The Average Crossing Number of Gaussian Random Walks and Polygons

Yuanan Diao and Claus Ernst

Abstract. In this paper, we extend results about the average crossing number of equilateral random walks and polygons to the average crossing number of the Gaussian random walks and polygons. We show that the asymptotical behavior of the ACN for the two models are very similar. More precisely, we show that the mean average crossing number ⟨ACN⟩ of Gaussian random walks and polygons of length n is of the form $\frac{1}{2\pi} n \ln n + O(n)$.

1. Introduction

Random walks and random polygons are frequently used to model linear long-chain molecules or ring polymers. They are useful in the study of the behavior of polymers at thermodynamic equilibrium. The two most frequently used continuum models are the equilateral random walks and the Gaussian random walks. An equilateral random walk is composed of freely jointed line segments of equal length, whereas a Gaussian random walk is composed of freely jointed line segments whose length follow the Gaussian distribution. Numerous studies are devoted to the topological properties of random polygons. For example, it is investigated what types of knots are formed on polymer chains [1, 11, 14], and it is theoretically proven that knotting becomes inevitable when the length of a random polygon approaches infinity [2, 5, 12, 15]. Furthermore, for the equilateral and Gaussian random polygons, it has been shown that knots in the global sense do exist with a high probability if the length of the polygon is large [4, 10]. However, while the overall dimensions of equilateral and Gaussian random walks (polygons) are well understood (they scale with the number of segments n as $\sqrt{n}$ [6, 7, 9]), the problem is much harder and less is known for random polygons with certain topological constraints (such as a fixed knot type).

A Gaussian random vector $X = (x, y, z)$ is a random point whose coordinates x, y and z are independent standard normal random variables (with mean $= 0$ and

---

1991 *Mathematics Subject Classification.* Primary 57M25.

Key words and phrases. Keywords: Random Walks, Random Polygons, Crossing Number, Average Crossing Number, Gaussian Random Walks, Gaussian Random Polygons.

The authors are partially supported by NSF grant DMS-0310562.

variance $= 1$). The pdf (probability density function) of X is the joint pdf of x, y and z, which is

$$\begin{aligned} f(X) &= \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{x^2+y^2+z^2}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{|X|^2}{2}}. \end{aligned}$$

A Gaussian random walk of n steps consists of $n+1$ consecutive points $X_0 = (0,0,0) = O,\ X_1,\ X_2,\ \dots,\ X_n$ such that $Y_{k+1} = X_{k+1} - X_k$ $(k = 0, 1, ..., n-1)$ are independent Gaussian random vectors. It follows that the joint pdf for all the vertices is

$$\begin{aligned} f(X_1, X_2, ..., X_n) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} e^{-\frac{1}{2}(|Y_1|^2+|Y_2|^2+\cdots+|Y_n|^2)} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} e^{-\frac{1}{2}(|X_1|^2+|X_2-X_1|^2+\cdots+|X_n-X_{n-1}|^2)}. \end{aligned}$$

In this paper, we are interested in a particular measure of a polymer entanglement, i.e., the average crossing number of a random walk or a random polygon. A Gaussian random polygon is a conditioned random walk such that the last vertex coincides with the starting point, which can be always assumed to be the original point. For a given random walk (or random polygon), the crossing number associated with a particular projection of the random walk is the number of crossings one observes when the walk (or polygon) is projected to a plane under the given projection direction. The average crossing number of the walk (or polygon) is then defined as the average of this crossing number over all possible projection directions (which is equivalent to the unit sphere). In [3], it is shown that the mean average crossing number ($\langle\text{ACN}\rangle$) of all equilateral random walks and polygons of length n is of the form $\frac{3}{16} \cdot n \cdot \ln n + O(n)$. We will study the same problem for the Gaussian random walks and polygons in this paper. Since the end to end distance behavior of two vertices along an equilateral random walk is approximately Gaussian ([13]), it is probably not surprising that the mean ACN for the Gaussian random walks and polygons would also follow the $n \ln n$ growth rule (although the difficulties one encounter in the two cases are different technically). As in the case of the equilateral random walks and polygons, we are also able to determine the coefficient of the $n \ln n$ term for the Gaussian random walks and polygons. More precisely, we will show that the mean ACN's for all the Gaussian random walks and polygons of length n are both of the form $\frac{1}{2\pi} n \ln n + O(n)$.

2. The Gaussian Random Walks and Polygons

As defined in the last section, a Gaussian random vector $X = (x, y, z)$ is a random point whose coordinates x, y and z are independent standard normal random variables (with mean $= 0$ and variance $= 1$) and a Gaussian random walk of n steps consists of $n+1$ consecutive points $X_0 = (0,0,0) = O,\ X_1,\ X_2,\ \dots,\ X_n$ such that $Y_{k+1} = X_{k+1} - X_k$ $(k = 0, 1, ..., n-1)$ are n independent Gaussian random vectors.

A Gaussian random walk of n steps is denoted by GW_n. The line segment that joins X_{k-1} and X_k (which has length $|Y_k| = |X_k - X_{k-1}|$) is called the k-th step of GW_n. A Gaussian random polygon GP_n is a conditioned GW_n of n steps such that the last vertex X_n coincides with the starting point $X_0 = O$. Thus, if we let $g_n(X_n)$ be the pdf of X_n for a GW_n, then the joint pdf of X_1, X_2, ..., X_{n-1} of a GP_n is

$$h(X_1, X_2, ..., X_n) = f(X_1, X_2, ..., X_n)/g_n(O).$$

Let A be a three by three real orthonormal matrix. Under the transformation $(x, y, z) \longrightarrow (x, y, z)A^T = (x', y', z')$ (where A^T is the transpose of A), the new variables x', y' and z' are apparently also independent standard normal random variables. So the vector $X' = (x', y', z')$ is also a Gaussian random vector. If we let $r = |X| = |X'|$, θ be the angle between X' and the z'-axis ($0 \leq \theta \leq \pi$) and ϕ be the angle between X' and the positive x'-axis ($0 \leq \phi < 2\pi$), then one can show that the random variables r, θ and ϕ are also independent. Furthermore, the probability density functions of r, θ and ϕ are given by

$$f_1(r) = \sqrt{\frac{2}{\pi}} r^2 e^{-\frac{r^2}{2}}, \tag{2.1}$$

$$f_2(\theta) = \frac{1}{2}\sin\theta, \text{ and} \tag{2.2}$$

$$f_3(\phi) = \frac{1}{2\pi} \tag{2.3}$$

respectively.

In the following, we will list several probability density functions that we will be using in the next section. These results are all well known and we will leave the proofs of these formulas to our reader since they are straight forward.

The pdf of the k-th vertex X_k (not to be confused with $Y_k = X_k - X_{k-1}$) in a GW_n is given by $g_k(X_k) = \int\int \ldots \int f(X_1)f(X_2 - X_1)f(X_3 - X_2)\ldots f(X_k - X_{k-1})dX_1dX_2\ldots dX_{k-1}$. From this one obtains:

$$g_k(X_k) = \left(\frac{1}{\sqrt{2\pi k}}\right)^3 e^{-\frac{|X_k|^2}{2k}}. \tag{2.4}$$

Furthermore, the pdf of the random vector $X_{j+k} - X_j$ (where X_j and X_{j+k} are vertices from the same GW_n) is given by $g_k(X_{j+k} - X_j) = \int\int \ldots \int f(X_{j+1} - X_j)f(X_{j+2} - X_{j+1})\ldots f(X_{j+k} - X_{j+k-1})dX_{j+1}dX_{j+2}\ldots X_{j+k-1}$, which is,

$$g_k(X_{j+k} - X_j) = \left(\frac{1}{\sqrt{2\pi k}}\right)^3 e^{-\frac{|X_{j+k}-X_j|^2}{2k}}. \tag{2.5}$$

On the other hand, if the above vertices are from a GP_n, then the pdf of X_k is given by $h_k(X_k) = \int\int \ldots \int \frac{1}{g_n(O)} f(X_1)f(X_2 - X_1)f(X_3 - X_2)\ldots f(O - X_{n-1})dX_1\ldots dX_{k-1}dX_{k+1}\ldots dX_{n-1}$. From this integral one obtains:

$$h_k(X_k) = \left(\frac{1}{\sqrt{2\pi}\sigma_{n,k}}\right)^3 e^{-\frac{|X_k|^2}{2\sigma^2_{n,k}}}, \tag{2.6}$$

where $\sigma_{n,k} = \sqrt{\frac{k(n-k)}{n}}$. Similarly, the pdf of $X_{j+k} - X_j$ in a GP_n is:

$$h_k(X_{j+k} - X_j) = \left(\frac{1}{\sqrt{2\pi}\sigma_{n,k}}\right)^3 e^{-\frac{|X_{j+k}-X_j|^2}{2\sigma_{n,k}^2}}. \tag{2.7}$$

Finally, the joint pdf of X_1, X_{k+1} and X_{k+2} in a GP_n is

$$\begin{aligned} & h'(X_1, X_{k+1}, X_{k+2}) \\ = & \int\int\ldots\int \frac{f(X_1)f(X_2-X_1)\ldots f(O-X_{n-1})}{g_n(O)} dX_2\ldots dX_k dX_{k+3}\ldots dX_{n-1}) \\ = & \left(\sqrt{2\pi n}\right)^3 f(X_1)f(X_{k+2}-X_{k+1})g_k(X_{k+1}-X_1)g_{n-k-2}(X_{k+2}) \\ = & \left(\sqrt{2\pi n}\right)^3 \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{|X_1|^2}{2}} \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{|X_{k+2}-X_{k+1}|^2}{2}} \cdot \\ & \left(\frac{1}{\sqrt{2\pi k}}\right)^3 e^{-\frac{|X_{k+1}-X_1|^2}{2k}} \cdot \left(\frac{1}{\sqrt{2\pi(n-k-2)}}\right)^3 e^{-\frac{|X_{k+2}|^2}{2(n-k-2)}}. \end{aligned} \tag{2.8}$$

3. The Main Results and their Proofs

The following two theorems are the main results of this paper.

Theorem 3.1. *Let χ_n be the ACN of a Gaussian random walk of n steps; then*

$$E(\chi_n) = \frac{1}{2\pi} n \ln n + O(n).$$

Theorem 3.2. *Let χ'_n be the ACN of a Gaussian random polygon of n steps; then*

$$E(\chi'_n) = \frac{1}{2\pi} n \ln n + O(n).$$

The following lemma is key to the proofs of the theorems.

Lemma 3.3. *Let P, Q, P_1 and Q_1 be four points in $\mathbf{R}^3$ such that P, Q are fixed and $P_1 - P$, $Q_1 - Q$ are two independent Gaussian random vectors. 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*

$$E(a(\ell_1, \ell_2)) = \frac{1}{2\pi r^2} + O(\frac{1}{r^{2.5}}), \tag{3.1}$$

where $r = |P - Q|$.

Proof. Without loss of generality, let us assume that $P = O$ and Q is on the positive z-axis (recall from last section that a rotation of the 3 space transforms independent Gaussian random vectors to independent Gaussian random vectors). Let θ_1 be the angle between $V_1 = \overrightarrow{PP_1}$ and the z-axis and θ_2 be the angle between $V_2 = \overrightarrow{QQ_1}$ and the z-axis. Furthermore, let φ be the angle between the projections of V_1 and V_2 on the xy-plane. See Figure 1 below.

Notice that φ is the angle between two plane vectors and its measure is between 0 and π. We leave it to the reader to show that φ is a random variable uniformly

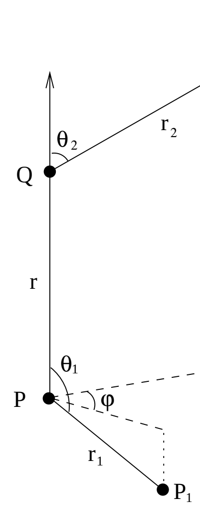

FIGURE 1. The ACN of two random edges

distributed on the interval $[0, \pi]$. In [8], 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

$$a(\ell_1, \ell_2) = \frac{1}{2\pi} \int_{\ell_1} \int_{\ell_2} \frac{|(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))|}{|\gamma_1(t) - \gamma_2(s)|^3} dtds, \tag{3.2}$$

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

$$\begin{aligned}
\gamma_1(t) &= t \cdot \frac{V_1}{|V_1|},\ 0 \le t \le |V_1|, \\
\gamma_2(s) &= \overrightarrow{OQ} + s \cdot \frac{V_2}{|V_2|},\ 0 \le s \le |V_2|.
\end{aligned}$$

By an elementary calculation, we have

$$\begin{aligned}
&\int_{\ell_1} \int_{\ell_2} |(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))| dtds \\
= &\int_0^{|V_1|} \int_0^{|V_2|} |(\frac{V_1}{|V_1|}, \frac{V_2}{|V_2|}, \overrightarrow{OQ})| dtds \\
= &|(V_1, V_2, \overrightarrow{OQ})| = r r_1 r_2 \sin\varphi \sin\theta_1 \sin\theta_2,
\end{aligned}$$

where $r = |\overrightarrow{OQ}|$, $r_1 = |V_1|$ and $r_2 = |V_2|$. Recall from the last section that θ_1 and θ_2 are independent and their pdf's are $\frac{1}{2}\sin\theta_1$ and $\frac{1}{2}\sin\theta_2$ respectively. For r large enough, if r_1, $r_2 \le \sqrt{r}$, then we have $r - 2\sqrt{r} \le |\gamma_1(t) - \gamma_2(s)| \le r + 2\sqrt{r}$. From this we obtain the following:

$$\frac{1}{|\gamma_1(t) - \gamma_2(s)|^3} = \frac{1}{r^3} + O(\frac{1}{r^{3.5}}).$$

It follows that

$$a(\ell_1,\ell_2) = \frac{1}{2\pi r^2} r_1 r_2 \sin\varphi \sin\theta_1 \sin\theta_2 + r_1 r_2 O(\frac{1}{r^{2.5}}) \tag{3.3}$$

and

$$\begin{aligned} & E(a(\ell_1,\ell_2)) \\ = & \int\int a(\ell_1,\ell_2) f(V_1) f(V_2) dV_1 dV_2 \\ = & \int_0^\pi \int_0^\pi \int_0^\pi \int_0^\infty \int_0^\infty a(\ell_1,\ell_2) p(\theta_1,\theta_2,\varphi,r_1,r_2) d\theta_1 d\theta_2 d\varphi dr_1 dr_2, \end{aligned} \tag{3.4}$$

where

$$p(\theta_1,\theta_2,\varphi,r_1,r_2) = \frac{1}{2\pi^2} \sin\theta_1 \sin\theta_2 r_1^2 r_2^2 e^{-\frac{r_1^2+r_2^2}{2}} \tag{3.5}$$

is the joint pdf of θ_1, θ_2, φ, r_1 and r_2. Thus, $E(a(\ell_1,\ell_2))$ can be estimated by splitting the integral in (3.4) in two parts: One with both r_1, $r_2 \le \sqrt{r}$ (which we will call I_1) and the other with either $r_1 > \sqrt{r}$ or $r_2 > \sqrt{r}$, which we will call I_2. Since $a(\ell_1,\ell_2) \le 1$, and if one assumes that $r_1 > \sqrt{r}$ one can bound I_2 from above as follows:

$$\begin{aligned} I_2 \le & \int_0^\pi \int_0^\pi \int_0^\pi \int\int_{r_1 \text{ or } r_2 > \sqrt{r}} p(\theta_1,\theta_2,\varphi,r_1,r_2) d\theta_1 d\theta_2 d\varphi dr_1 dr_2 \\ = & \frac{2}{\pi} \int\int_{r_1 \text{ or } r_2 > \sqrt{r}} r_1^2 r_2^2 e^{-\frac{r_1^2+r_2^2}{2}} dr_1 dr_2 \\ = & \frac{4}{\pi} \int_{\sqrt{r}}^\infty r_1^2 e^{-\frac{r_1^2}{2}} dr_1 \int_0^\infty r_2^2 e^{-\frac{r_2^2}{2}} dr_2 \\ = & 2\sqrt{\frac{2}{\pi}} \int_{\sqrt{r}}^\infty r_1^2 e^{-\frac{r_1^2}{2}} dr_1 \\ = & 2\sqrt{\frac{2}{\pi}} \int_{\sqrt{r}}^\infty e^{-\frac{r_1^2}{4}} (r_1^2 e^{-\frac{r_1^2}{4}}) dr_1 \\ \le & 2\sqrt{\frac{2}{\pi}} e^{-\frac{r}{4}} \int_{\sqrt{r}}^\infty r_1^2 e^{-\frac{r_1^2}{4}} dr_1 \\ = & O(e^{-\frac{r}{4}}) \end{aligned}$$

Hence I_2 is bounded above by $O(\frac{1}{r^{2.5}})$. On the other hand, using (3.3) for $a(\ell_1,\ell_2)$, we split I_1 into two integrals. The error term $O(\frac{1}{r^{2.5}})$ results the following:

$$O(\frac{1}{r^{2.5}}) \int_0^\pi \int_0^\pi \int_0^\pi \int_0^{\sqrt{r}} \int_0^{\sqrt{r}} r_1 r_2 p(\theta_1,\theta_2,\varphi,r_1,r_2) d\theta_1 d\theta_2 d\varphi dr_1 dr_2 = O(\frac{1}{r^{2.5}}).$$

The remaining integral

$$\int_0^\pi \int_0^\pi \int_0^\pi \int_0^{\sqrt{r}} \int_0^{\sqrt{r}} \frac{1}{4\pi^3 r^2} \sin\varphi \sin^2\theta_1 \sin^2\theta_2 r_1^3 r_2^3 e^{-\frac{r_1^2+r_2^2}{2}} d\theta_1 d\theta_2 d\varphi dr_1 dr_2$$

can be written as the product of the following five integrals:

$$\frac{1}{4\pi^3 r^2}\int_0^{\pi}\sin\varphi d\varphi = \frac{1}{2\pi^3 r^2},$$
$$\int_0^{\pi}\sin^2\theta_1 d\theta_1 = \int_0^{\pi}\sin^2\theta_2 d\theta_2 = \frac{\pi}{2},$$

and

$$\int_0^{\sqrt{r}} r_1^3 e^{-\frac{r_1^2}{2}} dr_1 = \int_0^{\sqrt{r}} r_2^3 e^{-\frac{r_2^2}{2}} dr_2 = 2-(2+r)e^{-r/2} = 2+O(\frac{1}{r^{2.5}}).$$

The result now follows. □

We will now give a proof for Theorem 3.1.

PROOF. Let ℓ_k be the k-th segment of a GW_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 for the average crossing number χ_n of GW_n

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

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

$$E(\chi_n) = \sum_{3\le j\le n}(n-j+1)E(a(\ell_1,\ell_j)), \tag{3.6}$$

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

$$E(a(\ell_1,\ell_j)|X_{j-1}-X_1) = \frac{1}{4\pi r_j^2} + O(\frac{1}{r_j^{2.5}})$$

for any fixed X_j such that $r_j \ge 10$ by Lemma 3.3. 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 pdf as X_{j-2}, it follows that

$$\begin{aligned}
E(a(\ell_1,\ell_j)) &= \int E(a(\ell_1,\ell_j)|X_{j-1}-X_1)g_{j-2}(X_{j-1}-X_1)d(X_{j-1}-X_1)\\
&= \int_0^{\infty} E(a(\ell_1,\ell_j)|X_{j-1}-X_1)4\pi r_j^2(\frac{1}{\sqrt{2\pi(j-2)}})^3 e^{-\frac{r_j^2}{2(j-2)}}dr_j\\
&= I_1 + I_2,
\end{aligned}$$

where

$$I_1 = \int_{10}^{\infty} E(a(\ell_1,\ell_j)|X_{j-1}-X_1)4\pi r_j^2(\frac{1}{\sqrt{2\pi(j-2)}})^3 e^{-\frac{r_j^2}{2(j-2)}}dr_j$$

and

$$I_2 = \int_0^{10} E(a(\ell_1,\ell_j)|X_{j-1}-X_1)4\pi r_j^2(\frac{1}{\sqrt{2\pi(j-2)}})^3 e^{-\frac{r_j^2}{2(j-2)}}dr_j.$$

Since $E(a(\ell_1,\ell_j)|X_{j-1}-X_1)$ is bounded above by one, it is easy to see that I_2 is bounded above as follows:

$$I_2 \le \int_0^{10} 4\pi r_j^2 (\frac{1}{\sqrt{2\pi(j-2)}})^3 e^{-\frac{r_j^2}{2(j-2)}}\, dr_j = O(\frac{1}{j^{1.5}}) \int_0^{10} r_j^2 e^{-\frac{r_j^2}{2(j-2)}}\, dr_j = O(\frac{1}{j^{1.5}}).$$

On the other hand, using

$$E(a(\ell_1,\ell_j)|X_{j-1}-X_1) = \frac{1}{2\pi r_j^2} + O(\frac{1}{r_j^{2.5}})$$

from Lemma 3.3, we can split I_1 into two integrals. The one containing the error term $O(\frac{1}{r_j^{2.5}})$ is bounded above as follows:

$$\begin{aligned}
& 4\pi(\frac{1}{\sqrt{2\pi(j-2)}})^3 \int_{10}^{\infty} O(\frac{1}{\sqrt{r_j}}) e^{-\frac{r_j^2}{2(j-2)}}\, dr_j \\
\le\ & O(\frac{1}{(j-2)^{1.5}}) \int_0^{\infty} O(\frac{1}{\sqrt{r_j}}) e^{-\frac{r_j^2}{2(j-2)}}\, dr_j \\
=\ & O(\frac{1}{(j-2)^{1.5}}) O((j-2)^{.25}) \\
=\ & O(\frac{1}{(j-2)^{1.25}}).
\end{aligned}$$

Finally, the remaining integral is estimated as:

$$\begin{aligned}
& 2(\frac{1}{\sqrt{2\pi(j-2)}})^3 \int_{10}^{\infty} e^{-\frac{r_j^2}{2(j-2)}}\, dr_j \\
=\ & (\frac{1}{\sqrt{2\pi(j-2)}})^3 \int_{-\infty}^{\infty} e^{-\frac{r_j^2}{2(j-2)}}\, dr_j - (\frac{1}{\sqrt{2\pi(j-2)}})^3 \int_{-10}^{10} e^{-\frac{r_j^2}{2(j-2)}}\, dr_j \\
=\ & (\frac{1}{\sqrt{2\pi(j-2)}})^3 \sqrt{2\pi(j-2)} + O(\frac{1}{j^{1.5}}) \\
=\ & \frac{1}{2\pi(j-2)} + O(\frac{1}{j^{1.5}}) = \frac{1}{2\pi j} + O(\frac{1}{j^{1.5}}).
\end{aligned}$$

Combining the above bounds, we get

$$E(a(\ell_1,\ell_j)) = \frac{1}{2\pi j} + O(\frac{1}{j^{1.25}}).$$

So

$$E(\chi_n) = \frac{1}{2\pi} n \sum_{3\le j\le n} \frac{1}{j} - \frac{1}{2\pi}(n-3) + nO(\sum_{3\le j\le n} \frac{1}{j^{1.25}})$$

by (3.6). The result of Theorem 3.1 follows since $\sum_{3\le j\le n} \frac{1}{j} - \ln n$ and $\sum_{3\le j\le n} \frac{1}{j^{1.25}}$ both converge to a constant independent of n. This finishes the proof of Theorem 3.1. □

Finally, we will prove Theorem 3.2.

PROOF. Let χ'_n be the ACN of a Gaussian 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 a GP_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

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

We have $E(a(\ell_{i_1}, \ell_{j_1})) = E(a(\ell_{i_2}, \ell_{j_2}))$ whenever $|j_1 - i_1| = |j_2 - i_2|$ or $|j_1 - i_1| = n - |j_2 - i_2|$ by symmetry. It follows that

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

j starts at 3 in the above formula since $a(\ell_1, \ell_2)$ is always 0. Let $r_j = |X_{j-1} - X_1|$. We have

$$E(a(\ell_1, \ell_j)) = \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} a(\ell_1, \ell_j) h'(X_1, X_{j-1}, X_j) dX_1 dX_{j-1} dX_j, \tag{3.7}$$

where

$$h'(X_1, X_{j-1}, X_j) = \left(\sqrt{2\pi n}\right)^3 f(X_1) f(X_j - X_{j-1}) g_{j-2}(X_{j-1} - X_1) g_{n-j}(X_j) \tag{3.8}$$

is the joint pdf of X_1, X_{j-1} and X_j (see (2.8)). In the proof of Theorem 3.1 Lemma 3.3 was used to estimate $E(a(\ell_1, \ell_j))$. However we cannot apply Lemma 3.3 directly since the random vectors X_1, $X_j - X_{j-1}$ and X_j are not independent Gaussian random vectors in the case of a polygon GP_n. In the following, we will show that the pdf $h'(X_1, X_{j-1}, X_j)$ in (3.7) can be replaced by $f(X_1) f(X_j - X_{j-1}) h_j(X_j)$ (which is the pdf of the three corresponding independent random vectors X_1, $X_j - X_{j-1}$ and X_j), while keeping the error term small enough so the result of the theorem will still hold.

By (2.6), the pdf of X_j is

$$h_j(X_j) = \left(\frac{1}{\sqrt{2\pi}\sigma_{n,j}}\right)^3 e^{-\frac{|X_j|^2}{2\sigma_{n,j}^2}}, \tag{3.9}$$

where $\sigma_{n,j} = \sqrt{j}\sqrt{\frac{n-j}{n}} = O(j^{0.5})$ for $j \leq (n+1)/2$. It follows that $(\frac{1}{\sqrt{2\pi}\sigma_{n,j}})^3 = O(j^{-1.5})$ and therefore $P(|X_j| \leq 10) = O(j^{-1.5})$. By the fact that $a(\ell_1, \ell_j) \leq 1$, we get the following inequality:

$$\begin{aligned}
& \int_{|X_j| \leq 10} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} a(\ell_1, \ell_j) h'(X_1, X_{j-1}, X_j) dX_1 dX_{j-1} dX_j \\
\leq & \int_{|X_j| \leq 10} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} h'(X_1, X_{j-1}, X_j) dX_1 dX_{j-1} dX_j \\
= & \ P(|X_j| \leq 10) = O(j^{-1.5}).
\end{aligned}$$

On the other hand, we have

$$
\begin{aligned}
P(|X_j| \geq j^{5/8}) &= \int_{|X_j|\geq j^{5/8}} (\frac{1}{\sqrt{2\pi}\sigma_{n,j}})^3 e^{-\frac{|X_j|^2}{2\sigma_{n,j}^2}} \\
&= O(j^{-1.5}) \int_{j^{5/8}}^{\infty} r^2 e^{-\frac{r^2}{2\sigma_{n,j}^2}} dr \\
&\leq O(j^{-1.5}) \int_{j^{5/8}}^{\infty} r^2 e^{-\frac{r^2}{2j}} dr \\
(\text{let } r = \sqrt{j}u) \quad &= O(j^{-1.5}) \int_{j^{1/8}}^{\infty} j^{1.5} u^2 e^{-\frac{u^2}{2}} du \\
&= O(1) \int_{j^{1/8}}^{\infty} u^2 e^{-\frac{u^2}{4}} e^{-\frac{u^2}{4}} du \\
&\leq O(e^{-\frac{j^{0.25}}{4}}) \int_0^{\infty} u^2 e^{-\frac{u^2}{4}} du \\
&\leq O(j^{-1.5}).
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\int_{|X_j|\geq j^{5/8}} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} a(\ell_1, \ell_j) h'(X_1, X_{j-1}, X_j) dX_1 dX_{j-1} dX_j \\
\leq \quad & P(|X_j| \geq j^{5/8}) \leq O(j^{-1.5})
\end{aligned}
$$

as well. Now consider $P(|X_1| \geq r^{1/4})$ where $r = |X_j|$. The joint pdf of X_1 and X_j is

$$
\begin{aligned}
& h'(X_1, X_j) \\
= \quad & \int \int \ldots \int \frac{f(X_1)f(X_2 - X_1)\ldots f(O - X_{n-1})}{g_n(O)} dX_2 \ldots dX_{j-1} dX_{j+1} \ldots dX_{n-1} \\
= \quad & \left(\sqrt{2\pi n}\right)^3 f(X_1) g_{j-1}(X_j - X_1) g_{n-j}(X_j) \\
= \quad & \left(\sqrt{2\pi n}\right)^3 \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{|X_1|^2}{2}} \left(\frac{1}{\sqrt{2\pi(j-1)}}\right)^3 e^{-\frac{|X_j - X_1|^2}{2(j-1)}} \left(\frac{1}{\sqrt{2\pi(n-j)}}\right)^3 e^{-\frac{|X_j|^2}{2(n-j)}} \\
\leq \quad & \left(\frac{1}{\sqrt{2\pi(j-1)}}\right)^3 \left(\sqrt{\frac{n}{2\pi(n-j)}}\right)^3 e^{-\frac{|X_1|^2}{2}}.
\end{aligned}
$$

Using the facts that $j \leq (n+1)/2$, $\int_{\mathbf{R}^3} e^{-\frac{|X_1|^2}{4}} dX_j = 8\pi^{3/2}$ and that $\int_{\mathbf{R}^3} e^{-\frac{|X_j|^{0.5}}{4}} dX_j$ converges to a constant independent of j, we get

$$
\begin{aligned}
& P(|X_1| \geq r^{1/4}) \\
\leq \quad & \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \left(\frac{1}{\sqrt{2\pi(j-1)}}\right)^3 \left(\sqrt{\frac{n}{2\pi(n-j)}}\right)^3 e^{-\frac{|X_1|^2}{4}} e^{-\frac{|X_j|^{0.5}}{4}} dX_1 dX_j \\
= \quad & \int_{\mathbf{R}^3} \left(\frac{1}{\sqrt{j-1}}\right)^3 \left(\sqrt{\frac{n}{\pi(n-j)}}\right)^3 e^{-\frac{|X_j|^{0.5}}{4}} dX_j \\
= \quad & O(j^{-1.5})
\end{aligned}
$$

Similarly,

$$P(|X_j - X_{j-1}| \geq r^{1/4}) \leq O(j^{-1.5}).$$

It now follows that

$$E(a(\ell_1, \ell_j)) = I + O(j^{-1.5}), \tag{3.10}$$

where

$$I = \int_\Omega a(\ell_1, \ell_j) h'(X_1, X_{j-1}, X_j) dX_1 dX_{j-1} dX_j \tag{3.11}$$

and Ω is the region defined by $10 < |r| < j^{5/8}$, $|X_1| < r^{1/4}$ and $|X_j - X_{j-1}| < r^{1/4}$. Under the constrain conditions of Ω, we have

$$r - 2r^{1/4} < |X_{j-1} - X_1| < r + 2r^{1/4}.$$

This leads to

$$\frac{|X_{j-1} - X_1|^2}{2(j-2)} = \frac{r^2}{2j} + O(j^{-7/32})$$

and

$$e^{-\frac{|X_{j-1}-X_1|^2}{2(j-2)}} = (1 + O(j^{-7/32})) e^{-\frac{r^2}{2j}}$$

since $e^{O(j^{-7/32})} \leq 1 + O(j^{-7/32})$. Thus the function $g_{j-2}(X_{j-1} - X_1)$ (which is part of $h'(X_1, X_{j-1}, X_j)$) in (3.11) can be written as

$$\begin{aligned} g_{j-2}(X_{j-1} - X_1) &= \left(\frac{1}{\sqrt{2\pi(j-2)}}\right)^3 e^{-\frac{|X_{j-1}-X_1|^2}{2(j-2)}} \\ &= (1 + O(j^{-7/32})) \left(\frac{1}{\sqrt{2\pi(j-2)}}\right)^3 e^{-\frac{r^2}{2j}} \\ &= (1 + O(j^{-7/32})) \left(\frac{1}{\sqrt{2\pi j}}\right)^3 e^{-\frac{|X_j|^2}{2j}}. \end{aligned}$$

Hence in (3.11) the function $h'(X_1, X_{j-1}, X_j)$ can be written as

$$\begin{aligned} & \left(\sqrt{2\pi n}\right)^3 f(X_1) f(X_j - X_{j-1}) g_{j-2}(X_{j-1} - X_1) g_{n-j}(X_j) \\ = \; & (1 + O(j^{-7/32})) \cdot f(X_1) f(X_j - X_{j-1}) \cdot \\ & \left(\frac{1}{\sqrt{2\pi j}}\right)^3 e^{-\frac{|X_1|^2}{2j}} \cdot \left(\frac{1}{\sqrt{2\pi(n-j)}}\right)^3 e^{-\frac{|X_j|^2}{2(n-j)}} \\ = \; & (1 + O(j^{-7/32})) \cdot f(X_1) f(X_j - X_{j-1}) \cdot \left(\frac{1}{\sqrt{2\pi}\sigma_{n,j}}\right)^3 e^{-\frac{|X_j|^2}{2\sigma_{n,j}^2}} \\ = \; & (1 + O(j^{-7/32})) f(X_1) f(X_j - X_{j-1}) h_j(X_j). \end{aligned} \tag{3.12}$$

Using (3.10), (3.11), and (3.12) we now have

$$E(a(\ell_1, \ell_j)) = O(j^{-1.5}) + (1 + O(j^{-7/32})) \cdot J_\Omega \tag{3.13}$$

where

$$J_\Omega = \int_\Omega a(\ell_1, \ell_j) f(X_1) f(X_j - X_{j-1}) h_j(X_j) dX_1 dX_{j-1} dX_j.$$

Now observe that the integral

$$J=\int_{\mathbf{R}^3}\int_{\mathbf{R}^3}\int_{\mathbf{R}^3} a(\ell_1,\ell_j)f(X_1)f(X_j-X_{j-1})h_j(X_j)dX_1dX_{j-1}dX_j \tag{3.14}$$

is the mean average crossing number of the two edges ℓ_1, ℓ_j under the conditions that X_1, X_j, $X_{j-1}-X_j$ are independent Gaussian random vectors such that the variance of X_1 and $X_{j-1}-X_j$ is one and the variance of X_j is $\sigma^2_{j,n}$. Therefore, we can apply Lemma 3.3 and the argument used in the proof of Theorem 3.1 to estimate J. More precisely, we have

$$\begin{aligned} J &= \int_{\mathbf{R}^3} E(a(\ell_1,\ell_j)|X_j)\left(\frac{1}{\sqrt{2\pi}\sigma_{n,j}}\right)^3 e^{-\frac{|X_j|^2}{2\sigma^2_{n,j}}}dX_j \\ &= \int_0^\infty \left(\frac{1}{2\pi r^2}+O(\frac{1}{r^{2.5}})\right)4\pi r^2\left(\frac{1}{\sqrt{2\pi}\sigma_{n,j}}\right)^3 e^{-\frac{r^2}{2\sigma^2_{n,j}}}dr \end{aligned}$$

As in the proof of Theorem 3.1 this integral can be split into two:

$$\int_0^{10}\left(\frac{1}{2\pi r^2}+O(\frac{1}{r^{2.5}})\right)4\pi r^2\left(\frac{1}{\sqrt{2\pi}\sigma_{n,j}}\right)^3 e^{-\frac{r^2}{2\sigma^2_{n,j}}}dr = O(j^{-1.5})$$

and

$$\int_{10}^\infty\left(\frac{1}{2\pi r^2}+O(\frac{1}{r^{2.5}})\right)4\pi r^2\left(\frac{1}{\sqrt{2\pi}\sigma_{n,j}}\right)^3 e^{-\frac{r^2}{2\sigma^2_{n,j}}}dr$$

The second integral has two parts. The one containing the error term $O(\frac{1}{r_j^{2.5}})$ can be bounded above as

$$\begin{aligned} & 4\pi\left(\sqrt{\frac{n}{2\pi j(n-j)}}\right)^3\int_{10}^\infty O(\frac{1}{\sqrt{r}})e^{-\frac{nr^2}{2j(n-j)}}dr \\ =\ & O(\frac{1}{j^{1.5}})\int_{10}^\infty\frac{1}{\sqrt{r}}e^{-\frac{nr^2}{2j(n-j)}}dr \\ =\ & O(\frac{1}{j^{1.5}})O(j^{.25}) = O(j^{-1.25}) \end{aligned}$$

and the other part is estimated as follows:

$$\begin{aligned} & 2\left(\sqrt{\frac{n}{2\pi j(n-j)}}\right)^3\int_{10}^\infty e^{-\frac{nr^2}{2j(n-j)}}dr \\ =\ & \left(\sqrt{\frac{n}{2\pi j(n-j)}}\right)^3\int_{-\infty}^\infty e^{-\frac{nr^2}{2j(n-j)}}dr-\left(\sqrt{\frac{n}{2\pi j(n-j)}}\right)^3\int_{-10}^{10} e^{-\frac{nr^2}{2j(n-j)}}dr \\ =\ & \left(\sqrt{\frac{n}{2\pi j(n-j)}}\right)^3\sqrt{\frac{2\pi j(n-j)}{n}}+O(j^{-1.5}) \\ =\ & \frac{1}{2\pi}\frac{n}{j(n-j)}+O(j^{-1.5}) = \frac{1}{2\pi}(\frac{1}{j}+\frac{1}{n-j})+O(j^{-1.5}). \end{aligned}$$

Combining the above bounds, we get

$$J=\frac{1}{2\pi}\left(\frac{1}{j}+\frac{1}{n-j}\right)+O(j^{-1.25}). \tag{3.15}$$

Following similar arguments used in the beginning of this proof, we also have

$$J = O(j^{-1.25}) + J_\Omega. \tag{3.16}$$

Combining (3.13), (3.15) and (3.16), we obtain the following inequality:

$$\begin{aligned} E(a(\ell_1,\ell_j)) &= O(j^{-1.5}) + \left(1 + O(j^{-7/32})\right) \cdot (\frac{1}{2\pi}\left(\frac{1}{j} + \frac{1}{n-j}\right) + O(j^{-1.25})) \\ &= \frac{1}{2\pi}\left(\frac{1}{j} + \frac{1}{n-j}\right) + O(j^{-39/32}). \end{aligned}$$

As in the proof of Theorem 3.1 we now estimate

$$\begin{aligned} E(\chi'_n) &= n \sum_{3\le j\le (n+1)/2} E(a(\ell_1,\ell_j)) \\ &= n \sum_{3\le j\le (n+1)/2} \frac{1}{2\pi}\left(\frac{1}{j} + \frac{1}{n-j}\right) + nO(\sum_{3\le j\le (n+1)/2} j^{-39/32}) \\ &= \frac{1}{2\pi} n \ln n + O(n). \end{aligned}$$

□

References

[1] T. Deguchi and K. Tsurusaki, *A statistical study of random knotting using the Vassiliev invariants*, Journal of Knot Theory and Its Ramifications, Vol.3, No.3 (1994), 321–353.

[2] Y. Diao, *The Knotting of Equilateral Polygons in* $\mathbf{R}^3$, Journal of Knot Theory and its Ramifications, Vol.4, No.2 (1995), 189–196.

[3] Y. Diao, A. Dobay, R. B. Kusner, K. Millett and A. Stasiak, *The Average Crossing Number of Equilateral Random Polygons*, Journal of Physics A: Mathematical and General, Vol. 36, No. 46 (2003), 11561 - 11574.

[4] Y. Diao, J. Nardo and Y. Sun, *Global Knotting in Equilateral Random Polygons*; Journal of Knot Theory and its Ramifications, Vol.10, No.4 (2001), 597–607.

[5] Y. Diao, N. Pippenger and D. W. Sumners, *On random knots*, Journal of Knot Theory and its Ramifications, Vol.3, No.3 (1994), 419–429.

[6] M. Doi and S.F. Edwards, *The theory of polymer dynamics*, Oxford University Press, 1986.

[7] J. P. Flory, *Principles of Polymer Chemistry*, Cornell University Press, 1953.

[8] M.H. Freedman and Z. He, *Divergence-free Fields: Energy and Asymptotic Crossing Number,* Annals of Math., 134 (1991), 189–229.

[9] P. G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, 1979.

[10] D. Jungreis, *Gaussian Random Polygons Are Globally Knotted,* Journal of Knot Theory and its Ramifications, Vol.3, No.4 (1994), 455–464.

[11] V. Katritch, W. K. Olson, A. V. Vologodskii, J. Dubochet and A. Stasiak, *Tightness of random knotting*, Phys. Rev. E. 61 (2000), 5545–5549.

[12] N. Pippenger, *Knots in Random Walks,* Discrete Appl. Math. 25 (1989), 273-278.

[13] L. Rayleigh, *On the problems of random vibrations, and of random flights in one, two, or three dimensions*, Phil. Mag. S. 6. Vol.37, No.220 (1919), 321–347.

[14] V. V. Rybenkov, N. R. Cozzarelli and A. V. Vologodskii, *The probability of DNA knotting and the effective diameter of the DNA double helix*, Proc. Nat. Acad. Sci. USA, Vol.90 (1993), 5307–5311.

[15] D.W. Sumners and S.G. Whittington, *Knots in Self-Avoiding Walks,* J.Phys.A: Math.Gen. 21(1988), 1689-1694.

Department of Mathematics, University of North Carolina at Charlotte, Charlotte, NC 28223

E-mail address: `ydiao@uncc.edu`

Department of Mathematics, Western Kentucky University, Bowling Green, KY 42101

E-mail address: `claus.ernst@wku.edu`