The Knot Spectrum of Confined Random Equilateral Polygons

Research Article

Y. Diao[1*], C. Ernst[2†], A. Montemayor[2], E. Rawdon[3‡], U. Ziegler[4§]

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

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

3 Department of Mathematics
University of St. Thomas
St. Paul, MN 55105, USA

4 Department of Computer Science
Western Kentucky University
Bowling Green, KY 42101, USA

Abstract: It is well known that genomic materials (long DNA chains) of living organisms are often packed compactly under extreme confining conditions using macromolecular self-assembly processes but the general DNA packing mechanism remains an unsolved problem. It has been proposed that the topology of the packed DNA may be used to study the DNA packing mechanism. For example, in the case of (mutant) bacteriophage P4, DNA molecules packed inside the bacteriophage head are considered to be circular since the two sticky ends of the DNA are close to each other. The DNAs extracted from the capsid without separating the two ends can thus preserve the topology of the (circular) DNAs. It turns out that the circular DNAs extracted from bacteriophage P4 are non-trivially knotted with very high probability and with a bias toward chiral knots. In order to study this problem using a systematic approach based on mathematical modeling, one needs to introduce a DNA packing model under extreme volume confinement condition and test whether such a model can produce the kind of knot spectrum observed in the experiments. In this paper we introduce and study a model of equilateral random polygons confined in a sphere. This model is not meant to generate polygons that model DNA packed in a virus head directly. Instead, the average topological characteristics of this model may serve as benchmark data for totally randomly packed circular DNAs. The difference between the biologically observed topological characteristics and our benchmark data might reveal the bias of DNA packed in the viral capsids and possibly lead to a better understanding of the DNA packing mechanism, at least for the bacteriophage DNA. The purpose of this paper is to provide information about the knot spectrum of equilateral random polygons under such a spherical confinement with length and confinement ratio in a range comparable to circular DNAs packed inside bacteriophage heads.

MSC: 57M25, 92B99

Keywords: DNA packing• topology of circular DNA • random polygons • DNA knots • knot spectrum of random polygons

* *E-mail: ydiao@uncc.edu*
† *E-mail: claus.ernst@wku.edu*
‡ *E-mail: ejrawdon@stthomas.edu*
§ *E-mail: uta.ziegler@wku.edu*

1. Introduction

It is well known that genomic materials (long DNA chains) of living organisms are often packed compactly under extreme confining conditions using macromolecular self-assembly processes. For example, in the prototypic case of the P4 bacteriophage virus, the $3\mu m$-long double-stranded DNA is packed within a viral capsid with a caliper size of about $50nm$, corresponding to a 70-fold linear compaction [1]. In the case of bacteriophages, the organization of the condensed DNA facilitates the process of DNA packing, provides stability to the capsid while packaged and also facilitates the release of DNA upon infection. The general DNA packing mechanism is a very important question in molecular biology that remains largely open. It is plausible and has been suggested in [2] that the topology of the packed DNA may be used to study the DNA packing mechanism. In the case of bacteriophage P4, DNA molecules packed inside the bacteriophage head are considered to be circular since the two sticky ends of the DNA are close to each other. Researchers are able to extracted the DNAs from the capsid without separating the two ends, essentially preserving the topology of the (circular) DNA. Since further relaxation of the DNA (needed in order to analyze the topology of the DNA) does not allow strand passage, the topology of the DNA remains intact. It is reported in [2] that the circular DNA extracted from bacteriophage P4 are non-trivially knotted with very high probability. In particular, a quantitative analysis of these DNA knots revealed that they are likely to be chirally organized based on a coarse knot spectrum of the extracted DNA knots [2]. In order to study this problem using a systematic approach based on mathematical modeling, one needs to introduce a DNA packing model under extreme volume confinement condition and test whether such a model can produce the kind of knot spectrum observed in the experiments.

The topology of the DNA packed inside a bacteriophage head is what motivated this study. Random equilateral polygons confined inside a sphere are a coarse model of circular DNAs packed inside bacteriophage heads where the confinement sphere models the virus head [3–5]. This model is chosen because of its clean and simple mathematical formulation and that it contains no bias (packing pattern); it is not meant to generate polygons that model DNA packed in a virus head directly. Instead, the average topological characteristics of this model may serve as benchmark data for totally randomly packed circular DNAs. The difference between the biologically observed topological characteristics and our benchmark data might reveal the bias of DNA packed in the viral capsids and possibly lead us to a better understanding of the DNA packing mechanism, at least for the bacteriophage DNA. The purpose of this paper is to provide information about the knot spectrum of equilateral random polygons in spherical confinement with length and confinement ratio in a range comparable to circular DNAs packed inside bacteriophage heads. The knot spectrum based on knots with a most seven crossings based on a simulation of viral packing of DNA in a virus has been reported in [6]. However the model used in [6] is complicated, and contains a number of choices to model actual DNA. Furthermore it models open DNA chains that use a closure at infinity schema to circularize. The choices made in such a model introduce a strong bias (with a strong effect on a knot distribution) and do not allow for a clean and simple mathematical formulation.

Unconfined equilateral random polygons (also known as ideal random polygons) have long been used to model ring polymers under the θ-conditions where polymer segments that are not in direct contact neither attract nor repel each other. It is a model that has been fairly thoroughly studied, both theoretically and numerically. Consequently, many theoretical aspects of the equilateral random polygons are well understood. For example, the mean squared distance between two vertices on an equilateral random polygon of length n that is k vertices apart is $k(n-k)/(n-1)$, the mean squared radius of gyration of such a random polygon is $(n+1)/12$ [7] and its mean ACN (average crossing number) behaves as $\frac{1}{16}n\ln n + o(n)$ [8, 9]. There are also several well tested algorithms for generating these unconfined equilateral random polygons. For example the crankshaft algorithm [10, 11], the hedgehog method [10, 12], the generalized hedge hog method [13] and the recently developed method by J. Cantarella *et al.* [14]. Topologically, it has been shown that the FWD conjecture holds true for the equilateral random polygons, *i.e.*, the probability that an equilateral random polygon of length n is a nontrivial knot goes to one as n goes to infinity [15]. However, detailed information on the knot spectrum of unconfined equilateral random polygons is still rather scarce. Numerical results reported in [16] show that knotting probability is about 0.1 for $n = 50$, about 0.3 for $n = 100$ and gradually increases to about .85 for $n = 500$.

Unlike equilateral random polygons without confinement, the confined equilateral random polygons have not been thoroughly studied and there are many unanswered questions. The first issue is how to define the models to reflect the various packing properties that DNA or polymer chains may have. Once a confined random polygon model is defined, the next issue is determining the probability distributions of the random polygons based on the model, and the third issue is the actual generation of the random polygons in accordance with these (theoretical) probability distributions. In a series of papers, four authors of this paper have developed algorithms for several models to generate equilateral random polygons that are confined inside a sphere of fixed radius [3–5]. The model presented in [4] is the one chosen for the study presented in this paper. The model can be described as follows: Consider equilateral random polygons that are "rooted" at the origin and assume that there is an algorithm that samples such objects with uniform probability. Now consider a confinement sphere S_R of radius $R \geq 1$ with its center at the origin. Only those randomly generated equilateral polygons are kept that are contained in the confinement sphere S_R. (Note that the algorithm used actually generates polygons in confinement using conditional probability density functions. The reader should note that using our algorithm each polygon generated is totally independent of the previous polygon and no de-correlation is necessary.) There is no biological or other reason for the polygons to be rooted at the center. It is rather a choice for simplicity: as it turns out, equilateral random polygons defined this way are much easier to generate due to the symmetry of the confining sphere (relative to the root) imposed on the equilateral random polygons.

In this paper we only discuss the effects of confinement and length of the random polygons on the topology (*i.e.* information based on knot types). In a future paper we shall discuss the effects of confinement, length and knotting complexity of the random polygons on the geometry (*i.e.* geometric quantities such as the average crossing number, curvature and torsion, and the writhe).

2. Polygon sampling and knot identification

In our sampling of the random polygons we considered different polygon lengths and different radii of confinement. For each fixed confinement radius and each fixed polygon length we constructed a sample of $10,000$ different polygons. We choose the different radii and lengths using three simple criteria: (i) the probability of knotting is relatively high; (ii) the knots should not be so complex that the identification of knot types becomes computationally impossible for almost all polygons and (iii) the length/confinement radius ratio should be a range that contains the linear compaction rate of a typical bacteriophage virus DNA (up to 70 for the P4 bacteriophage virus). Using these criteria we constructed our sample space as follows:

The lengths of the random polygons range from 10 to 90 in increments of 10. The confinement radii range from $R = 1$ to $R = 4.5$. Increments of $1/10$ are used between radii $R = 1$ to $R = 3$ and additionally radii $R = 3.5$, 4 and 4.5 are used. The reason for a denser sampling of R-values between one and three is that the knot-spectra change rapidly at smaller radii, but this change levels off for larger radii. For the longer polygons of lengths 60 to 90 we did not sample polygons over all confinement radii due to the fact that at small confinement radii the extreme knot complexity made it impossible for us to "identify" these knots. For example, polygons of lengths 70, 80 and 90 are only sampled at radii from $R = 1.5$ to $R = 4.5$ with increments of $1/2$. In total our sample space consists of 164 sets each containing $10,000$ polygons, yielding a total of $1,640,000$ random polygons.

One tool we used for knot identification is the HOMFLYPT polynomial; interested readers can refer to a knot theory textbook such as [17]. However there are many knot types that share a common HOMFLYPT polynomial. Thus the HOMFLYPT polynomial calculation provides only a list of chiral knot types (with 16 or fewer crossings) with that HOMFLYPT polynomial. The second tool we used was *knotfind* within knotscape [18], which uses Dowker codes to compute the exact knot type. However, *knotfind* does not compute the chirality of a knot type (*i.e.* it does not differentiate between the two chiral versions of a trefoil).

More specifically, we used the following process to extract the knot identifications from the polygons. For each polygon *unraveller* [19] is used to simplify the crossing information. *Unraveller* creates a projection into the plane from which a crossing code is extracted. Via a collection of simplification moves (based on Reidemeister moves) this crossing code is (potentially) simplified. *Unraveller* produces (at least) two types of output: a DT-code which is used by *knotfind* and crossing information which is used to compute the HOMFLYPT polynomial. For each polygon P we compute the HOMFLYPT polynomial H using a program written by Ewing and Millett [20]. We have a table of HOMFLYPT polynomials for all chiral knot types, prime and composite, with 16 or fewer crossings (under the generally accepted assumption that the crossing number of composite knots is additive). For each P, we obtain a list of chiral knot types $\{K_1, \ldots, K_s\}$ from the table all of which have H as their HOMFLYPT polynomial. Note that if this list is empty then the polygon does not represent any prime or composite knot with less than 17 crossings. In addition to the HOMFLYPT polynomial calculation, for each polygon we also used *knotfind* to compute the (non-chiral) knot type. We use the simplified DT-code D generated by *unraveller*, which

might be the DT-code of a minimal diagram or might be the DT-code of a diagram close to the minimal diagram. Then *knotfind* takes D and creates a "canonical" DT-code D'. If D' represents a knot within the DT-code knot table (which only contains prime knot types) then that uniquely identifies that knot (up to chirality).
If the non-chiral knot type from *knotfind* matches a chiral knot type in our list $\{K_1, \ldots, K_s\}$ based on the HOMFLYPT polynomial then this (chiral) knot type is identified as the correct one. This procedure should be totally reliable in identifying knots with less than or equal to 16 crossings. We had only one disagreement where the HOMFLYPT and the *knotfind* processes 'identified' incompatible knot types with 16 or fewer crossings. This disagreement was resolved using a second simplification process based on the planar-diagram code (PD-code) of the polygon, modifying the PD-code into a DT-code and looking it up with knotspace. However there are still several additional issues which we address in the following paragraphs.

(i) The polygon P might represent a knot that is not prime and therefore it is not in the DT-code knot table. At every stage of the simplification process using DT-codes, *knotfind* attempts to identify factors of composite knots. If there is an "obvious" connected sum then a part of the DT-code maps onto itself. If such a situation is detected the simplification process is applied separately to each DT-code of the two factors. In the end we obtain a collection of simplified DT-codes from which it is possible to reconstruct the original composite knot type. If the HOMFLYPT polynomial of this reconstruction agrees with the originally computed HOMFLYPT polynomial of the polygon then we identify P as the appropriate composite knot.

(ii) The polygon P might produce a simplified DT-code but the simplified DT-code does not match any knots with 16 or fewer crossing, and additionally the HOMFLYPT polynomial H of P leads to a list of one or more knot types with 16 or fewer crossings. In such a case, either *unraveller* and *knotfind* were not able to simplify the crossings to the point that the knot could be identified, or the knot has more than 16 crosssings but shares its HOMFLYPT polynomial with one or more knot types with 16 or fewer crossings. Cases for which the DT-code indicates that the knot has 16 or fewer crossings, are resolved based on running the simplification process based on the planar diagram-code (PD-code) of the polygon. In all cases, the PD-simplification confirmed the crossing number provided through the DT-code and P is identified as a knot with 16 or fewer crossings (which means that the translation from simplified DT-code to canonical DT-code failed). In cases for which the DT-code indicates that the knot has more than 16 crossings we identify the knot similar to (iii).

(iii) If the initial calculation of H indicates that P does not represent a knot in the table, then the simplified DT-code (or the set of simplified DT-codes in the case an obvious connected sum was identified) provides an upper bound on the crossing number. In these cases we double-checked by computing the HOMFLYPT polynomial from the simplified DT-code(s) to ensure that it matches H. We identify P as having the crossing number provided by the simplified DT-code.

We observed that the approach to knot simplification based on DT-codes is extremely reliable for the knots that are within the knot table. There is no reason to believe that the simplification of DT-codes becomes suddenly

unreliable once the actual crossing number exceeds sixteen. The simplification method using PD-codes (which also results only in an upper bound) was executed on a sample of polygons with contradictory information (see ii). It did find PD-codes with fewer crossings (up to 8 fewer crossings) than the simplified DT-codes from *knotfind* but only for polygons for which the *knotfind* had an upper bound between 33 and 53. The PD-code simplification could not produce a better result (fewer crossings in the code) than the DT-code simplification for crossing numbers between 17 and 33. Thus we believe that a simplified DT-code, while technically only providing an upper bound on the actual crossing number, gives a value that in most cases is actually the topological minimal crossing number. In particular this is true for crossing numbers that are not far above 16 (the largest number in the knot table.) Therefore we report these approximated crossing numbers as if they are the actual crossing numbers. We also report a knot as prime if no composition was detected during the simplification process.

As a final comment, we remark that in some instances the above procedures fail. This could happen if the calculation of either the HOMFLYPT polynomial or the various simplification processes crash or simply "time-out" (i.e. we allow only a finite predetermined computation time and if this is exceeded then we call this a time-out). We note that only 79 (or 0.005%) random polygons for which the HOMFLYPT calculation timed-out. There were many more cases where the simplification process crashed or timed-out. In total such cases are rare (28,349 polygons representing 1.7286 percent of the total) and have only a small effect on the overall statistics. Most of these cases (28,115 to be exact) occurred in a few sets, see Table 1. The missing 166 are scattered overt many sets with no number exceeding 20 and are not reported in the table. Thus it seems safe to assume that most of these 28,349 polygons represent very complicated knots. However we cannot rule out that some of these actually are of a very simple knot type.

In summary, the reader should keep in mind that all numbers involving topological information are subject to small errors (as explained). Additionally, the crossing number estimate and the claim of a prime knot become less and less reliable as the crossing number increases beyond 16. In this paper we simply state these numbers without any indication of a potential error.

Table 1. The number of random polygons whose knot type/crossing number we failed to identify/estimate. 10,000 polygons were generated for each set.

# observed	length	radius	# observed	length	radius	#observed	length	radius
7,711	90	1.50	1,159	70	1.50	120	40	1.00
5,504	60	1.10	1,011	50	1.10	100	80	2.00
4,349	80	1.50	502	90	2.00	68	50	1.30
2,836	60	1.20	428	60	1.40	42	60	1.60
2,787	50	1.00	287	50	1.20			
1,186	60	1.30	135	60	1.50			

3. Overview of the knot spectrum

Overall Summary Out of the 1,640,000 random polygons sampled:

- 938,417 (57.22%) are unknots;
- 177,602 (10.83%) are trefoils;
- 611,383 (37.28%) are prime knots and 61,851 (3.77%) are composite knots (the crossing numbers range from 3 to 53);
- 127,698 (7.79%) are knots which have more than 16 crossings (out of the 673,234 knotted polygons).

More detailed information is shown in Table 2 for polygons with crossing numbers up to 23 (even though we have data up to 53). A graphic display of all crossing numbers is shown in Figure 1.

Table 2. **The number of observed knots**

crossing number	# observed knots	# observed prime knots	# observed composite knots	percentage composite knots
0	938,417	938,417	0	
3	177,602	177,602	0	
4	49,549	49,549	0	
5	57,449	57,449	0	
6	50,805	35,233	15,572	30.65
7	31,731	24,477	7,254	22.86
8	36,023	28,086	7,937	22.03
9	25,282	19,401	5,881	23.26
10	23,976	19,988	3,988	16.63
11	20,544	16,717	3,827	18.63
12	17,665	15,022	2,643	14.96
13	16,386	14,017	2,369	14.46
14	13,903	12,057	1,846	13.28
15	13,046	11,487	1,559	11.95
16	11,575	10,223	1,352	11.68
17	10,313	9,377	936	9.08
18	9,544	8,674	870	9.12
19	8,537	7,857	680	7.97
20	7,917	7,323	594	7.50
21	7,328	6,770	558	7.61
22	6,827	6,324	503	7.37
23	6,317	5,912	405	6.41

The knot spectrum in pictures In the following we show the knot spectrum in a sequence of graphs for selected crossing numbers that are multiples of three. The x-axis is the confinement radius and the y-axis is the number of knots encountered with the given property. For each of the nine different lengths in our sample we have connected the points of the same length with lines.

Note that in Figure 2 many curves have a distinct maximum. For example, for length 90 the number of knots with 15 crossings peaks at radius $R = 2.5$. For smaller radii the number of 15 crossing knots drops because at length

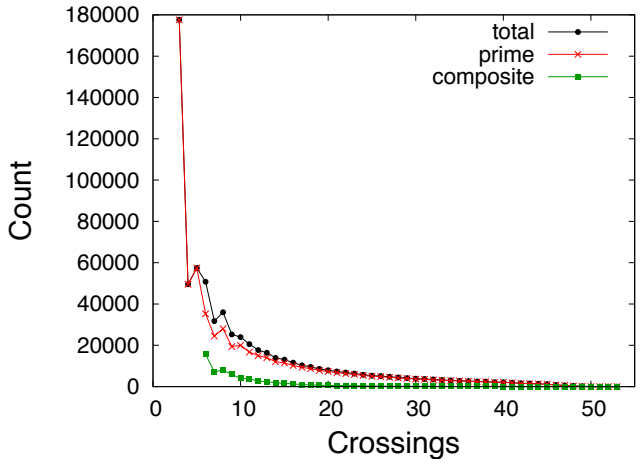

Figure 1. The number of knots with a fixed crossing number together with the number of prime and composite knots.

90 knotting complexity exceeds 15 crossings and knots of low complexity (such as knots of 15 crossings) become less likely. In Figure 3 we see the knot spectrum of composite knots and of unknown knots, that is polygons for which we neither were able to identify the knot nor to determine an upper bound on its crossing number.

4. Topological crossing number of confined random polygons

In this section we investigate the average topological crossing number TCN, *i.e.* the average minimal crossing number of our polygons. We want to remind the reader that for knotted polygons with more than 16 crossings we only have upper bound estimates for their crossing numbers. Moreover there are knotted polygons for which we do not even have an estimate for their crossing numbers, see Table 1 and the averages given here do not take into account these later polygons. The data is shown in Figure 4. In both graphs we see nine curves representing the polygons of length 10 to 90. The graph on the left contains only the information coming from knotted polygons and as the confinement radius increases all curves level off at crossing number three, that is to say the only knots that are left are trefoils. In the graph on the right we conclude in the TCN-average the unknotted polygons. Here we see that the TCN approaches zero as the radius R increases. For small radii R there is not much difference between the two graphs. On the left in Figure 5 there we show the 2D data from the left of graph in Figure 4 in a 3D view where the surface (the expected TCN discounting the unknots) is given by the equation

$$E(\mathrm{TCN}) \approx \left(-0.000796389 + \frac{0.011998}{\mathrm{R}^2}\right)\mathrm{L}^2 + \left(1.24109 - \frac{1.09742}{\mathrm{R}}\right)\ln(\mathrm{L}), \tag{1}$$

with R being the radius of confinement and L being the length of the polygon.

This equation fits the data with a $R^2 \approx 0.9890$ and a max and a mean difference between the surface and the data of about 2.956 and -0.072 respectively. This fitting equation is closely related to the fitting equation given in [9]

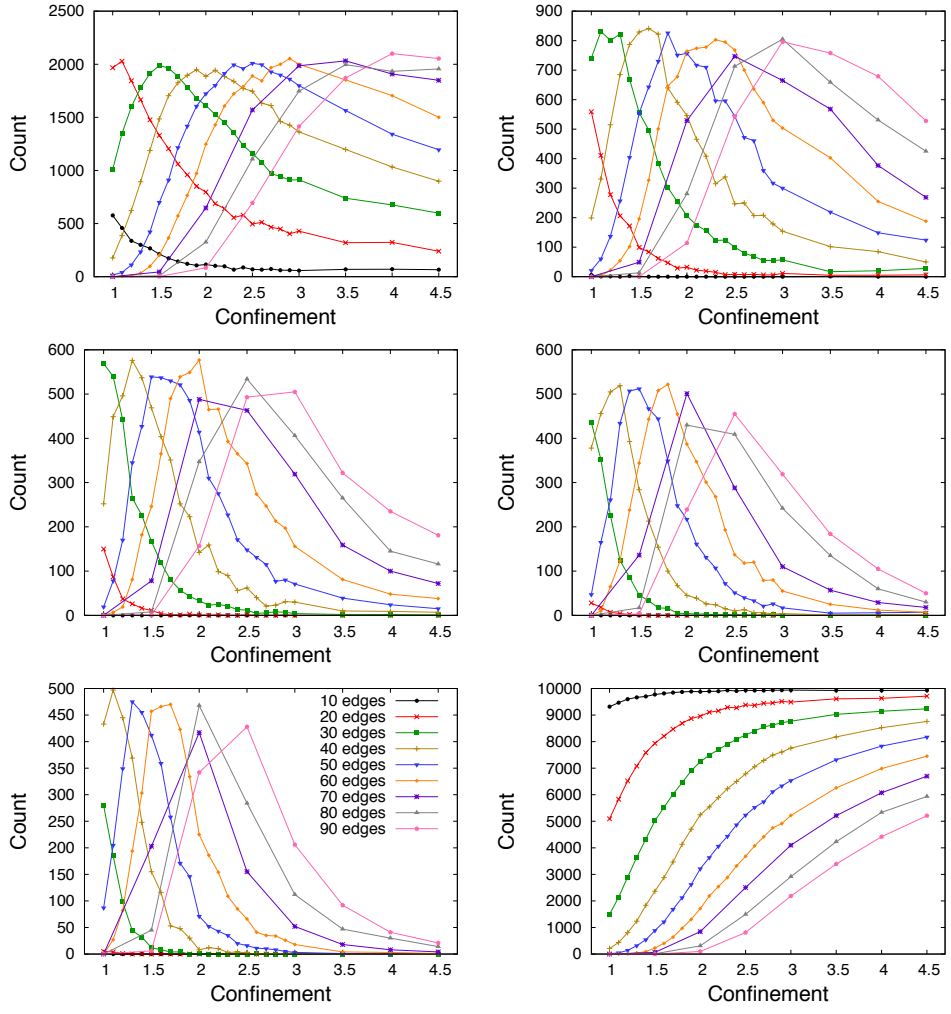

Figure 2. The knot spectrum for 3, 6, 9, 12 and 15 crossings. The last graph shows the number of unknots. In all cases except the case of the unknot, the curves are aligned at $R = 4.5$ in the decreasing order starting at length 90 on the top. In the case of the unknot, the curves are aligned at $R = 4.5$ in the increasing order starting at length 10 on the top.

for the mean average crossing number (mean ACN) of random polygons under confinement. It was established that the mean ACN of a random polygon in confinement grows quadratically with respect to the length of the polygon [9], and that the mean ACN scales as $\frac{3}{16}L\ln(L) + O(L)$ in the unconfined case (where L is the length of the polygon). A fitting formula of the form $a(R)L^2 + b(R)L\ln(L)$ is thus proposed in [9] where $a(R)$ is a decreasing function of R that starts as a positive number when R is small and decreases to near 0 as R increases to pass the average radius of gyration of the polygons (that is when the confining effect disappears); similarly, $b(R)$ is an increasing function that starts from near 0 when R is small and increases to the constant 3/16 as R passes the average radius of gyration. It is plausible that in tight confinement the mean TCN behaves similarly to the mean ACN (and our data clearly indicates that both the ACN and the TCN grow quadratically with length under a confinement condition). However it is rather intuitive and obvious to see that in the unconfined case, the mean

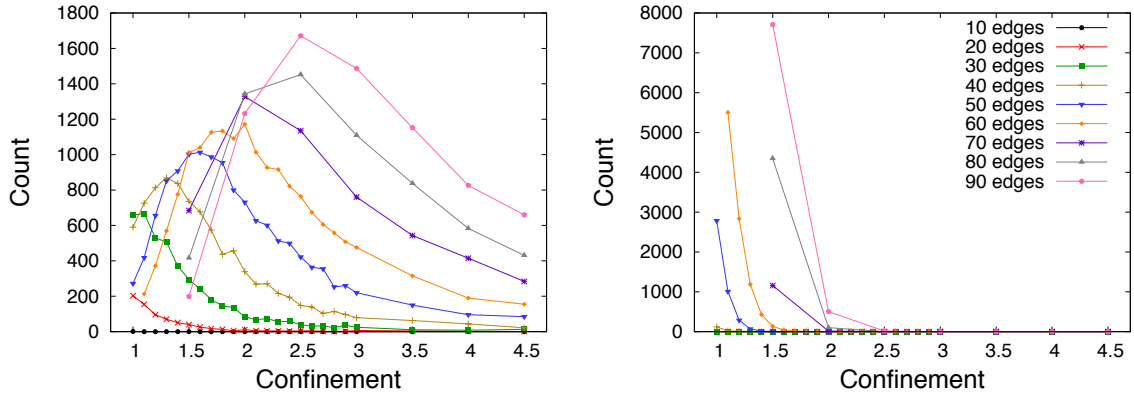

Figure 3. The knot spectrum for composite knots on the left and the number of unknown polygons (i.e. not even an upper bound could be identified) on the right. Again the curves are aligned at $R = 4.5$ in the decreasing order starting at length 90 on the top.

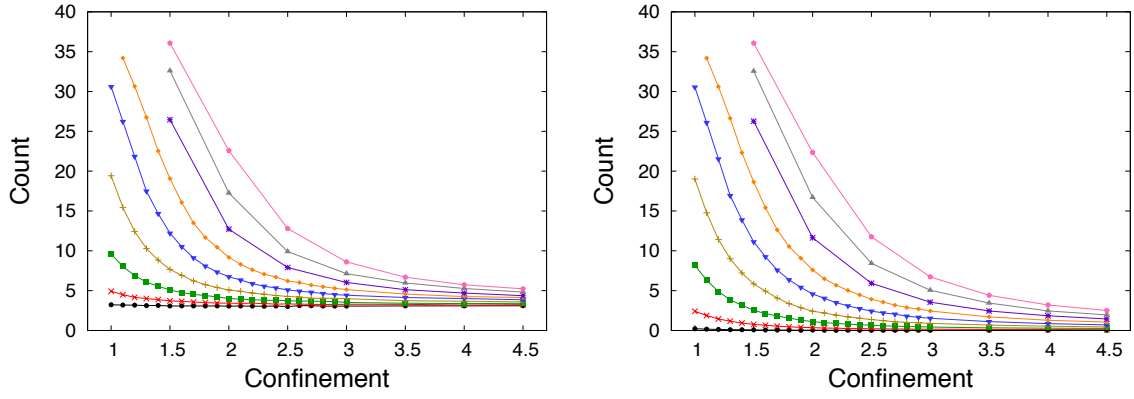

Figure 4. The average TCN dependence on the length and the confinement radius. The right side includes unknots, the left does not.

TCN would be much smaller than the mean ACN. Indeed, a fitting formula of the form $a(R)L^2 + b(R)L\ln(L)$ yielded a worse R^2 value. This actually raises an interesting question: Is it true that the mean TCN of a random equilateral polygon of length L scales as $O(\ln(L))$?

On the right in Figure 5 we analyze knot complexity in a different way. If we look at an individual curve in Figure 2 we notice that these curves often have a single maximum. For example in Figure 2 (top left) for polygons of length 30 and crossing number 3 this maximum occurs around $R = 1.5$. If we go to the top right then we find that for 6 crossings the maximum occurs at about $R = 1.1$. We have extracted this information for polygons of lengths 30 to 60 in Figure 5.

For some values we approximated the maximum by using an interpolation function through given data points to get a single value for the radius R. We did not use lengths 70 to 90 because we do not have enough data points and we did not use lengths 10 and 20 because for these short lengths we get essentially a constant radius $R = 1$.

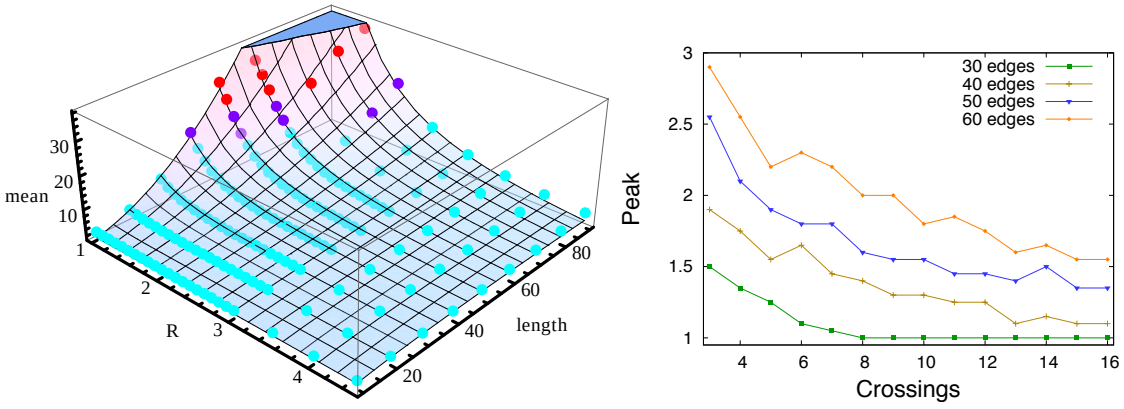

Figure 5. Left: The average TCN dependence on the length and the confinement radius in a 3D view. Right: The dependence of the maximum on length and confinement radius for lengths 30 (bottom) to 60 (top).

How do we interpret this data? For example our result indicates that for length 50 and crossing number 10 the maximum occurs at $R \approx 1.5$. This means that in order to generate 10 crossing knots of length 50, we have the highest yield if we generate them with a confinement radius $R \approx 1.5$. It also means that many knots are generated whose complexity is much higher than 10 crossings. In general, for a fixed polygon length, the more complex the knot, the smaller the confinement radius at which the maximum probability of generating such knots. Also, for a fixed crossing number, the longer the polygon the larger the confinement radius for this maximum. So for a fixed length, a very complicated knot either can never be generated or has a highest probability to be generated at the smallest allowable radius $R = 1$. The data points in Figure 5 do not tell us the magnitude of the probabilities. For example, if we consider length 30 then both 9-crossing knots and 15-crossing knots are generated with maximal probability when $R = 1$. However from Figure 2 is can be seen that the actualy number of knots generated is larger for 9-crossing kntos than for 15-crossing knots and is even smaller for knots with a larger crossing number. Using $R = 1$ does not help to generate knots with a very large crossing number because this maximal probability may be essentially zero.

5. Small polygonal knots: trefoils and figure eight knots

In our sample there are $177,602$ trefoils, that is, 10.83% of all polygons sampled are trefoils (or 26.38% of all knotted polygons sampled with identified crossing numbers are trefoils). The distribution of trefoils for each given length is shown in the top left of Figure 2. We found that there are $88,221$ positive trefoils (49.77%) and $89,381$ negative trefoils (50.33%). This data can be used to detect a possible bias in our polygon generating process, since a positive trefoil has the same probability to be generated than a negative trefoil if the generating process is unbiased. Given that a polygon generated is a trefoil, let p be the true conditional probability of a positive trefoil ($p = .5$ if the generating process is unbiased). The 95% confidence interval for p based on our sample data is $(.495372, .500028)$. Thus no obvious bias is detected. Here is another way to check on this. The average

writhe of all these trefoil-polygons is -0.018 with a sample variance of 16.613. From this one can construct a 95% confidence interval of the true mean writhe of all trefoil polygons, which is $(-.03696, .00096)$. This also seems reasonable as the mean writhe of all trefoil-polygons is zero for a unbiased generating process. Our sample of random polygons contained 761 triple trefoils ($3_1\#3_1\#3_1$), see Figure 6 There were even 26 quadruple trefoils in our sample space ($3_1\#3_1\#3_1\#3_1$). However strong confinement (*i.e.* small values of R) makes composite knots less likely as we shall see in Subsection 6.5 on composite knots.

In our sample there are 49,549 figure-eight knots, which accounts for 3.02% of all polygons sampled (or 7.36% of all knotted polygons sampled with identified crossing numbers). These knots have an average writhe of -0.0128, which is close to the expected value of zero.

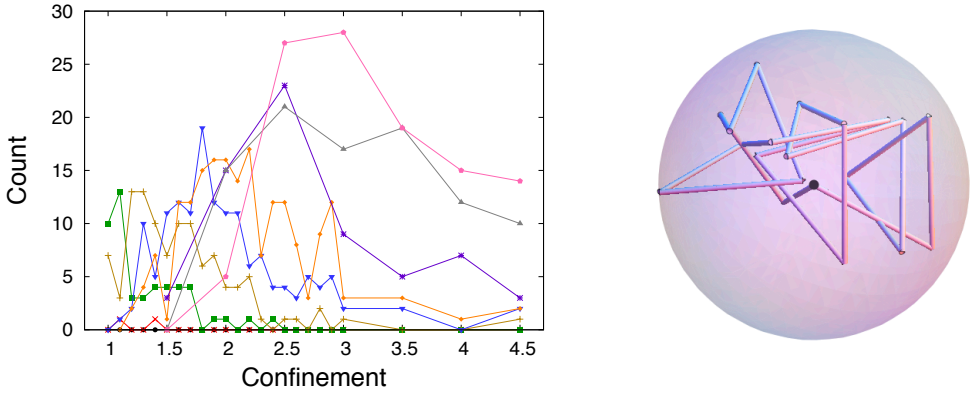

Figure 6. One the left the distribution of triple trefoils, on the right a 20 step triple trefoil in a confinement sphere of radius $R = 1.1$.

6. The frequencies of different knot types

6.1. Overview

In our sample there are 57,449 five-crossing knots: 20,984 of which are 5_1 knots (36.53%) and 36,465 of which are 5_2 knots (63.47%). There are more 5_2 knots almost across the entire range of different radii and lengths. In fact the only exceptions occur when the number of both knots is very small, which happens only for the shortest length $n = 10$ with one additional single exception at length $n = 90$ and $R = 2$. The distribution of the 50,805 six crossing knots is 12,191 of the 6_1 knots (24.00%), 14,488 of the 6_2 knots (28.52%), 8,554 of the 6_3 knots (16.84%) and 15,572 of the $3_1\#3_1$ knots (30.65%). Notice that the phenomenon that 5_2 is more likely than 5_1 and 6_2 is more likely than 6_1 or 6_3 has also been observed for unconfined random polygons [21, 22]. However in these studies the probability of generating more complicated knots is very low due to the lack of confinement, which weakens the statistical significance of these results. Notice that this is in strong contrast to a simulation

that models actual DNA in a viral capsid [6]. Here 5_2 is more likely than 5_1 and 6_1 is more likely than 6_2, while 6_3 almost does not occur at all.

For seven-crossing knots we have a total of 31,731 of them: of which 1,823 (7.75%) are 7_1 knots, 3,746 (11.80%) are 7_2 knots, 3,455 (10.88%) are 7_3 knots, 2,016 (6.35%) are 7_4 knots, 4,900 (15.44%) are 7_5 knots, 5,809 (18.31%) are 7_6 knots, 2,728 (8.60%) are 7_7 knots and 7,254 (22.86%) are $3_1\#4_1$ knots.

The frequencies of all 8 crossing polygonal knots are listed in Table 3. We note that the non-alternating prime knots 8_{19}, 8_{20} and 8_{21} clearly are more likely to appear than any of the alternating prime 8 crossing knots. We also observe that the least likely knot 8_{18} has a minimal diagram based on the 8^* basic Conway polyhedron. If each of the prime, alternating (non-alternating) 8-crossing knot types were equally likely, we would expect each to occur 1017.61 (3256.22) times on average.

Table 3. The frequencies of observed 8 crossing polygonal knots. Info in bold is related to knottypes related to Conway polyhedrons. Underlined knot types are fully amphicheiral.

8_1	8_2	$\underline{8_3}$	8_4	8_5	8_6	8_7	8_8	$\underline{8_9}$	8_{10}	8_{11}	$\underline{8_{12}}$
1,107	1,117	474	1,077	594	1,638	1,342	1,833	628	1,279	1,493	912
8_{13}	8_{14}	8_{15}	8_{16}	8_{17}	$\mathbf{8_{18}}$	8_{19}	8_{20}	8_{21}	$3_1\#5_1$	$3_1\#5_2$	$4_1\#4_1$
1,410	1,686	731	461	488	**47**	2,239	4,775	2,755	2,614	4,581	742

The frequencies of all 9 crossing polygonal knots are listed in Table 4. Again the knots with the largest frequencies 9_{42}, 9_{43}, 9_{44} and 9_{45} are non-alternating. We note that the lowest frequency is the knot 9_{40}, whose minimal diagram is based on the 9^* basic Conway polyhedron. Other knots with low frequencies are 9_{34} and the non-alternating knot 9_{47} both of which are obtained from the 8^* basic Conway polyhedron via a tangle substitution. If each of the prime, alternating (non-alternating) 9-crossing knot types were equally likely, we would expect each to occur 285.20 (963.50) times on average.

Table 4. The frequencies of observed 9 crossing polygonal knots. Info in bold is related to knottypes related to Conway polyhedrons. Underlined knot types are fully amphicheiral.

9_1	9_2	9_3	9_4	9_5	9_6	9_7	9_8	9_9	9_{10}	9_{11}	9_{12}	9_{13}	9_{14}
126	294	270	235	273	340	488	555	302	169	377	361	362	363
9_{15}	9_{16}	9_{17}	9_{18}	9_{19}	9_{20}	9_{21}	9_{22}	9_{23}	9_{24}	9_{25}	9_{26}	9_{27}	9_{28}
477	200	258	410	481	374	379	374	205	509	425	370	482	255
9_{29}	9_{30}	9_{31}	9_{32}	9_{33}	$\mathbf{9_{34}}$	9_{35}	9_{36}	9_{37}	9_{38}	9_{39}	$\mathbf{9_{40}}$	9_{41}	9_{42}
79	379	205	221	232	**57**	61	353	189	81	80	**8**	34	1,774
9_{43}	9_{44}	9_{45}	9_{46}	$\mathbf{9_{47}}$	9_{48}	9_{49}	$3_1\#6_1$	$3_1\#6_2$	$3_1\#6_3$	$4_1\#5_1$	$4_1\#5_2$	$3_1\#3_1\#3_1$	
1,336	1,843	1,495	696	**127**	245	192	1,349	1,514	865	499	893	761	

Finally we list the frequencies of 10 crossing polygonal knots in Table 5. Again the knots with the largest frequencies are non-alternating. The only non-alternating torus knot 10_{124} has a relatively large frequency. We note that the knot 10_{123} is the only knot that does not appear and that its minimal diagram is based on the 10^*

basic Conway polyhedron. The knots 10_{112} to 10_{120} are obtained from the 8^* basic Conway polyhedron via a tangle substitution, and the knots 10_{121}, 10_{122}, 10_{164} and 10_{165} are obtained from the 9^* basic Conway polyhedron via a tangle substitution. All these knots have a relatively low frequency. If each of the prime, alternating (non-alternating) 10-crossing knot types were equally likely, we would expect each to occur 63.65 (289.50) times on average.

Table 5. The frequencies of observed 10 crossing polygonal knots. Info in bold is related to knottypes related to Conway polyhedrons. Underlined knot types are fully amphicheiral.

10_1	10_2	10_3	10_4	10_5	10_6	10_7	10_8	10_9	10_{10}	10_{11}	10_{12}	10_{13}	10_{14}
87	83	63	56	84	94	83	66	69	89	65	96	88	83
10_{15}	10_{16}	$\underline{10_{17}}$	10_{18}	10_{19}	10_{20}	10_{21}	10_{22}	10_{23}	10_{24}	10_{25}	10_{26}	10_{27}	10_{28}
83	86	25	71	76	114	119	111	103	93	100	101	92	108
10_{29}	10_{30}	10_{31}	10_{32}	$\underline{10_{33}}$	10_{34}	10_{35}	10_{36}	$\underline{10_{37}}$	10_{38}	10_{39}	10_{40}	10_{41}	10_{42}
107	89	98	94	40	143	136	131	60	119	115	114	113	110
$\underline{10_{43}}$	10_{44}	$\underline{10_{45}}$	10_{46}	10_{47}	10_{48}	10_{49}	10_{50}	10_{51}	10_{52}	10_{53}	10_{54}	10_{55}	10_{56}
58	105	42	67	66	68	82	79	103	72	94	88	129	95
10_{57}	10_{58}	10_{59}	10_{60}	10_{61}	10_{62}	10_{63}	10_{64}	10_{65}	10_{66}	10_{67}	10_{68}	10_{69}	10_{70}
97	47	106	52	22	73	36	30	70	33	97	42	47	103
10_{71}	10_{72}	10_{73}	10_{74}	10_{75}	10_{76}	10_{77}	10_{78}	10_{79}	10_{80}	10_{81}	10_{82}	10_{83}	10_{84}
139	101	134	44	18	79	166	74	36	64	33	34	72	51
10_{85}	10_{86}	10_{87}	10_{88}	10_{89}	10_{90}	10_{91}	10_{92}	10_{93}	10_{94}	10_{95}	10_{96}	10_{97}	10_{98}
46	45	66	30	20	35	38	45	35	45	40	20	22	20
$\underline{10_{99}}$	10_{100}	10_{101}	10_{102}	10_{103}	10_{104}	10_{105}	10_{106}	10_{107}	10_{108}	10_{109}	10_{110}	10_{111}	$\mathbf{10_{112}}$
7	23	25	42	25	18	15	37	30	14	17	25	**13**	**11**
$\mathbf{10_{113}}$	$\mathbf{10_{114}}$	$\mathbf{10_{115}}$	$\mathbf{10_{116}}$	$\mathbf{10_{117}}$	$\mathbf{10_{118}}$	$\mathbf{10_{119}}$	$\mathbf{10_{120}}$	$\mathbf{10_{121}}$	$\mathbf{10_{122}}$	$\underline{\mathbf{10_{123}}}$	10_{124}	10_{125}	10_{126}
15	**10**	**10**	**14**	**15**	**10**	**22**	**6**	**6**	**7**	**0**	438	395	303
10_{127}	10_{128}	10_{129}	10_{130}	10_{131}	10_{132}	10_{133}	10_{134}	10_{135}	10_{136}	10_{137}	10_{138}	10_{139}	10_{140}
323	418	464	320	336	791	586	391	426	372	448	183	184	422
10_{141}	10_{142}	10_{143}	10_{144}	10_{145}	10_{146}	10_{147}	10_{148}	10_{149}	10_{150}	10_{151}	10_{152}	10_{153}	10_{154}
182	168	339	171	254	227	365	451	372	461	406	117	239	111
10_{155}	10_{156}	10_{157}	10_{158}	10_{159}	10_{160}	10_{161}	10_{162}	10_{163}	$\mathbf{10_{164}}$	$\mathbf{10_{165}}$	$3_1\#7_1$	$3_1\#7_2$	$3_1\#7_3$
126	197	37	102	166	241	288	142	64	**73**	**60**	175	359	346
$3_1\#7_4$	$3_1\#7_5$	$3_1\#7_6$	$3_1\#7_7$	$4_1\#6_1$	$4_1\#6_2$	$4_1\#6_3$	$5_1\#5_1$	$5_1\#5_2$	$5_2\#5_2$				
189	418	559	256	226	263	147	89	284	212				

The fact that knots with minimal diagrams close to basic Conway polyhedra all have low probabilities seems to imply that it is rare for a random polygon to possess this kind of order or symmetry. To examine this further, we looked at knot $12a1019$, which also has a special polyhedron called $12L$ [23]. It turns out that this knot was not observed in our sample either. Is this significant statistically? Actually we think it is. There are 2,176 different prime knots with 12 crossings and we have 15,022 prime knots with 12 crossings in our sample. Assuming that $12a1019$ has an equal chance to be sampled just like any other 12 crossing prime knot, then the probability that it is sampled each time a 12 crossing prime knot is selected is 1/2176. The probability that 15,022 independent selections all miss this knot is only .1%! However at this point we do not have a good explanation of this.

6.2. Alternating and non-alternating knots

Table 6 shows that prime alternating knots are becoming rarer (among all prime knots observed) as the knot complexity increases. It is known that as the crossing number goes to infinity, the percentage of prime alternating knots among all prime knots goes to 0 [24]. This trend has been picked up clearly by our data. However there is more to observe in our data. In the columns we list: the crossing number, the number of alternating knot types and the number of non-alternating knot types with that crossing number, followed by the the percentage of knot typess that are alternating. Then in the next two columns we show the number of observed alternating knots and the percentage of observed alternating knots. Finally, in the last column we compute the quotient of the two percentages. (That is the percentage of actual sampled alternating knots divided by the percentage of alternating knots from the knot table.) If all knot types would be equally likely to occur this number would be close to 1 for all entries. Thus, for example, the entry of 0.76 in the last column at crossing number 8 means that we only see 76% of the number of alternating knots that we would expect to see if all types would be equally likely. Clearly the table shows that on the average alternating knots are less and less likely to occur than non-alternating knots as the crossing number increases. Furthermore, this bias towards non-alternating knots seems to become stronger as the crossing number increases (as indicated by the generally decreasing order of the entries in the last column).

Table 6. **The frequencies of alternating and non-alternating prime knots**

crossing number	alternating knots	non-alt. knots	percentage alt. knots	observed alt. knots	obser. perc. alt. knots	quotient.
3	1	0	100.	177,602	100.	1.
4	1	0	100.	49,549	100.	1.
5	2	0	100.	57,449	100.	1.
6	3	0	100.	35,233	100.	1.
7	7	0	100.	24,477	100.	1.
8	18	3	86.	18,317	65.	0.76
9	41	8	84.	11,693	60.	0.72
10	123	42	75.	7,829	39.	0.53
11	367	185	66.	4,870	29.	0.44
12	1,288	888	59.	3,078	20.	0.35
13	4,878	5,110	49.	1,810	13.	0.26
14	19,536	27,436	42.	1,051	8.7	0.21
15	85,263	168,030	34.	546	4.8	0.14
16	379,799	1,008,906	27.	313	3.1	0.11

6.3. Fully amphichiral knots

If a knot is isotopic to both its reverse and its mirror image, it is fully amphichiral. The simplest knot with this property is the figure-eight knot. Just as alternating knots we can show that fully amphichiral knots are rare in our sample. To do this we use a similar method as in the case of alternating knots. However we only consider data up to 12 crossing since after that the percentage of these knots becomes too low for the collected data to

be statistically meaningful. The summary is shown in Table 7 where we do not show the odd crossing numbers since there are no fully amphichiral knots with an odd crossing number less or equal to 12.

Table 7. **The number of observed fully amphichiral prime knots.**

crossing number	all prime knots	amphi. knots	percentage amphi knots	observed amphi knots	% of observed amphi knots	quotient of percentages
4	1	1	100.	49,549	100.	1.
6	3	1	33.	8,554	24.	0.73
8	21	4	19.	2,061	7.3	0.39
10	165	7	4.2	232	1.2	0.27
12	2,176	17	0.78	31	0.21	0.26

6.4. Torus knots

Only the $T(n, 2)$ torus knots are alternating, all other torus knots $T(n, m)$ where $\gcd(m, n) = 1$ and $n > m > 2$ are highly non-alternating. As shown in Table 8 the non-alternating torus knots occur with surprisingly high frequencies compared to their alternating counterparts. In Table 8 we show the following in the columns moving from left to right: the crossing number; the torus knot symbol; the knot notation from the knot table of [18]; the number of observed torus knots; the expected frequency of a non-alternating knot if all non-alternating knots would be equally likely sampled; and finally the quotient of the observed frequency over the expected frequency. We can clearly see that the larger the crossing number becomes, the more pronounced the effect of increased probability for a non-alternating torus knot to appear becomes. Even at the small sample size, the effect is significant. After the horizontal bar in Table 8 we list the alternating $(n, 2)$-torus knots. Here we see the opposite effect, these knots are less frequent as they should be under the assumption that all alternating knots are equally likely. As with the non-alternating case this effect increases as the crossing number goes up.

Table 8. **The number of observed torus knots.**

crossing number	T(n,m)	knot notation	frequency observed	expected frequency	quotient
8	T(4,3)	8_{19}	2,239	3256.33	0.6876
10	T(5,3)	10_{124}	438	289.5	1.513
14	T(7,4)	14n21881	9	0.401	22.4354
15	T(5,4)	15n41185	7	0.065	107.505
16	T(8,3)	16n783154	4	0.01	407.31
3	T(3,2)	3_1	177,602	177,602.	1.
5	T(5,2)	5_1	20,984	28,724.5	0.7305
7	T(7,2)	7_1	1,823	3496.71	0.5213
9	T(9,2)	9_1	126	285.195	0.4418
11	T(11,2)	11a367	5	8.387	0.5962
13	T(13,2)	13a4878	0	0.371	0.
15	T(15,2)	15a85263	0	0.006	0.

6.5. Composite knots

Table 2 shows that the percentage of observed composite knots decreases as the crossing number increases. However it is generally expected (although it has not been theoretically established to the knowledge of the authors) that the number of composite knots grows faster than the number of prime knots as the crossing number increases. This tells us that large composite knots are less likely to be sampled under a confinement condition. In other words, confined equilateral random polygons are more likely to be prime knots. Similar information can be obtained from Figure 3. For example the top curve in both graphs in Figure 3 represents the polygons of length 90. On the left we see that the number of composite knots is declining as the radius decreases, while on the right we see that the number of knots so complicated that we cannot identify them increases drastically.

7. Ending remarks

Our numerical study has yielded many interesting results, some of these are expected and can be readily explained, while some other are surprising and are without a clear explanation. We end this paper with a few remarks on a few highlights of the paper.

• Confinement drastically increase knotting probability. This lends some explanation to as why a high percentage of bacteriophage DNAs are knotted.

• The spectrum of confined polygonal knots is wide, covering almost all small crossing knots up to 10 crossings. From a purely mathematical point of view, this means that confined equilateral polygons would be good tools to generate complicated knots (especially prime knots).

• The observed knot spectrum shows no obvious bias in knots with positive or negative writhe, but is clearly biased against alternating knots and some other knots with certain symmetry. One special case of this is the torus knot family and another special case is for knots with special symmetries -such as those obtained from a Conway polyhedron.

• The observed knot spectrum is heavily tilted towards prime knots rather than towards composite knots despite the fact that there are more composite knots in the knot table, indicating that confinement does not favor composite knots in general. This is in sharp contrast to the case of unconfined random equilateral polygons, where it has been shown that the probability for the polygon to be a composite knot tends to one as the length of the polygon goes to infinity [15]. The inevitability of knotting (namely the FWD conjecture) for confined equilateral random polygons has been a long standing open question. The numerical results here partly explain the difficulty: if a large proportion of the polygons consists of prime knots, then the technique used in [15] cannot be applied as it depends on the fact that a long, unconfined equilateral random polygon has a knotted component with high probability.

Acknowledgments

This work is supported in part by NSF Grants #DMS-0920880 and #DMS-1016460 (Y. Diao), by NSF grant #DMS-1016420 (C. Ernst, A. Montemayor, and U. Ziegler), and by NSF grant #DMS-1115722 (E. Rawdon)).

References

[1] P. J. Jardine and D. L. Anderson. Dna packaging in double-stranded dna phages. In Richard Calendar, editor, *The bacteriophages*, pages 49–65. Oxford University Press, 2006.

[2] Javier Arsuaga, Mariel Vazquez, Paul McGuirk, Sonia Trigueros, De Witt Sumners, and Joaquim Roca. DNA knots reveal a chiral organization of DNA in phage capsids. *Proc. Natl. Acad. Sci. USA*, 102:9165–9169, 2005.

[3] Yuanan Diao, Claus Ernst, Anthony Montemayor, and Uta Ziegler. Generating equilateral random polygons in confinement. *J. Phys. A: Math. Theor*, 44:405202, 2011.

[4] Y. Diao, C. Ernst, A. Montemayor, and U. Ziegler. Generating equilateral random polygons in confinement ii. *J. Phys. A: Math. Theor*, 45:275203, 2012.

[5] Y. Diao, C. Ernst, A. Montemayor, and U. Ziegler. Generating equilateral random polygons in confinement iii. *J. Phys. A: Math. Theor*, 45:465003, 2012.

[6] D. Marenduzzo, E. Orlandini, A. Stasiak, D.W. Sumners, L. Tubiana, and C. Micheletti. Dna-dna interactions in bacteriophage capsids are responsible for the observed dna knotting. *Proc Natl Acad Sci U S A.*, 106(52):22269–74, 2009.

[7] L. Zirbel and K. Millett. Characteristics of shape and knotting in ideal rings. *J. Phys. A: Math. Theor*, 45:225001, 2012.

[8] Yuanan Diao, Akos Dobay, Robert B. Kusner, Kenneth C. Millett, and Andrzej Stasiak. The average crossing number of equilateral random polygons. *J. Phys. A*, 36(46):11561–11574, 2003.

[9] J. Arsuaga, B. Borgo, Y. Diao, and R. Scharein. The growth of the mean average crossing number of equilateral polygons in confinement. *J. Phys. A: Math. Theor*, 42:465202, 2009.

[10] K. V. Klenin, A. V. Vologodskii, V. V. Anshelevich, A. M. Dykhne, and M. D. Frank-Kamenetskii. Effect of excluded volume on topological properties of circular DNA. *J. Biomol. Struct. Dyn.*, 5:1173–1185, 1988.

[11] K. C. Millett. Monte carlo explorations of polygonal knot spaces. In C. McA. Gordon, V. F. R. Jones, L. H. Kauffman, S. Lambropoulou, and J. H. Przytycki, editors, *Proceedings of the International Conference on Knot Theory and its Ramifications held in Delphi, August 7–15, 1998*, volume 24 of *Ser. Knots Everything*, pages 306–334, Singapore, 2000. World Sci. Publishing.

[12] Patrick Plunkett, Michael Piatek, Akos Dobay, John C. Kern, Kenneth C. Millett, Andrzej Stasiak, and Eric J. Rawdon. Total curvature and total torsion of knotted polymers. *Macromolecules*, 40(10):3860–3867, 2007.

[13] R. Varela, K. Hinson, J. Arsuaga, and Y. Diao. A fast ergodic algorithm for generating ensembles of equilateral random polygons. *J. Phys. A*, 42(9):1–13, 2009.

[14] J. Cantarella, T. Deguchi, and C. Shonkwiler. Probability theory of random polygons from the quaternionic viewpoint. *Communications on Pure and Applied Mathematics*, to appear.

[15] Yuanan Diao. The knotting of equilateral polygons in $\mathbf{R}^3$. *J. Knot Theory Ramifications*, 4(2):189–196, 1995.

[16] Kenneth C. Millett. Physical knot theory: an introduction to the study of the influence of knotting on the spatial characteristics of polymers. In *Ser. Knots Everything*, volume 46, pages 346–378. World Sci. Publ., Hackensack, NJ, 2012.

[17] P. Cromwell. *Knots and Links*. Cambridge University Press, 2004.

[18] Jim Hoste and Morwen Thistlethwaite. Knotscape. http://www.math.utk.edu/ morwen/knotscape.html. Program for computing topological information about knots.

[19] M. Thistlethwaite. The unraveler. code provided through private communication.

[20] Bruce Ewing and Kenneth C. Millett. Computational algorithms and the complexity of link polynomials. In *Progress in knot theory and related topics*, pages 51–68. Hermann, Paris, 1997.

[21] Tetsuo Deguchi and Kyoichi Tsurusaki. A statistical study of random knotting using the Vassiliev invariants. *J. Knot Theory Ramifications*, 3(3):321–353, 1994. Random knotting and linking (Vancouver, BC, 1993).

[22] Akos Dobay, Pierre-Edouard Sottas, Jacques Dubochet, and Andrzej Stasiak. Predicting optimal lengths of random knots. *Lett. Math. Phys.*, 55(3):239–247, 2001. Topological and geometrical methods (Dijon, 2000).

[23] S. V. Jablan, L. M. Radovi, and R. Sazdanovi. Basic polyhedra in knot theory. *Kragujevac J. Math.*, 28:155–164, 2005.

[24] M. Thistlethwaite. On the structure and scarcity of alternating links and tangles. *Journal of Knot Theory and Its Ramifications*, 07(07):981–1004, 1998.