

AMSI
VACATION
RESEARCH
SCHOLARSHIPS

2018-2019



The Transfer Matrix Approach to Polymer-Modelling Dyck Paths

Leo Li

Supervised by Dr Nicholas Beaton and Dr Thomas Wong
The University of Melbourne

Vacation Research Scholarships are funded jointly by the Department of Education and
Training and the Australian Mathematical Sciences Institute.



1 Introduction

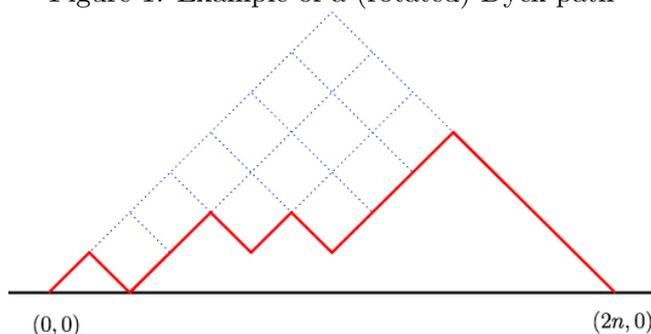
For most chemistry students, polymers are first taught in the later years of highschool as an exercise in the ability for a single unit (or monomer) to combine with itself to form a larger, more complex and more functional product. Examples of this include amino acids joining via peptide linkages to form the proteins of life, alkenes undergoing addition reactions to form plastics such as polyethylene, and substances pivotal in biophysics (such as the non-stick phenomenon observed in polytetrafluoroethylene). Whilst the nomenclature regarding the monomers and method of joining may change, these polymers all share the property that the linkages occur at fixed chemical angles. These, however, vary as well, with polyethylene and other aliphatic polymers exhibit sp^3 or tetrahedral geometry at their backbone carbons (resulting in bond angles of 109.5°) whilst the peptide linkages in proteins result in a trigonal planar geometry with bond angles of 120° . Nevertheless, these geometries are fixed within a given polymer, and thus the overall shape can be modelled as a directed walk beginning at one end and nodally moving towards the other (assuming it doesn't branch). Since there is plenty of literature regarding the behaviour of such paths, this opens the door to a combinatorial insight into the behaviour of polymers in real-world chemical and physical scenarios.

2 Dyck Paths

Definition 2.1. A *Dyck path* is a staircase walk beginning at $(0, 0)$ and ending at (n, n) that lies strictly below (but may touch) $y = x$.

Note that this path necessarily has length $2n$ (n up-steps and n right steps). For our purposes, we rotate this definition by $3\pi/4$ anticlockwise and dilate by a factor of $\sqrt{2}$, resulting in a Dyck path with $2n$ steps from a *Cardinal step set* of $(1, 1)$ and $(1, -1)$. For convenience, we label the start point $(0, 0)$ and end point $(2n, 0)$, as shown in the example below.

Figure 1: Example of a (rotated) Dyck path





One should note that the sole restriction on the Dyck path based on our definition is the inability for it to pass below the line $y = 0$. Rather nicely, the number of paths is given by the well-known Catalan numbers

$$C_n = \frac{1}{n+1} \binom{2n}{n}$$

This is only a potential model for a long-chain aliphatic polymer with no side chains, functional groups or interactions between consecutive carbons or the environment, rendering it incredibly restrictive and basic to the point of being useless. The model explored in this project aims to better approximate real-world scenarios.

3 Transfer Matrix Method

Definition 3.1. A sequence a_n is said to be of exponential order K^n if and only if $\limsup |a_n|^{1/n} = K$. The value K is referred to as the growth rate of the sequence.

It has been found that deducing an expression for the number of directed paths given interaction conditions is often impossible, so we instead turn our attention to finding an expression for the growth rate of the sequence. We first define a transfer matrix as follows:

Definition 3.2. Given a directed graph $G = (V, E)$, we attribute each edge a *weight* by using a weight function $w : E \rightarrow \mathbb{N}$, and a size given by size function $\sigma : E \rightarrow \mathbb{N}$. We then associate a transfer matrix $T[i, j]$ entry-wise by

$$T(z)[i, j] = \sum_{e \in E} w(e) z^{\sigma(e)}$$

Note that in the models below all steps (the 'edges') are of size 1, and hence the transfer matrix simplifies to $T(z) = zA$ for some adjacency matrix A with non-negative entries.

Previous work by Flajolet[1] and Wong[2] illustrates how to now proceed with the *Transfer Matrix Method*.

Theorem 3.1 (Transfer Theorem). *Let $T(z)$ be an irreducible transfer matrix. Then all entries*

$$F^{<i,j>}(z) = ((I - T(z))^{-1})[i, j]$$

have the same radius of convergence in z , R_z , which can be defined as $R_z = \lambda^{-1}$ where λ is the largest (or dominant) positive eigenvalue of $T(z)$. Moreover, for any i, j in an irreducible and aperiodic transfer matrix, the coefficient of z^n in $F^{<i,j>}(z)$ is of the same exponential order as λ .

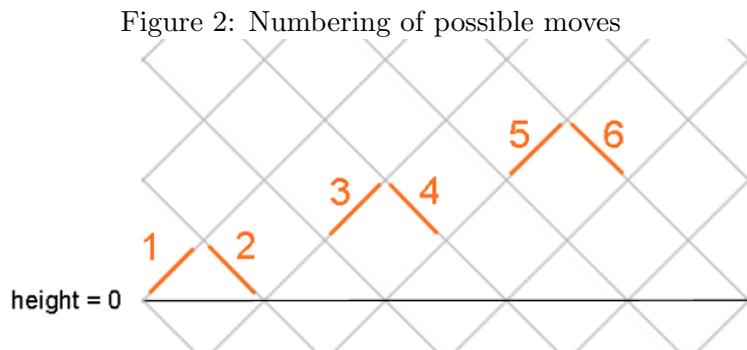


Theorem 3.2 (Perron-Frobenius Theorem). *Let A be an irreducible matrix with non-negative entries. The eigenvalues of A can be ordered in such a way that there is a largest real eigenvalue (though not necessarily unique in having the largest modulus). The number of eigenvalues that share this modulus is the period of the associated graph, and hence when there is an aperiodic graph, the dominant eigenvalue is unique.*

Corollary 3.2.1. *The dominant eigenvalue of a matrix zA , where A is an adjacency matrix with non-negative entries, has corresponding (left and right) eigenvectors with strictly positive entries.*

The Perron-Frobenius Theorem and Transfer Theorem together reduce the task of deducing the growth rate of our system of directed walks to deducing the dominant eigenvalue of a transfer matrix associated with said system.

Consider the directed walk as a dot moving rightward about nodes on a diagonal lattice similar to that pictured in Figure 1. We denote moves in the following way: the 'up-step' from height n to height $n + 1$ is denoted move $2n + 1$, whilst the 'down-step' from height $n + 1$ to height n is denoted move $2n + 2$. This results in the move numbering pictured below.



We then construct a transfer matrix setting the entry $T[i, j] = w$ if a move from move i to move j is possible and has weight w , and setting $T[i, j] = 0$ if such a move is impossible. The weights depend on interactions within the model being analysed. For example, for a slit of width 4 with no interactions, the transfer matrix, T , is:



$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

4 Free Energy and Force

Working within a physical context, a common situation is that of a polymer confined between two walls, and consequently a force of interest is that exerted by the walls upon the polymer due to interactions within this system. To understand this force we must first understand the free energy of the system, $\kappa(u)$, which is linked to the dominant singularity and eigenvalue.

Theorem 4.1 (Pringsheim's Theorem). *If $f(z)$ is representable at the origin by a series expansion with non-negative coefficients and the radius of convergence is R_z , then $f(z)$ has a singularity at $z_c = R_z$, known as the dominant singularity.*

We recall from the Transfer Theorem that $\lambda^{-1} = R_z$.

Definition 4.1. In a physical context, the (*thermodynamic*) free energy of a system, κ , is an energy-like property or state function that can be used to predict the change in a system and its ability to do work. Loosely speaking, it is the energy in a system free to do work. Mathematically speaking, it can be found by the relation $\kappa = -\log(z_c)$.

We therefore have $\kappa = -\log(z_c) = -\log(R_z) = -\log(\frac{1}{\lambda}) = \log(\lambda)$

Definition 4.2. The effective force between the walls of our system is given by a function of the width $\mathcal{F}(w)$:

$$\mathcal{F}(w) = \frac{\partial \kappa(w)}{\partial w}$$

In the context of polymers, a $\mathcal{F} > 0$ indicates the walk exerting an expansive force upon the walls of its containers, whilst $\mathcal{F} < 0$ indicates a contractive force. Combining these definitions with



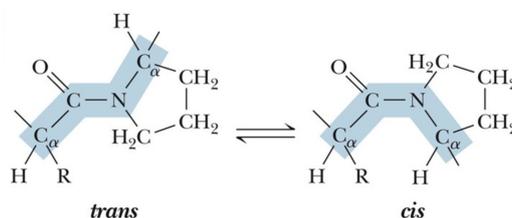
above theorems (as well as the chain rule) gives us an expression to determine the force from transfer matrices alone:

$$\mathcal{F}(w) = \frac{\partial \kappa(w)}{\partial w} = \frac{\partial \kappa(w)}{\partial \lambda} \frac{\partial \lambda(w)}{\partial w} = \frac{\partial}{\partial \lambda} (\log(\lambda)) \frac{\partial \lambda(w)}{\partial w} = \frac{1}{\lambda} \frac{\partial \lambda(w)}{\partial w}$$

5 Model

The model we choose to analyse is a path with weights allocated for consecutive steps in the same direction/orientation confined within a slit with interactive upper and lower walls. This is a chemically relevant scenario since many polymers with planar linkages between monomers favour a *trans* isomerisation over *cis* (due to electrostatic instability in the latter). A key example of this are the 20 naturally occurring amino acids which all almost **exclusively** exist as *trans* residues (with the exception of proline which favours *trans* only 4:1). Moreover, since proteins more often than not exist within natural confines such as a nucleus or cell wall, where interactions such as adsorption can occur, the entrapment within a slit is a logical inclusion.

Figure 3: *cis* and *trans* isomerisations of the peptide linkage between consecutive amino acids within a protein



More generally, when a linear, non-branching polymer is confined between two parallel plates, it has been observed that it loses configurational entropy and exerts a repulsive force upon the walls. However, due to aforementioned interactions such as adsorption, electrostatic attraction and Van Der Waal's forces, this repulsion can be lessened or, depending on the degree of interaction, even have a net attractive force upon the confining walls. This is exactly the force we described above as a derivative of free energy with respect to width.[3]

For convention, from a current position in the walk (ie. in a given row of the transfer matrix), if the next step will be in the same direction as the previous one, we multiply that entry by weight c . In some literature, this quantity is known as the *stiffness* of the molecule. Furthermore, if the next move is into a wall (of which there are two: move 2 into the bottom wall and move $2n - 1$ into the top wall), we multiply that entry by weight a . A move that is both consecutive directionally and into



a wall (eg. from move 4 to move 2) will have weight ac .

It is not difficult to notice a pattern in these matrices. For example, they bear a sort of "rotational" symmetry in that $T[i, j] = T[2n + 1 - i, 2n + 1 - j]$. Using Python, it is then an easy task to construct a transfer matrix from given a and c weights.

5.1 Finding the Dominant Eigenvalue

Given that a directed walk of length n results in a square transfer matrix of size $2n$ and proteins can be well into hundreds of amino acid residues in length, it would be absurd to attempt to solve for the dominant eigenvalue by a characteristic polynomial. Instead we opt for the "Power Iteration" method below, a simple algorithm which gives the eigenvalue of largest absolute value of matrix \mathbf{A} and its corresponding eigenvector:

1. Calculate $\mathbf{A}v$ for some non-eigenvector v_0
2. Let the largest entry in v_0 be m . Set $\lambda = m$. Let $v_1 = v_0/m$
3. Repeat steps 1 and 2 indefinitely for successive v_i . λ converges on the dominant eigenvalue whilst v_i converges on the corresponding eigenvector

Since we know that row 2 of \mathbf{A} is a 1 followed by zeroes, whilst there will certainly also be row with a sum of entries greater than 1, it is clear that $v_0 = [1, 1, 1, \dots]$ will suffice as the initial non-eigenvector. This Power Iteration method is efficient, though one must define the level of tolerance between successive λ to indicate when the loop should be terminated. It also faces the issue of not necessarily being a convergent process (step 3) if \mathbf{A} is a periodic matrix. This, however, is easily fixed by adding the identity, I , to \mathbf{A} , then subtracting 1 from the subsequent eigenvector since λ is an eigenvector $\iff \mathbf{A}v = \lambda v \iff (\mathbf{A} + I)v = (\lambda + 1)v$. Moreover, v_i will still converge to the eigenvector of λ . The precision of our estimation can be rigorously bounded based on the following theorem by Lothar Collatz:

Theorem 5.1. *Given a (finite) adjacency vector \mathbf{A} and a positive vector \mathbf{x} of corresponding size, we define $v_i(\mathbf{x}) = \frac{(\mathbf{A}\mathbf{x})_i}{x_i}$ and further define $m(\mathbf{x}) = \min_i v_i(\mathbf{x})$ and $M(\mathbf{x}) = \max_i v_i(\mathbf{x})$. Then, the dominant eigenvalue of \mathbf{A} , $\lambda(\mathbf{A})$, is bounded:*

$$m(\mathbf{x}) \leq \lambda(\mathbf{A}) \leq M(\mathbf{x})$$

with both equalities holding when \mathbf{x} is the positive dominant eigenvector of \mathbf{A} .



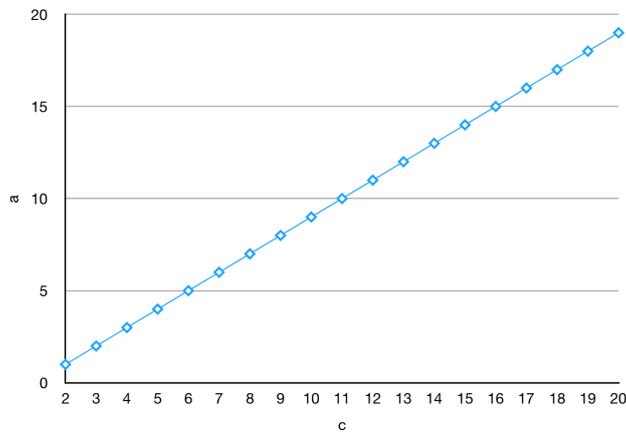
5.2 Zero Force Curve

Recall that $\mathcal{F}(w) = \frac{1}{\lambda} \frac{\partial \lambda(w)}{\partial w}$. Since the requirement of knowing λ as a function of w is a difficult task, the practical approach is to use a discrete approximation of the partial derivative:

$$\mathcal{F}(w) = \frac{1}{\lambda(w)} \frac{\lambda(w+2) - \lambda(w)}{2}$$

Since $\lambda \neq 0$, this approximation of force is only equal to zero when $\lambda(w + 2) = \lambda(w)$. This brings about the idea of the "zero force curve", a graph (known to be continuous) showing values of a and c that produce the zero-force result. We consider the curve as a function of a and c , with a on the x-axis and c on the y-axis. Using Python 3.7, by programming the construction of relevant transfer matrices, iteratively fixing values of a and searching (through a bisection method) for the "zero force" value of c , we are able to find data that suggests that the equation for the zero force curve in this model approximates $c = a - 1$ at all widths w . This result is supported by analysis of the system using generating functions[6].

Figure 4: The Zero Force Curve for a model with weight a corresponding to a wall interaction and weight c corresponding to a consecutive step



6 Sampling

Aside from assessing the growth rate and zero force curve, we also seek to find a method of generating walks probabilistically based on weights a and c . To do so, we use an algorithm of constructing weighted walks from Beaton et al. [4] in their work on polygons in a lattice. We now label the path $\pi = \pi_1 \pi_2 \pi_3 \dots \pi_n$. We now consider the moves as 'states', numbered the same as above, and let T be the transfer matrix, λ the dominant eigenvalue and ζ the corresponding eigenvector. We let $F(i)$ be the set of states that can follow state i .



We define the following probability functions:

$$\chi(i) = \begin{cases} a & \text{if } i \text{ is a state that touches the wall} \\ 1 & \text{otherwise} \end{cases}$$

$$t_1(i) = \sum_{j \in F(i)} T_{i,j} \zeta_j$$

$$t_s(i) = \frac{\sum_{j \in F(i)} T_{i,j}}{\zeta_i}$$

$$r_1(i) = 1 - \frac{t_1(i)}{\max_j t_1(j)}$$

$$r_s(i) = 1 - \frac{t_s(i)}{\max_j t_s(j)}$$

The algorithm then roughly reads as follows:

1. Choose π_1 with probability $\frac{\chi(\pi_1)}{\sum_j \chi(j)}$
2. With probability $r_1(\pi_1)$, reject this first step and return to (1). Else, proceed.
3. Choose π_2 with probability $\frac{T_{i,j} \zeta_{\pi_2}}{t_1 \pi_1}$
4. With probability $r_1(\pi_1)$, reject this first step and return to (1). Else, move to next step.
5. For $k = 3, 4, 5, \dots, n - 1$, choose π_k with probability $\frac{\pi_{k-1} \pi_k}{\lambda} \frac{\zeta_{\pi_k}}{\zeta_{\pi_{k-1}}}$.
6. Choose final step, π_n with probability $\frac{T_{\pi_{n-1}, \pi_n}}{\sum_{j \in F(\pi_{n-1})} T_{\pi_{n-1}, j}}$

We use this to produce a sequence of digits based on the probability distribution of each move that encodes the path taken, and translate this to a rudimentary ASCII path. As expected, when a high a value is used, the path tends to cling to the wall, much like the adsorption of a long-chain polymer. When a high c value is used, the path seems to 'bounce' from wall to wall in straight-line motion.

Figure 5: A probabilistically generated path with $w = 12, n = 50, a = 10, c = 1$

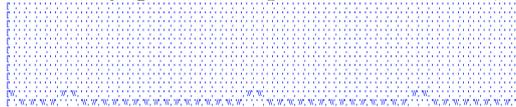


Figure 6: A probabilistically generated path with $w = 12, n = 50, a = 1, c = 10$





Furthermore, we then define a function (say $genpath()$) that takes parameters w , a , c , and l (the length of the path), and generates a walk. Iterating over this sampling function many times allows us to see the distribution of paths under these parameters. We expect these distributions to be Boltzmann distributions.

Definition 6.1. A *Boltzmann Distribution* is expressed in the form:

$$p_i \propto \exp\left(\frac{-\epsilon_i}{k_b T}\right)$$

where ϵ_i is the state, k_b is the Boltzmann constant and T is the absolute thermodynamic temperature.

The weight a represents a Boltzmann weight of $\exp[\frac{\epsilon_a}{k_b T}]$, where in statistical mechanics terms $-\epsilon_a$ is the energy attributed to a wall visit. Similarly, the weight c corresponds to a weight $\exp[\frac{\epsilon_c}{k_b T}]$ where $-\epsilon_c$ is the energy associated with a "stiffness point" between consecutive collinear steps. We can therefore expect that in our distributions of paths π ,

$$p_\pi \propto a^{\chi(\pi)} c^{\sigma(\pi)}$$

where $\chi(\pi)$ is the number of stiffness points and $\sigma(\pi)$ is the number of wall contacts. To confirm this, we take samples of $n = 100,000$ for $w = 2$, $l = 5$, and $(a, c) = (1, 1), (1, 2)$ and $(2, 1)$ respectively.

Figure 7: A sample of size $n = 100000$ with $w = 2$, $l = 5$, $(a, c) = (1, 1)$

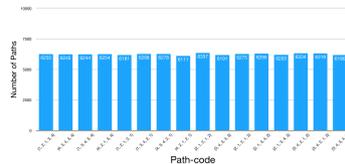


Figure 8: A sample of size $n = 100000$ with $w = 2$, $l = 5$, $(a, c) = (2, 1)$

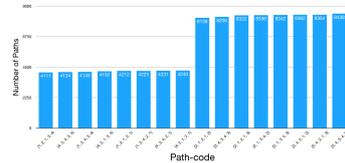


Figure 9: A sample of size $n = 100000$ with $w = 2$, $l = 5$, $(a, c) = (1, 2)$

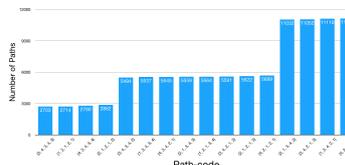
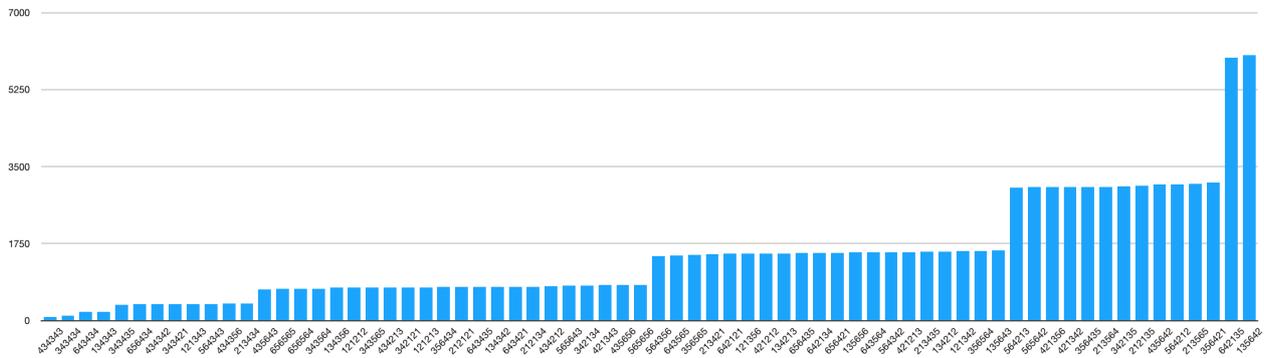




Figure 7 being a uniform distribution confirms that both our sample and model work as intended, since $(a, c) = (1, 1)$ translates to interactions with consecutive steps and the walls are neither preferred nor avoided, and thus all paths should have an equal probability. Figures 8 and 9 confirm that the distribution is indeed Boltzmann since in Figure 8, walks with n wall contacts appear with probability proportional to 2^n (with stiffness points having no bearing on weight), whilst in Figure 9, walks with n stiffness points appear with probability proportional to 2^n independent of wall contacts. Another more complex example of this (resulting in 7 probability classes as predicted by a Boltzmann distribution) is below:

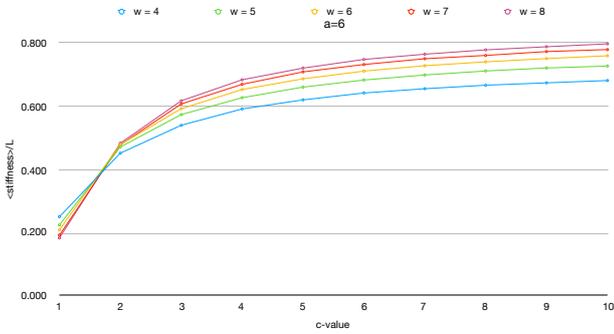
Figure 10: A sample of size $n = 100000$ with $w = 3, l = 6, (a, c) = (2, 2)$



Sampling is of interest because it not only provides a concrete model for the distribution of polymeric walks based on weights, length and width, but also brings the idea of finding the points (a, c) where the system switches from preferring wall contacts to stiffness points. We fix either a or c whilst varying the other interaction parameter, and then measure the mean number of stiffness points divided by the length of the walk (denoted $\frac{\langle stiffness \rangle}{L}$) or the mean number of wall contacts (denoted $\frac{\langle wall \rangle}{L}$), respectively. We do this at various values at multiple slit widths to better understand the change in the system. Below is data for $a = 6, 7, 8, 12, 20$ with $c \in [1, 10]$, as well as data for $c = 3, 6, 9, 21$ with $a \in [1, 10]$. Both sets of data are for widths $w \in [4, 8]$, $L = 100$ and sample size 1000, and do not include error bars since the 95% confidence intervals are all < 0.01 and invisible on our plots.

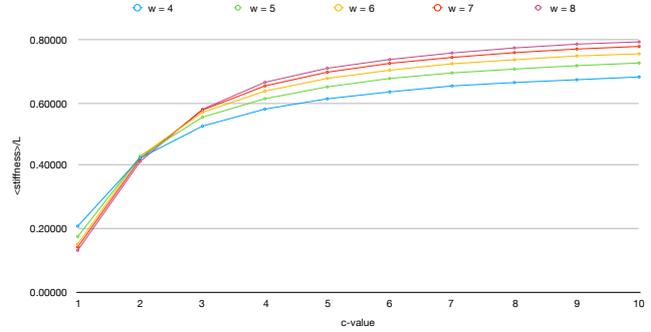
a=6

c	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.251	0.224	0.209	0.192	0.183
2	0.452	0.472	0.479	0.478	0.484
3	0.54	0.574	0.593	0.608	0.618
4	0.592	0.628	0.654	0.67	0.685
5	0.621	0.661	0.688	0.710	0.722
6	0.643	0.684	0.712	0.73	0.749
7	0.66	0.700	0.729	0.751	0.766
8	0.667	0.713	0.741	0.76	0.780
9	0.675	0.722	0.752	0.77	0.789
10	0.682	0.728	0.761	0.780	0.799



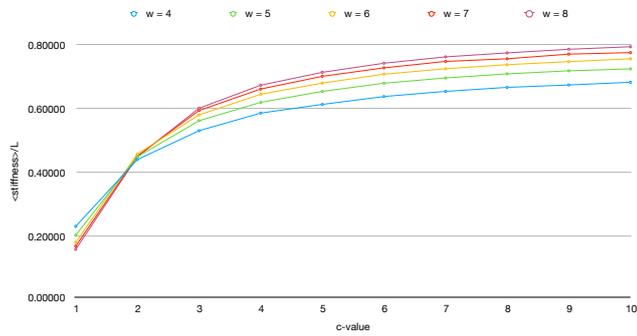
a=8

c	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.20919	0.17615	0.15204	0.14286	0.13253
2	0.42513	0.43013	0.43108	0.42257	0.41418
3	0.52632	0.55474	0.56983	0.57711	0.57971
4	0.58034	0.61319	0.63723	0.65378	0.66533
5	0.61298	0.65088	0.67763	0.69733	0.70991
6	0.63474	0.67743	0.70358	0.72531	0.73758
7	0.65380	0.69501	0.72413	0.74426	0.75827
8	0.66461	0.70735	0.73674	0.75936	0.77421
9	0.67349	0.71807	0.74893	0.77082	0.78631
10	0.68229	0.72646	0.75547	0.77907	0.79351



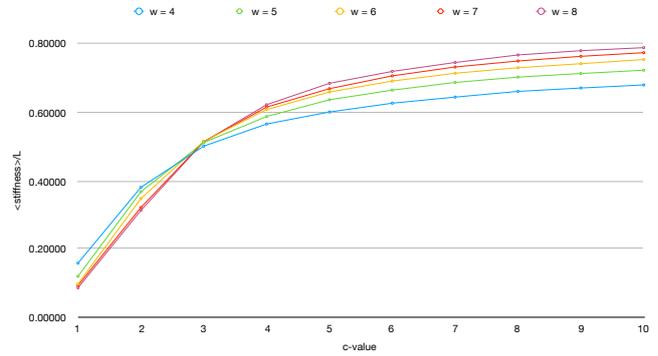
a=7

c	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.22719	0.19917	0.17648	0.16450	0.15431
2	0.43891	0.44762	0.45644	0.45031	0.44770
3	0.53023	0.56199	0.58037	0.59432	0.60151
4	0.58623	0.62020	0.64593	0.66240	0.67451
5	0.61372	0.65495	0.68131	0.70278	0.71543
6	0.63868	0.68090	0.70945	0.72962	0.74423
7	0.65502	0.69755	0.72680	0.74999	0.76442
8	0.66771	0.71061	0.73968	0.75842	0.77703
9	0.67536	0.72029	0.74933	0.77309	0.78848
10	0.68402	0.72633	0.75835	0.77800	0.79652



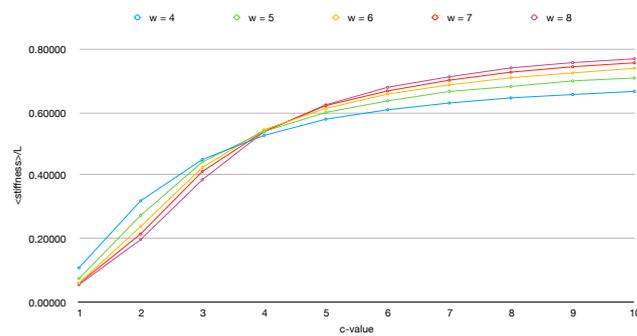
a=12

c	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.15970	0.12131	0.09897	0.09471	0.08787
2	0.38006	0.36706	0.34717	0.32133	0.31306
3	0.49996	0.51047	0.51416	0.51295	0.51245
4	0.56426	0.58803	0.60624	0.61346	0.62036
5	0.59924	0.63485	0.65740	0.66704	0.68268
6	0.62488	0.66285	0.68926	0.70434	0.71731
7	0.64254	0.68521	0.71190	0.73019	0.74326
8	0.65920	0.70072	0.72801	0.74754	0.76514
9	0.66902	0.71121	0.73973	0.76118	0.77761
10	0.67792	0.72079	0.75186	0.77177	0.78637



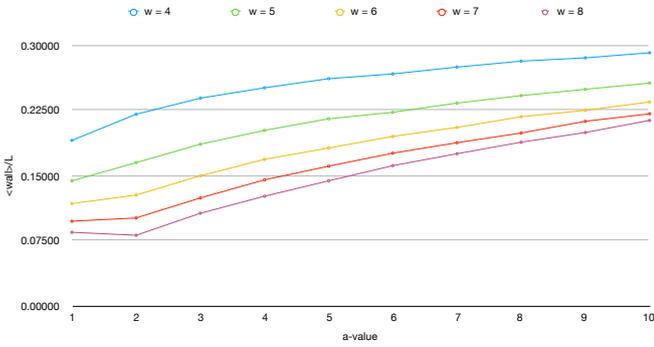
a=20

c	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.10744	0.07367	0.05921	0.05766	0.05402
2	0.32008	0.27375	0.23822	0.21491	0.19703
3	0.45104	0.44329	0.42594	0.41292	0.38692
4	0.52729	0.53986	0.54531	0.54070	0.53884
5	0.57869	0.60004	0.61266	0.62174	0.62414
6	0.60847	0.63660	0.65816	0.66815	0.67960
7	0.62996	0.66648	0.68732	0.70221	0.71308
8	0.64620	0.68236	0.70997	0.72763	0.74117
9	0.65674	0.70005	0.72529	0.74489	0.75790
10	0.66674	0.70913	0.74006	0.75725	0.77011



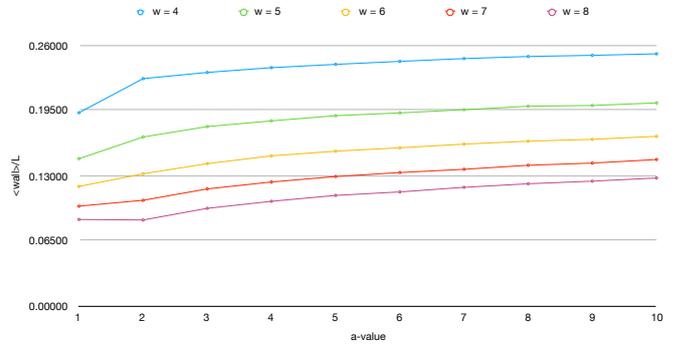
c=3

a	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.19078	0.14428	0.11808	0.09783	0.08506
2	0.22081	0.16520	0.12788	0.10156	0.08167
3	0.23929	0.18635	0.15005	0.12458	0.10696
4	0.25118	0.20227	0.16897	0.14537	0.12655
5	0.26169	0.21557	0.18200	0.16109	0.14424
6	0.26715	0.22309	0.19533	0.17608	0.16171
7	0.27503	0.23366	0.20568	0.18807	0.17541
8	0.28184	0.24222	0.21792	0.19922	0.18868
9	0.28565	0.24949	0.22529	0.21270	0.19983
10	0.29148	0.25664	0.23488	0.22128	0.21372



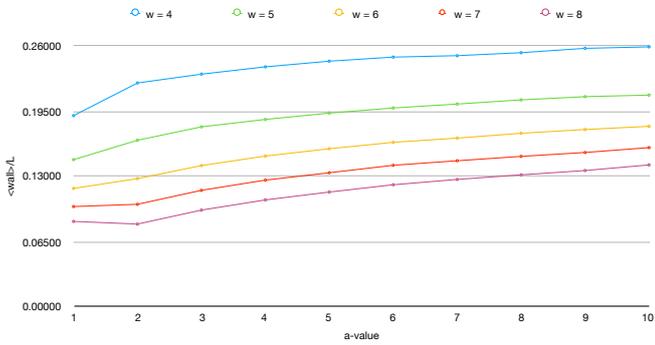
c=9

a	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.19292	0.14707	0.11941	0.09979	0.08650
2	0.22686	0.16871	0.13204	0.10565	0.08601
3	0.23304	0.17908	0.14214	0.11697	0.09762
4	0.23779	0.18477	0.14991	0.12387	0.10462
5	0.24108	0.18993	0.15486	0.12932	0.11050
6	0.24405	0.19270	0.15793	0.13326	0.11397
7	0.24684	0.19591	0.16160	0.13650	0.11856
8	0.24894	0.19937	0.16450	0.14054	0.12209
9	0.25007	0.20012	0.16638	0.14270	0.12470
10	0.25158	0.20265	0.16936	0.14624	0.12782



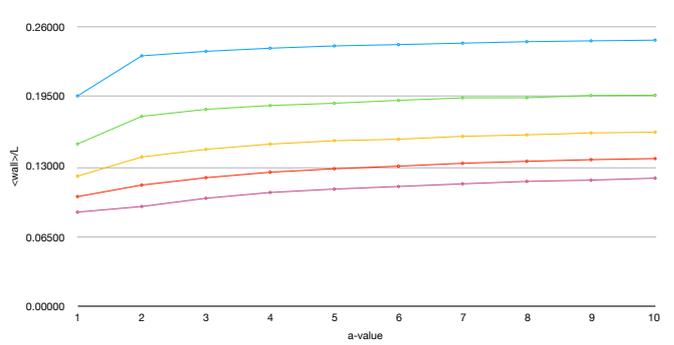
c=6

a	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.19087	0.14699	0.11835	0.10024	0.08546
2	0.22339	0.16634	0.12809	0.10253	0.08282
3	0.23219	0.17967	0.14109	0.11644	0.09675
4	0.23948	0.18710	0.15059	0.12663	0.10688
5	0.24512	0.19336	0.15788	0.13394	0.11473
6	0.24918	0.19849	0.16427	0.14139	0.12198
7	0.25062	0.20241	0.16848	0.14591	0.12730
8	0.25359	0.20655	0.17329	0.15026	0.13187
9	0.25791	0.20978	0.17705	0.15415	0.13616
10	0.25939	0.21134	0.18020	0.15898	0.14175

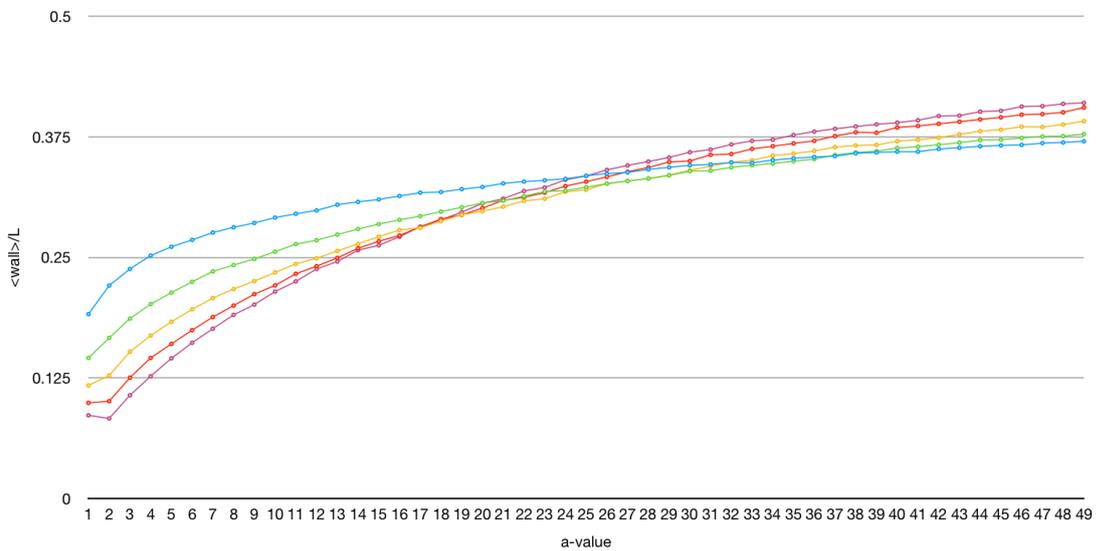


c=21

a	w = 4	w = 5	w = 6	w = 7	w = 8
1	0.19527	0.15086	0.12112	0.10220	0.08784
2	0.23229	0.17636	0.13883	0.11289	0.09307
3	0.23644	0.18286	0.14588	0.11963	0.10067
4	0.23934	0.18838	0.15071	0.12472	0.10600
5	0.24147	0.18847	0.15386	0.12792	0.10915
6	0.24269	0.19114	0.15523	0.13033	0.11154
7	0.24403	0.19343	0.15793	0.13299	0.11399
8	0.24539	0.19355	0.15924	0.13483	0.11626
9	0.24616	0.19566	0.16105	0.13632	0.11741
10	0.24680	0.19593	0.16172	0.13735	0.11925



c = 3 (extended)





6.1 Discussion

For the varying c values, the behaviour of the model is as expected. The interpolated line graphs for distinct w intersect at a certain point (for a low value of a). The larger widths having proportionally less stiffness points since there is a larger "potential space" for entropic movement and no preference for stiffness points, whilst in smaller widths the probabilistic attraction of the closer opposite wall causes consecutive steps. However, when $c \gg 1$, longer consecutive walks are favoured and this potential space now allows for long diagonal walks to occur in wide slits, whilst a thin slit will be impeded by the inevitable fact that the top and bottom walls are frequently hit. Furthermore, the aforementioned intersection shifts to the right (larger values of c) for greater fixed values of a , as a result of the greater "competition" between wall contacts and stiffness points. At one extreme, there is the walk that clings to the wall (eg. $(1, 2, 1, 2, 1, \dots)$) to reap the benefits of adsorption, whilst at the other we have the walk "bouncing" between the walls in straight diagonal lines (eg. $(1, 3, 5, \dots, 2w - 1, 2w, 2w - 3, 2w - 5, \dots)$) that exploits the c weight. We note that it appears only a small increase in c is required relative to a to both dramatically increase the proportion of stiffness points and decrease the proportion of wall contacts (the data for this is not shown in this report), suggesting that the c weight has a greater sway on the distribution of walks than a . A possible line of reasoning for this is that the number of state changes which trigger these weights. For width w , there are a total of $2(w - 1)$ state changes that have weight c (namely $1 \rightarrow 3, 3 \rightarrow 5, \dots, 2w - 3 \rightarrow 2w - 1$ and $4 \rightarrow 2, 6 \rightarrow 4, \dots, 2w \rightarrow 2w - 2$). For width a , there are only 4: $1 \rightarrow 2, 4 \rightarrow 2, 2w \rightarrow 2w - 1$ and $2w - 3 \rightarrow 2w - 1$. A final point to make regarding this set of data is the (slow) convergence of $\frac{\langle stiffness \rangle}{L}$ to $\frac{(w-1)}{w}$ and of $\frac{\langle wall \rangle}{L}$ to $\frac{1}{w} - \frac{1}{L}$, based upon the maximal walk for $\frac{\langle stiffness \rangle}{L}$ where the polymer bounces wall to wall. It should be noted that that $\frac{\langle wall \rangle}{L}$ appears to approach its proposed limit faster than $\frac{\langle stiffness \rangle}{L}$, since there is only one maximal walk satisfying the limit case, whereas there are multiple walks with $\frac{\langle wall \rangle}{L} = \frac{1}{w} - \frac{1}{L}$.

For varying a values with a fixed c , it appears that the lines of different widths very slowly approach each other, which we would expect given that we observed the effect of increasing c to be dramatic relative to a . Increasing the range of a to $[1, 49]$ shows that an intersection between the widths occurs in the region $[18, 32]$ (Figure 11). Again, we have the case where greater widths begin lower then rise higher. This is because for low values of a , there is no stickiness of walls and intuitively a thinner wall results in more collisions. However, for large values of a , we need to consider the entropic chance that the walk leaves the wall, which is equal for all widths given fixed (a, c) . For a thinner width, since the opposite wall is closer, there is a greater probability of it making the journey across compared to a wider slit. When sampling walks visually, it was confirmed that wider slits result in more walks that



cling to a single wall, whereas widths < 5 often result in the walk traversing the slit then clinging to the other wall, wasting steps that could be registering χ contacts. The maximal case for any walk of length 100 is 50 wall contacts, and we propose that $\frac{\langle wall \rangle}{L} \rightarrow 0.5$ as a becomes large. This, as with the limits proposed in the previous paragraphs, can be justified by the Boltzmann distribution of our sample. Recall that $p_\pi \propto a^{\chi(\pi)} c^{\sigma(\pi)}$ for paths π . We define $Z = \sum_\pi p_\pi$, and so $p_\pi = \frac{a^{\chi(\pi)} c^{\sigma(\pi)}}{Z}$. It is clear that for paths π_k , $a^{\chi(\pi_k)} c^{\sigma(\pi_k)} \sim a^{\chi(\pi_k)}$ and $Z \sim a^{\chi(\pi_{max})}$. Since in submaximal cases, $\chi(\pi_k) < \chi(\pi_{max})$, it follows that $p_{\pi_k} = \frac{a^{\chi(\pi_k)} c^{\sigma(\pi_k)}}{Z} \rightarrow 0$. Therefore, the only paths with $p > 0$ as $a \rightarrow \infty$ are those with the maximal number of contacts, and so $\frac{\langle wall \rangle}{L} \rightarrow 0.5$. Similar reasoning can be used to justify the approach $\frac{\langle stiffness \rangle}{L}$ to $\frac{(w-1)}{w}$ in the above paragraph.

6.2 Clingy/Stiff Transition Point

As mentioned earlier, some points of interest in our model are the (a, c) values for which the model appears to transition from walks of long, collinear steps (stiff) and those that move along the wall (clingy). In reality, this transition is likely gradual over numerous a values for a given c , but for ease of calculation we impose a clear-cut distinction.

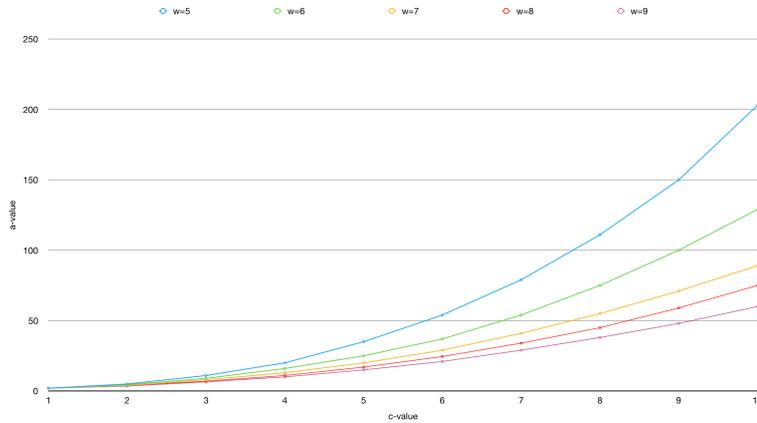
Definition 6.2. For a given walk, if $\frac{\langle wall \rangle}{4L} > \frac{\langle stiffness \rangle}{2(w-1)L}$, we call it *clingy*. If $\frac{\langle wall \rangle}{4L} < \frac{\langle stiffness \rangle}{2(w-1)L}$, we say it is *stiff*.

The rationale behind this definition is the normalisation using $\frac{1}{4}$ and $\frac{1}{2(w-1)}$, as these represent the relative proportion of ways for the c and a weight to be triggered (see Discussion above), or equivalently, the number of occurrences of a and c in the relevant transfer matrix. We solve for (a, c) for integers $c \in [1, 10]$ that produce equal numbers of stiff and clingy paths (we say these are *balanced*).

Using regression graphs for possible trend-lines, a power relationship between balanced a and c arises with $R^2 > 0.996$, indicating a good fit. However, the exponent of this relationship is dependent on width, and whilst our data can be perfectly represented by the quartic exponent $= -0.0028w^4 + 0.0769w^3 - 0.7471w^2 + 2.9028w - 1.3668$, this is likely not the inherent relation of the system. The explicit relation is left for future research, however, one should note that a power-relation of degree ≈ 2 is almost 'expected' if we consider the extreme cases of a -dominant and c -dominant walks. If we consider the weights acting in a multiplicative manner, the dominant walk of length $2l$ when a is large has weight a^l (the clingy walk), whereas the dominant walk when c is large has weight $c^{2l - \lceil \frac{2l}{w} \rceil}$, which approaches c^{2l} as $w \rightarrow \infty$ for fixed l . Hence, to balance this, it is perhaps intuitive that $a^l \approx c^{2l}$, and so $a = c^2$.



Figure 11: Balanced (a, c) for integers $c \in [1, 10]$



Also of interest is the distribution of the walks at balance points, which we find to be normally distributed when the walks are grouped by (χ, σ) , with dominant walks being those that best satisfy $\frac{\chi\pi}{4} = \frac{\sigma\pi}{2(w-1)}$. In other words, the data is normally distributed about parameters that satisfy the transition point from a clingy path to a stiff path. They spend time both clinging to walls and in diagonal, straight-line motion, with few steps 'wasted' changing directions in the middle of the slit. For example, for $(a, c) = (55, 8), w = 7, L = 100$, the dominant walks are those with $(\chi, \sigma) = (21, 64), (21, 66)$ and $(22, 66)$. One such walk generated is shown below:

Figure 12: Dominant balanced walk of length 100



7 Conclusion

This report has explored a model of polymers within an interactive slit using the previously developed Transfer Matrix Method. Numerical evidence and interpolation strongly suggests the Zero Force Curve for this model is $c = a - 1$, and this result is supported through contemporary work in generating functions (Liu). We furthered our numerical analysis using a sampling algorithm (Beaton) and repeated this iteratively to understand a population of such walks, confirming the Boltzmann distribution involved and bringing ideas of balance between a and c weights. It is important to remember that whilst we presented these Dyck paths as models of interactive polymers, they may fail in their applicability to real-world molecules due to generality. However, not only can this be amended by further refining of the model, but also the discrete and simple Dyck path allows for application to

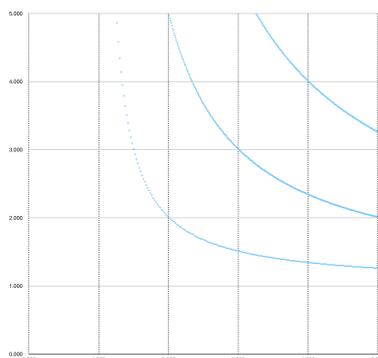


completely unrelated fields. For example, rotating back to the staircase definition of a Dyck path, this model could represent an election where one candidate always has more votes than the other (the bottom wall, or not crossing $y = x$), they are never winning by more than a certain amount (the top wall), and have a tendency to be either a landslide or a tight race (the wall interactions). Perhaps most interesting in this analogue is the consecutive steps representing consecutive votes for the same candidate, which could very well model the impact of human behaviour or the effect of sources of political influence. The further applications, developments and analysis are limited by one's imagination, and as always, left as an exercise for the reader.

8 Addendum and Acknowledgements

As of the submission of this report, it has been found that if the weights of the top and bottom walls are distinguished into a and b respectively, we have the case that the Zero Force Curve (or rather Zero Force Manifold) is given by $ab - a - b + 1 - c^2 = 0$. This is consistent with previous results, such as $ab - a - b = 0$ being the Curve for the $c = 1$ case (Wong [2]), and reduces to $c = \pm(a - 1) = a - 1$ (since we only take positive values of a and c), as we found in this report. This has since been proven analytically.

Figure 13: Zero Force Curves for $c = 1, 2$, and 3 , with a and b on the horizontal and vertical axes, respectively



I would like to thank AMSI for the opportunity to take part in this project and the AMSIConnect conference, which have both been rewarding experiences, as well as Dr Beaton and Dr Wong for their supervision, motivation and enthusiasm throughout the research period. I'd also like to thank Jonathon Liu for not only alerting me of this opportunity, but also being helpful, available and friendly source of banter within the office. You're a good lad.



References

- [1] P. Flajolet and R. Sedgewick *Analytic Combinatorics*. Cambridge University Press, 2009
- [2] T. Wong *Enumeration Problems in Directed Walk Models*. The University of British Columbia, 2015
- [3] R. Brak, A. L. Owczarek, A. Rechnitzer and S. G. Whittington *A directed walk model of a long chain polymer in a slit with attractive walls*. Journal of Physics A: Mathematical and General, 2005
- [4] N. R. Beaton, J. W. Eng and C. E. Soteris *Knotting statistics for polygons in lattice tubes*. Submitted 2018
- [5] L. Collatz *Einschließungssatz für die charakteristischen zahlen von matrizen*. (German) [*Inclusion theorem for the characteristic numbers of matrices*]. Mathematische Zeitschrift, 1942.
- [6] J. Liu *Directed Walk Model for Polymer Propagation*. Australian Mathematical Sciences Institute, 2018.