The Average Crossing Number of Equilateral Random Polygons

Y. Diao[†]§, A. Dobay[‡], R. B. Kusner[*], K. Millett[**] and A. Stasiak[‡]

[†]Department of Mathematics
University of North Carolina at Charlotte
Charlotte, NC 28223, USA
[‡]Laboratory of Ultrastructural Analysis
University of Lausanne
Lausanne, CH 1015, Switzerland
[*]Department of Mathematics and Statistics
University of Massachusetts at Amherst
Amherst, MA 01003, USA
[**]Department of Mathematics
University of California at Santa Barbara
Santa Barbara, CA 93106, USA

Abstract. In this paper, we study the average crossing number of equilateral random walks and polygons. We show that the mean average crossing number ACN of all equilateral random walks of length n is of the form $\frac{3}{16}\cdot n\cdot \ln n + O(n)$. A similar result holds for equilateral random polygons. These results are confirmed by our numerical studies. Furthermore, our numerical studies indicate that when random polygons of length n are divided into individual knot types, the $\langle \mathrm{ACN}(\mathcal{K})\rangle$ for each knot type $\mathcal{K}$ can be described by a function of the form $\langle \mathrm{ACN}(\mathcal{K})\rangle = a\cdot(n-n_0)\cdot\ln(n-n_0)+b\cdot(n-n_0)+c$ where a, b and c are constants depending on $\mathcal{K}$ and n_0 is the minimal number of segments required to form $\mathcal{K}$. The $\langle \mathrm{ACN}(\mathcal{K})\rangle$ profiles diverge from each other, with more complex knots showing higher $\langle \mathrm{ACN}(\mathcal{K})\rangle$ than less complex knots. Moreover, the $\langle \mathrm{ACN}(\mathcal{K})\rangle$ profiles intersect with the $\langle \mathrm{ACN}\rangle$ profile of all closed walks. These points of intersection define the equilibrium length of $\mathcal{K}$, i.e., the chain length $n_e(\mathcal{K})$ at which a statistical ensemble of configurations with given knot type $\mathcal{K}$ — upon cutting, equilibration and reclosure to a new knot type $\mathcal{K}'$ — does not show a tendency to increase or decrease $\langle \mathrm{ACN}(\mathcal{K}')\rangle$. This concept of equilibrium length seems to be universal, and applies also to other length-dependent observable for random knots, such as the mean radius of gyration $\langle \mathrm{R_g}\rangle$.

AMS classification scheme numbers: 57M25

Submitted to: *J. Phys. A: Math. Gen.*

§ To whom correspondence should be addressed (ydiao@uncc.edu)

1. Introduction

Random polygons are frequently used to model the behaviour of polymers at thermodynamic equilibrium. Probably the simplest but also the most fundamental type of random polygon is that composed of freely jointed segments of equal length (equilateral) where the individual segments have no thickness. Such a random polygon is known as an ideal random walk and it is used to model the behaviour of polymers under so-called theta conditions where polymer segments that are not in a direct contact neither attract nor repel each other. The behaviour of ideal random walks is thoroughly researched by now and it is well established, for example, that the overall dimensions of ideal random walks (such as the average end to end distance or the average radius of gyration) scale with the number of segments n as n^{ν} where $\nu = 0.5$ [12], [15], [11]. Although the overall dimensions of random walks provide important information about the modelled polymers, it is frequently the case that additional characteristics of polymers need to be investigated. One such characteristic is a measure of polymer entanglement.

There are numerous studies that investigated what types of knots are formed on polymer chains [21], [6], [17]. The determination of the knot type of a circular polymer can tell us, for example, what is the topological (minimum) crossing number of the given circular polymer, i.e., the minimum number of crossings one will see no matter how this polymer is artificially stretched, twisted, or bent. In contrast to the minimum crossing number, the average crossing number (ACN) is a more natural geometric measure of polymer entanglement as it refers to the actual number of crossings that can be perceived while observing a non-perturbed trajectory of a given polymer [18]. If a given trajectory of a polymer or of a random walk is orthogonally projected onto a plane along a given direction, one can count the number of crossings that are visible in this particular projection of the trajectory. To be independent of the choice of a particular projection, we use the average crossing number (ACN), which is defined as the average of crossing numbers over all orthogonal projections.

We are particularly interested in $\langle \mathrm{ACN} \rangle$, the average of ACN over the whole statistical ensemble of ideal random walks (or polygons) with a given number of segments. $\langle \mathrm{ACN} \rangle$ was shown to be an interesting measure of physical behaviour of knotted polymers and, in contrast to the minimum crossing number, it also correlates well with the experimentally observed speed of electrophoretic migration of knotted DNA molecules of the same size but of various knot types [22]. Furthermore, $\langle \mathrm{ACN} \rangle$ correlates well with the expected sedimentation coefficient of different types of DNA knots formed on the same size DNA molecules [23], and with relaxation dynamics of modelled knotted polymers [16]. In the case of protein chains, the ACN provides an interesting measure of their compactness [1] and how ACN in proteins scales with the length of polypeptide chain was investigated [2]. Another scaling aspect of ACN was discussed in the case of the so called ideal geometric ropelength minimizing representations of knots [3],[5],[18], and this has generated a great deal of mathematical work recently (see [5] and the references therein).

In this paper we investigate how $\langle \mathrm{ACN} \rangle$ scales with the length of various types of random walks and polygons. We apply two independent approaches: analytical derivations and numerical simulations. The agreement between the two approaches is remarkably close!

2. Equilateral Random Walks and Polygons

Let $U = (u, v, w)$ be a three-dimensional random vector that is uniformly distributed on the unit sphere S^2, i.e., the density function of U is

$$\varphi(U) = \begin{cases} \frac{1}{4\pi} & \text{if } |U| = \sqrt{u^2 + v^2 + w^2} = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Let $\vec{v}_1$ and $\vec{v}_2$ be any two vectors based at the origin that are perpendicular to each other. Let Σ be a plane normal to $\vec{v}_1$ containing $\vec{v}_2$. Let θ be the angle between $\vec{v}_1$ and U (as a unit vector based at the origin), and let ϕ be the angle between $\vec{v}_2$ and the projection of U onto Σ. The values of θ and ϕ are both between 0 and π. One can show that θ and ϕ are independent random variables. Furthermore, ϕ is uniformly distributed on $[0, \pi]$ and the probability density function of θ is $\frac{1}{2}\sin\theta$. It follows that the mean of $\sin\phi$ is $\frac{2}{\pi}$ and the mean of $\sin\theta$ is $\frac{\pi}{4}$. We will need this result later.

Suppose U_1, U_2, ... , U_n are n independent random vectors uniformly distributed on S^2. An equilateral random walk of n steps, denoted by EW_n, is defined as the sequence of points in the three dimensional space $\mathbf{R}^3$: $X_0 = O$, $X_k = U_1 + U_2 + \cdots + U_k$, $k = 1, 2, ..., n$. Each X_k is called a vertex of the EW_n and the line segment joining X_k and X_{k+1} is called an edge of EW_n (which is of unit length). If the last vertex X_n of EW_n is fixed, then we have a conditioned random walk $EW_n|X_n$. In particular, EW_n becomes a polygon if $X_n = O$. In this case, it is called an equilateral random polygon and is denoted by EP_n. Note that the joint probability density function $f(X_1, X_2, ..., X_n)$ of the vertices of an EW_n is simply $f(X_1, X_2, ..., X_n) = \varphi(U_1)\varphi(U_2)\cdots\varphi(U_n) = \varphi(X_1)\varphi(X_2 - X_1)\cdots\varphi(X_n - X_{n-1})$.

Let X_k be the k-th vertex of an EW_n ($n \geq k > 1$), its density function is defined by

$$f_k(X_k) = \int\int\cdots\int \varphi(X_1)\varphi(X_2 - X_1)\cdots \\ \varphi(X_k - X_{k-1})\mathrm{d}X_1\mathrm{d}X_2\cdots\mathrm{d}X_{k-1} \tag{2}$$

and it has the closed form $f_k(X_k) = \frac{1}{2\pi^2 r}\int_0^\infty x \sin rx \left(\frac{\sin x}{x}\right)^k dx$ [20]. It is easy to see from here that $f_k(X_k)$ is approximately normal for large values of k. The following lemma gives a fairly accurate estimate of $f_k(X_k)$. Its proof and a few other related topics can be found in [7], [8] and [9].

Lemma 1 *For $k \geq 10$, we have*

$$\left| f_k(X_k) - \left(\sqrt{\frac{3}{2\pi k}}\right)^3 \exp\left(-\frac{3|X_k|^2}{2k}\right) \right| < \frac{0.5}{k^{\frac{5}{2}}}. \tag{3}$$

In the case of equilateral random polygons, the density function of a vertex is still approximately Gaussian, but its estimation is slightly harder. We have the following lemma.

Lemma 2 *Let X_k be the k-th vertex of an EP_n and let h_k be its density function, then*

$$h_k(X_k) = \left(\sqrt{\frac{3}{2\pi\sigma_{nk}^2}}\right)^3 \exp\left(-\frac{3|X_k|^2}{2\sigma_{nk}^2}\right) + O\left(\frac{1}{k^{5/2}} + \frac{1}{(n-k)^{5/2}}\right), \tag{4}$$

where $\sigma_{nk}^2 = \frac{k(n-k)}{n}$.

Proof. First, the joint density function of the vertices of EP_n is of the form

$$g(X_1, X_2, ..., X_{n-1}) = \frac{1}{f_n(X_n)} \varphi(X_1)\varphi(X_2 - X_1) \cdots \varphi(X_n - X_{n-1})$$

with $X_n = X_0 = O$. Integrate the above over X_1, X_2, ..., X_{n-1} except for X_k, we get

$$h(X_k) = \frac{1}{f_n(O)} \cdot f_k(X_k) \cdot f_{n-k}(X_k).$$

The result follows by applying the formula in Lemma 1.

Remark. Notice that by symmetry, the probability density function $f_k(X_k)$ and $h_k(X_k)$ of X_k (for the equilateral random walks and the equilateral random polygons respectively) depend only on $|X_k|$ so they can be written as $f_k(|X_k|)$ and $h_k(|X_k|)$. Furthermore, if we let $\rho = |X_k|$ (which is also a random variable), then the probability density function of ρ is $4\pi\rho^2 f_k(\rho)$ for the equilateral random walks and $4\pi\rho^2 h_k(\rho)$ for the equilateral random polygons. It follows that the mean diameter of an EW_n or an EP_n is of order $\sqrt{n}$.

Let X_{k+1} and X_{k+2} be two consecutive vertices of an equilateral random polygon EP_n, then the joint probability density function $h(X_1, X_{k+1}, X_{k+2})$ of X_1, X_{k+1} and X_{k+2} is defined by

$$\int \cdots \int \frac{\varphi(X_1)\varphi(X_2 - X_1) \cdots \varphi(X_n - X_{n-1})}{f_n(O)} \widehat{\mathrm{d}X_1}\widehat{\mathrm{d}X_{k+1}}\widehat{\mathrm{d}X_{k+2}}, \quad (5)$$

where the integral is taken over all variables except X_1, X_{k+1} and X_{k+2}. The following lemma can be proved in a similar fashion to that of Lemma 2 and the details are left to the reader.

Lemma 3 *Let X_1, X_{k+1} and X_{k+2} be the first, $(k+1)$-st and $(k+2)$-nd vertices of an EP_n; then their joint probability density function $h_k(X_1, X_{k+1}, X_{k+2})$ can be approximated by*

$$\varphi(X_1)\varphi(X_{k+2} - X_{k+1}) \left(\frac{3}{2\pi\sigma_{nk}^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3|X_{k+1}|^2}{2\sigma_{nk}^2}\right), \quad (6)$$

where $\sigma_{nk}^2 = \frac{k(n-k)}{n}$ and the error term is at most of order $O\left(\frac{1}{k^{5/2}} + \frac{1}{(n-k)^{5/2}}\right)$.

3. Bounds on the Mean Average Crossing Number

We begin this section by considering a special case concerning the calculation of ACN, the average crossing number, when there are only two random steps involved. Assume that P and Q are two fixed points in $\mathbf{R}^3$ such that $r = |P - Q| \geq 4$. Let P_1 and Q_1 be two random points in $\mathbf{R}^3$ such that $U_1 = P_1 - P$ and $U_2 = Q_1 - Q$ are uniformly distributed on the unit sphere S^2. (See Figure 1 below.)

Lemma 4 *Let P, Q, P_1 and Q_1 be as defined above and let $a(\ell_1, \ell_2)$ be the average crossing number between the two line segments $\ell_1 = PP_1$ and $\ell_2 = QQ_1$, then we have*

$$E(a(\ell_1, \ell_2)) = \frac{1}{16r^2} + O(\frac{1}{r^3}). \quad (7)$$

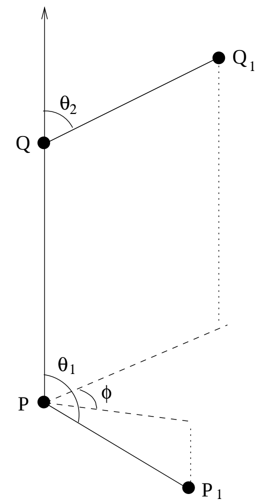

Figure 1. The case of two random edges

Proof. Without loss of generality, let us assume that $P = O$ and Q is on the positive z-axis. Let θ_1 be the angle between $U_1 = \overrightarrow{PP_1}$ and the z-axis and θ_2 be the angle between $U_2 = \overrightarrow{QQ_1}$ and the z-axis. Furthermore, let ϕ be the angle between the projections of U_1 and U_2 on the xy-plane. In [13], it is shown that for fixed P_1 and Q_1, the average crossing number $a(\ell_1, \ell_2)$ between the edges ℓ_1 and ℓ_2 is given by

$$\frac{1}{2\pi}\int_{\ell_1}\int_{\ell_2}\frac{|(\dot\gamma_1(t),\dot\gamma_2(s),\gamma_1(t)-\gamma_2(s))|}{|\gamma_1(t)-\gamma_2(s)|^3}\mathrm{d}t\mathrm{d}s, \tag{8}$$

where γ_1 and γ_2 are the arclength parameterizations of ℓ_1 and ℓ_2 respectively, and $(\dot\gamma_1(t), \dot\gamma_2(s), \gamma_1(t)-\gamma_2(s))$ is the triple scalar product of $\dot\gamma_1(t)$, $\dot\gamma_2(s)$, and $\gamma_1(t)-\gamma_2(s)$. We can write

$$\begin{aligned}\gamma_1(t) &= t\cdot U_1,\ 0\le t\le 1,\\ \gamma_2(s) &= \overrightarrow{OQ} + s\cdot U_2,\ 0\le s\le 1.\end{aligned}$$

By an elementary calculation, we have

$$\begin{aligned}&\int_{\ell_1}\int_{\ell_2}|(\dot\gamma_1(t),\dot\gamma_2(s),\gamma_1(t)-\gamma_2(s))|\mathrm{d}t\mathrm{d}s\\ &= \int_0^1\int_0^1 |(U_1,U_2,\overrightarrow{OQ})|\mathrm{d}t\mathrm{d}s\\ &= |(U_1,U_2,\overrightarrow{OQ})| = r\sin\phi\sin\theta_1\sin\theta_2.\end{aligned}$$

Using the first paragraph in Section 2, the reader can verify that ϕ is uniformly distributed over $[0,\pi]$ and that ϕ, θ_1, and θ_2 are independent; furthermore, the probability density functions for θ_1 and θ_2 are $\frac{1}{2}\sin\theta_1$ and $\frac{1}{2}\sin\theta_2$. Since $||\gamma_1(t)-\gamma_2(s)|-r|\le 2$ and $r\ge 4$, we have

$$\frac{1}{|\gamma_1(t)-\gamma_2(s)|^3} = \frac{1}{r^3} + O(\frac{1}{r^4}).$$

It follows that

$$a(\ell_1, \ell_2) = \frac{1}{2\pi r^2} \sin\phi \sin\theta_1 \sin\theta_2 + O(\frac{1}{r^3})$$

and

$$E(a(\ell_1, \ell_2)) = \int\int a(\ell_1, \ell_2)\varphi(U_1)\varphi(U_2)\mathrm{d}U_1\mathrm{d}U_2$$

$$= \frac{1}{8\pi^2 r^2}\int_0^\pi\int_0^\pi\int_0^\pi \sin\phi\sin^2\theta_1\sin^2\theta_2\mathrm{d}\phi\mathrm{d}\theta_1\mathrm{d}\theta_2 + O(\frac{1}{r^3})$$

$$= \frac{1}{16r^2} + O(\frac{1}{r^3}).$$

□

We are now ready to state and prove our first main theorem.

Theorem 1 *Let χ_n be the ACN of an equilateral random walk of n steps; then*

$$E(\chi_n) = \frac{3}{16} n \ln n + O(n).$$

Proof. Let ℓ_k be the k-th segment of an EW_n, that is, $\ell_k = \overline{X_{k-1}X_k}$ $(1 \le k \le n)$. Let $a(\ell_i, \ell_j)$ be the average crossing number between ℓ_i and ℓ_j; then we have

$$\chi_n = \frac{1}{2}\sum_{1\le i,j\le n} a(\ell_i, \ell_j),$$

and

$$E(\chi_n) = \frac{1}{2}\sum_{1\le i,j\le n} E(a(\ell_i, \ell_j)) = \sum_{1\le i<j\le n} E(a(\ell_i, \ell_j)).$$

By symmetry, $E(a(\ell_{i_1}, \ell_{j_1})) = E(a(\ell_{i_2}, \ell_{j_2}))$ whenever $|j_1 - i_1| = |j_2 - i_2|$. It follows that

$$E(\chi_n) = \sum_{3\le j\le n} (n - j + 1)E(a(\ell_1, \ell_j)), \tag{9}$$

where j starts at 3 since $a(\ell_1, \ell_2) = 0$. Letting $r_j = |X_{j-1} - X_1|$, $P = X_1$, $P_1 = O$, $Q = X_{j-1}$ and $Q_1 = X_j$, we obtain

$$E(a(\ell_1, \ell_j)|r_j) = \frac{1}{16r_j^2} + O(\frac{1}{r_j^3})$$

for any fixed $r_j \ge 4$ by Lemma 4. Since r_j is a random variable depending only on $X_{j-1} - X_1$, and since $X_{j-1} - X_1$ has the same density distribution function as X_{j-2}, it follows that

$$E(a(\ell_1, \ell_j)) = \int E(a(\ell_1, \ell_j)|r_j) f_{j-2}(|X_{j-1} - X_1|)\mathrm{d}(X_{j-1} - X_1)$$

$$= \int E(a(\ell_1, \ell_j)|r_j) 4\pi r_j^2 f_{j-2}(r_j)\mathrm{d}r_j$$

$$= \int_{r_j<4} E(a(\ell_1, \ell_j)|r_j) 4\pi r_j^2 f_{j-2}(r_j)\mathrm{d}r_j$$

$$+ \int_{r_j\ge4} E(a(\ell_1, \ell_j)|r_j) 4\pi r_j^2 f_{j-2}(r_j)\mathrm{d}r_j.$$

Since $a(\ell_1, \ell_j)$ is the average crossing number between two straight edges, it is at most 1. So if $j \geq 12$, then by Lemma 1 we have

$$\begin{aligned}
&\int_{r_j<4} E(a(\ell_1,\ell_j)|r_j) 4\pi r_j^2 f_{j-2}(r_j) \mathrm{d}r_j \\
\leq &\int_{r_j<4} 4\pi r_j^2 f_{j-2}(r_j) \mathrm{d}r_j \\
\leq &\, 4\pi \int_0^4 \left(\left(\frac{3}{2\pi(j-2)} \right)^{\frac{3}{2}} r_j^2 \exp\left(-\frac{3r_j^2}{2(j-2)} \right) + \frac{r_j^2}{2(j-2)^{\frac{5}{2}}} \right) \mathrm{d}r_j \\
= &\, O\left(\frac{1}{j^{\frac{3}{2}}} \right).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\int_{r_j \geq 4} E(a(\ell_1,\ell_j)|r_j) 4\pi r_j^2 f_{j-2}(r_j) \mathrm{d}r_j \\
= &\, \frac{\pi}{4} \int_4^{j-2} \left(1 + O(\frac{1}{r_j}) \right) \left(\frac{3}{2\pi(j-2)} \right)^{\frac{3}{2}} \exp\left(-\frac{3r_j^2}{2(j-2)} \right) \mathrm{d}r_j \\
+ &\, \frac{\pi}{4} \int_4^{j-2} \left(1 + O(\frac{1}{r_j}) \right) \left(\frac{1}{2(j-2)^{\frac{5}{2}}} \right) \mathrm{d}r_j \\
= &\, \frac{3}{16j} + O(\frac{\ln j}{j^{\frac{3}{2}}}).
\end{aligned}$$

(The details of the calculations leading to the above equation are left to the reader.) Combining the above results, we obtain

$$E(a(\ell_1,\ell_j)) = \frac{3}{16j} + O(\frac{\ln j}{j^{\frac{3}{2}}}).$$

So

$$E(\chi_n) = \frac{3}{16} n \sum_{3 \leq j \leq n} \frac{1}{j} - \frac{3}{16}(n-3) + nO(\sum_{3 \leq j \leq n} \frac{\ln j}{j^{\frac{3}{2}}})$$

by (9). The result follows since $\sum_{3 \leq j \leq n} \frac{1}{j} - \ln n$ and $\sum_{3 \leq j \leq n} \frac{\ln j}{j^{\frac{3}{2}}}$ both converge. □

Our next theorem is an analogue of Theorem 1 with equilateral random walks replaced by equilateral random polygons. Given that the average diameter of equilateral random polygons is smaller than the average diameter of equilateral random walks, one probably would expect to see a larger mean ACN. So the result of Theorem 2 is a bit surprising.

Theorem 2 *Let χ'_n be the ACN of an equilateral random polygon of n steps; then*

$$E(\chi'_n) = \frac{3}{16} n \ln n + O(n).$$

Proof. Although the result of Theorem 2 is similar to that of Theorem 1, its proof requires more technical treatment since the equilateral random polygons no longer have the property of an unconditioned Markov chain. Let χ'_n be the ACN of an equilateral random polygon of $n \geq 4$ steps. As we did in the proof of Theorem 1, let ℓ_k be the k-th segment of an EP_n, that is, $\ell_k = \overline{X_{k-1}X_k}$ $(1 \leq k \leq n)$. Let $a(\ell_i, \ell_j)$ be the average crossing number between ℓ_i and ℓ_j; then we have

$$\chi'_n = \frac{1}{2} \sum_{1 \leq i,j \leq n} a(\ell_i, \ell_j),$$

and

$$E(\chi'_n) = \frac{1}{2} \sum_{1 \leq i,j \leq n} E(a(\ell_i, \ell_j)).$$

In the case of an EW_n, we have $E(a(\ell_{i_1}, \ell_{j_1})) = E(a(\ell_{i_2}, \ell_{j_2}))$ whenever $|j_1 - i_1| = |j_2 - i_2|$, where ℓ_{i_1}, ℓ_{j_1} may belong to some EW_n and ℓ_{i_2}, ℓ_{j_2} may belong to some EW_m such that $n \neq m$. This is no longer the case for an EP_n. However, if ℓ_{i_1}, ℓ_{j_1}, ℓ_{i_2}, and ℓ_{j_2} are indeed segments of some EP_n, then we still have $E(a(\ell_{i_1}, \ell_{j_1})) = E(a(\ell_{i_2}, \ell_{j_2}))$ whenever $|j_1 - i_1| = |j_2 - i_2|$ or $|j_1 - i_1| = n - |j_2 - i_2|$ by symmetry. It follows that

$$E(\chi'_n) = n \sum_{3 \leq j \leq (n+1)/2} E(a(\ell_1, \ell_j)).$$

Again, j starts at 3 in the above formula since $a(\ell_1, \ell_2)$ is always 0. Let $r_j = |X_{j-1} - X_1|$. Since $a(\ell_1, \ell_j)$ depends only on X_1, X_{j-1} and X_j, by Lemmas 3 and 4, we have

$$\begin{aligned}
& E(a(\ell_1, \ell_j)) \\
= & \int\int\int a(\ell_1, \ell_j) h_{j-2}(X_1, X_{j-1}, X_j) \mathrm{d}X_1 \mathrm{d}X_{j-1} \mathrm{d}X_j \\
= & \int\int a(\ell_1, \ell_j) \varphi(X_1) \varphi(X_j - X_{j-1}) \mathrm{d}X_1 \mathrm{d}X_j \cdot \\
& \int \left(\left(\sqrt{\frac{3}{2\pi\sigma^2_{n(j-2)}}} \right)^3 \exp\left(-\frac{3|X_{j-1}|^2}{2\sigma^2_{n(j-2)}} \right) + O\left(\frac{1}{j^{5/2}} \right) \right) \mathrm{d}X_{j-1} \\
= & \int\int a(\ell_1, \ell_j) \varphi(U_1) \varphi(U_2) \mathrm{d}U_1 \mathrm{d}U_2 \cdot \\
& \int_{r_j < 4} \left(\left(\frac{3}{2\pi\sigma^2_{n(j-2)}} \right)^{\frac{3}{2}} \exp\left(-\frac{3|X_{j-1}|^2}{2\sigma^2_{n(j-2)}} \right) + O\left(\frac{1}{j^{5/2}} \right) \right) \mathrm{d}X_{j-1} \\
+ & \int\int a(\ell_1, \ell_j) \varphi(U_1) \varphi(U_2) \mathrm{d}U_1 \mathrm{d}U_2 \cdot \\
& \int_{r_j \geq 4} \left(\left(\frac{3}{2\pi\sigma^2_{n(j-2)}} \right)^{\frac{3}{2}} \exp\left(-\frac{3|X_{j-1}|^2}{2\sigma^2_{n(j-2)}} \right) + O\left(\frac{1}{j^{5/2}} \right) \right) \mathrm{d}X_{j-1} \\
= & O(\frac{1}{j^{\frac{3}{2}}}) + \frac{\pi}{4} \int_4^{j-2} \left(1 + O(\frac{1}{r_j}) \right) \cdot \\
& \left(\frac{3}{2\pi\sigma^2_{n(j-2)}} \right)^{\frac{3}{2}} \exp\left(-\frac{3r_j^2}{2\sigma^2_{n(j-2)}} \right) \mathrm{d}r_j \\
= & \frac{3}{16\sigma^2_{n(j-2)}} + O(\frac{\ln j}{j^{\frac{3}{2}}})
\end{aligned}$$

$$= \frac{3}{16}\left(\frac{1}{j-2} + \frac{1}{n-j+2}\right) + O(\frac{\ln j}{j^{\frac{3}{2}}})$$
$$= \frac{3}{16j} + O(\frac{1}{n}) + O(\frac{\ln j}{j^{\frac{3}{2}}}),$$

where $U_1 = X_1$ and $U_2 = X_j - X_{j-1}$. Since $\sum_{3\le j\le n/2} \frac{1}{j} - \ln n$ and $\sum_{3\le j\le n/2} \frac{\ln j}{j^{\frac{3}{2}}}$ both converge, the result follows. □

Remark. Notice that $E(\chi_n)$ can be approximated by $\frac{3}{16}\sum_{3\le j\le n}\frac{n-j}{j}$ and $E(\chi'_n)$ can be approximated by $\frac{3}{16}n\sum_{3\le j\le n}\frac{1}{j}$. The error terms involved in both cases are at most linear in n. Notice that

$$\frac{3}{16}n\sum_{3\le j\le n}\frac{1}{j} - \frac{3}{16}\sum_{3\le j\le n}\frac{n-j}{j} = \frac{3}{16}(n-3).$$

Interestingly, our simulation result also reveals that $E(\chi'_n) - E(\chi_n) \approx \frac{3}{16}(n-3)$. This strongly suggests that when $E(\chi_n)$ and $E(\chi'_n)$ are approximated by $\frac{3}{16}\sum_{3\le j\le n}\frac{n-j}{j}$ and $\frac{3}{16}n\sum_{3\le j\le n}\frac{1}{j}$ respectively, the errors involved are of (approximately) the same magnitude.

The next two theorems concern the extreme values of the ACN.

Theorem 3 *Let $\mathcal{K}$ be any given knot type and let $\chi'_n(\mathcal{K})$ be the ACN of an equilateral random polygon of n steps that is of knot type $\mathcal{K}$. Then $E(\chi'_n(\mathcal{K})) \le n^2$.*

Proof. This is obvious since $a(\ell_1, \ell_j) \le 1$ for each j. □

Theorem 4 *For n large enough, there always exists a knot type $\mathcal{K}$ such that $E(\chi'_n(\mathcal{K})) \ge c'n^2$ for some constant $c' > 0$.*

Proof. If n is large enough, one can always construct a (p, q) torus knot with pq as its crossing number such that p and q are both around $n/4$. Since $pq \ge c'n^2$ for some constant c' in this case and the ACN of any EP_n with this knot type is at least pq, the result follows. The details of the construction is left to the reader. □

4. Simulation Methods

Although the average crossing number of a random walk or polygon W can be calculated by the modified Gauss formula

$$\text{ACN}(W) = \frac{1}{4\pi}\int_W\int_W \frac{|(\dot{\gamma}(t), \dot{\gamma}(s), \gamma(t)-\gamma(s))|}{|\gamma(t)-\gamma(s)|^3}\text{d}t\text{d}s, \tag{10}$$

where γ is the arclength parameterization of W, the application of this formula leads to problems when some non-consecutive segments get very close to each other. For this reason our numerical determination for the ACN of random chains was based on counting the crossings in numerous projections of analyzed trajectories. We calculated the number of crossings in individual projections and then averaged over 50 randomly chosen directions of projections to obtain a good approximation of the actual ACN value for a given trajectory. To generate random equilateral walks (open walks) we first created a set of unit vectors with the same origin that randomly equisampled the surface of the unit sphere. The vectors were then joined sequentially while maintaining their original directions.

To generate random equilateral polygons (closed walks) we followed the approach of Dykhne [19]. For example, to construct a 100 segment-long closed trajectory we first created a set of 50 unit vectors randomly equisampling the surface of a unit sphere. Subsequently we added to this set another 50 unit vectors that are opposite to the original set. This procedure assures that the sum of the 100 vectors is zero and that the trajectory obtained by any random sequential joining of all 100 vectors results in a closed trajectory. To eliminate correlated parallel vectors in random trajectories, the set of 100 vectors was de-correlated by multiple rotations of random pairs of vectors around their respective sum vectors. Finally all 100 randomized vectors were sequentially joined to create an equilateral random polygon. Knot types of the resulting random polygons were recognized by calculation of their HOMFLY polynomials [14].

5. Numerical Results

Figure 2A shows the $\langle \text{ACN} \rangle$ values obtained in numerical simulations of ideal random walks in a non-constrained linear form (EW or open) or with a constraint of closure (EP or closed). We have analyzed walks with up to 1000 segments and each of the $\langle \text{ACN} \rangle$ data points was obtained by averaging the ACN values from 10^5 independent random configurations of open or closed random walks of the corresponding size.

To check our analytical predictions we have fitted the computed data points with the function $a \cdot n \cdot \ln(n) + b \cdot n$, leaving the two parameters a and b free. In both cases (closed and open walks) we have obtained an excellent fit (with $R = 1$) where the prefactor a was practically equal to $\frac{3}{16}$ ($\frac{3}{16} + 0.00065$ and $\frac{3}{16} - 0.00105$, respectively). Therefore, we have proceeded with another fit where we have fixed the prefactor a to $\frac{3}{16}$ and have left only one free parameter b. These fits are presented here and it is visible that the quality of these fits remains excellent as the fitted functions pass almost perfectly through all the data points. These results therefore confirm our theoretical prediction that the ACN of open and closed random walks show the above dependence on the number of segments n.

We decided therefore to check whether the difference of the ACN between closed and open random walks of the same chain length n can be described by the linear relation $\langle \text{ACN} \rangle_{closed} - \langle \text{ACN} \rangle_{open} \approx \frac{3}{16} \cdot n$. As can be seen in Figure 2B, this prediction is also entirely confirmed by our numerical results.

Figure 3 illustrates how the $\langle \text{ACN} \rangle$ values scale with the chain length of random closed walks representing various types of knots. It is clearly visible that random configurations of more complex knots have higher $\langle \text{ACN} \rangle$ values than the random configurations of simpler knots. For short chain lengths, the difference between the $\langle \text{ACN} \rangle$ of random configurations of a given knot and the $\langle \text{ACN} \rangle$ of random configurations of unknots with the same chain length is well approximated by the actual ACN value of ideal (ropelength minimizing) geometric representation of a given knot type [18]. For example, for 6 and 14 segment-long random walks, the difference

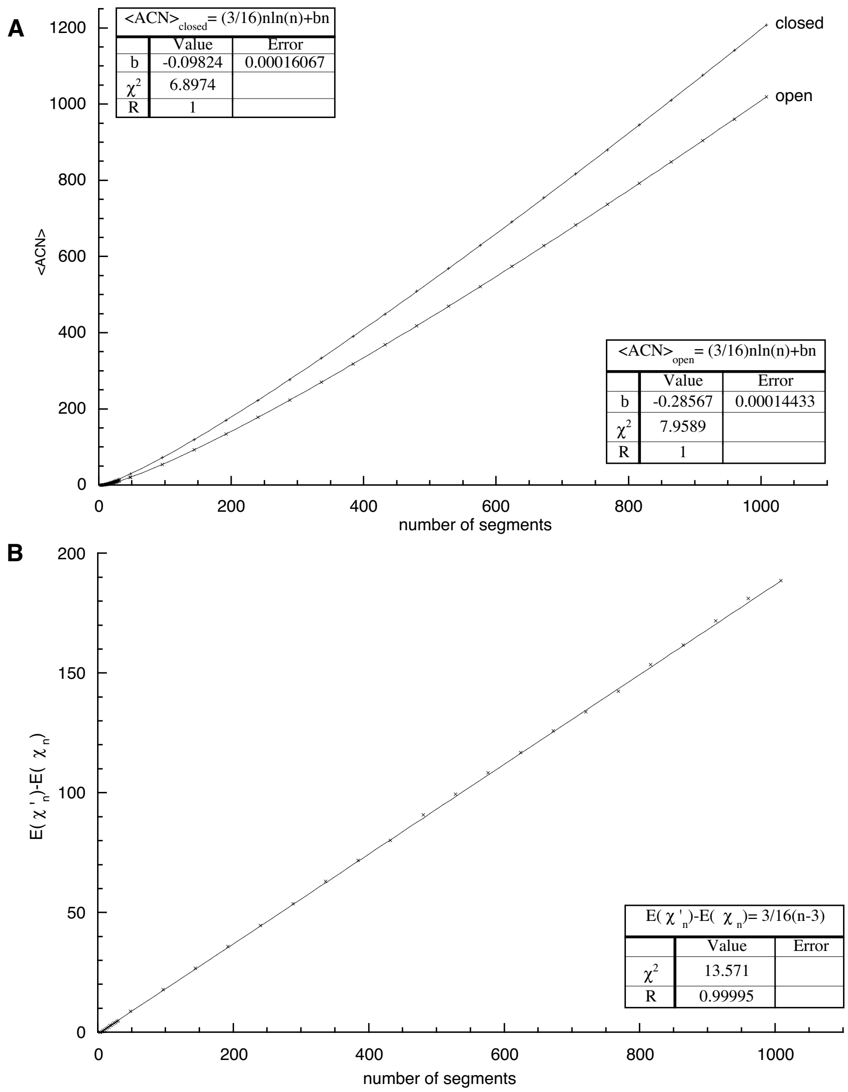

Figure 2. Comparison of the mean average crossing number values ⟨ACN⟩ for corresponding chain lengths of closed and open ideal random walks (equilateral random polygons and equilateral random walks). The standard deviation is about the size of the data points. A. The ⟨ACN⟩ values obtained in numerical simulations of closed and open random walks are marked as data points and the fitting functions are listed. The analyzed sample sizes of simulated configurations were bigger than 10^5 for each chain length. B. The difference of ⟨ACN⟩ between closed and open ideal random walks of the same length.

between the ⟨ACN⟩ of random trefoils and that of random unknots amounts to 4.14 and 5.21, while the ACN of an ideal trefoil amounts to 4.26. This relation between the ACN of ideal knots and the ⟨ACN⟩ of random walks with relatively small chain length was noticed earlier [18]. However, as the length of the analyzed random walks increases the $\langle\text{ACN}\rangle$ values of random knots of different type diverge from each other as noticed earlier in [16].

To analyze how the $\langle\text{ACN}\rangle$ values of random knots of various knot types scale with their chain lengths, we have fitted the computed data points with the function $a \cdot (n - n_0) \cdot \ln(n - n_0) + b \cdot (n - n_0) + c$, where a, b and c are free parameters, n is the number of segments in a walk and n_0 is the minimal number of segments required to form a given knot type [4]. In all cases analyzed by us, the fitted curves nearly pass through all the data points. The actual fitted parameters for different knot types are listed in Figure 3 together with the quality of the fit.

Let us denote by $\langle\text{ACN}(\mathcal{K})\rangle$ the mean average crossing number for all closed walks with knot type $\mathcal{K}$. Comparing the fitted parameters for different knot types with those for all closed walks analyzed in Figure 2, it is clear that the $n \cdot \ln(n)$ part has a much weaker contribution in the case of individual knot types while the linear contribution is much stronger. Consequently, the $\langle\text{ACN}(\mathcal{K})\rangle$ profiles of random polygons of each knot type $\mathcal{K}$ shows a lower growth rate than the $\langle\text{ACN}\rangle$ profiles of all closed walks that are grouped together independently of their knot type. It is interesting to consider the values of the free parameter c obtained for different knots. These values are close to the ACN values of ideal configurations of the knot type $\mathcal{K}$, but may in fact better correspond to the $\langle\text{ACN}\rangle$ of all configurations that can be realized using the minimal number (n_0) of segments for $\mathcal{K}$.

Comparing the $\langle\text{ACN}(\mathcal{K})\rangle$ profiles for random polygons of knot type $\mathcal{K}$ with that of all closed walks, it is visible that (with the exception of the unknots) the $\langle\text{ACN}(\mathcal{K})\rangle$ profiles intersect with the profile for all closed walks (see Figure 3). This is due to the fact that individual knot types show a smaller growth rate than all closed walks grouped together (see discussion above), while each individual knot type (with an exception of the unknots) initially have higher $\langle\text{ACN}(\mathcal{K})\rangle$ values than the $\langle\text{ACN}\rangle$ values for the ensemble of all closed walks. The more complex the knot, the later its $\langle\text{ACN}(\mathcal{K})\rangle$ profile intersects with the $\langle\text{ACN}\rangle$ profile of all closed walks.

The intersection of the $\langle\text{ACN}(\mathcal{K})\rangle$ and $\langle\text{ACN}\rangle$ profiles determines an $\langle\text{ACN}\rangle$-based equilibrium length $n_e(\mathcal{K})$, an interesting characteristic of a given knot type $\mathcal{K}$. Below the equilibrium length $n_e(\mathcal{K})$, a given knot shows an excess of $\langle\text{ACN}(\mathcal{K})\rangle$ as compared with the $\langle\text{ACN}\rangle$ of all possible walks realized with the same chain length. Therefore if one would cut a knot realized with a chain shorter than its equilibrium length $n_e(\mathcal{K})$, let it equilibrate and then promote reclosure of the ends, one would observe a tendency to form simpler knots than the original knot. Above the equilibrium length $n_e(\mathcal{K})$, the situation reverses. If one would cut a knot realized with the chain length longer than its equilibrium length, one would observe after chain reclosure a tendency to form more complex knots. At the equilibrium length, however, a knot would show no tendency to decrease or increase its $\langle\text{ACN}\rangle$ by forming less or more complex knots after cutting and reclosure.

Interestingly, these equilibrium lengths determined by intersections of the $\langle\text{ACN}(\mathcal{K})\rangle$ profiles of individual knot types with the $\langle\text{ACN}\rangle$ profile of all closed walks practically coincide with the equilibrium lengths determined by intersections of the corresponding profiles of the mean radius of gyration $\langle\text{R}_\text{g}\rangle$ (see Figure 3). The equilibrium length for trefoils and figure-eight knots based on measurements of $\langle\text{ACN}\rangle$ amounted to 176 ± 10 and 258 ± 10 segments respectively, while the equilibrium length of these knots based on measurements of $\langle\text{R}_\text{g}\rangle$ [10] amounted to 174 ± 14 and 270 ± 17

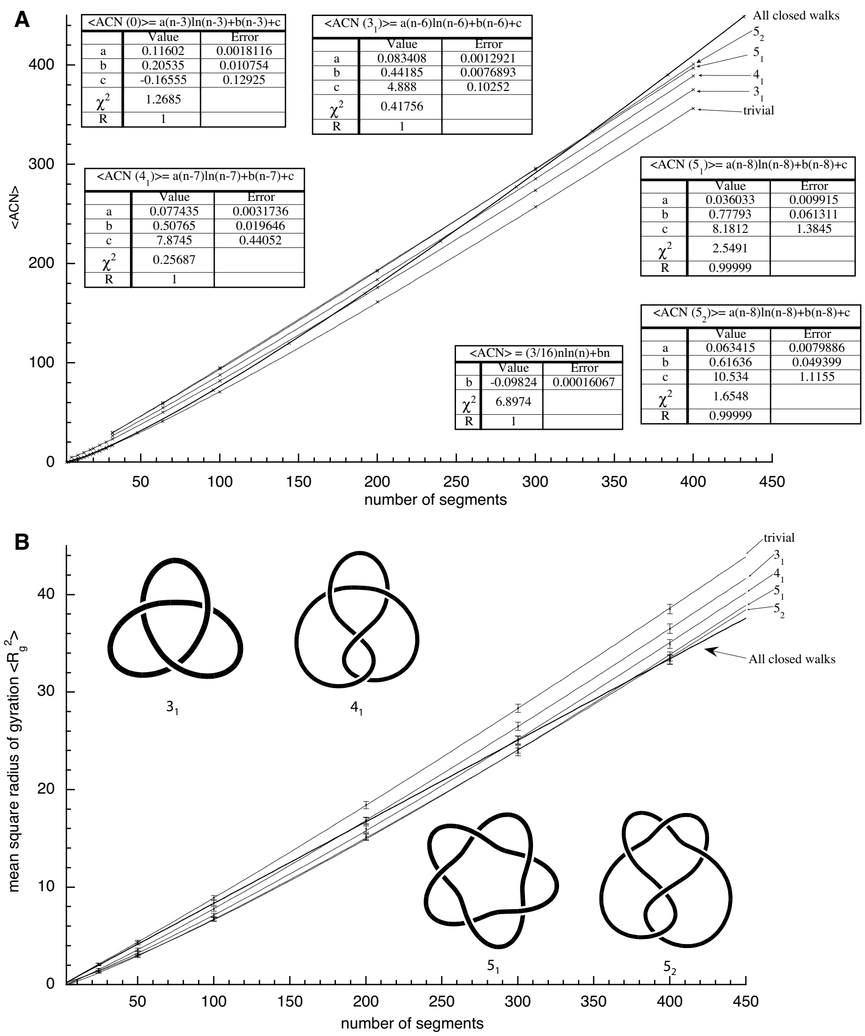

Figure 3. Effect of the topology on the average crossing number values $\langle \mathrm{ACN} \rangle$ for different types of closed random walk. The standard deviation is about the size of the data points. A. The $\langle \mathrm{ACN} \rangle$ values obtained in numerical simulations of closed random walks representing different knot types (trivial, 3_1, 4_1, 5_1 and 5_2) and all closed walks. The data points are marked and the fitting functions are listed. The statistical sets for different knots and different chain size were not the same. Highest quality data are for unknots with $2,319,455$ configurations analyzed in total and the poorest data set was this of 5_1 knots with $11,406$ configurations analyzed in total. Notice that $\langle \mathrm{ACN}(\mathcal{K}) \rangle$ profiles for individual knot types intersect with the $\langle \mathrm{ACN} \rangle$ profile for all closed random chains of the same length. B. The profiles for radius of gyration $\langle \mathrm{R}_\mathrm{g}^2 \rangle$ of individual random knots intersect with the $\langle \mathrm{R}_\mathrm{g}^2 \rangle$ profile of all closed random walks. The points of intersections define an R_g-based equilibrium length (the data in the inset are taken from [10]).

segments respectively (less robust statistical sampling in the case of R_g measurements caused the bigger error range). It was observed in [10] that the $\langle \mathrm{R}_\mathrm{g} \rangle$-based equilibrium

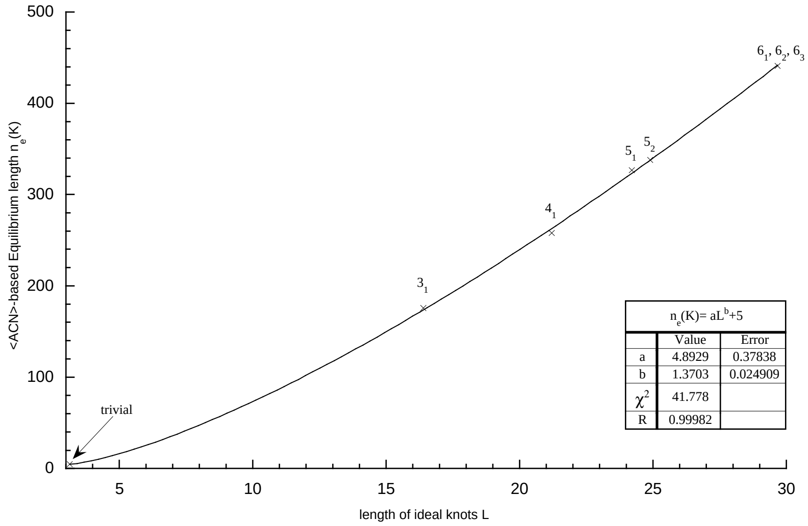

Figure 4. Equilibrium lengths $n_e(\mathcal{K})$ of random knots $\mathcal{K}$ based on measurements of $\langle$ACN$\rangle$ scale with the ropelength ($\frac{l}{d}$ ratio) of ideal geometric (ropelength minimizing) representations of corresponding type of knots. The data of the ropelength used here are from [22]. The standard deviation is about the size of the data points.

lengths of different knots show a power law relation with the ropelength (length to diameter ratio $\frac{l}{d}$) of ideal (ropelength minimizing) geometric representations of the corresponding knots.

Since statistical sampling in the present study permitted us to define more accurately the equilibrium lengths of various knots, we decided to check how well the relation between the $\langle$ACN$\rangle$-based equilibrium length and the ropelength ($\frac{l}{d}$ ratio) of the corresponding knots can be described by a simple power law function. Figure 4 shows that the fit is very good and that the length of ideal (ropelength minimizing) knots scales with the equilibrium length of random knots of the corresponding type.

Based on the observation that the $\langle$ACN$\rangle$-based equilibrium lengths coincide with $\langle R_g \rangle$-based equilibrium lengths, we propose that the concept of the equilibrium length may be universal, and thus may constitute an important characteristic of different knot types. We plan to investigate what other types of length dependent observable of random walks may be used to determine the equilibrium length of a given knot type, and whether the resulting equilibrium lengths will coincide with those obtained by studying the $\langle$ACN$\rangle$- or $\langle R_g \rangle$-scaling properties of knots.

6. Conclusions

We have provided an analytical proof that for long equilateral open and closed random walks their average crossing number calculated over all statistical ensembles of walks with the same number of segments (n) can be expressed by the formula

$\langle \mathrm{ACN} \rangle = \frac{3}{16} n \ln n + O(n)$. Subsequently, we have used numerical simulations to demonstrate that the analytically predicted scaling of $\langle \mathrm{ACN} \rangle$ with equal number of segments in a closed or open equilateral random walk holds not only for long chains, but also for short ones. It is expected that the closed walks would have a higher mean ACN value than that for the open walks. We have observed that the difference between $\langle \mathrm{ACN} \rangle$ for closed and open equilateral random walks with the same number of segments n follows the linear pattern $\langle \mathrm{ACN}_{\mathrm{closed}} \rangle - \langle \mathrm{ACN}_{\mathrm{open}} \rangle = \frac{3}{16} n$. This relation, interestingly, can be explained by looking at the main terms used to derive the general $\langle \mathrm{ACN} \rangle$ formula. We have also analyzed the scaling of $\langle \mathrm{ACN} \rangle$ with the number of chain segments n for individual knot types $\mathcal{K}$ and observed that in each case the observed relation can be described by a formula $\langle \mathrm{ACN}(\mathcal{K}) \rangle = a(n-n_0)\ln(n-n_0)+b(n-n_0)+c$, where n_0 is the minimal number of equilateral segments needed to form the given knot type $\mathcal{K}$. Interestingly, the $\langle \mathrm{ACN}(\mathcal{K}) \rangle$ profiles show slower growth rates than the corresponding $\langle \mathrm{ACN}_{\mathrm{closed}} \rangle$ profile of all closed walks. Our simulation result indicates that, as the complexity of the knot type $\mathcal{K}$ increases, the coefficient a in our fitting formula decreases and the coefficient b increases. Finally, the intersections of $\langle \mathrm{ACN}(\mathcal{K}) \rangle$ profiles with the $\langle \mathrm{ACN}_{\mathrm{closed}} \rangle$ profile define the so called equilibrium length of a given knot type $\mathcal{K}$, i.e., the length at which an ensemble of knots of a given type upon cutting and reclosure would show no tendency to decrease or increase its $\langle \mathrm{ACN} \rangle$ by forming less or more complex knots.

Acknowledgments

The authors wish to thank Gustavo Arteca, Greg Buck, Jacques Dubochet and Alexander Grosberg for their insights and comments. This work was supported in part by NSF grant DMS-0310562 to Y. Diao, by NSF grant DMS-0076085 to R. Kusner and by Swiss National Science Foundation Grants 31-68151.02 and 31-58841.99 to A. Stasiak. We especially thank the Bernoulli Centre of Swiss Federal Institute of Technology in Lausanne for sponsoring a joint meeting of all the authors during final stages of this work. A preliminary version of this paper was presented for the first time at the conference "Knots, random walks and biomolecules" (co-organized by John H. Maddocks, Andrzej Stasiak and sponsored by the Bernoulli Centre of Swiss Federal Institute of Technology in Lausanne) held in Les Diablerets, Switzerland during July 14-17, 2003.

References

[1] Arteca G 1995, *Phys. Rev. E.* **51** 2600–10.
[2] ——1997, *Phys. Rev. E.* **56** 4516–20.
[3] Buck G and Simon J 1999, *Topology Appl.* **91** 245–57.
[4] Calvo J and Millett K 1998, *Ideal Knots,* Ser. Knots Everything, vol 19, ed A Stasiak *et al* (World Sci. Publishing, River Edge, NJ) p 107.
[5] Cantarella J, Kusner R B and Sullivan J M 2002, *Invent. Math.* **150**(2) 257–86.
[6] Deguchi T and Tsurusaki K 1994, *J. Knot Theory Ramifications* **3**(3) 321–53.
[7] Diao Y 1995, *J. Knot Theory Ramifications* **4**(2) 189–96.
[8] ——, Nardo J and Sun Y 2001, *J. Knot Theory Ramifications* **10**(4) 597–607.
[9] ——, Pippenger N and Sumners D W 1994, *J. Knot Theory Ramifications* **3**(3) 419–29.
[10] Dobay A, Dubochet J, Millett K, Sottas P E and Stasiak A 2003, *Proc. Nat. Acad. Sci. USA* **100**(10) 5611–5.
[11] Doi M and Edwards S F 1986, *The theory of polymer dynamics*, Oxford University Press.
[12] Flory J P 1953, *Principles of Polymer Chemistry*, Cornell University Press.

[13] Freedman M H and He Z 1991, *Annals of Math.* **134** 189–229.
[14] Freyd P, Yetter D, Hoste J, Lickorish W B R, Millett K and Oceanu A 1985, *Bulletin of AMS* **12** 239–46.
[15] Gennes P G de 1979, *Scaling Concepts in Polymer Physics*, Cornell University Press.
[16] Huang J Y and Lai P Y 2001, *Phys. Rev. E.* **63**, 021506-1–6.
[17] Katritch V, Olson W K, Vologodskii A V, Dubochet J and Stasiak A 2000, *Phys. Rev. E.* **61** 5545–9.
[18] ——, Bednar J, Michoud D, Scharein R G, Dubochet J and Stasiak A 1996, *Nature* **384** 142–5.
[19] Klenin K V, Vologodskii A V, Anshelevich V V, Dykhne A M and Frank-Kamenetskii M D 1988, *J. Biomol. Struct. Dyn.* **5** 1173.
[20] Rayleigh L 1919, *Phil. Mag. S. 6.* **37**(220) 321–47.
[21] Rybenkov V V, Cozzarelli N R and Vologodskii A V 1993, *Proc. Nat. Acad. Sci. USA,* **90** 5307–11.
[22] Stasiak A, Katritch V, Bednar J, Michoud D and Dubochet J 1996, *Nature* **384** 122.
[23] Vologodskii A V, Crisona N, Laurie B, Pieranski P, Katritch V, Dubochet J and Stasiak A 1998, *J. Mol. Biol.* **278** 1–3.