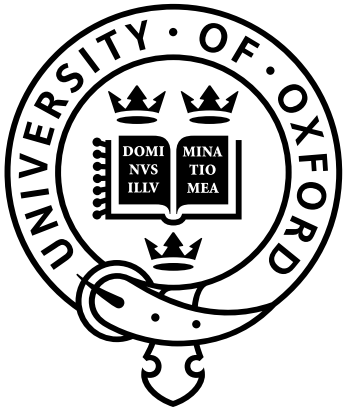
Computational Models of Assembly, Structure and Damage to Collagen IV Networks

Matthew H. J. Bailey

University College

University of Oxford



A Thesis Submitted for the Degree of Doctor of Philosophy

Oxford, Trinity 2022

Abstract.

Collagen IV networks play a key structural role in the human ocular lens capsule. However, their structure is not well understood, and is hypothesised to undergo significant changes over a lifetime, resulting in a loss of ability to accommodate different focal depths. In this thesis, a range of computational models and methods are developed, which aim to help better understand the properties of 2D networks of biopolymers, especially inspired by collagen IV.

The available experimental images of collagen-containing networks are assessed and an image analysis protocol is developed. A Monte Carlo method is designed, which introduces disorder into simplified graph-based networks by switching edges. In addition, Molecular Dynamics is used to simulate coarse-grained polymers which attract one another at their head groups. These coarse-grained polymers are shown to self-assemble into a variety of interesting network structures. The coarse-grained polymer approach is combined with a Molecular Statics scheme to calculate mechanical properties of the network, and understand how the networks behave when stretched. A model that can mimic the ageing of the network over a lifetime is developed. To better understand how lens capsule motion may damage collagen IV networks, the rupturing behaviour of the polymer networks is examined by stretching ordered hexagonal networks linearly and sinusoidally until they break. This is compared to the rupturing behaviour of networks containing defects.

The MD method is used to assess the surface roughness and dimensionality of polymer networks, which is quantified by a fractal dimension. Finally, a new type of persistence diagram is used to quantify and visualise damage to ring structures in a network over time.

Acknowledgements.

0.1 Institutional.

I am grateful to University College, Oxford, for the provision of a Radcliffe Scholarship to support my studies. I am also grateful to the Theory and Modelling in the Chemical Sciences EPSRC CDT.

0.2 Personal.

There are many people who I would like to thank for their help in this thesis and the last three years of my life that have led up to it.

- I would like to thank Mark Wilson, my supervisor, for keeping me on the right track even during the pandemic. He has successfully found just the right balance of leaving me to my own devices against giving me advice and insights on what to do next, and for that I am immensely grateful.
- I would like to thank the many characters of the Theory Wing for their fascinating discussions on science, Fermi estimations, or categorisation — including David, Max, Martin, Laszlo and Tim.
- My TMCS cohort have also been excellent help, answering questions and providing support, especially Rhos, Harry, Mike, Tristan, Jack, Dilhan, and Ava.
- My family — Caroline, David, and Moss, for their delivery of care packages of wine and tea and listening quietly on long walks while I explained.
- And finally, Anna, whose companionship and support has been invaluable. Especially over the last 36 months of being indoors in a small flat together and still tolerating me!

Publications.

0.3 Published.

This work has lead to the following publications, with associated credit statements:

- 'Simplified computational model for generating biological networks' M. H. J. Bailey *et al.*, *RSC Adv.*, 2020, **10**, 38275–38280

  Matthew H. J. Bailey: *Data Curation, Investigation, Software, Visualization, Writing — Original Draft*

- 'Self assembly of model polymers into biological random networks' M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262

  Matthew H. J. Bailey: *Data Curation, Investigation, Software, Visualization, Writing — Original Draft*

- 'Simulation of defects, flexibility and rupture in biopolymer networks' M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180

  Matthew H. J. Bailey: *Data Curation, Investigation, Software, Visualization, Writing — Original Draft*

0.4 Upcoming.

The following papers are in preparation or have been submitted:

- 'Persistence Diagrams to Visualise Damage to Biological Networks' M. H. J. Bailey & M. Wilson, *Submitted to Bulletin of Mathematical Biology*, 2022, DOI: `10.26434/chemrxiv-2022-k4thc`

  Matthew H. J. Bailey: *Data Curation, Investigation, Software, Visualization, Writing — Original Draft*

- 'Rumpling of 2D Biopolymer Networks' M. H. J. Bailey & M. Wilson, *In Preparation*, 2022

  Matthew H. J. Bailey: *Data Curation, Investigation, Software, Visualization, Writing — Original Draft*

Table of Contents

Abstract. **ii**

Acknowledgements. **iii**
0.1 Institutional. . . . iii
0.2 Personal. . . . iii

Publications. **iv**
0.3 Published. . . . iv
0.4 Upcoming. . . . iv

List of Figures **xi**

List of Tables **xv**

1 Introduction. **1**
1.1 Motivation. . . . 1
1.1.1 Ocular Lens Capsules and Biopolymer Networks. . . . 1
1.1.2 Importance of Models. . . . 4
1.1.3 Empirical Laws. . . . 6
1.2 Literature Review. . . . 8
1.2.1 Collagen IV. . . . 8
1.2.2 Polymer Models 9
1.2.3 Graph Based Models. . . . 11
1.3 Thesis Overview. . . . 13

2 Image Analysis. **15**
2.1 Introduction. . . . 15
2.2 Computational Methods. . . . 17
2.2.1 Noise Removal. . . . 17

2.2.2 Ridge Detection. . . . 17
2.2.3 Ring Identification. . . . 19
2.3 Results and Discussion. . . . 21
2.3.1 Single Image Analysis. . . . 21
2.3.2 Alternate Interpretations. . . . 22
2.3.3 Imaged Ring Statistics. . . . 24
2.4 Chapter Conclusions. . . . 24

3 Bond Switching. 27
3.1 Introduction. . . . 27
3.2 Computational Methods. . . . 29
3.2.1 Monte Carlo and Simulated Annealing. . . . 29
3.2.2 Potential Model. . . . 30
3.2.3 Geometry Optimisation. . . . 31
3.2.4 Bond Switching. . . . 32
3.2.5 Bond Switching and Stretching. . . . 35
3.3 Graph Classifications and Metrics. . . . 35
3.3.1 Node Degree Distributions. . . . 35
3.3.2 Maximum Entropy. . . . 36
3.3.3 Regularity Criterion. . . . 37
3.3.4 Assortativity. . . . 38
3.4 Results and Discussion. . . . 41
3.4.1 Generated Networks. . . . 41
3.4.2 Comparison to Experiment. . . . 43
3.4.3 Scale-Free *vs* Maximum Entropy. . . . 44
3.4.4 Regularity Criteria. . . . 46
3.4.5 Bond Switching and Stretching. . . . 49
3.4.6 Network Landscapes 53
3.5 Chapter Conclusions. . . . 55

4 Self-Assembly. 57
4.1 Introduction. . . . 57
4.2 Computational Methods. . . . 58

4.2.1 Molecular Dynamics. . . . 58
4.2.2 Minimum Image Convention. . . . 60
4.2.3 Coarse Graining and Potential Model. . . . 60
4.2.4 Temperature Control. . . . 64
4.2.5 Parameterisation and Simulation Protocol. . . . 65
4.3 Results and Discussion. . . . 66
4.3.1 Energetics of Idealised Network Structures. . . . 66
4.3.2 Self-Assembly. . . . 71
4.3.3 Control of the Network Structure. . . . 72
4.3.4 Regularity Criterion. . . . 79
4.3.5 The Network Landscape. . . . 84
4.4 Chapter Conclusions. . . . 86

5 Mechanical Properties. 88
5.1 Introduction. . . . 88
5.2 Computational Methods. . . . 90
5.2.1 Molecular Statics. . . . 90
5.2.2 Molecular Statics Potential Model. . . . 91
5.2.3 Graph to Polymer Conversion. . . . 92
5.2.4 Uniaxial vs Biaxial Stretching. . . . 93
5.2.5 Numerical Considerations. . . . 95
5.3 Results and Discussion. . . . 96
5.3.1 Uniaxial vs Biaxial. . . . 96
5.3.2 Fitted Parameters. . . . 99
5.3.3 Disordered Networks. . . . 101
5.3.4 Bond Switching Snapshots. . . . 103
5.4 Chapter Conclusions. . . . 107

6 Rupturing. 109
6.1 Introduction. . . . 109
6.2 Computational Methods and Metrics. . . . 111
6.2.1 Defect Introduction. . . . 111
6.2.2 Polymer Model. . . . 113

6.2.3 Stretching Until Rupture. . . . 115
6.2.4 Hessian Calculation. . . . 116
6.2.5 Velocity Autocorrelation Function. . . . 117
6.2.6 Inverse Participation Ratio. . . . 118
6.3 Results and Discussion. . . . 118
6.3.1 Hessian and Normal Modes. . . . 118
6.3.2 Linear Stretching. . . . 122
6.3.3 Sinusoidal Stretches. . . . 126
6.3.4 Edge Removal Defects. . . . 129
6.3.5 Combined Defects. . . . 133
6.3.6 Central Hole Defects. . . . 138
6.3.7 Thin Strip Defects. . . . 141
6.3.8 Repairable Damage. . . . 142
6.3.9 Burst Temperature. . . . 145
6.3.10 Varied Thermostat Parameters. . . . 146
6.4 Chapter Conclusions. . . . 151

7 Rumpling. 153
7.1 Introduction. . . . 153
7.2 Computational and Mathematical Methods. . . . 155
7.2.1 Scaling Laws and Surface Roughness. . . . 155
7.2.2 Simulation Protocol. . . . 156
7.3 Results and Discussion. . . . 157
7.3.1 Hexagonal Network Widths. . . . 158
7.3.2 Deformation Induced Rumpling. . . . 162
7.3.3 Defect Induced Rumpling. . . . 165
7.4 Chapter Conclusions. . . . 168

8 Persistent Homology. 169
8.1 Introduction. . . . 169
8.2 Computational Methods and Metrics. . . . 173
8.2.1 Ring Existence Function. . . . 173
8.2.2 Polymer Model and Simulation Protocol. . . . 174

8.2.3 Persistence Diagrams. 175
8.2.4 Birth-Death Diagrams. 176
8.2.5 Lifetime Metrics. 177
8.3 Results and Discussion. 178
8.3.1 Ordered Networks. 178
8.3.2 Disordered Networks. 181
8.3.3 Effect of Sampling Timestep. 186
8.3.4 Combined Trajectories. 190
8.4 Chapter Conclusions. 193

9 Conclusions and Future Work. 195
9.1 Conclusions. 195

10 References. 198

A Glossary. A1

B Algorithms. B1
B.1 Markov Chain Monte Carlo. B1
B.2 Non-Local Means Denoising B2

C Data. C1
C.1 Regularity Data C1

List of Figures

Chapter 1 **3**

1.1 Young's Modulus as measured by Fisher and Krag 3

1.2 Two possible graph structures proposed for collagen networks. 4

1.3 A single human collagen molecule showing curvature 10

1.4 An Erdős-Rényi graph and Barabási-Albert graph 12

Chapter 2 **19**

2.1 Example of detecting ridges in an image. 19

2.2 Removing Delaunay edges from a network to identify rings 20

2.3 Stages of image analysis for a single image 21

2.4 Ring structure extracted from example image 22

2.5 Two different interpretations of the same image 23

Chapter 3 **29**

3.1 A non-planar graph that would be rejected 30

3.2 Schematic WWW bond switch 33

3.3 Schematic extended bond switch 34

3.4 Periodic network wrapped into a torus 34

3.5 Networks generated by bond switching with stretching 36

3.6 Three hexagons of different shapes to show distinct regularity metrics. 39

3.7 Two graphs and their joint degree distribution 40

3.8 Images of generated Monte Carlo networks 42

3.9 Images of biological networks compared to similar-looking generated networks. 44

3.10 Distributions of k for Monte Carlo networks 45

3.11 Regularity of inorganic vs biopolymer network 47
3.12 Shape regularity criteria 49
3.13 Formation of defects over the course of a bond switching simulation with stretching 50
3.14 Formation of defects of stretched *vs* unstretched networks 51
3.15 Monte Carlo network landscapes 52

Chapter 4 **59**

4.1 Minimum image convention 61
4.2 Collagen IV *vs* a simplified schematic molecule. 62
4.3 Three frustrated arrangements of polymers 64
4.4 Comparison of two idealised networks 69
4.5 Energy minima for hexagonal or chain-link fence networks 70
4.6 Degree *vs* energy 72
4.7 Self-assembly as a function of t_{cool} 75
4.8 Self-assembly as a function of density 77
4.9 Self-assembly as a function of ε_{HH} 78
4.10 Two self-assembled networks coloured by regularity 80
4.11 Two failure modes of the attractive site model. 82
4.12 Exemplar pre-tangled molecules. 83
4.13 Networks formed with pre-tangled polymers 83
4.14 Landscape for self-assembled networks 86

Chapter 5 **91**

5.1 Binary search for Poisson's ratio 95
5.2 Identical hexagonal networks stretched either uniaxially or biaxially. 98
5.3 Traction-stretching ratio curves for the polymer model *vs* experimental data. 100
5.4 2D stress-strain curves for disordered networks 103
5.5 2D stress-strain curves for networks with defects 107

Chapter 6 **112**

6.1 Two schematics of damage types 112

6.2 A comparison of Harmonic and Morse potentials 114
6.3 Frequencies of the normal modes of the hexagonal network sorted by magnitude. 120
6.4 Eigenvectors visualised at the locations of the jumps in **Figure 6.3**. 121
6.5 Inverse participation ratios of the eigenvector for two networks. 121
6.6 Eigenvectors for hexagonal and defect-containing networks 122
6.7 Extremes of the network rupturing landscape. 124
6.8 Snapshots of a rupture of a hexagonal polymer network stretched linearly . . 125
6.9 Average node degree and edges per ring for rupturing networks 126
6.10 Snapshots of networks undergoing stretching with different parameters 128
6.11 Fraction of broken bonds as a function of frequency ω and amplitude A. . . . 129
6.12 Time to first damage and MSD with edges removed. 132
6.13 Time until first damage as a function of the system size. 134
6.14 Damage trajectories as a function of system size. 135
6.15 Three networks with different combinations of defects. 135
6.16 Snapshots across a sinusoidal stretching simulation 137
6.17 Damage metrics for networks with removed edges and/or defects. 138
6.18 Networks with holes in 139
6.19 Damage metrics for networks with holes in 140
6.20 Time until first damage and the damage trajectory over time for different sizes of strip. 142
6.21 Snapshots of repairable and irreparable damage 143
6.22 Damage metrics for networks stretched sinusoidally showing repair 144
6.23 Snapshots of a network rupturing with added kinetic energy 145
6.24 Damage metrics for networks with added energy 146
6.25 Velocity autocorrelation function (VACF) for two networks with energy added. 147
6.26 Effect of damping parameter γ on network damage. 148
6.27 Damage trajectories with different added energy. 149
6.28 Inverse Participation Ratio (IPR) for two networks with energy added. 150
6.29 VACF for two networks with energy added. 150

Chapter 7 **161**

7.1 Three views of a network and corresponding roughness. 161

7.2 Hexagon-rectangle transformation 163
7.3 Roughness of networks undergoing stretching 164
7.4 Widths during a stretching simulation at two frequencies 164
7.5 Height-height correlation by distance from defect 166
7.6 Rumpled width against number of defects 166

Chapter 8 **171**

8.1 A hierarchy of features across the same network. 171
8.2 A time persistence diagram for a simple ring system 176
8.3 Snapshots of ring damage to a hexagonal network 178
8.4 Persistence and birth-death diagrams for a stretched hexagonal network . . . 179
8.5 Birth-death diagrams and persistence diagrams for networks with different areas per molecule. 182
8.6 Magnified sections of persistence diagrams 184
8.7 Persistence and birth-death diagrams for systems sampled at different frequencies. 187
8.8 Fraction of time between first birth and last death of rings. 188
8.9 Fraction of time alive between first birth and last death. 189
8.10 Combined birth-death diagrams for trajectories repeated with the same conditions 192

Appendix C **C2**

List of Tables

Chapter 1 **3**

Chapter 2 **19**

2.1 Experimental data table 25

Chapter 3 **29**

3.1 Statistics for images of biological networks 43

Chapter 4 **59**

4.1 Self-assembly as a function K_θ 73

4.2 Self-assembled networks at different densities 76

4.3 Network statistics as a function of $\varepsilon_{\mathrm{HH}}$ 79

4.4 Regularity data for t_{cool} and $\varepsilon_{\mathrm{HH}}$ 80

4.5 Network statistics for tangled networks 84

Chapter 5 **91**

5.1 Fitted parameters for different sets of experimental data 101

5.2 Correspondence between the 2D stress-strain curves after the bond switching aging process and the matching 3D stress-strain curves 106

Chapter 6 **112**

7.1 Roughness fractal parameters at varied temperatures 160

Chapter 7 **161**

Chapter 8 **171**

8.1 Number of unique rings and rebirths per ring 184

Appendix C **C2**

C.1 Experimental regularity data C2

C.2 Self-assembled regularity data C3

C.3 Self-assembled regularity data seeded with defects C4

List of Algorithms

1 The shape identification algorithm using a Delaunay Triangulation. 20

2 The algorithm describing the Monte Carlo (MC) framework surrounding single bond-switching step. 29

3 The steepest descent algorithm used in the MC simulations. 32

4 The Velocity Verlet (VV) algorithm, where δt is a suitably small timestep. . . 59

5 the Molecular Statics algorithm. 91

6 Algorithm to convert a graph network into a polymer network. 93

7 Binary search algorithm to find zero stress scaling factor λ_x. 94



Chapter 1: Introduction.



This chapter contains details of the motivation for this thesis, including the biological origins of the problem of studying complex networks. The history and importance of modelling biopolymer networks are discussed, with a particular focus on the role of empirical laws. The current literature is also discussed, covering the state of the art in modelling collagen IV (and limitations thereof), as well as a brief history of polymer and graph-based models.



1.1 Motivation.

1.1.1 Ocular Lens Capsules and Biopolymer Networks.

Networks of biomolecules are ubiquitous in complex living beings, and occur in many forms. Constructs such as wood or bone are made up of hierarchies of networks combined by nature to be strong, or flexible, or both[1,2]. The lowest layers of these hierarchical materials are often two dimensional (2D) networks. The most common 2D network in animal life is the basement membrane, made of collagen IV[3]. Collagen IV forms a variety of collective structures in biology, either forming extensive 2D networks or forming fibrils (by joining longitudinally) that then form into fibres[4]. It is networks that are of most interest in this thesis, especially the age-related changes to networks occuring in the human ocular lens capsule.

It has been known for over 100 years that the ability of the human eye to accommodate

focal depths changes as the eye ages[5]. There are two competing theories as to how the human eye accommodates different focal depths. The classic theory, developed by Helmholtz (of thermodynamics fame), is that as the eye focusses on far-away objects a series of zonular fibres round the edge of the lens capsule become tenser to flatten the lens, and *vice versa* when the eye focusses on a close-up object[6,7]. More recent experiments and simulations have shown that mechanism to be unfeasible and support a new theory attributed to Schachar & Bax. The Schachar & Bax theory suggests the lens is in fact more curved in the centre when looking at close objects, and only flatter in the periphery of the lens[8].

Testing which of these competing theories more accurate is difficult as it is not possible to measure the lens thickness in live eyes. Instead, computer modelling techniques of the lens capsule must be used to help ascertain the truth[9,10]. The parameters of these models have been set by comparing to experimental data, including stretching tests on human lens capsules that have been removed after cataract surgery or animal lens capsules[11]. However, parameterising computer models from macroscopic measurements of lens capsules is fraught with its own issues. One such issue is that two different sets of experiments, the first due to Fisher, and the second due to Krag & Andreassen, disagree about how the stiffness of the lens capsule changes over time[12,13]. The data from their experiments are shown in **Figure 1.1**.

To better inform the large scale models of the ocular lens capsule, a microscopic model is required. Most of the change in stiffness of the lens capsule comes from a change in the stiffness of the network of collagen IV filaments[14]. With a microscopic model of collagen IV filaments, future work could inform macroscopic models of the lens capsule, which could allow a better understanding of the behaviour and properties of the human eye. It is hoped that the microscopic models used in thesis can lay the groundwork for future macroscopic models, and provide a better understanding of the properties of materials within the lens

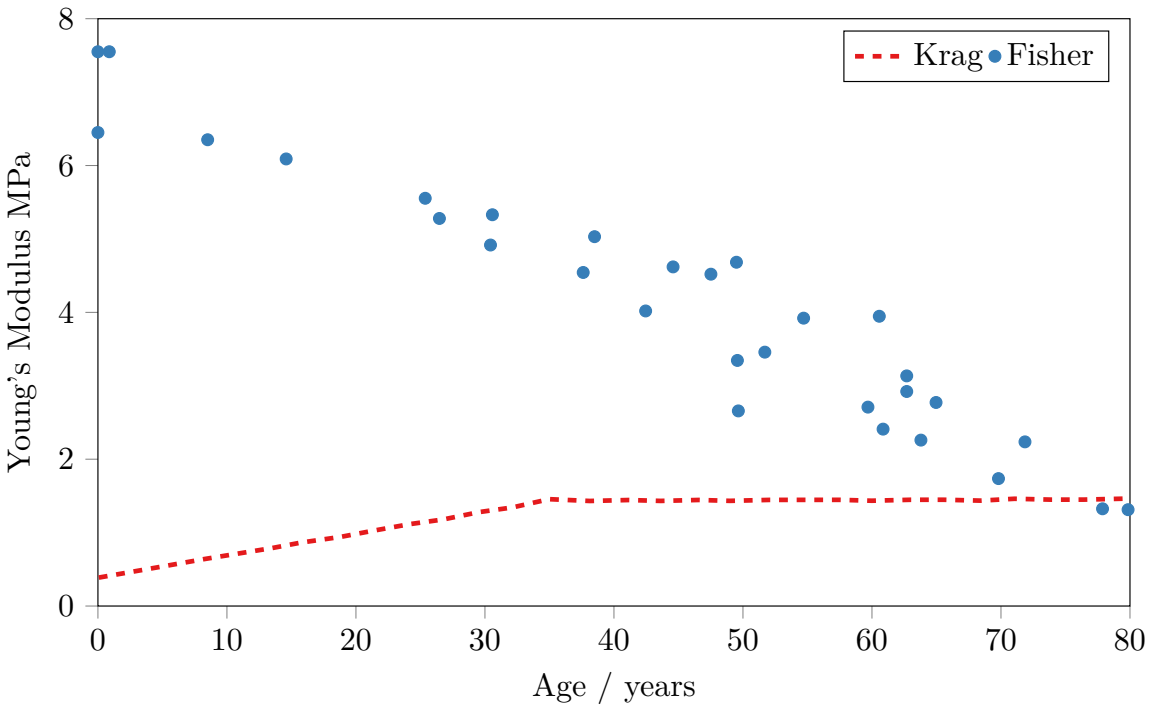
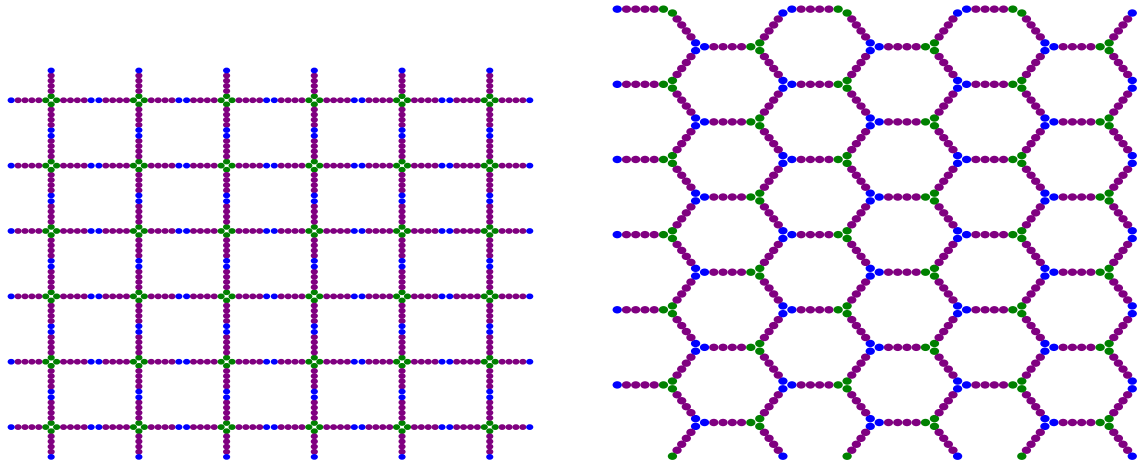


Figure 1.1: The Young's Modulus as a function of age measured by Fisher and using the Young's modulus behaviours proposed by Krag & Andreassen, redrawn from Burd[12–14].

capsule. An understanding of the material properties of the lens capsule would be helpful medically, as artificial intra-ocular lenses must be implanted in human patients who have had their original intra-ocular lens removed (often for cataract surgery). A better match for the material properties of the biological lens could allow for artificial lenses that outperform the aged lenses and lead to improved vision[15].

However, a third layer of confusion arises in the microscopic model, as there is some dispute in the literature as to exactly how the collagen IV molecules assemble themselves. Collagen IV polymers are formed of three triple helices joining two distinct ends, referred to as the 7S and the Non-Collagenous 1 (NC1) domains. The NC1 domains tend to dimerise, and the 7S domains tend to quadrimerise, leading many scientists to conclude that collagen IV networks are a chain-link fence shape[16]. However, studies where collagen IV molecules are allowed to self-assemble into a network show distinctly hexagonal motifs with an average



(a) A chain-link fence network structure with alternating NC1 dimers (blue) and 7S quadrimers.

(b) A hexagonal network structure where all nodes are trimers.

Figure 1.2: Two possible graph structures proposed for collagen IV networks.

degree of interaction sites of approximately three[17]. By comparing mechanical responses of different collagen containing networks, Burla *et al.* found a range of average degrees in the range 2.80 to 3.80[18]. The two different proposed network structures are shown in **Figure 1.2**.

1.1.2 Importance of Models.

Where 2D networks occur in nature, they are often hard to study for a multitude of reasons. Microscopy techniques are useful to provide a great deal of data on small samples of biological networks, but have their own sets of limitations.

The first and most obvious issue with studying a biopolymer network is its biological origins; there are often ethical and practical reasons why experimental data are hard to come by. Secondly, there is no guarantee that a biological network retains its structure when removed from the *in vivo* environment. Early studies used robust biological networks such as plant epithelial cells which are easily prepared and imaged[19]. However, more delicate collagen-containing networks must undergo significant work to prepare them for study. Extracting and preparing delicate networks has been shown to disturb their delicate structure, with

significant knock-on effects for their observed properties[20].

Thirdly, there are limits in the amount of detail that can be imaged, due to device resolution and signal-to-noise ratio constraints. Optical microscopy is limited in resolution due to diffusion of the beam[21], and electron microscopy or laser microscopy have a risk of damaging the delicate networks with high-power beams[20,22].

Fourthly, even when those difficult experiments have been performed and a high-quality image has been obtained, there is ambiguity in how to analyse the image. Each edge in a network is a large protein molecule which may have over 10 000 atoms in it. Those atoms, and contaminents, undergo Brownian motion while the image is being taken, resulting in blurry or unclear images. In more complex cases, there can be multiple layers of 2D networks imaged on top of one another. This means that there may be multiple interpretations of a single image, which can be disambiguated by a computer model calculating which interpretations are more physically reasonable.

These limitations combined are even worse when trying to observe damage over time. The aging process of a lens capsule occurs over a human lifetime, meaning that it cannot be observed in a laboratory timescale. Large-scale atomic simulations are also unfeasible, and can capture at most nanoseconds of a relevant trajectory. The computational expense of all-atom simulations will be discussed in **Section 1.2.1** To that end, an easy way to access a greater computational timescale is through coarse-grained simulation. A coarse-grained simulation of individual polymers can capture microseconds of an assembly process, and a graph model could capture an entire lifetime of aging.

The grand detail in a microscope image or the averaged ensemble data in the measurement of mechanical properties tell us little about the dynamical behaviour of a network. With a chemically informed computer model, dynamical processes can be simulated easily which can

aid interpretation of sparse experimental data.

1.1.3 Empirical Laws.

Traditionally, experimental studies and computer modelling of 2D networks has relied significantly on empirical laws. Empirical laws have been widely used in the studies of 2D glasses as they are easy to fit to (and verify against) experimental data. For example, simply tracking the area of rings against their number of sides provides a valuable insight into a system without requiring a complex model. Furthermore, examining fitted models provides a useful comparison when comparing and constrasting related networks.

Lewis' Law

The earliest empirical law in this field is Lewis' law[19]. Lewis was studying hexagonal cucumber cells and noted that '[t]he excess of six sides in the central cell is met by a deficiency of six in the cells in contact with it.' (Lewis[19]) He continued to develop this idea and 2 years later developed an equation to link the average area A_n of a ring to the number of sides n via

$$A_n = \alpha(n-2), \tag{1.1}$$

where α is a constant to be determined empirically[23].

Aboav-Weaire Law

In studying the arrangement of grains in a polycrystal of MgO, Aboav noticed that '[t]he grains of a polycrystal are found not to be disposed at random, but to be arranged in a way that can be quantitatively expressed in simple terms' (Aboav[24]). He observed, backing up Lewis' earlier observation, that if a grain had more than six sides its neighbours had fewer than six sides[24]. Quantifying this, he formulated a law that became known as the Aboav law

for the average number of sides m_n of all rings neighbouring a ring with n sides

$$m_n = 5 + \frac{8}{n}. \tag{1.2}$$

This law was further developed by Weaire, who derived a similar law by combining Euler's characteristic for polyhedra with the fact that the angles around a node must add up to 360°[25]. This lead to an improved formulation termed the Aboav-Weaire law

$$m_n = (6 - a) + \frac{6a + \mu_2}{n}, \tag{1.3}$$

in which a is a fittable parameter and μ_2 is the second moment of the distribution of edges per ring[26,27]. A number of other authors, including Edwards & Pithia, Fortes, and Wang & Liu attempted to extend the Aboav-Weaire law to 3D, also often empirically, respectively obtaining[27–29]

$$m_n = a + \frac{b}{cn + d} \tag{1.4}$$

$$m_n = \langle n \rangle + \frac{2\mu_2}{n} + a - \left(\frac{\langle n \rangle a + \mu_2}{\langle n \rangle^2} \right) n \tag{1.5}$$

$$m_n = 13 + \frac{14}{n}. \tag{1.6}$$

The conflict between these different forms of the Aboav-Weaire law was resolved by Mason *et al.*, who pointed out that 'a linear model is a good approximation of any smooth function over a small domain, and that the success of the law of Aboav–Weaire does not necessarily indicate that the average excess curvature is actually linear in n' (Mason *et al.*[30]). Mason's comment brings up a critical point about empirical laws. The range of edges per ring in physical systems is strongly restricted; one of Aboav's initial systems only had rings with 5, 6 or 7 edges[26]. For systems with few (and ambiguous) data points, it is easy to fit an arbitrary function to the data.

To resolve that problem, much of the more recent work in network studies has replaced the Aboav-Weaire law with assortativity[31]. The assortativity r represents a correlation function

between neighbouring rings. $r = 1$ indicates that there is a perfect correlation for a given ring to be neighboured by similar rings, $r = -1$ indicates there is a perfect correlation for a given ring to be neighboured by dissimilar rings, and $r = 0$ indicates no correlation at all[32]. The assortativity is useful as a more statistically interpretable and less empirical descriptor of networks compared to the Aboav-Weaire a parameter[33], and will be discussed in more detail in **Section 3.3.4**.

1.2 Literature Review.

Models of biological networks have been attempted in many forms at many different scales using a variety of techniques. This thesis details work using a mixture of graph-based models to perform large-scale simulation of biological networks, and polymer models to treat dynamical processes. While each upcoming chapter in this thesis will have a detailed literature review at the start (for example, a literature review of bond switching can be found in **Section 3.1**), this section serves as a general overview of previous modelling attempts in the field.

1.2.1 Collagen IV.

An individual collagen IV molecule comprises over 10 000 atoms. To accurately study a protein, hundreds of thousands (if not millions) of solvent atoms must also be simulated. When there was enough computational power available to meaningfully attempt large-scale protein simulation in 1999, Klein & Huang generated 'collagen-like peptides' and simulated 1 ns of their motion in a solvent using an all-atom Molecular Dynamics (MD) simulation powered by the AMBER-94 force field[34,35]. 12 years later, the MD simulations from the same research group were expanded to cover the full-length of the collagen IV triple helix (not including 7S or NC1 domains at either end) and involved approximately 1.8 million atoms

over 10.0 ns[36]. The cost of these simulations remains extremely high, as the aforementioned simulation took almost 11 months on 250 000 CPUs distributed across the Folding@Home project. Because of this, detailed atomistic simulation of an entire network is unfeasible (to say nothing of more accurate electronic structure calculations) and we must turn to polymer models.

1.2.2 Polymer Models

To capture all of the large-scale physics of interest, a simpler computational model informed by the small-scale atomistic detail must be designed.

One such simple model is to represent a single collagen IV molecule as a series of beads joined by springs[37,38]. As the triple helical backbone of a collagen IV molecule is semi-flexible, freely-moving beads with distance constraints as are often used for carbon-based polymers are unsuitable. A better model is the mathematical worm-like chain, which has been shown to accurately simulate the curvature and mechanical properties of individual collagen molecules[39,40]. One example of a worm-like chain as a measure of the folding behaviour of a single collagen molecule is shown in **Figure 1.3**. By controlling the interaction potentials between beads in different molecules, it is possible to generate complex cross-linked networks of biopolymers that mimic reality[41].

Effect of Salt.

The structure of collagen IV is not fixed, and is strongly dependent upon environmental conditions. For example, the flexibility of a single collagen molecule (of any type) in solution is strongly dependent on the concentration of dissolved salts and pH. Studies across different types of collagen in different conditions have found a range of persistence lengths (a measure of flexibility) in the range 10.0 nm to 167 nm which are difficult to reconcile[42,43]. Rezaei *et al.*

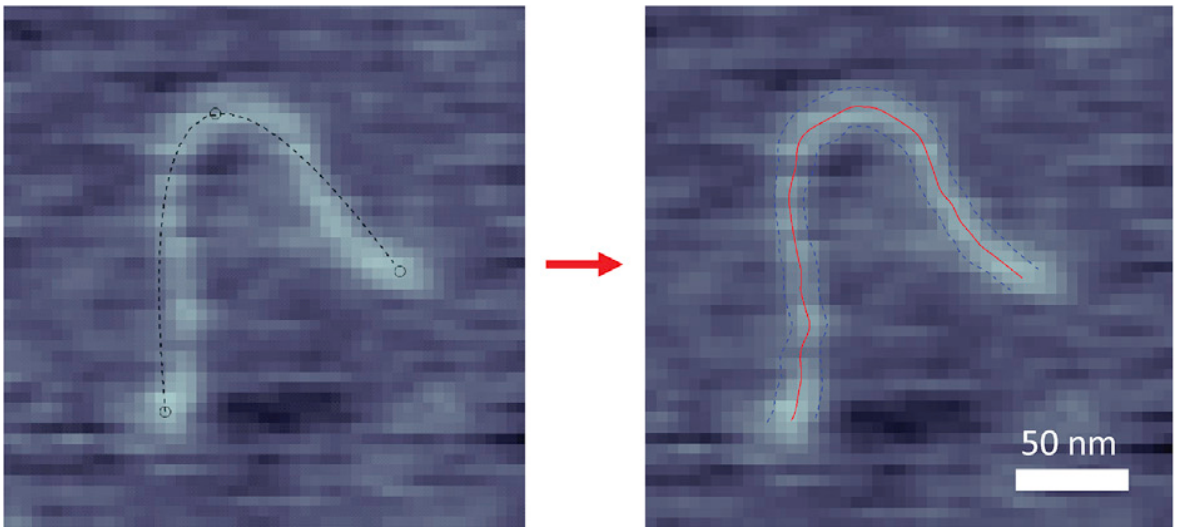


Figure 1.3: A single human collagen molecule showing curvature, with a fit worm-like chain (left panel) the central line of the polymer identified (right panel). Reproduced from N. Rezaei *et al.*, *Biophys. J.*, 2018, **115**, 1457–1469, Copyright (2018), with permission from Elsevier.

used a worm-like curve model to analyse collagen molecules of types I, II and III[39]. They found that increasing salt concentration or pH causes a naturally bent collagen molecule to become straighter and less flexible, with a change in persistence lengths from 60.0 nm to 120 nm.

Similarly, dissolved ions mediate the interactions of NC1 domains at the one end of a collagen IV molecule. When two NC1 trimers on different collagen molecules come together, intramolecular salt bridges of chloride ions rearrange to become intermolecular salt bridges within hexamers[44–46]. Different cations can also affect the assembly of collagen into larger structures: Cr^{3+} has been shown to form clusters between fibrils without significantly affecting ring structure[47]. However, Wang *et al.* found that Na^{+}, K^{+}, Mg^{2+} and Ca^{2+} had significant effects on the structure of collagen networks deposited on a mica surface[48]. This was rationalised in terms of the Hofmeister series dictating which ions 'salt out' proteins from solution[49]. It is difficult to extricate experimentally whether the effects of salt on stiffness or head group interactions are what lead to the difference in structure, as both effects occur

simultaneously.

1.2.3 Graph Based Models.

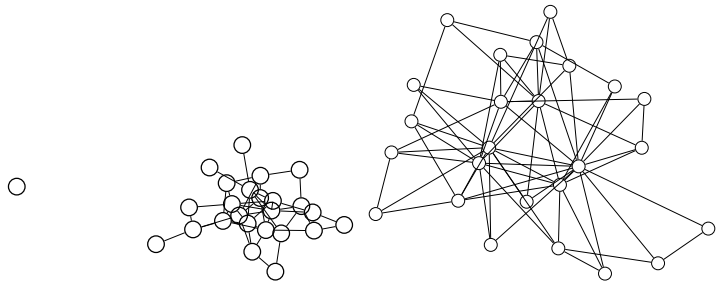
Complex networks feature a high degree of randomness and a great deal of unnecessary detail — the precise positions of atoms within a protein chain are often unimportant when it comes to assessing behaviour on a large scale. The extra detail can be removed by considering which pairs of polymers interact (or, more abstractly, pairs of interacting objects in the network). In a protein network, an edge represents a single collagen molecule, with a node representing a site where two or more molecules interact. Network theory is not limited to such concrete examples, and has been used in more abstract ways to great success: for example, an edge in a social network may represent a friendship, or an edge in a protein interaction network may represent a protein and enzyme reacting.

The complexity of many networks in nature often arises from random phenomena, which can be included using random graph models. The earliest random graph models were developed contemporaneously by Gilbert (who was considering telephone networks for Bell Labs) and also Erdős & Rényi (who were coming from a more abstract mathematical viewpoint)[50,51]. The resulting random graphs are called Erdős-Rényi graphs, and they can be generated in two ways:

1. Generating N nodes and randomly assigning m edges between them, called a $G(N, m)$ graph.

2. Generating N nodes and drawing edges between them with a probability p, called a $G(N, p)$ graph.

One example of a $G(N, p)$ graph is shown in **Figure 1.4a**. The number of edges connected to each node (referred to by graph theoreticians as the degree k or by chemists as the coordination



(a) An Erdős-Rényi graph with $p = 0.15$ and $n = 25$.

(b) A Barabási-Albert graph with $m = 3$ and $n = 25$.

Figure 1.4: Exemplar Erdős-Rényi and Barabási-Albert graphs with the same number of nodes. The Erdős-Rényi was constructed by randomly attaching two edges with a probability p, but the Barabási-Albert was constructed by a preferential attachment process.

number) in a $G(N, p)$ random graph follows a binomial distribution

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \tag{1.7}$$

which tends towards a Poisson distribution as N tends to infinity.

When the field of random networks had matured, there was some debate in the literature as to how random the networks in nature truly were[52]. Barabási & Albert noticed that in many actual networks the distribution of degrees followed a power law of the form

$$P(k) \propto k^{-\gamma}, \tag{1.8}$$

with γ a per-network constant[53]. They described networks like this as being 'scale-free' and argued that Erdős-Rényi graphs miss key features of actual random networks[53]. A Barabási-Albert network is shown in **Figure 1.4b**, described by a number of nodes n and an initial degree of $m = 3$. The scale-free networks are constructed by adding nodes one-by-one each with initial degree m and with a probability that each new node is connected to previous node i (with coordination number k_j) given by

$$p_i = \frac{k_i}{\sum_j k_j}. \tag{1.9}$$

This leads to a small number of nodes that are highly connected in a 'rich get richer' approach. The Barabási-Albert model showed that computer models of complex networks can be easily

constructed by biased random processes. However, more recent work by Khanin & Wit has questioned whether biological networks (specifically protein interaction networks) are truly scale-free[54]. Khanin & Wit showed that in many cases, networks had been described as being scale-free, but this was not statistically certain. This leads to a need for better, and more physically informed, models of biological networks. It is those models of biological networks, informed by microscopic chemical factors, that will form the remainder of this thesis.

1.3 Thesis Overview.

The rest of this thesis will examine the problems raised in the introduction and explore a number of ways to model different properties of complex protein networks. As each approach involves different simulation methods, each chapter will be relatively self-contained and detail the relevant literature and computational methods.

- **Chapter 2** will discuss a process to convert microscope images of biopolymer networks into graph representations.
- **Chapter 3** will discuss the usage of bond switching methods to generate biologically feasible networks.
- **Chapter 4** will discuss using a polymer model that self-assembles into collagen IV-like networks.
- **Chapter 5** will discuss the mechanical properties of polymer networks.
- **Chapter 6** will discuss the rupturing behaviour of collagen IV-like networks when stretched.
- **Chapter 7** will discuss the width of 2D networks when constraints are relaxed.

- **Chapter 8** will discuss the usage of persistence diagrams to analyse rupturing behaviour.

Chapter 2: Image Analysis.



Before attempting to simulate collagen-like networks, it is first important to establish the properties of experimental networks. Microscope images are the most common form of experimental data, and contain a high level of detail. These are often prepared under significantly different conditions or imaged using differing techniques. This chapter establishes a process that converts images of biopolymer networks into graph representations which can then be analysed. Graph representations are free from ambiguity and can be treated computationally with the same techniques as simulation results.



2.1 Introduction.

Computational models can be most useful when they provide insight into real physical phenomena backed up by experimental evidence. Experimental evidence for complex phenomena often provides very precise information about a small system (for example an electron microscope image) or averaged information over a large system (for example, mechanical properties of an eye). Computer models are useful because they can provide information about systems between those two extremes, and can be easily interrogated to extract other properties of interest.

When it comes to collagen networks, the most common experimental evidence about the

network structure is in the form of microscope images. Imaging experiments are difficult to perform, as they require studying a very thin membrane from a biological system which can be easily damaged. Historically, limitations on image resolution and clarity meant that only heavily-processed collagen networks provided high-quality images. For example, Bos *et al.* used a freeze-etching technique, in which they froze a collagen network, formed a replica by depositing a mixture of platinum and carbon, then finally imaged that replica[55]. Yurchenco & Ruben used a similar freeze-etching process, and also used a complex purification process to remove non-collagen components from the basement membrane[56].

There is also a great diversity in the collagen (and collagen-like) networks that have been imaged in the literature. Amongst others, Cavalcante *et al.* imaged collagen networks in lung tissue[57]; Reznikov *et al.* imaged collagen networks in bones[2]; and Wang *et al.* imaged collagen networks that had self-assembled in the presence of cations[48]. The question of structure is further complicated because separating individual two dimensional (2D) layers of collagen networks is extremely hard. For example, the often-cited image created by Barnard *et al.* shows multiple layers of a collagen network[58]. Finally, the phrase 'collagen network' can have different meanings depending on the length scale of the image:

- a network of individual collagen molecules (as used in this work),
- a network of collagen fibrils (themselves made up of collagen molecules),
- a network of collagen fibres (themselves made up of fibrils)[59].

The diversity in images taken of different networks over 50.0 years of publications means that there is no 'one size fits all' approach to analysing them so that they can be compared to simulation results. A more tailored algorithm can be applied to networks that are imaged and prepared consistently, such as the process designed by D'Amore *et al.* to characterise

connectivity in planar collagen fibre networks[60]. But for a variety of different images, as seen in this thesis, a flexible but consistent set of rules must be developed.

2.2 Computational Methods.

2.2.1 Noise Removal.

Microscopy experiments must be extremely sensitive to be able to identify small height or brightness differences in the network of interest, which is potentially on a rough surface. The sensitivity of the equipment can also lead to a significant amount of random noise in the final image. Noise can present issues for further analysis, so must be removed before any further processing is performed. In order to reduce the visual noise in a network image, the colours of the constituent pixels can be averaged across a neighbourhood. The details of the non local means denoising (NLMD) algorithm can be found in **Section B.2**. This process increases the pixel signal-to-noise ratio enough to allow further image processing steps to proceed smoothly. The NLMD algorithm was used as implemented in the plugin for ImageJ[61,62].

2.2.2 Ridge Detection.

A critical step in analysing a polymer network is identifying where the polymers are in space. There are multiple ways to approach this problem, including the Ridge Detection Algorithm[63] and the Hough Transform[64]. Established methods to identify edges in an image are not suitable for detecting polymers of finite width. This is because edge detection will identify the sides of the polymer and not the central ridge. In a greyscale image, polymers are either consistently brighter or consistently darker than the background. Here only the case where polymers are consistently brighter than the background will be considered for simplicity, as the alternative can easily be treated by negating the image. This is especially useful for

certain types of microscopy (such as atomic force microscopy) where the brightness of a pixel directly represents height above the scanning surface. Along a polymer the brightness $z(x, y)$ (or height) remains constant along the length of the polymer, and drops off either side of the polymer. Detecting the polymer then becomes a case of finding a parameterised curve $s(t)$ where $\frac{\mathrm{d}z}{\mathrm{d}t} = 0$ in the direction along that curve[63]. As the images are often noisy (even after noise reduction using the NLMD algorithm) the derivatives can be more accurately estimated by convolving the image with the derivative of a Gaussian kernel. However, a purely Gaussian kernel will calculate the wrong position of a ridge that has different heights either side[65]. To resolve this, the Gaussian kernel can be combined with a bar shape. In a parameterised direction t the kernel is given by

$$k(t) = \int_{\infty}^{t+w} \exp\left(-\frac{x}{2\sigma^2}\right) \mathrm{d}x - (a-1) \int_{\infty}^{t-w} \exp\left(-\frac{x}{2\sigma^2}\right) \mathrm{d}x, \tag{2.1}$$

in which σ, w and a are parameters chosen to best extract ridges from the image of interest.

A curve can be identified by calculating the Hessian matrix of the brightness

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 z}{\partial x \, \partial x} & \frac{\partial^2 z}{\partial x \, \partial y} \\ \frac{\partial^2 z}{\partial y \, \partial x} & \frac{\partial^2 z}{\partial y \, \partial y} \end{bmatrix}. \tag{2.2}$$

The two eigenvectors of this matrix point respectively to directions along and perpendicular to the curve. A curve can then be constructed by starting at the point with the largest value of $\frac{\mathrm{d}z}{\mathrm{d}t}t$ and connecting adjacent pixels in the appropriate directions. Where two curves meet, or a curve extends longer than a chosen maximum length, a node is placed. This algorithm is used as implemented by the RidgeDetection plugin for ImageJ[62,66]. Finally, edges can be straightened out by connecting nodes and a graph representation can be formed of the overall network.

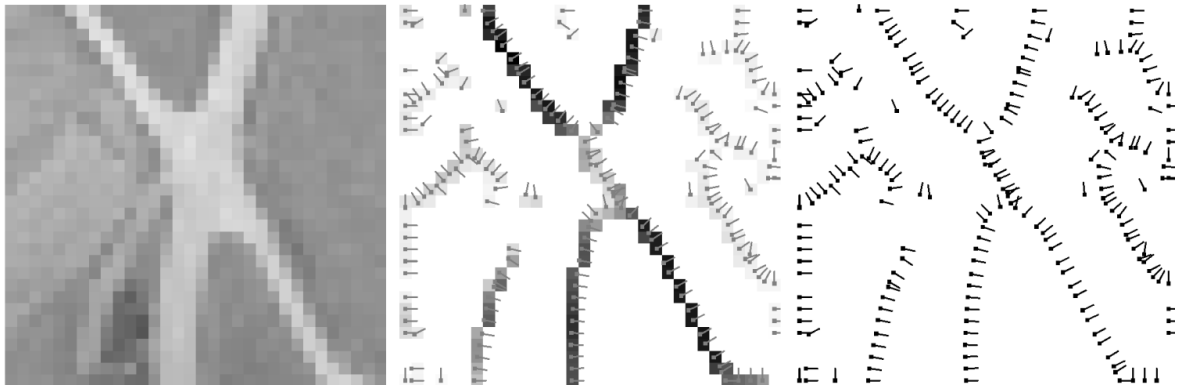


Figure 2.1: Example of detecting ridges in an image. The first panel shows an aerial image of roads. The second panel shows which points are detected as lying on the ridge, with directions superimposed. The off-ridge eigenvectors are shown in the third panel, and superimposed on the second panel. © 1998 IEEE. Reprinted, with permission, from G. Steger, *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, **20**, 113–125.

2.2.3 Ring Identification.

The edges forming a network are interesting on their own, but it is their medium range ordering in the form of ring structure that is of most interest in this thesis. Detecting rings from a 2D graph is a complex task, albeit considerably simpler than the comparable task of detecting rings in a three dimensional (3D) network. The rings in a 2D network can be assigned unambiguously if the network is planar (*i.e.* edges only meet at nodes and do not overlap), and if the coordinates of all nodes are known (described in graph theory as an embedding). The ridge detection algorithm described above provides a 2D embedding in the form of pixel coordinates for each node and edges between them. These can be turned into 'real' coordinates by multiplying by the scale factor of the image. With the coordinates calculated, an anticlockwise ordering of neighbours around a given node can be assigned. The anticlockwise ordering means that rings can be found by starting at a given node and traversing the graph by moving anticlockwise at each neighbour until the starting node is revisited[67,68]. Some issues occur when studying periodic graphs, as assigning the cycling

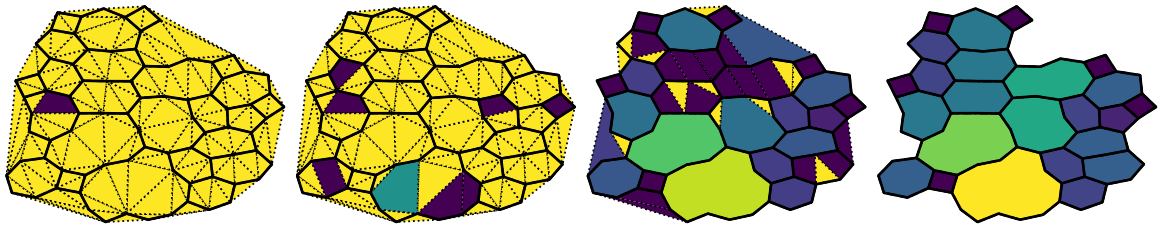


Figure 2.2: The process of removing Delaunay edges from a network to identify rings. The solid lines represent edges in the both the original network and the Delaunay triangulation, whereas the dotted lines exist only in the Delaunay triangulation. From left to right, dotted lines are removed to join yellow triangles into larger rings, coloured by their number of edges.

ordering around a node on the edge can be complex.

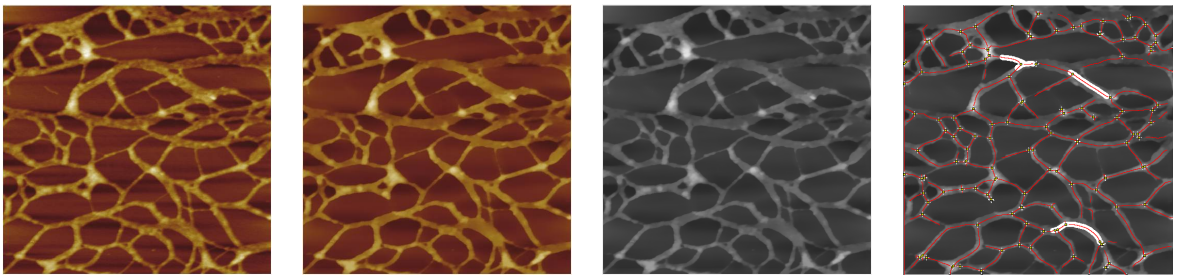
Instead, a method based on the Delaunay triangulation will be used throughout this thesis to identify rings in both periodic and aperiodic networks. The Delaunay triangulation and its dual, the Voronoi partition, are commonly calculated when studying networks so using them to assign rings costs little extra computational effort. This method works in a planar aperiodic 2D network by the steps described in **Algorithm 1**. The process is shown graphically in

Algorithm 1: The shape identification algorithm using a Delaunay Triangulation.

```
Compute the Delaunay triangulation D of all the coordinates of each node in a given
  embedding;
Find closed loop of 3 edges in D, and label as a triangle;
for edge (u, v) in D do
    if (u, v) is not in the original network then
        Remove (u, v) and join the two shapes that contain (u, v) into a larger shape;
    else
        Do nothing;
```

Figure 2.2 with a small aperiodic network.

A small amount of extra work is required to identify rings in a periodic network (as generated by computer simulation), because the edges will appear to wrap around the network when embedded in the plane. This can be resolved by applying the minimum image convention to remove edges that span one periodic boundary and adding virtual edges as if the network were unwrapped. The graph creation and ring identification steps were performed by a custom



(a) The raw image before any analysis.

(b) The raw image with random noise removed with the NLMD algorithm as described in **Section 2.2.1**. Here the parameters were chosen to be $\sigma_{\text{NLMD}} = 10$ and $h = 1$.

(c) The image with random noise removed converted to an 8-bit greyscale format for further analysis.

(d) The image with edges highlighted using the ridge detection algorithm. Where the process did not pick up ambiguous edges, they were manually highlighted (see **Section 2.3.2** for further discussion).

Figure 2.3: Stages of image analysis for a single image, with the base image reproduced with permission from Wang *et al.*[48].

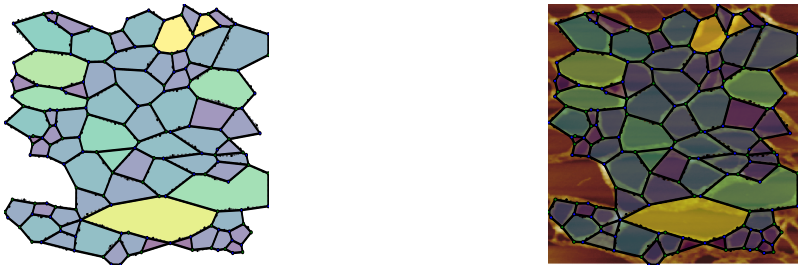
PYTHON script, available on `github.com/WilsonGroupOxford`.

2.3 Results and Discussion.

2.3.1 Single Image Analysis.

The steps described in **Section 2.2** can be applied to a single image to demonstrate the protocol. An image of a collagen network, reproduced from Wang *et al.*, is shown in **Figure 2.3a**[48]. This is an atomic force microscopy image where brightness directly represents height above the surface, with the difference between brightest and darkest pixels being equivalent to a length scale of 5.00 nm. The small difference between brightest and darkest pixels implies that this is a single layer of collagen IV, deposited on a mica surface. A section of collagen network 2.05 µm by 2.05 µm is shown at a resolution of 525 pixel by 524 pixel.

The steps in the image analysis process are shown in the panels of **Figure 2.3**. This image first has random noise removed using the NLMD algorithm, and is then converted to



(a) Ring structure from **Figure 2.3**.

(b) Ring structure from **Figure 2.3** overlaid on the image from **Figure 2.3a**.

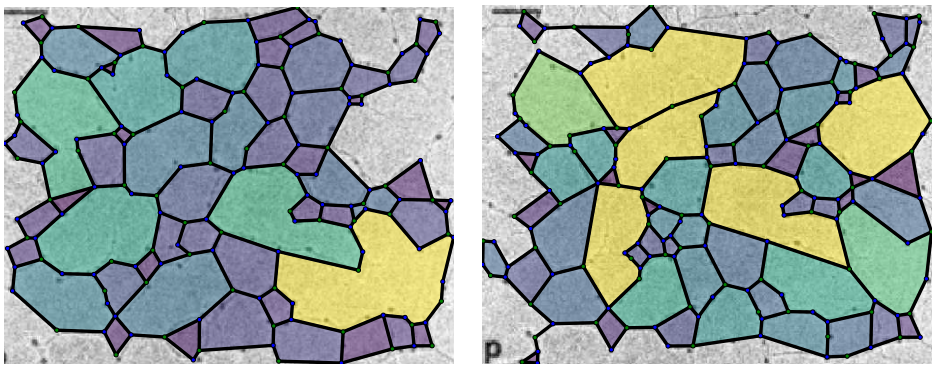
Figure 2.4: The ring structure extracted from that example image after all the stages had been completed. Faint dashed lines can be seen where the approximation of straight edges is inaccurate. The background in the right panel is reproduced from Z. Wang *et al.*, *J. Food Qual.*, 2017, **2017**, 1–10 with permission.

a greyscale heightmap in which each pixel has a z value in the range 0 to 255. Next, the ridge detection algorithm is used to identify the ridges and the nodes where they join. Some ridges in this image are unusually wide, which causes issues for the analysis process. It was observed in the original paper that 'the fibres gathered into bundles at high concentrations [of KAc],' (Wang *et al.*[48]) which implies the troublesome regions are bundles of multiple collagen IV polymers. To resolve this, the regions where an ridge is visible to the eye but cannot be identified algorithmically can be highlighted. The detected ridges are converted into edges of a graph, and then converted into rings.

Final ring structures are shown in **Figure 2.4**, both on their own and overlaid on the original image. This ring structure has a mean node coordination of $\langle k \rangle = 2.87$ and a mean of $\langle n \rangle = 5.55$ edges per ring.

2.3.2 Alternate Interpretations.

The process demonstrated in **Section 2.3.1** created a single ring structure from the image, and disregarded the issue of ambiguity by manually highlighting ridges. Ambiguity in analysing experimental images arises from a number of factors: random noise in the image, limited



(a) **(b)**

Figure 2.5: Two different interpretations of the same image, coloured by the number of edges in each ring. Reprinted with permission from P. D. Yurchenco & H. Furthmayr, *Biochemistry*, 1984, **23**, 1839–1850. Copyright 1984 American Chemical Society.

contrast, or non-2D networks being imaged as 2D. Often an image represents many layers of a complex network and it is extremely difficult to tease them apart. Thanks to this, there is no single network structure that can be assigned to a given image, but a range of possible structures. Different selections of noise removal parameters may exclude valid but noisy edges, or include spurious edges. Similarly, different selections of ridge detection parameters may also include or exclude edges. When ridges have been selected, assumptions are made about which molecules lie in a given plane of the network. Furthermore, post-processing may unify many short ridges into one edge in the graph, or split up a long ridge into many edges in the graph. **Figure 2.5** shows exactly the same network image and two different ring analyses, created by manually highlighting different edges and selecting different parameters. This has relatively significant effects for the ring statistics: the network on the left has a mean ring size of $\langle n \rangle = 6.25$ and a second moment of the ring distribution $\mu_2(n) = 7.00$, whereas the network on the right has $\langle n \rangle = 6.59$ and $\mu_2(n) = 10.5$.

2.3.3 Imaged Ring Statistics.

Ring statistics for a number of networks imaged by Yurchenco & Furthmayr, Wang *et al.*, Bos *et al.*, Barnard *et al.*, Fabris *et al.* and Yurchenco & Ruben are available in **Table 2.1**[17,48,55,56,58,69]. The data show a number of similarities across the networks; most have an average degree in the range $\langle k \rangle = 2.33$ to 3.01. If the networks were constructed purely from hexagons, we would expect $\langle k \rangle$ to be slightly below 3 as nodes on the edge of the image would be under-coordinated. The second moment of the node distribution $\mu_2(k)$ is relatively small, often around 0.320 — it is sometimes dramatically higher if the network is more complex. The mean ring size $\langle n \rangle$ is commonly in the range 5 to 7 but is considerably more variable with $\mu_2(n)$ being up to 126.

2.4 Chapter Conclusions.

This chapter described a process to extract ring structures from experimental images of biological networks. The process is general, and can be applied to a range of networks of different scales and preparation methods. It is difficult to fully automate this process, as the variety of images means that processing parameters must be chosen per-image. When false positives occurred (identifying ridges where there are none) or false negatives (failing to identify real ridges) then the relevant portions of the image were highlighted manually. Suitable images for analysis are relatively rare in the literature; only 19 were identified and analysed in this chapter. Many of those images were ambiguous, and it was difficult to assign graph structures to analyse them. Where images did exist and were easy to analyse, they provided a great deal of detail about only a small section of network, often less than 100 rings. Thankfully, the range of available images provides a useful baseline and comparison point for

Author	Figure	$\langle k \rangle$	$\mu_2(k)$	r_k	$\langle n \rangle$	$\mu_2(n)$	r_n	N_{node}	N_{ring}	$p_{k,3}$	$p_{n,6}$
Barnard *et al.*[58]	2	2.53	0.406	0.142	6.99	24.6	−0.138	385	91.0	0.385	0.0770
Barnard *et al.*[58]	4 c	2.96	0.378	−0.0710	5.57	4.11	−0.109	283	135	0.696	0.185
Barnard *et al.*[58]	6	3.15	0.824	0.0120	5.12	3.60	−0.0480	520	278	0.529	0.169
Bos *et al.*[55]	2 a	2.78	0.335	−0.0540	5.83	7.89	−0.209	100	40.0	0.620	0.225
Bos *et al.*[55]	5 c	2.88	0.198	0.171	5.31	4.32	−0.228	86.0	39.0	0.791	0.128
Bos *et al.*[55]	5 d	2.97	0.396	0.188	5.39	6.46	−0.107	228	109	0.632	0.119
Fabris *et al.*[69]	3 f	2.34	0.315	−0.108	5.89	15.2	0.162	155	19.0	0.246	0.316
Fabris *et al.*[69]	4 b	2.37	0.285	−0.0770	7.04	27.2	−0.147	230	32.0	0.313	0.0940
Wang *et al.*[48]	1 c	2.66	0.681	0.0640	7.25	126	−0.124	621	185	0.353	0.0810
Wang *et al.*[48]	2 a	2.53	0.372	−0.0270	8.50	31.8	−0.100	618	163	0.416	0.153
Wang *et al.*[48]	2 d	2.87	0.277	−0.0220	5.55	3.02	−0.131	150	66.0	0.727	0.258
Wang *et al.*[48]	3 b	2.95	0.805	−0.0190	5.74	15.3	−0.0570	1070	498	0.458	0.122
Wang *et al.*[48]	4 a	3.01	0.317	0.0850	5.52	5.39	−0.120	348	173	0.684	0.139
Wang *et al.*[48]	5 b	2.53	0.301	−0.0310	8.76	27.6	−0.147	350	94.0	0.480	0.128
Wang *et al.*[48]	5 c	2.99	0.484	0.231	5.36	4.60	−0.0290	216	97.0	0.611	0.175
Yurchenco & Furthmayr 1984[17]	4 b	2.67	0.287	−0.0330	6.59	10.5	−0.184	159	54.0	0.604	0.241
Yurchenco & Furthmayr 1984[17]	4 b ⋆	2.72	0.284	0.0740	6.25	7.00	−0.181	172	61.0	0.657	0.213
Yurchenco & Ruben 1987[56]	1 c	2.79	0.395	−0.0280	5.09	2.61	0.0280	106	42.0	0.595	0.0950
Yurchenco & Ruben 1987[56]	2 f	2.70	0.303	0.170	5.88	6.44	−0.363	69.0	25.0	0.610	0.200

Table 2.1: A full table of the experimental data, with the mean node degree $\langle k \rangle$, the second moment of the degree distribution $\mu_2(k)$, assortativity of nodes r_k, mean ring size $\langle n \rangle$, the second moment of the ring size $\mu_2(n)$, assortativity of rings r_n, number of nodes and rings N_{Node} and N_{ring}, as well as the fraction of degree 3 nodes and 6-members rings $p_{k,2}$ and $p_{n,6}$. Data is taken from Barnard *et al.*, Bos *et al.*, Fabris *et al.*, Wang *et al.*, and Yurchenco & Furthmayr, Yurchenco & Ruben. This table is adapted from the Supplementary Information of M. H. J. Bailey *et al.*, *RSC Adv.*, 2020, **10**, 38275–38280.

future models. Most collagen IV networks have an average ring size of approximately six, and an average node degree of just under three.

The analysis of experimental images highlights the need for computer models of biopolymer networks like these. A computer model can provide unambiguous network structures at a larger scale than is possible to image experimentally. Furthermore, computer models can help disambiguate which of the interpretations of a network image is more likely to be correct by establishing which combinations of ring statistics are most favourable. Given that each network is changed (or possibly destroyed) when it is imaged, it is extremely hard to get dynamic experimental information on the position of polymers within a network. Dynamic models such as the Monte Carlo (MC) used in **Chapter 3** and Molecular Dynamics (MD) used in **Chapter 4** can provide information on how an individual network changes over time. The data presented in this chapter and visual analysis of biopolymer networks backs up the assumption that collagen IV networks in the lens capsule are approximately hexagonal, which will be used in the simulation methods for the rest of this thesis.

Chapter 3: Bond Switching.



This chapter details the effect of adding disorder to ordered graph-based structures. A bond-switching method is developed that introduces defects one at a time, based on a Monte Carlo (MC) approach. The MC approach allows for physically reasonable defects to be built up into a disordered network. The disordered networks thus formed are compared to experimental images, and the available range of network structures is discussed. This chapter is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].



3.1 Introduction.

Images of collagen IV networks show a great deal of disorder. The observed degree of disorder is hypothesised to increase with age. This has significant consequences for network structure — it may be that real biopolymer networks start off relatively ordered and are damaged over time.

However, microscope images of biopolymer networks are sparsely available in the literature, and tend to provide a great deal of information about only a very small portion of a given network (as discussed in **Chapter 2**). This means it is very difficult to draw statistical conclusions about the range of possible network arrangements and the frequency at which they might be generated. Instead, we can reinforce the few experimental images and mimic the

damage processes by creating a simple computational model to introduce disorder informed by the underlying chemistry of the system. There is a key balance to strike between introducing realistic physical disorder and over-imposing artificats on a complex process like network aging. Any simulation method must therefore be flexible and physically motivated. Previous computational models of biopolymer networks have been powered by graph theory approaches, especially random-network models such as Erdős-Rényi graphs, Mikado networks and bond-switching of ordered graphs[70,72,73]. Such random graph methods by their central nature introduce very little physical ordering into the formed network. This chapter will focus on bond switching as a method to introduce meaningful disorder into a structured network.

The bond switching method was first introduced by Wooten, Winer and Weaire for simulations of disorder in inorganic glasses, specifically α-Si[74,75]. Bond switching methods were designed to match the Radial Distribution Functions (RDFs) of amorphous silica and germania, and show how those RDFs were related to the RDFs of ordered silica structures. To do so, the Wooten-Winer-Weaire (WWW) method starts with a given atomic structure and permutes the bonds randomly with a probability of successful permutation dependent on energetic factors. The WWW method works by tracking the edges between nodes as well as the centres of rings ('the dual graph') to maintain consistency of the ring structure. It has been used to great effect to model interfaces between states of α-Si[76], ring networks with certain constraints such as no odd rings[77], mechanical properties of glass networks[78], and to power persistent homology (PH) studies[79] (which will themselves be revisited in **Chapter 8**).

In this chapter, an extended bond switching method similar to the WWW algorithm is shown to be effective at generating networks that are comparable to experimental examples. Then, the range of possible networks that can be formed with an extended bond switching method is shown, and limitations in comparison to experimental data are discussed.

3.2 Computational Methods.

3.2.1 Monte Carlo and Simulated Annealing.

The MC technique that powers the bond switching method is a stochastic enhanced sampling method that favours sampling regions of energetic interest. At each step, a given bond switching move (discussed in **Section 3.2.4**) is made, the positions of the atoms are optimised (discussed in **Section 3.2.3**) and the change in energy is assessed according to an imposed potential model (discussed in **Section 3.2.2**). Schematically, the algorithm performs the steps shown in **Algorithm 2**. A more detailed discussion of the mathematical basis of the MC method is available in **Section B.1**.

Algorithm 2: The algorithm describing the MC framework surrounding single bond-switching step.

for *each step* **do**
  Calculate the current energy of the network U_{old};
  Identify and perform a valid switching move to generate a new network;
  Optimise the geometry to minimise the energy of the new network U_{new};
  Generate a random number p between 0 and 1;
  if $p < \exp\left(-\frac{U_{new}-U_{old}}{k_B T}\right)$ **then**
    Accept the move and continue;
  else
    Replace the new network with the old network;
  end
end

A significant benefit of the MC approach is its flexibility in choosing moves or rejecting them. For example, Ormrod Morley & Wilson use a similar MC approach but instead of selecting moves based on an energy U, they accept or reject moves based on a cost function χ measuring the difference between the current structure and a desired structure in order to reproduce properties of interest[80]. Extending this, it is also possible to enforce arbitrary constraints on the network such as that all rings must be convex[33].

The work in this thesis focusses on two dimensional (2D) planar networks, as the ring structure is well defined. To that end, a constraint is applied to reject any moves that result in a non-planar 2D network (*i.e.* one where two edges intersect, not at a node), but accept concave rings.

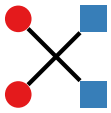


Figure 3.1: A non-planar graph that would be rejected.

A series of MC moves can be combined in a simulated annealing process, which will be used throughout this chapter. The simulated annealing process starts by performing MC moves at an effectively infinite temperature, thereby accepting all moves to 'thermalise' the starting lattice. After that, the temperature is lowered every N moves (where N is a varied parameter) from the high thermalising temperature to a near-zero temperature. The final near-zero temperature effectively rejects all moves that increase the energy.

3.2.2 Potential Model.

The underlying graph that is permuted by the bond switching model contains an explicit set of edges $\mathcal{E}$, detailing which nodes (in a set of nodes $\mathcal{V}$) are connected to which other nodes. This makes a potential model originally designed for chemical systems with explicit bonds attractive. The simplest such potential model uses harmonic bonds for the edges with an ideal equilibrium length r_{eqm}. An edge (u, v) connects two nodes u and v at positions $\boldsymbol{r}_u$ and $\boldsymbol{r}_v$ respectively in a 2D plane. The energy of bonds is then given by

$$U_{\text{bond}} = \sum_{(u,v)\in\mathcal{E}} \frac{1}{2} k_{u,v} \left|\boldsymbol{r}_u - \boldsymbol{r}_v\right|^2 , \tag{3.1}$$

where $k_{u,v}$ is the bond force constant for that specific edge. However, there are no constraints on the arrangement of edges around a given node, which can often lead to unphysical arrangements with very irregular shapes forming. The addition of an angular potential between two adjacent edges connected to a specific node allows for control over how regular

rings are, in the form

$$U_{\text{angle}} = \sum_{(u,v)\in\mathcal{E},(u,w)\in\mathcal{E}} \frac{1}{2} K_{u,v,w}(\cos(\theta_{u,v,w}) - \cos(\theta_{\text{eqm}}))^2, \tag{3.2}$$

where $\theta_{u,v,w}$ is the angle between edges (u, v) and (u, w), $K_{u,v,w}$ is the angular force constant, and θ_{eqm} is the equilibrium angle. This is sometimes referred to as a 'simplifed Keating potential', and has been used effectively to model amorphous silica bilayers[33,81]. It is also very similar to a form of the DREIDING potential used for molecular simulation which will be revisited in **Chapter 4**[82].

Finally the constraints mentioned in the previous section can be included in the potential model:

$$U_{\text{planarity}} = \begin{cases} 0 & \text{if} \quad \text{no edges intersect} \\ \\ \infty & \text{otherwise} \end{cases} \tag{3.3}$$

The total energy is then given by

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{planarity}}. \tag{3.4}$$

3.2.3 Geometry Optimisation.

After a bond switching move is completed, a network can often end up in an extremely high energy state, where edges are stretched or compressed considerably past their equilibrium length and the formed rings are extremely irregular. This can be resolved by optimising the geometry around the bond switching site. In this thesis, a line search steepest descent method is used for geometry optimisation. A vector $\boldsymbol{R}$ is constructed with elements

$$\boldsymbol{R} = \begin{bmatrix} r_{1x} & r_{1y} & r_{2x} & \dots & r_{ny} \end{bmatrix}.$$

The energy as a function of this vector is evaluated $U(\boldsymbol{R})$, and the gradient of that vector $\nabla U(\boldsymbol{R})$ is evaluated. Then, the line described by $-\nabla U(\boldsymbol{R})$ is searched in small steps of size λ until a turning point is found. This process is repeated over multiple iterations under until

convergence is achieved as documented in **Algorithm 3**.

Algorithm 3: The steepest descent algorithm used in the MC simulations.

```
for i < i_max do
    G = ∇U(R);
    while |U(R_new) − U(R_old)| < tolerance do
        R_new = R_old − λG;
```

However, optimising the geometry for the entire network after every step can be extremely computationally wasteful with a local potential model like the one used here. On relaxation, the node positions far from the bond switching site will remain mostly unchanged, whereas the node positions close to the site will need to change significantly. To reduce this computational strain, it more efficient to only optimise positions of nodes within m bonds of the bond switching site (here chosen to be $m = 5$). The rest of the node positions are frozen and their interactions ignored. Periodically, to avoid build-up of numerical error, the geometry of the entire system can is re-optimised with no nodes frozen.

3.2.4 Bond Switching.

The choice of bond switching moves is critical as they can control the overall pattern of disorder. There are two different bond switching moves used in this thesis: the Wooten-Winer-Weaire approach, and the extended bond switching approach.

Wooten-Winer-Weaire.

Traditionally, bond switching moves are chosen to preserve the degree of all nodes before and after the move, as this is well suited to inorganic chemical networks with strict coordination rules. This is known as the Wooten-Winer-Weaire method. A WWW bond switching move is performed by selecting a random edge (u, v) which is part of two rings, $\mathcal{A}$ and $\mathcal{B}$. A neighbour a of v that is part of ring $\mathcal{A}$ is chosen, and then a neighbour b of u that is part of ring $\mathcal{B}$ is

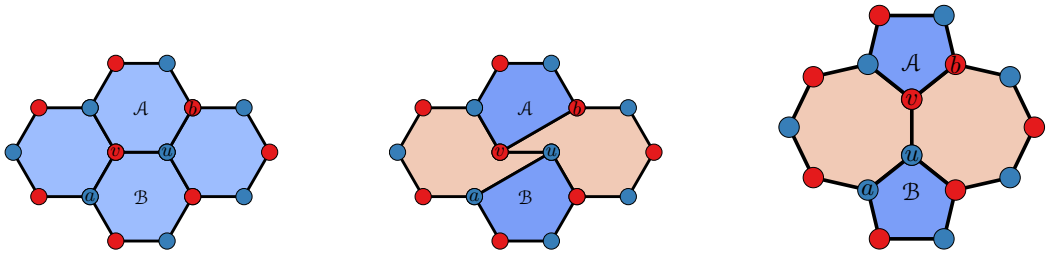


Figure 3.2: Schematic WWW bond switch for a simple ring arrangement. The central figure shows the network after the switch but before the positions are optimised. The nodes are moved along their lines to force to an energetic minimum in the rightmost figure.

chosen. The edges (a, v) and (b, u) are deleted, and two new edges (a, u) and (b, v) are added. This is shown in **Figure 3.2**.

Extended Bond Switching.

For semiflexible biopolymer moves, the WWW approach is inappropriate as it does not allow for variable degree. Instead, by selecting bond switching moves that preserve the average degree $\langle k \rangle$ individual nodes can gain or lose edges to represent polymers disconnecting or reconnecting from one another. An bond switching move is performed by similarly selecting a random edge (u, v) which is part of two rings, $\mathcal{A}$ and $\mathcal{B}$. This is shown in **Figure 3.3**. Next, one of the nodes forming that edge is chosen — for example, u. The neighbours of v that participate in either of the rings $\mathcal{A}$ or $\mathcal{B}$ are assessed and labelled a, b. Next, the edge (u, v) is deleted and a new edge is drawn between u and a randomly chosen neighbour of v that participates in one of the rings, either (u, a) or (u, b). This process decreases the degree of u by 1 and increases the degree of either a or b by 1. The rings are then updated, with one ring becoming one edge larger (a hexagon becomes a heptagon) and one ring becomes one edge smaller (a hexagon becomes a pentagon).

MC rules can then reject particularly unphysical moves that were randomly selected. In this chapter, only moves that creates notes with degree 2 to 8 and rings with 3 to 20 edges were accepted, unless otherwise stated.

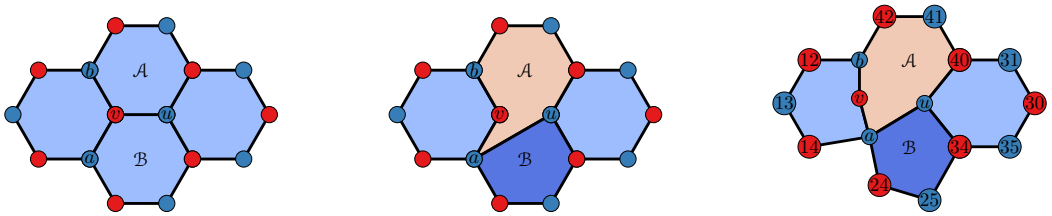


Figure 3.3: Schematic bond switch of a simple network. The central figure shows the network after the switch but before the positions are optimised. The nodes are moved along their lines to force to an energetic minimum in the rightmost figure.

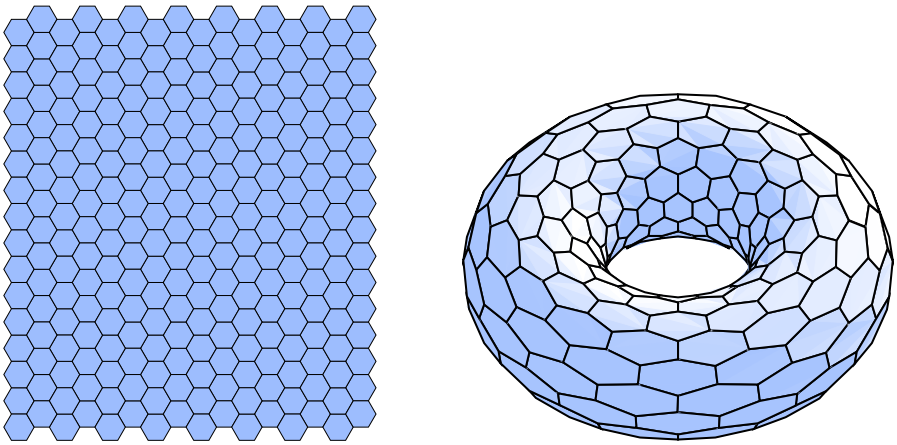


Figure 3.4: A periodic network shown in the plane, where the top edge connects to the left edge, and the same network shown wrapped around onto a torus.

By constraining the mean degree $\langle k \rangle$, the mean number of edges per ring is also constrained by Euler's equation for polyhedra

$$V - E + F = \chi, \tag{3.5}$$

where V is the number of nodes (referred to often as vertices), E is the number of edges, F is the number of rings (representing faces if the network were drawn as a polyhedron) and χ the Euler characteristic for the surface. In 3D space, $\chi = 2$ but a periodic 2D plane can be represented as a torus with $\chi = 0$ — this is shown in **Figure 3.4**.

With this rule, if the mean degree is constrained to $\langle k \rangle = 3$ the mean number of edges per ring is constrained to be $\langle n \rangle = 6$. This does not hold for networks imaged experimentally, which are of finite size and are aperiodic. However, the effect of this is small: in an aperiodic arrangement of 100 rings with $\langle k \rangle = 3$ there is an average number of edges per ring of $\langle n \rangle =$

5.94.

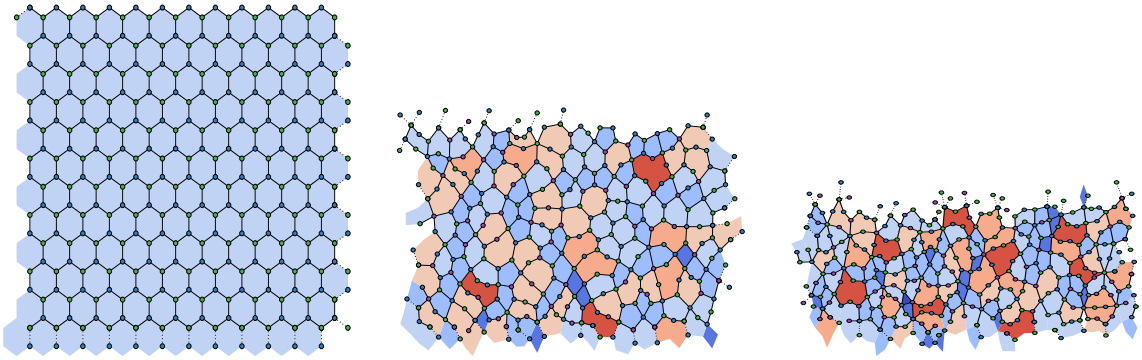
3.2.5 Bond Switching and Stretching.

The bond switching method detailed here can be useful to examine what sorts of defects are physically likely under a variety of conditions. One set of conditions that is of interest is whether networks under stretching favour different defects to networks that are not stretched. Stretching can be modelled during a bond switching simulation by regularly resizing the periodic simulation cell and rescaling the positions of all nodes within it. A stretching procedure is shown graphically in **Figure 3.5**. It is not feasible to extract meaningful mechanical properties from the bond switching simulations — after each step, the system is at an energy minimum thanks to the geometry optimisation which relieves stress. Additionally, the simulation protocol must be carefully designed to avoid biasing to favour certain types of MC moves. If the system is rescaled too often, then its evolution becomes dominated by the rescaling moves and not by the (more interesting) bond switching moves. For this reason, rescaling occurs every N bond switching moves where N is the number of rings in the simulation, such that on average one bond switch is attempted per ring before rescaling. Rescaling without preserving the volume favours bond switching moves that create rings with many edges which are less stressed for a given area. Because of this, the simulation cell is rescaled in x by a small factor λ_x and scale it in y by a factor $\lambda_y = \frac{1}{\lambda_x}$.

3.3 Graph Classifications and Metrics.

3.3.1 Node Degree Distributions.

Scale-free networks are a keenly discussed topic in network theory, as mentioned in **Section 1.2.3**. A scale-free network is one where the distribution of degrees k is well described



(a) A hexagonal network, where any bond switching move is enthalpically unfavourable.

(b) The same network at a stretching ratio of $\lambda^{10} \approx 1.41$

(c) The same network at a stretching ratio of $\lambda^{20} = 2.0$

Figure 3.5: Example networks of 144 hexagons generated by bond switching with regular rescaling every 144 steps. These networks were scaled to double the length of the simulation cell in x while halving the length in y.

by the equation

$$P(k) = k^{-\gamma}, \tag{3.6}$$

where γ is a constant per-network. Barabási & Albert found that this was a good match for networks of webpages ($\gamma_{\text{web}} \approx 2.1$), scientific citations ($\gamma_{\text{cite}} = 3$) or power grids ($\gamma_{\text{power}} = 4$)[53].

However, the idea of fitting a function to degree distributions of graphs is not without controversy. Recently, there has been much debate in the literature and online arguing either that 'scale-free networks are rare'[83], or that scale-free networks are both 'rare and everywhere'[84]. The cause of the dispute is that scale-free networks often observe small ranges of k for extremely complex finite size networks, which leads to ambiguity in defining what an edge is (for example, in social networks an edge is extremely hard to define), or statistical fluctuations.

3.3.2 Maximum Entropy.

In a discussion of MC simulation, it would be remiss to not include the effects of entropy. With randomness being a major factor in the simulation, it would be reasonable for the

distribution of some property (for example the distribution of degree k or edges per ring n) to be the distribution that maximises the entropy of the system, given some constraints. This approach was originally proposed by Rivier & Lissowski to demonstrate Lewis' law of areas per cell, and extended to distributions of edges per ring by Gervois *et al.*[85,86]. For a distribution of edges per ring, n, we wish to maximise the entropy

$$S = -\sum_n p_n \ln p_n, \tag{3.7}$$

according to the three constraints

$$\sum_n p_n = 1 \tag{3.8}$$

$$\sum_n n p_n = \langle n \rangle \tag{3.9}$$

$$\sum_n \frac{p_n}{n} = c. \tag{3.10}$$

The first two constraints arise naturally from the definition of p_n and Euler's formula, while the third is somewhat more mysterious. Gervois *et al.* propose that it arises from the area of each ring being distributed according to the equation $A_n = an + b + \frac{c}{n}$ where a, b and c were all constants[86]. This runs directly counter to Lewis' law, although the only comment given in the original maximum entropy formulation was 'Let us assume that no further constraint exists'[86]. We would also not expect a single set of maximum entropy constraints to describe both node distributions and ring size distributions well. Instead, Ormrod Morley shows that the third constraint can be derived from the assortativity of the network[33].

3.3.3 Regularity Criterion.

One further complicating factor is that biological networks are typically constructed from significantly irregular rings, which is a key difference from typical inorganic chemical networks. Inorganic glasses demonstrate a preference for convex polygonal structures owing to the

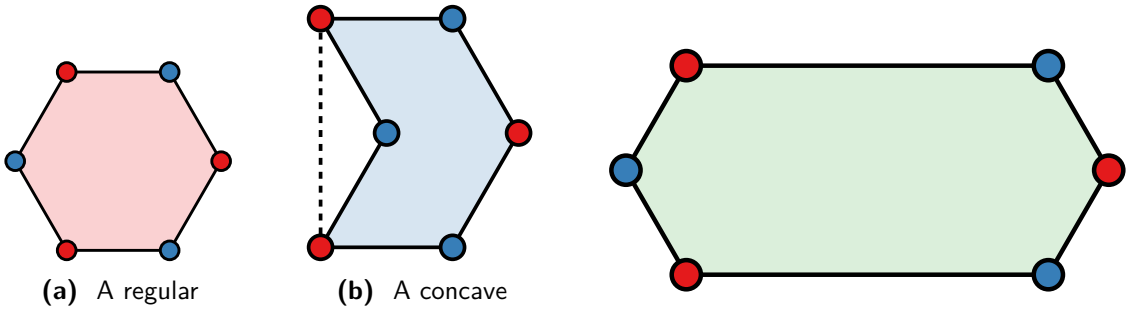
strong energetic penalty for irregular angles and fixed degree. Polymer networks, however, can have smaller energetic penalties for irregular angles, variable degrees, and even curved edges, leading to rings that are more likely to be concave or otherwise distorted. There are a number of metrics that have been used historically to quantify the irregularity of rings, relating a ring to its convex hull, the polygonality, and the eutaxy[87,88] These metrics are often troublesome as they do not vary dramatically — a rectangle will have the same perimeter and area as its convex hull, but be very different from a square. To rectify that, work in this thesis employs the Shape Regularity Criterion (SRC) as used by Giraud *et al.*:

$$SRC(\mathcal{R}) = \frac{P(\mathcal{CH})}{P(\mathcal{R})} \times \frac{A(\mathcal{R})}{A(\mathcal{R})} \times \sqrt{\frac{\max(\sigma_x, \sigma_y)}{\min(\sigma_x, \sigma_y)}} \tag{3.11}$$

in which $\mathcal{R}$ is the shape, $\mathcal{CH}$ is its convex hull, P is the perimeter, A is the area and σ_x, σ_y are the standard deviation of vertex x and y coordinates in the plane[89]. Regular shapes (such as squares, equilateral triangles and regular hexagons) are characterised by $SRC = 1.00$. Less-regular shapes (such as ellipses or rectangles) have $0 \leq SRC < 1$. Some example shapes are shown in **Figure 3.6** demonstrating how the metrics combine to give a useful overall regularity criterion.

3.3.4 Assortativity.

As initially discussed in **Section 1.1.3**, the assortativity is a useful metric to measure medium-range structure. This was originally introduced by Newman to study social networks, computer networks and biological networks, which have nodes with a high degree often connected to other nodes with a high degree[32]. Newman used a correlation coefficient r to characterise this preference for a variety of networks. Edges in graphs connect two nodes, so the probability of a node with degree j being connected to a node with degree k is described by the joint probability distribution e_{jk}[90]. This obeys the normalisation rules $\sum_{j,k} e_{jk} = 1$



(a) A regular hexagon.

$\frac{P(\mathcal{CH})}{P(\mathcal{R})}$	1.00
$\frac{A(\mathcal{R})}{A(\mathcal{R})}$	1.00
$\sqrt{\frac{\min(\sigma_x,\sigma_y)}{\max(\sigma_x,\sigma_y)}}$	1.00
SRC	1.00

(b) A concave hexagon.

$\frac{P(\mathcal{CH})}{P(\mathcal{R})}$	0.955
$\frac{A(\mathcal{R})}{A(\mathcal{R})}$	0.800
$\sqrt{\frac{\min(\sigma_x,\sigma_y)}{\max(\sigma_x,\sigma_y)}}$	0.884
SRC	0.676

(c) An elongated hexagon.

$\frac{P(\mathcal{CH})}{P(\mathcal{R})}$	1.00
$\frac{A(\mathcal{R})}{A(\mathcal{R})}$	1.00
$\sqrt{\frac{\min(\sigma_x,\sigma_y)}{\max(\sigma_x,\sigma_y)}}$	0.648
SRC	0.648

Figure 3.6: Three hexagons of different shapes to show distinct regularity metrics. The dashed line in **Figure 3.6b** represents the convex hull. The elongated hexagon is non-regular as it has a low balanced repartition, whereas the concave hexagon is non-regular due to a combination of a smaller area than its convex hull and its balanced repartition.

and $\sum_j e_{jk} = q_k$, where q_k is the probability of a node having 'remaining degree' (the number of edges leaving the node that was just connected to) k.

If the edges had no preference in their attachments and were purely random, the joint degree would have the value $e_{jk} = q_j q_k$. It is then possible to characterise the difference from the purely random distribution with a correlation function $\langle jk\rangle - \langle j\rangle \langle k\rangle = \sum_{j,k} jk(e_{jk} - q_j q_k)$. The correlation function can take a range of values depending on the maximum degree found in the graph, so it is helpful to scale the correlation function by its maximum value. This maximum value is conveniently the variance of the degree distribution σ_q^2, leading to a Pearson-style correlation coefficient:

$$r = \frac{1}{\sigma_q^2} \sum_{j,k} jk(e_{jk} - q_j q_k). \tag{3.12}$$

In an assortative network, nodes with high degree are likely to be connected to other nodes with high degree. The differences in the joint degree distributions are shown in **Figure 3.7** for two types of random graphs, one of which is slightly assortative and the other is slightly disassortative.

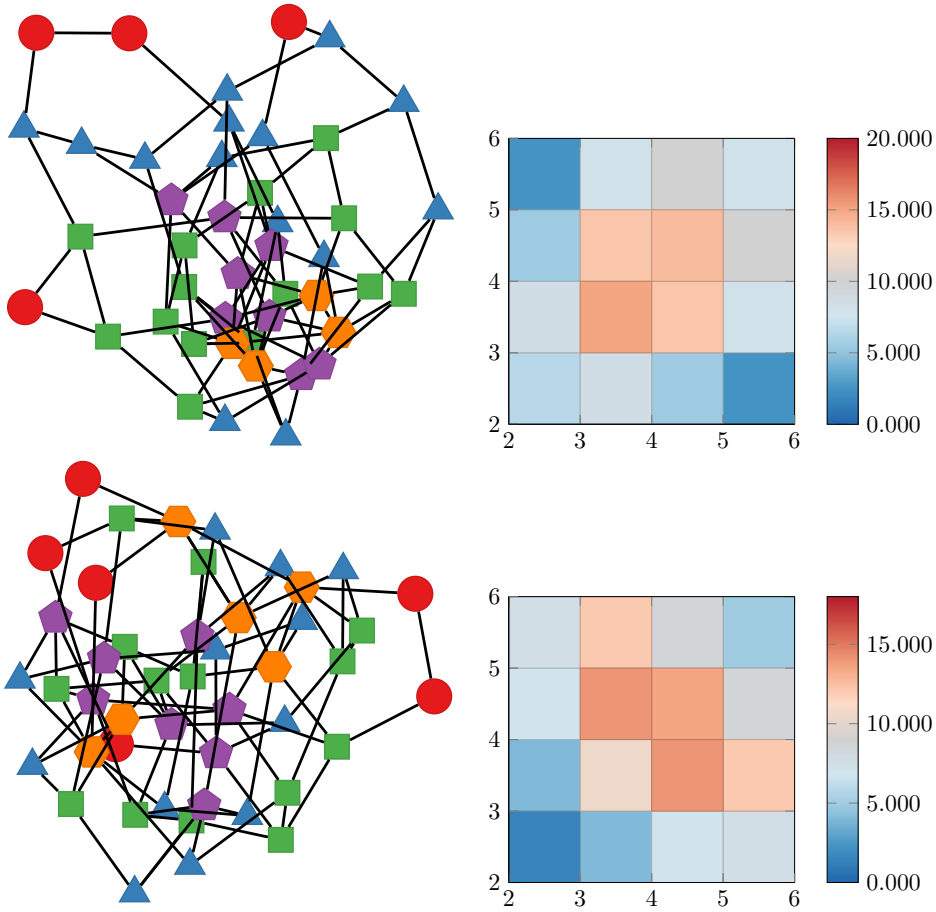


Figure 3.7: Two graphs with a maximum entropy degree distribution but different assortativities. Their joint degree distributions are shown as heatmaps, indicating the number of connections between nodes of degree k and nodes of degree l. The top graph has $r = 0.300$, which shows as nodes of the same colour clustering together, and much of the joint degree distribution lying along the diagonal. The bottom graph has $r = -0.300$, which shows as nodes of the same colour avoiding one another, and much of the joint degree distribution lying off the diagonal.

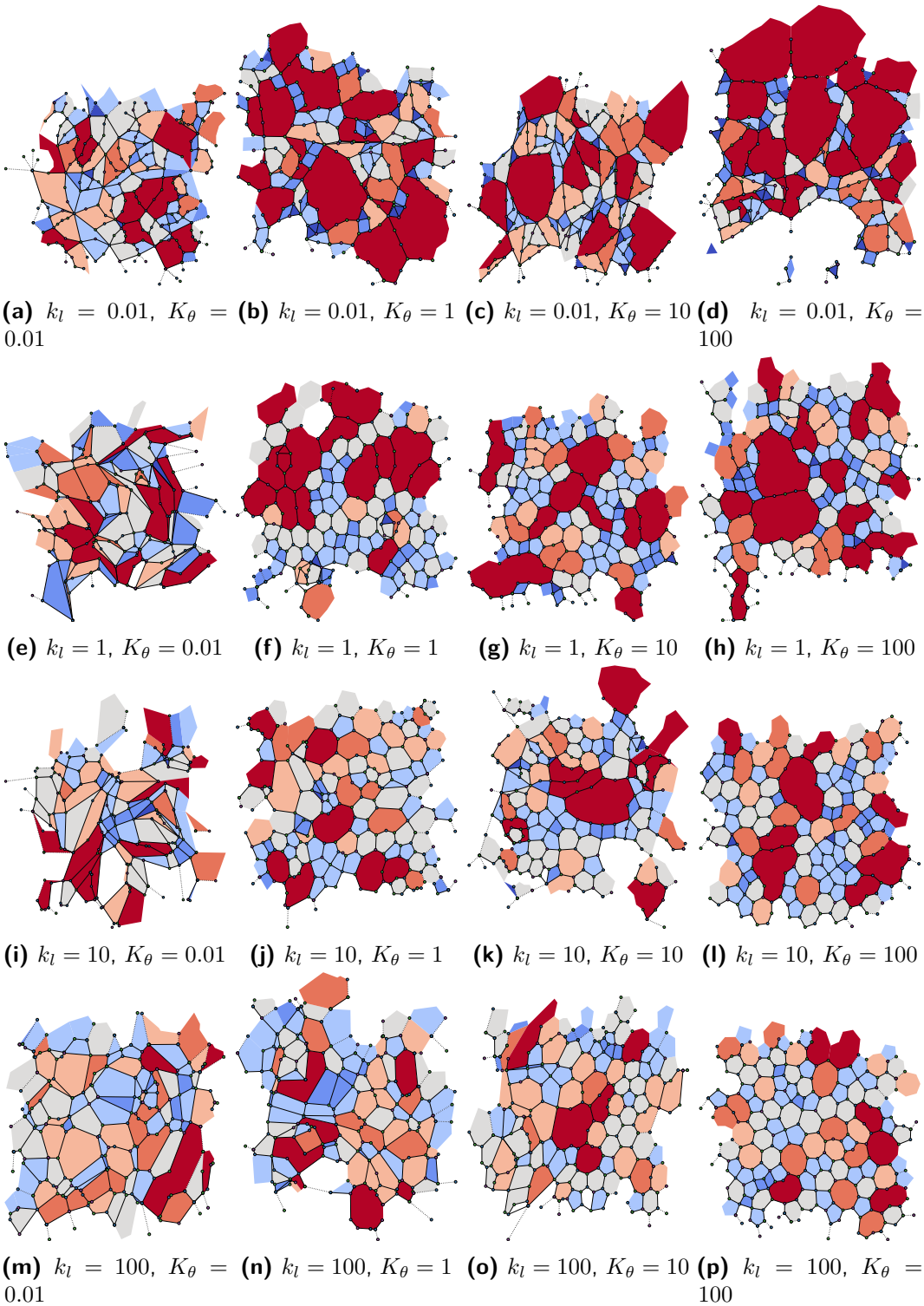
3.4 Results and Discussion.

3.4.1 Generated Networks.

A useful way to understand the properties of networks that can be formed by this extended bond switching technique is simply to generate many networks with a variety of parameters. Parameters from the potential model in **Section 3.2.2** provide an easy entry point to this. The balance of bond strength k_l, angular strength K_θ and MC temperatures T affect the overall structure of the network. Over a simulated annealing simulation the temperature will vary, so the bond and angular strength parameters are sorted into 'weak' (~ 0.0100 arb. units), 'medium' (1.00 arb. units to 10.0 arb. units) or 'strong' (~ 100 arb. units) regimes. These are presented in reduced units for ease of comparison; for example, a move attempted at $T = 1.00$ arb. units that results in an energy change of $\Delta U = 1.00$ arb. units has an acceptance probability of $1/e$. Final network structures can be explored by combining bond and angular strength parameters from each of these regimes.

Figure 3.8 shows a range of networks generated by the bond switching algorithm with different bond strength k_l (established in **Equation 3.1**) and angular strength K_θ parameters (established in **Equation 3.2**).

In these images, we can observe that systems with low angle force constants (the first column) the resultant networks feature large holes and are visually extremely irregular. This is especially pronounced for the networks in the lower left where k_l is much larger than K_θ, where angular distortions lead to extremely long or narrow shapes to avoid distorting bonds. If the opposite happens and K_θ is much larger than k_l, we can observe concave shapes with edges of different lengths; the loss of convexity and stretching bonds allows the system to keep angles around their equilibrium value. Systems where k_l and K_θ are both strong (the bottom



(a) $k_l = 0.01,\ K_\theta = 0.01$ **(b)** $k_l = 0.01,\ K_\theta = 1$ **(c)** $k_l = 0.01,\ K_\theta = 10$ **(d)** $k_l = 0.01,\ K_\theta = 100$

(e) $k_l = 1,\ K_\theta = 0.01$ **(f)** $k_l = 1,\ K_\theta = 1$ **(g)** $k_l = 1,\ K_\theta = 10$ **(h)** $k_l = 1,\ K_\theta = 100$

(i) $k_l = 10,\ K_\theta = 0.01$ **(j)** $k_l = 10,\ K_\theta = 1$ **(k)** $k_l = 10,\ K_\theta = 10$ **(l)** $k_l = 10,\ K_\theta = 100$

(m) $k_l = 100,\ K_\theta = 0.01$ **(n)** $k_l = 100,\ K_\theta = 1$ **(o)** $k_l = 100,\ K_\theta = 10$ **(p)** $k_l = 100,\ K_\theta = 100$

Figure 3.8: Many images of networks generated under the same simulated annealing scheme but with different potential model parameters.

Image	$\langle k \rangle$	$\mu_2(k)$	$\langle n \rangle$	r	N
Barnard *et al.*[58]	2.961	0.378	5.574	-0.109	135
Bos *et al.*[55]	2.780	0.335	5.825	-0.209	40
Wang *et al.*[48]	2.945	0.805	5.735	-0.057	498
Yurchenco & Furthmayr[17]	2.721	0.284	6.254	-0.181	61
Yurchenco & Furthmayr[17]	2.667	0.287	6.593	-0.184	54
$k_l = 0.01$, $K_\theta = 0.01$	3.000	1.440	6.000	-0.012	100
$k_l = 0.01$, $K_\theta = 0.01$	3.000	1.440	6.000	-0.012	100
$k_l = 10$, $K_\theta = 0.01$	3.000	1.660	6.000	-0.004	100
$k_l = 100$, $K_\theta = 100$	3.000	1.000	6.000	-0.037	100

Table 3.1: Statistics collected for a number of images of biological networks and simulations showing the mean node degree, $\langle k \rangle$, the second moment, $\mu_2(k)$, the mean number of ring edges, $\langle n \rangle$, the number of rings in each configuration, N, and the assortativity, r. Two different analyses of the networks imaged by Yurchenco & Furthmayr are shown in order to highlight the sensitivity of the obtained metrics. A more detailed table is available in **Chapter 2**.

right corner) appear more regular with fewer extremely large rings. This is due to a reduced influence of entropy in the melting phase (as the fixed thermal energy is less significant for stronger values of k_l and K_θ) and a greater drive back towards the energetic minimum of the ordered network in the cooling phase of the annealing process.

3.4.2 Comparison to Experiment.

While a visual analysis to show the range of networks is useful, it is most useful when compared to experimental data. An extract of the full experimental data table in **Table 2.1** is shown in **Table 3.1** and compared to ring statistics for four simulated networks. The simulated networks all show $\langle k \rangle = 3$ and $\langle n \rangle = 6$ as a consequence of Euler's formula in a periodic system, but the finite sized images from experiments show smaller $\langle k \rangle$ due to missing edges around the perimeter. Generally, the simulated networks show a larger value of $\mu_2(k)$ than the experimental images. This is partly because the image analysis process is unlikely to assign nodes with high degree and will sometimes split them into two $k = 3$ nodes. All networks in that table are disassortative, but the simulated networks are closer to a random ordering because of the tendency of MC processes to favour a maximum entropy system.

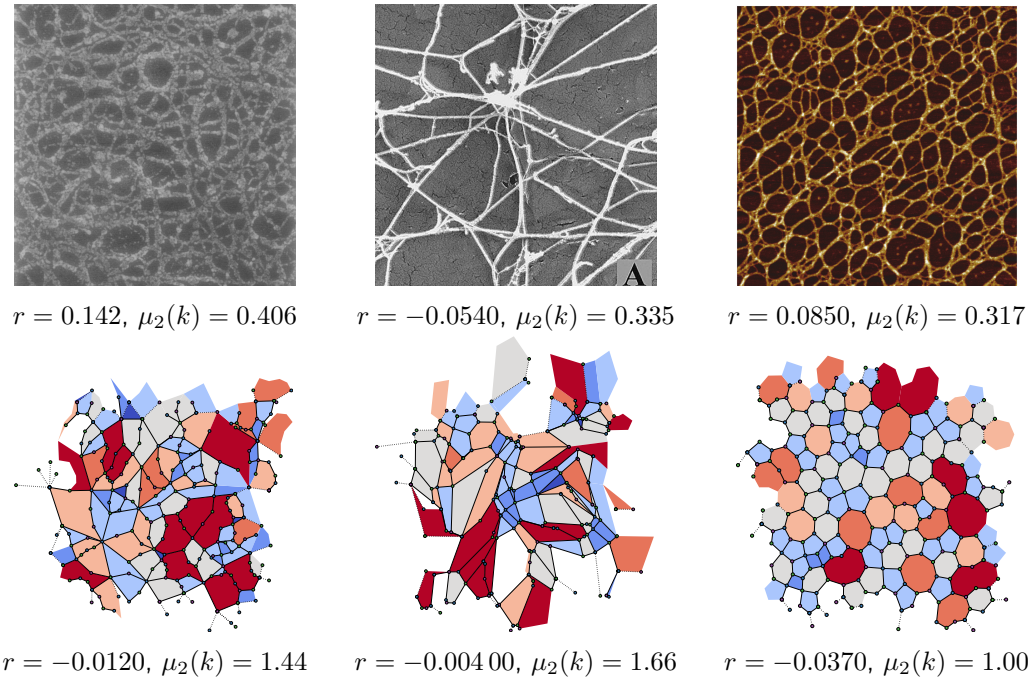
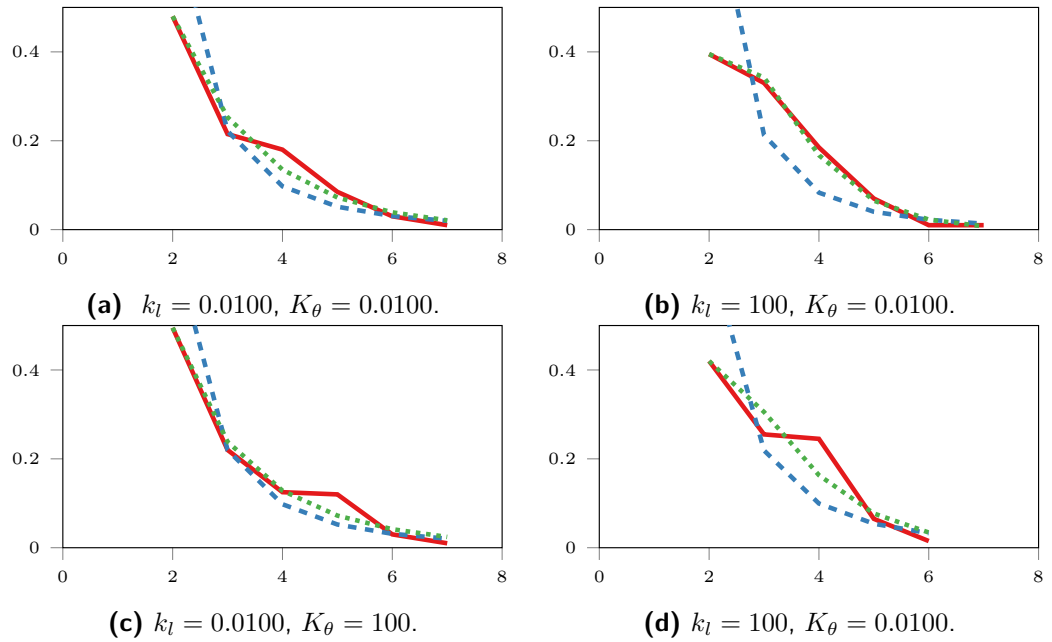


Figure 3.9: Images of biological networks from (left to right) Barnard *et al.*, Bos *et al.*, and Wang *et al.* respectively. The lower panels show networks generated by our method that appear similar, from left to right with $\{k_l, K_\theta\} = \{0.01, 0.01\}$, $\{10, 0.01\}$ and $\{100, 100\}$. Reproduced from from Barnard *et al.* and Bos *et al.*, Copyright (1992) and (2001), with permission from Elsevier. Reproduced from from Wang *et al.*, Copyright (2017) with permission. This figure is adapted from M. H. J. Bailey *et al.*, *RSC Adv.*, 2020, **10**, 38275–38280[70].

A graphical comparison with selected experimental networks with simulated networks is shown in **Figure 3.9**. The simulated networks were chosen to look qualitatively similar to the accompanying experimental networks. However, the structural metrics are significantly different despite the observed networks appearing similar. This suggests that a more rigorous comparison method based on structural metrics is needed.

3.4.3 Scale-Free *vs* Maximum Entropy.

The applicability of a scale-free distribution to collagen-like networks was assessed by running MC simulations that introduced disorder into a hexagonal network at varying bond strength and angle strength parameters (themselves discussed in **Section 3.2.2**). Each network was initially hexagonal, and was thermalised at a temperature of 10^2 arb. units for 1000 steps



(a) $k_l = 0.0100$, $K_\theta = 0.0100$. **(b)** $k_l = 100$, $K_\theta = 0.0100$.

(c) $k_l = 0.0100$, $K_\theta = 100$. **(d)** $k_l = 100$, $K_\theta = 0.0100$.

Figure 3.10: Distributions of degree k for networks with varied k_l and K_θ parameters under the same simulated annealing process. These green dotted line is the maximum entropy fit, and blue dashed line is the scale-free fit. The networks analysed are shown graphically in **Figure 3.8**.

to remove the memory of the original form. The thermalisation was assessed both visually and by observing that network metrics (such as energy) equilibrated. Next, the networks were cooled down to a temperature of 10^{-2} arb. units at logarithmically spaced temperature increments of $10^{-0.2}$ arb. units, with 250 steps taken at each temperature increment. The final degree statistics for both the networks and the dual network were analysed, and the best scale-free fit was found.

Figure 3.10 shows the results of this analysis, comparing the distributions of degree k to both a maximum entropy fit and a scale-free fit. We can see that even in the high entropy case, the scale-free fit significantly overestimates the number of nodes with small k — furthermore, if the range of allowed k values were larger, it would also overestimate the number of nodes with large k as it drops off polynomially, not exponentially. The maximum entropy fit is significantly better, especially for networks with low values of k_l and K_θ (as there is a smaller enthalpic contribution to their structure at a given T). This reinforces the

hypothesis that biopolymer networks generated by this method are not scale-free but are instead better described by a maximum entropy distribution.

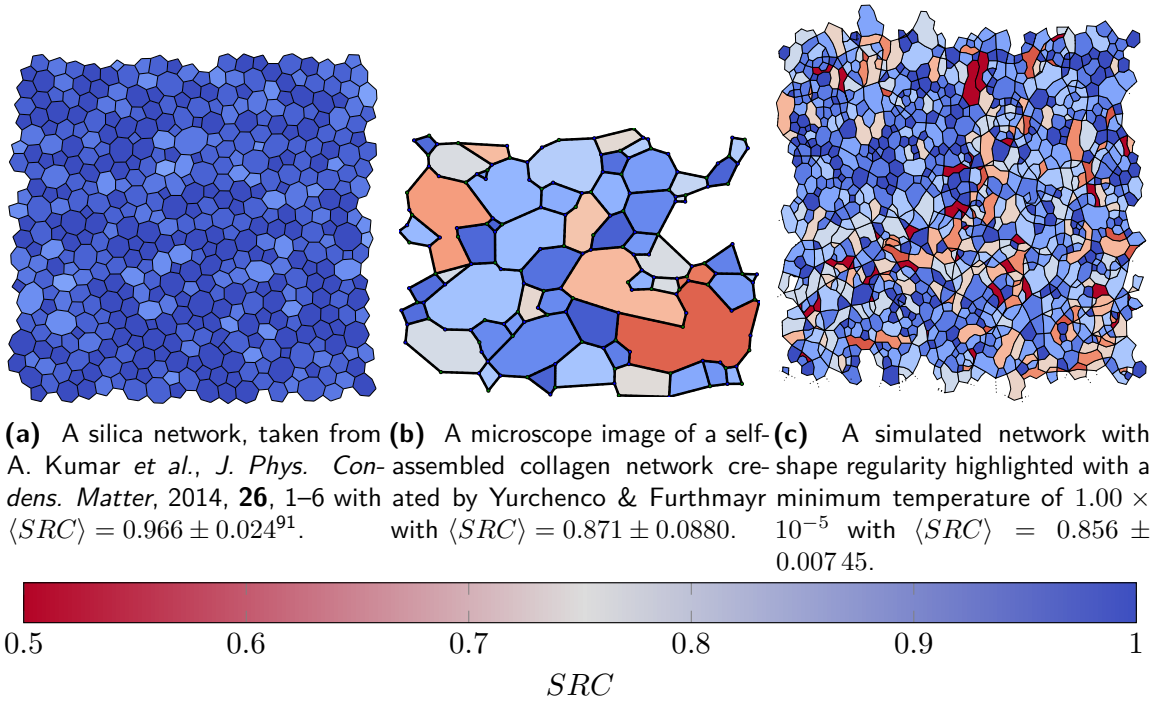
Given the constraint that $\langle k \rangle = 3$, the maximum entropy distribution leads to many nodes with $k = 2$. This is physically because it is easy to make a $k = 2$ node using the extended bond switching procedure, and hard to get rid of it: it can only gain an edge from a neighbour, whereas nodes with any other k can gain or lose edges. The proliferation of nodes with $k = 2$ has consequences for structure, as they tend to group together to reduce angular strain; this results in rings with many edges forming and demixing. Demixing of large rings can be seen in **Figure 3.8d** where many large red rings gather together in the top right corner.

3.4.4 Regularity Criteria.

As discussed in **Section 3.3.3**, the biopolymer networks that are simulated in this thesis are considerably more flexible and concave than the inorganic networks traditionally modelled using computational chemistry techniques.

Regularity criteria can also be calculated for experimental networks regardless of their origin as long as a ring structure can be determined. **Figure 3.11** shows a comparison of an inorganic network highlighted by regularity with a experimental biopolymer network and a simulated ring network. The silica network, shown in **Figure 3.11a**[91], demonstrates the tendency of inorganic glasses to be extremely regular as there is a strong energetic penalty to moving silicon or oxygen atoms away from their regular positions. This results in a SRC value of 0.966 ± 0.024, which is both regular and shows a narrow distribution of regularities (a small standard deviation).

The example biological network shown in **Figure 3.11b** is much more irregular owing to the greater flexibility of the polymers, with $\langle SRC \rangle = 0.871 \pm 0.0880$ — both with a lower



(a) A silica network, taken from A. Kumar *et al.*, *J. Phys. Condens. Matter*, 2014, **26**, 1–6 with $\langle SRC \rangle = 0.966 \pm 0.024$[91]. **(b)** A microscope image of a self-assembled collagen network created by Yurchenco & Furthmayr with $\langle SRC \rangle = 0.871 \pm 0.0880$. **(c)** A simulated network with shape regularity highlighted with a minimum temperature of 1.00×10^{-5} with $\langle SRC \rangle = 0.856 \pm 0.007\,45$.

Figure 3.11: A comparison of an inorganic network, an experimentally imaged network and a bond switched network. The colour ramp of red to blue represents the range of irregular ($SRC = 0.5$) to regular ($SRC = 1$) shapes. Panels of this image are reproduced from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

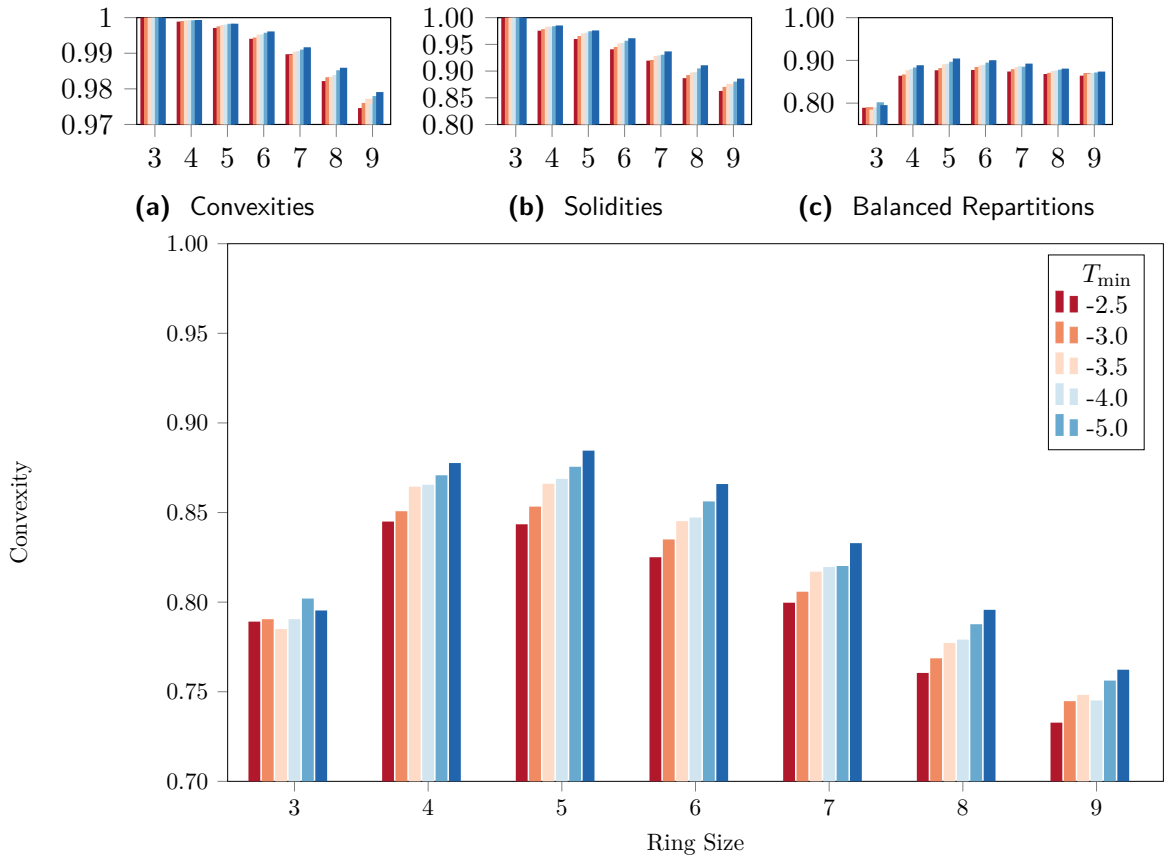
SRC and a greater standard deviation. This is matched well by the simulated network in **Figure 3.11c** which has a similar $\langle SRC \rangle = 0.856 \pm 0.007\,45$ and a qualitative match with inclusions of irregular red shapes.

The lower SRC values in biopolymer networks compared to inorganic networks are also affected by the existence of nodes with degree $k = 2$ which can lead to complex concave shapes. These are seen both in **Figure 3.11b** and **Figure 3.11c**, whereas all rings in **Figure 3.11a** are convex and very regular. A full table of data showing SRC values is in **Section C.1** showing that experimental images have SRC values in the range 0.693 to 0.966.

Next, we can examine the effect of simulation parameters on SRC to understand how it is linked to the physical properties of the system. A series of networks were created using the simulated annealing process described in **Section 3.2.1** but running to a lower minimum annealed temperature T_{min}. The three individual components of the SRC for networks

generated with this protocol are shown in **Figure 3.12a**, **Figure 3.12b** and **Figure 3.12c**. Each bar represents the average value of a given metric for 25 repeated simulations with different random seeds of 1024 rings. The SRC, created as a combination of metrics, is shown in **Figure 3.12d** at a variety of different ring sizes and temperatures. We can observe a few overall trends by looking at the components:

- Firstly, the convexity is in a small numerical range but with a pronounced trend of 0.970 to 1.00. The cooler the final temperature, the more convex the resultant rings are.
- Solidity follows a similar trend, but over a considerably wider range of 0.850 to 1.00.
- The balanced repartition however is independent for all temperatures. It is not 1 for any shape, indicating that even the triangles in the network are rarely equilateral. Rings with more edges tend to show flatter trends with T as they more likely to be approximately circular.
- The general trend is for all regularity metrics to decrease for rings with more edges. This is due to the greater flexibility of a larger ring which allows more distortion owing to the existence of more degrees of freedom. A triangle is necessarily its own convex hull, but this is not necessarily true for a quadrilateral.
- The SRC combining these factors shows a general trend of cooler final temperatures leading to more regular shapes as the influence of enthalpy is more important. This is not true for triangles which have convexity and solidity both = 1, and therefore have their regularity dominated by the balanced repartition, which is independent of temperature.



(a) Convexities **(b)** Solidities **(c)** Balanced Repartitions

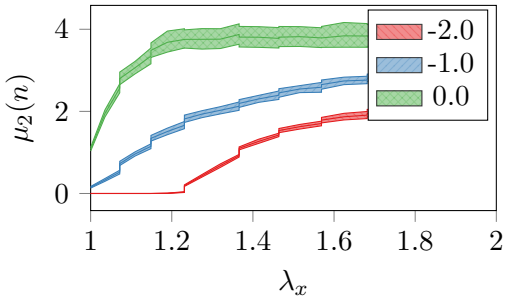
(d) Combined SRC statistics.

Figure 3.12: Components of SRC and the overall metric for rings of different sizes. Panels **(a)** to **(c)** have difference y-scales. Each bar represents the average metric across all rings with a given number of members across 20 simulations. The colouration of the bar shows the minimum temperature of the simulated annealing process.

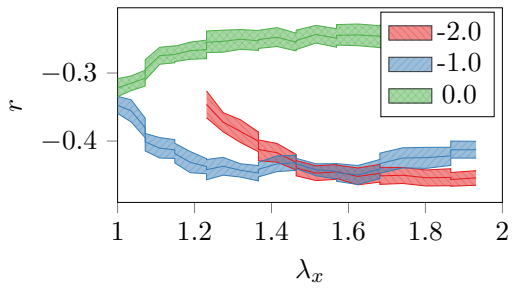
3.4.5 Bond Switching and Stretching.

The defects that form in an unstretched network may be different to the defects formed in a network under tension. For example, a stretched network may favour anisotropic rings that are stretched themselves. This would favour larger rings, which require less energy to deform.

To test that, the bond switching procedure was repeated and the network was stretched by regularly rescaling the simulation cell in one direction. A system of 144 hexagons undergoing bond switching was rescaled every 144 attempted moves by a factor of $\lambda = \sqrt[20]{2} \approx 1.04$ (such that the scaled axes doubled in length at the end of the simulation).



(a) The variance of the number of edges of rings in the network. When more rings with many or few ages appear $\mu_2(n)$ increases. When all rings are hexagons, $n = 6$ and $\mu_2(n) = 0$.



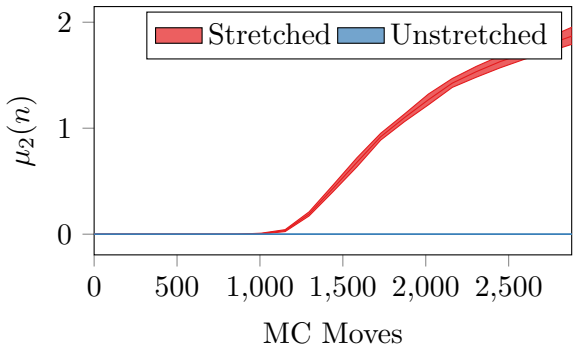
(b) The assortativity of rings in the network. When all rings are hexagons, r is undefined.

Figure 3.13: Formation of defects over the course of a bond switching simulation with stretching. With no stretching, all rings remain as hexagons. The filled area represents one standard deviation of the metric across the repeated simulations.

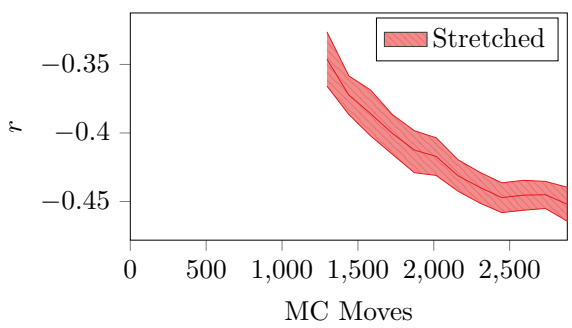
The introduction of defects due to the bond switching process is shown in **Figure 3.13**. **Figure 3.13a** shows the range of different ring metrics (represented by $\mu_2(n)$, the second moment of the ring edge count distribution) against the stretching ratio. In the range λ =1.00 to 1.20 for the lowest temperature $T = 1.00 \times 10^{-2}$, $\mu_2(n) = 0$ as the network is entirely made of hexagons, as was seen earlier in **Figure 3.5a**. As the stretching ratio increases, the range of different rings gets broader, represented by a larger $\mu_2(n)$ and more varied network appearances as shown in **Figure 3.5b** and **Figure 3.5c**. This effect is less significant as the temperature increases, and at higher temperatures the distribution rapidly approaches the maximum entropy limit.

The high temperature stretched networks in **Figure 3.13b** become less disassortative (a less strong tendency for dissimilar rings to neighbour one another) as more moves are made but low temperature networks become more disassortative. Usually this would be because of dilute defects, but visual inspection of the networks shows this is not the case.

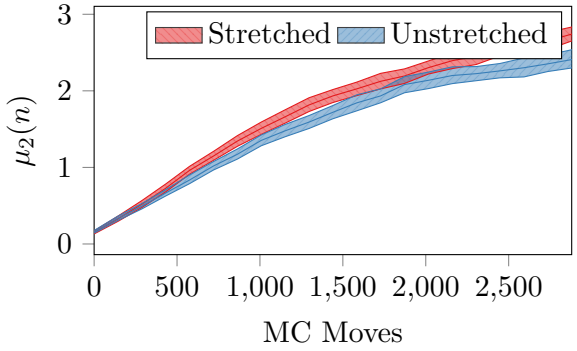
It is possible to piece apart how much of this is due to stretching from **Figure 3.14**, which compares networks generated at the same temperatures either stretched or unstretched. The difference between unstretched bond switching simulations and stretched bond switching



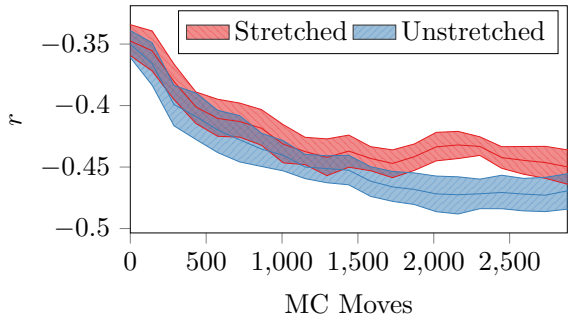
(a) $\mu_2(k)$ as a function of number of MC moves attempted at $T = 1.00 \times 10^{-2}$.



(b) r as a function of number of MC moves attempted at $T = 1.00 \times 10^{-2}$.



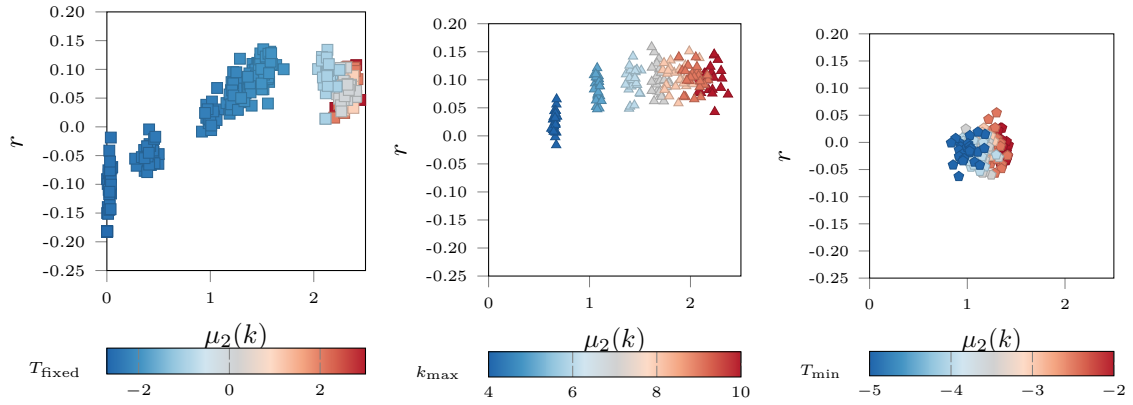
(c) $\mu_2(k)$ as a function of number of MC moves attempted at $T = 1.00 \times 10^{-1}$.



(d) r as a function of number of MC moves attempted at $T = 1.00 \times 10^{-1}$.

Figure 3.14: Defect metrics comparing stretched and unstretched networks at two different temperatures. At a low temperature (1.00×10^{-1} arb. units), the unstretched network shows no defects ($\mu_2(n) = 0$) whereas the stretched network does show defects. At a higher temperature (1.00×10^{-2} arb. units)„ the unstretched network shows fewer defects and are more disassortative. The filled area represents one standard deviation of the metric across the repeated simulations.

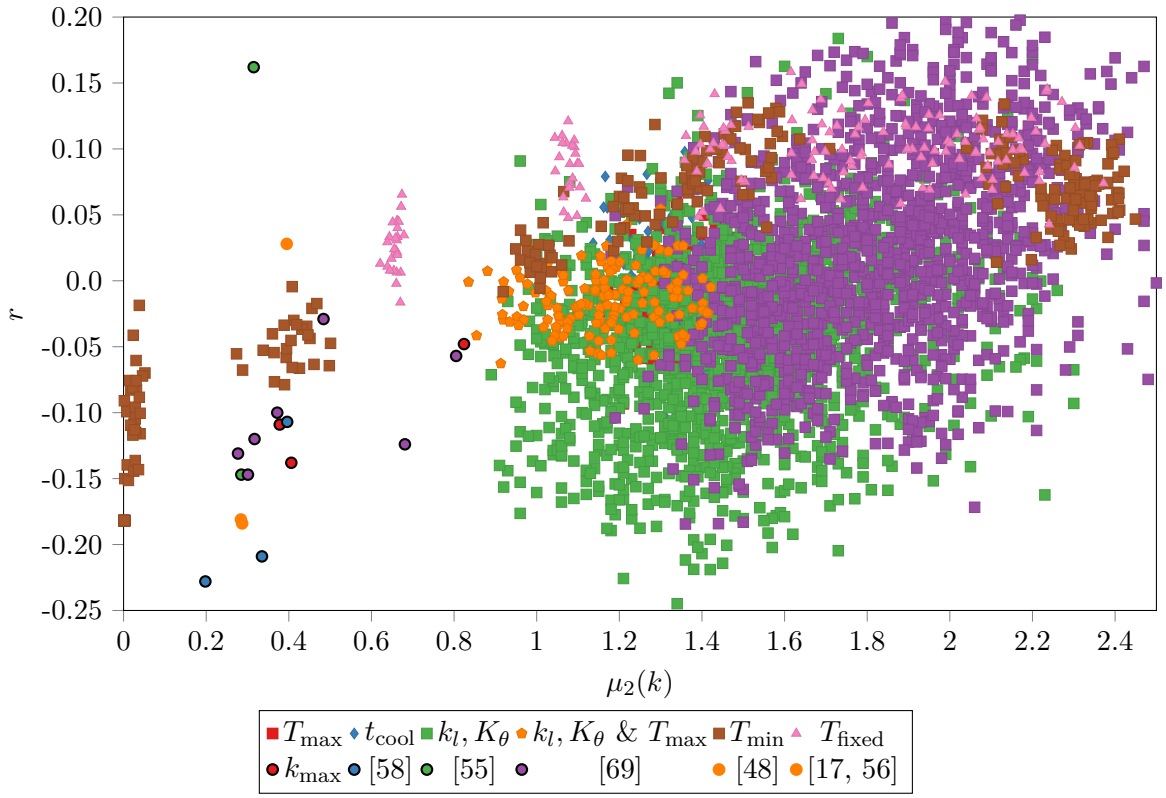
simulations becomes less significant at higher MC temperatures. At low temperatures, the difference is significant as the unstretched network exhibits no defects at all. At higher temperatures, the difference becomes less pronounced (but is still visible) as the metrics for stretched and unstretched networks track one another with a slight offset. This is because the stretching favours introduction of defects to lower the strain energy of the network, which becomes less significant in an entropically dominated regime at high temperature.



(a) $\mu_2(k)$ and r for networks generated with different partial thermalisation temperatures T_{fixed}, not undergoing the simulated annealing process. Bluer points were partially thermalised at a lower temperature, and redder points were thermalised at a higher temperature.

(b) $\mu_2(k)$ and r for networks generated with different maximum node degree k. Bluer points had a small maximum k and redder points had a large maximum k.

(c) $\mu_2(k)$ and r for networks generated at different minimum cooling temperatures. A redder point was cooled to a higher temperature, and a bluer point was cooled to a lower temperature.



(d) The total landscape with each set of coloured points representing a different simulation protocol. Icons remain the same as in the subpanels but have been recoloured. Outlined circles show experimental data.

Figure 3.15: Network landscapes for networks generated with a bond switching technique. This figure is adapted from M. H. J. Bailey *et al.*, *RSC Adv.*, 2020, **10**, 38275–38280[70].

3.4.6 Network Landscapes

To systematically understand what sorts of networks can be formed by a bond switching simulation process, many of the input parameters were varied across a wide range. These included varying the minimum cooling temperature during simulated annealing, varying the accessible degrees available to each node during simulated annealing, or only partially thermalising the networks.

The panels in **Figure 3.15** show the evolution of ring size variance μ_2 and the assortativity r. The subpanels are combined in the main figure, and compared to experimental data. Firstly, **Figure 3.15a** shows the $\mu_2(k)$ and r for networks generated by thermalising at a fixed temperature. The potential model parameters k_l and K_θ were both fixed at 0.100 arb. units, then 25 000 bond switching moves were attempted at a temperature of T_{fixed}.

At low temperatures the networks formed are disassortative, while at high temperatures the networks are mildly assortative. The two main areas in this landscape are found at the bottom left [negative r and smaller $\mu_2(k)$] and the top right [slightly positive r and larger $\mu_2(k)$]. The low-temperature bottom-left region represents networks that feature few uncorrelated defects in a regular lattice structure. Defects in the hexagonal network most commonly manifest like Stone-Wales defects in carbon networks[92], and since these defects are surrounded by the bulk of the regular hexagonal network this leads to disassortative mixing.

The number of defects is controlled by the balance of entropy and enthalpy. As the temperature increases more defects are introduced, leading to an increase in $\mu_2(k)$. The width of the node degree distribution function k increases up to a maximum, as the distribution of k tends to the maximum entropy distribution.

Figure 3.15b explores the effect of changing the accessible node degrees distribution by running simulated annealing simulations with different values of $k_{\max}$ in the range 3 to

10. Each simulation was thermalised at 1×10^4 arb. units for 5000 steps then cooled from 1×10^3 arb. units to 1×10^{-5} arb. units in increments of $1 \times 10^{-0.2}$ arb. units with 500 steps at each temperature. Potential model parameters k_l and k_θ were again set to 0.100 arb. units. For a low $k_{\max} = 4$, the networks have small values of $\mu_2(k)$ as the maximum width of the node degree distribution is necessarily narrow, and ring assortativities r of approximately 0, representing random mixing. As $k_{\max}$ increases towards ten, $\mu_2(k)$ increases (more values of k are accessible) and the networks become assortative with small positive values of r. This is because highly coordinated nodes with large k were surrounded by narrow triangles to relieve angular strain. Given that there are k neighbouring triangles around each node, this leads to a positive assortativity.

Figure 3.15c shows the effect of cooling the networks at the same rate for different lengths of time. Each simulation was again thermalised at 1.00×10^4 arb. units for 5000 steps then cooled from 1.00×10^3 arb. units to $T_{\min}$ in increments of $1 \times 10^{-0.2}$ arb. units with 500 steps at each temperature, and k_l, k_θ as before. This shows how much control is possible over the formed networks by relatively minor tweaks to the annealing procedure. Networks that are cooled for longer keep their assortativity. This is because the differences between simulations are in the low-temperature regions where there is not enough thermal energy to significantly change ring structure. However, the values of $\mu_2(k)$ drop as the highly coordinated nodes take many steps to break down. The high k nodes are enthalpically unfavourable because of significant angular strain.

Finally, **Figure 3.15d** combines all of these with experimental data (outlined circles) and the networks from **Figure 3.8** (with different k_l, K_θ). The experimental data points taken from references [17, 48, 55, 56, 58, 69] are generally found in the low $\mu_2(k)$ and negative r region, which is occupied sparsely by simulated networks. This is due to two main

factors: firstly, highly coordinated nodes with large k are relatively rare experimentally, as values of $k = 4$ are geometrically unstable with respect to splitting into two $k = 3$ nodes. Secondly, many biopolymer networks are constructed on a regular lattice or designed with some overarching structure[45,93]. It may be that **Chapter 2** had lower $\mu_2(k)$ than some of the simulated networks because the experimental networks were originally more hexagonal and had themselves gained defects through aging or preparation for imaging. The position of the experimental data points in the landscape reinforces the intuition that real biopolymer networks are unlikely to be randomly constructed. Entropy limits the amount of useful work that can be extracted from a system, and a network with maximised entropy is thus less useful to a living organism.

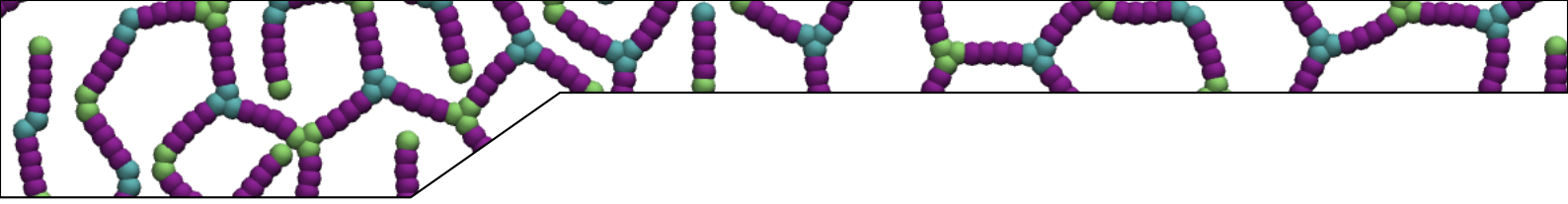
The simulated networks themselves show interesting physical properties as well. These include demixing, with a tendency towards assortative mixing. Similar rings tend to cluster together and become more convex when the networks were cooled for longer lengths of time. The increased cooling favours an enthalpic demixing, which allows a relaxing of angular strain. For example, nodes with $k = 2$ are least strained when they form a long boundary between large rings, or nodes with $k = 4$ are least strained when part of a region of squares.

3.5 Chapter Conclusions.

In conclusion, the bond switching method discussed in this chapter can generate physically reasonable 2D biopolymer networks that are similar to those found in biology. The strength of the bond switching method is that it can create a complex network by repeatedly introducing physically motivated defects without artificially imposing structure on those defects. It is important for a network generation method to generate networks with a wide range of ring and node distributions to match the diversity seen in biological networks.

Results in this chapter showed that the distributions of node degrees k often followed a maximum entropy distribution, and not a scale-free distribution. Experimental images often contain too few rings to fit a distribution with statistical certainty, especially given the ambiguities in assigning a graph representation to a network image. A key difference of biopolymer networks from inorganic glasses that have traditionally been simulated with bonds switching methods is the presence of concave rings. The extended bond switching method presented in this chapter can capture the irregular nature of biopolymer networks well, and the degree of irregularity can be controlled by changing simulation parameters (such as the cooling rate). Similarly, the ring size distribution and number of defects can also be controlled by the extended bond switching method, leading to a large variety of accessible networks.

The biological networks seen experimentally often show disassortative mixing in their ring structure, but simulated networks showed assortativities in the range $r = -0.200$ to 0.200. This is a weakness in the bond switching method, as it is noticeably difficult to generate networks that accurately match images of biological networks. The bond switched networks that best matched experimental networks were those that were partially thermalised with a low concentration of defects; networks that were generated exclusively by bond switching often showed structures dominated by entropy. This will inform later chapters of this thesis, which will use the bond switching method in conjunction with more physically motivated simulation techniques to generate and study collagen-like biopolymer networks.



Chapter 4: Self-Assembly.



This chapter details a Molecular Dynamics (MD) approach to simulating collagen IV networks. The mathematical underpinning of MD is discussed, as well as the importance of coarse-grained models. A coarse-grained model is used to assess the energetics of hexagonal *vs* chain-link fence network structures, and then to assess self-assembly processes. The effects of density, potential model parameters, and simulation cooling time on self-assembly are explored, and compared to results from Monte Carlo (MC) simulation and experiments. This chapter is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].



4.1 Introduction.

Chapter 3 was based on graph theoretical approaches to model biopolymer networks. Graph models of complex networks have many advantages, especially their speed and ease of interpretability. However, they must be parameterised based on sparse experimental data and small-scale simulations, and only provide a partial link between detailed molecular behaviour and the network behaviour.

Graph based models particularly cannot replicate one key experimental observation about collagen IV networks: that a network can spontaneously form from individual polymers[17,94]. Studying self-assembly computationally is an attractive proposition, as assembly occurs on

more accessible timescales than aging and the process is strongly dependent on hard-to-control environmental conditions such as the concentration of salts and cooling rates[45,48].

In this chapter, a coarse-grained biopolymer model is shown to self-assemble like collagen IV into two dimensional (2D) network structures. Coarse-grained models have been used to assemble extremely complex structures, including chains, rings and knots[95,96]. In this chapter, a polymer model is tuned to match experimental images of collagen IV networks both self-assembled and in samples of the ocular lens capsule. The coarse-grained nature of this model is important, as it allows long simulation timescales and can incorporate different interaction types, all with physically meaningful parameters. This chapter will show how much control is achievable over network structure by varying simulation parameters: the effects of cooling rates, molecular densities and strength of attractions are examined. Finally, the self-assembled networks are compared and contrasted to the networks generated with the bond switching approach of **Chapter 3** and the experimental data of **Chapter 2**. The self-assembled networks are shown to be more appropriate for reproducing experimental networks when compared with previous continuous random (entropically-driven) networks.

4.2 Computational Methods.

4.2.1 Molecular Dynamics.

MD is a method of simulating a physical system by repeatedly numerically integrating Newton's equations of motion. One mathematical formalism to describe it is the Lagrangian approach, which begins by describing a series of N particles with positions $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ and momenta $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$. These contribute to a potential energy function $\mathcal{U}(\boldsymbol{x})$ and kinetic energy function $\mathcal{T}(\boldsymbol{p})$ which combine to give a Lagrange function $\mathcal{L}(\boldsymbol{x}, \boldsymbol{p}) = \mathcal{U}(\boldsymbol{x}) - \mathcal{T}(\boldsymbol{p})$.

This Lagrange function then evolves according to the equation of motion

$$\frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = 0 \tag{4.1}$$

The natural framework to study the evolution of the system under that Lagrangian is to compute the forces on bodies $\frac{\partial \mathcal{U}}{\partial \boldsymbol{x}}$ according to a potential model. By solving the differential equations that arise numerically, the positions of the bodies over time can be predicted. Solving the differential equations is performed by applying the Velocity Verlet (VV) algorithm given in **Algorithm 4**[97].

Algorithm 4: The VV algorithm, where δt is a suitably small timestep.

Assign initial positions $\boldsymbol{x}(t=0)$;
Draw initial velocities $\boldsymbol{v}(t=0)$ from a Boltzmann distribution;
Calculate initial accelerations $\boldsymbol{a}(t=0)$ from the potential model and initial positions;
repeat
  Update positions with the 'half-step' velocities $\boldsymbol{x}(t+\delta t) = \boldsymbol{x}(t) + \boldsymbol{v}(t)\,\delta t + \frac{1}{2}\boldsymbol{a}\,\delta t^2$;
  Calculate a new $\boldsymbol{a}(t)$ from the potential model and positions $\boldsymbol{x}(t)$;
  Update velocities from the accelerations $\boldsymbol{v}(t+\delta t) = \boldsymbol{v}(t) + \frac{\boldsymbol{a}(t)+\boldsymbol{a}(t+\delta t)}{2}\delta t$;
until *simulation concluded*;

The VV algorithm is commonly used for MD over other numerical integration methods (such as Euler's method or Runge-Kutta). This is because the VV algorithm is a symplectic integration scheme, meaning it preserves volumes in phase space described by $\boldsymbol{x}$ and $\boldsymbol{p}$ over time. Symplectic integrators also preserve a perturbed form of the original Hamiltonian of the system, with perturbation of the order $\mathcal{O}\left(\delta t^2\right)$. Keeping a perturbed form of the original Hamiltonian also leads to desirable energy conservation properties.

There are many packages that implement MD simulations, with their own strengths and weaknesses. The work described in this thesis uses LAMMPS for MD simulations. This is because LAMMPS is well documented, commonly used in the literature, and it supports 2D simulation well[98,99].

4.2.2 Minimum Image Convention.

In order to get an accurate representation of a real physical system, the number of simulated atoms (or coarse-grained beads) N must be as large as possible. However, the computational time to calculate the interactions of bonded atoms scales as $\mathcal{O}(N)$, and to calculate the pairwise interactions of non-bonded atoms scales as $\mathcal{O}(N^2)$. This rapidly limits the system size and simulation timescale that can be simulated in a reasonable wall-clock time. Even the largest computer simulations still pale in comparison to nature; to date the largest atomistic simulations are approximately 1×10^9 atoms or 1×10^{-3} s[100–102], whereas the aging process of the human eye involves approximately 1×10^{19} atoms over 1×10^9 s. In simulations with a small N, a significant fraction of the constituent atoms are on the boundary of the simulation. This is different to nature, where most atoms are in the bulk of a material, and may lead to significant simulation artefacts. This can be resolved by treating a simulation periodically, similar to the approach used in **Figure 3.4**. Atoms leaving the left-hand side of the simulation re-enter on the right, atoms leaving the top re-enter on the bottom and *vice versa*. The introduction of periodic images allows for an effectively infinite simulation cell, but comes with a difficulty: if one molecule (or interaction) spans a simulation cell boundary, there is a discontinuity in the bond length which suddenly becomes very large. This is resolved by the minimum image convention, shown in **Figure 4.1**, where each atom only interacts with the closest periodic image.

4.2.3 Coarse Graining and Potential Model.

The model biopolymers in this chapter (and the following chapters of this thesis) were inspired by a combination of 'worm-like' chains and patchy particles. Patchy particles are colloids with carefully designed attractive regions that show a powerful tendency to self-assemble[103,104],

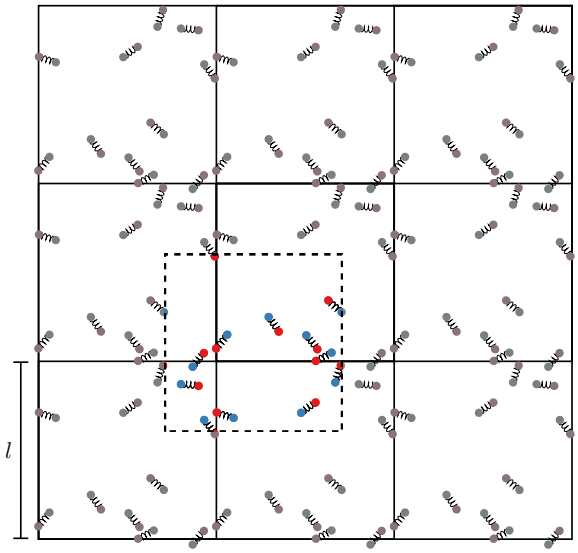


Figure 4.1: A schematic of the minimum image convention. Greyed-out molecules are periodic images of those in the central cell, and the highlighted molecules will only interact with the closest images (which are coloured).

and significant control can be achieved over the formed structures by fine tuning the nature of the attractive regions. Even more usefully, patchy particles have been shown to be an accurate model for crystallisation of proteins in solution[105]. Worm-like chains are often used to coarse-grain polymers into bead and spring chains, retaining the flexibility and complexity of the original, more detailed, polymer[37,40]. A simplified model is required as a single collagen IV molecule is extremely computationally expensive to simulate. Some of the detail and a process of simplification is shown in **Figure 4.2**.

The model biopolymers representing simplified collagen molecules are here treated as a worm-like chain constructed of series of beads joined by springs, shown in the bottom panel of **Figure 4.2**. Each spring joining two beads is modelled with a stiff harmonic potential leading to energy U_{spring}, and each pair of adjacent springs has their angle θ controlled by an angular potential leading to energy U_{angle}:

$$U_{\text{spring}} = k_l(r_{ij} - r_{\text{eqm}})^2 \qquad \text{and} \qquad U_{\text{angle}} = K_\theta[\cos(\theta_{ijk}) - \cos(\theta_{\text{eqm}})]^2. \tag{4.2}$$

The harmonic part features k_l, a bond force constant; r_{ij}, the distance between beads i and j;

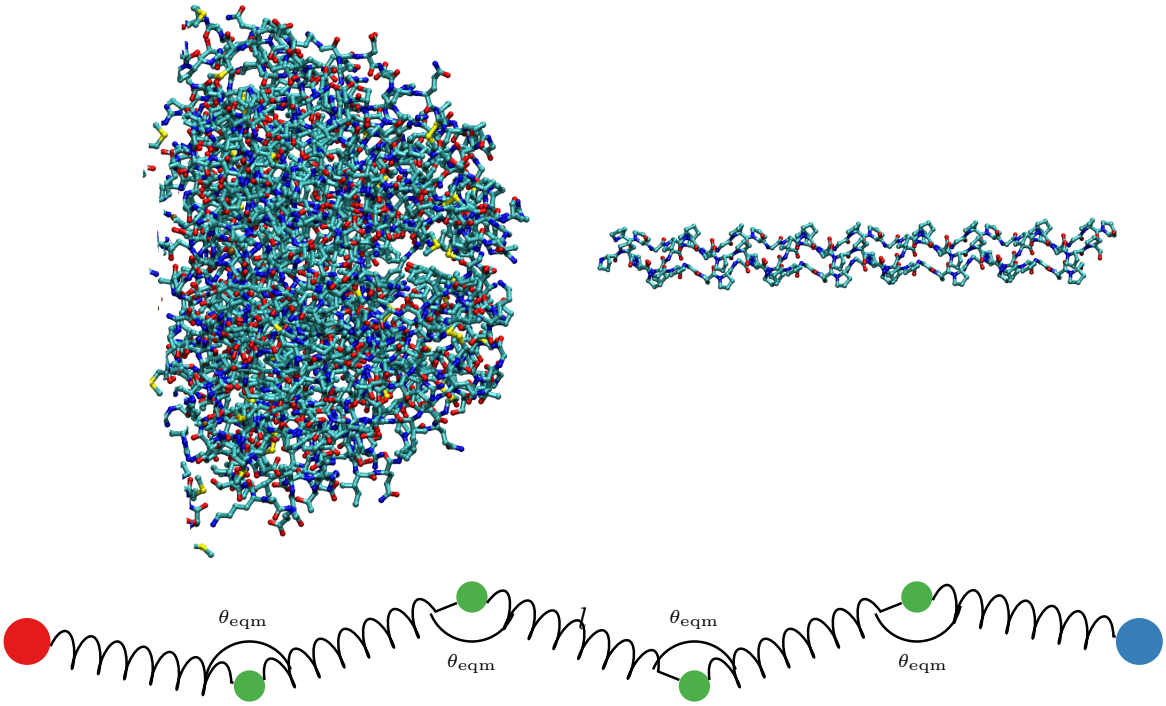


Figure 4.2: Two schematic parts of a collagen IV molecule, representing the simplification in the coarse graining process. The left-hand image represents a trimer in an Non-Collagenous 1 (NC1) domain, approximately 5000 atoms which are simplified into a single green head bead in the polymer representation[106]. The right hand image (shown on a different scale) shows one representative section of a triple helix joining the two ends of the collagen IV molecule, which is represented by two purple body beads and a joining spring[107]. The blue head bead at the other end of the polymer represents a 7S domain, which is not pictured here. This image is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

and r_{eqm}, the equilibrium separation between those two beads. The angular part features K_θ, an angle force constant; θ_{ijk}, the angle between bonds ij and jk; and θ_{eqm}, the equilibrium angle between those bonds.

The patchy nature of the polymer is then taken into account by attractive or repulsive potentials between each bead. The beads come in two types (although this can be extended), here body (B) and head (H) beads.

The body beads repel one another according to a truncated Lennard-Jones potential,

$$U_{\mathrm{BB}}(r) = \begin{cases} 4\varepsilon_{\mathrm{BB}}\left[\left(\frac{\sigma_{\mathrm{BB}}}{r}\right)^{12} - \left(\frac{\sigma_{\mathrm{BB}}}{r}\right)^{6}\right] & r < \sigma \\ 0 & r \geq \sigma. \end{cases} \tag{4.3}$$

The repulsive term is important to prevent polymers from interpenetrating one another, as a real polymer excludes a volume around it with an entropic force. The truncation distance σ_{BB} was chosen to be at the first zero of the U_{BB} curve, selecting only the repulsive part of the interaction. The head beads attract one another to encourage self-assembly, also according to a Lennard-Jones potential with energy parameter $\varepsilon_{\mathrm{HH}}$ and range parameter σ_{HH}

$$U_{\mathrm{HH}}(r) = 4\varepsilon_{\mathrm{HH}} \left[\left(\frac{\sigma_{\mathrm{HH}}}{r} \right)^{12} - \left(\frac{\sigma_{\mathrm{HH}}}{r} \right)^{6} \right]. \tag{4.4}$$

The range parameters σ_{HH} and σ_{BB} are chosen to encourage an average head bead coordination of $k = 3$ and to prevent molecules from interpenetrating. The body-head interactions are set to zero throughout with $\varepsilon_{\mathrm{BH}} = 0$. These interactions were chosen for simplicity, although future models may instead encourage a specific k by introducing attractive patches with a specialised potential model. For example, the Kern-Frenkel model or a generalised anisotropic Lennard-Jones model would be suitable[108,109], but are not implemented in LAMMPS.

By changing the energy and range scales of the Lennard-Jones potentials it is possible to encourage or inhibit assembly. Different parameters can also mimic the screening effects of salts dissolved in the solvent, without explicitly taking them into account, thereby saving significantly on computational complexity. Real collagen IV molecules have two different head groups: 7S and NC1 domains. These head groups only interact with other groups of the same type, and have different preferred degrees $k = 2$ or $k = 4$. Applying this interaction scheme to a simple simulation would lead to only even-sized rings forming, similar to the magnetic frustration observed in ferromagnetic systems. Forbidden odd-size rings are shown in **Figure 4.3**. To avoid that, the two head groups used here both have preferred degree $k = 3$ and interact regardless of type.

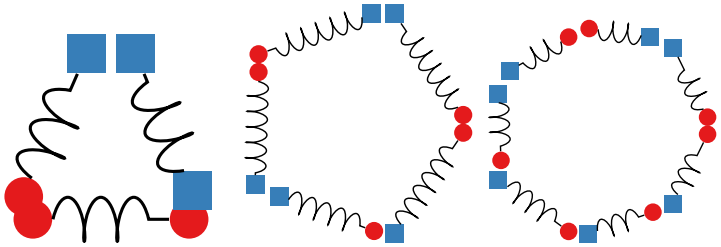


Figure 4.3: Three frustrated arrangements of polymers. Each frustrated arrangement has an odd number of edges, and thus one vertex must have feature different head groups (here red and blue). If head groups can only interact with other head groups of the same type, these structures never form.

4.2.4 Temperature Control.

The MD scheme described in **Section 4.2.1** is useful to simulate an ensemble at a constant number of particles, volume and energy (NVE), but that is often unhelpful in simulating physical systems. Instead, it is more common to simulate systems at a constant number of particles, volume and temperature, NVT. To achieve this, the equations of motion must be changed to add terms that control the temperature. There are multiple simulation approaches, each with their own benefits and downsides: Andersen thermostats, which randomly change the velocity of one particle; Nosé-Hoover thermostats, which couple the system to a heat bath; and Langevin thermostats, which simulate Brownian motion. The MD simulations in this chapter make use of a Langevin thermostat (unless otherwise specified), which modifies the forces $\boldsymbol{F}$ on each atom according to

$$\boldsymbol{F} = \underbrace{\boldsymbol{F}_{\text{Bond}} + \boldsymbol{F}_{\text{Angle}} + \boldsymbol{F}_{\text{LJ}}}_{\text{Conservative Forces}} - \underbrace{\frac{m\boldsymbol{v}}{\gamma}}_{\text{Drag}} + \underbrace{\xi\sqrt{\frac{k_{\text{B}}Tm}{\gamma \text{d}t}}}_{\text{Brownian}}. \tag{4.5}$$

This includes the 'conservative forces' between bonds, angles, and head groups denoted $\boldsymbol{F}_{\text{Bond}}$, $\boldsymbol{F}_{\text{Angle}}$, and $\boldsymbol{F}_{\text{LJ}}$. These are then modified by the addition of a drag term with coefficient γ and a Brownian motion term with parameter ξ (here chosen to be 1).

4.2.5 Parameterisation and Simulation Protocol.

Assigning physically meaningful parameters to simplified models is often performed by comparing to either experiment or more detailed models. Detailed models of collagen IV are too computationally expensive to effectively use for parameterisation, and experimental data rarely capture useful dynamical information. Where experimental data are available, they tend to be of mechanical properties such as elasticity and stress-strain relationships — **Chapter 5** will discuss the pitfalls when using mechanical data to parameterise a polymer model.

Instead, the general properties of self-assembling polymer models will be explored and compared to experiment. To that end, the potential model parameters will be varied significantly to understand the range of different possible behaviours and simulation outcomes. Generally, individual coarse-grained polymers were chosen to be made of 6 beads separated by 50.0 nm bonds for a total length of 300 nm; this is slightly shorter than the maximum length of collagen IV molecules recorded in solution. The shorter length was so the attractive regions at both end are approximately 350 nm apart, matching the results from Timpl *et al.* and Dölz *et al.*[110,111]. Each bond had an elastic constant of $k_l = 1.66 \times 10^{-4}\,\mathrm{N\,m^{-1}}$ to allow an extension of approximately 5 % at 300 K.

The beads are joined by springs with equilibrium length 50.0 nm, which is in the range of measured persistence lengths for collagens (measured to be 10.0 nm to 167 nm)[39,44]. Each pair of adjacent springs is joined by an angular potential, with a default value of $K_\theta = 200 \times 10^{-21}\,\mathrm{J}$ to roughly reproduce the angular correlation between segments in Rezaei *et al.*'s model, although this is extremely sensitive to environmental conditions[39].

Unless otherwise specified, the head beads had attractive length scales $\sigma_{\mathrm{HH}} = 50.0\,\mathrm{nm}$ and repulsive scales $\sigma_{\mathrm{BB}} = 138\,\mathrm{nm}$ to create the right total length of a molecule and encourage

$k = 3$. σ_{BB} is significantly larger than the measured width of a collagen IV triple helix ($\approx 1.50\,\mathrm{nm}$) as the simplified model must also take into account an entropic excluded volume effect. The energy scales of head group parameters were often $\varepsilon_{\mathrm{HH}} = 4.14 \times 10^{-21}\,\mathrm{J}$ and $\varepsilon_{\mathrm{BB}} = 16.1 \times 10^{-21}\,\mathrm{J}$ to encourage self-assembly at approximately 300 K.

The simulation protocol, unless otherwise specified, followed these steps:

- Place polymers on a square $N \times N$ lattice with random orientations such that none of them overlap.
- Simulate the polymers as an NVT ensemble until they form a proto-network for time t_{equil}.
- Fix the simulation cell size by equilibrating the network as an NpT ensemble under a barostat with a small external 2D pressure p for time t_{pressure}.
- Return to the NVT ensemble and melt the network at 300 K for time t_{heat}.
- Cool the network from 300 K to 0 K over time t_{cool}.

4.3 Results and Discussion.

4.3.1 Energetics of Idealised Network Structures.

There is significant debate in the literature about the actual network structure adopted by collagen IV (and collagen-like) networks. Authors such as Timpl *et al.* and Cummings *et al.* argue in favour of a 'chain-link fence' (sometimes referred to as 'chickenwire') network[45,110], which is a network of octagons where half the head-groups have degree $k = 4$ and the other half have degree $k = 2$. However, other authors such as Burd argue in favour of hexagonal networks, based off self-assembly results from Yurchenco & Furthmayr[14,17].

Imaging experiments, like those analysed in **Chapter 2**, have shed light on the structure of collagen networks. But as **Chapter 2** showed, ambiguity in images and the presence of complicating factors such as surfaces, dissolved salts, and preparation for study means that the matter has not been fully resolved. A computer model can help predict which interpretations are more likely by analysing how energetically favourable and kinetically stable different network structures are.

To do so, ideal hexagonal and 'chain-link fence' networks were constructed by first creating a graph representation of the network and then overlaying model polymers. The hexagonal network and 'chain-link fence' networks are shown in **Figure 4.4a** and **Figure 4.4b** respectively. The coordinates of nodes in idealised networks were then scaled in the xy plane to assess the range of preferred stability.

The energy-area curves showed that the hexagonal network is always lower in energy than the 'chain-link fence' network with these potential models. Further explorations showed that this was true for all probed combinations of polymer parameters k_l, K_θ and $\varepsilon_{\mathrm{HH}}$.

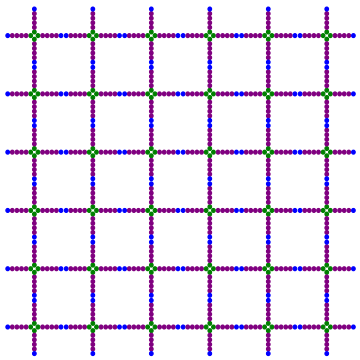
The energy-area curves are shown in **Figure 4.4c** for a single simulation with parameters $\varepsilon_{\mathrm{HH}} = 4.14 \times 10^{-21}\,\mathrm{J}$, $k_l = 1.66 \times 10^{-4}\,\mathrm{N\,m^{-1}}$, and all other parameters set to the defaults discussed in **Section 4.2.5**. This shows that the hexagonal network is more energetically stable up to an area of $1.07 \times 10^5\,\mathrm{nm^2}$ per molecule. This is partly because the hexagonal networks are denser than the chain-link fence networks, each individual hexagon is made of fewer molecules than the squares in the chain-link fence network (made of 8 molecules). **Figure 4.4d** shows that the energetic penalty mostly comes from stretching of bonds. As each molecule is straight during these simulations, the energy is independent of K_θ.

At larger stretching ratios, the chain-link fence network becomes more favourable as it relieves bond stretching strain. However, a hexagonal to square transition is unlikely to

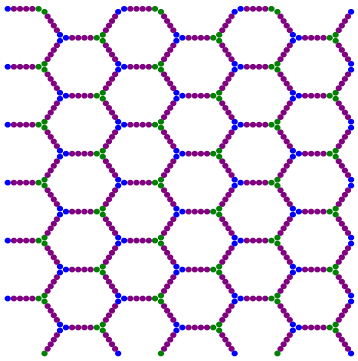
occur in nature because it involves a large concerted rearrangement in which all nodes must switch from being $k = 3$ to $k = 2$ or 4. The presence of a surface or scaffold would make this rearrangement even more unlikely.

Figure 4.5a shows that at all values of k_l and $\varepsilon_{\mathrm{HH}}$, the hexagonal network is of lower energy than the square network. **Figure 4.5b** also shows that the hexagonal network is at an increased density, with a smaller optimal area, for all polymer parameters. The optimal area for the square network increases slightly with $\varepsilon_{\mathrm{HH}}$ owing to the more complex interactions at degree 4 nodes. However, the hexagonal network retains the same optimal area regardless of potential model parameters.

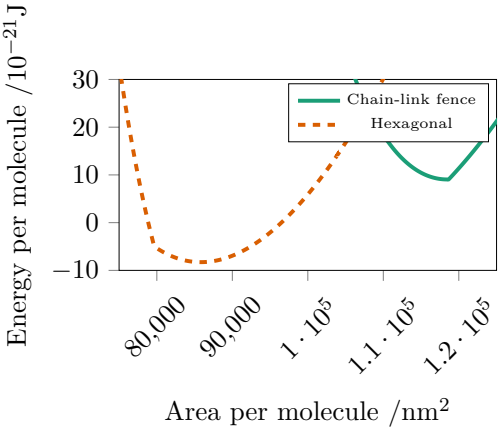
This does not rule out a natural preference for the 'chain-link fence' network. A more sophisticated potential model that discriminates between 7S and NC1 domains, and takes the existence of a scaffold or surface into account, may lead to a different preferred phase, but is outside the scope of this work.



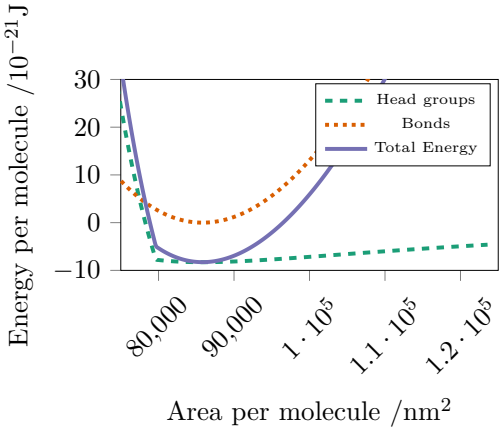
(a) An idealised 'chain-link fence' network.



(b) An idealised hexagonal network.

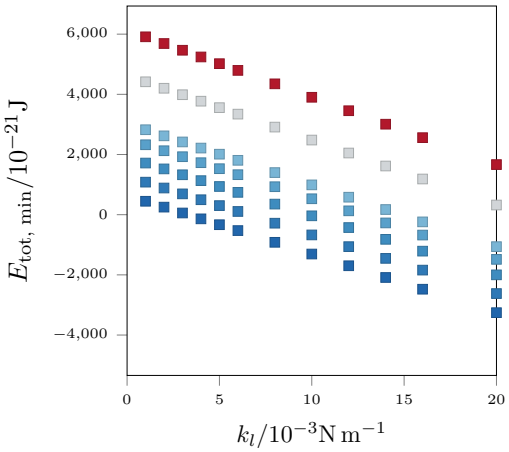


(c) Energy per molecule as a function of area per molecule for two different arrangements. The hexagonal network structure is lower energy than the chain link network structure until $1.08 \times 10^5\,\mathrm{nm}^2$.

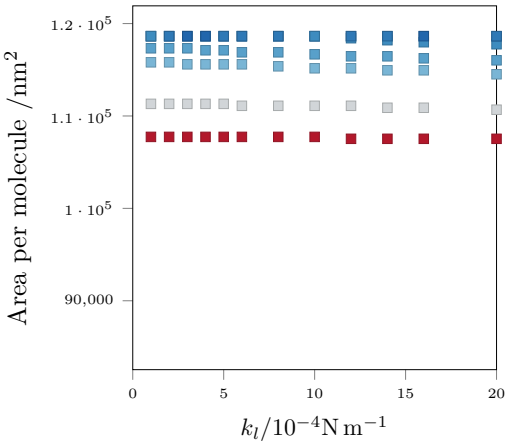


(d) Components of the energy per molecule as a function of area per molecule for a hexagonal network. Repulsion in the low area regime arises from head groups repelling one another, as the polymers can buckle. Resistance to stretching, however, mostly arises from increased energy of bond stretches.

Figure 4.4: A comparison of two idealised networks, one constructed of hexagons (where each edge is one molecule long) akin to a honeycomb and one constructed of squares (where each edge is two molecules long) akin to a chain-link fence. The networks were scaled isotropically with scale factor $\lambda = 0.900$ to 1.20. This was converted into an area per molecule as the networks had different initial areas. This image is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].



(a) Energy at the minimum volume as a function of polymer bond strength, k_l and coloured by head group interactions $\varepsilon_{\mathsf{HH}}$.



(b) Volume that minimises the energy of a network as a function of polymer bond strength k_l and coloured by head group interactions $\varepsilon_{\mathsf{HH}}$.

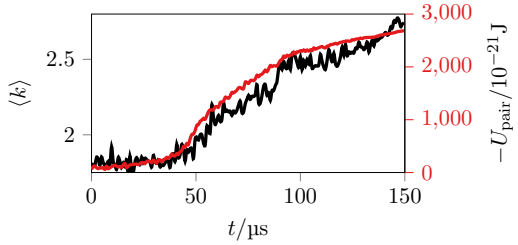
Figure 4.5: Hexagons represent idealised hexagonal networks, and squares represent chain-link fence networks. The chain-link fence are coloured by $\varepsilon_{\mathsf{HH}}$. A redder colour indicates a larger $\varepsilon_{\mathsf{HH}}$, and a bluer colour represents a lower $\varepsilon_{\mathsf{HH}}$. Hexagonal networks are not coloured as their energy is independent of $\varepsilon_{\mathsf{HH}}$ and only depends on k_l. These images are adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

4.3.2 Self-Assembly.

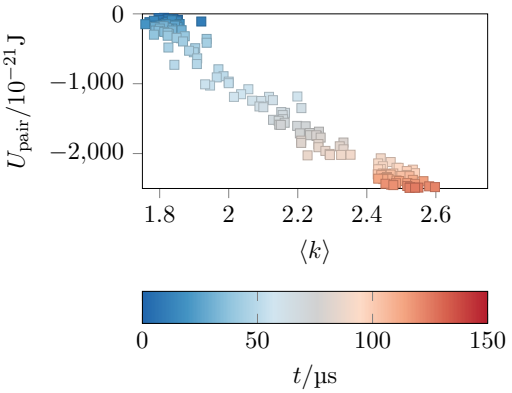
Degree and Internal Energy.

The energetics of the network structure as explored in **Section 4.3.1** are useful but clearly dependent on the relatively arbitrary choice of parameters k_l, K_θ and $\varepsilon_{\mathrm{HH}}$. These parameters are hard to assign physical values to, owing to the difficulty in collecting data suitable for use in coarse-graining process. Instead, this dependence can be simplified thanks to the short-range nature of the attractive potentials. Each head bead in a cluster only interacts with other head beads within that cluster, allowing us to use the degree k as a proxy for the actual internal energy of the network, with an added benefit of providing information on network structure. This simplifies the study of the network formation, as it removes complicating thermal factors and intra-polymer factors.

The correlation between degree k and internal energy U is shown in **Figure 4.6a** for a single simulation cooled over 100 µs. A scatter graph with coordinates $\langle k \rangle$ and U_{pair} from **Figure 4.6a** is shown in **Figure 4.6b** showing a strong linear correlation between the two. This correlation was observed across all simulations performed and is extremely useful to smooth out noise.



(a) Mean node degree $\langle k \rangle$ and head group energy over time for a system undergoing network self-assembly



(b) A scatter plot of mean node degree $\langle k \rangle$ against head group interaction energy U_{pair} to emphasize their correlation. Each point is coloured by simulation time, with a bluer point corresponding to a sample from near the start of the simulation and a redder point indicating a sample near the end of the simulation.

Figure 4.6: Two figures to highlight the strong correlation between the mean degree $\langle k \rangle$ and the pairwise head-group interaction energy U_{pair}. Since $\langle k \rangle$ dominates the internal energy, it can be used as a simple proxy for the energy.

4.3.3 Control of the Network Structure.

Polymer Potential Parameters.

As discussed in **Section 4.2.5**, the focus of this chapter is on determining what range of network structures can be generated by the polymer model, with hopes of providing a comparison point for the large range of networks seen experimentally. The difficulties in parameterisation lead to a great deal of uncertainty in the 'best' values to assign. To resolve this, polymers were allowed to self-assemble across simulations with different polymer parameters: k_l and K_θ. k_l had very little effect on the nature of self-assembled rings formed, with the mean length of an edge in the network remaining at $\langle l \rangle = 317\,\text{nm} \pm 53.0\,\text{nm}$. This is in agreement with the measurements made by Yurchenco & Furthmayr[17], but slightly longer than Barnard *et al.*'s measurements[112]. The low sensitivity of network structure to k_l means that it can be chosen for computational convenience — the minimum timestep of the

K_θ / 1.00×10^{-21} J	$\langle l \rangle$ / nm	N_{ring}	$\sigma_{N_{\text{ring}}}$
5.00	295	79.2	8.64
50.0	312	78.3	4.76
100	315	71.1	6.06
150	317	68.2	5.83
200	318	71.6	4.48
250	318	69.0	7.04
300	318	69.9	7.58
400	319	66.0	3.65

Table 4.1: The mean edge length $\langle l \rangle$, number of rings N_{ring}, and the standard deviation of the number of rings $\sigma_{N_{\text{ring}}}$ counted across 10 simulations as the angular strength K_θ changes. This table is reproduced from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

simulation δt capture vibrations in the stiffest part of the system, in this case the harmonic bonds. Network structure being independent of stiffness means that less stiff polymers can be used, allowing a larger δt and thus study of long time-scale properties of the system.

The angular strength parameter K_θ is slightly more interesting, and results of varying this are shown in **Table 4.1**. The first major effect is that the average length of an edge in the network increased from 295 nm for low K_θ to 319 nm for high K_θ, although this flattens off for K_θ in the range 200×10^{-21} J to 400×10^{-21} J. This is because the head-to-head distance between coordination sites is shorter if the backbone curves (a factor that cannot be taken into account in the previous MC model). In addition, more flexible polymers can bend to reduce steric strain and coordinate more easily. The second major effect is that simulations with larger K_θ formed fewer rings, from 79 when $K_\theta = 5.00 \times 10^{-21}$ J down to 66 when $K_\theta = 400 \times 10^{-21}$ J. This reinforces the importance of using a semiflexible worm-like curve model for biopolymers. Visual analysis of trajectories revealed that fewer rings where formed when K_θ was large because polymers lose the ability to bend towards one another during formation.

Dynamical Cooling Time.

Crystallisation phenomena (including self-assembly) are often strongly affected by cooling rates, and **Chapter 3** suggests that even an abstract MC 'cooling time' has significant consequences for network structure. Cooling rates are particularly hard to control precisely in a laboratory setting, but easy to control computationally by changing thermostat parameters. This comes with some caveats, as the short timescales accessible to simulation lead to cooling rates that are unphysically high. A coarse-grained model like the one in this chapter reduces the extremely high cooling rates, but does not eliminate the problem. For example, cooling from 300 K to 0 K in 100 µs is a rate of $3.00 \times 10^6\,\mathrm{K\,s^{-1}}$. In comparison, even a rapid experimental cooling rate from 300 K to 270 K over 10.0 s is just $3.00\,\mathrm{K\,s^{-1}}$.

To study the effect of cooling rate on polymer formation, the length of time the polymers had to cool (referred to as t_{cool}) from 300 K to 1.00 K was varied from 54.0 µs to 150 µs. All other parameters remain as discussed in **Section 4.2.5**. Longer t_{cool} periods had a significant effect on the forming of rings within the network, increasing from 31.6 on average to 105 on average. This is shown in **Figure 4.7**, with two example networks highlighted at different t_{cool}.

Fewer rings are formed when the network cools more rapidly as the drag from the simulated solvent prevents polymers from moving. The frozen polymers are thus locked into place, dangling, and cannot rearrange to form rings. If the polymer system is cooled more slowly, the individual polymers have more time in which to encounter one another and slot into place in the network.

Furthermore, longer values of t_{cool} causes the average node degree to increase to $\langle k \rangle = 3.00$ for $t_{\mathrm{cool}} = 150\,\mu\mathrm{s}$, up from $\langle k \rangle = 2.64$ for $t_{\mathrm{cool}} = 54.0\,\mu\mathrm{s}$. This is again due to polymers not having time to coordinate fully before being locked into place, and thus being left dangling.

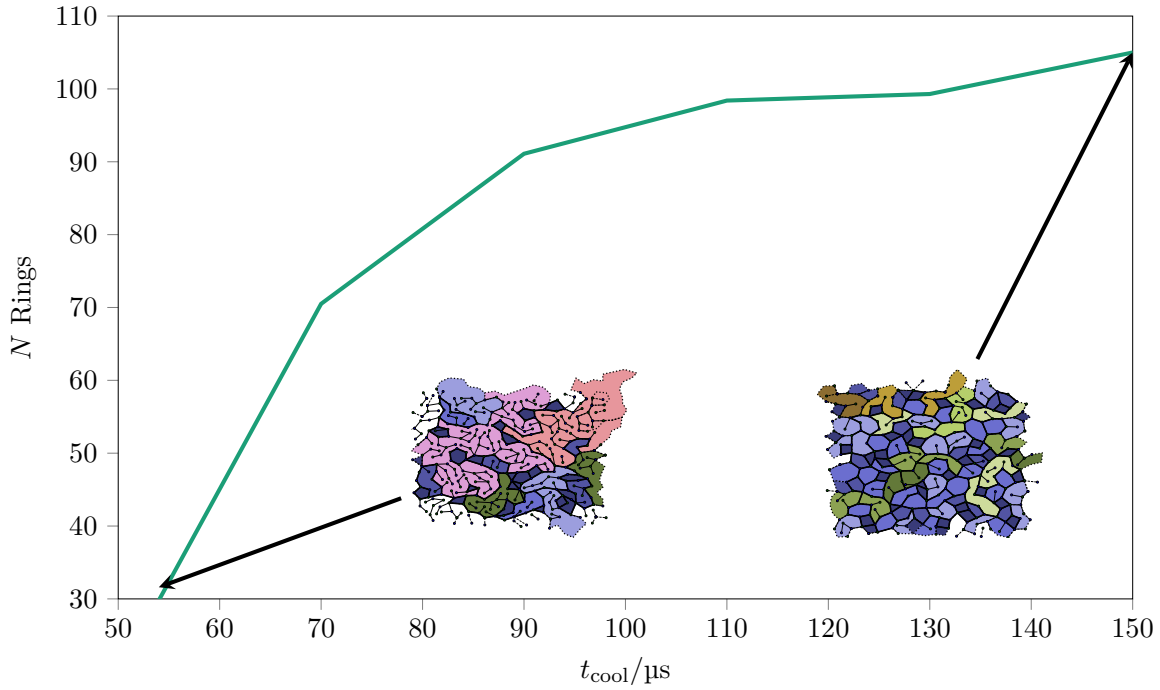


Figure 4.7: The number of rings formed, N, against the time taken to cool the networks from 300 K to 30.0 K (denoted t_{cool}) with two networks cooled at different rates inset as examples. Different colours are used to highlight rings with the same number of edges. This figure is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

When rings do form, the average area of rings with n edges remains unchanged, as does the average size of each edge. This suggests that the rings that form are equally energetically rewarding, it is just that they are prevented from forming when too cool.

Molecular Density.

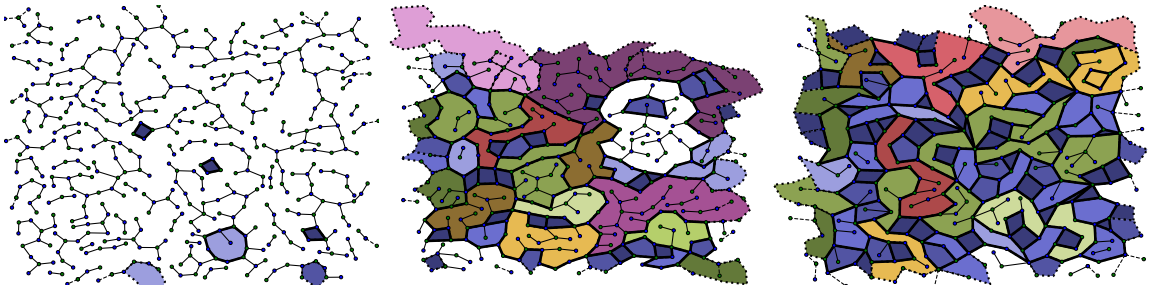
The emerging evidence from **Section 4.3.3** is that network structure is kinetically controlled, which deserves further study. An easy way to control the kinetics of network formation is to increase the density of polymers per unit area during formation — more polymers in a confined space increases their chance of interaction, and fewer polymers decreases their chance of interaction. The decreasing temperature during network formation, and high value of ε_{HH} compared to thermal energy, means that when two head groups stick together it is often irreversible. Studies of real collagen IV networks have suggested that the assembly of 7S

p/Pa	$\langle\rho\rangle$ /nm^{-2}	$\langle A\rangle$ /nm^2	$\langle k\rangle$	$\langle n\rangle$	D.E./%
5	4.60×10^{-4}	867 000	2.59	6.79	62.6
10	6.25×10^{-4}	640 000	2.69	7.04	47.5
20	1.01×10^{-3}	396 000	2.80	6.82	36.6
50	1.81×10^{-3}	221 000	3.06	6.03	23.9

Table 4.2: Assembly 2D pressure p, average area per ring $\langle A\rangle$, average molecules per unit area $\langle\rho\rangle$, average degree $\langle k\rangle$, average edges per ring $\langle n\rangle$, and the percentage of dangling edges which do not form parts of rings, as a function of molecular density. This table is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

head groups into tetramers is thermodynamically irreversible, which is critical for network formation[17]. Instead of assigning network densities directly, it is easier to change the 2D pressure p of the barostat for the proto-network; this prevents artefacts from the initial lattice from affecting the final network structure. 2D pressure p should be interpreted cautiously, as it is in fact a strain per unit area. The simulation is performed in standard three dimensional (3D) Cartesian coordinates, but each atom is constrained to $z = 0$ and images are well separated in z (see **Chapter 7** for a discussion of the consequences of this). The calculation of p is then corrected for the lower dimensionality. The remainder of this chapter will use the term 'pressure' because of its similarity to the 3D analogue. Furthermore, the pressure is only used during the NpT step in which the Langevin thermostat is disabled and a Nosé-Hoover thermostat with a barostat component is applied in its place. This neglects the effects of implicit solvent during the NpT step, which is done to computationally fix the simulation cell size.

Below a density of approximately $4.00 \times 10^{-4}\,\mathrm{nm}^{-2}$, no networks form (a 2D pressure roughly equivalent 2.00 Pa during formation). At a low molecular density, the likelihood of two polymers interacting before freezing is extremely small and the polymers are effectively well-dissolved. The average distance between interacting polymer beads is greater than the cut-off for assessing intermolecular interactions.



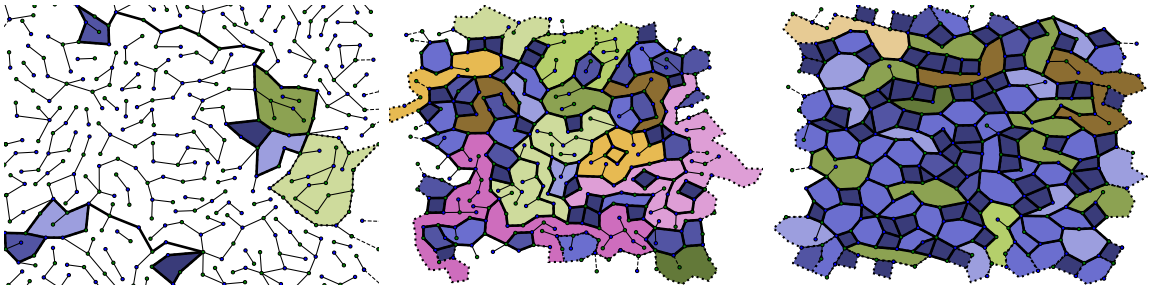
(a) $\rho = 7.90 \times 10^{-6}\,\mathrm{nm}^{-2}$ **(b)** $\rho = 6.25 \times 10^{-4}\,\mathrm{nm}^{-2}$ **(c)** $\rho = 1.81 \times 10^{-3}\,\mathrm{nm}^{-2}$

Figure 4.8: A comparison of three networks cooled at low, intermediate, and high molecule densities. Different colours are used to highlight rings with the same number of edges, and these networks are drawn at different scales. This image is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

As the density increases (an increased 2D pressure during equilibration) the mean degree per node $\langle k \rangle$ increases as there are more interactions between polymers. **Table 4.2** shows this, and also shows the decrease in mean number of edges per ring $\langle n \rangle$ as density increases; this is matched by a decrease in average area per ring. At low or intermediate densities, the networks often had a high fraction of dangling edges, defined as edges that were not part of any ring. The dangling edges arose when a loose polymer was trapped inside a larger ring as it formed, and were more common for low-density networks where rings with larger areas were common. As the density increased, the networks left little room for dangling nodes. Some example networks can be seen graphically in **Figure 4.8**. Finally, when the assembled density was too high, no well defined network formed. This is because at an extreme density the polymers could not adopt the correct orientations for network formation due to the lack of free space.

Lennard-Jones Interaction Strength.

The final axis of control over network structures is the strength of head group interactions. These are a useful proxy for the screening effects of salts dissolved in the solvent, which have been studied experimentally in great depth. Dissolved salts can affect the rate of formation,



(a) $\varepsilon_{\mathrm{HH}} = 1.00 \times 10^{-21}\,\mathrm{J}$ **(b)** $\varepsilon_{\mathrm{HH}} = 1.00 \times 10^{-20}\,\mathrm{J}$ **(c)** $\varepsilon_{\mathrm{HH}} = 2.00 \times 10^{-20}\,\mathrm{J}$

Figure 4.9: A comparison of three networks formed at low, intermediate, and high Lennard-Jones well depths. Rings that have the same number of edges are coloured the same. This image is adapted from M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262[71].

the network structure and the properties of individual polymers[39,45,48]. In an experiment, these are difficult to piece apart: it is difficult to vary the stiffness of polymers without also changing the energetics of network formation. Thankfully, these effects are easy to extricate computationally (**Section 4.3.3** has looked at the rate of formation and **Section 4.3.3** has looked at the effect of polymer stiffness parameters).

The effects of screening head-group interactions was studied by varying the attractive LJ energy scale $\varepsilon_{\mathrm{HH}}$ in the range $1.00 \times 10^{-21}\,\mathrm{J}$ to $2.00 \times 10^{-20}\,\mathrm{J}$. Example networks at low, intermediate and high values of $\varepsilon_{\mathrm{HH}}$ are shown in **Figure 4.9** Relatively weak head group interactions (a proxy for highly screened high-salt interactions) lead to a network with few well-defined rings. This is because $\varepsilon_{\mathrm{HH}}$ is smaller than the thermal energy for a significant fraction of the simulation and formed nodes are easily broken, then have little motivation to re-form. The nodes thus remained as isolated chains to maximise their entropy.

As the Lennard-Jones interactions became stronger (a proxy for lightly screened low-salt interactions), the number of successfully formed rings increased significantly and $\langle k \rangle$ increased concomitantly. Data for these phenomena are in **Table 4.3**. Simulations with a large value of $\varepsilon_{\mathrm{HH}}$ showed networks with many hexagons and squares formed to maximise the energetic reward for the interaction of head group. The data in **Table 4.3** bear a striking similarity to

$\varepsilon_{HH}/10^{-21}$ J	$\langle$Area$\rangle$ /nm^2	$\langle n \rangle$	$\langle k \rangle$
2.00	720 000	6.42	2.69
4.00	387 000	6.89	2.82
6.00	332 000	6.92	2.87
8.00	282 000	6.56	2.95
10.0	261 000	6.41	2.98
12.0	257 000	6.38	2.98
14.0	242 000	6.27	3.01
16.0	232 000	6.1	3.03
20.0	217 000	5.93	3.05

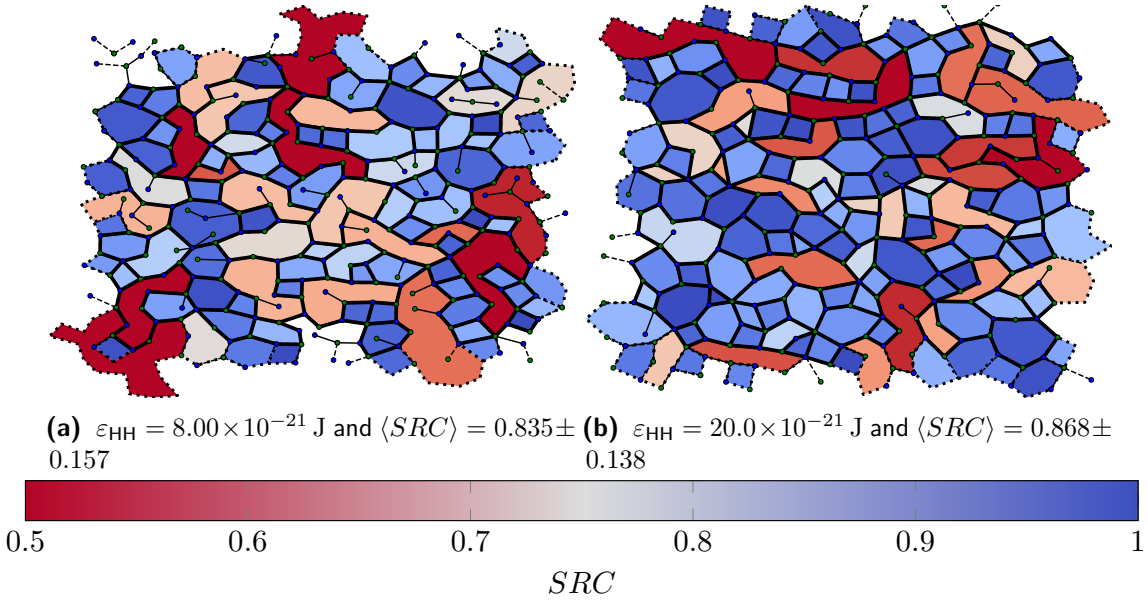
Table 4.3: Average ring area, average edges per ring $\langle n \rangle$, average degree $\langle k \rangle$ of networks as a function of Lennard-Jones well depth.

the data in **Table 4.2**. This is because both the network density and head-group interaction strength change the ratio between the kinetic energy of monomers and the potential energy they gain from network formation. By varying this ratio, the balance is shifted from enthalpy encouraging the polymers to be locked into a favourable network arrangement, towards entropy encouraging loose chains of molecules.

4.3.4 Regularity Criterion.

As discussed back in **Section 3.3.3** and **Section 3.4.4**, a key feature of biopolymer networks is their tendency to form irregular shapes. This is again measured with Shape Regularity Criterion (SRC). Two example networks generated with different values of ε_{HH} are shown in **Figure 4.10** These show one less regular network with $\langle SRC \rangle = 0.835$ and one slightly more regular network with $\langle SRC \rangle = 0.868 \pm 0.138$. For comparison, the images in **Figure 3.11** generated with a MC method had $\langle SRC \rangle = 0.856 \pm 0.007\,45$. Experimental images had SRC in the range 0.693 to 0.966.

Networks generated with the self-assembly method tend to match the observed range of regularities from both MC networks and experimental networks well, with data verifying this in **Table 4.4**. However, the standard deviations of the regularity are extremely large,



(a) $\varepsilon_{\mathrm{HH}} = 8.00 \times 10^{-21}$ J and $\langle SRC \rangle = 0.835 \pm 0.157$ **(b)** $\varepsilon_{\mathrm{HH}} = 20.0 \times 10^{-21}$ J and $\langle SRC \rangle = 0.868 \pm 0.138$

Figure 4.10: Two self-assembled networks coloured by regularity taken from **Section 4.3.3** with different head group attraction strengths.

$t_{\mathrm{cool}}/\mu\mathrm{s}$	$\langle SRC \rangle$	$\mathrm{Var}(SRC)$
54.0	0.806	0.158
70.0	0.781	0.184
90.0	0.805	0.174
110	0.822	0.166
130	0.828	0.162
150	0.831	0.159

(a) Regularity data for varied cooling times t_{cool}.

$\varepsilon_{\mathrm{HH}}/1.00 \times 10^{-21}$ J	$\langle SRC \rangle$	$\mathrm{Var}(SRC)$
1.00	0.844	0.114
2.00	0.769	0.179
4.00	0.780	0.185
8.00	0.835	0.157
16.0	0.857	0.140
20.0	0.866	0.138

(b) Regularity data for varied head group attraction strengths $\varepsilon_{\mathrm{HH}}$.

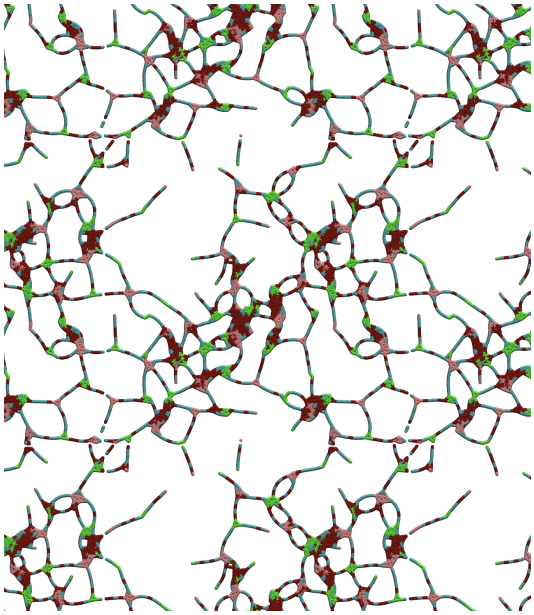
Table 4.4: Regularity data for networks formed while different parameters were varied (holding others constant).

indicating a significant range of regular and irregular rings. This is hard to compare to single images of experimental data as the images often cover few rings, but a normal experimental σ_{SRC} is approximately 0.100 (full data are available in **Appendix C**). Self-assembled networks have range of regularities that better match experimental data than the MC networks do.

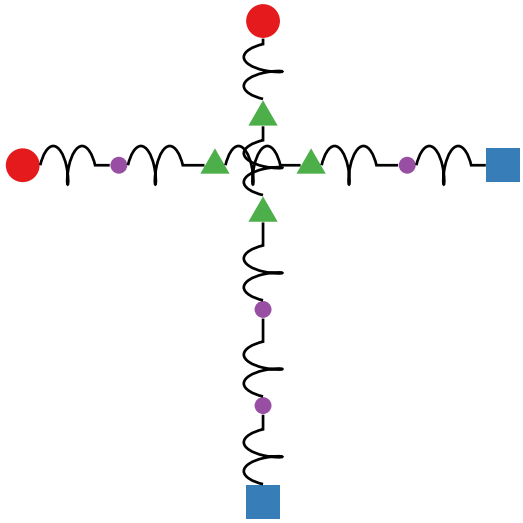
Molecular Tangling Interactions.

A significant failing of this model is its inability to capture the different interactions with NC1 and 7S head groups. If a strict interaction rule is applied such that NC1 can only interact with NC1 and 7S can only interact with 7S, then the resultant structures only have even-membered rings. An example of the frustrated rings that cannot form with strict interaction rules are shown in **Figure 4.3**. This implies that the wealth of different ring networks seen experimentally do not arise directly from 7S interacting with 7S and NC1 interacting with NC1, but instead a more complex interaction pattern that involves polymers becoming tangled with one another. This makes identifying rings in the network more difficult, as any number of polymers could be in a single ring.

One intuitive approach to mimic tangling is to allow body beads to interact with one another instead of purely repelling. If body beads were exclusively attractive, then no network would form as the minimum energy would be all body beads in a large cluster. Instead, some excluded volume can be included by only making a certain fraction of sites p_{S} attractive with a Lennard-Jones potential. Initial simulations varied the number of attractive sites along in the body from 0 % to 100 %, with $\varepsilon_{\mathrm{SS}} = 0.6\varepsilon_{\mathrm{HH}}$. All other parameters remained the same. These simulations revealed a number of significant issues with the attractive sites approach; the first of which being that the network structure rapidly breaks down even for moderate values of p_{S} and clumps, instead of forming a well-defined network. Lateral interactions in 3D with finer-grained beads would alleviate this issue at the cost of significantly more computational complexity. The next issue is that since tangling interactions are inherently a 3D phenomenon, they are hard to model in a 2D network approach. The addition of attractive sites breaks the assumption of planarity, and allows for two molecules to interpenetrate given enough energy. **Figure 4.11** shows the two failure modes of the attractive site model, where polymers either



(a) A network formed with 50.0 % of body beads being attractive; the polymers are clumping together significantly.



(b) A schematic interpenetration defect, breaking the assumption of planarity.

Figure 4.11: Two failure modes of the attractive site model. The left panel shows a network failure mode, where many attractive sites clump together and do not form a well-defined network. The right panel shows an interpenetration defect, where attractive sites do not repel one another enough to maintain planarity of the network.

clump together or interpenetrate.

At low values of p_S where few sites were sticky, not enough body-body interactions occured to significantly change the network structure. At high values of p_S where many sites were sticky, the networks did not form well defined structures. Clumping has been observed in collagen networks (see reference [48]), but requires a more detailed 3D model to study effectively. These failure regimes dominated across a range of different parameter values and as such this approach was deemed unsuitable.

Pre-seeding Lateral 'Tangling' Interactions.

Another approach to include tangling interactions is to 'pre-seed' them by placing pre-tangled molecules into the network and allowing them to self-assemble. Two such pre-tangled molecules

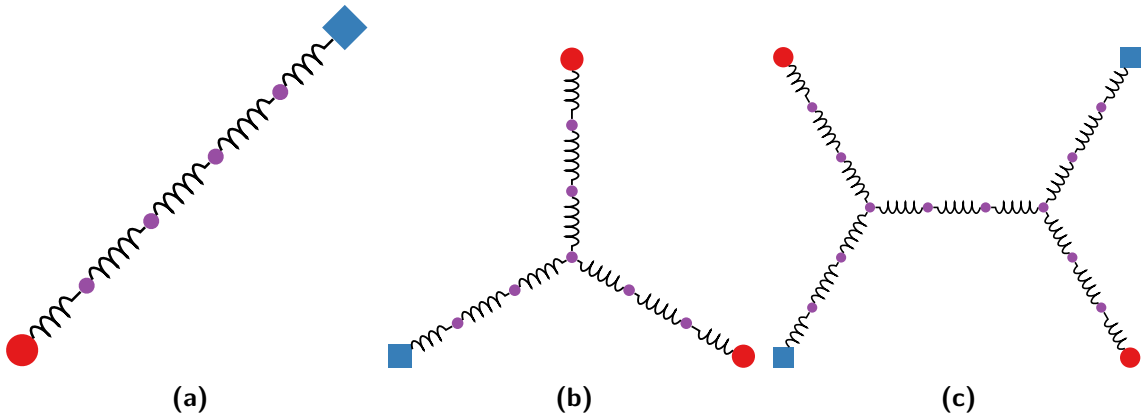
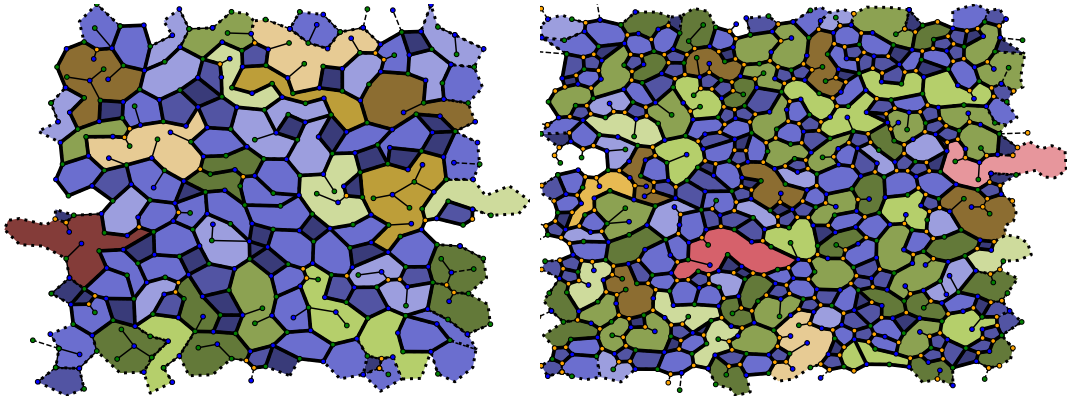


Figure 4.12: Three example molecules used in the simulations, demonstrating the ability to seed a simulation with pre-tangled interactions. Purple circles represent body beads, contributing to U_{BB}, and blue squares / red circles are head beads that contribute to U_{HH}, similar to those shown in **Figure 4.4a**. The gap from a blue/green head bead to a tangling site is $150\,\mathrm{nm}$, which matches a distance between interaction sites observed by Yurchenco & Furthmayr[17].



(a) A network with $5.00\,\%$ of molecules pre-tangled.

(b) A network with $50.0\,\%$ of molecules pre-tangled.

Figure 4.13: Comparison of networks formed with pre-tangled polymers.

are shown in **Figure 4.12**, chosen for their relative simplicity. These molecules were chosen such that they represented multiple polymers tangled together but with distances between interaction sites that matched experimental measurements. The pre-tangled molecules were seeded by replacing straight polymers in the initial lattice structure; up to $25.0\,\%$ of straight polymers were replaced with triangular molecules (**Figure 4.12b**) and up to $25.0\,\%$ straight polymers were replaced with cross molecules (**Figure 4.12c**). These were assembled using the same simulation protocol as before, and two such formed networks can be seen in **Figure 4.13**.

$n_{\text{Fig 4.12b}}/\%$	$n_{\text{Fig 4.12c}}/\%$	$\mu_2(k)$	r	$\langle\text{Area}\rangle\,/\text{nm}^2$
0	2.50	0.443	−0.116	245 473
0	25.0	0.257	−0.082	173 958
2.50	0	0.454	−0.114	247 026
2.50	2.50	0.417	−0.104	235 151
2.50	25.0	0.250	−0.068	203 734
25.0	0	0.281	−0.070	205 269
25.0	2.50	0.295	−0.084	190 625
25.0	25.0	0.215	−0.083	145 209

Table 4.5: Width of the node degree distribution as measured by its second moment $\mu_2(k)$ for simulations in which a certain percentage of linear molecules have been replaced with tangled molecules seen in **Figure 4.12**.

That figure shows that the addition of pre-tangled molecules leads to more rings with small area as the average side length decreases. The addition of pre-tangled molecules also reduces the width of the node degree distribution k; this is because each pre-tangled molecule introduces a new site with a fixed degree of $k = 3$. Data on these phenomena are shown in **Table 4.5**, alongside assortativity data. The addition of pre-tangled sites did not significantly affect the short-range ordering of the ring network, as networks with significantly different ranges of pre-tangled polymers added show very similar assortativities. The assortativities of all of these networks show a slight trend towards disassortativity, in strong contrast to the MC networks seen in **Chapter 3** which were slightly assortative. A full table of data for this is shown in **Appendix C**.

4.3.5 The Network Landscape.

The results in this chapter can be contrasted well with the results of MC networks in **Section 3.4.6**. In that chapter, the networks were driven extremely strongly by entropy, and tended to have assortativities in the range −0.200 to 0.200. Furthermore, the MC networks gravitated towards a wider node distribution than measured experimentally with $\mu_2(k) > 1$ for most networks. This left a gap in the network landscape that the MC networks could not

reproduce, and yet many experimental networks could be found in that region.

A new network landscape using the same metrics for self-assembled networks discussed in this chapter is shown in **Figure 4.14**, with data points from the previous landscape figure (**Figure 3.15d**) shown as shaded regions. Again, the values from experimental data show a wide range of assortativities r but relatively narrow node degree distributions $\mu_2(k)$. Each set of simulations from this chapter is represented as a different set of coloured squares. Compared to the previous landscape figure, there are many fewer data points across a much smaller range. This is because each self-assembly simulation is considerably more expensive to run, so fewer can be computed in the same length of time. The self-assembled networks show a smaller range of assortativities, all being negative, and $\mu_2(k)$ in the range 0.200 to 0.600. This is a much better match for experimental data than the MC method was, and shows a finer degree of control over network structure. Networks successfully formed over a wide parameter range, and there is often a strong link in this figure between r and $\mu_2(k)$ — networks become less assortative as the node distribution widens. That relationship between r and $\mu_2(k)$ can be explained by comparing to the assortativity of a Voronoi structure from a random point process, which is approximately $r \approx -0.150$[113]. The disassortative nature of a random point process comes entirely from geometrical constraints: when both the degree and average sides per ring are fixed, small rings are statistically likely to border large rings[70]. When that constraint on k is relaxed and $\mu_2(k) \neq 0$, the system can approach a considerably more random structure with $r \approx 0$. The data points corresponding to changed head group interaction strengths $\varepsilon_{\mathrm{HH}}$ (orange squares) or adding pre-tangled molecules (pink squares) show a slightly more complicated structure. $\varepsilon_{\mathrm{HH}}$ leads to a clustering of networks in the landscape around $r \approx -0.10$ with $\mu_2(k)$ increasing from 0.250 to 0.600. This is because as $\varepsilon_{\mathrm{HH}}$ increases there is a greater reward for high degree nodes but changing the energetics

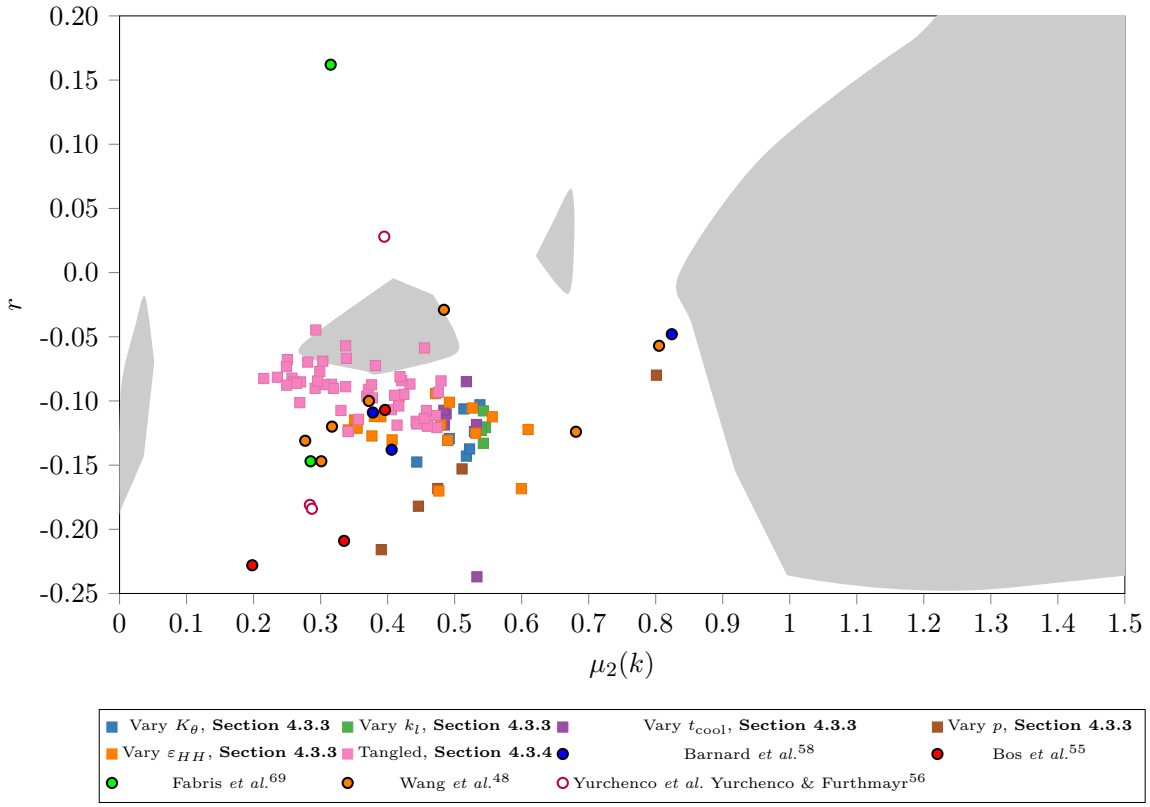


Figure 4.14: A 'network landscape', showing the assortativity r (representing short-range order of rings) against the second moment of the node degree distribution $\mu_2(k)$ (representing the range of node types). The simulated points in coloured squares match experimental data points (circles) better than previous landscape figure **Figure 3.15d**, which are represented as shaded regions.

leads to little change in the medium range order as measured by r. A similar effect is seen by introducing pre-tangled polymers into the mix: a number of points with similar r but varying $\mu_2(k)$, showing how ring structure and node structure can be varied independently.

4.4 Chapter Conclusions.

This chapter has demonstrated a simple worm-like polymer model that self-assembles over a wide range of potential model parameters. The simplified polymer model contains a number of parameters that are physically meaningful and can be varied to control the overall network structure. There is a large and rich range of networks that can be formed by this method, and it can form convex, irregular rings that match experimental collagen networks well.

Changing the potential model parameters to tweak the balance between enthalpy, entropy and kinetics leads to network structures that can match the richness and variability of experimental networks. For example, changing the internal polymer parameters can change the bending and stretching flexibilities of the polymers with consequences for network formation. Networks still formed over a large range of parameters, demonstrating the strong energetic reward for these polymers to self-assemble. Cooling rates, polymer density and head-group change the intra-polymer interactions and have consequences for the medium range ordering of the formed networks. Changing these variables shows that the kinetics of network formation are critical, where polymers become locked into place. If polymers are locked into place too soon or are short of flexibility, no network can be formed.

The polymer model is an improvement upon previous models as it produces meaningful trajectories, and can produce networks with structural metrics that match experimental data better than before. However, it has a few limitations. The first of which is that the polymer model is considerably more expensive than the MC model. The largest MC simulations were of 1024 hexagons, whereas the largest MD simulations produced 90 rings. Furthermore, the potential model parameters are extremely difficult to fit owing to the lack of experimental data. The model parameters used in this section were varied over wide ranges to explore general properties rather than reproduce specific collagen IV behaviour, and a great deal more experimental data would be required to fit the model to reproduce desired numerical properties. This is not the case for the MC model as it has many fewer free parameters. The issues in fitting potential model parameters are discussed in the next chapter, and the rest of this thesis will apply a combination of the polymer and graph-based models to understand the damage and properties of collagen-like networks.



Chapter 5: Mechanical Properties.



The mechanical properties of collagen-like networks are of interest, and the polymer model from the previous chapter is useful to investigate them. Molecular Statics is applied to probe the mechanical response of ideal hexagonal and disordered polymer networks. Bond switching is used to introduce realistic defects in stretched networks. A potential model which incorporates both a quadratic and quartic energy dependence on stretching is fitted to experimental data. The bond-switching and stretching algorithms are combined, and are shown (by direct comparison to experimental data) to mimic real time network aging processes over time-scales as long as 100 years, hugely beyond time-scales available from existing algorithms.



5.1 Introduction.

Mechanical properties of lens capsules are commonly studied from an engineering or medical perspective, and may provide a useful source of data to parameterise computer models. Mechanical data are available for networks made up purely of collagen, such as MatriGel[114,115]. Pure collagen networks are often amorphous, and do not provide a way to study how structure affects mechanical response. The only way to study the mechanical response of structured collagen networks is by studying them in tissue. Experiments to measure mechanical properties such as Young's Modulus involve stretching a section of lens capsule uniaxially, biaxially or

radially[116]. However, the complex nature of living tissue presents some difficulties. Biaxial stretching experiments, such as those performed by Fisher, show a decrease in Young's Modulus of he lens capsule as a function of age[12]. However, uniaxial stretching experiments, such as those performed by Krag & Andreassen, show a stiffening of the lens capsule as a function of age[13]. To add to the problem, radial stretching experiments, such as those performed by Michael *et al.*, showed no change in Young's modulus as a function of lens capsule age[116]. The complex tissues within the eye respond differently to different experimental setups, and can even change depending on how the lens was cut[117].

A computer model that can reproduce these behaviours would be useful to rationalise why they are so different. Furthermore, the experimental mechanical data could then be used to parameterise and improve the computer model. However, the Monte Carlo (MC) method is unsuitable for studying mechanical properties because it introduces non-physical random elements. Similarly, the Molecular Dynamics (MD) method is unsuitable because it is computationally expensive and suffers from thermal noise.

Thankfully by introducing a third simulation method, Molecular Statics (MS), those problems can be eliminated. MS, as first discussed by Vinogradov, can be used to calculate the mechanical properties of an atomic system (with a a special focus on two dimensional (2D) systems)[118]. A simulation powered by MS can re-use the polymer model from **Chapter 4** and calculate properties without introducing thermal noise. The algorithm will be discussed in more detail in **Section 5.2**. Bamer *et al.* and Ebrahem *et al.* have used MS to understand the mechanical properties of silica bilayers[78,119]. Furthermore, those authors combine bond switching methods with molecular statics to generate bilayers that exhibit specific properties[78,119].

In this chapter, potential methods to parameterise the coarse-grained model using ex-

perimental data will be explored. The difficulties in this approach will be examined, and the natural tension between self-assembly simulations and mechanical simulations will be highlighted. After that, the mechanical properties of non-hexagonal networks will be studied using results from previous chapters.

Finally, the aim of the work in this chapter is to combine bond-switched networks with that of experimental mechanical response data on the collagen network as a function of patient age. If possible, this would allow us to to associate the disordered configurations to a time-scale. Associating the damaged networks with real experimental data relies on a reliable parameterisation, which the rest of this chapter will show is difficult to achieve.

5.2 Computational Methods.

5.2.1 Molecular Statics.

Molecular Statics was developed for 2D systems and has been applied to study the defects and failure modes of silica glasses[78,118]. It has some advantages over MD simulation (as used in **Chapter 4**) as MS effectively removes thermal effects. Thermal effects can potentially lead to large error bars in calculation of mechanical properties. Another advantage of MS is that it is considerably more computationally efficient, especially when each step leaves the system close to a local minimum. MS works by repeating the steps in **Algorithm 5** using LAMMPS simulation software. This algorithm is sometimes also referred to as 'athermal quastistatic deformation'[121,122].

The ocular lens capsule contains a mixture of complex biomolecules, including networks of collagen IV and laminin. However, capturing the full atomistic detail of a network of biomolecules is computationally unfeasible[36]. As elsewhere in this thesis, a coarse-grained polymer model is used to examine mechanical properties of collagen-like networks.

Algorithm 5: the Molecular Statics algorithm.

Calculate the initial forces and energies within the system.;
repeat
  Expand the dimensions of the simulation cell by a small factor λ;
  Calculate the forces and the energies within the system according to a given potential model;
  Move the positions of coarse-grained particles in the system to a local energy minimum using the Polak-Ribiere Conjugate Gradient method[120];
  Calculate stresses and strain energy at the new local minimum;
until *ruptured or simulation concludes*;

5.2.2 Molecular Statics Potential Model.

Preliminary investigations revealed that the harmonic bond potentials from **Chapter 4** were unsuitable to reproduce mechanical properties of collagen IV. Collagen IV networks have been shown to become stiffer at higher stretching ratios[13,123]. **Figure 5.3** shows an example of the failure of a purely harmonic model to reproduce experimental data, and will be discussed more fully in **Section 5.3.2**. In order to capture the stiffening effect with the next simplest computational model (which allows for better model transferability and interpretability) a quartic term can be introduced into the potential energy of the model, and a smooth windowing function can be used to gradually move between the 'harmonic regime' and the 'quartic regime'. The windowing function has the effect that the small-extension regime behaves purely harmonically and the large-extension regime behaves purely quartically. The new potential model for bond energies at a bond length r is given by

$$\begin{aligned} U(r) = &\operatorname{erfc}\left[\left|\frac{r - r_{\text{eqm}}}{r_{\text{stiff}}}\right|\right] K_{\text{Quad}}(r - r_{\text{eqm}})^2 \\ &+ \operatorname{erf}\left[\left|\frac{r - r_{\text{eqm}}}{r_{\text{stiff}}}\right|\right] K_{\text{Quart}}(r - r_{\text{eqm}})^4. \end{aligned} \tag{5.1}$$

in which erf is the standard error function defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2}\, \mathrm{d}t \tag{5.2}$$

and erfc is the complementary error function defined as $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$. This acts as a smooth windowing function with values of 0 to 1 for positive arguments. K_{Quad} is a force constant for the quadratic region, K_{Quart} is a force constant for the quartic region, and r_{stiff} tunes the length scale at which the bonds transition from behaving quadratically to behaving quartically.

Chapter 4 showed that the polymer model self-assembled into a collagen-like network over a wide range of parameters[71]. However, one single potential model cannot reproduce both stretching behaviour and self-assembly — they are at significantly different energy scales, with the energies of stretching being up to $10^{10}\times$ larger than the thermal energy at which self-assembly happens. To quantify that, if we were to treat a collagen molecule of length 300 nm as a harmonic spring with a force constant of $k = 1550\,\mathrm{N\,m^{-1}}$ (as calculated by Burd[14]) and stretch it to twice its original length (such that $\lambda = 2$), then this would require 1.39×10^{-10} J. The thermal energy available to atoms at a self-assembly temperature of 300 K is 4.14×10^{-21} J.

5.2.3 Graph to Polymer Conversion.

Disordered networks are easily created using the MC protocol from **Chapter 3**. However, the MD simulations require a system of polymers and not a graph representation of a network. Converting a graph network into coarse-grained polymers is relatively hard, as the graph network contains no information about beads. That extra information about bead positions must be generated heuristically. The heuristics here were designed to place polymers at an energy minimum for a regular network. At each node of degree k in the graph, a cluster of k head beads evenly spaced at a distance r_k from that node. r_k is calculated to be the circumcircle radius of a regular polymer with k sides, and ensures that the cluster of beads is

at an energy minimum. Along each edge, 5 body beads are placed with equal spacing. This may be stretched or compressed compared to the equilibrium arrangement of beads. The process is described in **Algorithm 6**.

Algorithm 6: Algorithm to convert a graph network into a polymer network.

Find the distance l for which two head beads are at an energy minimum;

for *each edge in the network* **do**

Calculate the degree k of the two nodes;

Calculate the radius r_k of a regular polygon with k sides of length l;

Place head beads at distance r_k from each node;

Place body beads evenly spaced between the head beads;

end

Once polymers had been constructed, the bead positions were energetically minimised using LAMMPS's implementation of the Polak-Ribière Conjugate Gradient method[99].

5.2.4 Uniaxial vs Biaxial Stretching.

When stretching a 2D network in a rectangular simulation cell there are two major modes of stretching. Firstly the network can be stretched uniaxially, here in the x direction. This leaves a choice about what to do with the orthogonal y axis. Real materials naturally distort in the direction orthogonal to stretching due to Poisson's effect. The ratio of distortion in the direction of stretching to distortion orthogonal to the direction of stretching is known as Poisson's ratio. Samples of collagen from cartilage have been found to have a Poisson's ratio of ≈ 0.200, but different measurements have been found in the range 0.150 to 0.300 depending on which part of the body the cartilage came from[124]. Poisson's ratio is difficult to apply in simulation, however, as it may not be constant throughout the stretching process (and indeed Ban *et al.* found that it is not constant for collagen)[125].

Deformations of a network can be captured by noting that during uniaxial stretching, the net strain in the orthogonal direction should be zero. To create zero net strain in the orthogonal direction, the simulation cell can be rescaled in only that direction — thereby capturing the Poisson effect. The scaling factor is not known *a priori*, but can be found by a binary search across simulation cell lengths. This is similar to the athermal pressure relaxation strategy developed by Bamer *et al.* and follows the steps in **Algorithm 7**[126]. An example of the binary search is shown in **Figure 5.1**, showing the range of scaling values sampled over the course of a simulation. Since MS does not take the solvent (or a potential inter-polymer matrix) into account, an extra term must be added in the style of Burd[14]. The term due to the matrix is given by

$$E_{\text{matrix}}(\lambda_x, \lambda_y) = \mu \left(\lambda_x^2 + \lambda_y^2 + \frac{1}{\lambda_x \lambda_y} - 3 \right), \tag{5.3}$$

where μ is a free parameter representing matrix stiffness, and the final term of 3 is such that $E_{\text{matrix}}(1, 1) = 0$.

Algorithm 7: Binary search algorithm to find zero stress scaling factor λ_x.

```
Calculate stress in orthogonal direction p_yy;
Set λ_low = 0.1, λ_high = 3.0;
while |p_yy| < tolerance do
    Set λ = (λ_high + λ_low)/2;
    Rescale simulation cell such that L_y = λ L_y(0);
    Recalculate p_yy;
    if p_yy < 0 then
        Set λ_high = λ;
    else if p_yy > 0 then
        Set λ_low = λ;
    end
end
```

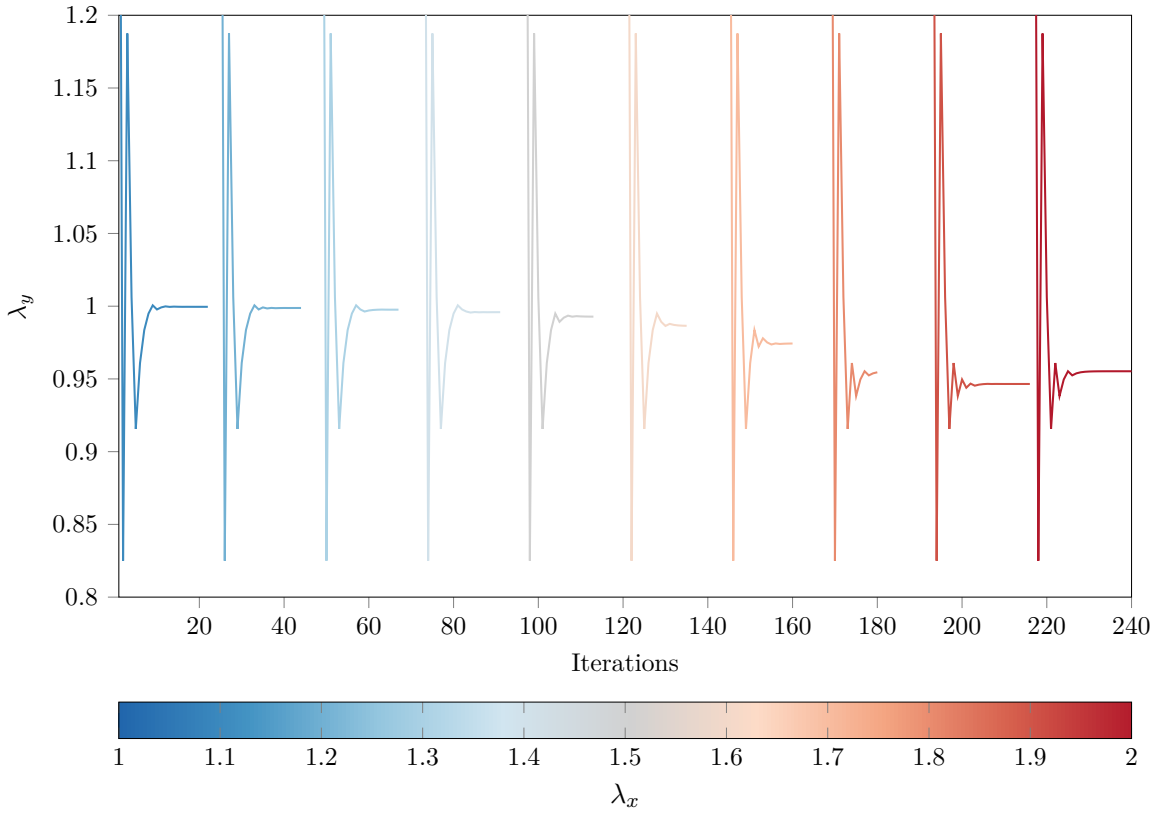


Figure 5.1: Stages of a binary search for λ_y in a uniaxial stretching simulation of a hexagonal network.

5.2.5 Numerical Considerations.

The potential model used for self-assembly was not suitable for reproducing mechanical properties. In order to self-assemble, the inter-polymer interaction energies had to be of a similar magnitude to thermal energy. The intra-polymer energies also had to be of a similar magnitude to thermal energy to be semi-flexible. However, preliminary stretching simulations placed strain energies on the network that were considerably larger than thermal energy. When the strain energy was larger than the inter-polymer interaction energy, the networks suddenly ruptured. The MS simulation method meant that the rupturing process was binary: a hexagonal network was either held together or it was entirely ruptured. Rupturing prevents the calculation of mechanical properties as all stresses are suddenly relieved, so rupturing

fell outside the scope of this chapter. Instead, rupturing will be discussed in more detail in **Chapter 6**.

To prevent the networks breaking down, the head group interaction strength needs to be effectively infinite. Sadly, extremely strong interactions present numerical issues. The MS procedure locally minimises the energy of the system numerically, until the energy between steps i and $i+1$ has converged to a tolerance, E_{tol} with equation $\left|\frac{E_{i+1}-E_i}{E_i}\right| \leq E_{\text{tol}}$. If the head group interactions are extremely strong, then they dominate E_i and the minimisation procedure only optimises the head groups, but not the bonds. The tolerance parameter E_{tol} controlled the maximum ratio of head groups strengths to bond strengths. The ratio cannot be too large before numerical precision errors arise. If the tolerance parameter was of the order of the floating point precision, $\varepsilon_{\text{double}} = 2^{-53} \approx 10^{-16}$, then the system could never converge. The tolerance parameter limited the range of K_{Quad} values available, as if the range of K_{Quad} was too large then the lowest values could not be minimised while the largest fractured. As a result, head group energy scale of $\varepsilon_{\text{HH}} = 10^{-11}\,\text{J}$ was chosen for this chapter (up from $\mathcal{O}\left(10^{-19}\,\text{J}\right)$ for previous chapters).

5.3 Results and Discussion.

5.3.1 Uniaxial vs Biaxial.

The first test of MS was to see if using it with the polymer model could reproduce mechanical properties. A commonly measured mechanical property is the stress-strain curve; in a 2D network undergoing stretching the closest analogue is a traction-stretching ratio curve. It is simpler to describe the traction-stretching ratio curve as a '2D stress-strain curve' for easier comparison with results for three dimensional (3D) networks in the literature. The highest quality experimental data available are from Krag & Andreassen, who studied uniaxial

stretching of lens capsules from patients of different ages[13].

Simulating uniaxial stretching with the polymer model presented difficulties as the network could undergo a traction-relieving rearrangement. The rearrangement is similar to a mechanical trellis opening, where all traction can be relieved by changing the angle between polymers with minimal extension; this is shown as inset figures **Figure 5.2**. The alternative is to stretch the network biaxially, in which both dimensions are stretched equally. For complex networks like collagen IV, other coarse-grained models have shown that the semiflexible nature of the polymers leads to strong coupling between stretching axes and a significant stiffening for biaxial stretching compared to uniaxial stretching[125]. This may be the cause of the difference between different measurements of lens capsule mechanical properties. The strong coupling and stiffening effects of these networks leads to extreme difficulty in effectively parameterising the coarse-grained polymer network used elsewhere in this thesis.

Biaxial stretching simulations using the polymer model had desirable computational properties when compared to uniaxial stretching simulations. These desirable properties include:

- An easier calculation of true stress (as calculating true stress over conventional stress for uniaxial stretching requires a correction for the change in area[127]).
- No need to impose an unknown and potentially varying Poisson's ratio on the material.
- Resistance to concerted stress-relieving rearrangements in idealised hexagonal network.

Uniaxial (expanding in x with y either held constant or calculated to have zero stress in that direction) and biaxial stretching simulations (expanding in x and y) were performed for an idealised hexagonal network using the MS method. In both forms, networks were stretched by a multiplicative factor of $\lambda = 1.01$ over 100 steps for a total stretch ratio of 1.91. 2D

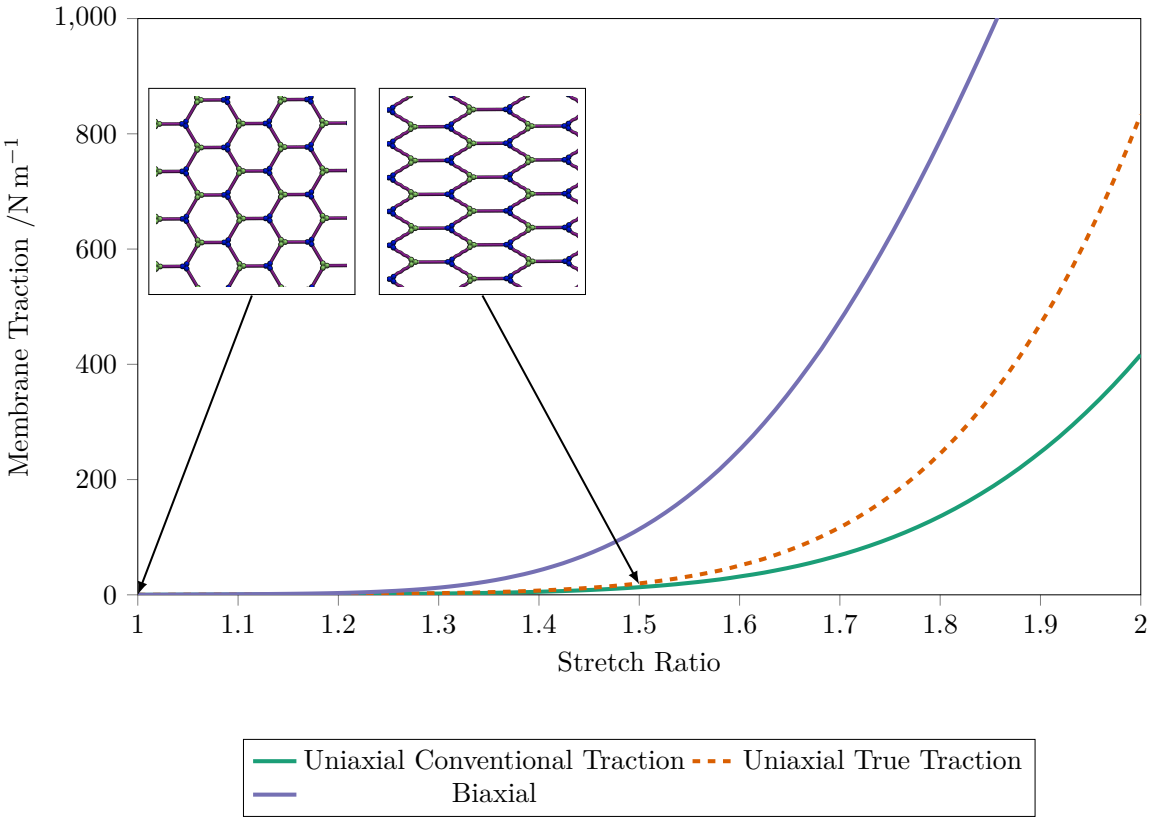


Figure 5.2: Identical hexagonal networks stretched either uniaxially or biaxially. The two curves show broadly similar behaviour, except that the uniaxially stretched network experiences almost zero traction (2D stress) up to $\lambda \sim 1.40$. This is because the uniaxially stretched network can elongate along the stretching axis in a rearrangement similar to a trellis opening, thereby relieving all the traction. Two snapshots of the trellis opening motion at the same scale are shown as inset figures.

stress-strain curves are shown in **Figure 5.2** using parameters from **Section 5.3.2**.

The limitations of the extremely simple polymer model in reproducing uniaxial experimental findings are clear. However, the trends in mechanical properties are still worth exploring. Biaxial simulated data will be compared to uniaxial experimental data to understand the effect of defects or the trends in amorphous networks, with an appropriate amount of caution taken in the interpretation.

5.3.2 Fitted Parameters.

Best values for the remaining polymer parameters, *i.e.* K_{Quad}, K_{Quart} and r_{stiff} in **Equation 5.1**, were assigned by fitting them using a least squares process to match the stresses observed by Krag & Andreassen[13]. Biaxial stretching with different parameters were compared the simulated tractions (effectively 2D stresses) to the 'Age 1' data produced by Krag & Andreassen[13]. The parameters were optimised to minimise the squared difference between the data from Krag & Andreassen and the calculated data. Experimental data were manually digitised and linearly interpolated between digitisation points to match the resolution of the simulated data. This lead to $r_{\text{stiff}} = 59.8\,\text{nm}$, $K_{\text{Quad}} = 39.3\,\text{N}\,\text{m}^{-1}$, $K_{\text{Quart}} = 1.64\times10^{18}\,\text{N}\,\text{m}^{-3}$. The resultant value for K_{Quad} is smaller than was used by Burd[14], due to their using a different potential model.

Figure 5.3 shows the resultant 2D stress-strain curve using the fitted parameters against lens capsules of various ages as measured by Krag & Andreassen[13]. This figure also highlights the effect of omitting the quartic term in the potential model. For the ideal hexagonal network, a purely quadratic potential energy model yields a linear dependence and so can only effectively reproduce the experimental data at low stretching ratios.

The gradient in the $\lambda = 1.00$ to 1.20 region of the 2D stress-strain curves represents an elastic modulus, and has a value of $9.23\,\text{N}\,\text{m}^{-1} \pm 1.05\,\text{N}\,\text{m}^{-1}$. If the 2D networks were extended to a 3D strip of 17.0 µm depth, the 2D elastic modulus would correspond to a 3D Young's Modulus of $0.543\,\text{MPa} \pm 0.0615\,\text{MPa}$. For a material that behaves non-linearly, the Young's Modulus must be interpreted carefully, and is only valid in the small extension region. The small extension region with strains of up to 1.10 matches the strains expected to be observed across normally accommodative motion[12,128]. Regardless, the simulated elastic modulus fits in the range of expected elastic moduli which have been recorded to be between

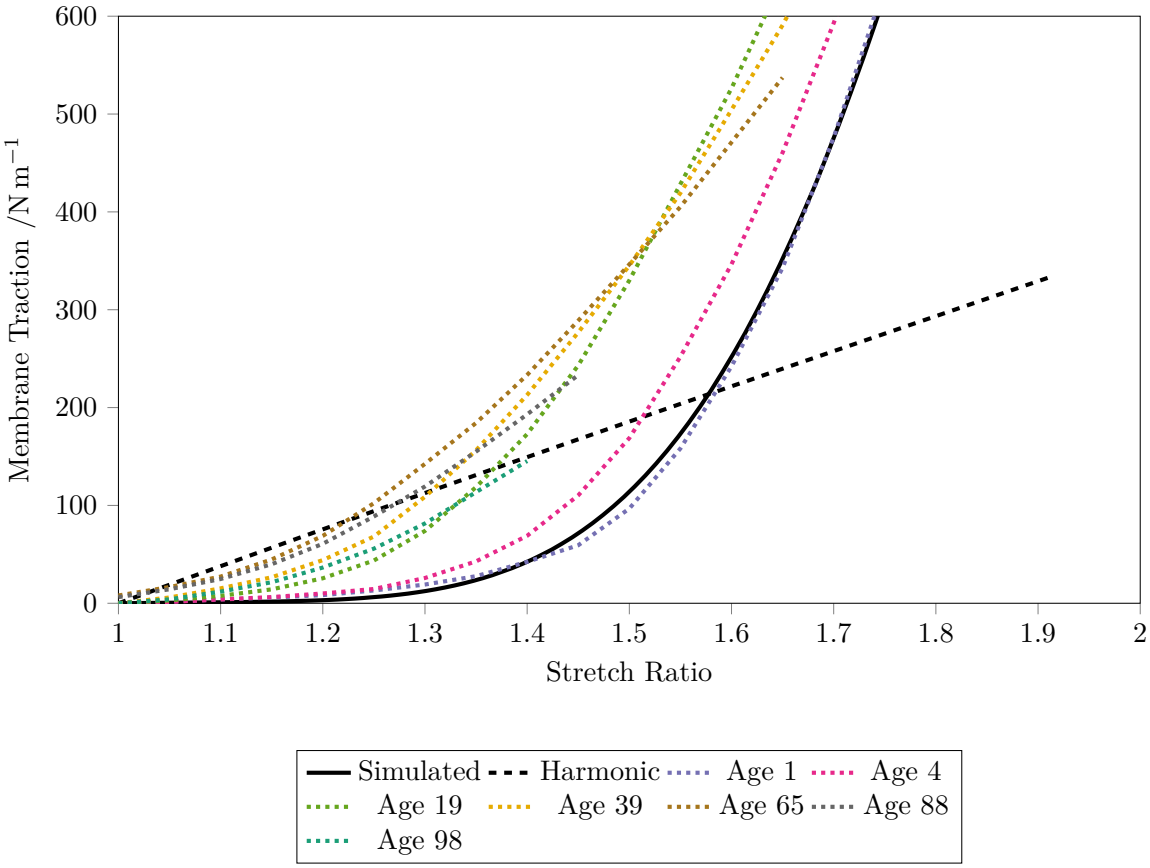


Figure 5.3: Traction-stretching ratio curves for the polymer model *vs* experimental data. Dashed lines are uniaxial experimental data as measured by Krag & Andreassen of lenses from people of different ages (shown in the legend in years)[13]. They were corrected from measured pressures to tractions by adjusting for twice the average width of a lens strip, taken to be 17.0 µm — see discussion by Burd[14]. The black dashed line represents the best fit using a solely quadratic potential.

0.300 MPa to 2.30 MPa by Krag & Andreassen and 0.250 MPa to 0.400 MPa by Michael *et al.*[13,116]. The elastic modulus of the lens capsule perpendicular to the 2D network plane (not tested in this 2D simulation) was found to be smaller by Last *et al.*, who measured an elastic modulus of 7.50 kPa[129].

Table 5.1 shows the potential model parameters that best fit each of the lens capsules measured by Krag & Andreassen[13]. The stiffening behaviour of the lens capsules is captured in the decreasing value for r_{stiff} and an increasing value for K_{quad}. However, the 2D stress-strain curves fit poorly for the older lens capsules. This is because the shorter r_{stiff} means the

Age	$r_{\text{stiff}}/\text{nm}$	$K_{\text{quad}}/\text{N m}^{-1}$	$K_{\text{quart}}/1.00 \times 10^{18}\,\text{N m}^{-3}$
1	59.8	31.6	1.64
4	46.9	268	1.61
19	25.4	478	2.07
39	2.06	11 400	1.99
65	3.13	26 500	2.09
88	1.80	30 200	2.44
98	0.748	28 400	1.96

Table 5.1: Fitted parameters for different sets of experimental data (categorised by age). For the older lens capsules, the fitting procedure is only weakly sensitive to a change in parameters; a range of parameters give fits of similar quality. This is because a lower r_{stiff} leads to the energy function becoming dominated by the quartic term.

polymers behave quartically for a large range of stretching ratios. The idealised hexagonal network is a poor model for the collagen network in an older lens capsule, as will be discussed in **Section 5.3.4**.

The polymer model gives excellent control over the rupturing properties of the network. By changing the head group attraction parameter ε_{HH}, the stretching ratio at which networks rupture can be controlled. MS only allows for an instantaneous rupture in a high-symmetry network, which is a poor description of real rupturing. More disordered networks can rupture partially to relieve stress, which will be seen in **Section 5.3.3** and **Section 5.3.4**. Furthermore, detailed studies of rupturing will be in **Chapter 6**.

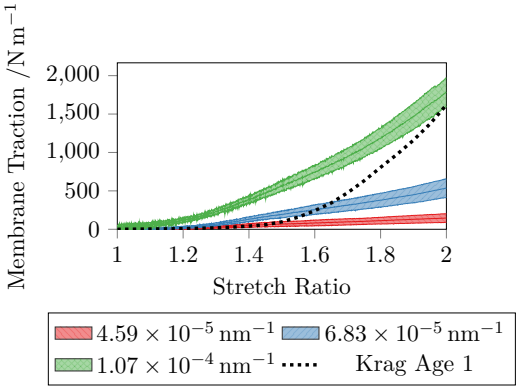
5.3.3 Disordered Networks.

Given that idealised hexagonal networks are a poor match for the collagen network in old lens capsules, a better network structure is required. **Chapter 3** and **Chapter 4** generated a useful library of disordered networks via bond switching and self-assembly. Those networks in those chapters were constructed using a different potential model, so transitioning to a new potential model may mean that the bead positions are not at an energy minimum. Network structures can be adapted to the new model by repeated minimisation, where the parameters

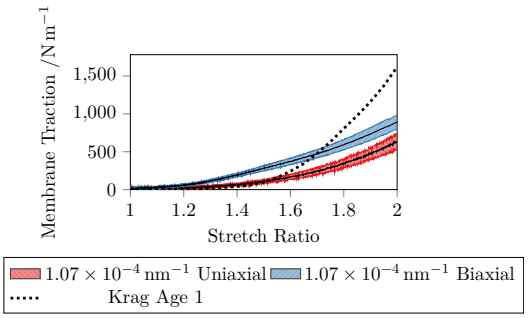
are scaled by a small multiplicative factor λ_{param} after each minimisation. Once $\lambda_{\text{param}} = 1$ and the bead positions were at a local minimum of the new model, the networks were stretched using the molecular statics procedure described in **Section 5.3**. **Figure 5.4a** shows the 2D stress-strain ratio for self-assembled networks formed at different densities. Examining the mechanical properties of a single network can be misleading, as mechanical properties are very sensitive to small numerical differences. Furthermore, the potential energy surfaces are not smooth, which can lead to minor artefacts when minimising. As a result, ten different network configurations were sampled at each density and stretched using the same procedure. The results show that looser networks respond less stiffly to stress, which is an intuitive result. The looser networks have more empty space and lower degrees (k) at the nodes, which allows the network to reorganise in response to a strain. This interpretation is reinforced by experimental data comparing Matrigel (which contains a disordered collagen and laminin molecules not in a network) to the human lens capsule. Those experiments found that the Matrigel responds much less stiffly to stretching than the lens capsules[115,129]. The amorphous networks show an almost linear response at high stretching ratios which has been observed experimentally, and the same linear response is observed in **Figure 5.5a**[13,14,127].

Finally, the denser networks cannot reorganise and instead respond more stiffly as the polymers are stretched. The stiffening increased at higher stretching ratios, because the polymers oriented themselves in the direction of the strains[14]. This is one of the reasons for the difference in biaxial and uniaxial mechanical response. When stretched uniaxially, the polymers can orient themselves with or against the stretching, and can bend to reduce stress. This is not possible during biaxial stretching.

To reconcile the difference between the data measured by Fisher, and Krag & Andreassen, Burd suggested ‘a fundamental difference in the behaviour of the capsule when tested in



(a) 2D stress-strain relationships for self-assembled networks generated at different network densities. The networks that were formed with denser initial conditions responded more stiffly to stretching. Each line is averaged over ten self-assembled networks, with the filled area representing one standard deviation either side of the mean.



(b) 2D stress-strain curves realised across ten samples of disordered networks at a given molecule density. The networks were stretched uniaxially (along x at constant y) or biaxially (along both x and y). There was no significant difference in behaviour for networks stretched uniaxially or biaxially.

Figure 5.4: 2D stress-strain curves for disordered networks, utilising a library of self-assembled networks created in previous chapters.

uniaxial and biaxial modes'[14]. To test this, uniaxial stretching of the disordered networks were compared to the biaxial stretching in **Figure 5.4a**. The results for networks at one density are shown in **Figure 5.4b**. In this model of the lens capsule, there was no significant difference in the generic network behaviour of uniaxial and biaxial stretching in the disordered networks. This similarity supports using the biaxial stretches in order to understand the underlying generic mechanical behaviour. Furthermore, it highlights the limitations of using a chemical model for an engineering problem.

5.3.4 Bond Switching Snapshots.

The polymer model was used to calculate detailed mechanical properties of the networks generated in **Chapter 3**. The stretching induced defects are of particular interest because they mimic the aging related changes of the eye. To introduce stretching defects, graph based networks were stretched uniaxially at constant volume while undergoing the bond switching

procedure, generating snapshots at regular intervals. The snapshots were then rescaled back to their original unstretched aspect ratio as discussed in **Section 3.2.5**.

The graph-based networks output from the bond switching simulations were converted into polymer networks, as was shown in **Algorithm 6**. However, as the purely graph based network in a bond switching method is more coarsely-grained than the already very coarsely-grained polymer networks, assumptions must be made on the positioning of beads in **Algorithm 6** The assumptions to convert graphs to polymer are most effective for a more regular network. When there are many defects in the network, the conversion algorithm leads to sudden rearrangements, and bonds that do not participate in the wider network (referred to as dangling bonds, with degree $k = 1$).

The networks with few defects responded more stiffly to stretching than the idealised hexagonal networks. The stress-strain curves for a single network undergoing the stretching process are shown in **Figure 5.5a**. Networks with many defects were more likely to experience individual nodes separating to relieve stress; this is visible as sharp drops on the 2D stress-strain curves. When there were very many defects, the network became considerably looser than the idealised network.

As with the self-assembled networks, the looser bond switching generated networks also responded less stiffly to stretching. These networks often showed a near-linear response to stretching, matching that observed in **Section 5.3.3** for highly disordered networks. This is partly because the disordered networks could relieve stress relatively easily by rearranging and breaking connections, which is not possible for a high-symmetry idealised network as was discussed in **Section 5.3.2**. A slice through the stress-strain curves at a constant stretching ratio of $\lambda = 1.15$ is shown in **Figure 5.5b**. This figure demonstrates the initially flat region where the networks are broadly hexagonal, followed by a rising region as defects

cause stiffening, followed by a drop-off as the networks break, as measured by the fraction of dangling bonds on the alternate axis in **Figure 5.5b**. This shows a flat region in the traction curve where there the networks haven't been aged enough to cause defects (shown by a low fraction of dangling bonds), followed by a steep rise as the defects cause a stiffening effect. For networks that have undergone many aging steps, the traction at $\lambda = 1.15$ drops off significantly due to the networks falling apart, as measured by a high fraction of dangling nodes.

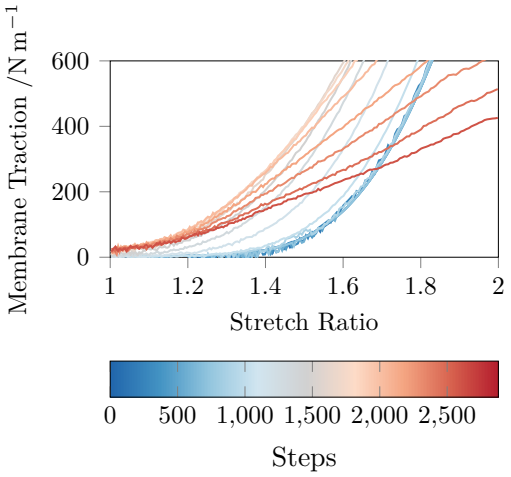
The blue line that has undergone zero aging steps in **Figure 5.5a** corresponds to the Age 1 line in **Figure 5.3**. The addition of few defects causing network stiffening demonstrates that the stretching-favoured defects function as a reasonable model for aging, as the resultant networks mimic the properties of older lens capsules also seen in **Figure 5.3**. **Table 5.2** contains data on the real lens as measured by Krag & Andreassen that best match the simulated aging, averaged over 25 networks[13]. These show a good fit as measured by R^2 of the simulated aging to the real data, increasing monotonically as more defects are formed in the simulated networks. Under these simulation conditions, a linear regression between aging steps and the corresponding age of a real lens capsule provides a slope of 20.9 ± 2.15 aging steps per year. This is specifically for the networks of 1024 rings studied here, and would generalise to other systems at a rate of 0.0479 aging steps per ring per year.

The quality of the match decreases as the age increases because the potential model used to form the defects is too simplistic to capture the detail of the enthalpic effects in the real lens capsule. Future work could improve the introduction of defects from the simple model presented here to better match the aging process, and ensure that an engineering model was used to calculate mechanical properties. However, given the relative simplicity of the coarse-grained model and the bond switching algorithm, the ability to reproduce generic

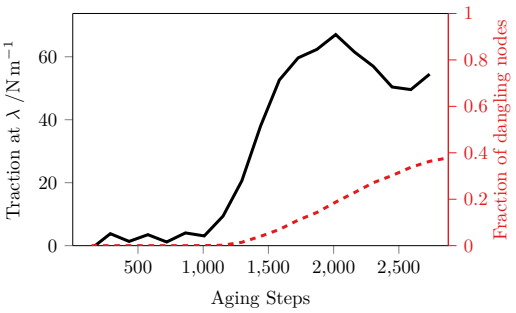
Aging Steps	Best Fit	R^2
288	1.00	0.916
432	1.00	0.905
576	1.00	0.905
720	1.00	0.905
864	1.00	0.915
1010	1.00	0.966
1150	4.00	0.998
1300	19.0	0.994
1440	19.0	0.994
1580	19.0	0.977
1730	19.0	0.978
1870	39.0	0.983
2020	65.0	0.993
2160	65.0	0.983
2300	88.0	0.992
2450	88.0	0.974
2590	88.0	0.947
2740	88.0	0.946

Table 5.2: The correspondence between the 2D stress-strain curves after the bond switching aging process and the matching 3D stress-strain curves of lens capsules as measured by Krag & Andreassen. The fits for the networks that have undergone few aging steps are worse as the networks have slightly distorted node positions to accommodate the stretching (discussed in **Section 6.2.1** and **Section 5.3.4**), but have not yet switched any bonds.

network aging across approximately 100 years of real-time network degradation is remarkable.



(a) The 2D stress-strain curves for repeated snapshots of a single network. Each snapshot represents stretching a single configuration of a network after a given number of MC moves.



(b) A slice through **Figure 5.5a** at a constant stretching ratio, here $\lambda = 1.15$ (black line, black axis) and the fraction of dangling nodes, which is a measure of network decay (red dashed line, red axis).

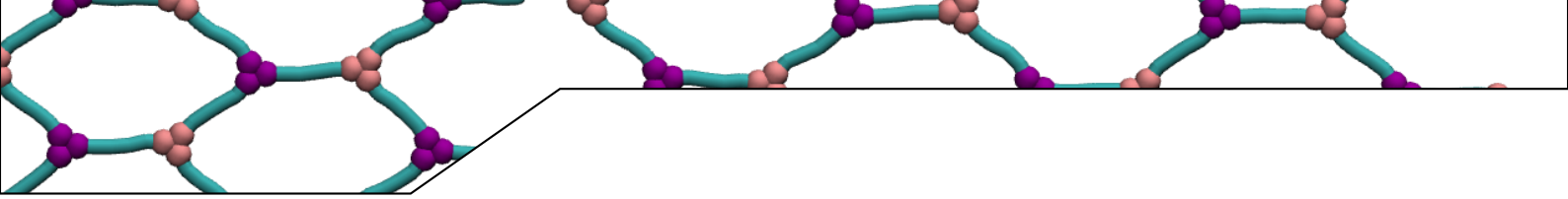
Figure 5.5: 2D stress-strain curves for snapshots of networks that had defects introduced via a bond switching process.

5.4 Chapter Conclusions.

This chapter detailed the mechanical response of generic biological networks. The coarse-grained model from **Chapter 4** combined with a MS approach allowed for relatively large length-scales to be accessed, well beyond the range of atomistic models. It also allows more direct contact with engineering-centred experimental measurements, albeit with some difficulty. Both ordered (hexagonal networks) and disordered (amorphous) materials were investigated, with configurations for the latter created with the self-assembly or bond switching methods detailed elsewhere in this thesis.

An attempt was made to parameterise the model by reference to detailed experimental measurements of collagen IV in the lens capsule. This turned out to be difficult, as the polymer model used here was too simplistic to capture mechanical response under uniaxial stretching. The difference in energy scales between mechanical stretching energy and self-assembly energy

indicated that the collagen network itself cannot fully explain the stretching phenomena observed in the human ocular lens capsule. Despite this, a parameterisation was established using biaxial stretching data, and this was compared to mechanical properties of lens capsules of different ages. Defects were systematically introduced by a combined bond-switching and system stretching algorithm. This showed a promising match for network aging effects observed experimentally, which are observed on the time-scale of around a century in human lens capsules.



Chapter 6: Rupturing.



In this chapter, the way that polymer networks rupture when stretched is explored. Hexagonal networks, both idealised and with defects, are stretched biaxially and uniaxially, with linear and sinusoidal expansion rates. The normal modes of defect containing networks are calculated, and the balance between structural strength, localised stress, and flexibility are discussed. Large holes and strips of finite width are studied to mimic surgical damage to collagen IV in the lens capsule, which are used to demonstrate the importance of flexibility. Damage is also introduced by adding kinetic energy to a single polymer, and the dispersal of energy through the network is tracked. This chapter is adapted from M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180[130].



6.1 Introduction.

The previous chapter on mechanical properties discussed the limitations of using Molecular Statics (MS) to study damage. Using MS, an idealised hexagonal polymer network was either fully ruptured or it was not. To avoid that issue, the potential model used in **Chapter 5** was tweaked to entirely prevent rupturing. However, the way that collagen IV networks undergo damage is an interesting topic in itself.

Injuries to the eye, including sporting or surgical injuries, can result in catastrophic rupture of the lens capsule[131]. The medical community is particularly interested in posterior

capsule rupture (PCR) as a consequence of cataract surgery[132]. PCR occurs in about 4.10 % of cataract surgeries, and leads to significant loss of visual acuity[133]. If the mechanism of rupture could be understood, it could help surgeons better predict when a surgery is likely to go awry.

The mechanism of rupture has previously been studied using engineering methods[134]. Finite element analysis is a common tool in that field, and is excellent at predicting the distribution of strains throughout a material. For example, Liu *et al.* used finite element analysis to examine the distribution of strains through a deforming lens capsule when impacted by objects travelling at different velocities[134]. The likelihood of PCR increases significantly with age, with one study estimating that a 70 year old patient has an 210 % increase in risk compared to a 50 year old patient[135]. This fits into the pattern of the lens capsule getting more brittle with age. Finite element techniques cannot study the molecular origins of that change, so a chemical model is required.

To understand the dynamics of rupture, we must return to the Molecular Dynamics (MD) approach from **Chapter 4**. Instead of starting with loose polymers that form a network, simulations in this chapter will start with ordered networks and break them. The MD approach will be extended to allow breaking both within and between polymers. In addition, ruptures may originate from weak points in the network. The effects of whether bond switching defects (introduced in **Chapter 3**) encourage or inhibit rupture will be examined. Bond switching defects have been shown to have complex effects on the rupturing behaviour of silica networks. This is especially true when there are multiple defects, as the loss of tensile strength depends on their number and relative arrangement[136].

While bond switching defects are appropriate for some types of damage, they are not the only form of damage found in collagen networks. In addition, the effect of removing edges

must be examined. This includes both removing edges one at a time from the network (to introduce a progressive weakening), and introducing large-scale damage. Surgical damage to the lens capsule occurs on a macroscopic scale, much larger than a single bond switching or edge removal defect. The simplest form of large-scale damage is to remove many edges around a central node to introduce a large hole. The limiting case of a large hole is where the boundary of the hole becomes infinitely large, leaving only a thin strip of network.

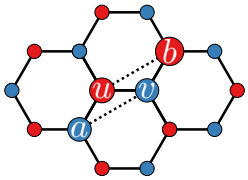
6.2 Computational Methods and Metrics.

6.2.1 Defect Introduction.

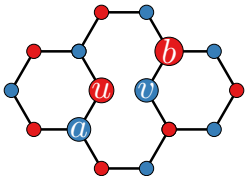
Studies of rupturing in ordered networks are useful, but **Chapter 2** showed that collagen IV networks in living issue are not ordered. Damage processes may be linked to the short- and medium-range ordering within a network, best represented by the ring structure. For example, the presence of a defect may introduce a weak point in the network, or the added flexibility may inhibit rupture. To explore how damaged networks rupture when stretched, a consistent approach to creating initial damage must be designed. For simplicity, work in this chapter will start with an idealised hexagonal network and incrementally introduce defects. The two main types of defects that will be used are the Wooten-Winer-Weaire (WWW) bond switch (first discussed in **Section 3.2.4**) and edge removals. A WWW defect is shown in **Figure 6.1a**, and a bond removal defect is shown in **Figure 6.1b**. Combining these defects can lead to extremely complicated network structures (as seen in **Chapter 3**), so a Monte Carlo (MC) process is used to reject moves that do not meet certain constraints.

Any defect introduction that causes one of the following criteria to be broken is rejected:

- A non-planar network is formed,



(a) A schematic bond switching step; the edges (a, u) and (v, b) are deleted and the new dashed (a, v) is deleted. This combines the central two edges (a, v) and (u, b) are added. The colours further highlight the defect with two blue-red pairs being replaced by blue-blue and red-red pairs. This transforms four hexagons into two heptagons and two pentagons.



(b) A schematic bond switching step; the edge (a, v) is deleted. This combines the central two edges (a, v) and (u, b) are added. The colours hexagons into a distorted decagon.

Figure 6.1: Two schematics of damage types to a section of ordered hexagonal network. After each move, the network geometry was re-optimised.

- Any node has a degree $k \leq 1$,
- The change in energy ΔU is too great,
- A ring with fewer than 3 or more than 8 edges is formed.

The change in energy due to a move is assessed with the DREIDING-inspired potential similar to that used elsewhere in this thesis[82]:

$$U = \underbrace{k_{\text{bond}}(r - r_{\text{eqm}})^2}_{\text{Bonds}} + \underbrace{K_{\text{angle}}(\cos(\theta) - \cos(\theta_{\text{eqm}}))^2}_{\text{Angles}} \tag{6.1}$$

The form of the angular part of the potential was chosen with $\theta_{\text{eqm}} = \frac{2\pi}{k}$ where k is the degree of that node. Bond and angular force constants k_{bond} and K_{angle} were chosen to be 1 in the defect introduction stage, and $r_{\text{eqm}} = 1$.

As the defect introductions disturb the equilibrium geometry of the network, the node positions must be moved to a new energetic minimum. In this chapter, the energetic minimisation was performed using a Broyden-Fletcher-Goldfarb-Shanno (BFGS)[137–140] routine implemented in SCIPY[141]. The relatively simple form of the potential is advantageous for finding the energetic minimum, as an analytical derivative can be calculated for the bond component

$$\frac{\mathrm{d}U}{\mathrm{d}\boldsymbol{r}} = 2k_{\text{bond}}(r - r_{\text{eqm}})\frac{\boldsymbol{r}}{|\boldsymbol{r}|}, \tag{6.2}$$

and the angular component

$$\left(\frac{\partial U}{\partial \boldsymbol{r}}\right)_{|r|} = 2K_{\text{angle}}\left(\cos(\theta) - \cos(\theta_{\text{eqm}})\right)\begin{bmatrix} \frac{\cos(\theta)x_{ab}}{|\boldsymbol{r_{ab}}|^2} - \frac{x_{ac}}{|\boldsymbol{r_{ac}}||\boldsymbol{r_{ab}}|} \\ \frac{\cos(\theta)y_{ab}}{|\boldsymbol{r_{ab}}|^2} - \frac{y_{ac}}{|\boldsymbol{r_{ac}}||\boldsymbol{r_{ab}}|} \end{bmatrix}. \tag{6.3}$$

In this equation the 'angular' derivative is calculated at a constant bond length $|r|$, where x_{ab} is the x component of the vector between sites a and b (and similarly for y_{ac}). As before, r_{eqm} is the equilibrium separation of the beads, k_{bond} is the bond force constant and K_{angle} is the angular force constant. Knowledge of the derivatives allows construction of a Jacobian matrix which greatly accelerates finding the minimum energy positions of the entire network. This means that the entire network can be optimised, as opposed to the local optimisation used in **Chapter 3**.

6.2.2 Polymer Model.

Detailed studies of collagen IV backbones with atomistic models have posited two breaking modes: a dissociation between head groups in a network, and a rupture along the backbone of a single molecule[142]. To achieve this, the polymer model introduced in **Chapter 4** was re-used (where head beads attract with a Lennard-Jones potential, body beads repel, bonds interact harmonically and angular potentials keep the polymer semi-flexible). However, the harmonic bonds in that model are non-reactive and do not allow for a backbone to break. The simplest change to create a reactive force field is again inspired by the DREIDING potential, and that is to replace the harmonic bond interaction with a Morse potential[82]:

$$U(r) = D_e[1 - \exp(-\alpha(r - r_{\text{eqm}}))]^2 - D_e, \tag{6.4}$$

with $D_e = 24.0 \times 10^{-21}\,\text{J}$, $r_{\text{eqm}} = 50.0\,\text{nm}$ and $\alpha = 0.150\,\text{nm}^{-2}$ (unless otherwise specified). The Morse potential was chosen as it is quadratic in the small stretching regime, resistant to compression and breaks at large values of r_{eqm}. A comparison of the two potentials is shown

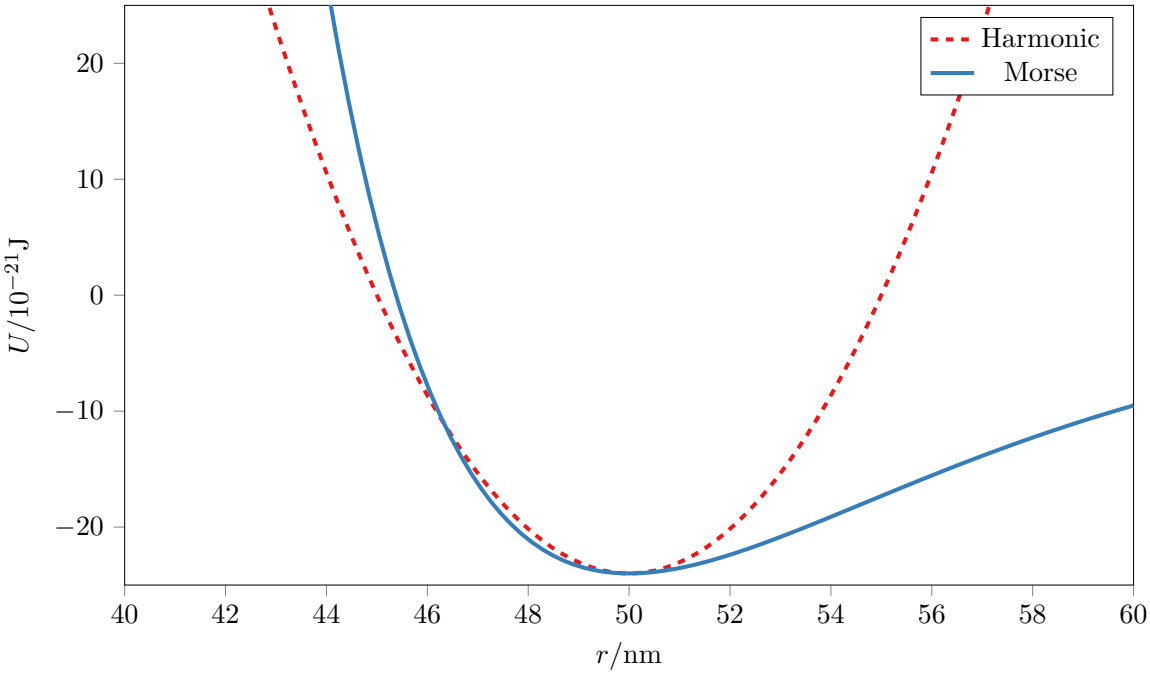


Figure 6.2: A comparison of shifted harmonic and Morse potentials used in this section. The shifted harmonic potential is **Equation 8.1** with appropriate parameters, and the Morse potential is **Equation 6.4**.

in **Figure 6.2**.

At every simulation step, the lengths of all bonds in the simulation were calculated and any bonds with a length greater than a cut-off $r_\mathrm{c} = 138\,\mathrm{nm}$ were deleted. This is valid as the Morse potential has the convenient property that $\lim_{r\to\infty} U(r) = 0$ and $\lim_{r\to\infty} \frac{\mathrm{d}U}{\mathrm{d}r} = 0$. Given that the forces and energies are effectively zero at large separations, the bonds can be broken without creating significant discontinuities. The cut-off distance must be chosen carefully to not only be sufficiently large to avoid energy discontinuities, but also not to trigger body bead repulsions. In LAMMPS there are no pairwise interactions between beads within the same molecule. However, when a bond is broken, the two fragments are treated as different molecules by the simulation software. If the cut-off distance is smaller than the body-body interaction distance, $r_c < \sigma_\mathrm{BB}$, then a broken bond would immediately cause the fragments to repel one another strongly.

The final issue is the angular part of the potential, which does not tend to zero as r tends to ∞. This means that if a group of 3 atoms a, b, c are arranged such that $\theta_{abc} \neq \theta_{\text{eqm}}$, then there is a discontinuity in the energy when bond ab is broken and the angular potential is deleted. Investigations of the total system energy over time suggest that this is not a problem for the semi-flexible polymers used in this chapter. The coarse-grained polymer model has a relatively large value for K_{angle}, leading to approximately straight molecules. This is especially true when molecules stretch under tension, as they are held straight, thereby avoiding issues with the angular potential.

6.2.3 Stretching Until Rupture.

Once a network is been formed, it can be forced to rupture by stretching it beyond its natural limits. Experimental studies of lens capsules suggest a maximum breaking strain of 100 % for young lenses and 40.0 % for old lenses[127]. This is similar to the maximum breaking strains of single collagen IV molecules studied by Buehler which were in the range 40.0 % to 60.0 %[142]. The similarity of the breaking strains for overall network damage and individual polymer damage indicates that the two failure modes are finely balanced.

Three different stretching regimes were applied to the networks in order to induce damage. The first stretching regime was to simulate the system undergoing a constant engineering strain rate $\frac{\mathrm{d}L_x}{\mathrm{d}t} = \lambda$ by rescaling the simulation cell (and atomic positions) by a small factor every step. The second stretching regime was sinusoidal, rescaling the simulation cell length L_x according to the equation

$$L_x(t) = L_x(0)\,[1 + A\sin(\omega t)] \tag{6.5}$$

in which A is the amplitude of stretching and ω is frequency of stretching.

In order to study cumulative damage, the sinusoidal damage equation was modified to

vary the amplitude A over time, increasing every cycle with time:

$$L_x(t) = L_x(0)\left[1 + A(t)\sin(\omega t)\right] \tag{6.6}$$

$$A(t) = 1.0 + 0.05\lfloor\frac{2\pi t}{\omega}\rfloor. \tag{6.7}$$

The orthogonal direction (here described as y) can be treated in a number of different ways, as discussed in **Chapter 5**. Work in this chapter focusses on dynamic rupturing behaviour rather than mechanical properties. This means that exactly reproducing Poisson's ratio to control compression in y is less important, and unless otherwise stated the stretching took place in x with the y dimension controlled to maintain constant area by shrinking each step.

6.2.4 Hessian Calculation.

While this chapter focusses on forcing networks to break, it is useful to compare to the 'natural' motion of the entire system. The natural motion can be mathematically expressed as a series of normal modes, which often involve a concerted motion of all beads across all polymers. Calculating the normal modes involves approximating the energy landscape around a point as a series of quadratics, and calculating the frequency of a harmonic spring with the same quadratic energy character. To find the normal modes of the harmonic springs, the Hessian matrix of second derivatives must be calculated:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 U}{\partial x_1 \partial x_1} & \frac{\partial^2 U}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 U}{\partial x_1 \partial x_n} \\ \frac{\partial^2 U}{\partial x_2 \partial x_1} & \frac{\partial^2 U}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 U}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 U}{\partial x_n \partial x_1} & \frac{\partial^2 U}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 U}{\partial x_n \partial x_n} \end{bmatrix}. \tag{6.8}$$

Calculating the Hessian from just the energy equation U is computationally expensive, requiring $\mathcal{O}\left(n^2\right)$ evaluations of the energy function and having poor numerical characteristics. Thankfully MD simplifies this problem, as the Jacobian is once again known analytically and

provides a vector of first derivatives.

The individual Hessian elements can be calculated by a finite difference method. This works by unravelling the positions array to a 1D vector $\boldsymbol{x}$ and offsetting the positions by $\varepsilon\hat{\boldsymbol{e}}_{\boldsymbol{i}}$, thereby calculating a new force $\frac{\partial U}{\partial x_j}$. The offset forces are then combined using a differentiation from first principles approach

$$\frac{\partial^2 U}{\partial x_i\,\partial x_j} = \frac{\left(\frac{\partial U}{\partial x_j}\right)_{\boldsymbol{x}+\varepsilon\hat{\boldsymbol{e}}_{\boldsymbol{i}}} - \left(\frac{\partial U}{\partial x_j}\right)_{\boldsymbol{x}-\varepsilon\hat{\boldsymbol{e}}_{\boldsymbol{i}}}}{2\varepsilon}, \tag{6.9}$$

where ε is a free parameter, and $\frac{\partial U}{\partial x_j}$ is the force in the j^{th} direction evaluated at the point $\boldsymbol{x} + \varepsilon\hat{\boldsymbol{e}}_{\boldsymbol{i}}$. The Hessian produced this way is not exactly symmetric due to numerical issues. It can be made symmetric by averaging out the $\frac{\partial^2 U}{\partial x_i\,\partial x_j}$ and $\frac{\partial^2 U}{\partial x_j\,\partial x_i}$ terms:

$$\mathbf{H} = \frac{\mathbf{H} + \mathbf{H}^{\mathrm{T}}}{2}. \tag{6.10}$$

The short-range nature of the interactions in the polymer model leads to an extremely sparse Hessian, which is easily diagonalisable. Diagonalising $\mathbf{H}$ provides a natural mapping for the eigenvectors to atomic motion (reshaping an eigenvector $\boldsymbol{v}_{\boldsymbol{i}}$ to the same shape as the positions vector). The eigenvectors provide two important pieces of information: low frequency modes corresponding to the network breathing, and imaginary frequency modes that correspond to a change in energy of the network.

6.2.5 Velocity Autocorrelation Function.

Energy is localised when rupturing begins, and spreads out through the network. This can be studied by measuring the evolution of kinetic energy throughout the network.

Kinetic energy is determined by the velocity of particles, which can be tracked via the velocity autocorrelation function (VACF). The VACF is defined as

$$\mathrm{VACF}(\Delta t) = \frac{1}{N}\sum_i \lim_{\tau\to\infty}\frac{1}{\tau}\int_{t=0}^{\tau} \boldsymbol{v}_i(t)\boldsymbol{v}_i(t+\Delta t)\,\mathrm{d}t, \tag{6.11}$$

where $\boldsymbol{v}_i(t)$ is the velocity of atom i at time t. This property is averaged over all times, and averaged over each atom i in the system.

The VACF measures how long the each atom has an 'memory' of its initial velocity. When VACF approaches zero, each atom's velocity is uncorrelated with its original velocity.

6.2.6 Inverse Participation Ratio.

Another metric that will be useful in this chapter is the Inverse Participation Ratio (IPR), $\mathcal{I}$, which calculates how diffuse a property is over the system. The IPR approximately reflects how many sites the property is delocalised over — that is, $\mathcal{I} = 1$ means that the property is localised on one site, and $\mathcal{I} = \frac{1}{N}$ means that the property is delocalised over approximately N sites. For kinetic energy, this can be calculated as

$$\mathcal{I} = \sum_i \left(\frac{{}^1\!/\!_2 m v_i^2}{\sum_j {}^1\!/\!_2 m v_j^2} \right)^2, \tag{6.12}$$

effectively the sum of the squares of the fraction of total kinetic energy assigned to each atom. By tracking $\mathcal{I}$ over time, the rate at which kinetic energy spreads out into the network can be seen.

Similarly, the IPR can be calculated for an eigenvector $\boldsymbol{v}$ spread across N sites

$$\mathcal{I} = \sum_{i=0}^{N} v_i^4. \tag{6.13}$$

6.3 Results and Discussion.

6.3.1 Hessian and Normal Modes.

The Hessian can be calculated from the finite difference scheme in **Section 6.2.4**. It provides some useful insights into the collective behaviour of the system when diagonalised.

Eigenvalues with positive values correspond to normal modes of the system with concerted

motion. The eigenvalues can be related to the frequencies ω of ideal normal mode springs by mass-weighting an eigenvalue λ_i according to the masses m of individual springs:

$$\omega = \sqrt{\frac{\lambda_i}{m^2}}. \tag{6.14}$$

Imaginary frequencies correspond to motions that lower the energy of the system. In this system there should be no imaginary frequency modes, as the system should be at a local minimum.

Picking the parameter ε (in **Equation 6.9**) can be difficult as there is a balance of numerical accuracy effects for small ε *vs* finite size step effects for large ε. A balance can be struck between these by sampling many ε values and seeing if the eigenvalues converge. Sampling ε values across a large range, here logarithmically in the form $2^k \times 10^{-3}$, can show which eigenvectors are spurious and which are physically meaningful. **Figure 6.3** shows a few imaginary frequencies as a function of ε, which become less prominent for slightly larger ε. This indicates that they are a numerical artefact and matches our intuition that there are no energy-reducing motions for the hexagonal network.

The eigenvalues approximately converged for many values of ε as seen in **Figure 6.3**. Broadly the frequencies for eigenvalues of different ranks i track one another regardless of ε, which indicates that they are physically meaningful.

The jumps in **Figure 6.3** correspond to the accessibility of new concerted modes. The largest four jumps at $i = 684$, 863, 971 and 1079 are shown in **Figure 6.4**. **Figure 6.4a** shows a mode where the network pulls towards some head groups and away from others, **Figure 6.4b** shows a molecular symmetric stretch normal mode, **Figure 6.4c** shows a molecular asymmetric stretch normal mode and **Figure 6.4d** shows a more complex stretching mode.

The IPR is plotted as a function of eigenvalue rank in **Figure 6.5**; the zero-frequency eigenvalues have moderately localised eigenvectors, but most eigenvectors in this network are

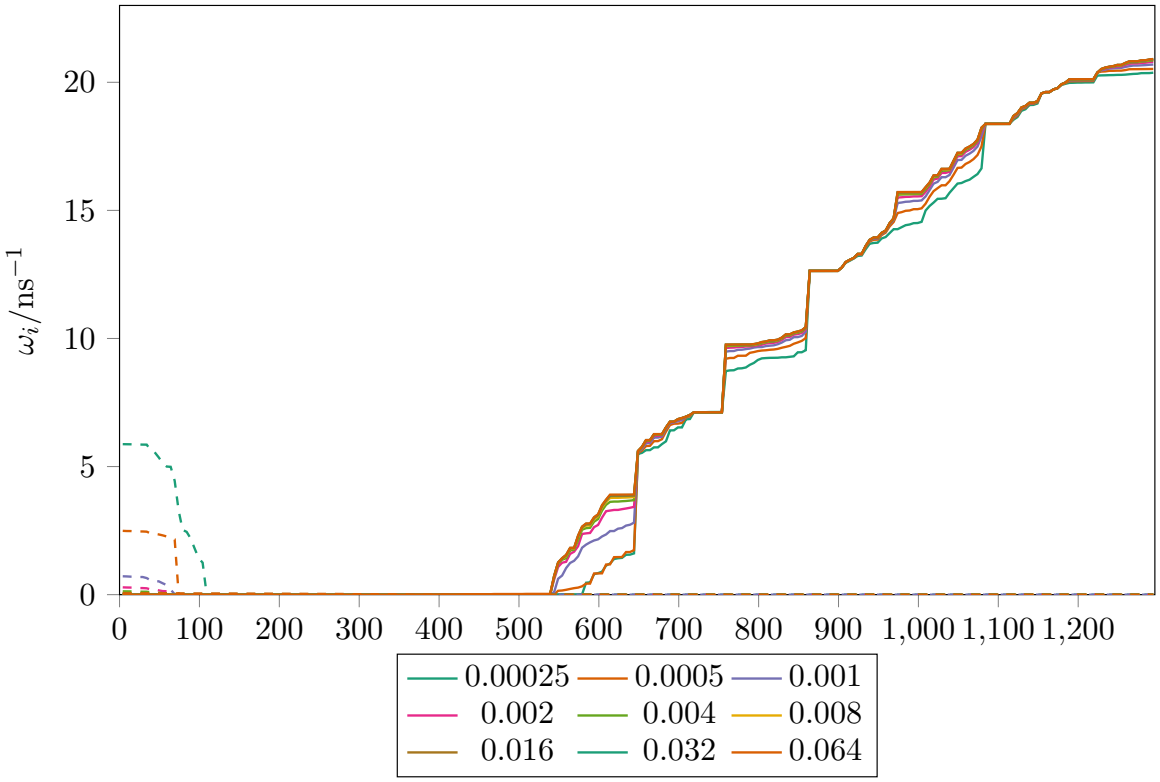
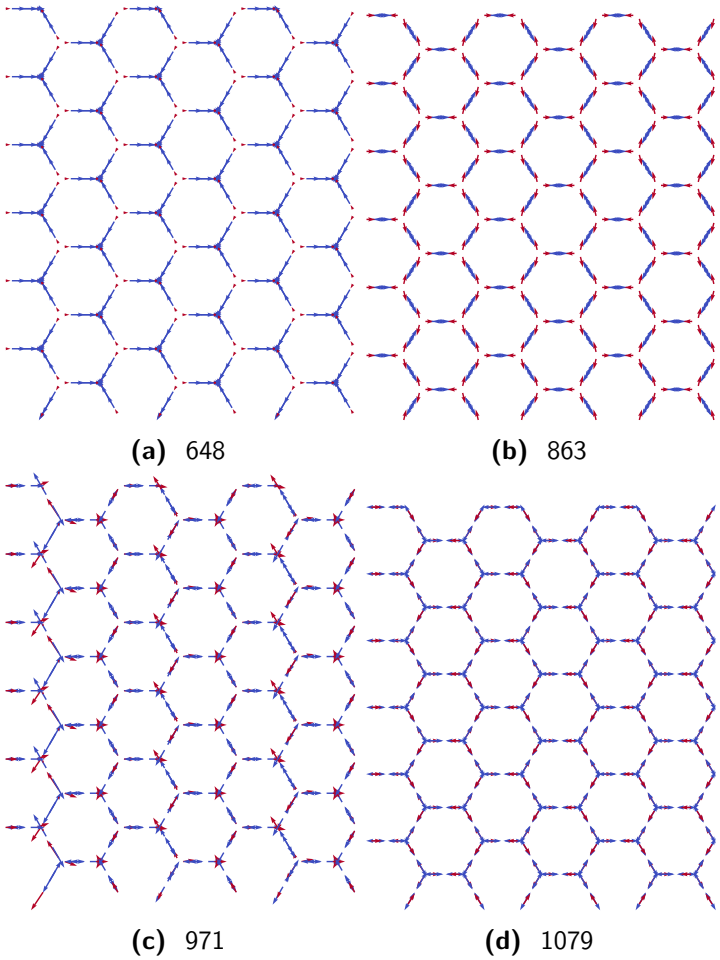


Figure 6.3: Frequencies of the normal modes of the hexagonal network sorted by magnitude. Dotted lines represent imaginary frequencies, and solid lines represent real frequencies.

delocalised. There is little correlation between ω_i and the corresponding IPR. However, the near zero-frequency modes are the most strongly localised. They are generally small motions localised to specific patterns of body beads. Body bead motions are low frequency owing to the semiflexibility of polymers in this model.

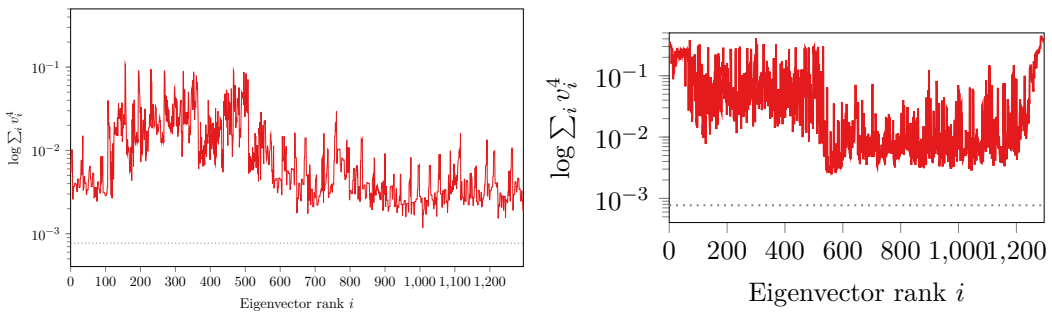
The IPR is most interesting when it shows the localisation of eigenvectors in networks featuring defects. **Figure 6.5** shows the degree of localisation for two idealised hexagonal network networks: one with no defects, and one with a single bond switching defect. The eigenvectors in **Figure 6.5** are clearly more localised, especially at the ends (representing large real and imaginary frequencies). Some visual examples of eigenvectors of ranks 1294 and 1295 (highest real frequencies) are shown in **Figure 6.6**. The idealised network clearly shows delocalised motion, especially around head groups. However, the equivalent eigenvectors for



(a) 648 **(b)** 863

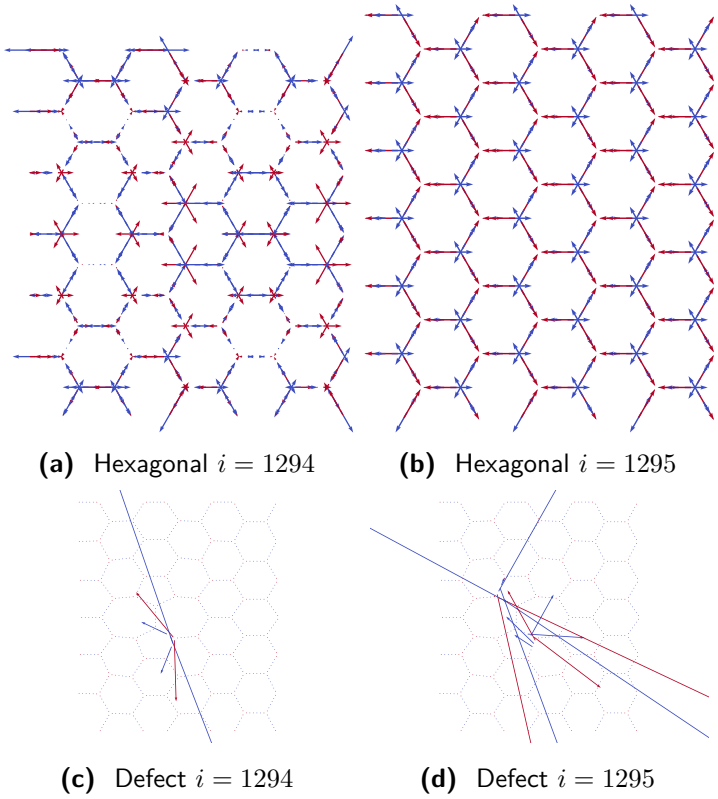
(c) 971 **(d)** 1079

Figure 6.4: Eigenvectors visualised at the locations of the jumps in **Figure 6.3**. Each jump represents a new concerted mode. Red arrows represent motion of head beads, blue arrows represent motion of body beads.



(a) IPR for an idealised hexagonal network. **(b)** IPR for a network with a defect.

Figure 6.5: Inverse participation ratios (IPR) of the eigenvectors for two networks. An IPR close to 1 means that the eigenvector is perfectly localised; an IPR close to $1/1296$ (dotted line) means the eigenvector is perfectly delocalised.



(a) Hexagonal $i = 1294$ **(b)** Hexagonal $i = 1295$

(c) Defect $i = 1294$ **(d)** Defect $i = 1295$

Figure 6.6: Eigenvectors visualised for the two highest frequency modes for a hexagonal network and a network with a single defect. The hexagonal eigenvectors are delocalised concerted motions, but the motion of the defect network is localised to the defect.

the defect-containing network are strongly localised. The localisation is extremely strong, as the vectors pointing out of frame in **Figure 6.6c** and **Figure 6.6d** show. There is dramatic motion around head groups on the edge of a defect. In the remainder of this chapter, how preference for motion around a defect affects rupturing will be examined.

6.3.2 Linear Stretching.

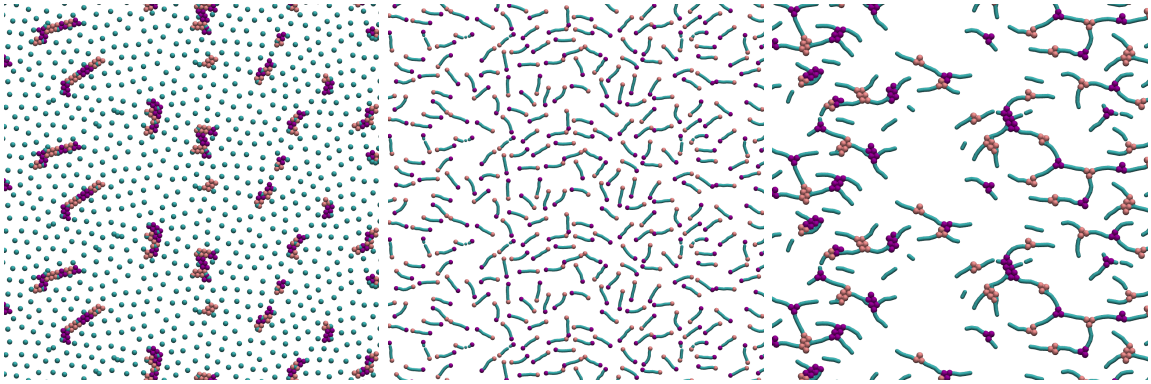
Rupturing phenomena are very complex, as they are extremely chaotic and sensitive to random effects. It is useful to explore the extremes of the rupturing behaviour with a simple model before exploring more complicated effects. The simplest rupturing behaviour is that of a hexagonal network stretched at a constant engineering strain rate $\frac{\mathrm{d}L_x}{\mathrm{d}t}$. This is a expansion of the same ratio every step from $\lambda = 1$ to $\lambda = 2$, here over 10.0 µs. This expansion rate

is still extremely fast by physical standards, which conveniently encourages the networks to rupture. For this section, much of the parameterisation from **Chapter 4** was re-used. This is because it allowed studies of dynamics over large timescales (up to 100 µs)[71], and comparison between chapters. Idealised hexagonal networks were constructed, and assigned varying dissociation energies D_e and varying head group energies ε. Both parameters were in the range 1.00×10^{-21} J to 128×10^{-21} J at a temperature of 100 K ($k_\mathrm{B}T \approx 1.38 \times 10^{-21}$ J).

The linear stretching simulations showed how the balance of bond damage to head group damage led to different outcomes. There were three final extreme scenarios: 'bead soup', 'polymer soup', and 'mixed soup'. Firstly, 'bead soup' occurred when D_e was small compared to $k_\mathrm{B}T$, which led to bonds breaking under normal thermal motion. Secondly, when ε_HH was small compared to D_e and $k_\mathrm{B}T$ the polymers broke at head groups and remained whole but uncoordinated. This led to a 'polymer soup' as seen in **Figure 6.7b** and is reminiscent of self-assembly results at low ε_HH in **Chapter 4**.

Finally, when D_e and ε_HH were of similar magnitude and both smaller than $k_\mathrm{B}T$ a 'mixed soup' arose. This was where the body-head bonds preferentially break as they are under the most strain, and is shown in **Figure 6.7c**. Once the body-head bonds broke, the head groups formed small clusters while the headless bodies floated free and repelled one another.

However, in the case where D_e and ε_HH were bigger than $k_\mathrm{B}T$, the networks ruptured under strain in interesting ways. A single hexagonal network with no defects and default Morse parameter potentials being stretched linearly is shown in **Figure 6.8**. The first panel **Figure 6.8a** shows rupturing started when one horizontal bond was broken — this was in the direction of stretching and randomly had enough energy to break and relieve the traction. Successive panels show how the rupture evolved, where the break initially formed a decagon and released the strain as kinetic energy into the neighbouring parts of the network. This



(a) 'Bead soup', for a ruptured network with low D_e

(b) 'Polymer soup', for a ruptured network with low $\varepsilon_{\mathsf{HH}}$.

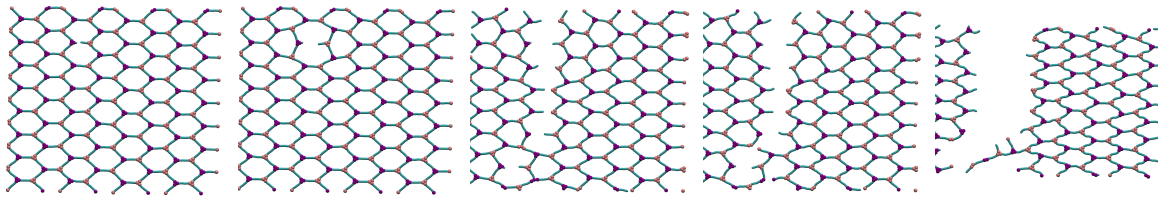
(c) 'Mixed soup', for a ruptured network with D_e and $\varepsilon_{\mathsf{HH}}$ comparable and low.

Figure 6.7: Extremes of the network rupturing landscape as three fully ruptured soups. The most interesting behaviour lies between these extremes.

additional kinetic energy and the associated distortion around the decagon caused the top and bottom bonds of the decagon to break, and thus the rupture propagated vertically above and below the initial defect. In **Figure 6.8d**, a new decagonal defect opened and this interrupted the rupture propagation. The new defects caused a chain of $k = 2$ nodes to form, adding enough flexibility to hold the two parts of the network together from 0.920 µs to 2.12 µs, when the network finally broke.

Altogether, this shows the importance of using semi-flexible polymers to study rupture. During rupture, there is complex motion as the polymers vibrate and bend, distorting their ring structure away from the ideal arrangement. This is especially obvious in **Figure 6.8e** as the hexagons clearly have curved edges. The polymer model naturally includes the non-elastic modes and coherent network motion to dissipate energy from a damage site.

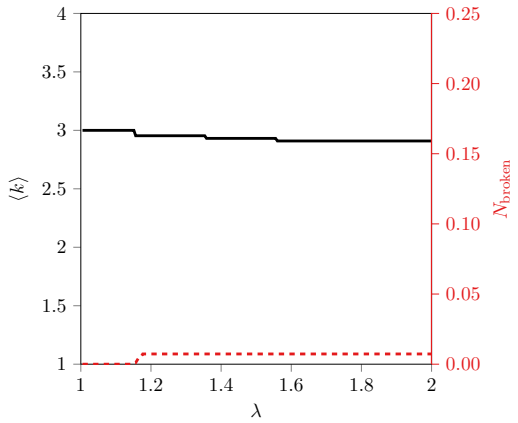
Elsewhere in this thesis, ring statistics are the primary descriptive metrics used for two dimensional (2D) networks. Ring statistics are unsuitable to describe rupture, so they are only calculated sparingly during this chapter. This is because ring statistics are heavily averaged, but rupture occurs in small local regions. Localisation is especially prominent in biopolymer



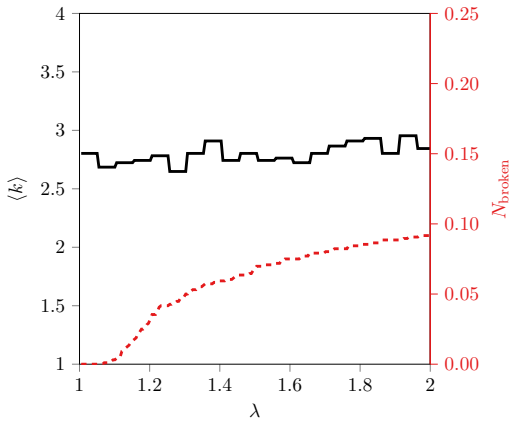
(a) 0.720 µs **(b)** 0.770 µs **(c)** 0.870 µs **(d)** 0.920 µs **(e)** 2.12 µs

Figure 6.8: Snapshots of a rupture of a hexagonal polymer network stretched linearly (left to right). The rupture occurs initially randomly along one edge to relieve stress, and then opens up orthogonal to the stretching direction. When the rupture is the length of the periodic simulation cell, all stress is relieved and the remaining network is again purely hexagonal. This figure is reproduced from M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180.

networks with short-range interactions. Therefore, the main metrics used in this chapter are the average degree $\langle k \rangle$, the fraction of broken bonds N_{broken}, and the time until the first damage arose. **Figure 6.9a** shows damage metrics tracked during the rupture shown in **Figure 6.8**. For comparison, the rupture of a network with a low D_e is shown in **Figure 6.9b**. When the first edge breaks in **Figure 6.8b**, this is reflected in a drop of $\langle k \rangle$ in **Figure 6.9a** from 3.00 to 2.95. The inclusion of a decagon leads to an average number of edges per ring of 6.19 as opposed to 6. While the tear propagates, $\langle k \rangle$ remains close to 3 and similarly $\langle n \rangle$ remains close to 6 — this is because the majority of the network remains undistorted. As the simulation ends and the tear joins up, $\langle n \rangle$ returns to 6 as the two separated halves of the network once more become idealised and hexagonal.



(a) A plot of $\langle k \rangle$ and N_{broken} against λ, the stretching ratio for a system with $D_e = 128 \times 10^{-21}$ J and $\varepsilon = 128 \times 10^{-21}$ J.



(b) A plot of $\langle k \rangle$ and N_{broken} against λ, the stretching ratio for a system with $D_e = 16.0 \times 10^{-21}$ J and $\varepsilon = 16.0 \times 10^{-21}$ J.

Figure 6.9: Average node degree and edges per ring for rupturing networks with different potential model parameters. The black solid line represents the mean node degree, $\langle k \rangle$, which is 3 for an ordered hexagonal network and 1 for 'loose' bonds. The red dashed lines represents N_{broken}, the fraction of broken bonds, which is 0 for an unruptured network and 1 for a system with no remaining bonds. This figure is adapted from M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180.

6.3.3 Sinusoidal Stretches.

The ocular lens capsule does not undergo linear stretches in the eye. Instead, the lens capsule is stretched by the action of ciliary muscles (according to Helmholtz) and relaxed over timescales of 200 ms to 500 ms, depending on the age of the eye[143,144]. The maximum amount of stretching that a lens capsule in the eye experiences is approximately 10.0 %[127].

However, it is very difficult to computationally match the frequencies and repeated patterns of stretching that real tissue undergoes. The natural movements of the eye take place over seconds or milliseconds. However, simulation is limited to (at the most) microseconds, even with a coarse-grained model. Thankfully, short timescales are actually helpful for studying rupture. Human eyes do not rupture every time they change focal depth and only accumulate damage slowly. A short timescale can accelerate rupturing and accumulation of damage in the network.

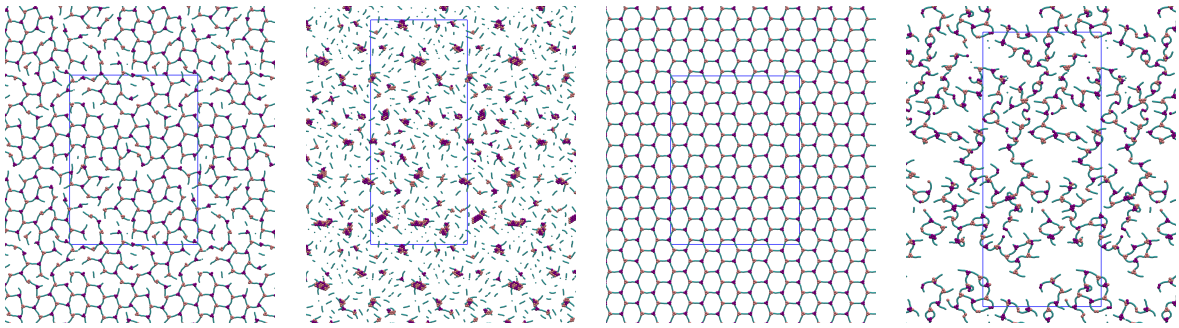
Stretching the simulation cell adds energy to the system; distorting the bead positions away from equilibrium requires the addition of potential energy. The energy that is pumped in is partially responsible for the damage to the network. The amount of energy added is related to the frequency and amplitude of simulation cell oscillations. By increasing the frequency, more energy is added to the network and the individual polymers have less time to accommodate their new positions. By increasing the amplitude, more energy is added to the network, because at maximum or minimum extension the beads are further from their equilibrium positions.

The increasing sinusoidal simulations varied the box length L_x from its initial length L_{x_0} according to **Equation 6.5**. with $A(t)$ the amplitude taking values in the range 0.100 to 0.500 and ω the frequency taking values in the range $0.100\,\mu s^{-1}$ to $10.0\,\mu s^{-1}$. Each run simulated a network over $10.0\,\mu s$, with potential model parameters in the range $\varepsilon_{HH} = 16.0 \times 10^{-21}\,J$ to $128 \times 10^{-21}\,J$ and $D_e = 16.0 \times 10^{-21}\,J$ to $128 \times 10^{-21}\,J$. Defaults for both of these parameters were set at $D_e = \varepsilon = 32.0 \times 10^{-21}\,J$.

Initial Exploration.

The initial exploration of sinusoidal stretching used hexagonal networks, similar to those used in **Section 6.3.2**. These were stretched sinusoidally at two amplitudes: firstly at $A = 1.1$ to mimic the range of normal accommodative motion[127] and secondly at $A = 1.5$ to mimic a destructive stretch. Each simulation covered $10.0\,\mu s$ of time, with $\omega = 1.00\,\mu s^{-1}$.

The $A = 1.5$ simulations always ruptured because the stretching was too large to be accommodated by deformation and too rapid to repair. However, the $A = 1.1$ simulations often stayed together due to a lower $\frac{d\lambda_x}{dt}$ and a small total stretch. This can be seen by inspecting the figures in **Figure 6.10**, where the $A = 1.1$ figures show a semblance of order but the $A = 1.5$ networks are fully broken.



(a) $A = 1.1$, $D_e = 16.0 \times 10^{-21}$ J, $\varepsilon = 16.0 \times 10^{-21}$ J.

(b) $A = 1.5$, $D_e = 16.0 \times 10^{-21}$ J, $\varepsilon = 16.0 \times 10^{-21}$ J.

(c) $A = 1.1$, $D_e = 128 \times 10^{-21}$ J, $\varepsilon = 128 \times 10^{-21}$ J.

(d) $A = 1.5$, $D_e = 128 \times 10^{-21}$ J, $\varepsilon = 128 \times 10^{-21}$ J.

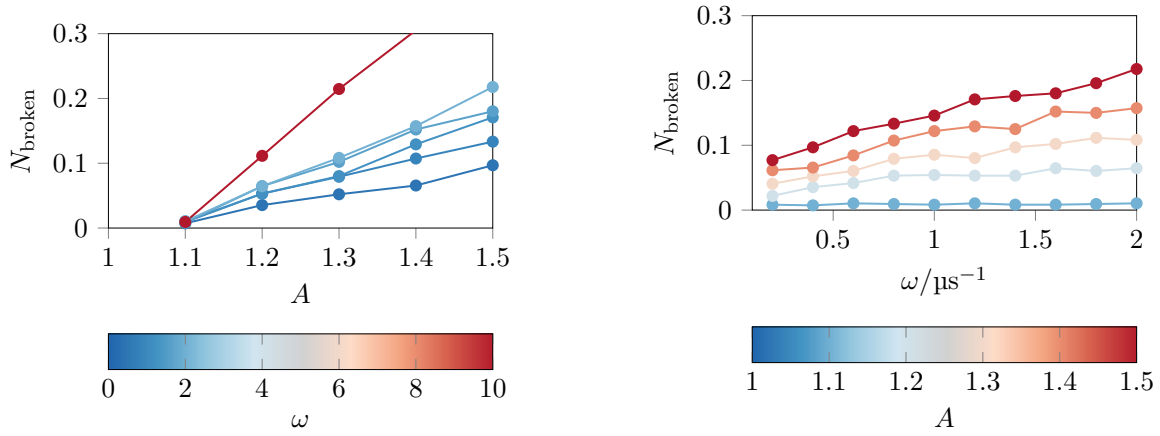
Figure 6.10: Ending snapshots of networks stretched with different potential model parameters and different amplitudes. The blue box represents a single repeating simulation cell.

Frequency and Amplitude.

With the limits of stretching behaviour established, it is also prudent to scan across the intermediate range to see how that behaviour changes. Simulations with amplitudes $A = 1.1$, 1.2, 1.3, 1.4 and 1.5 were performed, each with a range of stretching frequencies ω. The different values of ω change the number of oscillations per simulation.

Figure 6.11a shows the fraction of broken bonds as a function of amplitude, and **Figure 6.11b** shows the effect of varying frequency. When the stretching amplitude is low ($A = 1.1$) the networks suffer no damage and can oscillate forever. As there is no damage, the amount of damage is independent of ω. However, as ω increases, the networks suffer more damage. The increased damage with greater stretching frequency ω is because the polymers have less time to relax into new positions at the peaks and troughs of stretching, and more work is done per unit time by stretching. Instead, the polymers are forced out of position and break.

Larger amplitudes ($A = 1.20$ to 1.50) led to more damage in the networks at the end of the simulation. The damage, as measured by number of broken bonds, increases with amplitude of oscillation. The effect of increased amplitude can be seen in **Figure 6.11a**, where a larger



(a) Networks stretched at varying amplitude A but constant frequency ω.

(b) Networks stretched at varying frequency ω but constant amplitude A.

Figure 6.11: Fraction of broken bonds at the end of a simulation frequency ω and amplitude A. The fraction of broken bonds at the end of the simulation increases approximately linearly with both frequency and amplitude, as the most damage is introduced at the maxima and minima of a stretching cycle. The energy inserted into the network at the maxima and minima of stretching is linearly related to the frequency and amplitude.

A leads linearly to more damage at the end of the simulation.

The linear relationship between A and N_{broken} could be explained as the stretching forces are given by $\frac{\mathrm{d}U}{\mathrm{d}t} \propto A\omega\cos(\omega t)$, which are maximised with a value $\propto A\omega$ at peaks and troughs of the stretching.

6.3.4 Edge Removal Defects.

The initial part of this chapter, **Section 6.3.1**, discussed the motion around defect sites in a network. Eigenvectors of a defect-containing network showed dramatic motion around the defect site, leading to a hypothesis that those networks will rupture more easily. With the rupture of idealised networks under both linear and sinusoidal stretching established in **Section 6.3.2** and **Section 6.3.3**, it is possible to compare to the rupture of defect-containing networks.

The simplest defect to introduce is to remove an edge, as missing edges provide a weak point in the network. This was also inspired by how the rupturing motion in **Figure 6.8**,

which showed that a tear began by breaking an edge to form a decagon. To study the effect of removing edges, hexagonal networks were set up and then a given fraction N_{broken} of edges were removed. Edges that joined two nodes with degree $k = 3$ were selected randomly and deleted. This ensured that the final network was still a cohesive network, mimicking studies of procrystalline 2D lattices[145].

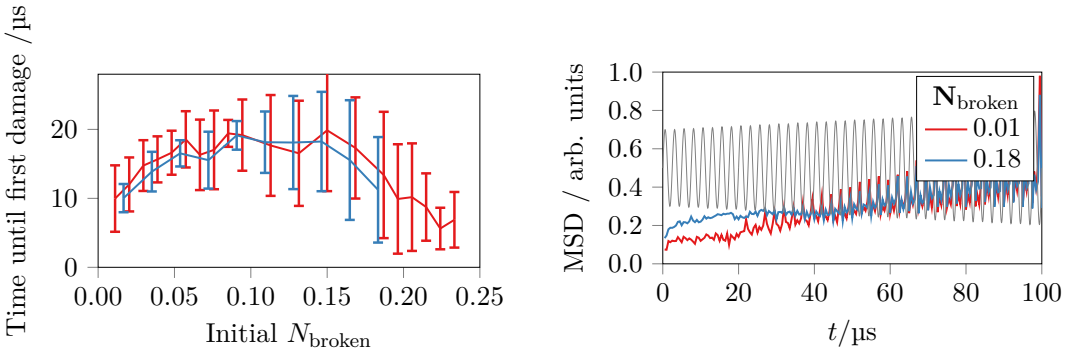
Each network was stretched sinusoidally with an increasing amplitude over time. This was because it is difficult to know the true amplitude of stretching, and **Section 6.3.3** established the importance of stretching amplitude. Each network was stretched for 100 µs, undergoing 40.0 oscillations. The amplitude per oscillation, A, increased linearly with each subsequent stretch in the range $A = 1.10$ to 1.50. Stretching like this showed the effect of cumulative damage to the network, and also provides a direct way to measure the critical amplitude at which the network fully ruptures. Each simulation was repeated ten times, and used potential model parameters $T = 100\,\text{K}$, $\varepsilon_{\text{HH}} = D_e = 16.0 \times 10^{-21}\,\text{J}$, and all other parameters set at the defaults established in **Section 6.2.2**. The sets of ten repeats were themselves performed in two groups. In first group of ten, described as 'random edge removal', different edges were removed between simulations. In the second group of ten, described as 'progressive edge removal', the same edges were removed each time. The edges were removed such that if an edge (u, v) was removed in an N_{broken} simulation, then (u, v) would also be removed in simulations with $N'_{\text{broken}} > N_{\text{broken}}$. No more than 25.0 % of the edges could be removed randomly without compromising the overall network structure.

Figure 6.12a shows how long it took a network with N_{broken} edge removal defects to suffer its first irreparable damage. The first irreparable damage is defined as when one bond within a polymer is over a threshold of 138 nm and is deleted by LAMMPS. The results show three regions with different behaviour:

- A flexible initial region, where each edge removed allows for greater flexibility, delaying damage. This is in the range 0 to 0.0750.
- A weak final region, where each edge removed weakens the overall network. This is in the range 0.150 to 0.250.
- A flattening off region where those two effects cancelled out. This is in the range 0.100 to 0.150.

The flexibility of a network can be measured by how far each bead has moved from its starting position over time, as measured by the Mean Square Displacement (MSD). The other panel in **Figure 6.12b** shows the MSD per bead for two networks. The red line shows a network with few edges removed, and the blue line shows a network with many edges removed. The network with most edges still intact shows a coupling to the stretching motion starting at 30.0 µs. This network first broke around 10.0 µs. By the end of the simulation at 100 µs, the motion of the beads was entirely dominated by the simulation cell motion. However, the network with more edges removed wasn't strongly coupled until 50.0 µs. This network first broke later, around 15.0 µs. Even close to the end of the simulation, the coupling of the more broken network to simulation cell motion was less dramatic.

Considering the differences in MSD of those two networks is helpful to understand the pattern in **Figure 6.12a**. The network with more edges removed was significantly more flexible than the network with fewest removed. Flexibility means the network could deform to absorb stretching or compression without rupturing. That energy had nowhere to go in an undamaged network, so it could only be released through damage.



(a) Time to first damage as a function of edges removed. The red points represent networks where all edges were removed randomly. The blue points were removed randomly in the same order; that is the edges removed for $N_{\text{broken}} = 0.0200$ are a subset of the edges removed for $N_{\text{broken}} = 0.200$ and so on.

(b) The MSD of beads over a simulation. The overlaid sin graph represents the size of the simulation cell over time, and the two coloured lines represent the MSD for two simulations with different values of N_{broken}.

Figure 6.12: Time to first damage and the MSD of networks undergoing sinusoidal stretching with differing numbers of edges removed. This figure is adapted from M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180.

System Size Effects

The preceding simulations showed relatively large error bars in their time until first damage, partly due to the influence of thermal damage. Even though the simulations were performed at a relatively low temperature (100 K), in a small system nodes may or may not have enough energy to break. Larger systems have an added benefit in that they allow a disentangling of effects. If the rupture begins exclusively at a defect, the damage metrics should be independent of system size. However, if the rupture begins in the bulk, the damage metrics should scale with system size. A larger system will be damaged much sooner than a smaller system, as the probability of at least one polymer having enough thermal energy to break increases.

Figure 6.13 reproduces the time until first damage figure (**Figure 6.12a**) with different system sizes. The larger systems show shorter times until first damage, indicating that damage began in the bulk. Larger systems also show smaller error bars, as random effects become more predictable. The smaller error bars help show the effect of flexibility more clearly. The

two counterbalanced effects of flexibility and structural integrity become clearer with smaller error bars.

Individual damage trajectories, featuring the number of broken bonds over time are shown in **Figure 6.14**. The reduced uncertainty for larger systems is shown more clearly in these figures, with shaded regions representing one standard deviation across ten simulations. The larger systems approximately follow the same damage trajectory as the smaller systems, but with a much smaller standard deviation.

Each of the three sub panels shows that irreparable damage increases monotonically over time. **Figure 6.14c**, from networks with 20.0 % of edges removed, shows first damage early on. However, it increases slowly after first damage, especially for larger system sizes. The trajectories become broadly similar at later times as a result of the bead motion coupling to simulation cell motion. The coupling is slower in **Figure 6.14c** than in **Figure 6.14a**, which results in a shallower curve.

6.3.5 Combined Defects.

Broken edges are important in network rupture, but they are not the only form of defect that can occur. One likely form of defect is the bond switching defect, used in **Chapter 3** to generate plausible biological networks. However, the defects in **Chapter 3** change the degree around a node, with consequences for head group coordination numbers at the ends of the constituent polymers. This introduces weak points that more rapidly break and complicate study. Instead of those defects, the networks are damaged with Stone-Wales defects which preserve the degree of both ends as mentioned in **Section 6.2.1**. A similar approach has been used successfully by Bamer *et al.* to introduce weak points into 2D silica networks[126].

Ruptures in an idealised network propagate along a line. Adding a flexible defect by

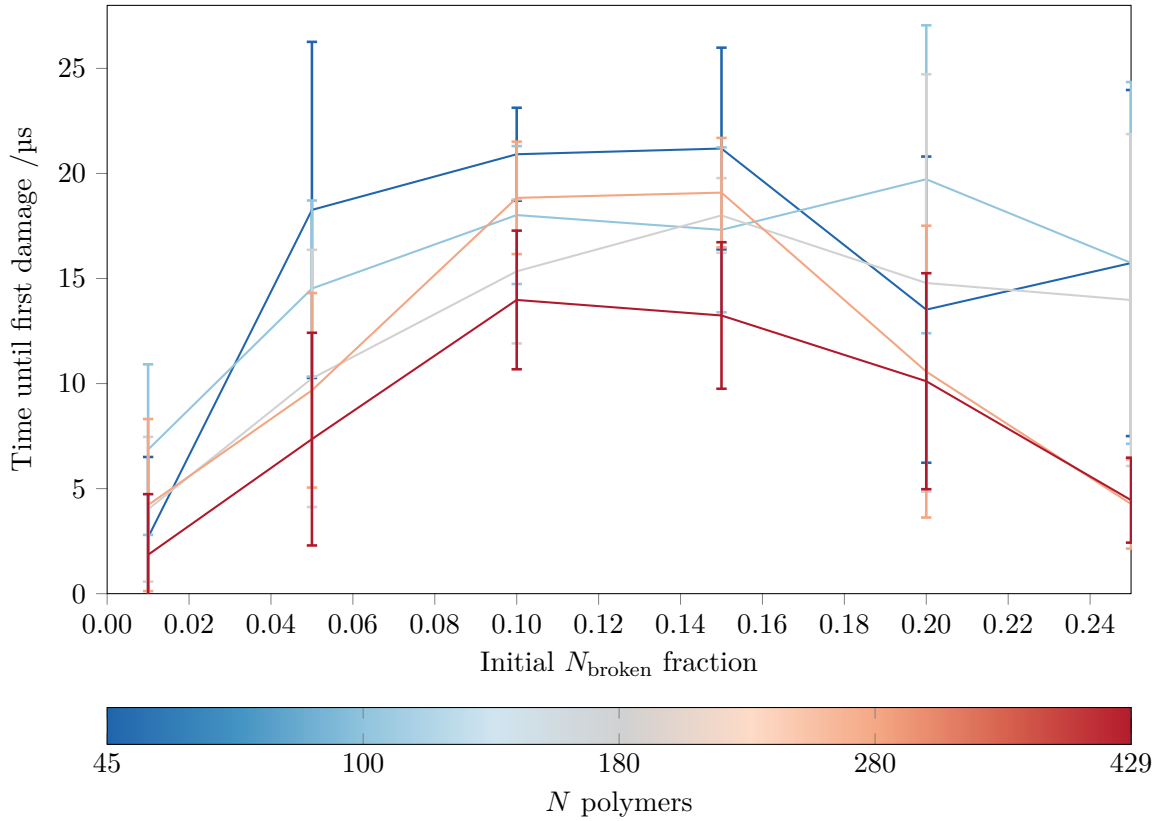


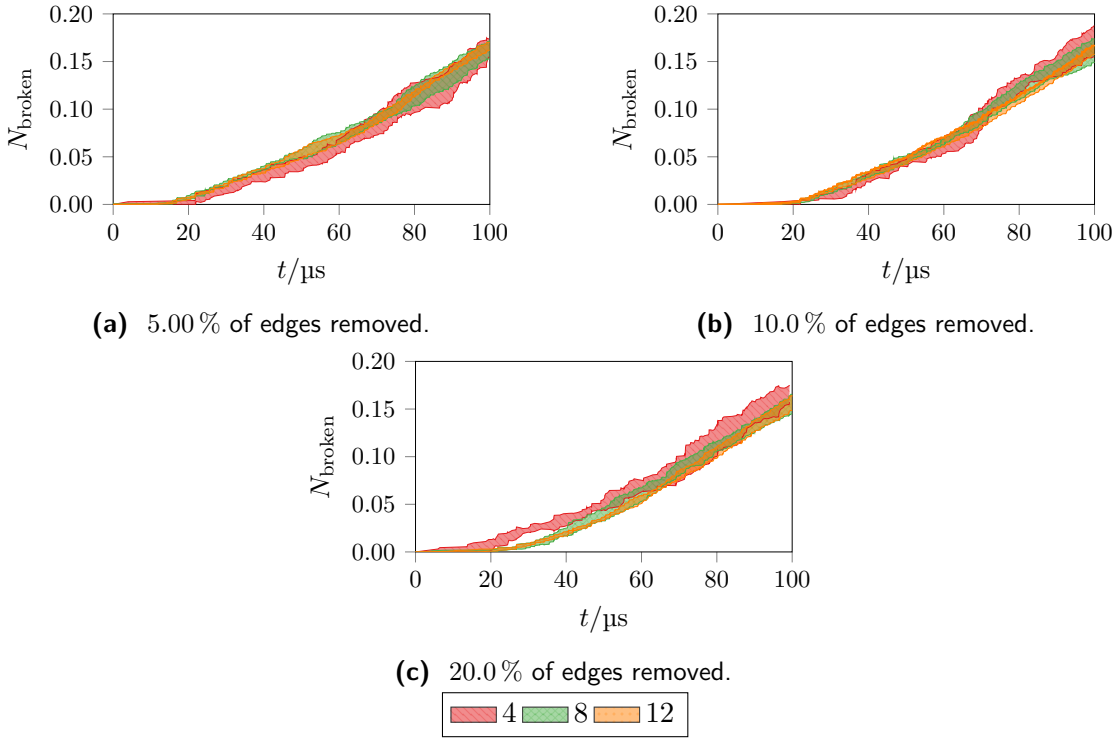
Figure 6.13: Time until first damage as a function of the system size. Each different line represents a simulation with $N_{\text{broken}}\%$ of edges are removed from a network of N polymers, where N is denoted in the legend. The larger systems break sooner because there are more constituent polymers undergoing random motions.

removing edges may delay the onset of a rupture, or even halt it if the defect is along that line. To test that hypothesis, the following polymer networks were constructed:

- Hexagonal network with one bond switching defect, shown in **Figure 6.15b**.
- Hexagonal network with one edge removed and one bond switching defect, shown in **Figure 6.15c**.

The defect sites were chosen such that the removed edge was never part of the bond switching site.

Each network then underwent the increasing sinusoidal stretching procedure from **Section 6.3.4**. Each simulation covered 10.0 µs, for 10.0 oscillations up to a maximum amplitude

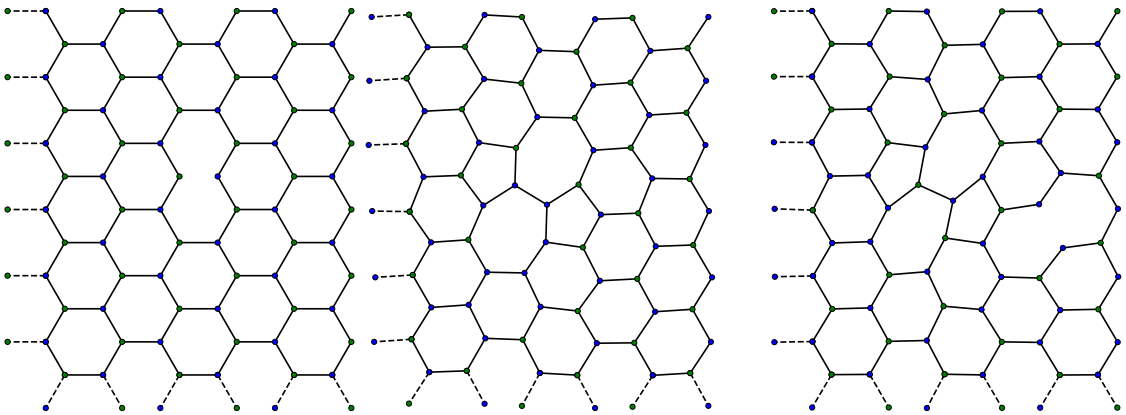


(a) 5.00 % of edges removed.

(b) 10.0 % of edges removed.

(c) 20.0 % of edges removed.

Figure 6.14: Damage trajectories for systems with the same fraction of edges removed but different system sizes. The shaded area in the bottom figures represents one standard deviation of the damage trajectory as described in the caption for **Figure 6.9b**, and the legend indicates the number of molecules in the network. This figure is adapted from M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180.



(a) A network with a single bond removed.

(b) A network with a Stone Wales defect.

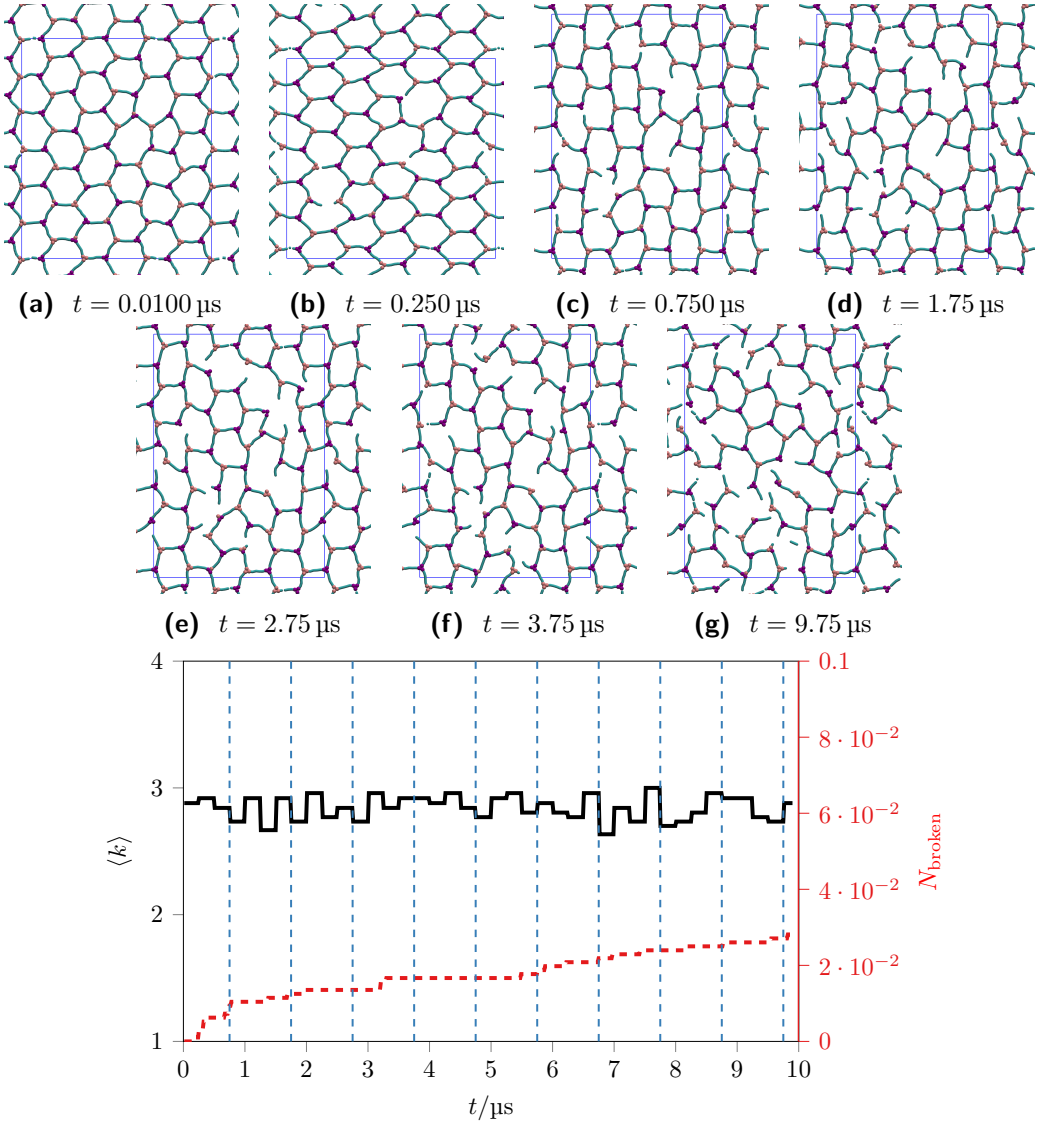
(c) A network with one edge removed and a Stone-Wales defect.

Figure 6.15: Three networks with different combinations of defects.

of $A = 1.1$. This was a shorter timescale than earlier simulations, chosen to only capture the times where the polymers had not coupled to simulation cell motion. Potential model parameters were varied in the range $D_e = 16.0 \times 10^{-21}\,\mathrm{J}$ to $64.0 \times 10^{-21}\,\mathrm{J}$ and $\varepsilon_{\mathrm{HH}} = 16.0 \times 10^{-21}\,\mathrm{J}$ to $64.0 \times 10^{-21}\,\mathrm{J}$. All other parameters were kept the same as before, including $T = 100\,\mathrm{K}$ and $\alpha = 0.150\,\mathrm{nm}^{-2}$.

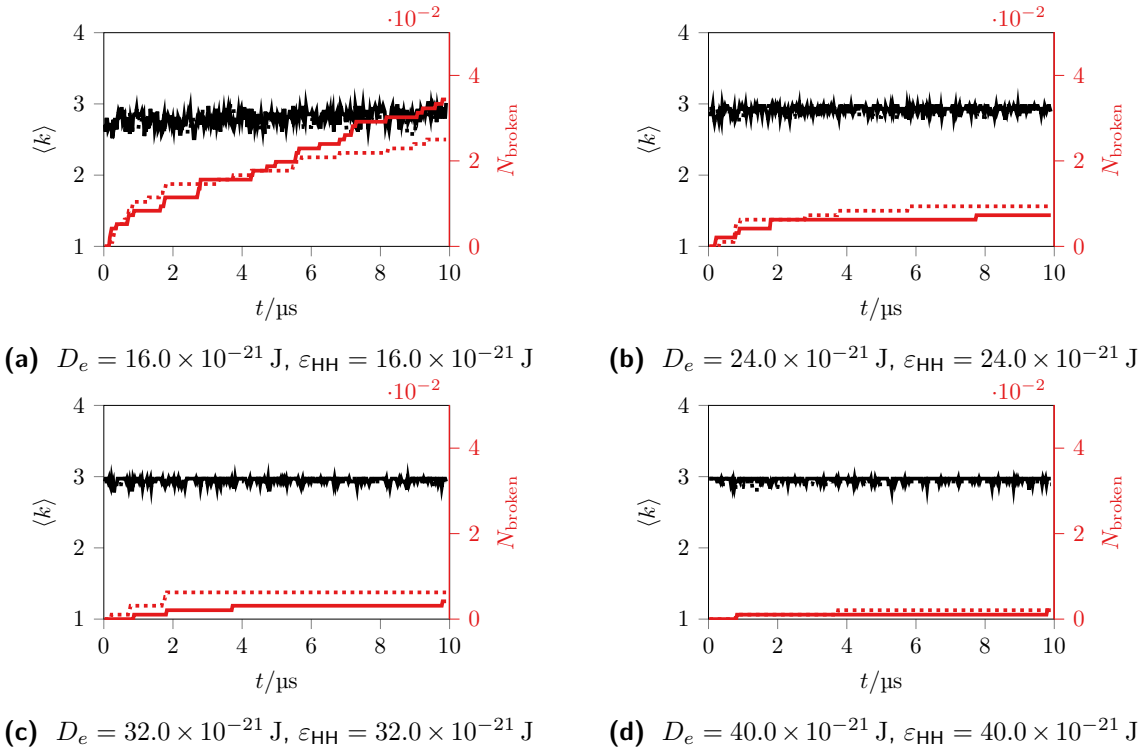
Snapshots of a network with a Stone-Wales defect undergoing sinusoidal stretching are shown in **Figure 6.16**, accompanied by two metrics of damage. These show that a bond initially breaks at $0.250\,\mu\mathrm{s}$ on the edge of the defect. Over time, more bonds break, mostly at the body-head connections. The average degree $\langle k \rangle$ tracks the number of head beads in each cluster. Damage to head beads in a cluster is repairable, and at the peaks of stretching (shown in snapshots) the head beads had often re-coordinated.

Damage trajectories from these simulations are shown in **Figure 6.17**, tracking the number of broken bonds and average degree over time. Comparing these figures does not show a clear trend. In **Figure 6.17a**, the network with both defects underwent more damage initially, but less damage towards the end of the simulation. However, in **Figure 6.17b**, the effect was reversed. A similar effect is seen in **Figure 6.17c**. In **Figure 6.17d** there was not enough damage to meaningfully analyse, because the bonds were so strong. If the bond switching defect prevents a tear from propagating, then we would expect to see similar initial damage but then metrics to plateau. Similarly, if the bond switching defect encouraged a rupture, we would expect to see sooner initial damage for the networks with those defects. Neither effect is readily apparent, which is most likely due to a cancellation of effects. Bond switching defects do provide flexibility (inhibiting rupture), but they also are local areas of high stress (accelerating rupture).



(h) Metrics tracking damage over time. Each blue line represents stretching reaching a maximum and often coincide with snapshots. The red line represents fraction of bonds broken, and the black line shows average node degree.

Figure 6.16: Snapshots of a network with $D_e = 16.0 \times 10^{-21}\,\text{J}$ and $\varepsilon_{\text{HH}} = 16.0 \times 10^{-21}\,\text{J}$ across a sinusoidal simulation with $\lambda_{\text{max}} = 0.100$ and period $1.00\,\mu s$. The periodic box is shown in blue. The rupture happens at the defect site initially at a body-head bond and cannot be repaired; each oscillation introduces more damage.



(a) $D_e = 16.0 \times 10^{-21}$ J, $\varepsilon_{\mathsf{HH}} = 16.0 \times 10^{-21}$ J

(b) $D_e = 24.0 \times 10^{-21}$ J, $\varepsilon_{\mathsf{HH}} = 24.0 \times 10^{-21}$ J

(c) $D_e = 32.0 \times 10^{-21}$ J, $\varepsilon_{\mathsf{HH}} = 32.0 \times 10^{-21}$ J

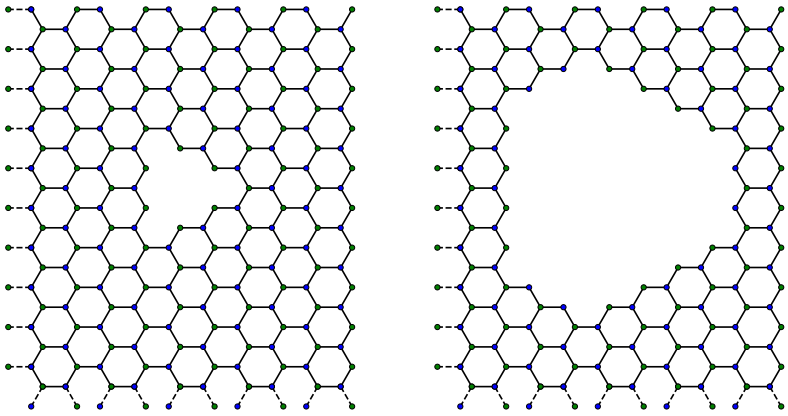
(d) $D_e = 40.0 \times 10^{-21}$ J, $\varepsilon_{\mathsf{HH}} = 40.0 \times 10^{-21}$ J

Figure 6.17: Damage metrics for networks with removed edges, or removed edges and defects. These lines track number of bonds broken N_{broken} and average degree $\langle k \rangle$ for networks with a removed edge (solid lines) or a removed edge and a Stone-Wales defect (dotted lines) stretched sinusoidally with different potential model parameters.

6.3.6 Central Hole Defects.

The defects studied thus far have shown some interesting effects, including the importance of flexibility. However, they have not provided much insight into rupturing in networks that have suffered significant damage. To study rupture after significant damage, the easiest solution is to introduce a larger defect: instead of studying the removal of a single edge, a hole can be introduced by removing many edges. Creating a central hole in a hexagonal network can be done by removing a single node from the graph, and then removing all other nodes that are within m edges from the initial node. Again, once nodes are removed, the positions must be re-optimised and edges converted into polymers.

This approach may be useful to understand surgical damage, as surgery involves cutting



(a) A hole of size 1 — an edge has been removed, as have all edges connected to it.

(b) A hole of size 5 — an edge has been removed, as have all edges within 5 connections of it.

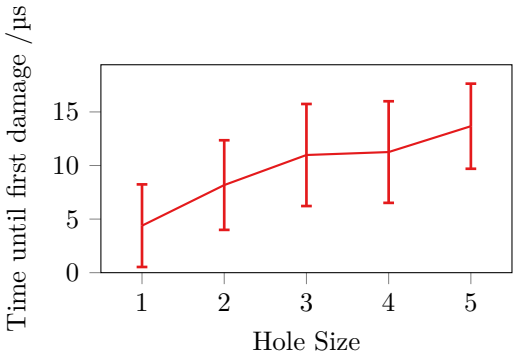
Figure 6.18: Graph representations of two networks with central holes of different sizes removed. The triangular nature of the holes is because all nodes within m edges of a central node have been deleted. This was chosen to potentially create weak points at the vertices.

a hole in the lens capsule (although the length scales are difficult to capture in computer simulation).

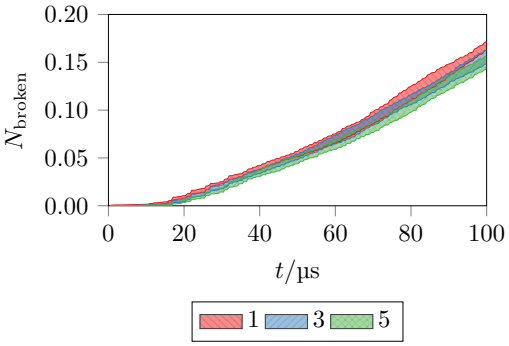
Two such networks with holes in are shown in **Figure 6.18a** and **Figure 6.18b**. The hole creation process described here creates triangular holes, chosen to potentially create weak points at their vertices. Holes of size 1 to 5 were set up, and the networks subjected to a strong stretching procedure of 40 increasing oscillations over 100 µs up to a maximum amplitude of $A = 1.5$. All other parameters remained set at the defaults, and each simulation was repeated 20 times with a different random seed.

Damage metrics calculated across these simulations are shown in **Figure 6.19**. The first observation is that networks with larger holes suffered initial damage later. This is not surprising, as networks featuring large holes contain fewer polymers. With fewer polymers, it becomes less likely that any one of them will break randomly.

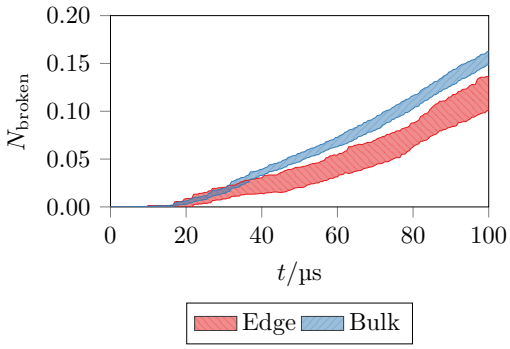
More interestingly, networks with larger holes suffered less damage both in absolute terms (a larger number of polymers) and in relative terms (a larger fraction of polymers). This effect can be seen in **Figure 6.19b**. Larger holes led to less damage over time because of flexibility.



(a) Time until the first bond is broken as a function of hole size. Larger holes led to later initial damage, as there are fewer polymers that have the chance to break.



(b) Damage over time as a function of hole size. Larger holes led to a lower fraction of polymers being damaged, despite system size effects. The legend lists the number of hexagons removed in the hole.



(c) Damage over time for a network with a hole of size five, tracking bonds on the edge of the hole against bonds ('edge') against bonds away from the hole ('bulk').

Figure 6.19: Three metrics of damage resulting from stretching networks with holes in. The damage metrics show the importance of flexibility around the central hole, as polymers near the edge break less, leading to less overall damage in a network with a larger hole removed. Filled areas represent one standard deviation over ten repeated simulations.

When the network was stretched in the x direction it was compressed in the y direction. The central hole provided a space that the nearby polymers could deform into during compression. If this space were not available, the network would have to rupture to relieve the stress.

This explanation is reinforced by analysing the rate of damage for polymers on the edge of the hole compared to the bulk. **Figure 6.19c** shows an average damage trajectory for a network with a hole of size 5. The polymers on the edge break much more slowly over time, although there is more uncertainty about when they will break. The bulk polymers break

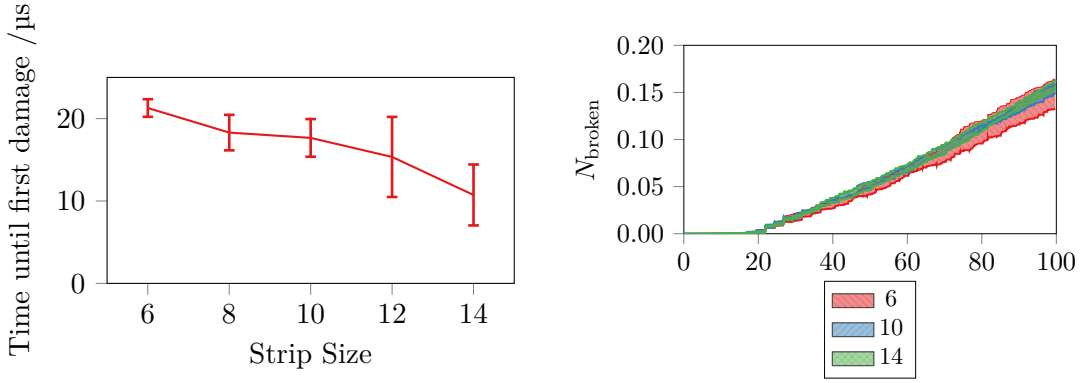
more rapidly over the course of the simulation as they cannot relieve any stress by deforming into the hole.

6.3.7 Thin Strip Defects.

A problem with putting holes in the network is that the holes appear in repeated periodic images of the simulation cell. If the hole is large compared to the simulation cell, then sections of network between holes can become very thin. To simulate macroscopic damage, an extremely large system must be used, which would be very computationally expensive. Thankfully, any truly macroscopic damage is effectively infinitely large compared to the length scale of a polymer. Simulating a thin strip of a network effectively mimics rupturing along the boundary of a macroscopic hole. Periodic boundary conditions are limiting here, as the network cannot only be open on one side. The network is repeated in the y dimension, with open regions repeated in the x direction. A wide strip must be used to prevent the interaction of two boundaries dominating.

The strips then underwent a stretching simulation repeating the process from **Section 6.3.3**. As with **Section 6.3.6**, each strip underwent 40 oscillations over 100 µs. The amplitude increased per oscillation up to $A = 1.5$.

Damage metrics for strips of different widths are shown in **Figure 6.20a** and **Figure 6.20b**. The time until first damage in **Figure 6.20a** shows that larger strips are damaged first. Earlier initial damage occurs for larger strips again because there are more constituent polymers. The relative fraction of polymers that break remains similar, but the absolute number of broken polymers increases. This is shown in **Figure 6.20b**. The effect becomes more significant as the strips get larger, as a greater fraction of polymers are found in the bulk. As was found for the simulations of networks with holes removed, polymers on the edge break more slowly



(a) Time until the first bond is broken as a function of the strip width (defined as the number of hexagons perpendicular to the edges). Larger strips see earlier instances of initial damage.

(b) Fraction of bonds broken over time as a function of strip width (as highlighted in the legend). Over time, all the strips behave similarly. The larger strips show smaller error bars in their trajectory there are more molecules to average over.

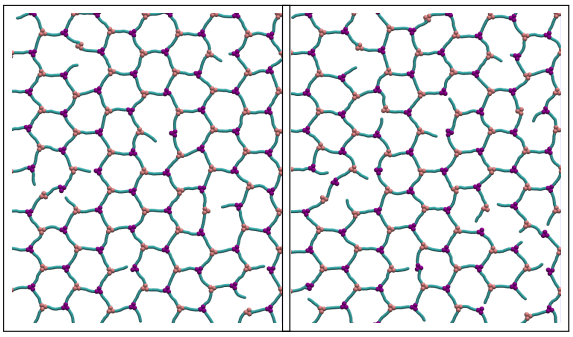
Figure 6.20: Time until first damage and the damage trajectory over time for different sizes of strip. The larger strips break sooner due to a larger number of constituent polymers, but the damage as a fraction of total bonds remains approximately constant. This figure is adapted from M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180.

as they have more available area to distort into. With a smaller fraction of polymers on the edge, breakages in the bulk become even more likely.

This finding agrees with the conclusions from studies of networks featuring holes and defects, showing that network flexibility is critical to prevent rupture. Strips of hexagonal networks showed the same trends as before, with the polymers on the edge of the network being less likely to break because they could relieve strain by distorting. One key difference between strip networks and networks with holes in is that the holes were approximately circular. Strip networks broke sooner, as they could only deform in one axis to relieve strain.

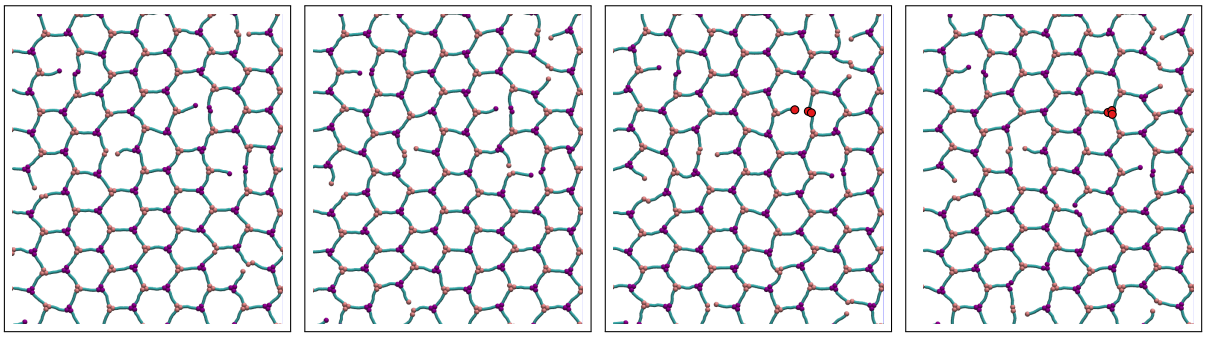
6.3.8 Repairable Damage.

This chapter has focussed on irreparable damage. However, variance in $\langle k \rangle$ shows that this is not the only form of damage that a network can suffer. A schematic comparison of damage that can be repaired against damage that cannot is shown in **Figure 6.21**. Panels in the



(a) Irreparable damage appears at $t = 0.500\,\mu s$.

(b) ... and is still visible at $t = 1.00\,\mu s$.



(c) Head group damage appears at $t = 0.500\,\mu s$.

(d) ... at $t = 1.00\,\mu s$.

(e) ... a new rupture appears at $t = 1.50\,\mu s$.

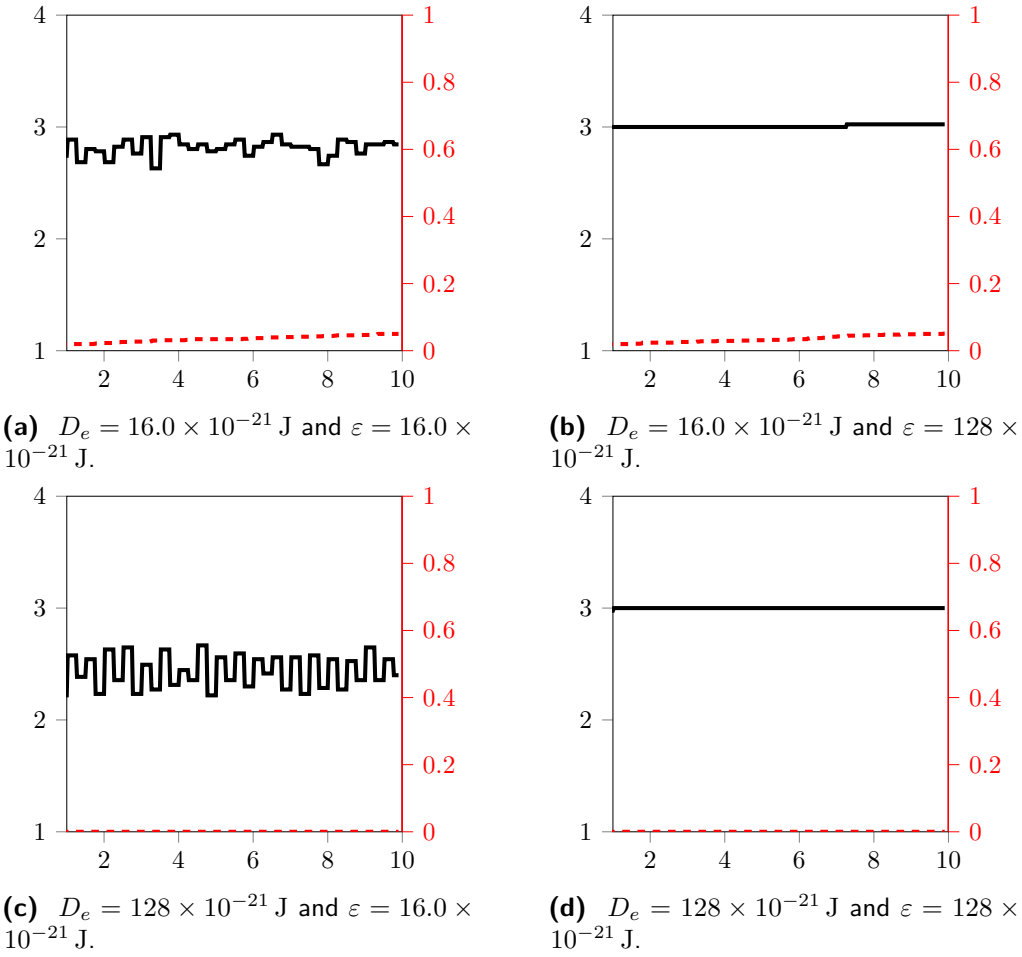
(f) ... and at $t = 2.00\,\mu s$ that damage has been repaired

Figure 6.21: Panels **(a)** & **(b)**: bond breaking damage in this network with $D_e = 16.0 \times 10^{-21}$ J and $\varepsilon_{\mathrm{HH}} = 128 \times 10^{-21}$ J cannot be repaired.
Panels **(c)**-**(f)**: head group damage in this network with $D_e = 128 \times 10^{-21}$ J and $\varepsilon_{\mathrm{HH}} = 16.0 \times 10^{-21}$ J can be repaired. Some nodes that undergo repair are shown as enlarged red circles.

top row show the damage to a network with $D_e = 16.0 \times 10^{-21}$ J and $\varepsilon_{\mathrm{HH}} = 128 \times 10^{-21}$ J. This network suffers damage within a polymer which cannot be repaired, and remains broken forever.

Panels in the bottom row show the damage to a network with parameters $D_e = 128 \times 10^{-21}$ J and $\varepsilon_{\mathrm{HH}} = 16.0 \times 10^{-21}$ J. It suffers damage at the nodes (made of clusters of head groups) which can be repaired during small oscillations as the head groups can re-coordinate. Damage at nodes can often be suffered and repaired repeatedly.

Damage metrics are shown in **Figure 6.22** for networks at the high and low ends of the parameter spectrum. Irreparable damage is seen in **Figure 6.22a** and **Figure 6.22b** as



(a) $D_e = 16.0 \times 10^{-21}$ J and $\varepsilon = 16.0 \times 10^{-21}$ J.

(b) $D_e = 16.0 \times 10^{-21}$ J and $\varepsilon = 128 \times 10^{-21}$ J.

(c) $D_e = 128 \times 10^{-21}$ J and $\varepsilon = 16.0 \times 10^{-21}$ J.

(d) $D_e = 128 \times 10^{-21}$ J and $\varepsilon = 128 \times 10^{-21}$ J.

Figure 6.22: Damage metrics for networks stretched sinusoidally with different parameters. The networks with high D_e and low ε are damaged preferentially at the nodes, leading to a drop in $\langle k \rangle$. However, this damage can be repaired leading to a sawtooth shape.

the red line rising monotonically. Self-repair is visible as a sawtooth shape in the black $\langle k \rangle$ curves, such as in **Figure 6.22c**. Networks that are only damaged a little and then repair can be seen as lines that are close to $\langle k \rangle = 3$ but briefly dip below that. It is clear that these metrics are averaged across too many rings and do not highlight the important differences in rupturing behaviour, or ring structure. A new set of metrics must be developed to cope with repairable damage, which will be discussed in **Chapter 8**.

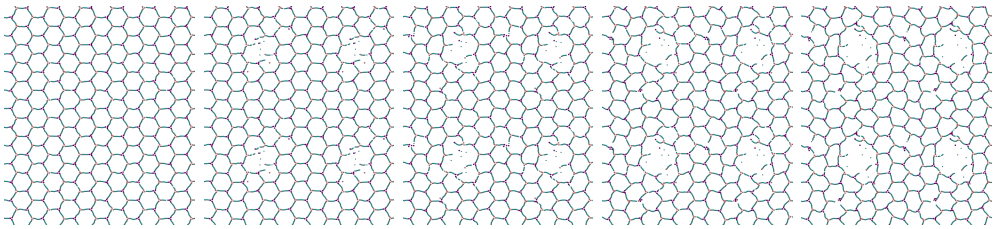


Figure 6.23: Snapshots of a network with $\gamma = 100\,\text{ns}$ with a burst temperature of $30\,000\,\text{K}$ over the first five microseconds.

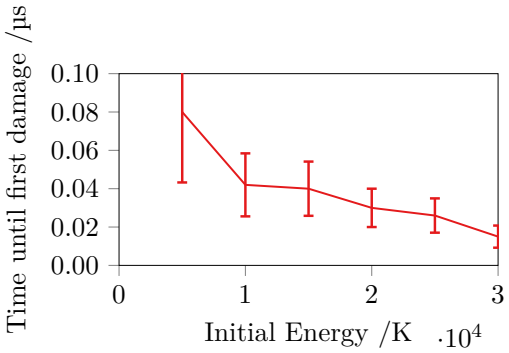
6.3.9 Burst Temperature.

The stretching process introduces damage evenly across a network, which is not suitable to study localised damage. One way to introduce localised damage is to add a large amount of energy into one polymer. This could either be potential energy by displacing the atoms or kinetic energy by assigning a large random velocity. The introduction of kinetic energy is easier than potential energy, as there is less chance of introducing artefacts into the simulation.

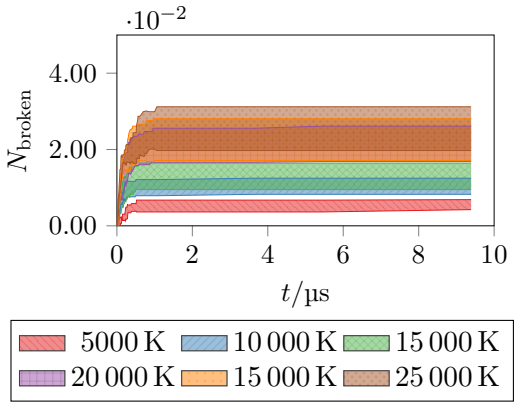
In this section, idealised hexagonal networks were set up with $D_e = \varepsilon = 24.0 \times 10^{-21}\,\text{J}$. The damping parameter for the Langevin thermostat (first discussed in **Section 4.2.4**) was set to $\gamma = 100\,\text{ns}$. In each network, one polymer given kinetic energy equivalent to $5000\,\text{K}$, $10\,000\,\text{K}$, $15\,000\,\text{K}$, $20\,000\,\text{K}$, $25\,000\,\text{K}$ or $30\,000\,\text{K}$, while the rest of the molecules were assigned kinetic energy equivalent to $100\,\text{K}$.

The snapshots over the first $5.00\,\mu\text{s}$ are shown in **Figure 6.23** for a simulation with a burst temperature of $30\,000\,\text{K}$. These show how a hole forms around the damage site and is propagated by fragments of the damaged molecule.

The time to initial damage is shown in **Figure 6.24a**, and the damage trajectory is shown in **Figure 6.24b**. These show that the initial burst temperature is correlated with the initial rate of damage, but the damage trajectories rapidly become parallel after $30.0\,\mu\text{s}$ to $50.0\,\mu\text{s}$. This is due to the effect of the Langevin thermostat removing thermal energy from the system.



(a) Time to first damage as a function of added kinetic energy. This counts the time it takes the first polymer (the one with added energy) to explode.



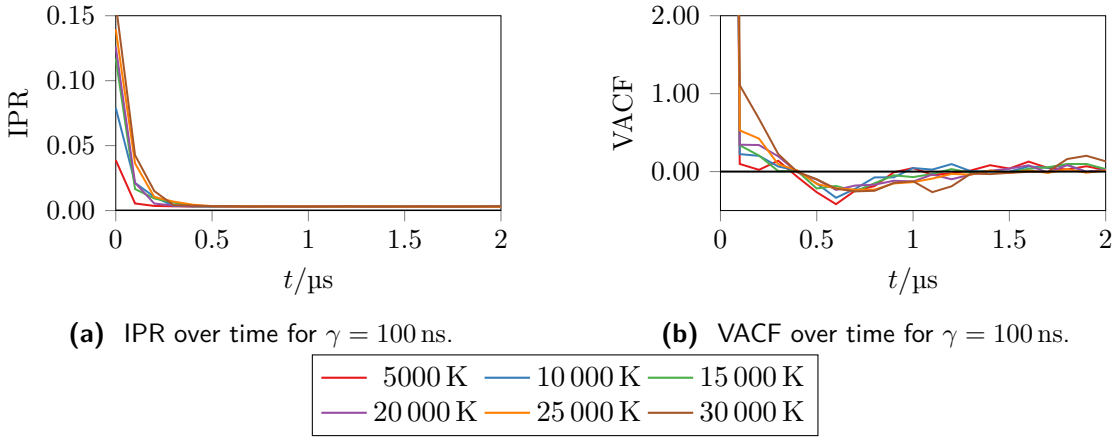
(b) Damage trajectories as a function of added kinetic energy. The filled area represents one standard deviation over ten simulations, and the legend indicates the effective added kinetic energy in temperature units.

Figure 6.24: Damage metrics for networks with added kinetic energy.

Figure 6.25a and **Figure 6.25b** show the IPR and VACF over time respectively. In this case, neither of the localisation metrics proves to be particularly useful. This is because they both require sampling considerably faster than the timescale of damping, γ. Sampling velocities very frequently takes up large amounts of disk space and processing time. VACF is slightly more useful than IPR, although both drop to near-zero rapidly. VACF shows a characteristic rebounding around 0.600 ns for all temperatures. This rebounding is where beads have stopped accelerating, often due to collisions, and are travelling in the opposite direction.

6.3.10 Varied Thermostat Parameters.

Section 6.3.9 showed that the spread of kinetic energy through the system is affected by the thermostat. This is problematic, as the parameters in the thermostat are only semi-physical and cannot be measured experimentally. When parameters cannot be known experimentally, it is useful to scan across a landscape and establish patterns of behaviour. The experiments



(a) IPR over time for $\gamma = 100\,\text{ns}$. **(b)** VACF over time for $\gamma = 100\,\text{ns}$.

Figure 6.25: IPR and VACF over the course of the simulation. VACF shows a large initial spike at $t \approx 0\,\text{ns}$, but quickly drops to just being random noise. IPR is also initially high, but rapidly drops to near zero.

from **Section 6.3.9** were repeated, but with different values of the thermostat damping parameter γ. γ was given values 1.00 ns (most damping), 10.0 ns, 100 ns or 1000 ns (least damping).

Damage metrics are shown in **Figure 6.26** and show a significant dependence of rupturing behaviour on damping parameter. A small γ causes the thermostat to absorb all energy before it can cause damage (**Figure 6.26a**), but a large γ (**Figure 6.26d**) leads to high energy fragments causing damage over long periods of time. This effect is less significant when less energy is added — the damage trajectories for 5000 K are similar across $\gamma = 10.0$ ns and 1000 ns. This is because the network structure can absorb that energy by deforming. However, for larger added temperatures such as 30 000 K the fragments can travel around the network for a long time and cause a significant amount of damage, as is seen in **Figure 6.26d**.

Damage trajectories at the same added temperature and different γ in **Figure 6.27** show this more clearly. This pattern is repeated if the same data are instead grouped by burst temperature in **Figure 6.27**. The shortest damping timescale often leads to no damage, as the damping effect occurs faster than the timescale of damage. The largest damping timescale leads to lots of damage, especially when more kinetic energy is added.

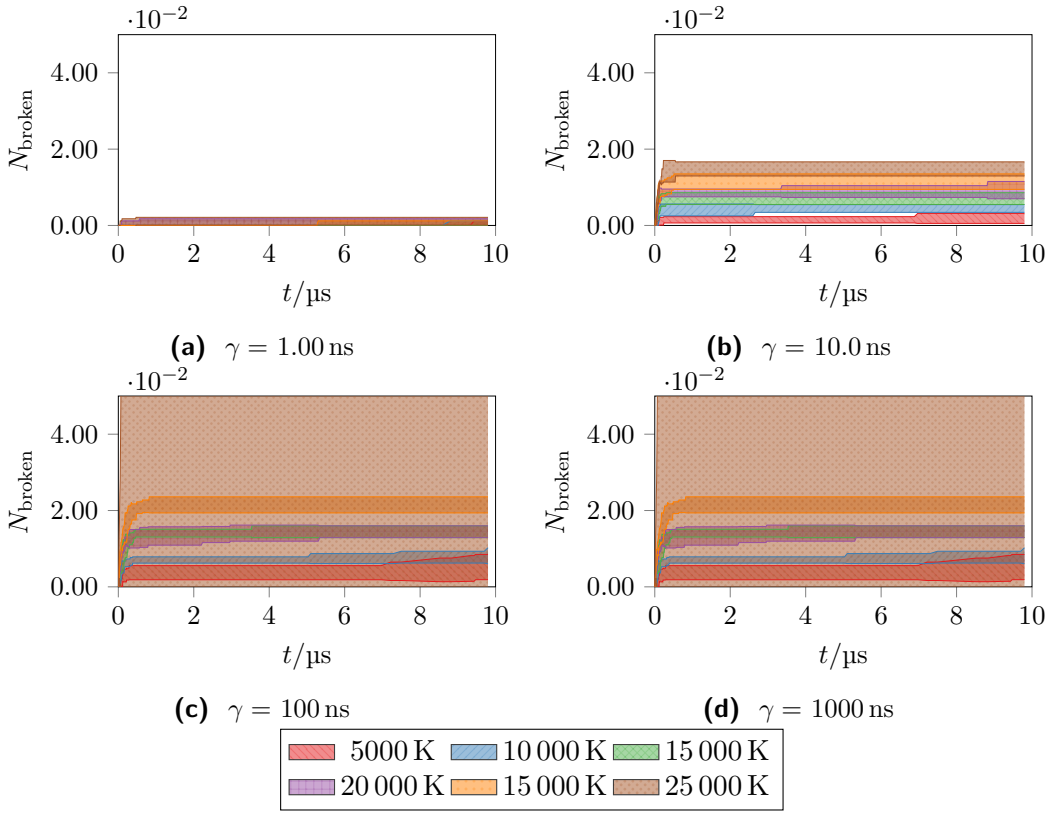


Figure 6.26: The effect of damping parameter γ on network damage. The filled area represents one standard deviation of the amount of damage over ten simulations.

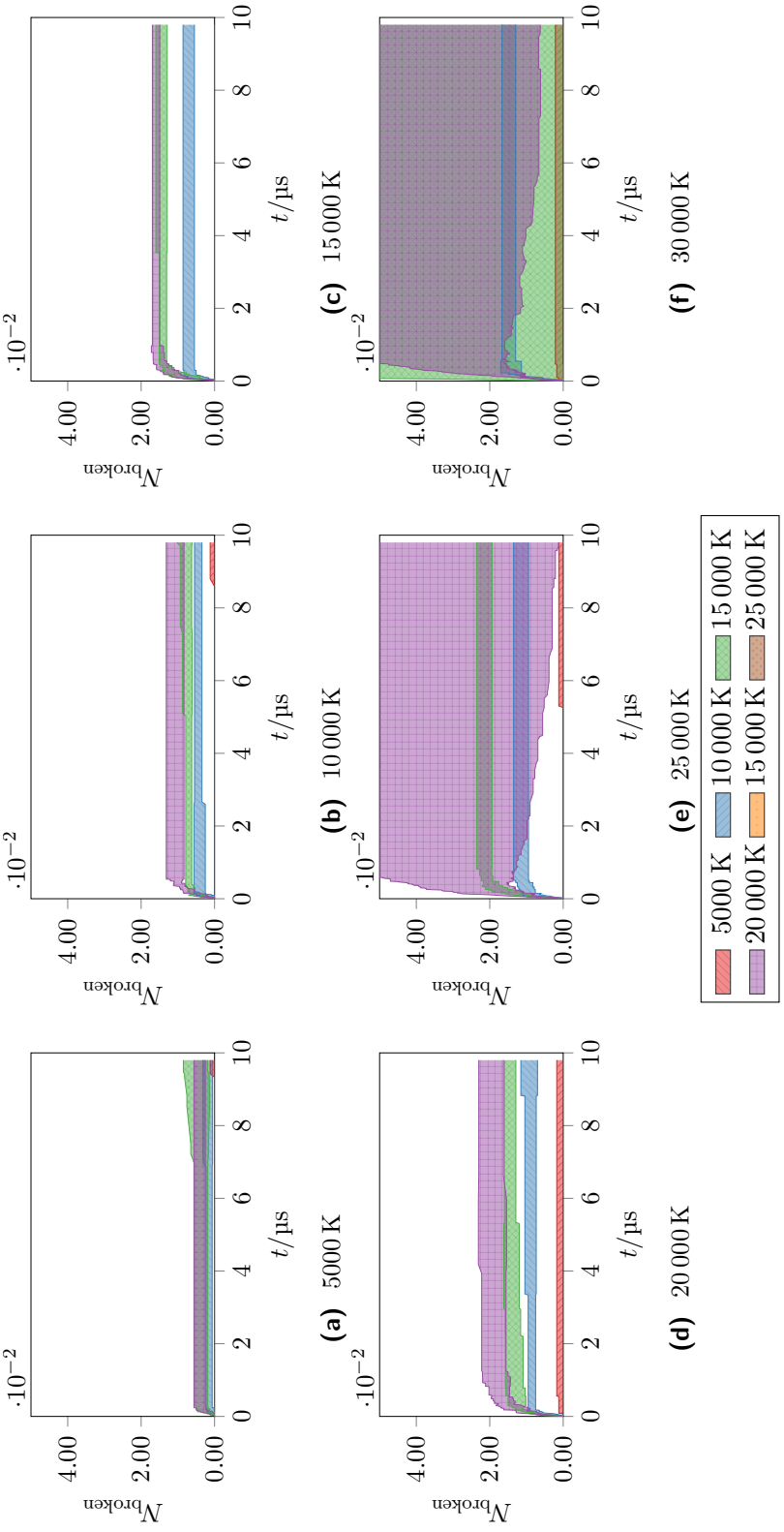
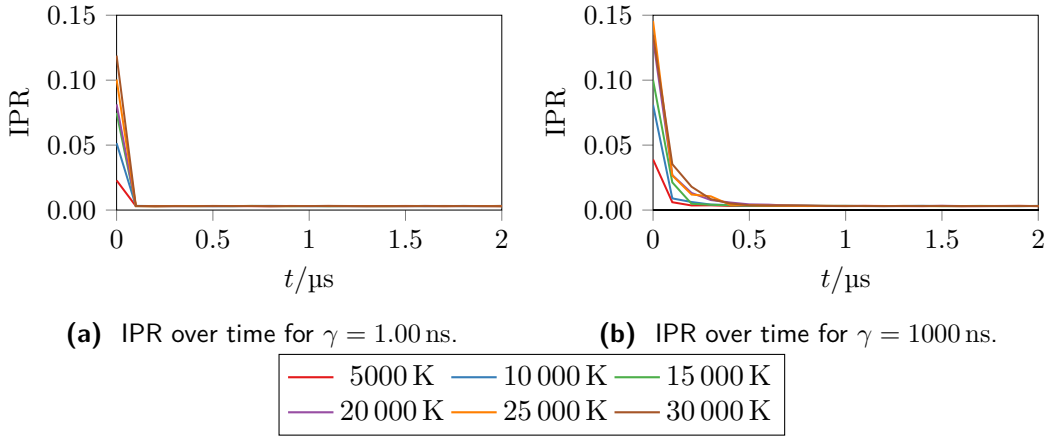
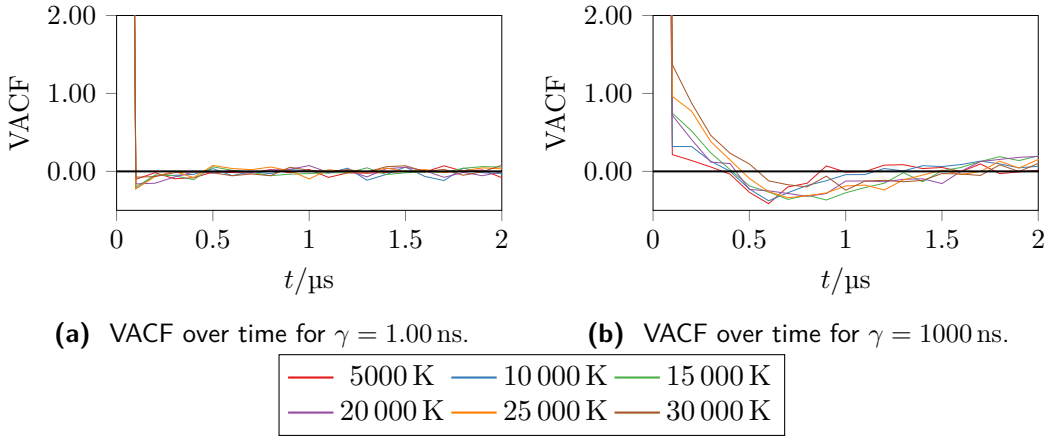


Figure 6.27: Damage trajectories at the same burst temperature but different γ. The filled area represents one standard deviation over the damage over ten repeated simulations.



(a) IPR over time for $\gamma = 1.00\,\mathrm{ns}$. **(b)** IPR over time for $\gamma = 1000\,\mathrm{ns}$.

Figure 6.28: IPR over the course of the simulation at different values of thermostat parameter γ.



(a) VACF over time for $\gamma = 1.00\,\mathrm{ns}$. **(b)** VACF over time for $\gamma = 1000\,\mathrm{ns}$.

Figure 6.29: VACF over the course of the simulation at different values of thermostat parameter γ. This shows a large initial spike at $t \approx 0\,\mathrm{ns}$, but quickly drops to just being random noise. The right panel, with a longer damping timescale, shows a characteristic rebounding effect.

6.4 Chapter Conclusions.

This chapter documented explorations of how coarse-grained polymer networks rupture. MD was used to simulate smooth damage trajectories, to improve on the sudden damage observed in the MC method of **Chapter 5**. The rupturing process was first examined through a lens of the normal modes of the system, examining the natural motion around a defect. The eigenvector analysis showed a clear preference for rupturing around a Stone-Wales bond switching defect.

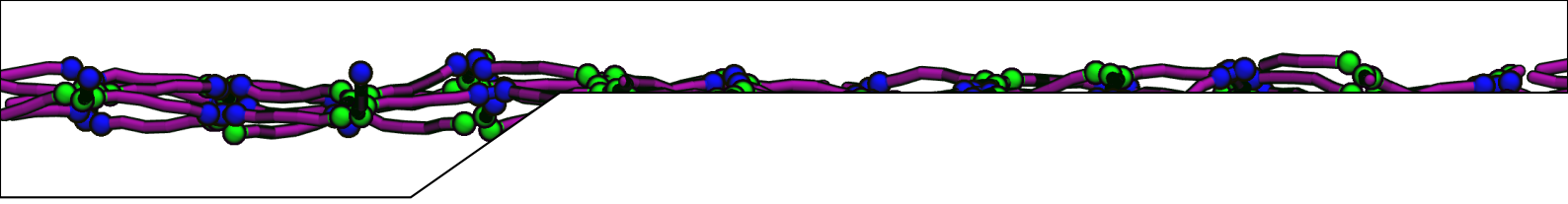
The rupturing behaviours were then examined in terms of a balance between intermolecular damage (head group dissociation) and intramolecular damage (bond breaking). Extremes of damage were established as the ratio of the breaking energies were changed with respect to thermal energy. A single rupturing trajectory was analysed with the energy required to activate the two damage modes balanced closely against the thermal energy.

Damage trajectories of idealised networks undergoing stretches at a constant engineering strain rate highlighted the need for more sophisticated metrics, which will be covered in **Chapter 8**. The idealised networks were then stretched sinusoidally at different frequencies and maximum amplitudes, showing a linear increase in damage over a fixed trajectory as a function of both ω and A. Moving on from idealised networks and simple stretching, the effect of flexibility was explored by removing bonds and stretching at increasing amplitudes every cycle. This showed a key trade-off between flexibility and structural strength, which affected how long it took a given network to be damaged.

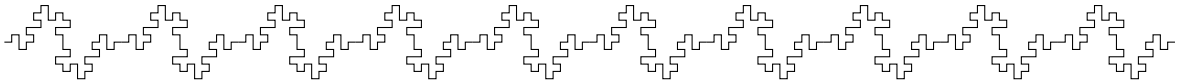
The bond removal defects were combined with bond switching defects and underwent the same increasing stretching procedure. This showed that bond switching defects provide areas of flexibility, but also high stress, which will be explored in **Chapter 7**. Large-scale defects

were then introduced of central holes and thin strips of network, which were also stretched until they ruptured. This reinforced conclusions about flexibility, as polymers on the edge of a large scale defect were less likely to break.

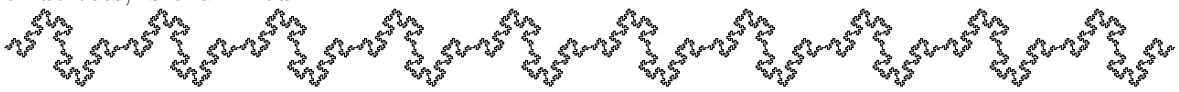
Finally, networks were damaged by adding a large amount of kinetic energy to a single polymer. The resultant damage was strongly dependent on thermostat parameter γ which dictated how energy spread through the network.



Chapter 7: Rumpling.



No real network is truly two-dimensional, but instead will 'rumple' into a third dimension to maximise its entropy. In this chapter, the degree to which collagen IV networks rumple into the third dimension is examined. This is used to assess how accurate it is to approximate networks as 2D. The nature of the rumpled surface is described using a fractal self-affine model, where the roughness is self-similar across length scales. A Molecular Dynamics (MD) approach is used to calculate the height-height correlations between beads in the network, with motion allowed in the z direction. A self-affine scaling law is fitted to the correlation functions, revealing characteristic fractal parameters. How those fractal parameters (especially surface width) vary as a function of temperature, stretching or compression, and the presence of defects, is examined.



7.1 Introduction.

Chapter 2 covered image analysis methods on microscope images of networks. Microscopes are powerful tools that give their users a great deal of detail about a small region within the network under study. However, the images taken via a microscope are inherently two dimensional (2D) (stacks of 2D slices through a three dimensional (3D) network). This is excellent for examining materials that comprise layers of 2D networks, such as collagen IV in

the ocular lens capsule[58]. However, the experimental data available inform the approaches used in **Chapter 3** and **Chapter 4**, which were two different 2D models of collagen-like networks. 2D models are often seen in the literature, both to match experimental results and for computational ease[14,71,146]. The removal of a degree of freedom makes calculations faster and easier to interpret.

However, no real network is purely 2D, but instead has a finite width owing to a combination of the physical size of the molecules and entropic effects, which favour utilising an extra degree of freedom. When the 2D constraint on a network is relaxed and the system is allowed to evolve (above 0 K) it is likely to rumple and become rough. Macroscopically, this is similar to how a piece of paper rolled into a ball then flattened out again will rumple. Close inspection of the roughness of many surfaces reveals that the patterns in local roughness are similar to the patterns of roughness across the wider network. This is described as being self-affine: patterns on different length scales are similar, but not necessarily in the same scaling ratio per axis.

Surface roughness (sometimes referred to as 'nanocorrugation') has been studied in great detail for 2D networks such as graphene[147,148]. For many atomic surfaces, the roughness affects thermal and electrical conductivity. It is often measured by atomic force microscopy (as seen in **Chapter 2** being used for collagen IV) which directly records the height of a surface, or where that method does not work, angle resolved light scattering can be used to reconstruct the height of a surface[149]. Thankfully, in MD simulations the height of a surface is obtained easily every step, and correlation functions can be computed without worrying about limited resolution.

Self-affine surfaces are neither truly 2D (they have a width into the 3$^{\text{rd}}$ dimension) nor fully 3D (they are primarily still a surface) but somewhere in-between. It is most useful to

describe self-affine surfaces using a fractal approach, as pioneered by Mandelbrot[150]. A fractal description of a surface provides a series of metrics, including the fractal dimension, which measure the degree to which it is 2D or 3D[151].

7.2 Computational and Mathematical Methods.

7.2.1 Scaling Laws and Surface Roughness.

Fractal surfaces are governed by statistical mechanics, with finite size scaling laws for some metric σ as a function of separation x often taking the form

$$\sigma(x) \propto L^{\beta} f\left(x/L^{\gamma}\right) \tag{7.1}$$

with L being a characteristic length scale, while β and γ are critical exponents[152].

The scaling of interest here is the roughness of the surface, which can be be measured by a height-height correlation function of the form

$$H(\boldsymbol{r}) = \left\langle [z(\boldsymbol{r}) - z(\boldsymbol{0})]^2 \right\rangle \tag{7.2}$$

with $z(\boldsymbol{0})$ being the height at some arbitrary point and $z(\boldsymbol{r})$ being the height at vector position $\boldsymbol{r}$ from that point. A flat surface has zero roughness, and a greater $H(\boldsymbol{r})$ indicates a rougher surface. For an isotropic surface $\boldsymbol{r}$ can be replaced with the separation r, and a histogram with bins of size δr can be used to interpret this function.

If the surface is truly self-affine then it is possible to fit a fractal scaling law to the height-height correlation of the form

$$H(r) = 2w^2 f\left(r/\xi\right) \tag{7.3}$$

in which f is a scaling function that obeys

$$f(x) = \begin{cases} x^{2\alpha} & \text{if} \quad x \ll 1 \\ 1 & \text{if} \quad x \gg 1 \end{cases} \tag{7.4}$$

A common function used for this purpose is[153,154]

$$f(x) = 2w^2 \left[1 - \exp\left(-(r/\xi)^{2\alpha}\right)\right], \tag{7.5}$$

which is chosen because it has the correct scaling properties and has smooth derivatives.

This leads to the three descriptors of the surface: w, the width of the surface; ξ, the local roughness length scale; and α, the Hurst exponent, which is related to the fractal dimension of the surface (the degree to which the surface is between 2D and 3D)[155,156]. By inspecting the values of w, α and ξ it is possible to assess the accuracy of the 2D approximation used in this thesis.

7.2.2 Simulation Protocol.

All previous simulations in this thesis forced the coarse-grained polymers or network models to be in 2D. This was achieved by setting the z component of positions and velocities to zero every simulation step, or simply excluding the z component. The 2D constraint was useful for network formation, as the reduced degrees of freedom accelerate assembly. It also meant that the resultant network is 2D and initially free from any 3D complexity.

Throughout this section, the coarse-grained model from **Chapter 4** is re-used. The head groups are again attracted to one another with a Lennard-Jones potential featuring energy parameter $\varepsilon_{\mathrm{HH}} = 24.0 \times 10^{-21}\,\mathrm{J}$, and length scale $\sigma_{\mathrm{HH}} = 50.0\,\mathrm{nm}$. The bonds within a polymer are harmonic, with energy parameter $k = 1.66 \times 10^{-4}\,\mathrm{N\,m^{-1}}$ and equilibrium length $r_{\mathrm{eqm}} = 50.0\,\mathrm{nm}$.

Initial simulations in this section first equilibrate a pre-formed network for $0.100\,\mu\mathrm{s}$ at a

fixed temperature T. After the initial equilibration, the $z = 0$ constraint is removed and the system is allowed to rumple into the new degree of freedom at a chosen temperature. The chosen temperatures are relatively cool to avoid damaging the network.

More complex simulations then involve scanning over bond strength parameters and temperatures. Some simulations in this section will involve applying a sinusoidal stretching force of up to 25.0 % maximum stretching or compression. The low amplitude of stretching is to preserve the structure of the network in the plane, while allowing for tension to flatten or compression to rumple the network.

These stretching simulations are performed in two types: the first are volume preserving, such that any change in x is accompanied by an inverse change in y. The second are volume changing, such that any change in x is matched with no change in y. Both sets of simulations are sinusoidal with the same maximum amplitude and frequency with L_x, the length of the simulation cell in the x direction at time t, governed by

$$L_x(t) = L_x(0)\left[1 + A\sin\left(\frac{2\pi}{T}t\right)\right]. \tag{7.6}$$

In this equation A is the amplitude of the oscillation (chosen here to be $A = 0$ to 0.250) and T is the timescale of the oscillation (here chosen to be 1.00×10^4 ns) unless otherwise specified in the relevant section.

Most networks in this section will be idealised hexagonal networks. Defects will be introduced using the Wooten-Winer-Weaire (WWW) scheme discussed previously in **Section 3.2.4**.

7.3 Results and Discussion.

As in previous chapters, an idealised hexagonal network of coarse-grained biopolymers was initially chosen to explore the effects of relaxing the 2D simulation constraint. The idealised

hexagonal network was chosen as it is a simple, stable network structure that is at an energy minimum when all edges are coplanar.

7.3.1 Hexagonal Network Widths.

To examine the natural width and roughness of a collagen IV-like network, a simple network was allowed to undergo thermal motion. The network was first equilibrated with the 2D constraint enabled for 0.100 µs; the constraint was removed for the remaining 100 µs of the simulation at 100 K and the roughness of the network was tracked over time.

Snapshots of the network evolving over time into the newly accessible third dimension are shown in **Figure 7.1d(a-c)**. In these figures, the network lies mostly in its original plane shortly after the constraint was relaxed. In **Figure 7.1b**, it is already showing some characteristic rumpling patterns very shortly after the constraint was relaxed. After a longer period of time, in **Figure 7.1c**, the network is more clearly lying outside its original plane with large undulations. However, these undulations simply move about the surface without changing total roughness. This is because the overall roughness of the surface is controlled by the balance of enthalpy (the energetic penalty to leave the minimum energy coplanar arrangement) and entropy (more possible arrangements of polymers outside of the plane).

The time evolution of the roughness of the surface formed by the hexagonal network is shown in **Figure 7.1d**, showing the roughness immediately after the constraint is relaxed to be near zero and increasing over time until a maximum roughness is reached. The maximum roughness is limited by the balance of thermal energy to bond strengths (both the bonds between pairs of beads and the angular bonds between each three beads) as the out-of-plane motion requires a lengthening of bonds and a distortion of angles.

An examination of **Figure 7.1** shows some of the key regions in the height-height

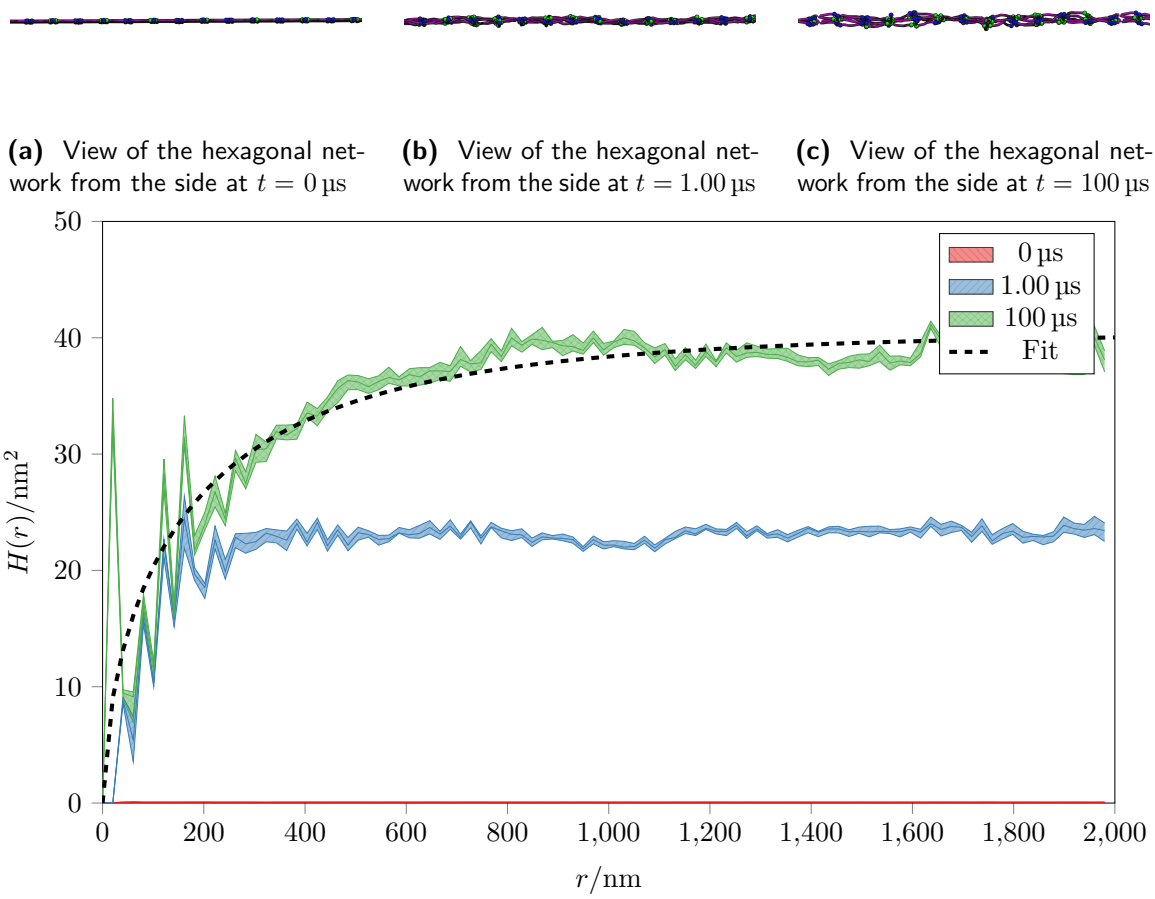
correlation graphs:

- An initial jagged region representing the local roughness. The jaggedness of this region is because there is a minimum separation between beads of 50.0 nm, and not many available samples. This provides information on ξ and α.
- A smooth transition between the jagged initial region and the bulk, showing the interaction between the local and global roughness. This also provides information on ξ and α.
- A flat region representing the non-local roughness. This provides information on the width w of the surface.

This process was repeated for different temperature combinations to show that the width of the surface is dominated by thermal factors, and the results are in **Table 7.1**. These show that the local roughness exponent α was consistently in the range 0.338 to 0.416, with a mean of 0.367. This is related to the fractal dimension D by the relationship $D = 2+1-\alpha$[156], implying a fractal dimension of 2.63. The size of the local roughness region ξ varies significantly, but here it is always similar to the length of a single polymer. Finally, the width of the network w increases with temperature to increase from 2.94 nm to 9.07 nm, which is small compared to the length of one polymer (300 nm).

T/K	ξ/nm	α/nm	w/nm
10.0	290	0.338	2.94
20.0	338	0.354	3.26
30.0	245	0.384	3.27
40.0	254	0.369	3.51
50.0	251	0.352	3.60
60.0	336	0.358	3.84
70.0	240	0.371	3.61
80.0	241	0.380	5.77
90.0	360	0.349	4.03
100	207	0.416	9.07

Table 7.1: Roughness fractal parameters for hexagonal networks with the 2D constraint relaxed at different temperatures.



(a) View of the hexagonal network from the side at $t = 0\,\mu s$

(b) View of the hexagonal network from the side at $t = 1.00\,\mu s$

(c) View of the hexagonal network from the side at $t = 100\,\mu s$

(d) Evolution of the roughness of a hexagonal network over time. The black dotted line is the best fit of the self-affine equation with parameters $w = 4.50\,\mathrm{nm}$, $\xi = 177\,\mathrm{nm}$ and $\alpha = 0.316$.

Figure 7.1: Three views of the hexagonal network, looking along the original 2D plane. The roughness of these networks is shown in **Figure 7.1d** as the height-height correlation function from **Equation 7.2**.

7.3.2 Deformation Induced Rumpling.

The change of 2D networks into partially 3D networks is intuitive on a macroscopic scale when considering surfaces such as cloth or paper — these are nearly perfectly 2D under tension, but 3D under compression. The same is true for surfaces on a microscopic scale, albeit with some small differences owing to the larger impact of thermal effects on a microscopic network[157]. To examine this, the ideal hexagonal network of coarse-grained polymers used in **Section 7.3.1** was stretched, as was previously shown to realistically induce damage in collagen-like networks. The stretching process was described in **Section 7.2.2**.

Volume preserving oscillations show the anisotropy of the hexagonal network and demonstrate the preference of a complex network to absorb strain by first deforming in the plane. When the networks are stretched along the long axis of the hexagons (and thus compressed along the short axis), the polygonal structure undergoes a hexagon-to-rectangle transition, often observed in inorganic glasses. The hexagon-to-rectangle transition is shown in **Figure 7.2**. This deformation means that the roughness of the network is effectively independent of the stretching ratio λ_x, as seen in **Figure 7.3a**.

However, when the network changes volume (that is, it maintains a constant L_y), the roughness changes significantly over time. This can be seen in the different $H(r)$ curves of **Figure 7.3b**. The curves show that the roughness increases dramatically when λ_x is small and decreases when λ_x is large, although not to zero. There is a significant increase of roughness shown in the small r region where the self-affine approximation breaks down due to the coarse-grained nature of the polymer beads used in this work; the polymer must form a zig-zag shape under compression with each 50.0 nm section being straight.

When the network is compressed, the roughness increases considerably as measured by the network width w increasing from 2.39 nm to 5.72 nm. This is shown in **Figure 7.4**, wherein

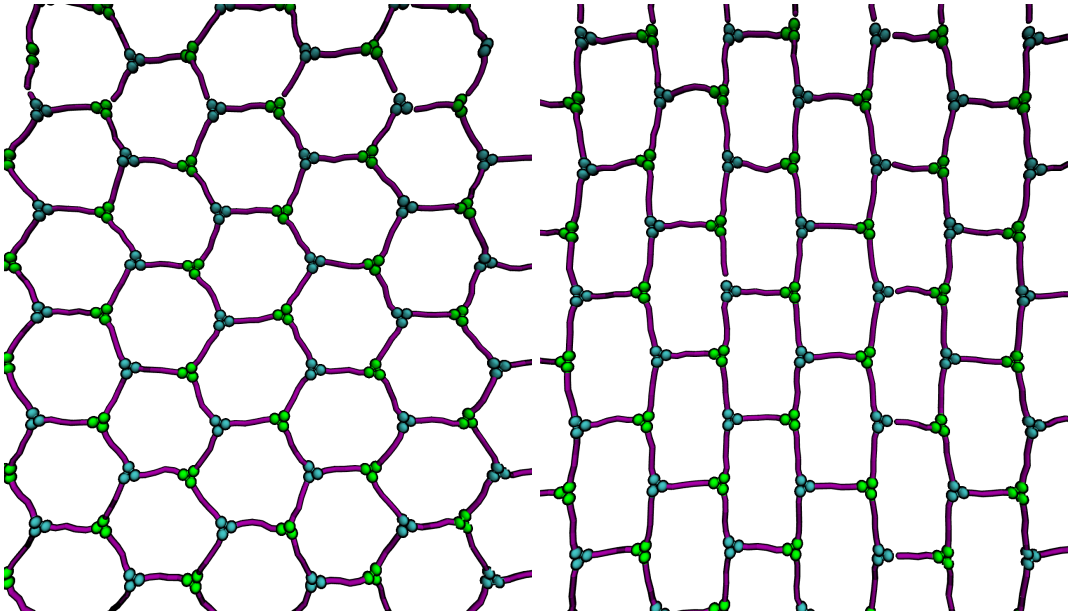
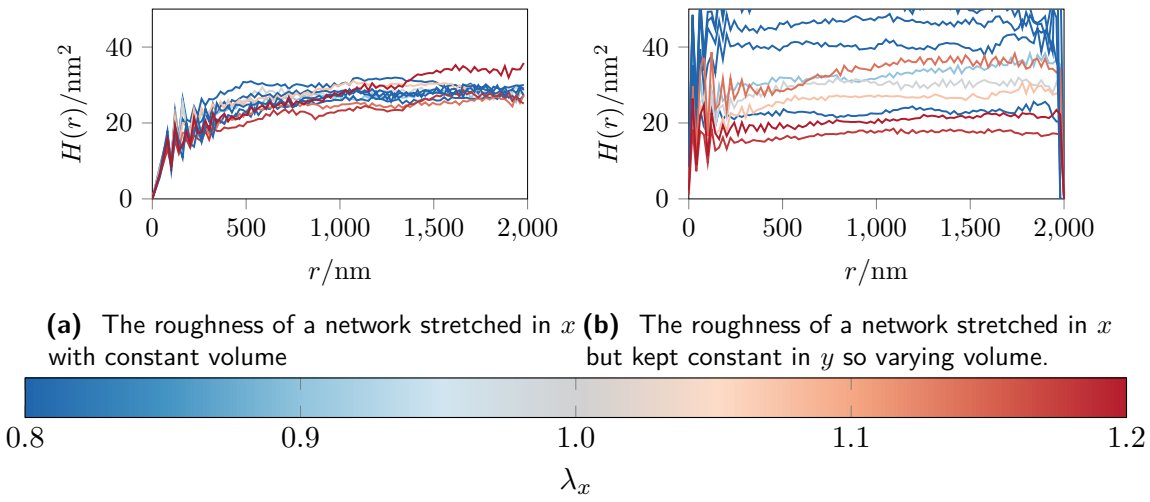


Figure 7.2: The hexagon-to-rectangle transition, shown as the same region of network under compression (hexagons) and tension (rectangles).

the largest w occurs slightly after the network is compressed. The maximum width is greater in the lower-frequency stretches than the high-frequency stretches owing to a small amount of lag as the polymers move after stretching. Because of the lag, the relatively soft network has not yet fully adjusted to the compressed simulation cell dimensions when the sinusoidal stretching begins expanding it again.

The minimum width under compression increases over time, because the stretching 'pumps' energy into the system. Short simulation timescales means that the network still has a memory of its previous rumpled arrangement when stretched out, and does not drop to zero width.



(a) The roughness of a network stretched in x with constant volume

(b) The roughness of a network stretched in x but kept constant in y so varying volume.

Figure 7.3: Roughness of the same network undergoing stretching at either constant (left panel) or changing (right panel) volumes. The lines are coloured by the simulation cell stretching ratio, $\lambda_x = L_x(t)/L_x(0)$.

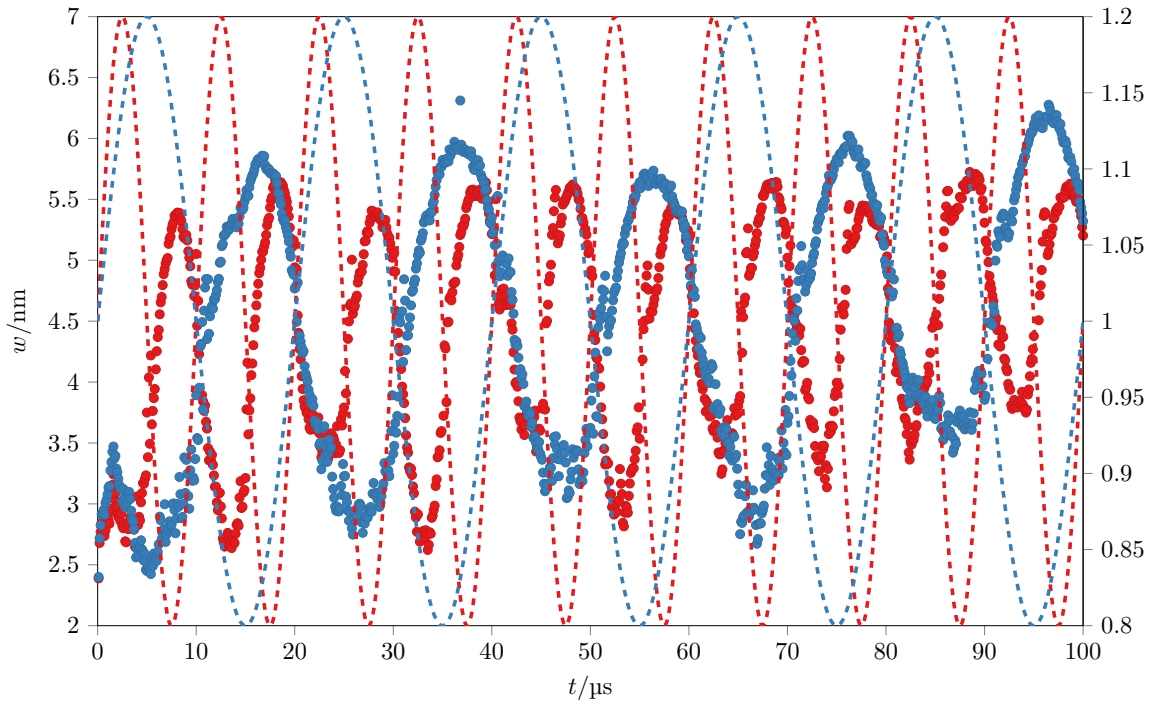


Figure 7.4: Widths during a stretching simulation at two frequencies, $0.0500\,\mu\mathrm{s}^{-1}$ (blue) and $0.100\,\mu\mathrm{s}^{-1}$ (red). The dots show the evolution of the width metric over time, and the dashed line shows the stretching ratio in x.

7.3.3 Defect Induced Rumpling.

As has been explored thoroughly in **Chapter 3** and **Chapter 4**, biopolymer networks have complex ring structures[70]. Given that the hexagonal network is the lowest energy arrangement for the coarse-grained polymers, the addition of any defect causes a site with a locally increased potential energy. The network can potentially relieve the strains around a given site by distorting out of the plane. For example, an overcrowded site may relieve steric strain by moving adjacent polymers out of the same plane, or relieve angular strain by tilting into the newly accessible third dimension.

The effect of defects on the 2D/3D nature of collagen-like biopolymer networks can be tested by introducing defects via a bond switching algorithm. Disordered networks were created by performing b bond switching moves on an ideal hexagonal network, then repeating the simulation protocol described in **Section 7.3.1**. Each set of b bond switching moves was repeated 10 times with a different random seed. Then, the networks were converted into polymers (see **Section 5.2.3** and ref [130]) and simulated with the 2D constraint removed at 100 K for 100 µs and all other parameters as before.

The effects of defect introduction can be explored by looking at a network with only one defect in it. **Figure 7.5** shows roughness around a single Stone-Wales defect in an otherwise hexagonal network after 100 µs of simulation. This figure demonstrates the distortion in z around the defect, with the nearby molecules being tilted significantly out of the plane as shown by the large gradient around the defect. The significant distortion of nearby biomolecules in a caldera shape around the defect also leads to large error bars in the roughness metric. The size of the caldera near the defect is again approximately one polymer, matching the observations in the patterns of ξ.

The width of networks after a number of bond switching moves is shown in **Figure 7.6**. A

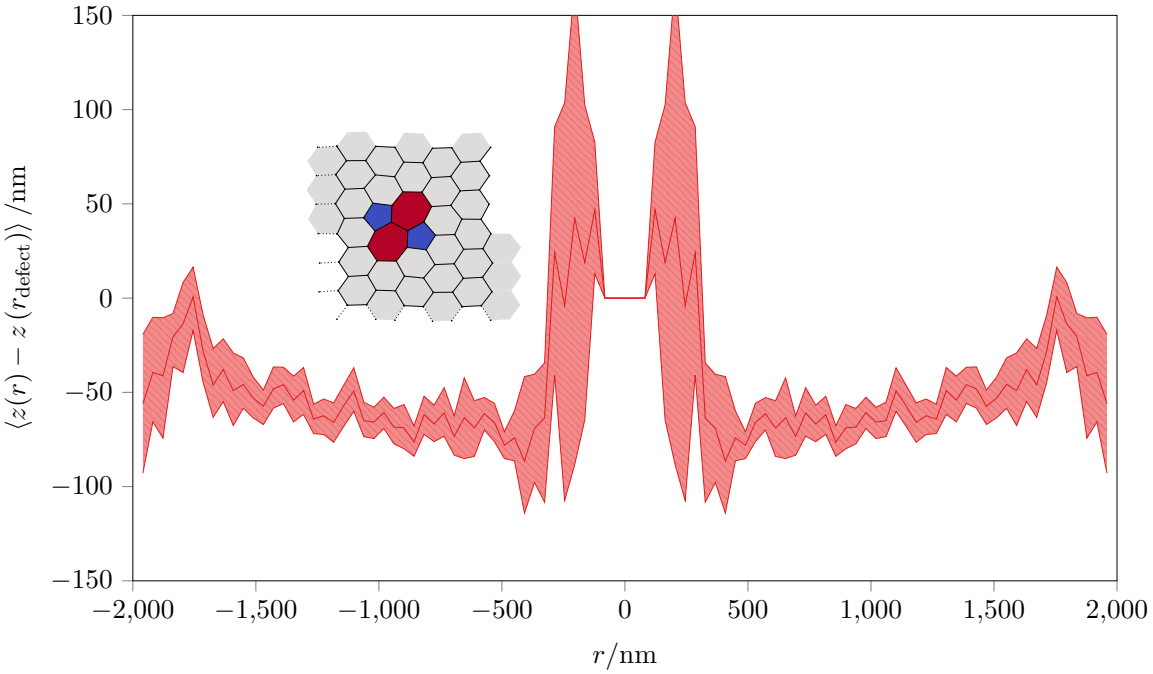


Figure 7.5: The height-defect height correlation function measuring the difference between bead z positions and mean defect z position as a function of distance from the centre of the defect. The shaded area represents a $95.0\,\%$ bootstrap confidence interval, and the network is embedded in the image.

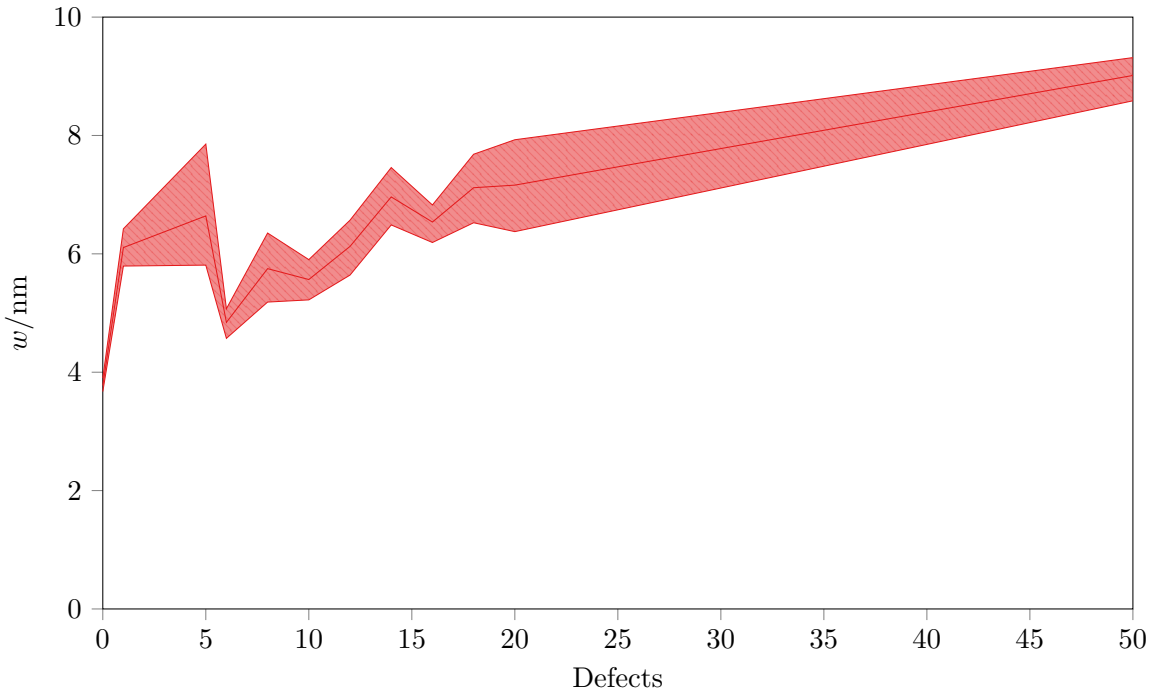


Figure 7.6: The self affine parameter w against the number of attempted bond switching moves. The shaded area represents a $95.0\,\%$ bootstrap confidence interval.

significant increase in the width w can be seen as more defects are inserted into the otherwise ideal hexagonal network. This figure shows two important trends. Firstly, a slight increase in the width of the network as the initial (0 to 5) defects are introduced. The increase in average width for few defects is because each defect gets its own caldera-shaped distortion which rise above the plane of the network. Secondly, there is a sudden drop in width after 5 to 15 bond switching moves have been attempted. The decrease in average width in this region is due to an increased likelihood of correlated defects, which leads to greater flexibility around the defect-containing region. The greater flexibility can absorb both the distortion due to the defects and allow the network to flatten out overall.

This effect becomes less significant as many more defects are introduced (20 to 50) because the network loses its hexagonal structure entirely. The defect-ridden networks can often break down as the energetic penalty of a defect can be relieved by breaking at a node; the resultant free polymer can move outside the plane of the network, appearing as 'hair', and leading to increase in roughness metrics.

7.4 Chapter Conclusions.

This chapter has detailed investigations into the rumpling of 2D networks allowed to move into 3D space. A fractal approach was used to describe the degree to which the networks are no longer truly 2D, using a self-affine scaling law to describe the trends in height-height correlation as a function of distance.

The fractal approach fits well, despite the extremely coarse-grained nature of the model. A fractal dimension in the range 2.60 to 2.70 was found to describe the hexagonal networks well across a range of parameters. The widths of the network depended on the balance between enthalpy and entropy, with more thermal energy leading to a rougher surface. The fractal length scale ξ was found to be approximately one coarse-grained polymer length, reflecting their semiflexible nature.

By stretching and compressing the networks sinusoidally, deformations into 3D space were explored. These suggested that if the volume were held constant then the polymer networks would preferentially distort in the plane, but if the volume changed then they would distort out of the plane. However, the networks did not flatten out fully even when stretched significantly — this was because they still had some thermal energy which led to distortion.

Distortion was more obvious when defects were added to the network. A Stone-Wales bond switching defect created a region of local stress in the network, which could be relieved by forming a caldera around it. The presence of a low concentration of correlated defects surprisingly allowed the networks to flatten out further than a single defect does. This was likely due to the influence of adding defects also adding flexibility.

Chapter 8: Persistent Homology.



Persistent homology (PH) can describe the damage to networks over time by characterising changes in their topological structure. Normally in persistent homology, topology is described by ring structure evolving as a function of a radial cut-off r, but here visualisation methods show ring structure over time t during self-assembly and rupturing. The underlying theory behind PH approaches is discussed, and the utility of persistence diagrams and birth-death diagrams are demonstrated. Metrics of ring lifetime and rebirths are used to quantify repair phenomena over time, and birth-death diagrams showing multiple combined simulations will be shown. This chapter is adapted from M. H. J. Bailey & M. Wilson, *Submitted to Bulletin of Mathematical Biology*, 2022, DOI: `10.26434/chemrxiv-2022-k4thc`[158].



8.1 Introduction.

Chapter 6 discussed how damage can occur to complicated networks when they are exposed to physical strains. Similarly, **Chapter 3** and **Chapter 5** discussed likely defects that may be introduced, and the consequences of those defects for mechanical properties.

In **Chapter 6**, polymer metrics were used to quantify the damage suffered over time, which were either too detailed or averaged across too many rings to be useful. However, one theme running through this thesis is the simplifying power of graph based networks. Graph based methods have seen usage in analysing inorganic two dimensional (2D) networks such as

amorphous graphene or glasses[113,159]. By treating molecules as edges, and their interaction sites as nodes, a graph of the network can be constructed. Graph theory approaches are more useful when we zoom out from polymer-level metrics to analyse larger structures. The edges in a planar 2D network form a set of rings, with medium-range ordering in the arrangement of those rings[160]. Much of the literature characterises networks by their constituent rings, referred to as 'ring statistics'[91,159].

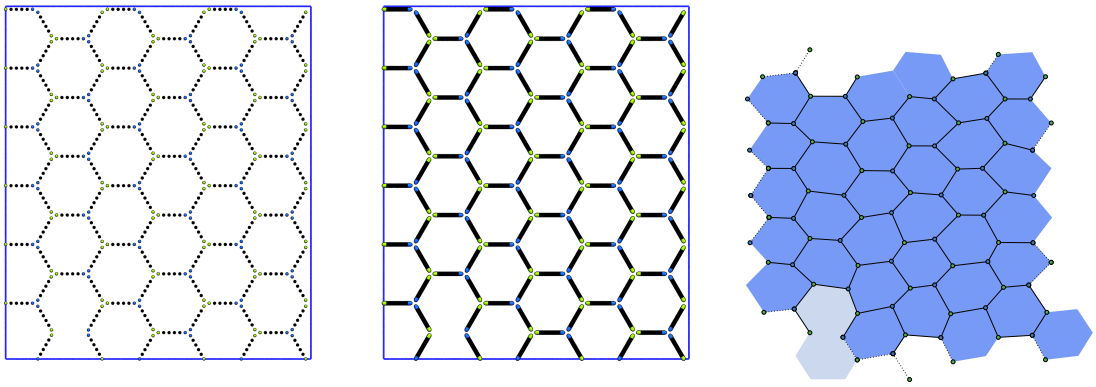
There is a hierarchy of features in a graph representation with a geometric meaning:

- Nodes by themselves are zero-dimensional features, shown in **Figure 8.1a**.
- Edges connecting nodes are one-dimensional features, shown in **Figure 8.1b**.
- Rings formed of edges are two-dimensional features, shown in **Figure 8.1c**.

In non-planar networks there are also three-dimensional features, seen as pores in three dimensional (3D) networks such as zeolites[161], but these are outside the scope of this chapter. In general these are examples of mathematical objects known as k-simplices, where k is the number of dimensions.

This hierarchy of detail is similar to the framework used in persistent homology (PH) methods. PH has been used to characterise amorphous inorganic networks, proteins and zeolites[79,161,162].

PH works by analysing 'homology groups', constructed by embedding a data set in a vector space with a well-defined distance metric. One way to conceptualise PH is to imagine that spheres are placed on every point in the dataset. The radius of the spheres, described as the 'filtration length' r, is then gradually increased. An edge is drawn between two points when two spheres overlap. Where many edges are connected, they are linked into simplices: edges in 1D, triangles in 2D, tetrahedra in 3D and so on. This is mathematically described as forming a Čech complex (a more specific case of a Vietoris-Rips complex). The filtration



(a) Atom positions, a zero dimensional feature with little spatial ordering information

(b) Molecules formed of bonds between atoms, a one dimensional feature with short-range spatial ordering information

(c) Rings formed by the gaps between molecules, a two dimensional feature with medium-range spatial ordering information

Figure 8.1: Rings can be seen as medium-range order in a network, and built up as the highest dimensional feature. Here, the same network is represented in three different ways: atom positions, molecules, and rings.

values at which simplices are formed, referred to as being 'born', or cease to exist, referred to as 'dying', are tracked and used for further analysis[163].

A homology group (generated from the full complex of simplices) is unwieldy and difficult to interpret[164]. To simply the problem, a number of useful metrics and diagrams can be created from the sequence of births and deaths, including Betti numbers, persistence diagrams and birth-death diagrams.

Betti numbers are a concept from algebraic topology, where β_k represents the number of k-dimensional rings that can be removed from the network without disconnecting it. A torus has $\beta_0 = 1$, as it is a single surface. It also has $\beta_1 = 2$, as could be cut either vertically or horizontally and remain as a single object. Finally, it has $\beta_2 = 1$, as there is a single 3D void inside the torus. Betti numbers map neatly onto the graphs used in this thesis, as the zeroth Betti number β_0 is the number of edges, the first Betti number β_1 is the number of rings, and so on.

Persistence diagrams chart the time between birth and death of a given set of features in

the system as a function of filtration value[165,166]. They are usefully robust against random noise, and show a characteristic barcode pattern. Birth-death diagrams are scatter diagrams with each point showing when a topological object is born and dies, and they often show clusters of features appearing or disappearing.

Persistence diagrams are sometimes criticised in that they reflect abstract properties of a network that are difficult to visualise, and are extremely hard to interpret quantitatively[79,167]. On the other hand, persistence diagrams can enable new methods of qualitative analysis[33,168]. This is especially true when the features in the persistence diagram map onto physical features, such as diffraction peaks in amorphous materials, which are strongly linked to the network topology[169].

As was seen in **Chapter 6**, the dynamic behaviours of biological networks are some of the most interesting. Simply analysing the ring statistics at a single time point loses critical information, and using polymer statistics over time provides no insight into medium range order. This is where the PH approach can be utilised, albeit with some modifications. While PH varies the length scale in the filtration function, biological networks already have meaningful length scales defined. The most useful length scale is approximately the length of one polymer, and it is most interesting to see how information varies across time instead. There is precedent for taking time into account in filtration functions: Dong *et al.* have used a filtration function of both length and time $f(x,t)$ to calculate Betti numbers for dynamic properties of zeolites[170]. Using time as a filtration function instead of spatial cut-off also makes interpreting the diagrams easier, as temporal evolution is easier to visualise.

In this chapter, time persistence diagrams are used to characterise damage over time to biological networks. The new diagrams are shown to be robust against noise and sampling parameters, and shown to make complex damage metrics easy to interpret.

8.2 Computational Methods and Metrics.

8.2.1 Ring Existence Function.

The PH method requires a strict definition of a 'topological feature'. First, a number of graph theory terms that have been used informally (and somewhat loosely) elsewhere in this thesis must be formalised. An *embedding* is a map that transforms a graph G into a set of positions in vector space V, while preserving its properties. The vector space chosen here is a 2D periodic surface, equally representable as the surface of a torus. There is room for choice in vector spaces in PH approaches generally, especially with the definition of a distance metric. Because collagen IV networks exist in Cartesian space, work in this chapter makes use of a standard Euclidean distance metric.

A *ring* is then defined as being a finite set of nodes linked by edges that form a face on the toroidal embedding of the graph (sometimes described as being a 'polygon'). If the graph is traversed starting from one node in the ring, following edges and picking the most anticlockwise direction at each subsequent node, each other node in the ring will be visited.

This definition becomes difficult when rings are broken or re-form as the graph (and associated embedding) evolves over time. Nodes in the graph are not numbered uniquely, and the number of nodes may change over time. Instead, the invariant is the polymers that define each edge. A ring is identified by the unique set of polymers associated with the constituent edges. For example, one square may be constructed of polymers $\mathcal{R}_1 = \{p_1, p_2, p_3, p_4\}$ and a second square may be constructed of polymers $\mathcal{R}_2 = \{p_4, p_5, p_6, p_7\}$. These two squares are distinct, and may exist separately at different times. It is then possible to define an existence function $f(t; \mathcal{R})$ for each ring $\mathcal{R}$ at each time t over a trajectory. The existence function has a key different from the filtration functions used elsewhere in the PH literature: it is not

exclusively increasing, as a given ring can be born, die, and be resurrected. As an example, the square $\mathcal{R}_1 = \{p_1, p_2, p_3, p_4\}$ could exist at t, die at t', and be reborn at t''.

8.2.2 Polymer Model and Simulation Protocol.

The coarse-grained polymers and Molecular Dynamics (MD) method established first in **Chapter 4** and extended in **Chapter 6** are re-used in this chapter. However, the harmonic potential for internal bonds from **Chapter 4** was combined with the Morse potential from **Chapter 6** to study the entire lifecycle of a network from assembly to rupture.

The Morse potential is unsuitable for self-assembly simulations, because a high temperature is used to randomise the polymers after they are initially placed. High temperatures would break all the Morse bonds in the system during the randomisation step. To resolve this problem, a shifted harmonic potential is applied during the randomisation step. That potential is implemented in LAMMPS as

$$U = \frac{U_{\min}}{(r_{\text{eqm}} - r_c)^2} \left[(r - r_{\text{eqm}})^2 - (r_c - r_{\text{eqm}})^2 \right]. \tag{8.1}$$

The HARMONIC/SHIFT potential is a harmonic potential with an equilibrium length of r_{eqm} and a force constant equivalent to $k = \frac{U_{\min}}{2(r_0 - r_c)^2}$. It is implemented in this unusual form for speed of evaluation, because using only squared distances avoids expensive square root operations. This potential model has energy $U = 0$ at $r = r_c$, chosen to be 45.0 nm. At the equilibrium distance of $r_{\text{eqm}} = 50.0\,\text{nm}$ the bond has energy $U_{\min} = -24.0 \times 10^{-21}\,\text{J}$. Values of the parameters U, r_{eqm} and r_c are chosen to match the first terms of the Taylor expansion of the Morse potential in **Chapter 6**.

For self-assembly simulations, the network underwent the following steps with a harmonic potential:

- Proto-network formation,

- Simulation cell size fixing using an NpT ensemble,
- Heating and randomisation,
- Cooling into final network.

These steps were discussed in more detail in **Section 4.2.5**. The bond and angular potentials established in **Chapter 6** were reused, with Morse potential parameters $D_e = 24.0 \times 10^{-21}\,\mathrm{J}$, $r_{\mathrm{eqm}} = 50.0\,\mathrm{nm}$ and $\alpha = 0.150\,\mathrm{nm}^{-2}$. Other parameters and conditions for each stage will be discussed in more detail in the relevant results section. During the course of the simulation, regular snapshots were taken and the graph structure was extracted.

8.2.3 Persistence Diagrams.

Persistence diagrams are the key tool introduced in this chapter. A persistence diagram shows vertical bars charting for which time values a given ring exists. First, rings are assigned an increasing numerical ID. These can be assigned arbitrarily, but for interpretable diagrams rings are assigned IDs according to the order of their first births. For instance, a ring born at $t = 0$ will have a lower ID than a ring born at $t = 1$.

Samples are taken from the trajectory at regular timesteps of Δt, which leads to batching of births and deaths. If multiple rings are born in the same sampling frame, then they are assigned IDs based on their total lifespan: a shorter lifespan means a smaller ID. For instance, a ring born at $t = 0$ and dying at $t = 1$ will have a lower ID than a ring born at $t = 0$ and dying at $t = 2$. If ties need to be broken further, the number of edges per ring is used, where rings with fewer edges get a lower ID.

Assigning ring IDs can be troublesome, as they are not necessarily consistent between analysis runs with different Δt. The ring IDs used throughout this work have been made consistent throughout and hence are independent of timestep.

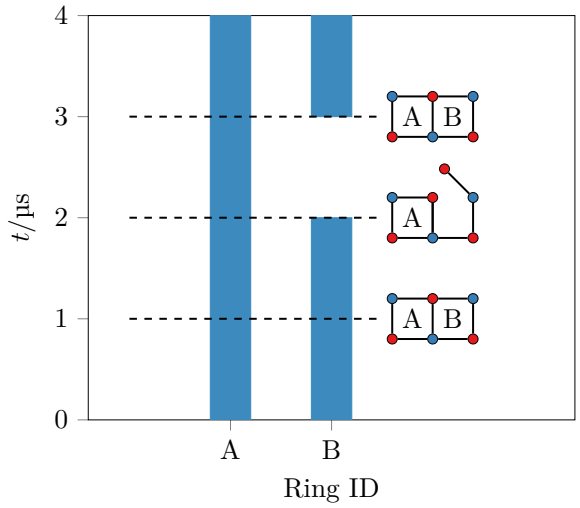


Figure 8.2: A time persistence diagram for a single ring system. Ring A exists continuously from times $t = 0$ to 4, but ring B exists only between $t = 0$ to 2 and $t = 3$ to 4.

A time persistence diagram showing the simple system of two squares mentioned in **Section 8.2.1** is shown in **Figure 8.2**. This diagram is striped in both horizontal and vertical directions. The horizontal striping is useful to identify how many rings exist at a given time, and the vertical striping is useful to analyse repair phenomena.

8.2.4 Birth-Death Diagrams.

The other visualisation used in this section is the birth-death diagram. For each ring, a point is plotted at $(t_{\text{birth}}, t_{\text{death}})$. This leads to an upper left triangular diagram, where a point on the diagonal line indicates an ephemeral ring that died immediately after being born.

As multiple births and deaths are possible for each ring, there are two conventions that could be chosen:

- Plot every birth-death pair for the ring, or
- Plot only the first birth and final death.

In a simple ring system with frequent repair events, the first convention is useful to track the birth-death trajectory of a given ring. However, work in this chapter will use the second

convention, as it leads to simpler diagrams and easier interpretation for the majority of rings. Furthermore, the second convention is robust to different values of the timestep Δt, and is not interrupted by repair events. Points plotted with the second convention end up spaced further from the diagonal, which may make critical phenomena more obvious.

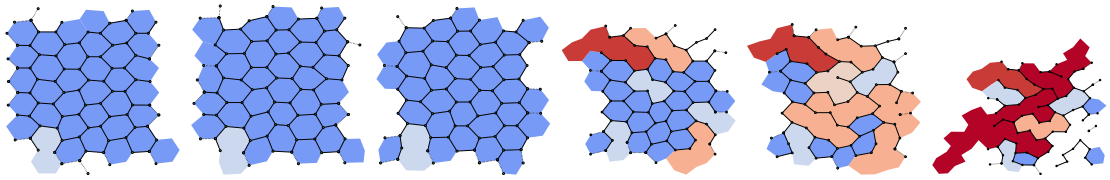
8.2.5 Lifetime Metrics.

Numerical metrics can be calculated from the existence patterns of individual rings, adding a quantitative element to the relatively qualitative approaches described thus far. The first metric is to measure the number of repair events a given ring undergoes, represented by how many vertical stripes there are. The number of repair events rings is useful to characterise structural changes over the course of a simulation.

The second metric is calculated from the lifespan L of a ring $\mathcal{R}$, and is determined by the vertical integral of its bar (or the equivalent summation):

$$L(\mathcal{R}) = \int_0^{t_{\max}} f(t;\mathcal{R})\,\mathrm{d}t \approx \sum_k f(k\,\Delta t;\mathcal{R}). \tag{8.2}$$

This lifespan can be compared to the total time between first birth and last death to calculate a 'fraction of time alive'. The fraction of time alive is a useful metric to measure how important repair events are to rings of different sizes or under different conditions. For example, a ring that is damaged once and repaired immediately afterwards will have a high fraction of time alive. Conversely, a ring that is damaged once and repaired only at the end of the trajectory will have a low fraction of time alive.



(a) $t = 0\,\mu s$ **(b)** $t = 10.0\,\mu s$ **(c)** $t = 20.0\,\mu s$ **(d)** $t = 30.0\,\mu s$ **(e)** $t = 40.0\,\mu s$ **(f)** $t = 50.0\,\mu s$

Figure 8.3: Snapshots of ring damage to a hexagonal network. Each snapshot was taken when the network was unstretched; between images the network is stretched in x and compressed in y, then *vice versa*.

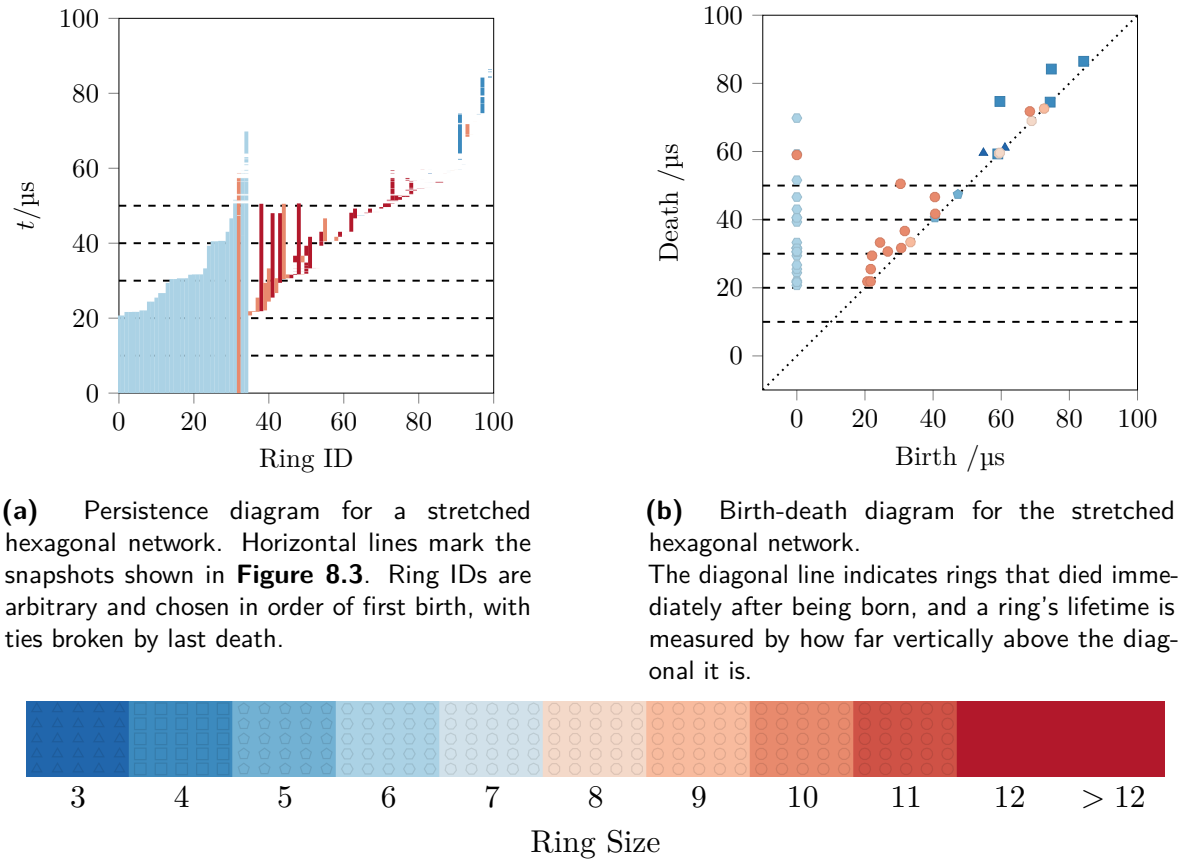
8.3 Results and Discussion.

8.3.1 Ordered Networks.

As persistence diagrams can be initially unintuitive, it is first helpful to establish a visual reference by analysing a simple system qualitatively. The simple system used throughout this thesis is an ordered hexagonal network constructed of coarse-grained polymers. **Chapter 5** showed that idealised hexagonal networks tend to rupture all at once because they are highly symmetrical. To avoid that issue, one bond was removed from an initial network of 36 hexagons; this left 34 hexagons and one decagon.

The polymer network was then stretched sinusoidally over 100 µs. Each individual stretching cycle had a time period of 5.00 µs, and each cycle linearly increased in amplitude from $A = 1.10$ to $A = 1.50$. This is the same method as was used in **Section 6.3.4** and broadly repeats simulations discussed there.

Snapshots of a single network undergoing that stretching procedure are shown in **Figure 8.3**. For the first 20.0 µs, in **Figure 8.3a** to **Figure 8.3c**, the network keeps its original topology. There are only small perturbations to the node positions in this initial section where the maximum amount of stretching is small. From 30.0 µs onwards, in **Figure 8.3d**, the network begins to suffer damage. A few rings with many sides appear where polymers have



(a) Persistence diagram for a stretched hexagonal network. Horizontal lines mark the snapshots shown in **Figure 8.3**. Ring IDs are arbitrary and chosen in order of first birth, with ties broken by last death.

(b) Birth-death diagram for the stretched hexagonal network.
The diagonal line indicates rings that died immediately after being born, and a ring's lifetime is measured by how far vertically above the diagonal it is.

Figure 8.4: Persistence and birth-death diagrams for a stretched hexagonal network.

come loose. When the stretching amplitude gets larger from 40.0 µs to 50.0 µs, the damage becomes more significant. The network loses memory of its original hexagonal nature, and extremely large rings appear. There are also a number of loose chains which do not contribute to the ring statistics.

The qualitative analysis from snapshots can be useful in understanding general trends, but is insufficient to gain a numerical understanding. Analysing individual snapshots may also over-weight ephemeral rings that only exist fleetingly — for example, **Figure 8.3e** features five 14-sided rings. From analysing snapshots, it is hard to tell which of those decagons lives the longest or is most important for overall structure.

A systematic approach can be constructed using a persistence diagram and a birth-death diagram. The persistence diagram for the rupturing process is shown in **Figure 8.4a**. The

initial hexagons are shown on the left hand side with Ring IDs 0 to 35 being born at $t = 0\,\mu s$, and one decagon coloured differently. These hexagons start dying at $20.0\,\mu s$, as can be seen in the change from **Figure 8.3c** to **Figure 8.3d**. The deaths of the hexagons are marked by the red cluster of hexagons in **Figure 8.4b**, which were all born at $0\,\mu s$. When the hexagons die, they are initially replaced by decagons; rings with many edges and IDs 36 to 55 are born immediately when their predecessors die. This is shown by a vertical bar starting where another bar ends in the persistence diagram. Some of these new rings are extremely short lived, and break down into larger rings immediately. This is shown by extremely short vertical bars, and by dots on the diagonal of **Figure 8.4b**. Seeing the short lifetime of these large rings is critical, as it undercuts the temptation to over-analyse the snapshots. Two of the 14-sided rings in **Figure 8.3e** have extremely short lifetimes, but the third does not.

Not all rings are short lived. For example, the 14-sided ring formed of three broken hexagons in the bottom right of **Figure 8.3d** and **Figure 8.3e**, lives for a long time and can be seen as the long vertical bar with ID 43. As the simulation progresses, almost all the original rings are broken by $t = 50.0\,ns$ with only one or two vertical bars passing through the dashed line at that point.

A few hexagons die in the range $50.0\,\mu s$ to $80.0\,\mu s$ shown by isolated hexagons on the left hand size of the birth death diagram. Afer $80.0\,\mu s$, only damage products remain. The few rings that survive longest after the network is almost entirely broken are rectangles, which are formed easily from the network detritus and small enough to resist further damage.

After $50.0\,ns$ we can observe some repair phenomena: this is especially clear in Ring 35. That ring briefly dies shortly after $50.0\,ns$ as it ruptures under the stretching. However, it rapidly reforms and its persistence bar continues, albeit interrupted. This repair phenomenon repeats at least 7 times over the rest of the simulation, often as part of a concerted phenomenon

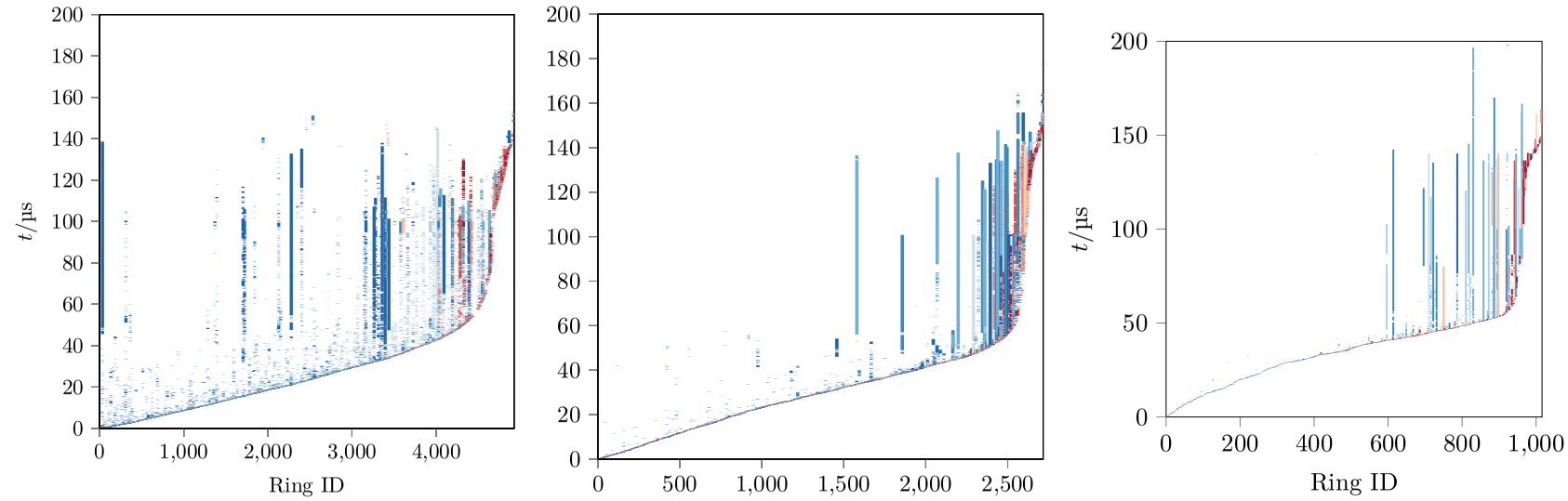
in which a handful of rings break and are repaired. Repair phenomena are not possible in traditional persistence diagrams, where new simplices can only ever be created, but they are possible in a temporal diagram.

8.3.2 Disordered Networks.

With an example established, the analysis process can be applied to a more complex simulation. By combining the self-assembly studies of **Chapter 4** with the rupturing studies of **Chapter 6**, patterns of assembly and destruction can be examined. Networks were constructed by first placing 400 polymers on an ordered square grid with random rotations. The polymers were allowed to form a proto-network by equilibrating at 100 K for 4.00 µs. The proto-network was used to set the simulation cell size by then continuing equilibration over 2.00 µs under a barostat in an NpT ensemble with a small 2D pressure p, temperature 100 K to fix the area per molecule. 2D pressures were varied to create a range of networks with different areas per molecule, and thus significantly different topologies (see **Section 4.3.3** for more discussion of this). Once the simulation cell area was equilibrated, it was fixed and the simulation moved back to an NVT ensemble with a Langevin thermostat. The network was heated up to 1000 K for 4.00 µs to ensure melting. The polymer liquid was then cooled back down to 100 K over 100 µs. Once the networks were cooled and a network was formed, the stretching process from **Section 8.3.1** was repeated.

Persistence diagrams for networks with three different densities are shown in **Figure 8.5a**, **Figure 8.5b** and **Figure 8.5c**. Accompanying birth-death diagrams are shown in **Figure 8.5d**, **Figure 8.5e** and **Figure 8.5f**. These diagrams can be used to track key stages in network formation and rupturing, and highlight qualitative differences between the simulations.

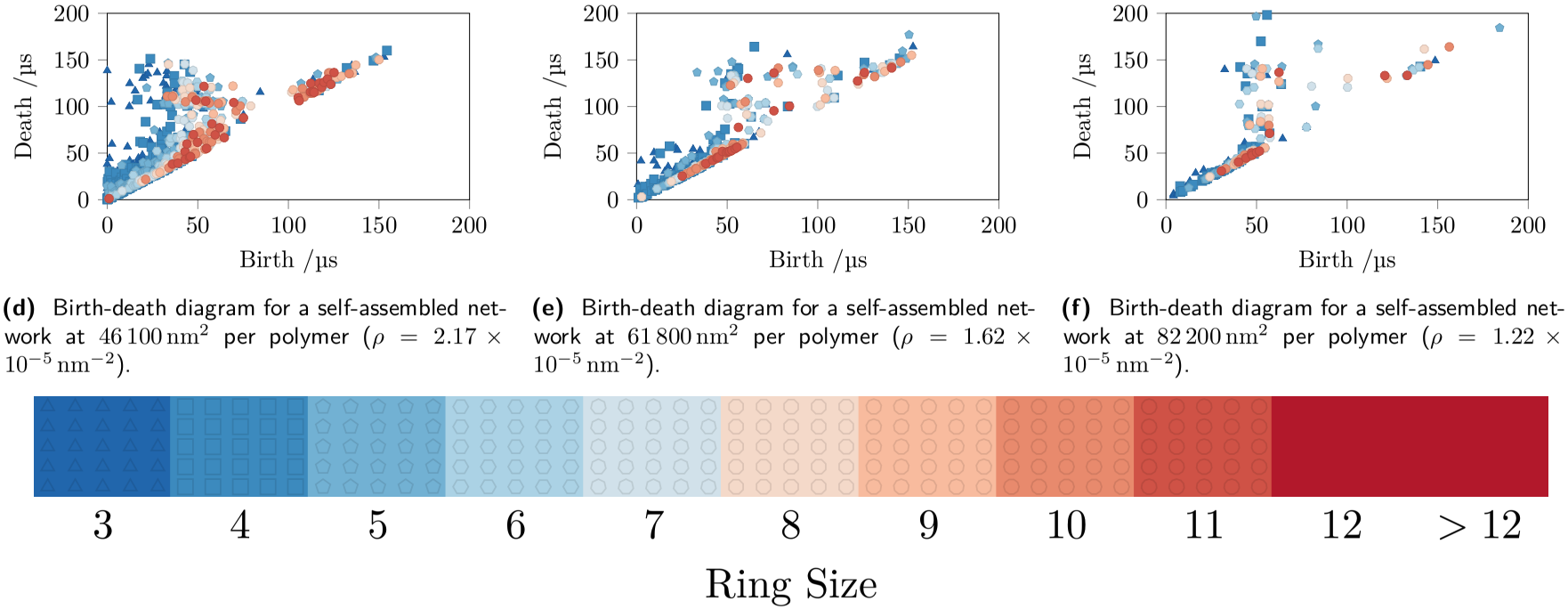
In the low density network seen in **Figure 8.5c** and **Figure 8.5f** only small rings with



(a) Persistence diagram for a self-assembled network at $46\,100\,\mathrm{nm}^2$ per polymer ($\rho = 2.17 \times 10^{-5}\,\mathrm{nm}^{-2}$).

(b) Persistence diagram for a self-assembled network at $61\,800\,\mathrm{nm}^2$ per polymer ($\rho = 1.62 \times 10^{-5}\,\mathrm{nm}^{-2}$).

(c) Persistence diagram for a self-assembled network at $82\,200\,\mathrm{nm}^2$ per polymer ($\rho = 1.22 \times 10^{-5}\,\mathrm{nm}^{-2}$).



(d) Birth-death diagram for a self-assembled network at $46\,100\,\mathrm{nm}^2$ per polymer ($\rho = 2.17 \times 10^{-5}\,\mathrm{nm}^{-2}$).

(e) Birth-death diagram for a self-assembled network at $61\,800\,\mathrm{nm}^2$ per polymer ($\rho = 1.62 \times 10^{-5}\,\mathrm{nm}^{-2}$).

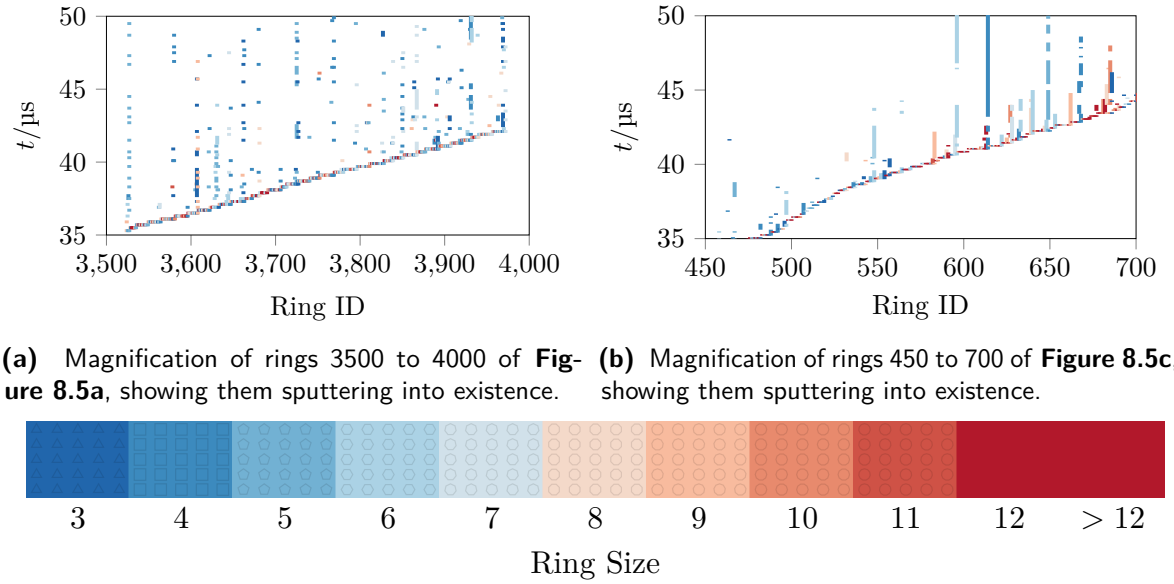
(f) Birth-death diagram for a self-assembled network at $82\,200\,\mathrm{nm}^2$ per polymer ($\rho = 1.22 \times 10^{-5}\,\mathrm{nm}^{-2}$).

Figure 8.5: Birth-death diagrams and persistence barcodes (lifetimes) for networks with simulation cells equilibrated at different areas per molecule. This figure is adapted from M. H. J. Bailey & M. Wilson, *Submitted to Bulletin of Mathematical Biology*, 2022, DOI: `10.26434/chemrxiv-2022-k4thc`.

short lifespans are formed in the initial period of 40.0 µs. The critical formation point is then between 40.0 µs and 60.0 µs where many rings form, and few rings are formed in the final 60.0 µs to 100 µs. This can be seen in a solid vertical band of rings born between 40.0 µs and 60.0 µs in **Figure 8.5f**, and a similar structure on the other axis in the persistence diagram. This is because most molecules are now coordinated, and little thermal energy is available to rearrange. The gradient of the persistence diagram indicates how many rings are formed per unit time; a shallow gradient is many rings per unit time and a steep gradient is few rings per unit time.

For the network formed at intermediate and high densities, the critical formation point is no longer so sharply defined. This can be seen in the birth-death diagrams of **Figure 8.5e** and **Figure 8.5f** as the vertical lines around formation time less well defined, with more rings formed before and after this point. The difference in assembly behaviour for the higher density networks is because semi-viable networks form more easily when molecules are in a confined space. When the forming network is loosely packed, many long chains form initially and slowly coalesce into rings. However, when network is densely packed, small rings form initially with short lifespans.

Magnified sections of the persistence diagrams for high and low area networks are shown in **Figure 8.6**. These highlight the key difference between the temporal persistence diagrams used here and the persistence diagrams used in other PH approaches. The diagrams in **Figure 8.6** are non-continuous, as rings may repeatedly be born and die. This limits us from using some more powerful PH analysis techniques, such as using the diagrams to measure similarity between trajectories[171] or to identify binding sites[172]. Thankfully, the vertical striping in the persistence bars allows for useful chemical analysis, especially of repair phenomena.



(a) Magnification of rings 3500 to 4000 of **Figure 8.5a**, showing them sputtering into existence.

(b) Magnification of rings 450 to 700 of **Figure 8.5c**, showing them sputtering into existence.

Figure 8.6: Magnified sections of persistence diagrams showing the sputtering and repair phenomena. The rings in the higher density simulation (left panel) are only alive for short periods, leading to a pale vertical pattern. The rings in the lower density simulation (right panel) are only broken a few times after formation and are alive for long stretches.

Area per polymer /nm^2	ρ/nm^{-2}	Unique Hexagons	Rebirths
46 100	2.17×10^{-5}	62.8	4.81
61 800	1.62×10^{-5}	56.7	3.66
82 200	1.22×10^{-5}	26.4	3.39

Table 8.1: Number of unique hexagons observed during simulations with a given area per polymer, and the average number of times each of those hexagons is reborn. This table is adapted from M. H. J. Bailey & M. Wilson, *Submitted to Bulletin of Mathematical Biology*, 2022, DOI: `10.26434/chemrxiv-2022-k4thc`[158].

Rings in **Figure 8.6a** (the high density network) continually sputter after initial formation. This is because in the high density network there are many polymers and attractive head groups per unit area. The density of polymers leads to frequent microscopic breaking and reforming events.

Rings in **Figure 8.6b** (the low density network) show different behaviour immediately after being formed. They often sputter slightly, being born and dying a few times as polymers with lots of energy being able to coordinate with one another. However, after some initial sputtering, rings show long stable lifetimes as they are well-separated from one another.

The number of reforming events is shown quantitatively in **Table 8.1**. The table quantifies the number of unique hexagons observed over a simulation, where a unique hexagon is defined as a ring in the graph made up of a unique set of six polymers. Ephemeral rings, defined as rings that only exist in one sample, are discarded. Ephemeral rings show up as single pixel tall bars in a persistence diagram, or points exactly on the diagonal of a birth-death diagram. The table also quantifies how often a given ring is reborn on average over its lifetime. One rebirth is defined as a ring dying at time t but being observed again at a later time t'.

Data in **Table 8.1** show that in higher density networks more unique hexagons are formed on average, and each of those individual hexagon undergoes more rebirths. As the density decreases, there are fewer hexagons observed and they are reborn less. Less frequent rebirth in lower density networks is because the overall network structure can absorb deformation more easily, so rings are not repeatedly destroyed and reformed. The presence of fewer unique hexagons in lower density networks is because there are fewer options for a polymer to coordinate to when it breaks, so it is likely to reform its original ring.

The final section of interest is the behaviour of self-assembled networks during rupturing, shown in the top right quartiles of the panels **Figure 8.5(a-c)** and top halves of the panels **Figure 8.5(d-f)**. All times in those diagrams after 100 µs represent rupturing times. There is a break visible in the birth death diagrams [**Figure 8.5(d-f)**] at 100 µs where the change in potential model and associated energy minimisation causes some rings to break or form. After that, a number of rings break in the range 100 µs to 150 µs, often forming short-lived rings with many edges. The point at which the networks rupture fully is seen as many deaths at 150 µs, which is better defined for the high density simulations. Few rings are formed after this point, but rings are more likely to form in the loose, low density networks after this time.

8.3.3 Effect of Sampling Timestep.

The nature of MD is that the continuous evolution of a system is split into many discrete time slices. Only a fraction of the time slices are then saved for future analysis, often skipping many steps to save on disk space and computing time. By controlling the timestep Δt between samples, it is possible to highlight and analyse phenomena occurring over different timescales.

Repeating the persistence analysis at different sampling rates can either smooth out noisy high-frequency patterns in the data by taking infrequent samples, or track short lived ephemeral phenomena by taking frequent samples. Alternatively, low-frequency sampling rates can make the analysis much faster as fewer steps need to be analysed and there are fewer unique rings to track.

Figure 8.7 shows the same persistence and birth-death diagrams as were seen in **Figure 8.5c** and **Figure 8.5f** but sampled either every 0.200 µs, 2.00 µs or 10.0 µs. All panels still show formation occurring at 40.0 µs and most rupture occurring at 150 µs. The lines showing formation and rupturing at 40.0 µs and 150 µs become clearer at shorter timesteps, but are present in all analyses. This highlights their importance as critical phenomena during the trajectory, and shows that the persistence analysis method is not sensitive to frequency of sampling.

One major difference for the analysis using a slower sampling frequency is that fewer unique rings are observed. If a sample is taken regularly at intervals spaced by Δt, then rings with lifespans shorter than Δt can be missed. This is what leads to gaps between bars in **Figure 8.7f**, and would lead to a different x axis if the ring IDs had not been adjusted for visualisation purposes.

Having fewer ephemeral rings in the analysis can often be clearer, as it draws attention only to the rings that live for at least a period of Δt. The key feature of formation beginning

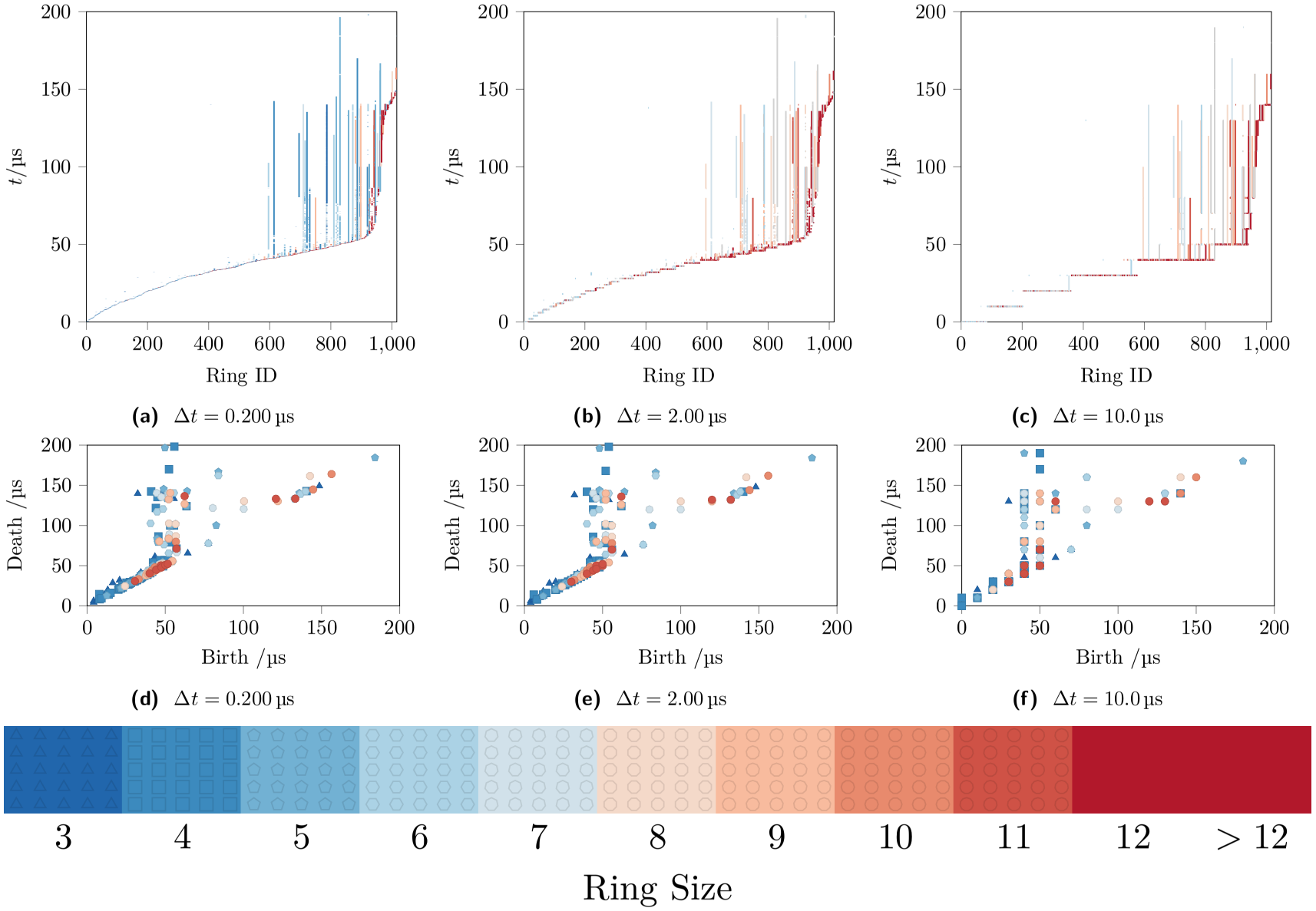
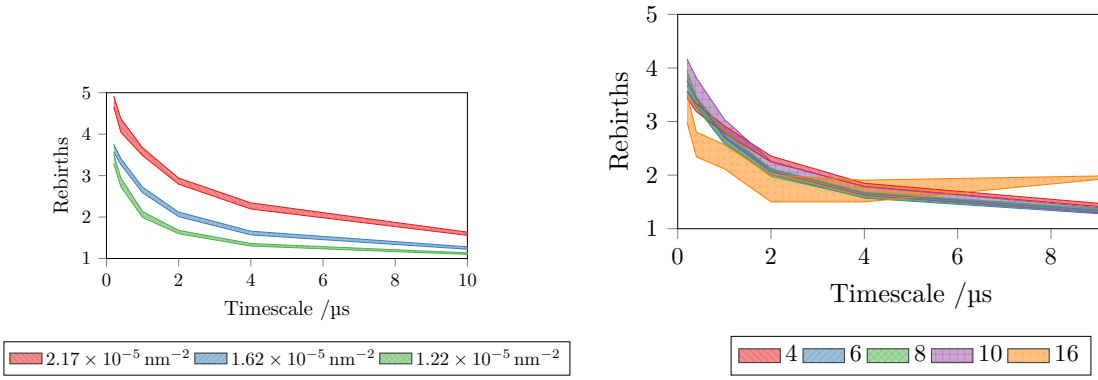


Figure 8.7: Persistence and birth-death diagrams for systems sampled at different frequencies. These are all from simulations with $82\,200\,\mathrm{nm}^2$ ($\rho = 1.22 \times 10^{-5}\,\mathrm{nm}^{-2}$) per polymer. Ring IDs have been constructed to be the same across different simulations for comparisons.



(a) Number of rebirths per unique ring as a function of timestep at three different areas per molecule.

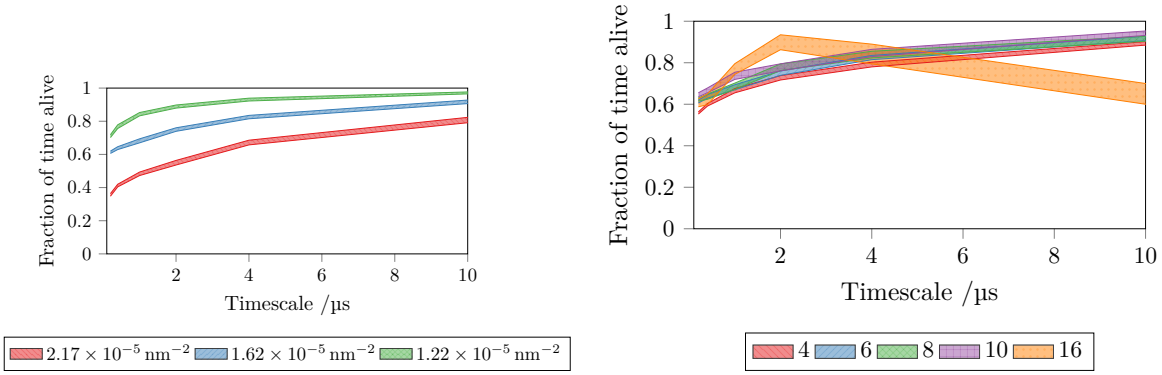
(b) Number of rebirths per unique ring with n edges at a medium area per molecule.

Figure 8.8: Number of rebirths per unique ring as a function of timestep. This excludes ephemeral rings that are alive for only one sampling step, and the shaded regions are a 50.0 % bootstrap confidence interval.

properly at 40.0 µs is obscured in **Figure 8.7d** with many fleeting rings being formed, but clearer in **Figure 8.7f**. While counting fewer rings can make analysis clearly, it can also miss short-lived phenomena. For example, repair phenomena are almost impossible to see in **Figure 8.7f** as most repair phenomena have happened many times over in 10.0 µs.

Missing out repair phenomena can be described quantitatively by measuring the average number of repair events per ring as a function of Δt, similar to **Table 8.1**. These statistics can be seen in the panels of **Figure 8.8**. The left panel, **Figure 8.8a**, shows that the analysis correctly picks out more frequent repair events for more dense networks at all timescales. At larger timesteps Δt, the number of rebirths per ring drops towards 1. This graphically shows the missing repair phenomena when a large sampling timestep is used. However, the ratio of repair events from one density to another does not remain constant, varying between 1.2 and 2.

The number of repair events for rings with n edges in the medium density simulations are shown in **Figure 8.8b**. This again shows the same overall trend as seen in the central blue line of **Figure 8.8a**, starting at an average of 3.90 rebirths per ring at $\Delta t = 0.200$ µs and



(a) Fraction of time between first birth and last death that hexagons are alive in simulations with different areas per molecule.

(b) Fraction of time between first birth and last death that rings of a given size are alive at a medium area per molecule.

Figure 8.9: Fraction of time between first birth and last death of rings as a function of sampling timestep. This excludes ephemeral rings that are alive for only one sampling step, and the shaded regions are a 50.0 % bootstrap confidence interval.

dropping to 1.50 rebirths per ring at $\Delta t = 10.0\,\mu s$. Hexagons, octagons and decagons all show a near identical change in rebirths as a function of Δt, while rectangles show slightly more rebirths per ring. This is most likely because rectangles are very strained rings so are easily broken, but are formed in complex sections of the network so are easily reformed.

The significance of repair events can also be measured by seeing for what fraction of its total lifespan a given ring is alive: a ring that is reborn 0.200 µs after dying is significantly different to a ring that is reborn 20.0 µs after dying. Fractions of time spent alive are shown in **Figure 8.9**, again split by area per molecule and size of ring at a medium density. We can see that as the density increases in **Figure 8.9a**, rings are alive for a smaller fraction of the time between first birth and last death. In a denser network there are more options for two rings to alternate between which of the two is alive, and more nearby polymers which can force rings open. In less dense networks, rings are less likely to reform. This is because there are fewer candidates to form viable rings nearby, and the rings are well separated from other polymers that might knock or damage them.

The next trend, in **Figure 8.9b**, is that rings with more edges show a larger fraction of

their total lifespan alive. This is counter to expectations that a ring with more edges is more easily damaged. The reason for large rings being alive for more of their lifespan is because they are more flexible, so can absorb damage without breaking, but small rings must break rather than deform. Both of these overall trends show that the limit of fraction of time alive tends to 1 as Δt gets larger, as fleeting repair phenomena are 'stepped over'.

8.3.4 Combined Trajectories.

Thus far in this chapter, the analysed results have all been of single trajectories. While a persistence diagram can only depict a single trajectory (as there is no way to consistently assign ring IDs between different trajectories), a birth-death diagram has no such constraint. By combining the data in birth-death diagrams over multiple simulations, the trends can become clearer.

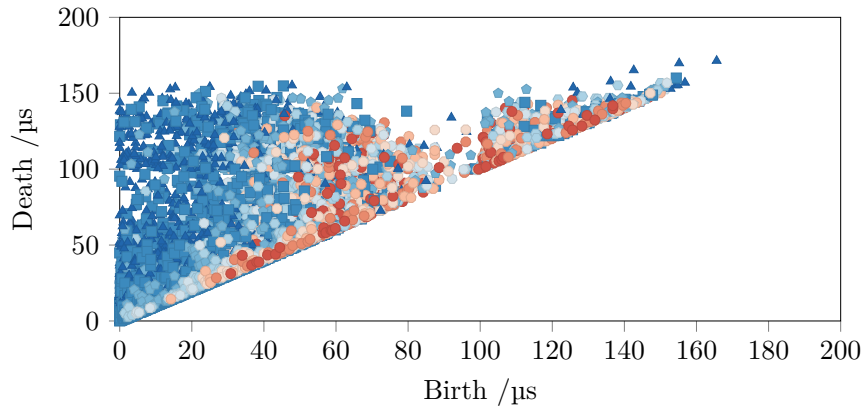
The self-assembly to rupture simulations described in **Section 8.3.2** and then visualised in **Figure 8.5** were repeated 20 times with the same simulation conditions but different random seeds. Birth-death data were collected for each ring observed over those simulations, and the resultant scatter plots are shown in the panels of **Figure 8.10**. By combining many simulations, a number of patterns become clear:

- At high density (low area per molecule) rings are formed immediately from the start of the simulation, but this is not true at low density where a gap is visible in the left hand side of **Figure 8.10c**.
- Most early rings died rapidly, shown by a cluster of data points near the bottom left diagonal.
- However, in the high density networks, triangles and rectangles that formed early on survived for long times.

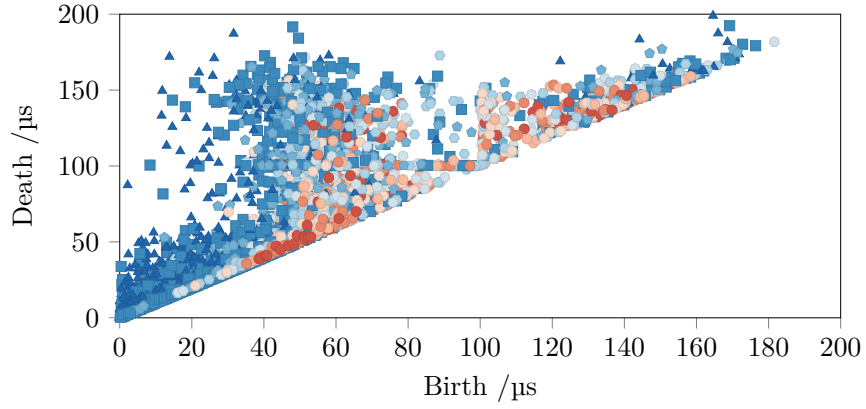
- Larger, and relatively long lived, rings form in the 40.0 µs to 80.0 µs period, as shown by a cluster of red points.

- There are visible horizontal and vertical lines where the simulation model changes from a harmonic to a Morse potential and was optimised. These lines were less obvious when looking at a single trajectory.

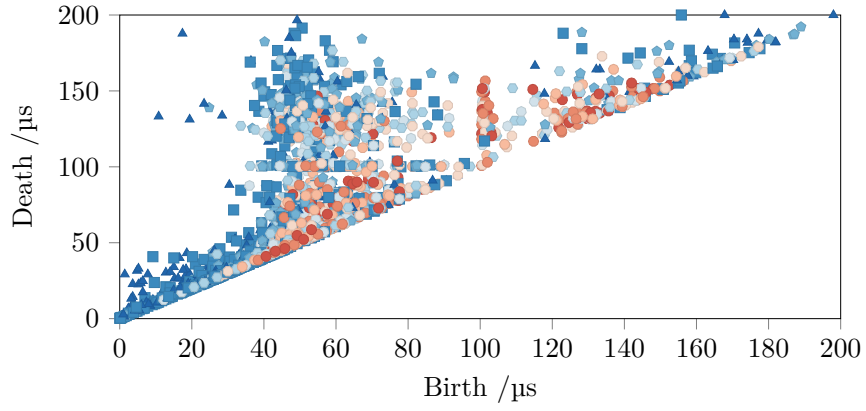
By repeating the simulations and combining the data, the overall trends in ring formation and damage behaviours are clearer than for a single simulation. In low density networks (such as **Figure 8.10b**) formation happens smoothly and rupture happens suddenly. This can be contrasted with high density networks (such as **Figure 8.10c**) where formation happens suddenly and rupture happens smoothly. Furthermore, the trends in which rings are formed or destroyed at various times can more clearly be seen in the combined diagrams. The overall trend along the diagonal shows small rings are initially formed, then stable large rings, then small rings again in the detritus.



(a) $46\,100\,\mathrm{nm}^2$ per polymer ($\rho = 2.17 \times 10^{-5}\,\mathrm{nm}^{-2}$).



(b) $61\,800\,\mathrm{nm}^2$ per polymer. ($\rho = 1.62 \times 10^{-5}\,\mathrm{nm}^{-2}$)



(c) $82\,200\,\mathrm{nm}^2$ per polymer. ($\rho = 1.22 \times 10^{-5}\,\mathrm{nm}^{-2}$)

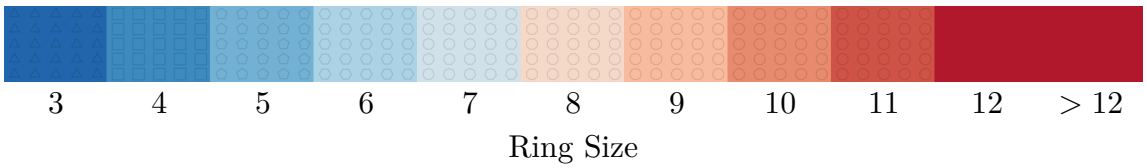


Figure 8.10: Combined birth-death diagrams for trajectories repeated with the same conditions but different random seeds. Each shows rings formed across 20 simulations for each set with the same area per molecule. Ephemeral rings that die as soon as they were born have been omitted for clarity.

8.4 Chapter Conclusions.

This chapter described a new set of visualisation and analysis methods. These new methods were inspired by PH, and can track damage to ring networks over time. The central mathematical idea was to track at which times rings (uniquely identified by their constituent edges) existed. Persistence diagrams were created in which vertical bars showed the lifespan of a given ring and lines drawn horizontally would show which rings existed at a given time point t. Birth-death diagrams were also created, in which the first birth and final death of each ring were plotted as scatter points.

Persistence diagrams were particularly useful to analyse repair effects, and allowed the evolution of structure to be seen visually over time. They were also useful in identifying and characterising repair phenomena: when a ring was broken and then repaired, it led to a vertically striped persistence bar. Persistence diagrams also indirectly showed the rate of ring formation, which could be seen in the gradient of the line described by the bases of the bars. Finally, persistence diagrams coloured by number of edges per ring were useful to identify the different stages in the formation or rupture of a ring. The progress in colouration of bars from left to right indicated whether larger rings or smaller rings were being formed, which provided physical insight into the processes occurring.

Birth-death diagrams were useful in identifying critical stages over a trajectory. Scatter points on the birth-death diagrams show where groups of rings were born or died, depending on the processes occurring. For example, birth-death diagrams combining data points from many trajectories showed a sharp line indicating a critical moment for ring formation or rupturing.

Two metrics were extracted from the trajectories, which allowed a characterisation of

damage using ring structure instead of polymer structure. The first of these was the number of rebirths per ring, which measured repair phenomena and was higher for high density networks which broke and re-formed small rings very often. The second of these was the fraction of its total lifespan that each ring spent alive. This was low when rings were often broken, and there were long gaps between repair events.

All of the metrics and diagrams were calculated for different sampling timesteps Δt. The same overall structural trends were visible regardless of the value of Δt, showing that the visualisation and analysis methods are robust. A small Δt was useful to study high frequency repair events, and a large Δt was useful to quickly study long-lived phenomena.

Finally, data extracted from different trajectories (under the same conditions) were combined. The combined birth-death diagrams showed critical phenomena (like sudden formation or rupture) clearly. They also highlighted important changes in the simulation and reduces the influence of individual rings on the analysis.

Chapter 9: Conclusions and Future Work.

9.1 Conclusions.

In this thesis, different computational methods have been developed in order to study the assembly, structure and damage to collagen IV-like networks. Each method demonstrated has explored a different aspect of collagen-like networks ranging from a high level graph theory approach (to understand structure), to a coarse-grained polymer model (to study assembly and damage).

To establish a point of comparison, an image analysis process was designed and used on existing microscope images, in which networks have been formed under a wide range of experimental conditions. Each image had to be analysed manually, owing to the ambiguity in assigning a graph structure to an image. In general, the networks had an average node degree of approximately three and hexagonal ring structures, so most of the work discussed in this thesis was started from an idealised hexagonal network.

The first model developed used a Monte Carlo bond switching algorithm that explored the range of possible network structures. This used an extended version of the Wooten-Winer-Weaire bond switching algorithm, inserting disorder into hexagonal networks by breaking and reforming edges. It was found that the bond switching model successfully introduced physically reasonable defects into a network, and could quickly generate many large network structures reminiscent of biological arrangements. However, the abstract graph based nature of this model made it less useful to understand assembly phenomena.

Next a coarse-grained polymer model was developed, in which individual collagen IV-like molecules were treated as semiflexible bead and spring polymers, and in which head beads attracted other head beads, and body beads repelled other body beads. The simple polymer model spontaneously self-assembled into complex network structures, and showed the importance of kinetics in network formation. The use of a coarse-grained model allowed the physical effects responsible for structural properties to be effectively untangled. For example, the concentration of is found salt experimentally to affect both polymer stiffness and attractions. In the model, these can be controlled controlled by varying different parameters K_θ and $\varepsilon_{\mathrm{HH}}$.

The coarse-grained polymer model was used to understand mechanical properties of collagen-like networks, replacing the Molecular Dynamics simulations with Molecular Statics for computational efficiency. This helped elucidate the experimentally observed differences in biaxial and uniaxial stretching, and indicated that the collagen network alone cannot explain the stiffness observed in human ocular lens capsules.

The rupturing of the lens capsule is a significant medical problem. By employing the model to stretch the collagen networks beyond their natural limits, the most likely rupturing mechanisms were obtained. Defects were introduced into an ordered graph-based network by switching bonds or removing edges, showing the importance of flexibility in the rupturing process.

A common theme across these models was the assumption that collagen networks are 2D. By relaxing the 2D constraint on the model polymers, it was possible to study how the surface rumpled. This revealed that the networks were significantly flat, and could be modelled well with a fractal self-affine scaling law. Adding defects caused the network to rumple around the defect site, significantly relieving strain.

The final results in this thesis were based in persistent homology, establishing a new analysis framework to study how rings formed over time. Time persistence diagrams were created, in which vertically striped bars show the birth, death and repair of rings over the course of a simulation. This was useful for visualising complex formation and rupturing phenomena, and provided more detail on what had previously only been described by heavily averaged metrics.

In this thesis, the emphasis was placed on simplicity of modelling. As a result, there are rich avenues for future studies available. Some of the most promising avenues include using temporal persistence diagrams to track ring structure in other, more general network systems such as amorphous atomic networks (including amorphous carbon, boron nitride or silica bilayers) or 2D arrangements of colloids. The random graph based models could create starting structures for engineering simulations of the mechanical responses of collagen networks. This could provide a powerful tool to study of how mechanical response changes as a function of age. The polymer model itself could be improved with a more detailed potential model taking different coordinations and lateral interactions into account. This may include specific parameterisations aiming to reproduce properties of interest, or a more complex potential model involve anisotropic attractions, or tangling in 3D space. Finally, the series of layers present in a real network could be included by allowing motion into the z axis, and either repeated layers or a carefully chosen periodic boundary condition.

Chapter 10: References.

(1) P. Fratzl & R. Weinkamer, *Prog. Mater. Sci.*, 2007, **52**, 1263–1334 (cit. on p. 1).

(2) N. Reznikov, H. Chase, V. Brumfeld, R. Shahar & S. Weiner, *Bone*, 2015, **71**, 189–195 (cit. on pp. 1, 16).

(3) R. Jayadev & D. R. Sherwood, *Curr. Biol.*, 2017, **27**, R207–R211 (cit. on p. 1).

(4) S.-W. Chang & M. J. Buehler, *Mater. Today*, 2014, **17**, 70–76 (cit. on p. 1).

(5) A. Duane, *Am. J. Ophthalmol.*, 1922, **5**, 865–877 (cit. on p. 2).

(6) H. Helmholtz, *Arch. für Opthalmologie*, 1855, **2**, 1–74 (cit. on p. 2).

(7) K. Wang, D. Venetsanos, J. Wang & B. K. Pierscionek, *Sci. Rep.*, 2016, **6**, 2–11 (cit. on p. 2).

(8) R. A. Schachar & A. J. Bax, *Compr. Ther.*, 2001, **27**, 122–132 (cit. on p. 2).

(9) H. J. Burd, S. J. Judge & J. A. Cross, *Vision Res.*, 2002, **42**, 2235–2251 (cit. on p. 2).

(10) E. A. Hermans, M. Dubbelman, G. L. van der Heijde & R. M. Heethaar, *Vision Res.*, 2006, **46**, 3642–3650 (cit. on p. 2).

(11) A. Pellegrino, H. J. Burd, L. Pinilla Cortés *et al.*, *Exp. Eye Res.*, 2018, **168**, 19–27 (cit. on p. 2).

(12) R. F. Fisher, *J. Physiol.*, 1969, **201**, 1–19 (cit. on pp. 2, 3, 89, 99, 102).

(13) S. Krag & T. T. Andreassen, *Prog. Retin. Eye Res.*, 2003, **22**, 749–767 (cit. on pp. 2, 3, 89, 91, 96, 97, 99, 100, 102, 105, 106).

(14) H. J. Burd, *Biomech. Model. Mechanobiol.*, 2009, **8**, 217–231 (cit. on pp. 2, 3, 66, 92, 94, 99, 100, 102, 103, 154).

(15) S. D. McLeod, V. Portney & A. Ting, *Br. J. Ophthalmol.*, 2003, **87**, 1083–1085 (cit. on p. 3).

(16) K. M. Mak & R. Mei, *Anat. Rec.*, 2017, **300**, 1371–1390 (cit. on p. 3).

(17) P. D. Yurchenco & H. Furthmayr, *Biochemistry*, 1984, **23**, 1839–1850 (cit. on pp. 4, 23–25, 43, 47, 52, 54, 57, 66, 72, 76, 83, 86, C2).

(18) F. Burla, S. Dussi, C. Martinez-Torres *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 8326–8334 (cit. on p. 4).

(19) F. T. Lewis, *Anat. Rec.*, 1926, **33**, 331–355 (cit. on pp. 4, 6).

(20) R. A. Boerboom, K. N. Krahn, R. T. A. Megens *et al.*, *J. Struct. Biol.*, 2007, **159**, 392–399 (cit. on p. 5).

(21) A. M. Barlow, L. B. Mostaço-Guidolin, E. T. Osei, S. Booth & T.-L. Hackett, *PLoS One*, 2020, **15**, e0229278 (cit. on p. 5).

(22) S. P. Reese, N. Farhang, R. Poulson, G. Parkman & J. A. Weiss, *Biophys. J.*, 2016, **111**, 1797–1804 (cit. on p. 5).

(23) F. T. Lewis, *Anat. Rec.*, 1928, **38**, 341–376 (cit. on p. 6).

(24) D. A. Aboav, *Metallography*, 1970, **3**, 383–390 (cit. on p. 6).

(25) D. Weaire, *Metallography*, 1974, **7**, 157–160 (cit. on p. 7).

(26) D. A. Aboav, *Metallography*, 1980, **13**, 43–58 (cit. on p. 7).

(27) S. F. Edwards & K. D. Pithia, *Phys. A Stat. Mech. its Appl.*, 1994, **205**, 577–584 (cit. on p. 7).

(28) M. A. Fortes, *Philos. Mag. Lett.*, 1993, **68**, 69–71 (cit. on p. 7).

(29) H. Wang & G. Q. Liu, *Epl*, 2012, **100**, DOI: 10.1209/0295-5075/100/68001 (cit. on p. 7).

(30) J. K. Mason, R. Ehrenborg & E. A. Lazar, *J. Phys. A Math. Theor.*, 2012, **45**, DOI: 10.1088/1751-8113/45/6/065001 (cit. on p. 7).

(31) A. Chremos & P. J. Camp, *Phys. Rev. E*, 2007, **76**, DOI: 10.1103/PhysRevE.76.056108 (cit. on p. 7).

(32) M. E. J. Newman, *Phys. Rev. Lett.*, 2002, **89**, 208701 (cit. on pp. 8, 38).

(33) D. Ormrod Morley, DPhil, University of Oxford, 2020 (cit. on pp. 8, 29, 31, 37, 172).

(34) T. E. Klein & C. C. Huang, *Biopolymers*, 1999, **49**, 167–183 (cit. on p. 8).

(35) W. D. Cornell, P. Cieplak, C. I. Bayly *et al.*, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197 (cit. on p. 8).

(36) D. L. Bodian, R. J. Radmer, S. Holbert & T. E. Klein, *Pacific Symp. Biocomput. 2011, PSB 2011*, 2011, 193–204 (cit. on pp. 9, 90).

(37) P. T. Underhill & P. S. Doyle, *J. Nonnewton. Fluid Mech.*, 2004, **122**, 3–31 (cit. on pp. 9, 61).

(38) P. T. Underhill & P. S. Doyle, *J. Rheol. (N. Y. N. Y).*, 2005, **49**, 963–987 (cit. on p. 9).

(39) N. Rezaei, A. Lyons & N. R. Forde, *Biophys. J.*, 2018, **115**, 1457–1469 (cit. on pp. 9, 10, 65, 78).

(40) M. W. H. Kirkness, K. Lehmann & N. R. Forde, *Curr. Opin. Chem. Biol.*, 2019, **53**, 98–105 (cit. on pp. 9, 61).

(41) Y. Zhang, E. P. Debenedictis & S. Keten, *Soft Matter*, 2019, **15**, 3807–3816 (cit. on p. 9).

(42) S. W. Chang, S. J. Shefelbine & M. J. Buehler, *Biophys. J.*, 2012, **102**, 640–648 (cit. on p. 9).

(43) E. J. Amis, C. J. Carriere, J. D. Ferry & A. Veis, *Int. J. Biol. Macromol.*, 1985, **7**, 130–134 (cit. on p. 9).

(44) A. Al-Shaer, Undergraduate Honours Thesis, Simon Fraser University, 2018 (cit. on pp. 10, 65).

(45) C. F. Cummings, V. Pedchenko, K. L. Brown *et al.*, *J. Cell Biol.*, 2016, **213**, 479–494 (cit. on pp. 10, 55, 58, 66, 78).

(46) V. Pedchenko, R. Bauer, E. N. Pokidysheva *et al.*, *J. Biol. Chem.*, 2019, **294**, 7968–7981 (cit. on p. 10).

(47) B. Wu, C. Mu, G. Z. And *et al.*, *Langmuir*, 2009, **25**, 11905–11910 (cit. on p. 10).

(48) Z. Wang, Q. Xiao, X. Song, Y. Wan & J. Zhu, *J. Food Qual.*, 2017, **2017**, 1–10 (cit. on pp. 10, 16, 21, 22, 24, 25, 43, 44, 52, 54, 58, 78, 82, 86, C2).

(49) K. D. Collins, *Methods*, 2004, **34**, 300–311 (cit. on p. 10).

(50) E. N. Gilbert, *Ann. Math. Stat.*, 1959, **30**, 1141–1144 (cit. on p. 11).

(51) P. Erdős & A. Rényi, *Magy. Tudományos Akadémia Mat. Kut. Intézetének Kőzleményei*, 1960, **5**, 17–61 (cit. on p. 11).

(52) R. Albert & A.-L. Barabási, *Rev. Mod. Phys.*, 2002, **74**, 47–97 (cit. on p. 12).

(53) A.-L. Barabási & R. Albert, *Science (80-.).*, 1999, **286**, 509–512 (cit. on pp. 12, 36).

(54) R. Khanin & E. Wit, *J. Comput. Biol.*, 2006, **13**, 810–818 (cit. on p. 13).

(55) K. J. Bos, D. F. Holmes, R. S. Meadows *et al.*, *Micron*, 2001, **32**, 301–306 (cit. on pp. 16, 24, 25, 43, 44, 52, 54, 86, C2).

(56) P. D. Yurchenco & G. C. Ruben, *J. Cell Biol.*, 1987, **105**, 2559–2568 (cit. on pp. 16, 24, 25, 52, 54, 86, C2).

(57) F. S. A. Cavalcante, S. Ito, K. Brewer *et al.*, *J. Appl. Physiol.*, 2005, **98**, 672–679 (cit. on p. 16).

(58) K. Barnard, S. A. Burgess, D. A. Carter & D. M. Woolley, *J. Struct. Biol.*, 1992, **108**, 6–13 (cit. on pp. 16, 24, 25, 43, 44, 52, 54, 86, 154, C2).

(59) H. Tang, M. J. Buehler & B. Moran, *Ann. Biomed. Eng.*, 2009, **37**, 1117–1130 (cit. on p. 16).

(60) A. D'Amore, J. A. Stella, W. R. Wagner & M. S. Sacks, *Biomaterials*, 2010, **31**, 5345–5354 (cit. on pp. 16, 17).

(61) T. Wagner & P. Behnel, *Non Local Means Denoising*, 2016 (cit. on p. 17).

(62) C. A. Schneider, W. S. Rasband & K. W. Eliceiri, *Nat. Methods*, 2012, **9**, 671–675 (cit. on pp. 17, 18).

(63) G. Steger, *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, **20**, 113–125 (cit. on pp. 17–19).

(64) R. O. Duda & P. E. Hart, *Commun. ACM*, 1972, **15**, 11–15 (cit. on p. 17).

(65) R. Deriche & G. Giraudon, *Int. J. Comput. Vis.*, 1993, **10**, 101–124 (cit. on p. 18).

(66) T. Wagner, M. Hiner & X. Raynaud, *Ridge Detection*, 2018 (cit. on p. 18).

(67) M. de Berg, *Computational geometry : algorithms and applications*, Springer, Berlin ; London, 2nd, rev., 2000 (cit. on p. 19).

(68) S. Schneider & I. F. Sbalzarini, *Finding faces in a planar embedding of a graph*, 2015 (cit. on p. 19).

(69) G. Fabris, A. Lucantonio, N. Hampe *et al.*, *Biophys. J.*, 2018, **115**, 1770–1782 (cit. on pp. 24, 25, 52, 54, 86, C2).

(70) M. H. J. Bailey, D. Ormrod Morley & M. Wilson, *RSC Adv.*, 2020, **10**, 38275–38280 (cit. on pp. 25, 28, 44, 52, 85, 165).

(71) M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262 (cit. on pp. 27, 47, 57, 62, 69, 70, 73, 75–78, 92, 123, 154, C2).

(72) C. P. Broedersz & F. C. Mackintosh, *Rev. Mod. Phys.*, 2014, **86**, 995–1036 (cit. on p. 28).

(73) R. C. Picu, *Soft Matter*, 2011, **7**, 6768–6785 (cit. on p. 28).

(74) F. O. Wooten, K. Winer & D. Weaire, *Phys. Rev. Lett.*, 1985, **54**, 1392–1395 (cit. on p. 28).

(75) F. O. Wooten & D. Weaire, in *Adapation Simulated Annealing to Chem. Optim. Probl.* Ed. J. H. Kalivas, Elsevier Science, 1995, ch. 15, pp. 329–349 (cit. on p. 28).

(76) Y. Tu, J. Tersoff, G. Grinstein & D. Vanderbilt, *Phys. Rev. Lett.*, 1998, **81**, 4899–4902 (cit. on p. 28).

(77) N. Y. Rivier, D. Weaire & R. de Römer, *J. Non. Cryst. Solids*, 1988, **105**, 287–291 (cit. on p. 28).

(78) F. Ebrahem, F. Bamer & B. Markert, *Mater. Sci. Eng. A*, 2020, **780**, 139189 (cit. on pp. 28, 89, 90).

(79) D. Ormrod Morley, P. S. Salmon & M. Wilson, *J. Chem. Phys.*, 2021, **154**, DOI: `10.1063/5.0040393` (cit. on pp. 28, 170, 172).

(80) D. Ormrod Morley & M. Wilson, *J. Phys. Condens. Matter*, 2018, **30**, 50LT02 (cit. on p. 29).

(81) A. Kuronen, K. K. K. Kaski & S. von Alfthan, *Phys. Rev. B*, 2003, 1–4 (cit. on p. 31).

(82) S. L. Mayo, B. D. Olafson & W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909 (cit. on pp. 31, 112, 113).

(83) A. D. Broido & A. Clauset, *Nat. Commun.*, 2019, **10**, 1–10 (cit. on p. 36).

(84) P. Holme, *Nat. Commun.*, 2019, **10**, 8–10 (cit. on p. 36).

(85) N. Y. Rivier & A. Lissowski, *J. Phys. A. Math. Gen.*, 1982, **15**, L143–L148 (cit. on p. 37).

(86) A. Gervois, J. P. Troadec & J. Lemaitre, *J. Phys. A. Math. Gen.*, 1992, **25**, 6169–6177 (cit. on p. 37).

(87) L. da Fontoura Costa, F. Rocha & S. M. Araújo de Lima, *Phys. Rev. E*, 2006, **73**, 011913 (cit. on p. 38).

(88) G. Contreras-Figueroa, L. Hernández-Sandoval & J. L. Aragón, *Theor. Biol. Med. Model.*, 2015, **12**, 27 (cit. on p. 38).

(89) R. Giraud, V. T. Ta & N. Papadakis, *Proc. - Int. Conf. Image Process. ICIP*, 2018, **2017-Septe**, 3455–3459 (cit. on p. 38).

(90) D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. Newman & S. H. Strogatz, *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, 2001, **64**, 7 (cit. on p. 38).

(91) A. Kumar, D. Sherrington, M. Wilson & M. F. Thorpe, *J. Phys. Condens. Matter*, 2014, **26**, 1–6 (cit. on pp. 46, 47, 170, C2).

(92) A. J. Stone & D. J. Wales, *Chem. Phys. Lett.*, 1986, **128**, 501–503 (cit. on p. 53).

(93) H. J. Burd & R. A. Regueiro, *Biomech. Model. Mechanobiol.*, 2015, **14**, 1363–1378 (cit. on p. 55).

(94) R. Kalluri & D. Cosgrove, *J. Biol. Chem.*, 2000, **275**, 12719–12724 (cit. on p. 57).

(95) M. Marenda, E. Orlandini & C. Micheletti, *Nat. Commun.*, 2018, **9**, DOI: 10.1038/s41467-018-05413-z (cit. on p. 58).

(96) J. M. Tavares, L. Rovigatti & F. Sciortino, *J. Chem. Phys.*, 2012, **137**, DOI: 10.1063/1.4737930 (cit. on p. 58).

(97) M. P. Allen & D. J. Tildesley, *Computer Simulations of Liquids*, Oxford University Press, First, 1987 (cit. on p. 59).

(98) A. P. Thompson, H. M. Aktulga, R. Berger *et al.*, *Comput. Phys. Commun.*, 2022, **271**, 108171 (cit. on p. 59).

(99) S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19 (cit. on pp. 59, 93).

(100) D. E. Shaw, P. Maragakis, K. Lindorff-Larsen *et al.*, *Science (80-.).*, 2010, **330**, 341–346 (cit. on p. 60).

(101) J. Jung, W. Nishima, M. Daniels *et al.*, *J. Comput. Chem.*, 2019, **40**, 1919–1930 (cit. on p. 60).

(102) J. Jung, C. Kobayashi, K. Kasahara *et al.*, *J. Comput. Chem.*, 2021, **42**, 231–241 (cit. on p. 60).

(103) Z. Zhang & S. C. Glotzer, *Nano Lett.*, 2004, **4**, 1407–1413 (cit. on p. 60).

(104) Y. Zhao, R. Berger, K. Landfester & D. Crespy, *Polym. Chem.*, 2014, **5**, 365–371 (cit. on p. 60).

(105) D. Fusco & P. Charbonneau, *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 2013, **88**, 1–9 (cit. on p. 61).

(106) R. M. Vanacore, S. Shanmugasundararaj, D. B. Friedman *et al.*, *J. Biol. Chem.*, 2004, **279**, 44723–44730 (cit. on p. 62).

(107) R. Berisio & L. Vitagliano, *Protein Sci.*, 2002, **11**, 262–270 (cit. on p. 62).

(108) N. Kern & D. Frenkel, *J. Chem. Phys.*, 2003, **118**, 9882–9889 (cit. on p. 63).

(109) J. P. Doye, A. A. Louis, I. C. Lin *et al.*, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2197–2205 (cit. on p. 63).

(110) R. Timpl, H. Wiedemann, V. Van Delden, H. Furthmayr & K. Kühn, *Eur. J. Biochem.*, 1981, **120**, 203–211 (cit. on pp. 65, 66).

(111) R. Dölz, J. Engel & K. Kühn, *Eur. J. Biochem.*, 1988, **178**, 357–366 (cit. on p. 65).

(112) K. Barnard, E. D. T. Atkins, M. A. Taylor & L. J. Gathercole, *Biopolymers*, 1993, **33**, 897–902 (cit. on p. 72).

(113) D. Ormrod Morley, A. Thorneywork, R. Dullens & M. Wilson, *Phys. Rev. E*, 2020, 1–13 (cit. on pp. 85, 170).

(114) P. A. Janmey, M. E. McCormick, S. Rammensee *et al.*, *Nat. Mater.*, 2007, **6**, 48–51 (cit. on p. 88).

(115) S. S. Soofi, J. A. Last, S. J. Liliensiek, P. F. Nealey & C. J. Murphy, *J. Struct. Biol.*, 2009, **167**, 216–219 (cit. on pp. 88, 102).

(116) R. Michael, M. Mikielewicz, C. Gordillo *et al.*, *Investig. Ophthalmol. Vis. Sci.*, 2012, **53**, 6109–6114 (cit. on pp. 89, 100).

(117) L. P. Cortés, H. J. Burd, G. A. Montenegro *et al.*, *Investig. Ophthalmol. Vis. Sci.*, 2015, **56**, 2926–2932 (cit. on p. 89).

(118) O. Vinogradov, *Int. J. Comput. Methods*, 2006, **3**, 153–161 (cit. on pp. 89, 90).

(119) F. Bamer, F. Ebrahem & B. Markert, *Materialia*, 2020, **9**, 1–16 (cit. on p. 89).

(120) E. Polak & G. Ribiere, *ESAIM Math. Model. Numer. Anal. - Modélisation Mathématique Anal. Numérique*, 1969, **3**, 35–43 (cit. on p. 91).

(121) D. Fiocco, G. Foffi & S. Sastry, *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 2013, **88**, 1–5 (cit. on p. 90).

(122) S. Patinet, D. Vandembroucq & M. L. Falk, *Phys. Rev. Lett.*, 2016, **117**, 1–5 (cit. on p. 90).

(123) L. J. Gathercole, K. Barnard & E. D. T. Atkins, *Int. J. Biol. Macromol.*, 1989, **11**, 335–338 (cit. on p. 91).

(124) P. Kiviranta, J. Rieppo, R. K. Korhonen *et al.*, *J. Orthop. Res.*, 2006, **24**, 690–699 (cit. on p. 93).

(125) E. Ban, H. Wang, J. Matthew Franklin *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 6790–6799 (cit. on pp. 93, 97).

(126) F. Bamer, F. Ebrahem & B. Markert, *Comput. Mater. Sci.*, 2019, **163**, 301–307 (cit. on pp. 94, 133).

(127) S. Krag, T. Olsen & T. T. Andreassen, *Investig. Ophthalmol. Vis. Sci.*, 1997, **38**, 357–363 (cit. on pp. 97, 102, 115, 126, 127).

(128) B. P. Danysh & M. K. Duncan, *Exp. Eye Res.*, 2009, **88**, 151–164 (cit. on p. 99).

(129) J. A. Last, S. J. Liliensiek, P. F. Nealey & C. J. Murphy, *J. Struct. Biol.*, 2009, **167**, 19–24 (cit. on pp. 100, 102).

(130) M. H. J. Bailey & M. Wilson, *RSC Adv.*, 2022, **12**, 2171–2180 (cit. on pp. 109, 125, 126, 132, 135, 142, 165).

(131) G. A. Aerni, *Clin. Sports Med.*, 2013, **32**, 289–301 (cit. on p. 109).

(132) N. Astbury, M. Wood, U. Gajiwala *et al.*, *Community Eye Heal. J.*, 2008, **21**, 6–8 (cit. on p. 110).

(133) A. Ionides, *Br. J. Ophthalmol.*, 2001, **85**, 222–224 (cit. on p. 110).

(134) X. Liu, L. Wang, C. Du, D. Li & Y. Fan, *Comput. Methods Biomech. Biomed. Engin.*, 2015, **18**, 914–921 (cit. on p. 110).

(135) I. T. Tsinopoulos, G. I. Karras, A. B. Haidich & K. T. Tsaousis, *J. Cataract Refract. Surg.*, 2015, **41**, 1783–1784 (cit. on p. 110).

(136) F. Ebrahem, F. Bamer & B. Markert, *J. Mater. Sci.*, 2020, **55**, 3470–3483 (cit. on p. 110).

(137) C. G. Broyden, *IMA J. Appl. Math.*, 1970, **6**, 76–90 (cit. on p. 112).

(138) R. Fletcher, *Comput. J.*, 1970, **13**, 317–322 (cit. on p. 112).

(139) D. Goldfarb, *Math. Comput.*, 1970, **24**, 23 (cit. on p. 112).

(140) D. F. Shanno, *Math. Comput.*, 1970, **24**, 647 (cit. on p. 112).

(141) P. Virtanen, R. Gommers, T. E. Oliphant *et al.*, *Nat. Methods*, 2020, **17**, 261–272 (cit. on p. 112).

(142) M. J. Buehler, *J. Mater. Res.*, 2006, **21**, 1947–1961 (cit. on pp. 113, 115).

(143) F. W. Campbell & G. Westheimer, *J. Physiol.*, 1960, **151**, 285–295 (cit. on p. 126).

(144) T. E. Lockhart & W. Shi, *Ergonomics*, 2010, **53**, 892–903 (cit. on p. 126).

(145) D. Ormrod Morley, A. L. Goodwin & M. Wilson, *Phys. Rev. E*, 2020, **102**, 1–8 (cit. on p. 130).

(146) A. J. Licup, S. Münster, A. Sharma *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9573–9578 (cit. on p. 154).

(147) C. H. Lui, L. Liu, K. F. Mak, G. W. Flynn & T. F. Heinz, *Nature*, 2009, **462**, 339–341 (cit. on p. 154).

(148) D. A. Kirilenko & P. N. Brunkov, *Ultramicroscopy*, 2016, **165**, 1–7 (cit. on p. 154).

(149) K. Fang, R. Adame, H. N. Yang, G. C. Wang & T. M. Lu, *Appl. Phys. Lett.*, 1995, **2077**, 2077 (cit. on p. 154).

(150) B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, New York, 1st edn., 1982, vol. 26, pp. 131–132 (cit. on p. 155).

(151) S. Alexander, in *Dyn. Fractal Surfaces*, ed. F. Family & T. Vicsek, World Scientific Publising Co., Singapore, 1st edn., 1991, ch. Chapter 2: pp. 52–64 (cit. on p. 155).

(152) F. Family & T. Vicsek, *J. Phys. A. Math. Gen.*, 1985, **18**, L75–L81 (cit. on p. 155).

(153) S. Labat, C. Guichet, O. Thomas, B. Gilles & A. Marty, *Appl. Surf. Sci.*, 2002, **188**, 182–187 (cit. on p. 156).

(154) T. Gredig, E. A. Silverstein & M. P. Byrne, *J. Phys. Conf. Ser.*, 2013, **417**, 1–5 (cit. on p. 156).

(155) V. W. Stone, A. M. Jonas, B. Nysten & R. Legras, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1999, **60**, 5883–5894 (cit. on p. 156).

(156) T. Gneiting & M. Schlather, *SIAM Rev.*, 2004, **46**, 269–282 (cit. on pp. 156, 159).

(157) A. Favata, A. Micheletti, P. Podio-Guidugli & N. M. Pugno, *Meccanica*, 2017, **52**, 1601–1624 (cit. on p. 162).

(158) M. H. J. Bailey & M. Wilson, *Submitted to Bulletin of Mathematical Biology*, 2022, DOI: 10.26434/chemrxiv-2022-k4thc (cit. on pp. 169, 182, 184).

(159) S. Le Roux & P. Jund, *Comput. Mater. Sci.*, 2010, **49**, 70–83 (cit. on p. 170).

(160) M. Sadjadi & M. F. Thorpe, *Phys. Rev. E*, 2016, **94**, DOI: 10.1103/PhysRevE.94.062304 (cit. on p. 170).

(161) Y. Lee, S. D. Barthel, P. Dłotko *et al.*, *Nat. Commun.*, 2017, **8**, 15396 (cit. on p. 170).

(162) K. Xia, Z. Zhao & G. W. Wei, *J. Chem. Phys.*, 2015, **143**, DOI: 10.1063/1.4931733 (cit. on p. 170).

(163) C. M. M. Pereira & R. F. De Mello, *Expert Syst. Appl.*, 2015, **42**, 6026–6038 (cit. on p. 171).

(164) L. Steinberg, PhD Thesis, University of Southampton, 2019 (cit. on p. 171).

(165) E. Carlsson, G. Carlsson & V. De Silva, *Int. J. Comput. Geom. Appl.*, 2006, **16**, 291–314 (cit. on p. 172).

(166) G. Carlsson, *Bull. Am. Math. Soc.*, 2009, **46**, 255–308 (cit. on p. 172).

(167) S. S. Sørensen, C. A. N. Biscio, M. Bauchy, L. Fajstrup & M. M. Smedskjaer, *Sci. Adv.*, 2020, **6**, DOI: 10.1126/sciadv.abc2320 (cit. on p. 172).

(168) L. Wasserman, *Annu. Rev. Stat. Its Appl.*, 2018, **5**, 501–532 (cit. on p. 172).

(169) Y. Onodera, S. Kohara, S. Tahara *et al.*, *J. Ceram. Soc. Japan*, 2019, **127**, 853–863 (cit. on p. 172).

(170) Z. Dong, J. Pu & H. Lin, *Comput. Aided Geom. Des.*, 2021, **88**, 102004 (cit. on p. 172).

(171) Y. M. Chung, C. S. Hu, Y. L. Lo & H. T. Wu, *Front. Physiol.*, 2021, **12**, 1–18 (cit. on p. 183).

(172) V. Kovacev-Nikolic, P. Bubenik, D. Nikolić & G. Heo, *Stat. Appl. Genet. Mol. Biol.*, 2016, **15**, 19–38 (cit. on p. 183).

(173) W. Humphrey, A. Dlake & K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

(174) A. A. Hagberg, D. A. Schult & P. J. Swart, *7th Python Sci. Conf. (SciPy 2008)*, 2008, 11–15.

(175) A. Kohlmeyer, *Topotools*, 2020.

(176) O. Tange, *GNU Parallel*, 2018.

(177) C. R. Harris, K. J. Millman, S. J. van der Walt *et al.*, *Nature*, 2020, **585**, 357–362.

(178) J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 90–95.

(179) A. Buades, B. Coll & J.-M. Morel, *Image Process. Line*, 2011, **1**, 208–212 (cit. on p. B2).

Appendix A: Glossary.

LAMMPS Molecular Dynamics software used in this thesis, standing for 'Large-scale Atomic/Molecular Massively Parallel Simulator'. 59, 63, 90, 93, 114, 130, 174

2D two dimensional 1, 5, 6, 13, 16, 19, 20, 23, 30, 55, 58, 59, 66, 76, 77, 81, 89, 90, 93, 96, 99, 100, 102, 124, 130, 133, 153–158, 160–162, 165, 168–170, 181, A3

3D three dimensional 19, 76, 81, 82, 96, 99, 153–156, 162, 165, 168, 170

7S One of the interacting ends of a collagen IV molecule. 3, 4, 8, 62, 63, 68, 75, 81

assortativity The correlation coefficient, often denoted r, between a given numeric property of neighbouring nodes in a graph. 85, 86

BFGS Broyden-Fletcher-Goldfarb-Shanno 112

bootstrap Statistical method to generate confidence intervals by resampling the original data array with replacement many times. 166, 188, 189

coarse-grained A simplified model of an atomic system, combining many discrete atoms into large 'beads'. ii, 5, 57, 58, 60, 65, 74, 89–92, 97, 105, 107, 115, 126, 151, 156, 157, 162, 165, 168, 174, 178, 195, 196

collagen IV Biopolymer of interest 1–4, 8–10, 13, 21, 22, 26, 27, 57, 58, 61–63, 65, 66, 75, 87, 90, 91, 97, 107, 109, 111, 113, 115, 153, 154, 158, 173, 195, 196, A1

degree The number of edges connected to a given node, denoted k during this thesis. In chemical literature, this is often referred to as the coordination number. 4, 11, 12, 24–26, 32–40, 43, 45, 47, 52, 53, 63, 66, 68, 71, 72, 74, 76, 77, 79, 84–86, 92, 93, 102, 104, 112, 125, 126, 130, 133, 136, 137, A2

IPR Inverse Participation Ratio xiv, 118–121, 146, 147, 150

maximum entropy A graph that is well described by the distribution of degrees following maximum entropy distribution. 37, 45, 46

MC Monte Carlo xviii, 26, 27, 29, 30, 32, 33, 35, 36, 41, 43, 44, 51, 57, 73, 74, 79, 80, 84, 85, 87, 89, 92, 107, 111, 151

MD Molecular Dynamics 8, 26, 57–59, 64, 87, 89, 90, 92, 110, 116, 151, 153, 154, 174, 186

Monte Carlo A simulation protocol involving accepting or rejecting random moves with a probability depending on the energy difference before and after that move. B1

MS Molecular Statics 89, 90, 94–97, 101, 107, 109

MSD Mean Square Displacement 131, 132

NC1 Non-Collagenous 1 3, 4, 8, 10, 62, 63, 68, 81

NLMD non local means denoising 17, 18, 21, B2

PCR posterior capsule rupture 109, 110

peptide A short chain of amino acids 8

PH persistent homology 28, 170, 172, 173, 183, 193

planar A planar graph is one that can be down in a two dimensional (2D) plane with no edges intersecting, and edges only meeting at nodes. 19

RDF Radial Distribution Function 28

scale-free A graph that is well described by a preferential attachment model, with degrees following a power law $p(k) \propto k^{-\alpha}$. 12, 13, 44–46, 56

self-affine A self-affine surface is similar to itself but with different scale factors in different dimensions, a weakened version of the self-similarity property. 153–155, 161, 168

self-assembly A process in which a disorganised system (here of polymers) spontaneously forms an ordered structure. 57, 66, 72, 79, 85, 92, 95, 107, 123, 169, 174, 181

SRC Shape Regularity Criterion 38, 39, 46–49, 79, 80

VACF velocity autocorrelation function xiv, 117, 118, 146, 147, 150

VV Velocity Verlet xviii, 59

WWW Wooten-Winer-Weaire 28, 32, 33, 111, 157

Appendix B: Algorithms.

B.1 Markov Chain Monte Carlo.

The Monte Carlo technique used in this thesis is a Markov chain model, which allows for enhanced sampling without every calculating the full partition function. Each possible state of the network (here, intuitively, a graph of rings) belongs to a finite set of possible outcomes. For example the idealised hexagonal network could be $\mathbf{\Gamma}_1$, and the networks after successive bond switches are denoted $\mathbf{\Gamma}_2, \ldots, \mathbf{\Gamma}_n$. Each individual state has no memory of its history, and has an energy only depending on the state $\mathbf{\Gamma}_i$, and the positions of each node $\boldsymbol{q}_i$. A bond switching move represents a jump between two states $\mathbf{\Gamma}_i$ and $\mathbf{\Gamma}_j$ with a probability π_{ij}, an element of the transition matrix $\boldsymbol{\pi}$. The probability matrix is normalised such that the sum of each row is unity $\sum_j \pi_{ij} = 1$. In the bond switching method, many states cannot be directly access from other states (if they require two bond switching moves), so most elements of this matrix are zero.

At each step, the system will be in a given state Γ_i. The probability that it will be in that state is described by a row vector $\boldsymbol{\rho}_s$, which evolves each step as $\boldsymbol{\rho}_{s+1} = \boldsymbol{\rho}\boldsymbol{\pi}$. Since any state of the system can be reached from any other state in a finite number of bond switching moves (*i.e.* the system is ergodic), the probability tends to a stationary distribution $\boldsymbol{\rho}$ which satisfies the eigenvalue equation

$$\boldsymbol{\rho} = \boldsymbol{\rho}\boldsymbol{\pi} \tag{B.1}$$

The limiting form of the probability matrix is too large to be used directly. Instead,

the expectation value of an observable can be calculated by taking a random walk through configurational states. The random walk is a chain of states $\boldsymbol{S} = \{\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \cdots, \boldsymbol{\Gamma}_S\}$. Since the elements of $\boldsymbol{\pi}$ are drawn from a limiting Boltzmann distribution, the Markov chain $\boldsymbol{S}$ can be constructed by using the Metropolis criterion. With this, the expectation value of some observable O is given by

$$\langle O \rangle = \lim_{S \to \infty} \frac{1}{S} \sum_{i \in \boldsymbol{S}} O(S_i) \tag{B.2}$$

Thus by making a series of Monte Carlo moves and tracking the measured values of $O(S_i)$ we can calculate properties of interest, such as energies or ring statistics.

B.2 Non-Local Means Denoising

Non-Local Means denoising is an algorithm to reduce visual noise in an image by averaging colours in a neighbourhood around a given pixel. Here, the neighbourhood, denoted $\mathcal{N}$, is chosen to be a 21×21 square around the central point x, y. The colour averaging can be performed according to the non local means denoising (NLMD) equation designed by Buades *et al.*

$$\boldsymbol{u'}(x, y) = \frac{\sum_{x', y' \in \mathcal{N}} \boldsymbol{u}(x, y) w(x, y, x', y')}{\sum_{x', y' \in \mathcal{N}} w(x, y, x', y')}, \tag{B.3}$$

with $w(x, y, x', y')$ a weighting function based on the two chosen pixels[179]. The weighting function is an exponential kernel taking the similarity of the two pixels into account

$$w(x, y, x', y') = \exp\left(-\frac{\max(d^2(x, y, x', y') - 2\sigma^2_{\mathrm{NLMD}}, 0)}{h^2}\right) \tag{B.4}$$

with h and σ_{NLMD} free parameters, and d^2 being a similarity metric. The similarity metric takes into account how similar two different pixel environments are by calculating the Euclidean distance (in RGB colour space) between pixels in blocks of size f surrounding (x, y) and

(x', y'):

$$d^2(x, y, x', y') = \frac{1}{3(2f+1)^2} \sum_{i \in \mathrm{R,G,B}} \sum_{\delta x=0}^{f} \sum_{\delta y=0}^{f} \left(u_i(x+\delta x, y+\delta y) - u_i(x'+\delta x, y'+\delta y)\right)^2 \tag{B.5}$$

Appendix C: Data.

C.1 Regularity Data

Name	Figure	$\langle SRC \rangle$	σ_{SRC}
Kumar *et al.*	2a	0.966	0.0240
Kumar *et al.*	2b	0.964	0.0280
Kumar *et al.*	2c	0.957	0.0320
Kumar *et al.*	2d	0.962	0.0250
Kumar *et al.*	2e	0.967	0.0230
Barnard *et al.*	2	0.792	0.147
Barnard *et al.*	4c	0.832	0.115
Barnard *et al.*	6	0.848	0.113
Bos *et al.*	2a (alternative)	0.853	0.146
Bos *et al.*	2a	0.834	0.123
Bos *et al.*	5c	0.836	0.103
Bos *et al.*	5d	0.854	0.0970
Fabris *et al.*	3f	0.837	0.114
Fabris *et al.*	4b	0.810	0.128
Wang *et al.*	1c	0.781	0.159
Wang *et al.*	2a	0.810	0.145
Wang *et al.*	2d	0.841	0.106
Wang *et al.*	3b alternative	0.693	0.218
Wang *et al.*	3b	0.830	0.133
Wang *et al.*	4a	0.847	0.128
Wang *et al.*	5b	0.770	0.151
Wang *et al.*	5c	0.847	0.107
Yurchenco & Furthmayr	4b alternative	0.874	0.0920
Yurchenco & Furthmayr	4b	0.848	0.105
Yurchenco & Ruben	1c	0.871	0.0880
Yurchenco & Ruben	2f	0.911	0.0810

Table C.1: Comparison of regularity data of experimentally imaged collagen IV networks. This table is reproduced from the Supplementary Information of M. H. J. Bailey & M. Wilson, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 1253–1262.

Parameter	Value	Unit	$\langle SRC \rangle$	σ_{SRC}	
θ_{eqm}	120	°	0.769	0.174	
θ_{eqm}	130	°	0.854	0.113	
θ_{eqm}	140	°	0.809	0.147	
θ_{eqm}	150	°	0.746	0.193	
θ_{eqm}	160	°	0.771	0.192	
θ_{eqm}	170	°	0.788	0.181	
θ_{eqm}	180	°	0.799	0.185	
K_θ	5.00×10^{-20}	J	0.763	0.179	
K_θ	1.00×10^{-19}	J	0.782	0.186	
K_θ	1.50×10^{-19}	J	0.783	0.176	
K_θ	2.00×10^{-19}	J	0.788	0.181	
K_θ	2.50×10^{-19}	J	0.800	0.182	
K_θ	3.00×10^{-19}	J	0.794	0.182	
K_θ	4.00×10^{-19}	J	0.786	0.177	
K_θ	5.00×10^{-19}	J	0.774	0.185	
k_l	2.00×10^{-5}	$\mathrm{N\,m^{-1}}$	0.790	0.177	
k_l	5.00×10^{-5}	$\mathrm{N\,m^{-1}}$	0.801	0.174	
k_l	1.00×10^{-4}	$\mathrm{N\,m^{-1}}$	0.787	0.186	
k_l	1.50×10^{-4}	$\mathrm{N\,m^{-1}}$	0.790	0.177	
k_l	2.50×10^{-4}	$\mathrm{N\,m^{-1}}$	0.788	0.178	
t_{cool}	54.0	µs	0.806	0.158	
t_{cool}	70.0	µs	0.781	0.184	
t_{cool}	90.0	µs	0.805	0.174	
t_{cool}	110	µs	0.822	0.166	
t_{cool}	130	µs	0.828	0.162	
t_{cool}	150	µs	0.831	0.159	
$\varepsilon_{\mathrm{HH}}$	1.00×10^{-21}	J	0.844	0.114	
$\varepsilon_{\mathrm{HH}}$	1.50×10^{-21}	J	0.796	0.161	
$\varepsilon_{\mathrm{HH}}$	2.00×10^{-21}	J	0.769	0.179	
$\varepsilon_{\mathrm{HH}}$	2.50×10^{-21}	J	0.748	0.194	
$\varepsilon_{\mathrm{HH}}$	3.00×10^{-21}	J	0.770	0.185	
$\varepsilon_{\mathrm{HH}}$	3.50×10^{-21}	J	0.772	0.189	
$\varepsilon_{\mathrm{HH}}$	4.00×10^{-21}	J	0.780	0.185	
$\varepsilon_{\mathrm{HH}}$	4.50×10^{-21}	J	0.796	0.181	
$\varepsilon_{\mathrm{HH}}$	5.00×10^{-21}	J	0.801	0.176	
$\varepsilon_{\mathrm{HH}}$	5.50×10^{-21}	J	0.798	0.178	
$\varepsilon_{\mathrm{HH}}$	6.00×10^{-21}	J	0.810	0.168	
$\varepsilon_{\mathrm{HH}}$	8.00×10^{-21}	J	0.835	0.157	
$\varepsilon_{\mathrm{HH}}$	1.00×10^{-20}	J	0.839	0.154	
$\varepsilon_{\mathrm{HH}}$	1.20×10^{-20}	J	0.855	0.141	
$\varepsilon_{\mathrm{HH}}$	1.40×10^{-20}	J	0.860	0.134	
$\varepsilon_{\mathrm{HH}}$	1.60×10^{-20}	J	0.857	0.140	
$\varepsilon_{\mathrm{HH}}$	2.00×10^{-20}	J	0.866	0.138	
p	0.200	Pa	0.950	0.0553	
p	0.500	Pa	0.928	0.0893	
p	2.00	Pa	0.871	0.142	
p	5.00	Pa	0.823	0.171	
p	20.0	Pa	0.787	0.187	
p	50.0	Pa	0.781	0.154	C3

Table C.2: Regularity data for self-assembled networks with different parameters from **Section 4.3.4**.

Cross	Plus	$\langle SRC \rangle$	σ_{SRC}
0	10	0.830	0.155
0	100	0.852	0.121
0	2	0.834	0.155
0	25	0.838	0.148
0	4	0.831	0.152
0	50	0.840	0.129
10	0	0.832	0.158
10	10	0.834	0.144
10	100	0.852	0.120
10	2	0.836	0.149
10	25	0.838	0.141
10	4	0.836	0.147
10	50	0.844	0.132
100	0	0.847	0.125
100	10	0.848	0.128
100	100	0.872	0.107
100	2	0.854	0.123
100	25	0.858	0.114
100	4	0.846	0.125
100	50	0.859	0.110
2	0	0.832	0.159
2	10	0.832	0.154
2	100	0.857	0.119
2	2	0.828	0.156
2	25	0.838	0.142
2	4	0.837	0.143
2	50	0.845	0.133
25	0	0.836	0.151
25	10	0.833	0.146
25	100	0.866	0.114
25	2	0.834	0.147
25	25	0.844	0.128
25	4	0.834	0.146
25	50	0.842	0.128
4	0	0.829	0.153
4	10	0.838	0.140
4	100	0.858	0.120
4	2	0.833	0.153
4	25	0.833	0.146
4	4	0.838	0.151
4	50	0.842	0.132
50	0	0.835	0.140
50	10	0.843	0.138
50	100	0.860	0.112
50	2	0.837	0.139
50	25	0.846	0.122
50	4	0.841	0.141
50	50	0.849	0.121

C4

Table C.3: Regularity data for self-assembled networks with different parameters from **Section 4.3.4** seeded with defects.