LATTICE MODELS OF THETA-SHAPED POLYMERS

Nicholas Beaton, James Gleeson and Aleks Owczarek

GuttmannFest 2025 Melbourne, VIC

SUMMARY

- Introduction
- Motivation
- Monte Carlo methods
- Results: 2D
- Results: 3D
- Future work

TONY



INTRODUCTION

SELF-AVOIDING WALKS

A self-avoiding walk (SAW) ω is a sequence of lattice vertices $(\omega_0, \omega_1, \dots, \omega_n)$ such that $\omega_i \neq \omega_j$ for $i \neq j$, and ω_i, ω_{i+1} are adjacent on the lattice for each $i = 0, \dots, n-1$. We say ω has size $|\omega| = n$.



Since the lattice is vertex-transitive, if one SAW is a translation of another we consider them to be the same. Equivalently assume $\omega_0 = 0$.

For a given lattice take c_n to be the number of SAWs of size n. Then e.g. on the square lattice

$$(c_n)_{n\geq 0} = (1,4,12,36,100,284,780,\ldots)$$

SELF-AVOIDING POLYGONS

A self-avoiding polygon (SAP) π is a sequence of lattice vertices $(\pi_1, \pi_1, \ldots, \pi_n)$ such that $\pi_i \neq \pi_j$ for $i \neq j$, and π_i, π_{i+1} are adjacent on the lattice, and so too are π_1, π_n . The size of π is n.

Equivalently a SAP is a SAW which ends adjacent to its initial vertex.



For our purposes two SAPs are the same if they are translates of one another. So there are 2n SAWs (of length n-1) corresponding to each SAP of size n. Let p_n be the number of size-n SAPs. Then e.g. on the square lattice

$$(p_n)_{n\geq 4}=(1,0,2,0,7,0,28,0,124,\ldots)$$

(On the triangular and fcc lattices SAPs can have odd size.)

THETAS AND TADPOLES

SAPS have all vertices of degree 2 while SAWs have two vertices of degree 1. Here we introduce vertices of degree 3: thetas (two vertices of degree 3) and tadpoles (one degree 1 and one degree 3).



Again counted up to translation. On the square lattice

$$egin{aligned} & (heta_n)_{n\geq 7} = (2,0,12,6,62,60,338,430,1966,\ldots) \ & (arphi_n)_{n\geq 5} = (8,24,84,244,740,2072,6004,\ldots) \end{aligned}$$

LATTICES













On the square and cubic lattices, the three "arms" of a theta must have the same parity \rightarrow strong parity effects. Not a problem on the triangular or fcc lattices.

MOTIVATION

POLYMER MODELS

Conception for SAWs is usually credited to Paul Flory (*J. Chem Phys.* 1949). He was interested in a theoretical model for long linear polymer chains in solution. In particular he wanted a model which accounted for the excluded volume principle: the idea that two monomers in a chain cannot occupy the same space.



poly(2-vinylpyridine) (Y. Roiter & S. Minko, J. Amer. Chem. Soc. 2005)



square lattice SAW of length 2^{25} (N. Clisby 201?)

SAWs (or SAPs, trees, etc., depending on the model) turn out to be a good model for polymers in a good solvent (at least for long ones).

KNOTS

Headphone cables, shoelaces, necklaces, and even polymer molecules can get tangled up in knots and links.

H.L. Frisch and E. Wasserman (*J. Amer. Chem. Soc.* 1961) and M. Delbruck (*Proc. Symp. Appl. Math.* 1962) conjectured that random simple closed curves in \mathbb{R}^3 will be knotted with high probability. More precisely, the probability of being unknotted decays exponentially with the length of the curve.





(S.A. Wasserman et al, Science 1985)

The FWD conjecture was first proved by D.W. Sumners and S.G. Whittington (*J. Phys. A: Math. Gen.* 1988) for SAPs on the cubic lattice. Their method used pattern theorems.

Subsequently proved for a number of other models (on and off the lattice).

THE SIZE OF KNOTS

Knots are very likely because small "knotted parts" can be inserted into a polygon and force it to be knotted. But in turn the "knotted part" of a typical large knotted polygon is relatively small.



Different ways to estimate the "size" of the knotted part. Marcone et al (*J. Phys. A* 2005) use two different methods:



Both methods lead to an estimate of the knotted part of an n-edge prime knot occupying $pprox n^{0.75}$ edges on average.

KNOTS VS. THETAS

Question: How do the different parts of a typical theta scale in size? Most likely two arms are small and one big:



R-LOOPS

During DNA transcription, single-stranded RNA can fuse with double-stranded RNA to form R-loops:



Jonoska et al, Using Mathematics to Understand Biological Complexity 2021



Carrasco-Salas et al, Nucleic Acids Res. 2019

The presence of R-loops can affect repair, eg. in human cells.

R-LOOPS

Topologically the DNA:RNA hybrid is a graph with two vertices of degree 3 or 4:



Different parts are single- or double-stranded so may have different stiffness etc.

KNOTTED THETA-GRAPHS

Thetas can also have non-trivial topology -- open questions about minimum number of edges required, etc.



Buck & O'Donnol 2018

MONTE CARLO METHODS

BFACF ALGORITHM

The BFACF algorithm (Berg & Foester 1981, Aragão de Carvalho, Caracciolo & Fröhlich 1983 and 1983) is a method for sampling SAWs and SAPs which preserves topology.



For 2D SAPs it is ergodic while in 3D the ergodicity classes are the knot types. For SAWs the ergodicity classes are walks between two fixed endpoints.

Also works on the triangular lattice.



BFACF ALGORITHM

For the fcc and bcc lattices, Janse van Rensburg and Rechnitzer (2011) found a set of moves which also give the knot types as the ergodicity classes.





BFACF FOR BRANCHED STRUCTURES

The above BFACF moves can shift vertices of degree 2 and leave vertices of degree 1 fixed.

To accommodate vertices of degree 3, Sokai (2018) found a new set of moves on the square/cubic lattices which can realise any ambient isotopic transformation:



Note that the degree 3 vertices can move around.

BFACF FOR BRANCHED STRUCTURES

Simpler on the triangular lattice:



Same moves on the fcc lattice since the faces are still triangles.

WANG-LANDAU METHOD

We combine the BFACF moves with the Wang-Landau method (2001) to estimate the number a_n of configurations of size n.

Start with f = 1 and arrays G and H both all 0.

1. For a given configuration γ , pick a random edge and a random adjacent plaquette. If there is a valid BFACF move (changing γ to γ'), accept with Metropolis probability

$$\mathbb{P}(\gamma o \gamma') = \min\left\{1, \exp(G(|\gamma|) - G(|\gamma'|)) rac{w(\gamma' o \gamma)}{w(\gamma o \gamma')}
ight\},$$

else stay in state γ . Here $w(a \rightarrow b)$ is the probability that a randomly chosen move on state a will change it to state b.

2. Now in state γ'' (either γ or γ'), increment $G(|\gamma''|)$ by f and $H(|\gamma''|)$ by 1.

- 3. If H is sufficiently flat, decrease f (e.g. $f \mapsto f/2$), reset H = 0 and start again. Otherwise continue with the current value of f.
- 4. When f reaches some small threshold, terminate.

If m is the size of the smallest configuration, set g=G(m) and then for each n set

$$A(n) = \exp(G(n) - g) \cdot a_m.$$

Then $A(n) \approx a_n$.

WANG-LANDAU METHOD

Easily generalises to measure more than just size -- e.g. we will estimate $\theta_{n,k}$, the number of thetas of size n with shortest arm of size k.

Many other variations -- e.g. change how f decreases, or tune G and then use those values to estimate some other quantity.

RESULTS: 2D

TEST CASE: THE NUMBER OF POLYGONS

The number p_n of SAPs on the lattice is $\sim {
m const.} n^g \mu^n$ where (Guttmann 2009)

$$g = egin{cases} -5/2 & d=2\ -2.76279 & d=3 \end{cases}$$

and μ is the connective constant of the lattice (Jensen 2004, Clisby 2022):

$$\mu_{
m tri} = 4.150797226(26) \ \mu_{
m fcc} = 10.03705785(14)$$

TEST CASE: THE NUMBER OF POLYGONS

Estimating the exponent g for the triangular lattice (10 independent runs, N=500):



Expect $heta_n \sim {
m const.} n^h \mu^n$ for the same μ .

Estimating the exponent h for the triangular lattice (10 independent runs, N=500):

 $hpprox g+1 \leftrightarrow O(n)$ places to position a small loop within a big polygon.

Expect $\theta_n \sim \text{const.} n^h \mu^n$ for the same μ .

Estimating the exponent h for the triangular lattice (10 independent runs, N = 500):



WL flattening over length and shortest arm:

 $h \approx g + 1 \iff O(n)$ places to position a small loop within a big polygon.

Expect $\theta_n \sim \text{const.} n^h \mu^n$ for the same μ .

Estimating the exponent h for the triangular lattice (10 independent runs, N = 500):



WL flattening over length and middle arm:

 $h \approx g + 1 \iff O(n)$ places to position a small loop within a big polygon.

Estimating the exponent h for the square lattice (10 independent runs, N=500):

 $hpprox g+1 \leftrightarrow O(n)$ places to position a small loop within a big polygon.

Estimating the exponent h for the square lattice (10 independent runs, N = 500):

WL flattening over length and shortest arm:



 $h \approx g + 1 \iff O(n)$ places to position a small loop within a big polygon.

Estimating the exponent h for the square lattice (10 independent runs, N = 500):

WL flattening over length and middle arm:



 $h \approx g + 1 \iff O(n)$ places to position a small loop within a big polygon.

MEAN LENGTH OF SHORTEST THETA ARM

If $\sigma(\theta)$ is the length of the shortest arm then expect (?) $\langle \sigma \rangle_n \sim \text{const.} n^s$ for some s. Estimating s for the triangular lattice (10 independent runs, N = 500):



MEAN LENGTH OF SHORTEST THETA ARM

Estimating s for the square lattice (10 independent runs, N = 500):



MEAN LENGTH OF SECOND-SHORTEST THETA ARM

Can also look at $\tau(\theta)$, the second-shortest (second-longest) arm. Triangular lattice (10 independent runs, N = 500):



Likely the same exponent as for shortest arm.

MEAN LENGTH OF SECOND-SHORTEST THETA ARM

Square lattice (10 independent runs, N = 500):



Likely the same exponent as for shortest arm.

RATIO OF SECOND-SHORTEST TO SHORTEST ARM

Can also look at the ratio of the two. Expect it to approach a (lattice-independent?) constant. Triangular lattice (10 independent runs, N = 500):



RATIO OF SECOND-SHORTEST TO SHORTEST ARM

Square lattice (10 independent runs, N=500):



DISTRIBUTION OF σ

Distribution of the length of the shortest arm for triangular lattice (mean value indicated in green)

Distribution of the length of the shortest arm for triangular lattice (mean value indicated in green)



n = 100:

Distribution of the length of the shortest arm for triangular lattice (mean value indicated in green)

 $\begin{array}{c}
10^{119} \\
10^{118} \\
10^{117} \\
10^{116} \\
10^{115} \\
10^{114} \\
10^{113} \\
10 20 30 40 50 60 \\
\end{array}$

n = 200:

Distribution of the length of the shortest arm for triangular lattice (mean value indicated in green)



n = 300:

Distribution of the length of the shortest arm for triangular lattice (mean value indicated in green)

 10^{242} 10^{240} 10^{238} 10^{236} 20 40 60 80 100 120

n = 400:

Distribution of the length of the shortest arm for triangular lattice (mean value indicated in green)



n = 500:

RESULTS: 3D

Estimating the exponent h for the fcc lattice (10 independent runs, N=400):

Recall that gpprox -2.76279
ightarrow again hpprox g+1.

Estimating the exponent *h* for the fcc lattice (10 independent runs, N = 400):

WL flattening over length and shortest arm:



Recall that $g \approx -2.76279 \rightarrow \text{again } h \approx g + 1$.

Estimating the exponent *h* for the fcc lattice (10 independent runs, N = 400):

WL flattening over length and middle arm:



Recall that $g \approx -2.76279 \rightarrow \text{again } h \approx g + 1$.

Estimating the exponent h for the cubic lattice (10 independent runs, N=400):

Again $h \approx g+1$.

Estimating the exponent h for the cubic lattice (10 independent runs, N = 400):

WL flattening over length and shortest arm:

Again $h \approx g + 1$.

Estimating the exponent h for the cubic lattice (10 independent runs, N = 400):

WL flattening over length and middle arm:

Again $h \approx g + 1$.

MEAN LENGTH OF SHORTEST THETA ARM

If $\sigma(\theta)$ is the length of the shortest arm then expect (?) $\langle \sigma \rangle_n \sim \text{const.} n^s$ for some s. Estimating s for the FCC lattice (10 independent runs, N = 400):



MEAN LENGTH OF SHORTEST THETA ARM

Estimating s for the cubic lattice (10 independent runs, N = 400):



MEAN LENGTH OF SECOND-SHORTEST THETA ARM

Can also look at $\tau(\theta)$, the second-shortest (second-longest) arm. FCC lattice (10 independent runs, N = 400):



Likely the same exponent as for shortest arm.

MEAN LENGTH OF SECOND-SHORTEST THETA ARM

Cubic lattice (10 independent runs, N=400):



Likely the same exponent as for shortest arm.

RATIO OF SECOND-SHORTEST TO SHORTEST ARM

Can also look at the ratio of the two. Expect it to approach a (lattice-independent?) constant. FCC lattice (10 independent runs, N = 400):



RATIO OF SECOND-SHORTEST TO SHORTEST ARM

Cubic lattice (10 independent runs, N=400):



DISTRIBUTION OF σ

Distribution of the length of the shortest arm for FCC lattice (mean value indicated in green)

This contrasts with observed knot size distributions (linear chains) which seem to have a peak at fixed length and then power-law decay (Tubiana et al 2013).

Distribution of the length of the shortest arm for FCC lattice (mean value indicated in green)



This contrasts with observed knot size distributions (linear chains) which seem to have a peak at fixed length and then power-law decay (Tubiana et al 2013).

Distribution of the length of the shortest arm for FCC lattice (mean value indicated in green)



This contrasts with observed knot size distributions (linear chains) which seem to have a peak at fixed length and then power-law decay (Tubiana et al 2013).

Distribution of the length of the shortest arm for FCC lattice (mean value indicated in green)



This contrasts with observed knot size distributions (linear chains) which seem to have a peak at fixed length and then power-law decay (Tubiana et al 2013).

n = 300:

Distribution of the length of the shortest arm for FCC lattice (mean value indicated in green)



This contrasts with observed knot size distributions (linear chains) which seem to have a peak at fixed length and then power-law decay (Tubiana et al 2013).

n = 400:

FUTURE WORK

- Other shapes? (e.g. tadpoles, combs)
- How do knots affect things?
- Minimum sizes of prime theta knots
- Introduce weights to model R-loops

Thanks for listening!