac{2n^2 + 3n + 1}{6n}.$$

Likewise the second term,

$$\frac{2}{n^2} \sum_{j=2}^{n} \sum_{i=1}^{j-1} \left((n - j + 1)(n - i + 1) \left(\frac{-1}{n-1} \right) \right)$$

is simplified to

$$\frac{-1}{n^2(n-1)} \sum_{j=2}^{n} (j^3 - (3n+4)j^2 + (2n^2 + 7n + 5)j - (2n^2 + 4n + 2)).$$

The fact that we are summing from $j = 2$ rather than $j = 1$ complicates this calculation. To fix this, we change the sum so that it starts at $j = 1$, and at the end, we will subtract off the $j = 1$ term:

$$1 - 3n - 4 + 2n^2 + 7n + 5 - 2n^2 - 4n - 2 = 0.$$

Because the $j = 1$ term is 0, we may change the summation to start at $j = 1$. This allows further simplification, and the second term is finally

$$-\left(\frac{3n^4 + 2n^3 - 3n^2 - 2n}{12n^2(n-1)}\right).$$

Combining these two terms, we have

$$\| c_P \|^2 = \frac{2n^2 + 3n + 1}{6n} - \frac{3n^4 + 2n^3 - 3n^2 - 2n}{12n^2(n-1)}$$

$$= \frac{(n+1)}{12}.$$

$\square$

3.5. Squared radius of gyration

With the average center of mass defined, we may ask for the average squared difference between the center of mass and the vertices. This average difference is precisely the average squared radius of gyration.

Definition 7. *Let $P \in \mathcal{P}_n$ with vertices $v_1, \ldots, v_n$. With c_P as above, define the squared radius of gyration of P to be*

$$R_G^2(P) = \frac{1}{n} \sum_{k=1}^{n} \| v_k - c_P \|^2 .$$

Definition 8. *We will define the average of the set $\{R_G^2(P) : P \in \mathcal{P}_n\}$ to be $R_{G,n}^2$, the average squared radius of gyration of $\mathcal{P}_n$.*

Theorem 6. *The average squared radius of gyration for all $P \in \mathcal{P}_n$ is $R_{G,n}^2 = \frac{n+1}{12}$.*

Proof. For any $P \in \mathcal{P}_n$, we have that

$$R_G^2(P) = \frac{1}{n} \sum_{k=1}^{n} \| v_k - c_P \|^2$$

$$= \frac{1}{n} \sum_{k=1}^{n} (\| v_k \|^2 - 2(v_k \cdot c_P) + \| c_P \|^2)$$

$$= \frac{1}{n} \sum_{k=1}^{n} \| v_k \|^2 - 2\left(\sum_{k=1}^{n} \frac{v_k}{n} \cdot c_P\right) + \frac{1}{n}(n \| c_P \|^2)$$

$$= \frac{1}{n} \sum_{k=1}^{n} \| v_k \|^2 - 2(c_P \cdot c_P) + \| c_P \|^2$$

$$= \frac{1}{n} \sum_{k=1}^{n} \| v_k \|^2 - \| c_P \|^2 .$$

From theorem 3 we have that the average value of $\| v_k \|^2$ is $\frac{k(n-k)}{n-1}$. From lemma 5 we have that the average value of $\| c_P \|^2$ is $\frac{n+1}{12}$. Replacing these, we can find the average squared radius of gyration over all $P \in \mathcal{P}_n$, that is, $R^2_{G,n}$ (rather than $R^2_G(P)$ as above)

$$R^2_{G,n} = \left(\frac{1}{n} \sum_{k=1}^{n} \frac{k(n-k)}{n-1} \right) - \frac{n+1}{12}$$

$$= \frac{n^2 - 1}{12(n-1)} = \frac{n+1}{12} = \frac{n}{12} + \frac{1}{12}.$$

$\square$

Therefore the average squared radius of gyration scales like $\frac{n}{12}$, agreeing with Zimm and Stockmayer's estimate [27].

3.6. Average squared radius of gyration of subsegments of length k

The scaling of the average squared radius of gyration of ideal rings and chains is influenced by the presence of knotting [16]. As a consequence, one expects that the average squared radius of gyration of subsegments will also be affected by the presence or absence of knotting.

Determining a correlation between knotting and shape characteristics will be affected by the nature of knotting. If the knotted sections of the rings are small, involving relatively few edges, we may not see the influence of those edges in the squared radius of gyration, which can be obscured by the behavior of the unknotted complement. In order to detect local behavior, we study characteristics of shape that are local.

In open chains, the average squared radius of gyration of subsegments of length k inside of larger chains of length n is identical to the average squared radius of gyration of segments of length k. That is, the ambient chain has no effect of the shape of the subsegment. This is not true for ideal rings, where we will see that the closure constraint plays a pivotal role in local shape. However, these characteristics may play a pivotal role in identifying and characterizing the knotted portions of open chains.

Definition 9. *Let $P \in \mathcal{P}_n$. Define $P_{i,k}$ to be the translated subsegment of P of length k beginning with the ith edge vector, e_i. That is, $P_{i,k}$ is a segment starting at the origin, where the jth vertex is given by $\sum_{m=i}^{i+j} e_m$. For ease of notation, we will relabel the position and edge vectors so that the jth vertex is given by $v'_j = \sum_{m=1}^{j} e'_m$.*

So $P_{i,k}$ is isomorphic to a subsegment of P, though we've done some relabeling. Now we would like to find the center of mass and the squared radius of gyration of $P_{i,k}$.

Definition 10. *Let $c_{P_{i,k}} = \frac{1}{k} \sum_{j=1}^{k} v'_j$, the translated center of mass of $P_{i,k}$. Then we define the squared radius of gyration of $P_{i,k}$ to be*

$$R^2_G(P_{i,k}) = \frac{1}{k} \sum_{j=1}^{k} \| v'_j - c_{P_{i,k}} \|^2 .$$

For $P \in \mathcal{P}$ and $k < n$, define the average squared radius of gyration of subsegments of length k as

$$R^2_{G,k}(P) = \frac{1}{n} \sum_{j=1}^{n} R^2_G(P_{i,k}).$$

Now, we have the analogous definitions when we take the averages of $c_{P_{i,k}}$ and $R^2_{G,k}(P)$ over all $P \in \mathcal{P}_n$.

Definition 11. *Let $c_{k,n}$ be average center of mass of a translated subsegment of length k, the average of $\{c_{P_{i,k}} : i \leqslant n \text{ and } P \in \mathcal{P}_n\}$. For some n, let $R^2_{G,n,k}$ be the average squared radius of gyration of subsegment of length k to be the average of $\{R^2_{G,k}(P) : P \in \mathcal{P}_n\}$.*

As in the last section, in order to compute $R^2_{G,n,k}$, we must first find $c_{k,n}$.

Lemma 7. $\| c_{k,n} \|^2 = \frac{(2k^2+3k+1)2n-3k(k+1)^2}{12k(n-1)}$.

Proof. Let us begin with some $\| c_{P_{m,k}} \|^2$.

$$\| c_{P_{m,k}} \|^2 = \frac{1}{k^2}\left(\sum_{j=1}^{k} v'_j \cdot \sum_{j=1}^{k} v'_j\right)$$

$$= \frac{1}{k^2}\left(\sum_{j=1}^{k}\sum_{i=1}^{j} e'_i \cdot \sum_{j=1}^{k}\sum_{i=1}^{j} e'_i\right)$$

$$= \frac{1}{k^2}\left(\sum_{i=1}^{k}(k-i+1)^2 + 2\sum_{j=2}^{k}\sum_{i=1}^{j-1}(k-j+1)(k-i+1)(e'_i \cdot e'_j)\right).$$

Because $(e'_i \cdot e'_j) = (e'_{i+m} \cdot e'_{j+m})$, the product is still independent of i and j, and we can replace $(e'_i \cdot e'_j)$ with $r_n = \frac{-1}{n-1}$. This will let us take the average over all such $P_{m,k}$ to obtain $\| c_{k,n} \|^2$:

$$\| c_{k,n} \|^2 = \frac{1}{k^2}\left(\sum_{i=1}^{k}(k-i+1)^2 + 2\sum_{j=2}^{k}\sum_{i=1}^{j-1}(k-j+1)(k-i+1)\left(\frac{-1}{n-1}\right)\right).$$

There are two sums to evaluate. The first is straightforward: $\sum_{i=1}^{k}(k-i+1)^2 = \frac{k(k+1)(2k+1)}{6}$. The second sum is $S = 2\sum_{j=2}^{k}\sum_{i=1}^{j-1}(k-j+1)(k-i+1)\left(\frac{-1}{n-1}\right)$.

$$S = \frac{-2}{n-1}\sum_{j=2}^{k}\sum_{i=1}^{j-1}(k-j+1)(k-i+1)$$

$$= \frac{-1}{n-1}\sum_{j=2}^{k}(j^3 - (3k+4)j^2 + (2k^2+7k+5)j - 2(k+1)^2).$$

Evaluating $j^3 - (3k+4)j^2 + (2k^2+7k+5)j - 2(k+1)^2$ at $j=1$ we have $1 - 3k + 4 + 2k^2 + 7k + 5 - 2k^2 - 4k - 2 = 0$. Thus we can replace the lower bound of our sum, $j=2$ with $j=1$ with no penalty.

$$S = \frac{-1}{n-1}\sum_{j=1}^{k}(j^3 - (3k+4)j^2 + (2k^2+7k+5)j - 2(k+1)^2)$$

$$= \frac{2k+3k^2-2k^3-3k^4}{12(n-1)}.$$

Combining these terms, we have

$$\| c_{k,n} \|^2 = \frac{1}{k^2}\left(\frac{k(k+1)(2k+1)}{6} + \frac{2k+3k^2-2k^3-3k^4}{12(n-1)}\right)$$

$$= \frac{(2k^2+3k+1)2n-3k(k+1)^2}{12k(n-1)}.$$

$\square$

We note that when $k = n$, we have

$$\| c_{n,n} \|^2 = \frac{(2n^2 + 3n + 1)2n - 3n(n+1)^2}{12n(n-1)}$$

$$= \frac{n^3 - n}{12n(n-1)} = \frac{n+1}{12}$$

which agrees with lemma 5.

Theorem 8. *The average squared radius of gyration of a subsegment of length k, taken over all such subsegments in all $P \in \mathcal{P}_n$ is $R^2_{G,n,k} = \frac{(k^2-1)(2n-k)}{12k(n-1)}$.*

Proof. For some $P_{i,k}$ we have

$$R^2_G(P_{i,k}) = \frac{1}{k} \sum_{j=1}^{k} \| v'_j - c_{P_{i,k}} \|^2$$

$$= \frac{1}{k} \sum_{j=1}^{k} \| v'_j \| -2 \left(\frac{1}{k} \sum_{j=1}^{k} v'_j \cdot c_{P_{i,k}} \right) + \frac{1}{k} \sum_{j=1}^{k} \| c_{P_{i,k}} \|^2$$

$$= \frac{1}{k} \sum_{j=1}^{k} \| v'_j \| -2(c_{P_{i,k}} \cdot c_{P_{i,k}}) + \frac{1}{k} k \| c_{P_{i,k}} \|^2$$

$$= \frac{1}{k} \sum_{j=1}^{k} \| v'_j \| - \| c_{P_{i,k}} \|^2 .$$

From lemma 7, $\| c_{P_{i,k}} \|^2$ is, on average, $\| c_{k,n} \|^2 = \frac{(2k^2+3k+1)2n-3k(k+1)^2}{12k(n-1)}$.

Likewise, average value of $\| v'_j \|^2$, the end to end distance of the segment $P_{i,k}$, is given in theorem 3. By replacing $\| c_{P_{i,k}} \|^2$ and $\| v'_j \|^2$ with the averages for these values, we can find $R^2_{G,n,k}$:

$$R^2_{G,n,k} = \frac{1}{k} \sum_{j=1}^{k} \frac{j(n-j)}{n-1} - \frac{(2k^2 + 3k + 1)2n - 3k(k+1)^2}{12k(n-1)}$$

$$= \frac{(k^2 - 1)(2n - k)}{12k(n-1)}.$$

$\square$

We note that when $k = n$, we have

$$R^2_{G,n,n} = \frac{(n^2 - 1)(n)}{12n(n-1)}$$

$$= \frac{n+1}{12}$$

which agrees with theorem 8.

Now for a fixed n, we have a function that returns the average squared radius gyration of a subsegment of length k.

3.7. Comparison between ideal rings and open ideal chains

For comparison, consider ideal chains. From lemma 4, which also holds for ideal chains, we know that for any $W \in \mathcal{W}_n$, the squared center of mass is given by

$$\| c_W \|^2 = \frac{1}{n^2} \left(\sum_{k=1}^{n} (n-k+1)^2 + 2 \sum_{1 \leqslant i < j \leqslant n} (n-j+1)(n-i+1)(e_i \cdot e_j) \right).$$

When the average is taken over all ideal chains $W \in \mathcal{W}_n$, define the average $c_{\mathcal{W}_n}$. Likewise, let the squared radius of gyration of the ideal chain W be $R_G^2(W)$, and let the average over all such ideal chains be $R_{G,\mathcal{W}_n}^2$.

Lemma 9. $R_{G,\mathcal{W}_n}^2 = \frac{n^2-1}{6n}$.

Proof. First, we find $\| c_{\mathcal{W}_n} \|^2$, which is simplified by that fact that in $\mathcal{W}_n$, the average value of $(e_i \cdot e_j)$ is 0 for $i \neq j$.

$$\| c_n \|^2 = \frac{1}{n^2} \left(\sum_{k=1}^{n} (n-k+1)^2 + 2 \sum_{1 \leqslant i < j \leqslant n} (n-j+1)(n-i+1)(e_i \cdot e_j) \right)$$

$$= \frac{1}{n^2} \left(\sum_{k=1}^{n} (n-k+1)^2 \right)$$

$$= \frac{2n^2 + 3n + 1}{6n}.$$

As in the first part of theorem 8, the squared radius of gyration of some open chain, $R_G^2(W)$, is

$$R_G^2(W) = \frac{1}{n} \sum_{k=1}^{n} \| v_k \|^2 - \| c_W \|^2 .$$

Replacing $\| c_W \|^2$ with the average $\| c_{\mathcal{W}_n} \|^2$ and $\| v_k \|^2$ with its average value, k, from lemma 1, we can find the average radius of gyration for open chains, $R_{G,\mathcal{W}_n}^2$:

$$R_{G,\mathcal{W}_n}^2 = \frac{1}{n} \sum_{k=1}^{n} k - \| c_{\mathcal{W}_n} \|^2$$

$$= \frac{1}{n} \left(\frac{n(n+1)}{2} \right) - \frac{2n^2 + 3n + 1}{6n}$$

$$= \frac{n^2 - 1}{6n}.$$

□

For open chains, the squared radius of gyration of a subsegment of length k is the same as for a chain of length k. So we may compare $R_{G,n,k}^2 = \frac{(k^2-1)(2n-k)}{12k(n-1)}$ and $R_{G,\mathcal{W}_n}^2 = \frac{k^2-1}{6k}$, as in figure 3.

As with squared end to end distance, we can see that the average squared radius of gyration of subsegments in an ideal ring is radically different from the average squared radius of gyration of subsegments in an open ideal chain. For ideal rings and chains of length n, in order to have the averages squared radius of gyration of a subsegment of length k for ideal rings to be within $\frac{1}{100}$ of the squared radius of gyration of a subsegment of length k in an ideal chain, we must have each other, we must have that the length of the subsegment considered, k, is less than $\frac{n}{50}$. So for $n = 1000$, k must be 20 or less.
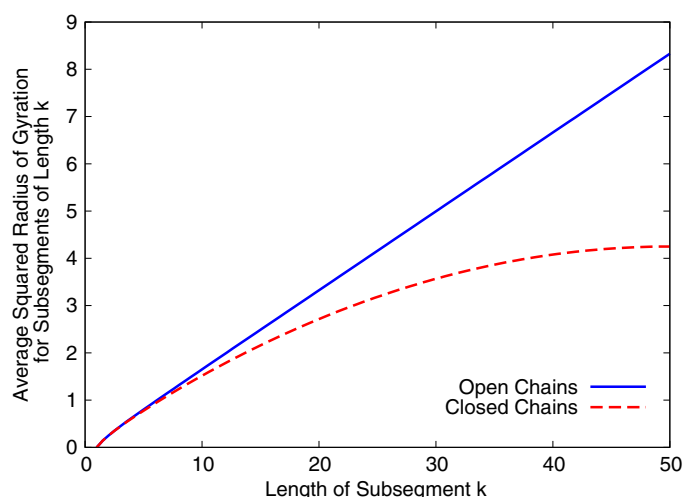
Figure 3. Here we compare the average squared radius of gyration of subsegments of length *k* for open chains (blue) and ideal rings (red and dashed).

Recall, with squared end to end distance, for the averages to be within 1% of each other, the segments had to be less than $\frac{1}{100}$th of the length, so *k* had to be length 10 or shorter if $n = 1000$. This suggests that the squared end to end distance characterization is more sensitive than squared radius of gyration of subsegments, as it is more affected by the closure condition. Once again, we observe that the local scale of an ideal ring differs significantly from that of a ideal chain at all but the smallest length scales.

4. Experimental methods

Generating a random sample from $\mathcal{P}_n$ is much trickier than generating a random walk. With the latter, we can uniformly sample points edge vectors on the unit sphere. The complication in $\mathcal{P}_n$ is the closure constraint: we cannot just generate random walks and hope that they close.

In order to randomly sample $\mathcal{P}_n$ we have two step process called the hedgehog method [1]. First, we find a starting point, then we use crankshaft rotations to sample the space. After some number of moves, our sample is independent of the starting point, and is a random element of $\mathcal{P}_n$.

4.1. Hedgehog method

The hedgehog method begins with selecting *n* points uniformly on the unit sphere, and label them $e_1, \ldots, e_n$. Then, we add to that list each e_is negative, $-e_i$. Thus we have $2n$ edge vectors, $e_1, -e_1, e_2, -e_2, \ldots, e_n, -e_n$, as in figure 4.

We randomly permute these, getting a list of edge vectors $e'_1, e'_2, \ldots, e'_{2n}$. Adding in each edge vector's negative ensures the all important closure condition is met.

4.1.1. Crankshaft rotations. This is a good starting point, but for each edge vector, its exact opposite is also an edge vector, which is undesirable.

We finish the hedgehog method by performing crankshaft rotations. We begin a crankshaft rotation by randomly selecting two non-parallel edge vectors e_j and e_k. These are rotated about
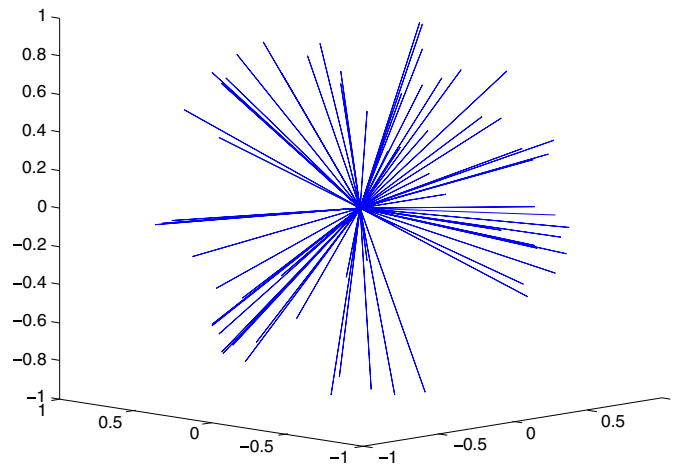
12

Figure 4. The edge vectors $e_1, -e_1, e_2, -e_2, \ldots, e_n, -e_n$ are plotted above, hence this is called the hedgehog method.

the axis determined by $e_j + e_k$, by a random angle θ. The form of these rotations is given by the following:

$$e_j \longmapsto \frac{e_j + e_k}{2} + \frac{e_j - e_k}{2} \cos(\theta) + \frac{e_j \times e_k}{\| e_j + e_k \|} \sin(\theta)$$

$$e_j \longmapsto \frac{e_k + e_j}{2} + \frac{e_k - e_j}{2} \cos(\theta) + \frac{e_k \times e_j}{\| e_k + e_j \|} \sin(\theta).$$

Because the sum $e_j + e_k$ is conserved, the modified sequence of edge vectors and vertices still satisfies the closure condition.

Because any equilateral polygon can be deformed by a finite sequence of crankshaft rotations to the regular polygon, a finite series of crankshaft rotations will take us from one polygon to another random polygon [1]. We performed $6n$ crankshaft rotations on each sample polygon.

5. Experimental results

We use the above methods to randomly generate ideal rings. Numerically, we will have two primary foci: how quickly does the average for an ensemble converge to these theoretical values, and how will knotting affect these characteristics?

5.1. Convergence

For various population sizes, 10^1, 10^2, 10^3, 10^4 and 10^5, we generated that many 50 edged polygons with 150 crankshaft rotations. For each polygon, we found the squared radius of gyration, and took the average over the population. For each population size, we did this ten times. We then compared these averages to the theoretical average. We define $E(n)$ to be the difference between the experimental average and the theoretical average with n samples. Figure 5 shows the convergence result.
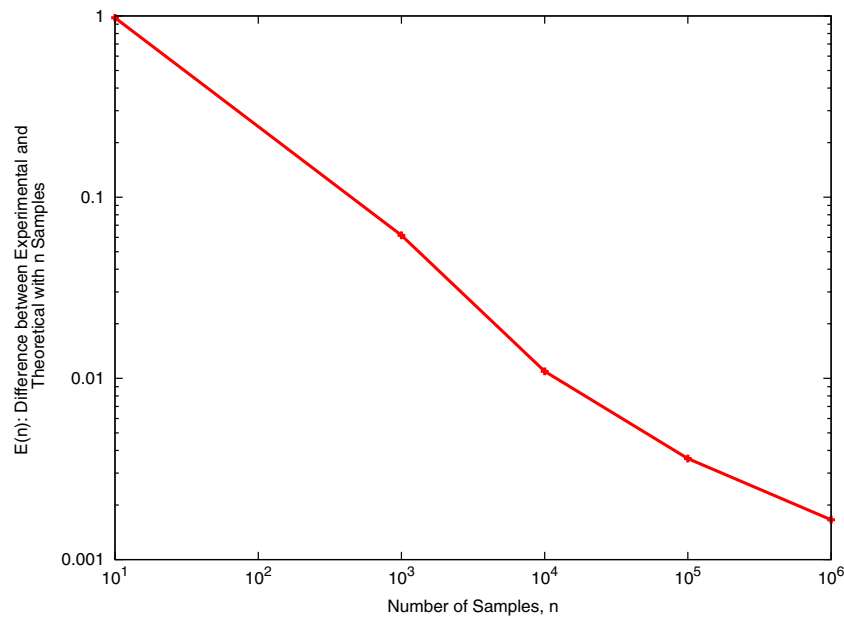
Figure 5. For population sizes 10^1, 10^2, 10^3, 10^4 and 10^5 we computed the average radius of gyration, and plotted it against the theoretical value for the same length. Note that this is a log scale plot of the data.

For population sizes of 10 000 and larger, we have excellent convergence to the theoretical average. Using linear regression, we can estimate that $E(n) \sim Cn^{-1.559}$, providing for excellent convergence for large n.

5.2. Knotting

Knotted ideal rings, specifically trefoils, were generated by first constructing ideal rings of length 50, then calculating their knot type, and saving those of the given knot type.

By the length of a knot in a ring we mean the length of the shortest subchain that contains the knot. The knot length is determined by examining subsegments of progressively longer length, starting at all possible locations. Each open segment is situated inside of a large sphere. A random collection of points on the surface of the sphere is selected. The ends of the segment are closed to each point, and the knot type of each of these closures is then calculated. This yields a spectrum of knot types, as in [14, 15, 18, 19, 23]. A segment is considered to be a trefoil if the closure is a trefoil with some tolerance (greater than 50%). The knotted portion is identified as the shortest segment which is a trefoil, for which the complement is unknotted.

In figure 6 we compare the average squared end to end distance for a phantom population of ideal rings and a population of randomly sampled trefoils. We can see that for length 50, the average for the trefoils is smaller than the average for the whole space. This suggests that for length 50, knotting compresses the polygon, making vertices closer together.

Looking at the maximum squared end to end distance, we may ask what length curve has the same maximum, 10.2291. Solving, we have that a curve of 39.89, approximately 40, has the given maximum. That suggests that the average shortening caused by knotting is about 10 edges of length, and that on average a trefoil of length 50 has an effective length of 40.
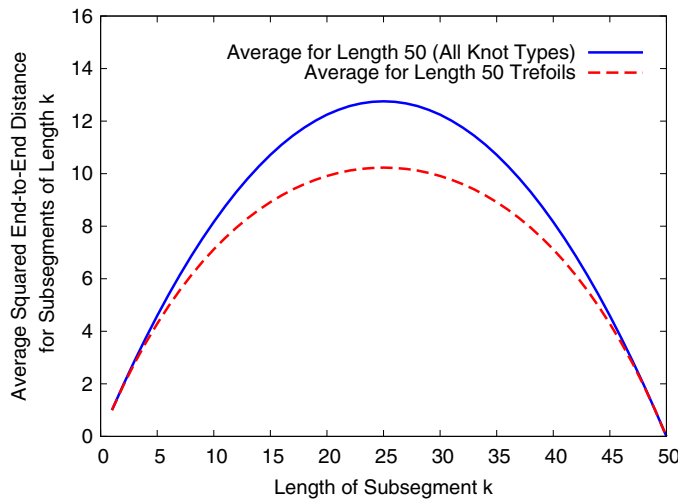
14

Figure 6. Here we compare the average squared end to end distance for a phantom population of ideal rings (blue) and a population of randomly sampled trefoils (red and dashed).
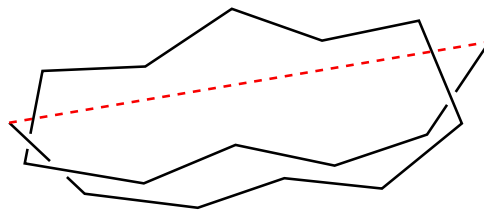

Figure 7. A hypothetical average knotted section with knot length 16 edges and end to end distance (red and dashed) of 6.

The average length of the knotted portion of these trefoils is 16.4. We predict that the difference between the length 10 reduction we saw above and the average trefoil length, 16.4, can be accounted for by examining the end to end distance of the knotted portion. As in figure 7, the difference could be explained by an average knotted portion of 16 to 17 edges with an average end to end distance around 6.

As with squared end to end distance, in figure 8 we see that the average for the trefoils is smaller than the average for the whole space. Again, this suggests that for length 50, knotting compresses the polygon, making vertices closer together. We know that ideal trefoils, and indeed any collection of some fixed knot type, will for some small number of edges have mean squared radius of gyration less than the mean for the total population, and for larger number of edges, they will have a greater average squared radius of gyration [5]. For example, the average squared radius of gyration of trefoils is smaller than the average squared radius of gyration of the whole population for lengths less than 175, and the opposite is true for lengths greater than 200 [5]. We expect similar behavior for squared end to end distance.

We can set the average squared radius of gyration of the population of trefoils, 3.5768 equal to our function for squared radius of gyration, $\frac{n+1}{12}$. Solving for n we have $n = 41.9216$, suggesting that the average shortening caused by knotting is about 8 edges of length. We can compare this with the average shortening prediction from squared end to end distance, 10. These differ by 4% of the total length, 50.
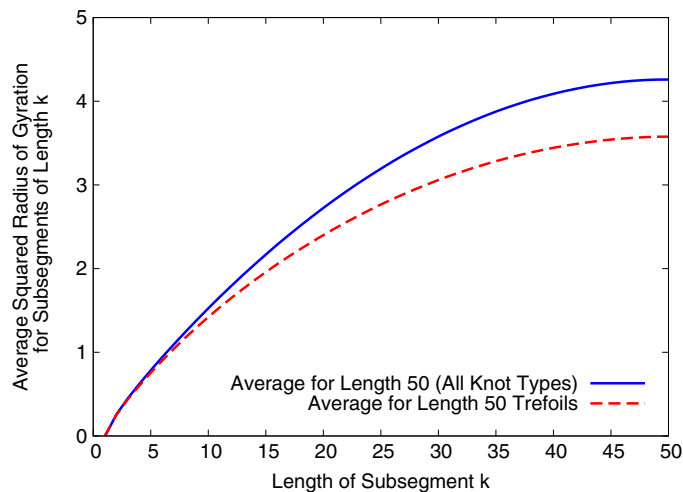
15

Figure 8. Here we compare the average squared radius of gyration of subsegments of a phantom population of ideal rings (blue) and of a population of randomly sampled trefoils (red and dashed).

6. Conclusion

These theoretical averages have many potential applications. Primarily, they can be used as a criterion to determine effectiveness of sampling methods. By comparing the squared radius of gyration of subsegments or the squared internal end to end distance of polygons generated a given sampling method and the theoretical average, we can determine the effectiveness of the generation scheme and corroborate numerical simulations. Further, in the above generation method, we can use this convergence to determine how many crankshaft rotations are needed to sample the space of polygons uniformly, as in [1].

As the previous section highlights, squared end to end distance and squared radius of gyration of subsegments may be used to predict knot length, which is computationally expensive to calculate. This will allow us to examine the growth of the knot length as $n \to \infty$, to determine if average knot length is bounded, or grows proportionally with $n^{1/2}$ or n, allowing us to ascertain if knotting is strongly local, local or global on average [4, 7, 12, 14, 15, 17, 20].

Acknowledgments

References

[1] Alvarado S, Calvo J A and Millett K C 2011 The generation of random equilateral polygons *J. Stat. Phys.* **143** 102–38
[2] Champoux J J 2001 DNA topoisomerases: structure, function and mechanism *Annu. Rev. Biochem.* **70** 369–413
[3] De Gennes P G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[4] Diao Y, Nardo J C and Sun Y 2001 Global knotting in equilateral random polygons *J. Knot Theory Ramifications* **10** 597–608

[5] Dobay A, Dubochet J, Millett K, Sottas P E and Stasiak A 2003 Scaling behavior of random knots *Proc. Natl Acad. Sci.* **100** 5611–5

[6] Doi M and Edwards S F 1988 *The Theory of Polymer Dynamics* vol 73 (Oxford: Oxford University Press)

[7] Ercolini E, Valle F, Adamcik J, Witz G, Metzler R, De Los Rios P, Roca J and Dietler G 2007 Fractal dimension and localization of DNA knots *Phys. Rev. Lett.* **98** 058102

[8] Fiers W and Sinsheimer R L 1962 The structure of the DNA of bacteriophage phi x174: III. Ultracentrifugal evidence for a ring structure *J. Mol. Biol.* **5** 424–34

[9] Flory P J 1953 *Principles of Polymer Chemistry* (Ithaca, NY: Cornell University Press)

[10] Grosberg A Y 2008 Total curvature and total torsion of a freely jointed circular polymer with $n \gg 1$ segments *Macromolecules* **41** 4524–7

[11] Kanaeda N and Deguchi T 2008 Universality in the diffusion of knots arXiv:0807.0304

[12] Katritch V, Olson W K, Vologodskii A, Dubochet J and Stasiak A 2000 Tightness of random knotting *Phys. Rev.* E **61** 5545–9

[13] Liu L F, Depew R E and Wang J C 1976 Knotted single-stranded DNA rings: a novel topological isomer of circular single-stranded DNA formed by treatment with *Escherichia coli* w protein *J. Mol. Biol.* **106** 439–52

[14] Marcone B, Orlandini E, Stella A L and Zonta F 2005 What is the length of a knot in a polymer? *J. Phys. A: Math. Gen.* **38** L15–21

[15] Marcone B, Orlandini E, Stella A L and Zonta F 2007 Size of knots in ring polymers *Phys. Rev.* E **75** 041105

[16] Matsuda H, Yao A, Tsukahara H, Deguchi T, Furuta K and Inami T 2003 Average size of random polygons with fixed knot topology *Phys. Rev.* E **68** 011102

[17] Micheletti C, Marenduzzo D and Orlandini E 2011 Polymers with spatial or topological constraints: theoretical and computational results *Phys. Rep.* **504** 1–73

[18] Millett K, Dobay A and Stasiak A 2005 Linear random knots and their scaling behavior *Macromolecules* **38** 601–6

[19] Millett K C and Sheldon B M 2005 Tying down open knots: a statistical method for identifying open knots with applications to proteins *Physical and Numerical Models in Knot Theory* (Singapore: World Scientific) pp 203–17

[20] Nakajima C H and Sakaue T 2011 Knot localization in a flexible polymer confined in a tube *Prog. Theor. Phys. Suppl.* **191** 205–9

[21] Shimamura M K, Kamata K, Yao A and Deguchi T 2005 Scattering functions of knotted ring polymers *Phys. Rev.* E **72** 041804

[22] Suzuki J, Takano A, Deguchi T and Matsushita Y 2009 Dimension of ring polymers in bulk studied by Monte Carlo simulation and self-consistent theory *J. Chem. Phys.* **131** 144902

[23] Tubiana L, Orlandini E and Micheletti C 2011 Probing the entanglement and locating knots in ring polymers: a comparative study of different arc closure schemes arXiv:1103.0475

[24] Valle F, Favre M, De Los Rios P, Rosa A and Dietler G 2005 Scaling exponents and probability distributions of DNA end-to-end distance *Phys. Rev. Lett.* **95** 158105

[25] Witz G 2010 DNA topology: a biophysical study of DNA rings, knots and catenanes by means of atomic force microscopy and numerical simulations *PhD Thesis* École Polytechnique Fédérale de Lausanne

[26] Yao A, Tsukahara H, Deguchi T and Inami T 2004 Distribution of the distance between opposite nodes of random polygons with a fixed knot *J. Phys. A: Math. Gen.* **37** 7993–8006

[27] Zimm B H and Stockmayer W H 1949 The dimensions of chain molecules containing branches and rings *J. Chem. Phys.* **17** 1301–14