The average inter-crossing number of equilateral random walks and polygons

2005 J. Phys. A: Math. Gen. 38 7601

(http://iopscience.iop.org/0305-4470/38/35/001)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **38** (2005) 7601–7616 doi:10.1088/0305-4470/38/35/001

The average inter-crossing number of equilateral random walks and polygons

Y Diao[1], A Dobay[2] and A Stasiak[3]

[1] Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

[2] Arnold Sommerfeld Center for Theoretical Physics, Ludwig-Maximilians University, 80333 Muenchen, Germany

[3] Laboratory of Ultrastructural Analysis, University of Lausanne, Lausanne, CH 1015, Switzerland

E-mail: ydiao@uncc.edu and adobay@theorie.physik.uni-muenchen.de

Received 22 April 2005, in final form 30 June 2005
Published 16 August 2005
Online at stacks.iop.org/JPhysA/38/7601

Abstract

In this paper, we study the average inter-crossing number between two random walks and two random polygons in the three-dimensional space. The random walks and polygons in this paper are the so-called equilateral random walks and polygons in which each segment of the walk or polygon is of unit length. We show that the mean average inter-crossing number ICN between two equilateral random walks of the same length n is approximately linear in terms of n and we were able to determine the prefactor of the linear term, which is $a = \frac{3\ln 2}{8} \approx 0.2599$. In the case of two random polygons of length n, the mean average inter-crossing number ICN is also linear, but the prefactor of the linear term is different from that of the random walks. These approximations apply when the starting points of the random walks and polygons are of a distance ρ apart and ρ is small compared to n. We propose a fitting model that would capture the theoretical asymptotic behaviour of the mean average ICN for large values of ρ. Our simulation result shows that the model in fact works very well for the entire range of ρ. We also study the mean ICN between two equilateral random walks and polygons of different lengths. An interesting result is that even if one random walk (polygon) has a fixed length, the mean average ICN between the two random walks (polygons) would still approach infinity if the length of the other random walk (polygon) approached infinity. The data provided by our simulations match our theoretical predictions very well.

PACS numbers: 02.10.Kn, 02.40.Sf, 05.40.Fb
Mathematics Subject Classification: 57M25

1. Introduction

The behaviour of ideal random walks and polygons is thoroughly researched and is now a well-established research area [6, 7, 9]. Probably the simplest, but also the most fundamental type of random walk (polygon), is that composed of freely jointed segments of equal length (equilateral) where the individual segments have no thickness. Although a random walk or polygon so defined may have self-intersections (in contrast to the self-avoiding property of physical polymers), the theoretical probability of this happening is zero. Thus, such a random walk could still provide a good model for a thin and long polymer chain. In fact, this type of random walk is frequently used to model the behaviour of polymers at thermodynamic equilibrium under so-called θ-conditions or in melt phase where polymer segments that are not in direct contact neither attract nor repel each other. Although the overall dimensions of random walks provide important information about the modelled polymers, it is often the case that additional characteristics of polymers need to be investigated. One such characteristic is a measure of polymer entanglement.

The average crossing number (ACN) is a natural geometric measure of polymer entanglement and it corresponds to the average number of crossings that can be perceived while observing the axial trajectory of a given polymer from a random direction [11]. The ACN $a(\ell)$ of a space curve ℓ is given by [8]

$$a(\ell) = \frac{1}{4\pi}\int_\ell\int_\ell \frac{|(\dot\gamma(t), \dot\gamma(s), \gamma(t)-\gamma(s))|}{|\gamma(t)-\gamma(s)|^3}\,\mathrm{d}t\,\mathrm{d}s, \tag{1}$$

where γ is the arclength parametrization of ℓ and $(\dot\gamma(t), \dot\gamma(s), \gamma(t)-\gamma(s)) = (\dot\gamma(t)\times\dot\gamma(s))\cdot(\gamma(t)-\gamma(s))$ is the triple scalar product of $\dot\gamma(t)$, $\dot\gamma(s)$ and $\gamma(t)-\gamma(s)$. Interestingly, it was observed that the speed of electrophoretic migration of knotted DNA molecules of the same size but of various knot types showed a quasi-linear correlation with the ACN values calculated for DNA molecules forming a given knot type [13]. A similar correlation was established between the expected sedimentation coefficient and ACN values of DNA molecules forming different knot types [14]. Furthermore, a correlation was established between the relaxation dynamics of modelled knotted polymers and their ACN values [10]. In the case of protein chains, the ACN provides an interesting measure of their compactness [1] and it was investigated how ACN in proteins scales with the length of polypeptide chain [2, 5].

The mean ACN of a single random walk and polygon was thoroughly studied and well understood in [3]. In this paper, we report further investigations on the geometric properties of random walks and polygons using the mean average inter-crossing number $\langle$ICN$\rangle$, the average of crossing numbers over all orthogonal projections between two random walks or two random polygons. For two spatial curves ℓ_1 and ℓ_2, the average inter-crossing number ICN(ℓ_1, ℓ_2) between them can be calculated using the following formula similar to (1) [8]:

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

where γ_1 and γ_2 are the arclength parametrizations of ℓ_1 and ℓ_2 respectively. The ICN can characterize the extent of entanglement of two (or more) independent polymer chains in abstraction to self-entanglement of each polymer chain. The ICN measure should be a valuable tool for monitoring 'intimacy' of the contacts between individual polypeptide chains in dimeric or multimeric proteins. Each individual polypeptide chain in a multisubunit protein naturally defines the size of the chain for which the ICN can be measured with another polypeptide chain. Measuring ICN changes resulting from different conformational states of active multimeric proteins should shed additional light on global and local rearrangements resulting from functional transitions of studied proteins.

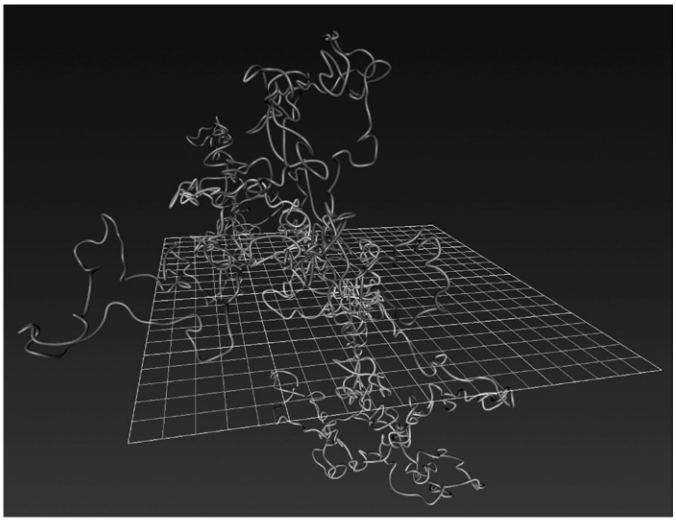

Figure 1. Two randomly generated equilateral random polygons in $\mathbb{R}^3$.

Unlike the linking number Lk between two random polygons, which is a link invariant, the mean average ICN is not a topological invariant. Thus, $\langle \text{ICN} \rangle$ does not (at least not directly) tell us the topological properties of the two random polygons in question, but it does measure an important geometric property of the two components (walks or polygons), which is their spatial closeness. Although it is possible to arrange two (long) walks or polygons so that they are close to each other in space with a very small average ICN, walks and polygons in such special positions are very unlikely when they are randomly placed. Thus, two randomly generated walks or polygons with small average ICN between them are a good indication that the two components are well separated in space. This provides a good measure to study the spatial interaction between different polymer chains. In this case, there is an additional parameter that we have to consider, namely the spatial distance between these two components. In our study, we first generate two random walks with the origin as their starting point, then we choose a random direction and translate one of the random walks in that direction by a distance ρ, where ρ is the preset parameter (distance). We then carry out the computer simulations to compare the result with our theoretical prediction. The case of two random polygons are similarly treated. Figure 1 shows two random polygons of the same length and same starting point generated by this method.

Let X_k be the (three-dimensional) random variable representing the position of the kth vertex of an equilateral random walk EW_n of n edges ($n \geqslant k > 1$), its density function $f_k(X_k)$ is approximately Gaussian and can be estimated by the following formula [4]:

$$\left| f_k(X_k) - \left(\sqrt{\frac{3}{2\pi k}}\right)^3 \exp\left(-\frac{3\,|X_k|^2}{2k}\right) \right| < \frac{0.5}{k^{\frac{5}{2}}}, \qquad k \geqslant 10. \tag{3}$$

In other words, the function $\left| f_k(X_k) - \left(\sqrt{\frac{3}{2\pi k}}\right)^3 \exp\left(-\frac{3|X_k|^2}{2k}\right) \right|$ is uniformly bounded by $\frac{0.5}{k^{\frac{5}{2}}}$ for any X_k and k with $k \geqslant 10$. Note that $f_k(X_k)$ only depends on $r = |X_k|$ (by symmetry) so we also write it as $f_k(r)$ sometimes. In the case of an equilateral random polygon EP_n of n

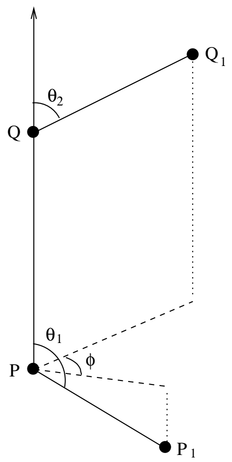

Figure 2. The case of two random edges.

edges, the density function $h_k(X_k)$ of a vertex X_k is still approximately Gaussian, but with a slightly different standard deviation. It can be approximated by the following formula [4]:

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

where $\sigma_{nk}^2 = \frac{k(n-k)}{3n}$. The argument we will use largely depends on the following lemma, which calculates the mean ACN between two unit random edges placed at a distance r apart. Throughout this paper, we will use the notation $O(x)$ for a term bounded between $-M|x|$ and $M|x|$ for some positive constant M that is independent of the variable x.

The lemmas below are needed in the next sections and we state them here without proofs. The proofs of lemmas 1 and 2 can be found in [3] and the proof of lemma 3 is similar to that of lemma 2.

Lemma 1. *Assume that P and Q are two fixed points in $\mathbf{R}^3$ such that $r = |P - Q| \geqslant 4$. 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 2). 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*

$$E(a(\ell_1, \ell_2)) = \frac{1}{16r^2} + O\left(\frac{1}{r^3}\right), \tag{5}$$

where $r = |P - Q|$ and $E(a(\ell_1, \ell_2))$ is the mean of $a(\ell_1, \ell_2)$.

Lemma 2. *If ℓ_1 and ℓ_2 are two edges from an equilateral random walk and there are j edges between ℓ_1 and ℓ_2 along the random walk, then*

$$E(a(\ell_1, \ell_2)) = \frac{3}{16j} + O\left(\frac{\ln j}{j^{\frac{3}{2}}}\right). \tag{6}$$

Lemma 3. *Let ℓ_1 and ℓ_2 be two independent random vectors uniformly distributed on the unit sphere. If the starting points X and Y of the two vectors are placed such that their difference $X - Y$ satisfies the normal distribution with density function*

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{3}{2}} \exp\left(-\frac{|X-Y|^2}{2\sigma^2}\right), \tag{7}$$

then we have

$$E\,(a(\ell_1, \ell_2)) = \frac{1}{16\sigma^2} + O\left(\frac{\ln \sigma}{\sigma^3}\right). \tag{8}$$

2. The case of two random walks

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

Theorem 1. *Let $EW_1(n)$ and $EW_2(n)$ be two equilateral random walks of n edges whose starting points are of a distance $\rho \geqslant 0$. Let $\xi_n(\rho)$ be the ICN between $EW_1(n)$ and $EW_2(n)$, then*

$$E(\xi_n(\rho)) = \frac{3\ln 2}{8} n + \epsilon(n, \rho), \tag{9}$$

where $\epsilon(n, \rho)$ is the error term and it is bounded by $M(\ln n + \rho + \rho^2)\sqrt{n}$ for some positive constant M that is independent of n and ρ.

Proof. Let us consider the case of $\rho = 0$ first. In this case, both equilateral random walks EW_1 and EW_2 will start at the original point. Let e_i be the ith edge of EW_1, e'_j be the jth edge of EW_2 and $a(e_i, e'_j)$ be the ACN between e_1 and e_2, then

$$E(\xi_n(0)) = \sum_{i,j=1}^{n} E(a(e_i, e'_j)). \tag{10}$$

Thus we need to find $E(a(e_i, e'_j))$ for each i, j. Apparently, if $i + j \leqslant 2$, then $a(e_i, e'_j) = 0$ so we only need to consider the case when $i + j \geqslant 3$. Let X_{i-1}, X_i be the end points of e_i and let Y_{j-1}, Y_j be the end points of e'_j, then the sequence $X_{i-1}, X_{i-2}, \ldots, X_1, O, Y_1, \ldots, Y_{j-1}$ can be regarded as an equilateral random walk of $k = i + j - 2$ edges started at point X_{i-1}. In other words, e_i and e'_j can be regarded as two edges from a random walk and there are $k = i + j - 2$ edges between them. By lemma 2, we have

$$E(a(e_i, e'_j)) = \frac{3}{16k} + O\left(\frac{\ln k}{k^{\frac{3}{2}}}\right). \tag{11}$$

It follows that

$$\begin{aligned} E(\xi_n(0)) &= \sum_{i,j=1, i+j\geqslant 3}^{n} E(a(e_i, e'_j)) \\ &= \sum_{i,j=1, i+j\geqslant 3}^{n} \frac{3}{16(i+j-2)} + O\left(\sum_{i,j=1, i+j\geqslant 3}^{n} \frac{\ln(i+j-2)}{(i+j-2)^{\frac{3}{2}}}\right) \\ &= \frac{3}{16}\sum_{k=1}^{n-1} \frac{k+1}{k} + \frac{3}{16}\sum_{k=0}^{n-2} \frac{n-(k+1)}{n+k} + O(\sqrt{n}\ln n) \\ &= \frac{3\ln 2}{8} n + O(\sqrt{n}\ln n). \end{aligned}$$

Note that in the above we used the following estimations:

$$\sum_{k=1}^{n-1} \frac{k+1}{k} = n + O(\ln n), \tag{12}$$

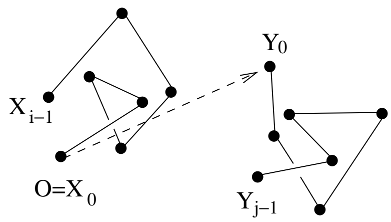

Figure 3. Two equilateral random walks that start at different points.

$$\sum_{k=0}^{n-2}\frac{n-(k+1)}{n+k}=n\int_0^1\frac{1-x}{1+x}+O(1)=(2\ln 2-1)n+O(1) \tag{13}$$

and

$$\sum_{i,j=1,i+j\geqslant 3}^{n}\frac{\ln(i+j-2)}{(i+j-2)^{\frac{3}{2}}}<\sum_{k=1}^{2n-2}\frac{k\ln k}{k^{\frac{3}{2}}}=\sum_{k=1}^{2n-2}\frac{\ln k}{k^{\frac{1}{2}}}=O(\sqrt{n}\ln n). \tag{14}$$

This proves the theorem for the case of $\rho = 0$. Now let us consider the case when $\rho > 0$ and is relatively small compared to n, the length of the random walks.

In this case, EW_1 is generated as before but there is an additional step for generating EW_2. We first generate an equilateral random walk of n edges starting at the original point. We then choose a random direction (that is, a random vector that is uniformly distributed on the unit sphere) and translate EW_2 in this direction for a distance of ρ. In other words, the vector $\vec{\rho}$ is added to each vertex of EW_2 (as a vector). For the sake of convenience, we will still call this translated random walk EW_2. The vertices of EW_1 will still be called X_{i-1} (with $i = 1$ to $n+1$) and the vertices of (the translated) EW_2 will still be called Y_{j-1} ($1 \leqslant j \leqslant n+1$). This situation is illustrated in figure 3. The vertices of EW_2 before the translation, on the other hand, will be denoted by Y'_{j-1}.

As before, if we let $\xi(\rho)$ be the ICN between EW_1 and EW_2, then we still have

$$E(\xi_n(\rho))=\sum_{i,j=1}^{n}E(a(e_i,e'_j)), \tag{15}$$

where e_i is the ith edge from EW_1 and e'_j is the jth edge from (the translated) EW_2. Lemma 2 cannot be applied in this case since the density function of $Y_{j-1} - X_{i-1}$ has changed. Let $r = |Y_{j-1} - X_{i-1}| = |\vec{\rho} + Y'_{j-1} - X_{i-1}|$ and $r_1 = |Y'_{j-1} - X_{i-1}|$. By lemma 1, we have (with fixed $\vec{\rho}$ and r_1)

$$E(a(e_i,e'_j)|\vec{\rho},r_1)=\frac{1}{16r^2}+O\left(\frac{1}{r^3}\right). \tag{16}$$

Let θ be the angle between $\vec{\rho}$ and $Y'_{j-1} - X_{i-1}$, as illustrated in figure 4. We have

$$r^2=r_1^2+\rho^2-2r_1\rho\cos\theta. \tag{17}$$

As we mentioned at the beginning of the last section, the probability density function of θ is $\frac{1}{2}\sin\theta$ and $0 \leqslant \theta \leqslant \pi$. So we have

$$\begin{aligned}E(a(e_i,e'_j)|r_1)&=\int_0^{\pi}\frac{1}{2}\sin\theta\left(\frac{1}{16r^2}+O\left(\frac{1}{r^3}\right)\right)\mathrm{d}\theta\\&=\frac{1}{64r_1\rho}\ln\left(\frac{r_1+\rho}{r_1-\rho}\right)^2+O\left(\frac{1}{r_1\left(r_1^2-\rho^2\right)}\right).\end{aligned}$$

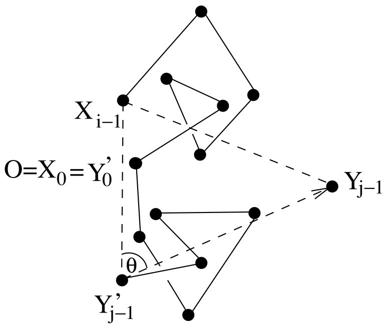

Figure 4. The relation between X_{i-1} and Y_{j-1}.

Since $a(e_i, e'_j) \leqslant 1$, we have $E(a(e_i, e'_j)|r_1) \leqslant 1$ for any r_1. In particular, we can bound $E(a(e_i, e'_j)|r_1)$ by 1 for any $r_1 < 2\rho$. On the other hand, if $r_1 \geqslant 2\rho$, then we have $\ln\left(\frac{r_1+\rho}{r_1-\rho}\right) = \frac{2\rho}{r_1} + O\left(\frac{\rho^2}{r_1^2}\right)$, $\frac{1}{r_1(r_1^2-\rho^2)} = O\left(\frac{1}{r_1^3}\right)$ and

$$E(a(e_i, e'_j)|r_1) = \frac{1}{16r_1^2} + \frac{1}{r_1^3}\left(O(1) + O(\rho)\right). \tag{18}$$

By definition,

$$\begin{aligned} E(a(e_i, e'_j)) &= \int_0^k 4\pi r_1^2 f_k(r_1) E(a(e_i, e'_j)|r_1)\, \mathrm{d}r_1 \\ &\leqslant \int_0^{2\rho} 4\pi r_1^2 f_k(r_1)\, \mathrm{d}r_1 + \int_{2\rho}^k 4\pi r_1^2 f_k(r_1) E(a(e_i, e'_j)|r_1)\, \mathrm{d}r_1, \end{aligned}$$

where

$$f_k(r_1) = \left(\left(\sqrt{\frac{3}{2\pi k}}\right)^3 \exp\left(-\frac{3r_1^2}{2k}\right) + O\left(\frac{1}{k^{\frac{5}{2}}}\right)\right) \tag{19}$$

as defined in (3). The first integral of the above is apparently bounded by $O(\rho^2)\frac{1}{k^{\frac{3}{2}}}$. On the other hand, the second integral above is estimated to be

$$\frac{3}{16k} + O\left(\frac{\ln k}{k^{\frac{3}{2}}}\right) + O(\rho)\frac{1}{k^{\frac{3}{2}}}. \tag{20}$$

The derivation of this estimation involves analysis similar to that used in the proof of theorem 1 in [3]. We leave the details to our reader. Please refer to [3]. Combining the above results, we obtain

$$E(a(e_i, e'_j)) = \frac{3}{16k} + (O(\ln k) + O(\rho) + O(\rho^2))\frac{1}{k^{\frac{3}{2}}}, \tag{21}$$

where $k = i + j - 2 \geqslant 1$. We can now follow the same proof as in the case of $\rho = 0$ and obtain the following result:

$$E(\xi_n(\rho)) = \frac{3\ln 2}{8}n + O\left(\ln n + \rho + \rho^2\right)\sqrt{n}. \tag{22}$$

$\square$

3. The case of two random polygons

Let $EP_1(n)$ and $EP_2(n)$ be two equilateral random polygons of n edges starting from the same point O. Let e_i be the ith edge of EP_1, and e'_j be the jth edge of EP_2. Without loss of generality, we may assume that $i \leqslant (n+1)/2$ and $j \leqslant (n+1)/2$. Let X_{i-1}, X_i be the end points of e_i and let Y_{j-1}, Y_j be the end points of e'_j, then the probability density functions of X_i and Y_j can be approximated by ([4])

$$\left(\sqrt{\frac{1}{2\pi\sigma_{ni}^2}}\right)^3 \exp\left(-\frac{|X_i|^2}{2\sigma_{ni}^2}\right) \tag{23}$$

and

$$\left(\sqrt{\frac{1}{2\pi\sigma_{nj}^2}}\right)^3 \exp\left(-\frac{|Y_j|^2}{2\sigma_{nj}^2}\right) \tag{24}$$

respectively, where $\sigma_{ni}^2 = \frac{i(n-i)}{3n}$ and $\sigma_{nj}^2 = \frac{j(n-j)}{3n}$. In this case the error bound estimation is more complicated and for the sake of simplicity we will only concentrate on obtaining the main terms in the estimation.

Using the above two functions, we see that the joint probability density function of X_i and Y_j can be estimated by

$$\left(\sqrt{\frac{1}{2\pi\sigma_{ni}^2}}\right)^3 \exp\left(-\frac{|X_i|^2}{2\sigma_{ni}^2}\right)\left(\sqrt{\frac{1}{2\pi\sigma_{nj}^2}}\right)^3 \exp\left(-\frac{|Y_j|^2}{2\sigma_{nj}^2}\right). \tag{25}$$

Let $Z = X_i - Y_j$ in the above and integrating over X_i, we obtain an estimation for the pdf of $X_i - Y_j$ of the following form (details are left to our reader):

$$\left(\sqrt{\frac{1}{2\pi\sigma_{nij}^2}}\right)^3 \exp\left(-\frac{|X_i - Y_j|^2}{2\sigma_{nij}^2}\right), \tag{26}$$

where $\sigma_{nij}^2 = \sigma_{ni}^2 + \sigma_{nj}^2 = \frac{i(n-i)}{3n} + \frac{j(n-j)}{3n}$. By lemma 3, we have

$$E(a(e_i, e'_j)) \approx \frac{1}{16\sigma_{nij}^2}. \tag{27}$$

Let ξ'_n be the ICN between $EP_1(n)$ and $EP_2(n)$, then

$$\begin{aligned} E(\xi'_n) &= 4\sum_{i,j=1}^{n/2} E(a(e_i, e'_j)) \\ &\approx n\sum_{i,j=1}^{n/2} \frac{3}{4}\frac{1}{i(n-i)+j(n-j)}. \end{aligned} \tag{28}$$

Let $n \to \infty$ (to extract the main term of the above double series), the above summation converges to the following double definite integral

$$b_1 = \frac{3}{4}\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}} \frac{1}{x(1-x)+y(1-y)}\,\mathrm{d}x\,\mathrm{d}y \approx 0.687. \tag{29}$$

Thus we estimate that $E(\xi'_n) \approx 0.687n$. This leads to our second theorem.

Theorem 2. *Let $EP_1(n)$ and $EP_2(n)$ be two equilateral random polygons of n edges whose starting points are of a distance $\rho \geqslant 0$ where ρ is relatively small. Let $\xi'_n(\rho)$ be the ICN* *between $EP_1(n)$ and $EP_2(n)$, then $E(\xi_n'(\rho))$ can be approximated by $b_1 n \approx 0.687n$, where b_1 is defined by the integral in (29).*

Our simulation yielded a remarkable match to this result. See the simulation section. Even though our theorem only deals with the case when ρ is small, the plot there gives a hint on how $E(\xi_n')$ changes when the distance between the starting points of the two polygons increases.

4. The case of two random walks and polygons with different lengths

Let us consider the case of two random walks with different lengths first.

Theorem 3. *Let EW_1 and EW_2 be two equilateral random walks. Let n be the length of the EW_1 and m be the length of EW_2. Without loss of generality, we may assume that $n \geqslant m$ and we will let $r = \frac{n}{m} \geqslant 1$. If the starting points of EW_1 and EW_2 are of a small distance $\rho \geqslant 0$, then the mean average ICN between EW_1 and EW_2 can be approximated by $a_r m$ where*

$$a_r = \frac{3}{16}\int_0^r \mathrm{d}y \int_0^1 \frac{1}{x+y}\mathrm{d}x = \frac{3}{16}\ln\frac{(1+r)^{1+r}}{r^r} \tag{30}$$

and the error term is of order at most $\sqrt{m}\ln m$.

Proof. Following the proof of theorem 1, it is not hard to see that the mean ICN in this case can be approximated by the following summation

$$\frac{3}{16}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{i+j}. \tag{31}$$

If we rewrite the above summation in the following form

$$\frac{3m}{16}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{\frac{i}{m}+\frac{j}{m}}\frac{1}{m^2}, \tag{32}$$

then we see that as $m \to \infty$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{\frac{i}{m}+\frac{j}{m}}\frac{1}{m^2} \longrightarrow \int_0^r \mathrm{d}y \int_0^1 \frac{1}{x+y}\mathrm{d}x. \tag{33}$$

Evaluating this integral yields the desired result. We omit the error estimate here since it is similar to the proof of theorem 1. □

Remark. Note that as $r \to \infty$, a_r grows proportionally with respect to $\ln r$. Thus, if m is fixed but $r = \frac{n}{m} \to \infty$, then the mean average ICN between EW_1 and EW_2 will go to infinity in the order of $O(\ln r)$.

In the case of two random polygons EP_1 and EP_2 with different lengths, the mean ICN between EP_1 and EP_2 can be derived similar to what we did in the last section, which is of the following form (see equation (28)):

$$\frac{3}{4}\sum_{i=1}^{n/2}\sum_{j=1}^{m/2}\frac{1}{\frac{i(n-i)}{n}+\frac{j(m-j)}{m}} = \frac{3m}{4}\sum_{i=1}^{n/2}\sum_{j=1}^{m/2}\frac{1}{\frac{i}{m}\left(1-\frac{i}{rm}\right)+\frac{j}{m}\left(1-\frac{j}{m}\right)}\frac{1}{m^2}. \tag{34}$$

As $m \to \infty$, the summation at the right-hand side of the above equation approaches the integral

$$\int_0^{r/2} \mathrm{d}y \int_0^{1/2} \frac{1}{x(1-x) + y\left(1 - \frac{y}{r}\right)} \,\mathrm{d}x. \tag{35}$$

Thus, the mean ICN between EP_1 and EP_2 can be approximated by $b_r m$, where

$$b_r = \frac{3}{4} \int_0^{r/2} \mathrm{d}y \int_0^{1/2} \frac{1}{x(1-x) + y\left(1 - \frac{y}{r}\right)} \,\mathrm{d}x. \tag{36}$$

We state this as our last theorem.

Theorem 4. *Let EP_1 and EP_2 be two equilateral random polygons. Let n be the length of the EP_1 and m be the length of EP_2. Without loss of generality, we may assume that $n \geqslant m$ and we will let $r = \frac{n}{m} \geqslant 1$. If the starting points of EP_1 and EP_2 are of a small distance $\rho \geqslant 0$, then the ICN between EP_1 and EP_2 can be approximated by $b_r m$ where*

$$b_r = \frac{3}{4} \int_0^{r/2} \mathrm{d}y \int_0^{1/2} \frac{1}{x(1-x) + y\left(1 - \frac{y}{r}\right)} \,\mathrm{d}x. \tag{37}$$

We end this section with the following remark.

Remark. If we rewrite the integral form of b_r as

$$\frac{3}{16} \int_0^{r} \mathrm{d}y \int_0^{1} \frac{1}{\frac{x}{2}\left(1 - \frac{x}{2}\right) + \frac{y}{2}\left(1 - \frac{y}{2r}\right)} \,\mathrm{d}x$$

and compare it to the integral form of a_r (given in (30)), then it is easy to see that $2a_r < b_r < 4a_r$ in general since for $0 < x < 1$ and $0 < y < 1$ we have

$$\frac{x}{4} < \frac{x}{2}\left(1 - \frac{x}{2}\right) < \frac{x}{2}$$

and

$$\frac{y}{4} < \frac{y}{2}\left(1 - \frac{y}{2r}\right) < \frac{y}{2}.$$

Thus, for fixed m, the mean average ICN between EP_1 and EP_2 will also approach infinity in the order of $O(\ln r)$.

5. Simulation methods

The average inter-crossing number ICN(W_1, W_2) between two random walks (or polygons) W_1 and W_2 is not an invariant. Although the ICN(W_1, W_2) can be (theoretically) calculated by the Gaussian formula

$$\mathrm{ICN}(W_1, W_2) = \frac{1}{2\pi} \int_{W_1} \int_{W_2} \frac{|(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))|}{|\gamma_1(t) - \gamma_2(s)|^3} \,\mathrm{d}t\,\mathrm{d}s \tag{38}$$

(where γ_1 and γ_2 are the arclength parametrizations of W_1 and W_2 respectively), the numerical application of the Gaussian formula sometimes leads to overflow problems when some segments from different components get very close to each other. For this reason our numerical determination for the ICN of random walks and polygons was based on directly counting the number of crossings in numerous projections of analysed trajectories. We calculated the

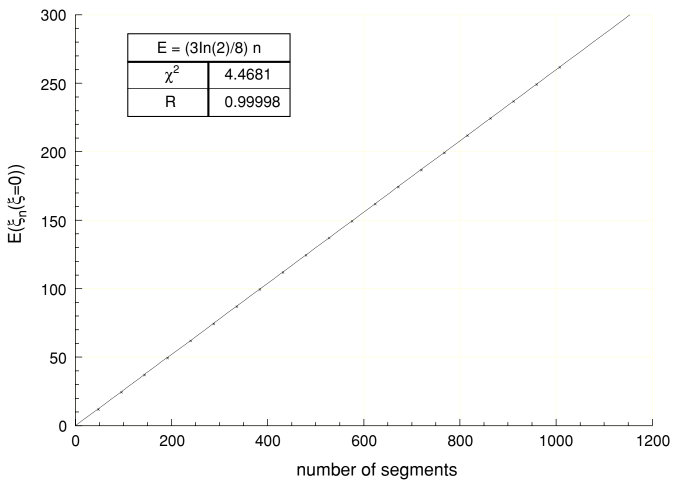

Figure 5. The mean average inter-crossing number between two equilateral random walks of the same length that start at the same origin. The statistical error bars are about the size of the data points and correspond to one standard deviation divided by the square root of the sample size. The χ^2 test statistic and the correlation coefficient R are given in the inserted table.

number of inter-crossings in individual projections of the random walks or polygons and then took the average over 50 randomly chosen directions of projections to obtain a good approximation of the actual ICN value for any given pair of randomly generated walks or polygons.

To generate random equilateral 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.

The random equilateral polygons were generated using the approach of Dykhne [12], namely the hedgehog method. For example, to construct an equilateral random polygon of 100 unit segments, 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.

6. Numerical results

6.1. The case of random walks

Figure 5 shows that $E(\xi_n(\rho = 0))$ (namely the mean average inter-crossing number between two equilateral random walks that are both of length n and both start at the same origin) behaves almost linearly as $\frac{3\ln 2}{8}n$ with very little variation. We have analysed walks with up to 1008 segments and each of the $E(\xi_n(\rho = 0))$ data points was obtained by averaging the ICN values from 10^5 independent random configurations of random walks of the corresponding size. The small variation indicates that the actual correction term to the main term of $E(\xi_n(\rho = 0))$ may be much smaller than the upper bound given in theorem 1.

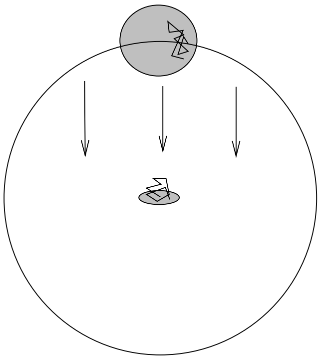

Figure 6. The inter-crossings occur only when the random walk starting on the sphere lies within the grey area. The diameters of the random walks are of the order of $O(\sqrt{n})$ and the radius of the sphere is ρ. The projection direction is indicated by the arrows.

We decided then to test by numerical simulations the ICN behaviour for walks that do not start from the same origin and we investigated the effect of the increasing distance between the origins of the two walks. For $\rho > 0$ that is small comparing to the chain length n, theorem 1 predicted that the ICN grows approximately linear in terms of n. Of course, any fixed ρ will at some point become small compared to the length n of the chains as the chains grow, and the ICN would tend to the linear behaviour $\frac{3\ln 2}{8}n$ eventually. We investigated therefore the ICN behaviour for three different values of $\rho = 20, 40$ and 90, for random walks of length up to 1000. As we will point below, these ρ values are in fact not small compared to the lengths of the random walks in our numerical study.

In order to choose a suitable fitting function, we need to take a look at what happens if ρ is much larger than $\sqrt{n}$. In this case, the two random walks would be (in most cases) far away from each other since the mean diameter of the random walk is of the order of $O(\sqrt{n})$. Thus, their projections would not intersect at all in most projection directions (so the inter-crossing numbers in these projections would all be 0), see figure 6. We can place the starting point of one random walk at the centre of a sphere of radius ρ and place the starting point of the other random walk on the sphere. Since the radius of gyration of a random walk of length n is of the order $O(\sqrt{n})$, it is not hard to see that the area of the projections that could result in inter-crossings between the two random walks is of the order of $O(n)$. It follows that $E(\xi_n(\rho))$ is of the order of $O(\frac{n}{\rho^2})$. On the other hand, if $\sqrt{n}$ is much larger than ρ, then $E(\xi_n(\rho))$ is of the order of $O(n)$ by theorem 1. Thus, we decided to use the function

$$\frac{3\ln 2}{8}n\frac{n+b\rho}{n+a\rho^3} \tag{39}$$

to fit our data. Here a and b are free parameters. Note that in the case that ρ is either 0 or small compared to n, the function provides us the essential linear behaviour in terms of n. For ρ comparable to n or larger than n, the function would provide us the desired behaviour of $O\left(\frac{n}{\rho^2}\right)$, at least in the asymptotic sense. Figures 7, 8 and 9 show the fitting for the cases of $\rho = 20, 40$ and 90 respectively for n up to 1000. Since $\sqrt{1000} \approx 32$, these choices included the cases $\rho \approx \sqrt{n}$ and $\rho \gg \sqrt{n}$ (the case $\rho \ll \sqrt{n}$ is covered by $\rho = 0$ by theorem 1). The nice fitting shown by the figures indicates that the function defined in (39) is a good model for the overall behaviour of $E(\xi_n(\rho))$. Although a and b are free parameters in the fitting, it is likely that the ratio b/a would converge when $\rho > n \to \infty$.

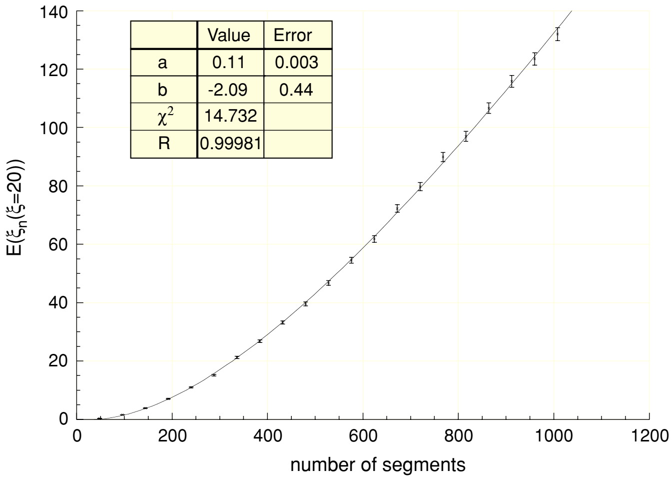

Figure 7. The mean average inter-crossing number between two equilateral random walks of the same length that start at two origins separated by a distance of 20 segments. The error bars correspond to one standard deviation divided by the square root of the sample size. The error margins of the coefficients computed by the fitting program are given in the inserted table.

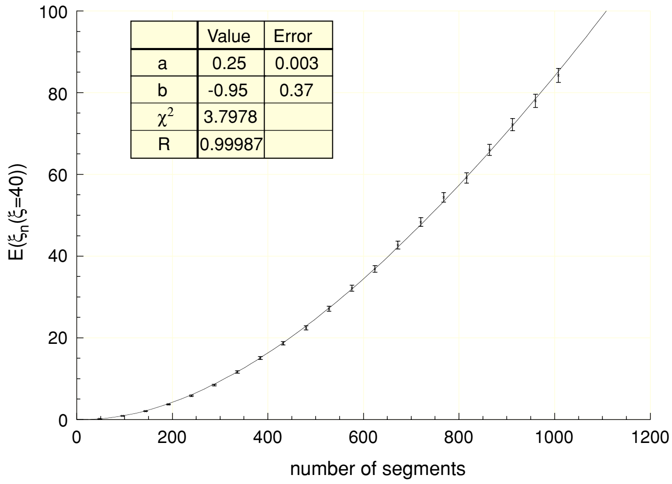

Figure 8. The mean average inter-crossing number between two equilateral random walks of the same length that start at two origins separated by a distance of 40 segments. The error bars correspond to one standard deviation divided by the square root of the sample size.

6.2. The case of random polygons

We have also analysed equilateral random polygons with up to 960 segments and each of the $E(\xi_n'(\rho = 0))$ data points shown in figure 10 was obtained by averaging the ICN values from 10^5 independent random configurations of random polygons of the corresponding size. Figure 10 shows that $E(\xi_n'(\rho = 0))$ also behaves almost linearly with respect to n, the length of the two equal length equilateral random polygons. The small variation indicates that the correction term of $E(\xi_n'(\rho = 0))$ is small compared to n.

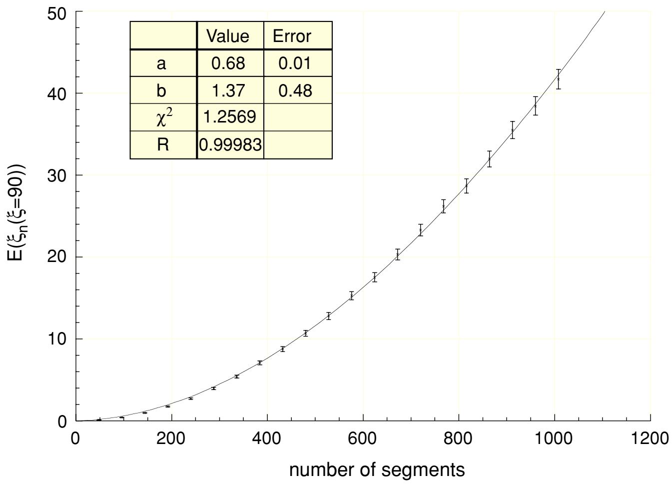

Figure 9. The mean average inter-crossing number between two equilateral random walks of the same length that start at two origins separated by a distance of 90 segments. The error bars correspond to one standard deviation divided by the square root of the sample size.

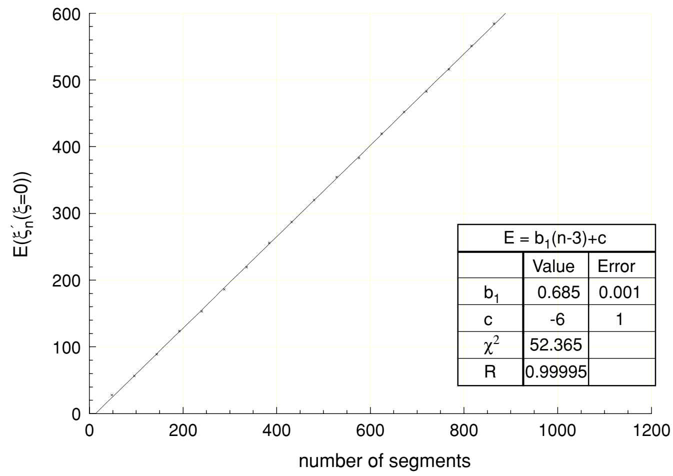

Figure 10. The mean average inter-crossing number between two equilateral random polygons of the same length that start at the same origin. The standard deviation divided by the square root of the sample size is about the size of the data points.

The simulation of the mean ICN between two equal length random polygons starting at different points can be similarly carried out as we did for the random walks in the last subsection. The fitting results are similar and we decide not to include the figures.

6.3. The case of random walks and polygons with different lengths but having the same origin

We generated random samples of walks and polygons of lengths $100, 200, \ldots, 1000$. In table 1, $\overline{\mathrm{ICN}^{\mathrm{r}}_{\mathrm{w}}}$ ($\overline{\mathrm{ICN}^{\mathrm{r}}_{\mathrm{p}}}$) stands for the mean average ICN between the random walks (polygons) of length 100 and the random walks (polygons) of length $100r$ ($r = 2, 3, \ldots, 10$) from the random sample and each sample contain 10 000 independent random pairs of configurations

Table 1. $\overline{\mathrm{ICN}_{\mathrm{w}}^{\mathrm{r}}}$ is the numerical mean averaged ICN between the random walks of length 100 and those of length $100r$. Similarly, $\overline{\mathrm{ICN}_{\mathrm{p}}^{\mathrm{r}}}$ is the numerical mean averaged ICN between the random polygons of length 100 and those of length $100r$. The intervals given are at the 95% confidence level.

	$\mathrm{ICN}_{\mathrm{w}}^{\mathrm{r}}$		$\mathrm{ICN}_{\mathrm{p}}^{\mathrm{r}}$	
r	$100a_r$	$\overline{\mathrm{ICN}_{\mathrm{w}}^{\mathrm{r}}}$	$100b_r$	$\overline{\mathrm{ICN}_{\mathrm{p}}^{\mathrm{r}}}$
2	35.80	35.34 ± 0.52	94.06	87.49 ± 0.87
3	42.18	41.74 ± 0.61	109.87	100.56 ± 1.01
4	46.91	46.17 ± 0.67	121.29	112.81 ± 1.13
5	50.69	50.18 ± 0.73	130.20	121.38 ± 1.17
6	53.82	52.68 ± 0.76	137.48	124.43 ± 1.25
7	56.52	55.48 ± 0.80	143.63	134.58 ± 1.31
8	58.87	58.32 ± 0.84	148.96	136.56 ± 1.35
9	60.95	60.40 ± 0.87	153.64	146.90 ± 1.39
10	62.83	61.98 ± 0.88	157.83	147.20 ± 1.44

of random walks or polygons of the corresponding sizes. The average ICN between each pair of random walks (polygons) is obtained by taking the average over 50 randomly chosen directions of projections as before. According to theorems 3 and 4, the mean ICN in these cases can be approximated by $a_r m$ and $b_r m$, where $r = 2, 3, \ldots, 10$ and $m = 100$. These theoretical values are listed in table 1 and besides them are the numerical results obtained from the simulations.

As one can see, the numerical results are mostly in line with the theoretical prediction in the random walk case. In the random polygon case, there is a visible gap though the overall trend of the numerical result matches the theoretical result. One can also observe the statistical fluctuations apparently caused by the sample at $r = 10$ in the case of random walks and at $r = 6$ in the case of random polygons. Although it is possible to improve this by increasing the sample size, it is also possible that the error correction terms in these cases are more significant. Furthermore, since the coefficients a_r and b_r are obtained in the asymptotic case, the size we chose for the shorter random walk and polygon (100) may not be large enough to accurately reflect the error correction terms. So a more in depth simulation may be needed in order to see a better fit.

7. Conclusions

We have provided an analytical proof that for long equilateral random walks, the average inter-crossing number between two random walks (calculated over all statistical ensembles of walks with the same number of segments (n)) can be expressed by the formula $\frac{3\ln 2}{8}n$, plus a correction term. This formula also works if the random walks are generated from different starting points that are not far away from each other. In the case that the two random walks are of different lengths, the average inter-crossing number between two random walks (calculated over all statistical ensembles of walks with m segments and n segments respectively such that $n/m = r \geqslant 1$) can be expressed by the formula $a_r n$, plus a correction term, where a_r is defined by $a_r = \frac{3}{16}\ln\frac{(1+r)^{1+r}}{r^r}$. In the case of random polygons, the corresponding mean average inter-crossing number can be expressed as $b_r n$ (plus a correction term) where b_r is defined explicitly by a definite integral and has the property $2a_r < b_r < 4a_r$. Subsequently, we have used numerical simulations to demonstrate that the analytically predicted scaling of the mean average ICN between random walks and polygons with equal number of segments holds not only for long chains, but also for short ones. In the case that the random walks are generated from different starting points that are of a distance $\rho > 0$ apart, our numerical simulation indicates that the mean average ICN behaves as $\frac{3\ln 2}{8} n \frac{n+b\rho}{n+a\rho^3}$ where a and b are free parameters. We end this paper by asking the following interesting question for future studies: If the random polygons generated are from certain knot families, how would that affect the overall behaviour of the mean average ICN among these random polygons (or other polygons)?

Acknowledgments

The authors thank Jacques Dubochet for his interest and encouragement. This work was supported in part by NSF grant DMS-0310562 to Y Diao, by Swiss National Science Foundation Grants PA00A-105000 to A Dobay and by Swiss National Science Foundation Grants 31-68151.02 to A Stasiak and 3100A0-103962 to A Stasiak and J Dubochet.

References

[1] Arteca G 1995 *Phys. Rev.* E **51** 2600–10
[2] Arteca G 1997 *Phys. Rev.* E **56** 4516–20
[3] Diao Y, Dobay A, Kusner R, Millett K and Stasiak A 2003 *J. Phys. A: Math. Gen.* **36** 11561–74
[4] Diao Y, Nardo J and Sun Y 2001 *J. Knot Theory Ramifications* **10** 597–607
[5] Dobay A, Diao Y, Dubochet J and Stasiak A 2005 *Physical and Numerical Models in Knot Theory Including Applications to the Life Sciences* ed J A Calvo, K C Millett, E J Rawdon and A Stasiak (Singapore: World Scientific) pp 219–31 (at press)
[6] Doi M and Edwards S F 1986 *The Theory of Polymer Dynamics* (Oxford: Oxford University Press)
[7] Flory J P 1953 *Principles of Polymer Chemistry* (Ithaca, NY: Cornell University Press)
[8] Freedman M H and He Z 1991 *Ann. Math.* **134** 189–229
[9] de Gennes P G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[10] Huang J Y and Lai P Y 2001 *Phys. Rev.* E **63** 021506-1–6
[11] Katritch V, Bednar J, Michoud D, Scharein R G, Dubochet J and Stasiak A 1996 *Nature* **384** 142–5
[12] Klenin K V, Vologodskii A V, Anshelevich V V, Dykhne A M and Frank-Kamenetskii M D 1988 *J. Biomol. Struct. Dyn.* **5** 1173
[13] Stasiak A, Katritch V, Bednar J, Michoud D and Dubochet J 1996 *Nature* **384** 122
[14] Vologodskii A V, Crisona N, Laurie B, Pieranski P, Katritch V, Dubochet J and Stasiak A 1998 *J. Mol. Biol.* **278** 1–3