

# Mean unknotting times of random knots and embeddings

Yao-ban Chan<sup>1</sup>, Aleksander L Owczarek<sup>1</sup>,  
Andrew Rechnitzer<sup>2</sup> and Gordon Slade<sup>2</sup>

<sup>1</sup> Department of Mathematics and Statistics, University of Melbourne,  
Victoria 3010, Australia

<sup>2</sup> Department of Mathematics, University of British Columbia, Vancouver,  
BC, V6T 1Z2, Canada

E-mail: [y.chan@ms.unimelb.edu.au](mailto:y.chan@ms.unimelb.edu.au), [a.owczarek@ms.unimelb.edu.au](mailto:a.owczarek@ms.unimelb.edu.au),  
[andrewr@math.ubc.ca](mailto:andrewr@math.ubc.ca) and [slade@math.ubc.ca](mailto:slade@math.ubc.ca)

Received 5 December 2006

Accepted 12 April 2007

Published 4 May 2007

Online at [stacks.iop.org/JSTAT/2007/P05004](http://stacks.iop.org/JSTAT/2007/P05004)

doi:10.1088/1742-5468/2007/05/P05004

**Abstract.** We study mean unknotting times of knots and knot embeddings by crossing reversals, in a problem motivated by DNA entanglement. Using self-avoiding polygons (SAPs) and self-avoiding polygon trails (SAPTs) we prove that the mean unknotting time grows exponentially in the length of the SAPT and at least exponentially with the length of the SAP. The proof uses Kesten's pattern theorem, together with results for mean first-passage times in the two-parameter Ehrenfest urn model. We use the pivot algorithm to generate random SAPTs of up to 3000 steps and calculate the corresponding unknotting times, and find that the mean unknotting time grows very slowly, even at moderate lengths. Our methods are quite general—for example, the lower bound on the mean unknotting time applies also to Gaussian random polygons.

**Keywords:** rigorous results in statistical mechanics, classical Monte Carlo simulations, stochastic processes (theory), mechanical properties (DNA, RNA, membranes, bio-polymers) (theory)

---

## Contents

|   |           |
|---|-----------|
| <b>1. Introduction</b>  | <b>2</b>  |
| 1.1. Mean unknotting times . . . . .                                | 2         |
| 1.2. The processes . . . . .  | 5         |
| 1.2.1. Unknotting SAPTs. . . . .                                    | 5         |
| 1.2.2. Unknotting SAPs. . . . .                                     | 5         |
| 1.3. Main analytic results . . . . .                                | 6         |
| <b>2. Lower bound on the mean unknotting time of SAPs and SAPTs</b> | <b>7</b>  |
| 2.1. Kesten's pattern theorem . . . . .                             | 7         |
| 2.2. Mapping to the Ehrenfest urn model . . . . .                   | 9         |
| 2.2.1. The two-parameter Ehrenfest urn model. . . . .               | 9         |
| 2.2.2. Lower bound for SAPTs. . . . .                               | 10        |
| 2.2.3. Lower bound for SAPs. . . . .                                | 11        |
| 2.2.4. Lower bound to 'nearly unknot'. . . . .                      | 11        |
| <b>3. Upper bound on the mean knotting time of SAPTs</b>            | <b>12</b> |
| <b>4. Numerical results</b>   | <b>13</b> |
| <b>5. Conclusion</b>  | <b>15</b> |
| <b>Acknowledgments</b>  | <b>15</b> |
| <b>References</b>   | <b>16</b> |

---

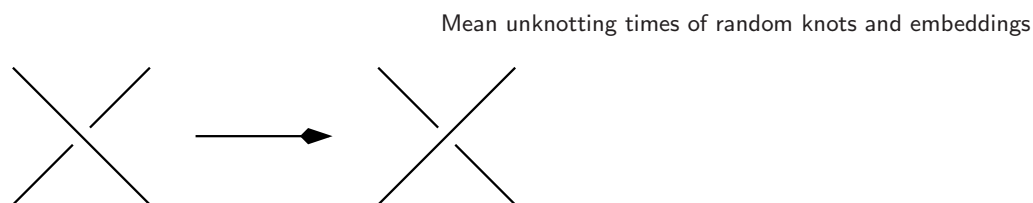
## 1. Introduction

### 1.1. Mean unknotting times

An embedded knot can be transformed into the unknot by a sequence of crossing reversals (see figure 1). The minimum number of crossing reversals needed to change a knot into the unknot is called its *unknotting number*. This quantity has been extensively studied (see, for example, [12, 5, 3, 25, 11]), but its value is known only for small knots. In this paper we study the average number of random crossing reversals required to transform a 'typical' knot to the unknot, where random knots are modelled by self-avoiding polygon trails (SAPTs) on the square lattice  $\mathbb{Z}^2$  and by self-avoiding polygons (SAPs) on the cubic lattice  $\mathbb{Z}^3$ . We call this average number of crossing reversals the *mean unknotting time*.

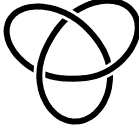
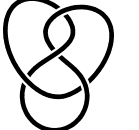


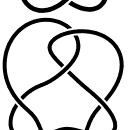
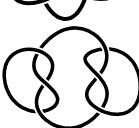
Consider the embeddings in table 1. At each time step, choose a crossing uniformly at random and reverse it. If the new embedding is the unknot then stop, otherwise repeat. For embeddings with such a small number of crossings, the mean unknotting times under this process can be computed exactly.

Although the unknotting number depends only on the knot type (since it is a minimum over all embeddings of that knot), the mean unknotting time depends strongly on the embedding. To see this, observe that the knot type is unchanged by adding a simple twist to any knot via a type I Reidemeister move [21] (see figure 2). Reversing this new crossing



**Figure 1.** Reversing a crossing.

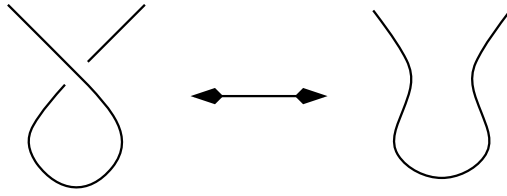
**Table 1.** Exact mean unknotting times and variances under random crossing reversals.

| Knot         | Embedding   | Unknotting<br>Number | Mean            | Var               |
|--------------|---|----------------------|-----------------|-------------------|
| $3_1$        |    | 1                    | 1               | 0                 |
| $4_1$        |   | 1                    | 1               | 0                 |
| $5_1$        |  | 2                    | $\frac{5}{2}$   | $\frac{5}{4}$     |
| $5_2$        |  | 1                    | $\frac{20}{11}$ | $\frac{105}{121}$ |
| $6_1$        |  | 1                    | $\frac{15}{8}$  | $\frac{51}{64}$   |
| $3_1 \# 3_1$ |  | 2                    | $\frac{10}{3}$  | $\frac{34}{9}$    |

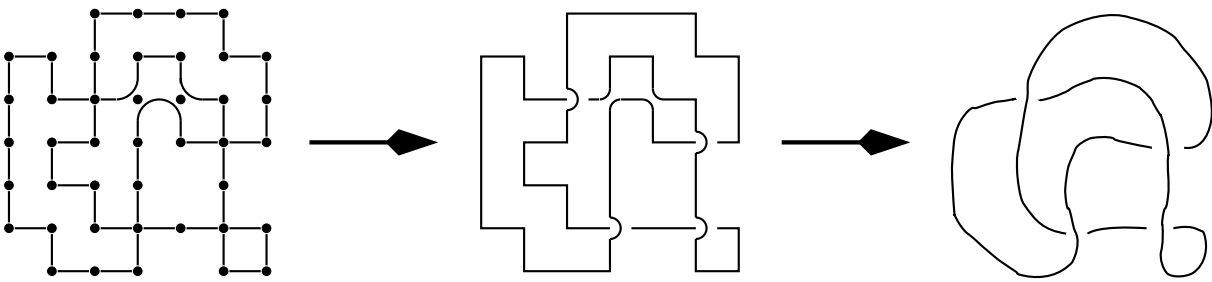
has no effect on the knot type and so its presence increases the mean unknotting time. In particular, the mean unknotting time of a *given* knot type is not well defined.

A *self-avoiding polygon* (SAP) is a path on the cubic lattice which starts and ends at the same point, in which each vertex has degree two. A *self-avoiding polygon trail* (SAPT) is a path on the square lattice which starts and ends at the same point, in which each edge may only be visited once, but sites may be visited multiple times. Note that SAPs and SAPTs are neither rooted nor directed. We declare each SAP of length  $n$  to be equally likely, and similarly for SAPTs.

An example of a self-avoiding polygon trail is depicted in figure 3. At the points where the trail ‘veers off’ from a vertex, it actually touches that vertex and moves off at



**Figure 2.** A type I Reidemeister move.

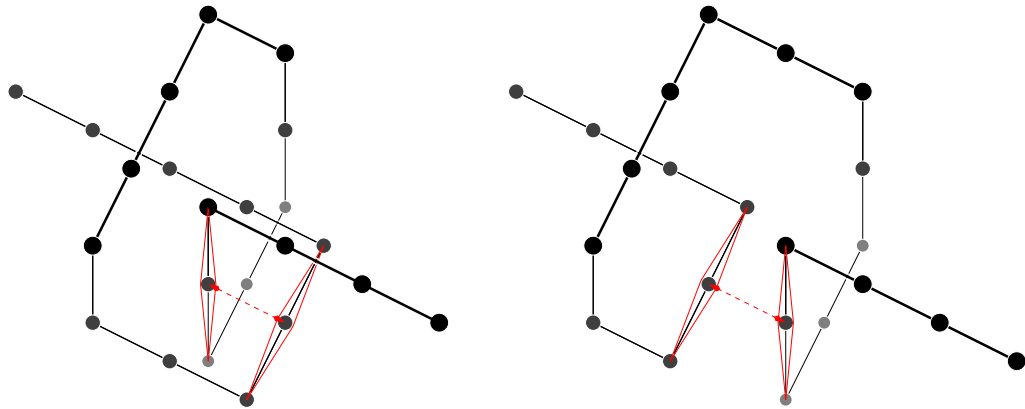


**Figure 3.** Converting a SAPT to a knot.

a right angle, but we draw it in this way to differentiate from the case where it crosses itself. At each crossing, we assign one strand to be the overpass (so that it lies above the other strand). This gives an embedding of a knot. Note that due to the necessity of choosing overpasses, a SAPT with  $m$  crossings can generate  $2^m$  distinct embeddings.

It is well known that SAPs are knotted with high probability [20, 26, 24]. The proof of this fact, which uses Kesten's pattern theorem [13], extends also to SAPTs [6] (as we explain in more detail in section 2.1). In addition, the pattern theorem implies that, with high probability, a typical SAP or SAPT contains a number of tight local trefoils that are linear in the length  $n$  of the polygon. It is the combination of a linear number of tight trefoils, together with a fixed positive probability that an unknotted trefoil might be reknotted, which allows us to prove that, with high probability, a randomly chosen SAP or SAPT will have a mean unknotting time that is at least exponentially large in  $n$ . For SAPs, the exponential behaviour of the mean unknotting time persists even when we bias our local unknotting processes to be much more likely to perform crossing reversals that unknot local knots than to create local knots, as long as there is a fixed positive probability to (mistakenly) create a knot. Moreover, the mean time to unknot all but a small linear fraction of the tight local trefoils is also at least exponentially large—there is no need to require complete unknotting to observe the exponential behaviour. The proof of this lower bound is based on an analysis of the two-parameter Ehrenfest urn model. For the case of SAPTs, we also prove an exponential upper bound on the mean unknotting time.

Our SAPT model is reminiscent of the *flat knots* studied by Gutter and Orlandini [9] and Metzler *et al* [16]. However, the focus of [9, 16] is on two-dimensional embeddings of knots of a fixed knot type, and on the localized nature of the knot. In contrast, in the SAP or SAPT models, the pattern theorem ensures that, with high probability, the knot type becomes increasingly complex as the size of the polygon increases. It is the increasingly large number of localized trefoils in a typical SAP or SAPT that lies at the heart of our analysis.



**Figure 4.** A trefoil arc in a self-avoiding polygon. One possible way it can be unknotted is by passing the highlighted strands through each other.

Our unknotting processes mimic the action of type II topoisomerases, which act in similar ways to reduce the topological complexity of DNA strands within cells in order for biological processes to take place. These enzymes simplify the topology of a DNA strand by breaking it at one point, passing another section of the strand through the break and then rejoining the broken ends—effectively the reversal of a crossing. See [29, 22, 28, 27] for examples. There is evidence [23, 30] that this enzyme drives the system towards the unknotted state rather than simply crossing and recrossing strands at random. The parameters in our SAP model allow for a bias towards unknotting, consistent with this ‘intelligent’ action of type II topoisomerase.

## 1.2. The processes

We consider two unknotting processes, one acting on two-dimensional SAPTs and the other acting on three-dimensional SAPs.

**1.2.1. Unknotting SAPTs.** Consider a SAPT with  $m$  crossings. Each crossing (be it initially an overpass or an underpass) is defined to be initially in state 0. When the crossing is reversed it moves from 0 to 1 or vice versa. At each time step one of the  $m$  crossings is selected uniformly at random and reversed. The process continues until the SAPT becomes unknotted. This models a naive topoisomerase enzyme which performs crossing reversals randomly. We will mimic more intelligent unknotting behaviour in our SAP model.

**1.2.2. Unknotting SAPs.** Gitter and Orlandini [9] and Metzler *et al* [16] have studied flat knots and found that the topological details of the knots are tightly localized. Metzler *et al* speculate that such behaviour continues in three dimensions. With such localized knots in mind, we define our unknotting process on three-dimensional SAPs by considering the probabilities of knotting and unknotting tight trefoil arcs (an example is given in figure 4) rather than a process on individual strands of the polygon. According to the pattern theorem [13, 20, 26], a SAP contains a linear number of tight trefoils with high probability.

Consider a SAP with at least  $m$  tight trefoil arcs. In their initial (knotted) state each of these  $m$  arcs is assigned state 0 and when they are unknotted we assign them state 1. At each time step one of the  $m$  arcs (be they knotted or unknotted) is selected uniformly at random. If the arc is knotted then it is unknotted with probability  $s$  (moving from state 0 to state 1). If the arc is in state 1 then it is reknotted with probability  $t$  (moving from state 1 to state 0). We do not consider the effect of this process on other knotted arcs within the polygon.

By reducing the chance that an arc in state 1 can be changed back to state 0 the process can be driven away from its initial (knotted) state more quickly than the unbiased  $s = t = 1$  process. Thus choosing  $s > t$  mimics the action of an ‘intelligent’ type II topoisomerase. For all  $0 < s, t \leq 1$  we find that the mean time to unknot or nearly unknot the SAP grows at least exponentially with length. These results are stated more precisely below.

### 1.3. Main analytic results

The following theorem proves an exponential lower bound on our two unknotting processes:

**Theorem 1.1.** (i) (Lower bound for SAPTs.) *There is a constant  $C > 1$  such that a randomly chosen two-dimensional SAPT of length  $n$  has mean unknotting time at least  $C^n$ , with a probability that tends to 1 exponentially rapidly in  $n$ . Further, there is a  $\delta > 0$  such that, again with exponentially high probability, there are at least  $\delta n$  tight knotted arcs initially present in a randomly chosen SAPT, and there exist  $\epsilon > 0$  and  $C_\epsilon > 1$  (both independent of  $n$ ) so that the mean time it takes to unknot all but  $\epsilon \delta n$  of these initially knotted arcs is at least  $C_\epsilon^n$ .*

(ii) (Lower bound for SAPs.) *Let  $s > 0$  be the probability of unknotting a tight trefoil arc in a SAP and let  $t > 0$  be the probability of it being reknotted. There is a constant  $D > 1$  (depending on  $s$  and  $t$ ) such that a randomly chosen three-dimensional SAP of length  $n$  has mean unknotting time at least  $D^n$  with a probability that tends to 1 exponentially rapidly in  $n$ . Further, there is a  $\delta > 0$  such that, again with exponentially high probability, there are at least  $\delta n$  tight knotted arcs initially present in a randomly chosen SAP, and there exist  $\epsilon > 0$  and  $D_\epsilon > 1$  (depending on  $s$  and  $t$  but independent of  $n$ ) so that the mean time it takes to unknot all but  $\epsilon \delta n$  of these initially knotted arcs is at least  $D_\epsilon^n$ .*

We also prove the following upper bound on the mean unknotting time of the two-dimensional process:

**Theorem 1.2.** (Upper bound for SAPTs.) *There is a  $\lambda > 1$  such that the mean unknotting time of every two-dimensional SAPT of length  $n$  is bounded above by  $\lambda^n$ .*

The upper bound does not extend to our SAP model. Indeed, consider any unknotting process that only works on either local crossings (where strands are at most some constant  $k$  lattice spacings apart) or on very localized knots (with diameter of at most  $k$ ), then by a pattern-theorem argument (see section 2.1 below), with very high probability a polygon will contain ‘bloated’ trefoil arcs whose strands are far more than  $k$  lattice spacings apart. The local unknotting process can never unknot these arcs. However, the lower bound

holds regardless of whether or not such large knotted arcs are present. On the other hand, the existence of an exponential upper bound for the naive crossing reversal SAPT model implies that any more intelligent unknotting process by definition should not have a larger mean unknotting time.

We have also measured the mean unknotting time of the unknotting process on SAPTs of lengths up to 3000, using numerical simulations. The mean unknotting time is quite small and its exponential growth is very weak.

In section 2, we give the proof of the lower bound of theorem 1.1; the proof relies on Kesten's pattern theorem and on an analysis of the two-parameter Ehrenfest urn model. The upper bound for SAPTs is proved in section 3. In section 4, we detail our simulation results from generating random SAPTs. Finally, in section 5 we summarize our conclusions.

## 2. Lower bound on the mean unknotting time of SAPs and SAPTs

We first use the pattern theorem to show that a typical SAP or SAPT contains a positive density of trefoil arcs (such as those illustrated in figures 4 and 6). We then describe the random unknotting process in terms of the mean first-passage time of the two-parameter Ehrenfest urn model. An exponential lower bound for the latter then implies that the mean unknotting time grows exponentially with the number of trefoil arcs and hence the length.

### 2.1. Kesten's pattern theorem

A crucial ingredient in our proof of the exponential lower bound on the mean unknotting time is Kesten's pattern theorem. The original pattern theorem for self-avoiding walks was proved by Kesten in 1963 [13]. The pattern theorem was applied to prove that SAPs are knotted with high probability in [20, 26]. In this section, we describe an extension to SAPTs; the corresponding discussion for SAPs is very similar.

Loosely speaking, the pattern theorem tells us that a SAPT or SAP of length  $n$  will contain a given pattern of edges at least  $\delta n$  times (for some  $\delta > 0$ ) with a probability that tends to 1 exponentially rapidly in  $n$ . We give the following definitions for SAPTs:

**Definition 2.1.** Let  $t = (x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$  be a self-avoiding trail in  $\mathbb{Z}^2$ . We call  $t$  a bridge if  $x_0 < x_i \leq x_n$  for all  $i > 0$ .

We call the line  $x = x'$  a cutting line of the bridge  $t$  if there exists an index  $j$  where  $0 < j < n$ ,  $x_j = x'$ , and the  $x$ -coordinates satisfy the conditions

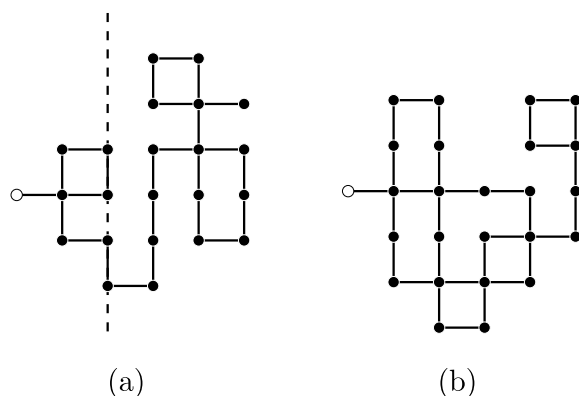
$$x_0 < x_i \leq x_j \quad \forall 0 < i < j \quad (1)$$

and

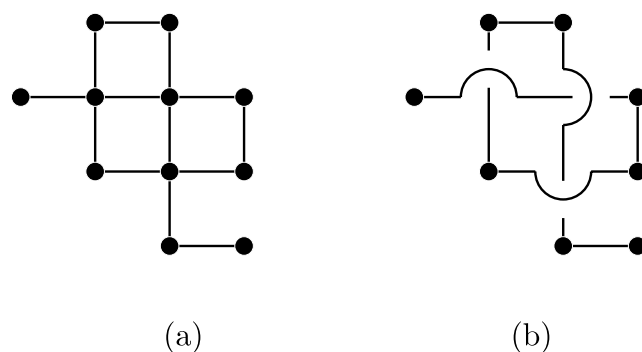
$$x_j < x_k \leq x_n \quad \forall j < k < n. \quad (2)$$

A cutting line divides a bridge into two separate non-empty bridges. If a non-empty bridge has no cutting line, we call it prime. Note that bridges can be trivial, but prime bridges cannot.

Examples illustrating the above definition are given in figure 5.



**Figure 5.** Two types of trails: (a) a bridge, with a cutting line; and (b) a prime bridge. The starting vertices are denoted by hollow circles.



**Figure 6.** Trefoil arc: (a) without crossings specified and (b) with crossings specified.

**Definition 2.2.** A pattern is a synonym for self-avoiding trail. It is used mainly in the context where it is part of another trail or SAPT. A prime pattern is a pattern which is a prime bridge.

The following is the pattern theorem for SAPTs. The proof for SAPTs is given in [6]; it is similar in spirit to proofs of the equivalent theorems for self-avoiding walks [13, 15] and lattice ribbons [10]. With the obvious modifications, the theorem (and its corollary) applies also to SAPs.

**Theorem 2.3.** Let  $T^o(n; p, m)$  be the number of SAPTs of length  $n$  in which the prime pattern  $p$  occurs at most  $m$  times. Then there exists a number  $\delta(p) > 0$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{2n} \ln T^o(2n; p, 2n\delta(p)) < \ln \mu^o \quad (3)$$

where  $\mu^o$  denotes the growth constant for SAPTs.

**Corollary 2.4.** Given a prime pattern  $p$ , let  $N_p$  denote the number of occurrences of  $p$  in a SAPT of length  $n$ . Then the probability that  $N_p$  is at least  $\delta(p)n$  goes to 1 exponentially rapidly as  $n \rightarrow \infty$ .



**Proof.** This is a restatement of theorem 2.3.  $\square$

Consider the pattern depicted in figure 6(a), which we call the *trefoil arc*  $\tau$ . This pattern is clearly a prime pattern, so corollary 2.4 implies that with overwhelmingly high probability it occurs in a SAPT of length  $n$  at least  $\delta(\tau)n$  times. If we allocate the crossings in the right way, e.g. as in figure 6(b), then this generates a trefoil in the knot. There are exactly two ways (out of eight possibilities) that the crossings can create a trefoil, so the probability of this occurring is  $\frac{1}{4}$ . Similarly, the trefoil arc in figure 4 (left) occurs a positive density of times in three-dimensional SAPs (except with exponentially small probability). These observations will be important in our proof of the lower bound on the mean unknotting time.

## 2.2. Mapping to the Ehrenfest urn model

We now use the pattern theorem results to map the mean unknotting time to the mean first-passage time in the two-parameter Ehrenfest urn model. The latter has been well studied [7, 14, 19], and the exponential growth of its mean first-passage time will imply exponential growth for the mean unknotting time.

*2.2.1. The two-parameter Ehrenfest urn model.* The two-parameter Ehrenfest urn is the Markov chain with state space  $\{0, 1, \dots, m\}$  and transition probabilities

$$q_{i,i-1} = b \frac{i}{m}, \quad q_{i,i} = 1 - q_{i,i-1} - q_{i,i+1}, \quad q_{i,i+1} = a \left(1 - \frac{i}{m}\right), \quad (4)$$

where  $a, b \in [0, 1]$  are parameters and  $i = 0, \dots, m$ . Explicit formulae for mean first-passage times are derived in [7, 14, 19]. We will state one of these formulae in (8) below, but to help keep our analysis self-contained, we also apply the following elementary reasoning.

The Markov chain (4) has the following interpretation: a total of  $m$  balls are divided between two urns  $A$  and  $B$ . A ball is chosen uniformly at random. If it is chosen from urn  $A$  then it is placed in urn  $B$  with probability  $b$  and otherwise it is returned to urn  $A$ . Similarly, if the ball is chosen from urn  $B$  then it is placed in urn  $A$  with probability  $a$  and otherwise it is returned to urn  $B$ . The state of the system is the number of balls in urn  $A$ .

It is easy to guess the stationary distribution of the two-parameter Ehrenfest urn, as follows. We label the balls from 1 to  $m$  and note that balls move independently. Each ball when selected moves from urn  $A$  to urn  $B$  with probability  $b$  and moves from urn  $B$  to urn  $A$  with probability  $a$ . In equilibrium, it is therefore in urn  $A$  with probability  $a/(a+b)$ . The number of balls in urn  $A$  in equilibrium therefore has a binomial distribution with parameters  $m$  and  $a/(a+b)$ , i.e. the stationary distribution of the Markov chain is given by

$$\pi_i = \binom{m}{i} \left(\frac{a}{a+b}\right)^i \left(\frac{b}{a+b}\right)^{m-i} \quad (i = 0, \dots, m). \quad (5)$$

Checking that detailed balance is satisfied (i.e.  $\pi_i q_{ij} = \pi_j q_{ji}$ ) then proves that this is indeed the stationary distribution. This stationary distribution can also be found in [7, 14].

For our purposes, we first note that in general the mean first-return time to a state  $i$  in a Markov chain is given by  $1/\pi_i$ , so that the mean first-return time  $M_{m,m}$  to state  $m$

starting from  $m$  is given by  $M_{m,m} = 1/\pi_m = ((a+b)/a)^m$ . By conditioning on the first step, we see that

$$M_{m,m} = (1-b) + b(1 + M_{m-1,m}) = 1 + bM_{m-1,m}, \quad (6)$$

where  $M_{m-1,m}$  is the mean first-passage time from state  $m-1$  to state  $m$ . Therefore

$$M_{m-1,m} = \frac{1}{b}M_{m,m} - \frac{1}{b} = \frac{1}{b} \left( \frac{a+b}{a} \right)^m - \frac{1}{b}, \quad (7)$$

which is a special case of theorem (b) on p 968 of [14], where more general hitting times are also considered. Since every passage from 0 to  $m$  must pass through  $m-1$ , the mean first-passage time from 0 to  $m$  is at least as large as the mean first-passage time from  $m-1$  to  $m$ .

In general, it is shown in [7] that the mean first-passage time from state  $i$  to state  $j$  is given by

$$M_{i,j} = \frac{m}{a+b} \sum_{k=1}^j \frac{1}{k(-m)_k} \left( \frac{a+b}{a} \right)^k ((-j)_k - (-i)_k), \quad (8)$$

where  $(n)_0 = 1$ ,  $(n)_k = n(n+1) \cdots (n+k-1)$  is the Pochhammer symbol. We will find this formula useful in section 2.2.4.

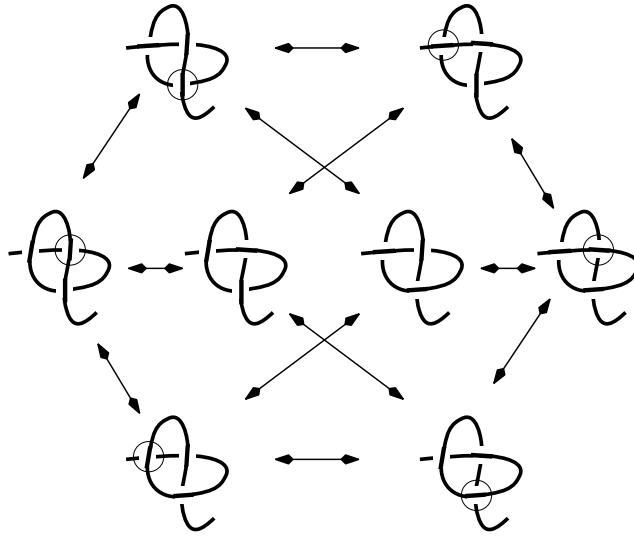
**2.2.2. Lower bound for SAPTs.** Let  $K_\tau$  denote the number of times that the trefoil arc  $\tau$  occurs in knotted form in a SAPT of length  $n$ . According to the remarks following corollary 2.4, with a probability that is exponentially close to 1,  $K_\tau$  is at least  $\delta n$ , for some  $\delta > 0$ . We will show that there is a  $c_1 > 1$  such that a SAPT with  $K_\tau$  knotted trefoil arcs has a mean unknotting time that is at least  $c_1^{K_\tau}$  and so prove the lower bound on the mean unknotting number stated in theorem 1.1 for SAPTs.

Given a SAPT of length  $n$  with  $K_\tau = m$  trefoil arcs whose crossings are all arranged to form trefoils (i.e. knotted), we associate to this knot the site  $\mathbf{0}$  of the  $m$ -dimensional unit cube. As we perform reversals, some of these trefoil arcs can become unknotted. To each trefoil arc, we assign the state 0 if it is knotted, and 1 if it is unknotted. We seek a lower bound on the time needed to move from state  $\mathbf{0}$  (all trefoil arcs knotted) to state  $\mathbf{1}$  (all trefoil arcs unknotted). This will also provide a lower bound on the mean unknotting time of the embedding, since each trefoil arc must be unknotted if the embedding is to be the unknot. It is only a lower bound because the entire knot may still be knotted even after all the trefoil arcs have been unknotted.

Analysing the possible crossing reversals of a trefoil, shown in figure 7, shows that if a trefoil arc is in state 0, a crossing reversal has probability 1 of unknotting it (and therefore sending it to state 1), but if the arc is unknotted, only one of the three crossings makes the arc knotted again when reversed. If the number of 1's (unknotted trefoil arcs) is  $i$ , then the crossing reversal process corresponds to a Markov chain with transition probabilities

$$p_{i,i-1} = \frac{1}{3} \frac{i}{m}, \quad p_{i,i} = \frac{2}{3} \frac{i}{m}, \quad p_{i,i+1} = 1 - \frac{i}{m} \quad (i = 0, \dots, m). \quad (9)$$

Comparing with (4), we see that this is the two-parameter Ehrenfest urn model with  $a = 1$ ,  $b = \frac{1}{3}$ . The mean unknotting number is bounded below by the mean first-passage time from state 0 to state  $m$ , which by (7) is at least  $3(\frac{4}{3})^m - 3$ .



**Figure 7.** All possible crossing reversals of a trefoil arc. The two knots in the centre are knotted; the outer knots are unknotted. The circled crossings are the ones which differ from the connected centre knot.

**2.2.3. Lower bound for SAPs.** The proof for SAPs is essentially identical. By the pattern theorem, a SAP of length  $n$  contains at least  $m = \delta n$  tight local trefoils, with probability exponentially close to 1. If we denote by  $i$  the number of these trefoils in a knotted state, then the transition probabilities of our process are

$$p_{i,i-1} = t \frac{i}{m}, \quad p_{i,i} = 1 - p_{i,i-1} - p_{i,i+1}, \quad p_{i,i+1} = s \left(1 - \frac{i}{m}\right). \quad (10)$$

This is equivalent to the urn model with  $a = s$  and  $b = t$ . The mean unknotting number is at least as large as the mean first-passage time  $M_{0,m}$  of the urn model, which by (7) grows exponentially in  $n$ , provided that the probability of reknitting a trefoil is nonzero (i.e.,  $t > 0$ ).

**2.2.4. Lower bound to ‘nearly unknot’.** We can now extend these results to compute the mean time to ‘nearly unknot’ a configuration; i.e. the mean time it takes to remove all but a small positive fraction  $\epsilon$  of the knotted arcs. This quantity is equivalent to the mean first-passage time from 0 to  $(1 - \epsilon)m$  in the two-parameter Ehrenfest urn model, with  $a = b = 1$  for the SAPT model, and with  $a = s$  and  $b = t$  for the SAP model. When  $i = 0$ , all the summands in (8) are positive and so any particular summand provides a lower bound. In particular,

$$M_{0,j} \geq \frac{m}{a+b} \frac{1}{j} \binom{m}{j}^{-1} \left(\frac{a+b}{a}\right)^j. \quad (11)$$

When  $j = (1 - \epsilon)m$  this gives

$$M_{0,(1-\epsilon)m} \geq \frac{m}{a+b} \frac{1}{(1-\epsilon)m} \binom{m}{(1-\epsilon)m}^{-1} \left(\frac{a+b}{a}\right)^{(1-\epsilon)m}. \quad (12)$$

The binomial coefficient is  $\binom{m}{\epsilon m} \sim A_\epsilon (\epsilon^{-\epsilon}(1-\epsilon)^{\epsilon-1})^m m^{-1/2}$ . By choosing  $\epsilon$  close to 0, we can make  $\epsilon^{-\epsilon}(1-\epsilon)^{\epsilon-1}$  arbitrarily close to 1. Hence for any given  $a, b > 0$  we can choose  $\epsilon > 0$  so that  $\epsilon^\epsilon(1-\epsilon)^{1-\epsilon}(1+b/a)^{1-\epsilon} > 1$  and the mean first-passage time from 0 to  $(1-\epsilon)m$  will grow exponentially. This completes the proof of theorem 1.1.

### 3. Upper bound on the mean knotting time of SAPTs

We now prove the upper bound of theorem 1.2 for the SAPT process. Consider a SAPT of length  $n$  which contains  $m$  crossings. Since a SAPT must visit each crossing exactly twice and cannot visit more than  $n$  distinct vertices, it follows that  $m \leq n/2$ . Each crossing (be it initially an overpass or an underpass) is in state 0. When the crossing is reversed it moves from 0 to 1 and vice versa. Hence the process of reversing crossings is equivalent to a simple random walk on a unit  $m$ -dimensional cube.

At least one way of allocating strands in the crossings must result in the unknot, and so at least one point in the  $m$ -cube is equivalent to the unknot. The mean unknotting time must therefore be less than or equal to the mean first-passage time from the origin to that point. We bound this quantity from above by the mean first-passage time from  $\mathbf{0}$  to  $\mathbf{1}$ , using the following lemma.

**Lemma 3.1.** *Let  $\mathbf{x}$  be a site in the  $m$ -cube. The mean first-passage time from  $\mathbf{0}$  to  $\mathbf{x}$  is less than or equal to the mean first-passage time from  $\mathbf{0}$  to  $\mathbf{1}$ .*

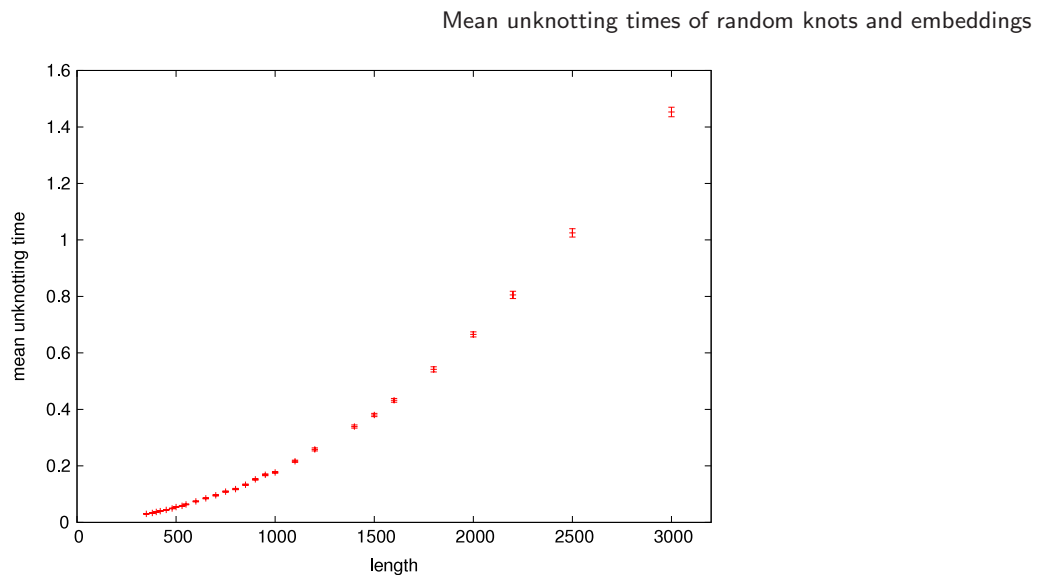
**Proof.** We define addition of sites  $\mathbf{x}$  and  $\mathbf{y}$  in the  $m$ -cube by componentwise addition mod 2. Let  $T_{\mathbf{x},\mathbf{y}}$  denote the first-passage time from  $\mathbf{x}$  to  $\mathbf{y}$ . By symmetry,  $T_{\mathbf{x},\mathbf{y}}$  and  $T_{\mathbf{x}+\mathbf{a},\mathbf{y}+\mathbf{a}}$  have the same probability distribution for any  $\mathbf{a}$  in the  $n$ -cube. In particular,  $ET_{\mathbf{0},\mathbf{x}} = ET_{\bar{\mathbf{x}},\mathbf{1}}$ , where  $\bar{\mathbf{x}} = \mathbf{1} - \mathbf{x}$  is the site obtained by interchanging coordinates 0 and 1 in  $\mathbf{x}$ . Therefore it suffices to show that  $ET_{\mathbf{0},\mathbf{1}} \geq ET_{\bar{\mathbf{x}},\mathbf{1}}$ .

Let  $A$  denote the set of sites  $\mathbf{y}$  which contain exactly the same number of coordinates 1 as  $\bar{\mathbf{x}}$ , and let  $X$  denote the first site in  $A$  that is hit by a walk started from  $\mathbf{0}$ . By symmetry, the conditional expectation  $E[T_{\mathbf{0},\mathbf{1}}|X = \mathbf{y}]$  is the same for every  $\mathbf{y} \in A$ , and therefore  $ET_{\mathbf{0},\mathbf{1}} = E[T_{\mathbf{0},\mathbf{1}}|X = \bar{\mathbf{x}}]$ . However, conditional on  $X = \bar{\mathbf{x}}$ , we can decompose a first-passage path from  $\mathbf{0}$  to  $\mathbf{1}$  into the path from  $\mathbf{0}$  to  $\bar{\mathbf{x}}$  responsible for  $X = \bar{\mathbf{x}}$ , followed by a first-passage path from  $\bar{\mathbf{x}}$  to  $\mathbf{1}$ . By neglecting the time taken to achieve  $X = \bar{\mathbf{x}}$ , we obtain the lower bound  $ET_{\mathbf{0},\mathbf{1}} \geq E[T_{\bar{\mathbf{x}},\mathbf{1}}|X = \bar{\mathbf{x}}]$ . But the random variable  $T_{\bar{\mathbf{x}},\mathbf{1}}$  is independent of the event that  $X = \bar{\mathbf{x}}$ , and hence  $E[T_{\bar{\mathbf{x}},\mathbf{1}}|X = \bar{\mathbf{x}}] = ET_{\bar{\mathbf{x}},\mathbf{1}}$ , which gives the desired result.  $\square$

The mean unknotting time is thus at most the mean first-passage time  $f(m)$  from  $\mathbf{0}$  to  $\mathbf{1}$  on the  $m$ -cube, with  $m$  at most  $n/2$ . It is well known that  $f(m)$  is equivalent to the mean first-passage time from state 0 to state  $m$  in the *classical* Ehrenfest urn model with  $m$  balls, which has the transition probabilities (4) with  $a = b = 1$ . The quantity  $f(m)$  has been studied in [1, 4, 17, 18]. In particular, the asymptotic formula  $f(m) \sim 2^m$  was obtained in [1]. Setting  $i = 0$  and  $j = m$  in equation (8) gives

$$f(m) = M_{0,m} = \frac{m}{2} \sum_{k=1}^m \frac{1}{k} 2^k \leq m 2^m. \quad (13)$$

We can then absorb the factor  $m$  into the exponential growth. Since  $m \leq n/2$ , we obtain an upper bound  $\lambda^n$ . This completes the proof of theorem 1.2.



**Figure 8.** Mean unknotting time versus length of generating SAPT.

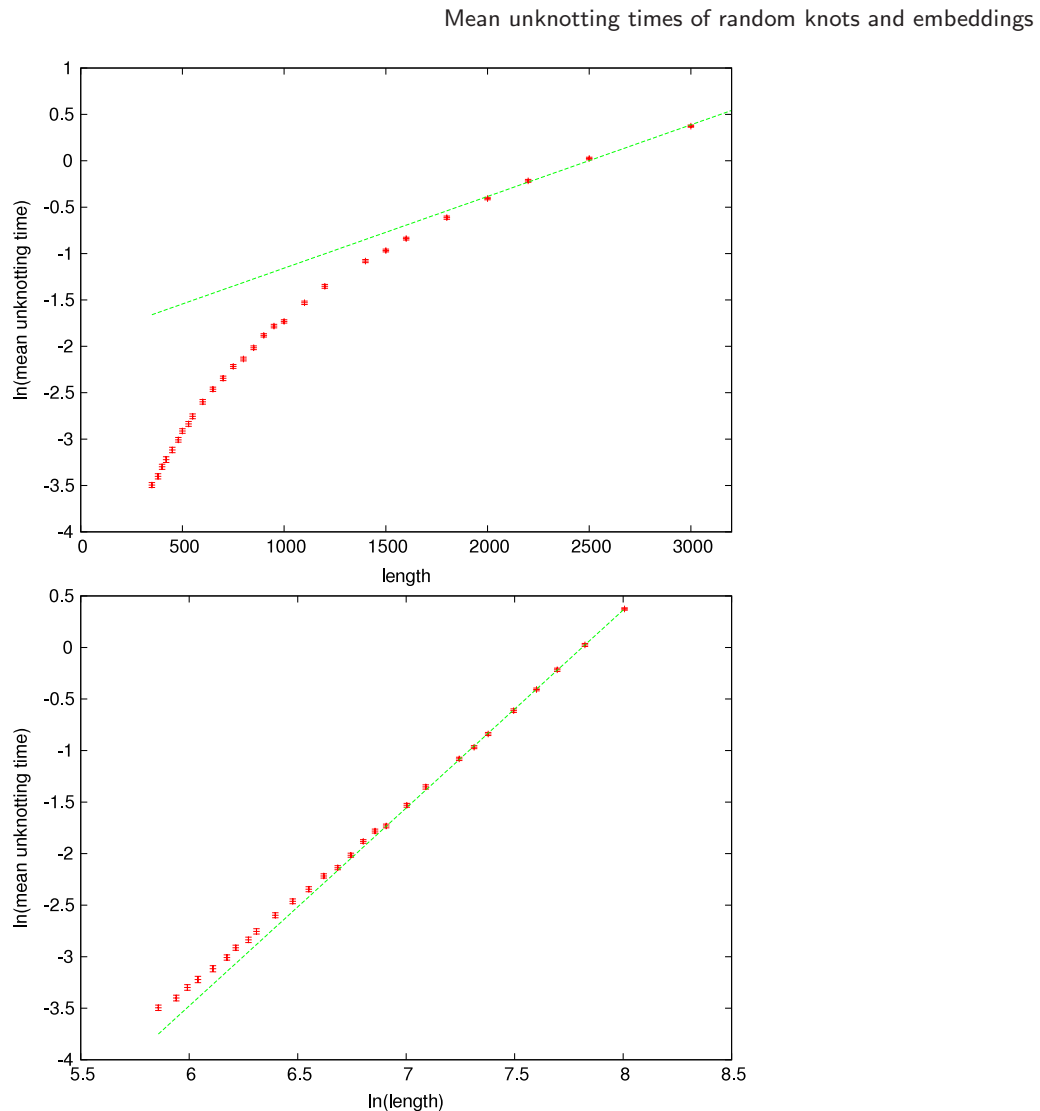
#### 4. Numerical results

We now turn to simulation to estimate the exponential growth rate of the mean unknotting time of SAPTs via random crossing reversals. The pivot algorithm is a standard algorithm for the simulation of self-avoiding walks and self-avoiding polygons [15]. It is explained in [6] how to adapt the pivot algorithm to SAPTs. In particular, it is shown in [6] that the pivot algorithm is a valid algorithm for SAPTs, in the sense that its stationary distribution is uniform. We apply the pivot algorithm to generate random SAPTs of a given length. Crossings are then randomly assigned to be over- or underpasses. We then observe how long they take to unknot by random crossing reversals, i.e. crossings are picked uniformly at random and reversed until the unknot is obtained.

To decide whether or not the current embedding is the unknot we use the Alexander polynomial. The Alexander polynomial of an embedding with  $n$  crossings can be computed in  $O(n^3)$  time (using the algorithm described in [2]). The Alexander polynomial is not a perfect knot invariant, and it is possible that it incorrectly identifies an embedding as the unknot; the smallest non-trivial knot whose Alexander polynomial is equal to that of the unknot contains 11 crossings. We observe that the vast majority of crossings in a SAPT are twists such as those in figure 2 and so can be removed without changing the knot type. Since other knot invariants, such as the Jones polynomial, are far more time consuming to compute, we have worked under the assumption that misclassifying the unknot is very unlikely.

We calculated the mean unknotting times for SAPTs of lengths up to 3000. The results are shown in figure 8, with error bars computed using autocorrelation times. Since our analytic results indicate that the mean unknotting time grows exponentially with length, we tried to fit our data to a curve of the form  $Ae^{\lambda n}$ . However, the exponential growth is extremely weak, and so we were unable to form precise estimates. For example, fitting the last four data points to an exponential curve gives:

$$\text{mean unknotting time} \sim 0.15 (1.0008)^n, \quad (14)$$



**Figure 9.** The mean unknotting time on a log-linear scale (top) and a log-log scale (bottom). The lines are the results of fitting the last four points of the data on these scales and correspond to exponential and power-law growth, respectively.

where  $n$  is the length of the SAPT. We do not give error bars for these parameters, as they should only be considered as a rough guide to demonstrate that the growth rate is indeed small. We did not attempt to fit corrections to scaling since our estimates of the exponential term are so imprecise. For the sake of comparison, we fitted the data using a power law  $Bn^\gamma$ ; using the last four data points we found that the mean unknotting time was approximately  $(3 \times 10^{-7})n^{1.9}$ . The fitted data is shown in figure 9. While the power law is a better fit to the data, we found that the fitted exponent increased with length; this is consistent with weak exponential growth of the mean unknotting time.

The exponential growth implied by our analytic results comes about because a typical very large SAPT has a positive density of trefoils, but we have no estimate of how large the SAPT has to be to observe this behaviour. At length 3000, for example, we find that the average number of crossings is approximately 50. One might expect that this is sufficient to guarantee that the embedding contains several trefoils. However, we found

that typically almost all of these crossings are twists that can be removed using a type I Reidemeister move (see figure 2). So while a typical configuration of length 3000 contains far more crossings than is necessary to produce a knot, we find that only about 1 in 10 configurations are actually knotted. Furthermore, the knotted configurations typically have low unknotting number—of the knotted configurations, a further 9 out of 10 are trefoils at this length. This is consistent with the highly localized nature of flat knots found in [9, 16].

To obtain a better estimate of the growth of the mean unknotting time, we would need to extend the simulations to far greater lengths, which would require far greater computing resources. Alternatively, one could skew the distribution used to sample SAPTs, to favour more compact configurations with a higher number of crossings. This would lead to more complicated knots for the same SAPT length.

## 5. Conclusion

We have studied the mean unknotting time of random self-avoiding polygons, and of random self-avoiding polygon trails with random crossing allocations, using both analytic and numerical methods. We have proved an exponential lower bound on the mean unknotting time of both SAPs and SAPTs by relating the models to the problem of the mean first-passage time in the two-parameter Ehrenfest urn model. We have also proved an exponential upper bound for the mean unknotting time of SAPTs. Simulations of random SAPTs of length up to 3000 show that the exponential growth is very weak: configurations of moderate lengths can be unknotted using only a small number of random crossing reversals.

The exponential lower bound for SAPs still holds in the case of  $s > t > 0$ , in which the process is driven towards the unknotted state. A long SAP will typically contain a linear number of trefoil arcs and the process will quickly unknot many of these. However, as long as there is a non-zero probability of reknotted a trefoil, it will always be difficult to find and unknot the last few trefoils without reintroducing knots. Indeed we have shown that the mean time to nearly unknot either a SAP or SAPT (by removing all but a small fraction of the knots) also grows exponentially.

Our lower bound is quite robust and will apply generally in situations in which a typical closed curve contains a positive density of tight trefoil arcs. In particular, Diao *et al* [8] have proved a continuum version of Kesten's pattern theorem for Gaussian random polygons (GRPs). A GRP is a piecewise linear circle in  $\mathbb{R}^3$  whose edges are normally distributed vectors. Theorem 3 in [8] shows that there exist positive  $\beta$  and  $\epsilon$  such that for large enough  $n$ , a GRP of  $n$  steps contains  $n^\beta$  trefoils with probability at least  $1 - e^{-n^\epsilon}$ . Our results then imply that the mean unknotting time of a randomly chosen GPR grows at least as fast as  $\lambda^{n^\beta}$  for some  $\lambda > 1$  (with probability at least  $1 - e^{-n^\epsilon}$ ).

Our upper bound has been proved for the case of two-dimensional SAPTs with uniformly random crossing reversals. Since random crossing reversal is a naive strategy for trying to reach the unknot, any more intelligent process should reach the unknot more quickly, and we expect our upper bound also to apply quite generally.

## Acknowledgments

Financial support from the Australian Research Council and the Centre of Excellence for Mathematics and Statistics of Complex Systems is gratefully acknowledged by YBC, ALO



and AR. The work of GS was supported in part by NSERC of Canada. YBC would like to thank the University of Melbourne, CSIRO and the Australian National University. GS is grateful to Tony Guttman for hospitality in Melbourne. We thank Stu Whittington for many helpful discussions and E J Janse van Rensburg for providing the computer program used to analyse autocorrelations in our simulations. Finally, we are grateful to the anonymous referees for their helpful suggestions which motivated us to extend our results for two-dimensional SAPTs to three-dimensional SAPs.

## References

- [1] Aldous D J, *Some inequalities for reversible Markov chains*, 1982 *J. Lond. Math. Soc.* **25** 564
- [2] Alexander J W, *Topological invariants of knots and links*, 1928 *Trans. Am. Math. Soc.* **30** 275
- [3] Bernhard J A, *Unknotting numbers and minimal knot diagrams*, 1994 *J. Knot Theor. Ramif.* **3** 1
- [4] Bingham N H, *Fluctuation theory for the Ehrenfest urn*, 1991 *Adv. Appl. Probab.* **23** 598
- [5] Bleiler S A, *A note on unknotting number*, 1984 *Math. Proc. Camb. Phil. Soc.* **96** 469
- [6] Chan Y, *Selected problems in lattice statistical mechanics*, 2005 PhD Thesis The University of Melbourne
- [7] Dette H, *On a generalization of the Ehrenfest urn model*, 1994 *J. Appl. Probab.* **31** 930
- [8] Diao Y, Pippenger N and Sumners D W, *On random knots*, 1994 *J. Knot Theor. Ramif.* **3** 419
- [9] Guitter E and Orlandini E, *Monte Carlo results for projected self-avoiding polygons: a two-dimensional model for knotted polymers*, 1999 *J. Phys. A: Math. Gen.* **32** 1359
- [10] Janse van Rensburg E J, Orlandini E, Sumners D W, Tesi M C and Whittington S G, *Entanglement complexity of lattice ribbons*, 1996 *J. Stat. Phys.* **85** 103
- [11] Kawamura T, *The unknotting numbers of  $10_{139}$  and  $10_{152}$  are 4*, 1998 *Osaka J. Math.* **35** 539
- [12] Kazez W H (ed), 1993 *Geometric Topology* vol 2 (Providence, RI: American Mathematical Society)
- [13] Kesten H, *On the number of self-avoiding walks*, 1963 *J. Math. Phys.* **4** 960
- [14] Krafft O and Schaefer M, *Mean passage times for tridiagonal transition matrices and a two-parameter Ehrenfest urn model*, 1993 *J. Appl. Probab.* **30** 964
- [15] Madras N and Slade G, 1993 *The Self-Avoiding Walk* (Basel: Birkhäuser)
- [16] Metzler R, Hanke A, Dommersnes P G, Kantor Y and Kardar M, *Equilibrium shapes of flat knots*, 2002 *Phys. Rev. Lett.* **88** 188101
- [17] Palacios J L, *Fluctuation theory for the Ehrenfest urn via electric networks*, 1993 *Adv. Appl. Probab.* **25** 472
- [18] Palacios J L, *Another look at the Ehrenfest urn via electric networks*, 1994 *Adv. Appl. Probab.* **26** 820
- [19] Palacios J L and Tetali P, *A note on expected hitting times for birth and death chains*, 1996 *Stat. Probab. Lett.* **30** 119
- [20] Pippenger N, *Knots in random walks*, 1989 *Discrete Appl. Math.* **25** 273
- [21] Reidemeister K, *Knoten und gruppen*, 1927 *Abh. Math. Sem. Univ. Hamburg* **5** 7
- [22] Roca J, *The mechanisms of DNA topoisomerases*, 1995 *Trends Biochem. Sci.* **20** 156
- [23] Rybenkov V V, Ullsperger C, Vologodskii A V and Cozzarelli N R, *Simplification of DNA topology below equilibrium values by type II topoisomerases*, 1999 *Science* **277** 690
- [24] Soteros C E, Sumners D W and Whittington S G, *Entanglement complexity of graphs in  $\mathbb{Z}^3$* , 1992 *Math. Proc. Camb. Phil. Soc.* **111** 75
- [25] Stoimenow A, *Some examples related to 4-genera, unknotting numbers and knot polynomials*, 2001 *J. Lond. Math. Soc.* **63** 487
- [26] Sumners D W and Whittington S G, *Knots in self-avoiding walks*, 1988 *J. Phys. A: Math. Gen.* **21** 1689
- [27] Wang J C, *DNA topoisomerases*, 1982 *Sci. Am.* **247** 94
- [28] Wang J C, *Cellular roles of DNA topoisomerases: a molecular perspective*, 2002 *Nat. Rev. Mol. Cell Biol.* **3** 430
- [29] Wasserman S A, Dungan J M and Cozzarelli N R, *Discovery of a predicted DNA knot substantiates a model for site-specific recombination*, 1985 *Science* **229** 171
- [30] Yan J, Magnasco M O and Marko J F, *A kinetic proofreading mechanism for disentanglement of DNA by topoisomerases*, 1999 *Nature* **401** 932