Inverse Materials Design Employing Self-folding and Extended Ensembles

by

Paul M. Dodd

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in The University of Michigan
2018

Doctoral Committee:

        Professor Sharon Glotzer, Chair
        Professor Nick Kotov
        Assistant Professor Greg van Anders
        Professor Robert Ziff

Paul Dodd

pdodd@umich.edu

ORCID iD: 0000-0002-6376-0774

ACKNOWLEDGEMENTS

A big thanks to my parents and my three sisters. Your love and support was always there for me when I needed it most.

Laura, there are no words to describe how grateful I am you are in my life. Without your love, support, and stern grip on reality, I would not be here today.

Finally, I would like to thank my other committee members, Bob Ziff and Nick Kotov, for their astute questions that always make me think.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Inverse Materials Design Employing Self-Folding and Extended Ensembles The development of new technology is made possible by the discovery of novel materials. However, this discovery process is often tedious and largely consists of trial and error. In this thesis, I present methods to aid in the design of two distinct model systems. In the first case study, I model the 43,380 nets belonging to the five platonic solids to elucidate a universal folding mechanism. I then correlate geometric and topological features of the nets with folding propensity for simple shapes (i.e., tetrahedron, cube, and octahedron), in order to predict the folding propensity of nets belonging to more complex shapes (i.e., dodecahedron and icosahedron). In the second case study, I develop Monte Carlo techniques to sample the alchemical ensemble of hard polyhedra. In general, the anisotropy dimensions (e.g, faceting, branching, patchiness, etc.) of material building blocks are fixed attributes in experimental systems. In the alchemical ensemble, anisotropy dimensions are treated as thermodynamic variables and the free energy of the system in this ensemble is minimized to find the equilibrium particle shape for a given colloidal crystal at a given packing fraction. The method can sample millions of unique shapes within a single simulation, allowing for efficient particle design for crystal structures. Finally, I employ the method to explore how glasses formed from hard polyhedra, which are geometrically frustrated systems, can utilize extra dimensions to escape the glassy state in the extended ensemble.

CHAPTER I

Introduction

At the root of most technological advancements is the discovery of new materials. New materials enable society to achieve the most complex tasks more efficiently. However, the discovery of novel materials is a slow and tedious process that has been driven by trial and error for over two millennia. In 1833, Faraday discovered the property of certain materials to be semiconductors, but it took over 100 years to take this discovery and turn it into the first silicon-based computer. The field of materials design strives to achieve two goals: (1) to discover new materials at a rapid rate, and (2) to quickly implement new technologies based on these discoveries.

In recent years, there has been a hard push to develop dynamic and adaptive materials[1], including molecular machines, DNA robots[2], and colloidal robots[3]. Such systems offer abundant information encoding through chemical bonds, complementarity, and external fields. In this work, we explore how information can be stored in geometric arrangements of (1) rigid bodies that are connected via a hinge, and (2) hard convex polyhedra.

1.1   Self-Folding Nets

The primary systems that we study are the nets of the Platonic solids. A net is defined as the resulting two-dimensional sheet obtained by cutting along the edges of a polyhedron in such a way that the faces can lay in the plane. There are two ways to unfold a tetrahedron, 11 ways to unfold a cube or octahedron, and 43,380 ways to unfold an icosahedron or dodecahedron. The main questions addressed in this work are: Out of all of the nets for a given shape, which one will fold best? Do all of the nets for a given shape fold using the same set of pathways? How does a net's geometry and topology affect the folding process?

By answering these questions, we gained valuable insights into the fundamental physics underpinning stochastic folding, and were then able to distill this knowledge into a set of basic design rules that accurately predict which nets of the dodecahedron will successfully fold. These findings were published in reference [4].

1.2   Extended Ensembles

In statistical mechanics, there exist numerous thermodynamic ensembles. These ensembles define which variables in the system are allowed to vary, and which ones must be held fixed. Extended ensemble methods can treat parameters of the system that are conventionally (or experimentally) constant as thermodynamic variables. In our case, we treat the shape of the hard particles that we study as thermodynamic variables. This treatment led us to develop a Monte Carlo method for the successful design of particles that assemble into target crystal structures. This work was developed in reference [5], and subsequently used in references [6] and [7]. Beyond materials design, we used this method to study glassy systems of hard polyhedra.

Critically, we found that the glass transition can be avoided in this ensemble through the reduction of faceting of shapes. This final study by P. Dodd and S. Glotzer will be published following the completion of this dissertation[8].

References for Chapter I

[1] Daniel M. Sussman, Yigil Cho, Toen Castle, Xingting Gong, Euiyeon Jung, Shu Yang, and Randall D. Kamien. Algorithmic lattice kirigami: A route to pluripotent materials. *Proceedings of the National Academy of Sciences of the United States of America*, 112(24):7449–7453, 2015. doi: 10.1073/pnas.1506048112.

[2] Anupama J. Thubagere, Wei Li, Robert F. Johnson, Zibo Chen, Shayan Doroudi, Yae Lim Lee, Gregory Izatt, Sarah Wittman, Niranjan Srinivas, Damien Woods, Erik Winfree, and Lulu Qian. A cargo-sorting dna robot. *Science*, 357(6356), 2017. ISSN 0036-8075. doi: 10.1126/science.aan6558.

[3] Koohee Han, C. Wyatt Shields, Nidhi M. Diwakar, Bhuvnesh Bharti, Gabriel P. López, and Orlin D. Velev. Sequence-encoded colloidal origami and microbot assemblies from patchy magnetic cubes. *Science Advances*, 3(8), 2017. doi: 10.1126/sciadv.1701108.

[4] Paul M. Dodd, Pablo F. Damasceno, and Sharon C. Glotzer. Universal folding pathways of polyhedron nets. *Proceedings of the National Academy of Sciences of the United States of America*, 2018.

[5] Greg van Anders, Daphne Klotsa, Andrew S. Karas, Paul M. Dodd, and Sharon C. Glotzer. Digital alchemy for materials design: Colloids and beyond. *ACS Nano*, 9(10):9542–9553, 2015. doi: 10.1021/acsnano.5b04181.

[6] Y. Geng, G. van Anders, P. M. Dodd, J. Dshemuchadse, and S. C. Glotzer.

Engineering entropy for the inverse design of colloidal crystals from hard shapes. *ArXiv e-prints*, 2017.

[7] Rose K. Cersonsky, Greg van Anders, Paul M. Dodd, and Sharon C. Glotzer. Relevance of packing to colloidal self-assembly. *Proceedings of the National Academy of Sciences*, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1720139115.

[8] Paul M. Dodd, Greg van Anders, and Sharon C. Glotzer. Avoiding the Glass Transition using the Alchemical Ensemble. *in preparation*.

CHAPTER II

Universal Folding Pathways of Polyhedron Nets

In the 16th century, the Dutch artist Albrecht Dürer investigated which two-dimensional (2D) cuts of non-overlapping, edge-joined polygons could be folded into Platonic and Archimedean polyhedra. Dürer cuts were later called "nets" but, for a long time, the interest around them was mostly restricted to the field of mathematics [1, 2, 3]. A newer concept, self-folding origami adds a modern twist to the ancient art of paper folding. By providing a mechanism to achieve complex three-dimensional (3D) geometries from low-dimensional objects – without the need for manipulation of the constituent parts – self-folding brings the concepts pioneered by Dürer to the forefront of many research fields, from medicine [4] to robotics [5],

Several recent works have leveraged physical forces to achieve controllable folding of 3D objects including light [6], pH [7], capillary forces [8], cellular traction [9] or thermal expansion [10]. Other works have investigated the relationship between geometric attributes of the object being folded and its propensity for successful folding. In the macroscopic folding of kirigami sheets – origami-like structures containing cuts and creases – the effect of different cut patterns on the material's stress-strain behavior has been elucidated [11, 12, 13] and the "inverse design problem" of finding cuts leading to the folding of a particular target structure has been solved [14]. For

sub-millimeter-sized capsules, formed via non-stochastic folding of nets into polyhedra, it has been suggested that nets fold with higher yield when they are compact [15, 8, 16], but the reason for this correlation remains unclear. In natural systems, the canonical example of self-folding occurs for proteins, where a string of amino acids navigate, thermodynamically, from a denatured (unfolded) state to a natured (folded) one. Even after many decades of study, however, a universal relationship between molecular sequence and folded state – which could provide crucial insight into the causes and potential treatments of many diseases – remains out of reach [11, 12, 13].

In this work, we study the thermodynamic foldability of 2D nets for all five Platonic solids. Despite being the simplest and most symmetric 3D polytopes, the family of Platonic shapes suffices to demonstrate the rapid increase in design space as shapes become more complex: a tetrahedron has two possible net representations, cubes and octahedra each have 11 nets, dodecahedra and icosahedra have, each, 43,380 distinct net unfoldings. Here we are interested in the thermodynamic self-folding of these nets. Our goal is to understand how topology affects yield in the stochastic folding of 3D objects. The advantage is three-fold. First, by using a collection of sheets folding into the same target shape, we isolate the geometric attributes responsible for high-yield folding. Second, the model allows exhaustive computation of the pathways followed by the nets during folding, elucidating how some nets achieve high-yield. Third, by studying increasingly more complex objects – from tetrahedra to icosahedra – we can use the folding mechanisms quantified in the simplest objects to predict, and potentially validate, their occurrence in the more complex shapes.
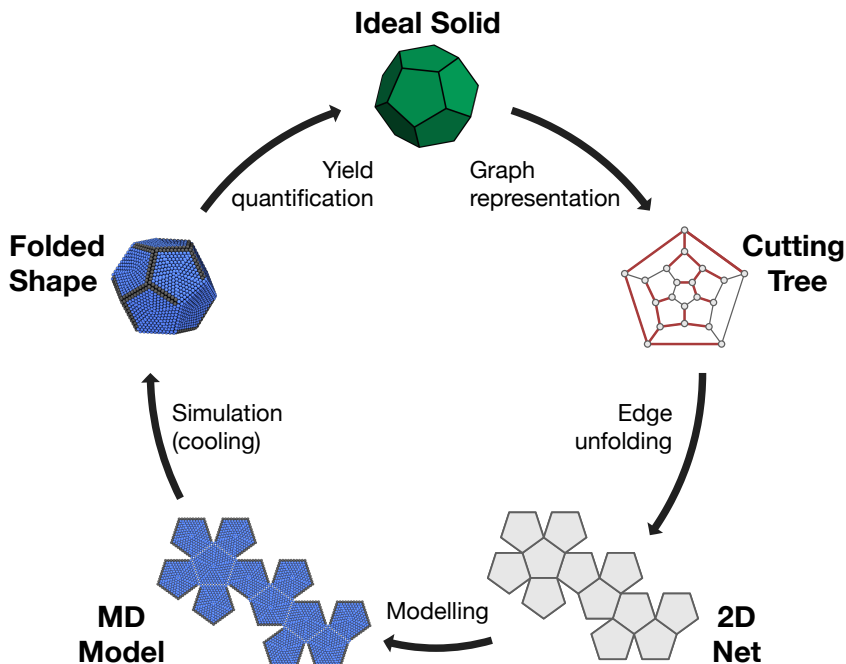
Figure 2.1: Workflow for polyhedra folding study. Starting with a target 3D polytope (top, and moving clockwise), a graph representation is used to enumerate different permutations of edge cuttings. The vertices (edges) in the so-called cutting tree graph correspond to the vertices (edges) of the polyhedron. Red edges in the cutting tree mark those being cut. From the cutting tree, a planar, hinged connection of non-overlapping faces (a net) results. Each face of the net is then modeled as a union of spheres rigidly held together, tethered along hinges via harmonic springs. In the MD model shown, gray spheres interact via a Lennard-Jones potential. Blue spheres interact with each other and with gray spheres via a purely repulsive WCA potential. The simulations are initialized at high temperature and brought to low temperature either following a fast quench or a slow annealing protocol. Once the final temperature is reached the final configuration is compared to the target shape. The folding yield is then calculated as the probability of achieving the desired 3D shape via the particular cooling protocol.

Beginning with a target Platonic shape (Fig. 2.1), we construct a graph whose vertices and edges correspond to those in the polyhedron. This mapping of the shape to a graph facilitates the exhaustive search of all distinct nets by allowing spanning tree enumeration [3]. A set of pre-chosen edges (the cutting tree) are then cut, in a process called edge unfolding, to create a single, contiguous and flat 2D sheet of non-overlapping faces: a net. For the Platonic shapes, whose nets are enumerated[17], this can be repeated exhaustively until all distinct nets are discovered (see Section 2.5 for more details). We note that for other shapes, while the process of computing all

nets might become computationally prohibitive, it has been recently demonstrated that a subset of interest for these nets can be computed algorithmically [18]. We list all 86,784 nets for the Platonic solids in a database [19].

Each net is modeled as a sheet of rigid polygons connected to adjacent polygons via harmonic springs. The polygons are composed of rigidly connected spheres and the influence of thermal fluctuations on a single net, suspended in implicit solvent, is modeled via Langevin molecular dynamics (more details in Section 2.5). We assign non-specific, short-ranged attractive (sticky) interactions between all edges not joined by springs of a net so that the polyhedron formed from folding is also the ground state configuration. This does not guarantee, however, the uniqueness of the ground state and, as we will see, other 3D foldings can arise. Unless explicitly noted otherwise, by "folded state" we refer to the original polyhedron.

As in proteins and other biomolecules, the non-specificity of the interactions between edges of the nets allows for both native and non-native contacts. As a consequence, when the system temperature is rapidly decreased (quenched), kinetic traps are possible and net misfolds are observed. This possibility for kinetic traps raises the question: among all nets generated by unfolding a polyhedron, which of them show the highest propensity to re-fold into the original polyhedron?

2.1 More Compact, "Leafy" Nets Fold More Reliably.

To identify the nets able to fold reliably into their polyhedron of origin we performed hundreds of cooling simulations for each net using both a fast and slow cooling protocol (see Section 2.5 for more details). The two distinct nets for the tetrahedron, hereafter referred to as the triangular net and the linear net (Fig. 2.3a), showed

remarkably different folding propensies for the fast cooling protocol: out of 126 simulations, all triangular nets folded into the target tetrahedron while only 54% of the linear nets succeeded – the other 46% experiments resulted in misfolded configurations. In general the slower cooling rate simulations yielded a higher folding probability for each net. This is expected as the net has more time to find the global minimum. Similar simulations for each of the 11 nets of both the cube and octahedron revealed even larger differences: for the cube nets (Fig. 2.3b), only three of the nets showed greater than 50% success in folding, and only at slow cooling rates; for the octahedron (Fig. 2.3c), none of the nets achieve higher than 50% success rate for either cooling rates.

The misfolded configurations for the tetrahedron and cube nets were incomplete three-dimensional geometries (showing, for instance, faces collapsed on top of each other or bonds incompatible with the formation of the respective target shape). In contrast, octahedron nets often folded into another three-dimensional shape: a concave, boat-like conformation with the same number of edge-edge contacts as the octahedron: in other words, a degenerate ground-state. This competing structure, which is less symmetric than the octahedron, has higher rotational entropy resulting in a lower free energy than the octahedron(see Fig. 2.2a for folding probabilities for the boat conformation). A competition between similar degenerate structures was reported for finite clusters of six attractive spherical colloids [20], where symmetry breaking leads to the formation of the same boat-like conformation.

a) boat   bi-tetrahedra   b)

bipyramid   flat polyhedra

Image	Database Id	oat Prob. slow cooling (%)
	0	3.2
	1	0.0
	2	5.6
	3	0.0
	4	0.0
	5	0.0
	6	0.0
	7	12.7
	8	0.0
	9	2.4
	10	9.5
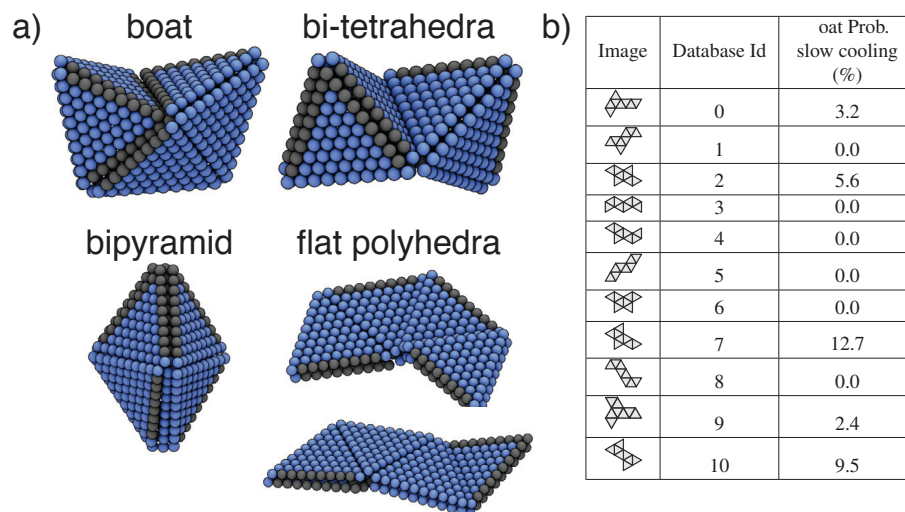
Figure 2.2: Octahedron net misfolds. (a) Examples of trapped states that can be achieved by octahedron nets. The boat conformation occurs more often. (b) Probabilities of the octahedron nets folding into the boat conformation for fast $(2.5{\times}10^{-6}T/t)$ and slow $(2.5{\times}10^{-8}T/t)$ cooling rates.

Figure 2.3 shows that, despite having the same ground-state energy, not all nets of a polyhedron are equivalent. In general, we observe that the nets that fold most reliably are the most compact and have the highest number of leaves on its cutting graph (green dots in nets in Fig. 2.3). A net is said to be more compact if it has a large number of leaves, the one degree vertices on the cutting tree, and a small diameter, the longest shortest path between any two faces on the face graph. Exact values for the leaves and diameter are shown in Table 2.1. Most strikingly, even nets differing only by the location of a single face can have folding probabilities reduced from 99% to 17%. What causes one shape to fold nearly perfectly every time while a slightly different one fails to do so almost as frequently? And why do net 'leafiness' and 'compactness' correlate with folding yield?

Image	Database Id	T_m	Diameter	Leaves	Paths
	0	2.50	3	2	2
	1	2.50	2	3	3
	0	4.10	4	3	9
	1	3.57	4	4	12
	2	3.10	4	3	6
	3	4.04	4	3	12
	4	4.16	5	2	6
	5	3.35	3	4	16
	6	4.15	5	2	2
	7	3.95	4	2	6
	8	3.99	5	4	8
	9	3.98	5	3	5
	10	3.74	3	4	18
	0	2.44	6	2	5
	1	2.49	6	3	7
	2	2.48	5	4	12
	3	2.56	7	3	6
	4	2.48	6	3	4
	5	2.49	7	2	6
	6	2.50	5	4	14
	7	2.45	5	3	12
	8	2.47	7	2	2
	9	2.50	5	3	12
	10	2.44	5	2	8

Table 2.1: Relevant data for the tetrahedron, cube, and octahedron nets. For each net, a database ID, melting temperature T_m, diameter, number of leaves, and number of paths connecting unfolded and folded states is shown.
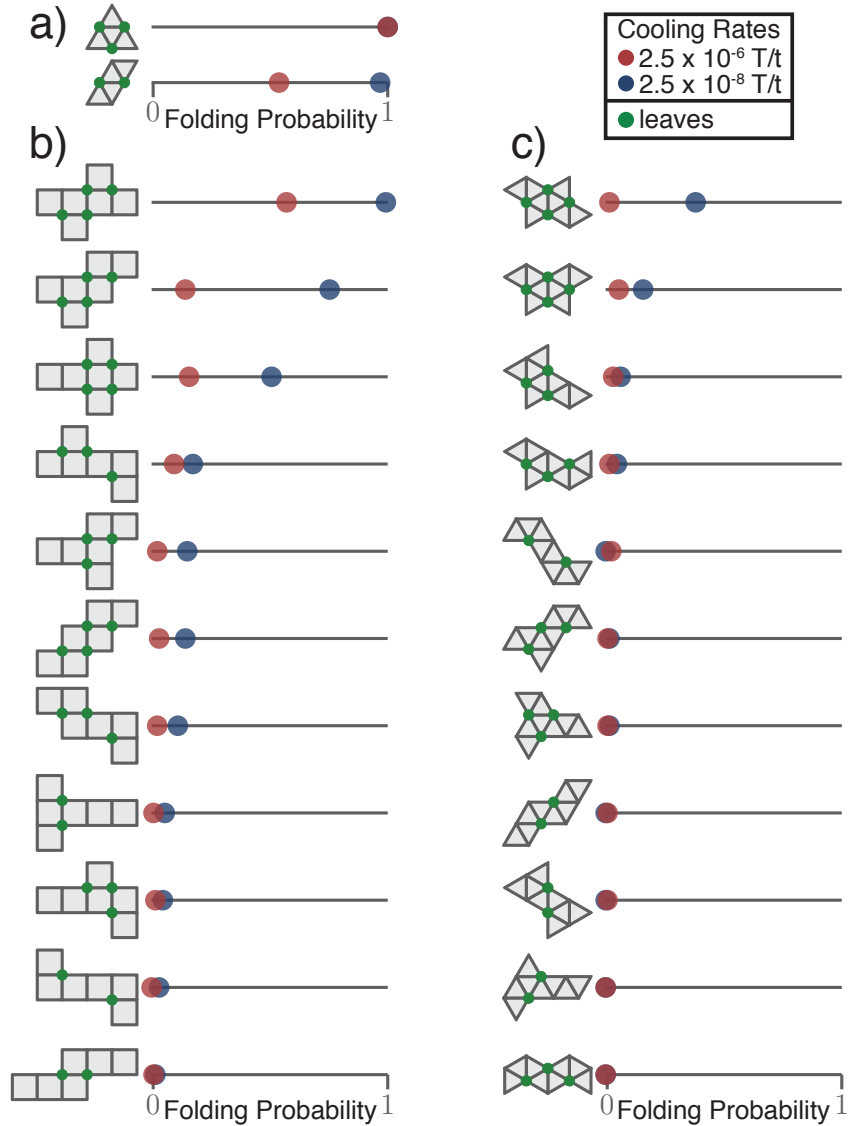
Figure 2.3: Effect of net topology on the folding probability for the simplest Platonic solids. Folding yield, defined by the fraction of simulations that reached the folded state, was calculated for two cooling rates. Nets for each polyhedron are ordered from highest to lowest yield at low cooling rate. a) For the two tetrahedron nets (triangular, top, and linear, bottom), noticeable difference in folding success rate is visible for rapid cooling rates, but similar folding propensity is found for slower cooling rates. b) For the cube nets 8/11 nets fold poorly (below 50%) even for slow folding reactions while for the octahedron (c), most nets are unable to fold into the original shape. Generally the more compact nets fold better: nets that have a high number of leaves on their cutting tree and a smaller diameter usually fold best.

2.2 High Temperature Folding Happens via Native Contacts.

To answer why small differences in net topology can have a large impact on the net's folding propensity, we used Markov state models (MSM) [21, 22, 23] to compute the pathways through which nets fold into their folded state. Since quench rate was observed to affect the folding propensity of the nets, we run constant temperature simulations of each net while computing the rate of transitions between two states (flux). A representative folding network created is shown in Fig. 2.4a (see Fig. 2.5 for other example networks). Arrows represent observed transitions between different states and each arrow has a thickness proportional to the probability flux of the transition being observed (see Section 2.5 for more details). To simplify, we show only the most visited pathways (i.e. those whose combined flux account for at least 50% of the total reactive flux between unfolded and folded states). Intermediate configurations can achieve the folded state via the formation of native (green) or non-native (red) contacts. If a pathway includes native contacts only, every newly formed bond is compatible with the final polyhedron and error correction is not needed. When the folded state is achieved via incompatible bonds, we observe that these non-native contacts can sometimes have an "active" role by bringing otherwise far-away native contacts closer together, facilitating folding. In other cases the non-active contacts contribute only passively and folding occurs sequentially after the misfold occurs until it reaches a point where the non-native contact must break for folding to continue. In either case, non-native contacts must eventually break before a folded configuration can be achieved.
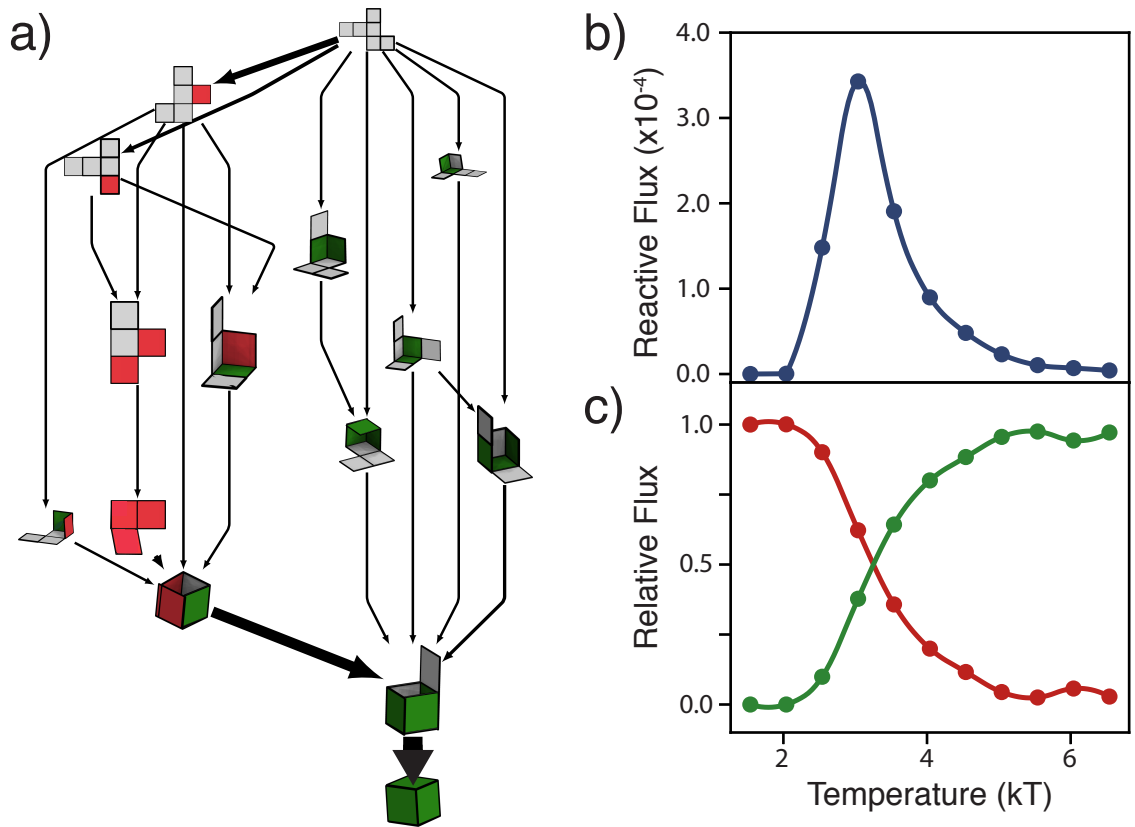
Figure 2.4: Folding pathways for a representative cubic net. a) Intermediate folding states (nets, in the diagram), arise when new bonds between edges are formed. States are connected with an arrow when a pathway from one state to another is observed in the simulation. The thickness of such a connection is proportional to the measured probability of this transition being observed. For better visualization, only the most visited states are shown. Red (green) faces correspond to a state where a non-native (native) contact is formed. The pathways containing only native contacts follow a "sequential" folding process, where one face folds at a time. Non-native pathways can lead to the native folded shape via one of two mechanisms: i) either the misfold helps bring otherwise far away faces together, or ii) the folding proceeds sequentially after the misfold occurs until it reaches a point where the non-native contact must break for folding to continue. In both cases, if the correct polyhedron is achieved in the end, the non-native contact is corrected along the pathway. The network represents 55% of the folding flux at $T = 3kT$. b) The total probability flux, defined as the sum of the fluxes along all pathways, for the representative net in a) as a function of temperature. The peak around temperature $T = 3kT$ shows that there is a temperature at which there is a maximum number of expected transitions from the unfolded to folded state per unit time. (c) The relative amount of flux going through pathways that use the native (green curve) and non-native (red curve) pathways. The folding occurs mostly via native contacts when the system is kept at higher temperatures.
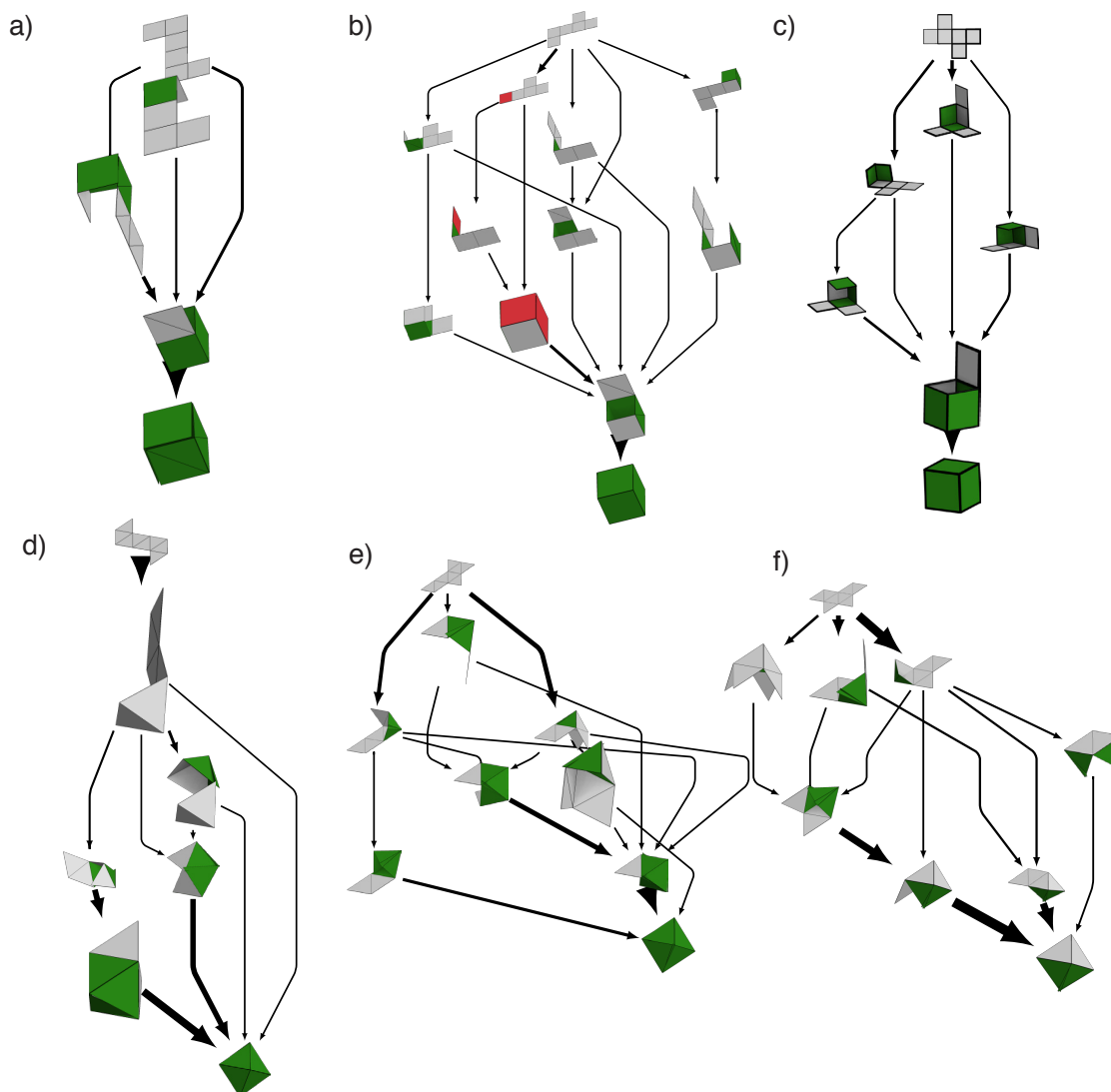
Figure 2.5: Folding networks for representative cube and octahedron nets that have two (a,d), three (b,e), and four leaves (c,f). Networks were calculated at the melting temperature reported in Table 2.1.

The total flux connecting unfolded and folded states (Fig. 2.4b) first increases with temperature and then decreases to zero at high temperature. At low T the folding flux is low because the states are mostly trapped into a few configurations, i.e. the slow kinetics inhibits bond breaking. As the temperature increases, bonds can now break and the folding/unfolding process occurs at a higher rate. At intermediate

T, a maximum in reactive flux is observed, meaning that there is a temperature at which there is a maximum number of expected transitions from unfolded to folded state per unit time τ. The representative network shown in Fig. 2.4a was calculated at that peak temperature. Finally, at high T the unfolded state is preferred and again the flux vanishes. These trends are also true for the other nets studied see Fig. 2.6. If we separate the flux into those following native and non-native contacts, we see (Fig. 2.4c) that the folding pathways at high temperature mostly follow the formation of native contacts while the behavior inverts for low temperatures, and mostly non-native contacts are observed. There were only two nets that a crossover temperature was not observed. The triangular tetrahedral net does not exhibit a crossover temperature since it has no traps. The best folding cubic net also does not have a cross over temperature in the range of temperature we study, while traps exist, the fraction of native contacts decreases and the temperature decreases but never falls below 50% (see Fig. 2.6). This T dependency is also observed in colloids, where the assembly of an icosahedron is monomeric at high temperatures, but, at low temperature, particles first aggregate into large clusters (not necessarily compatible with icosahedral symmetry) and those later rearrange into the ground state structure [24].
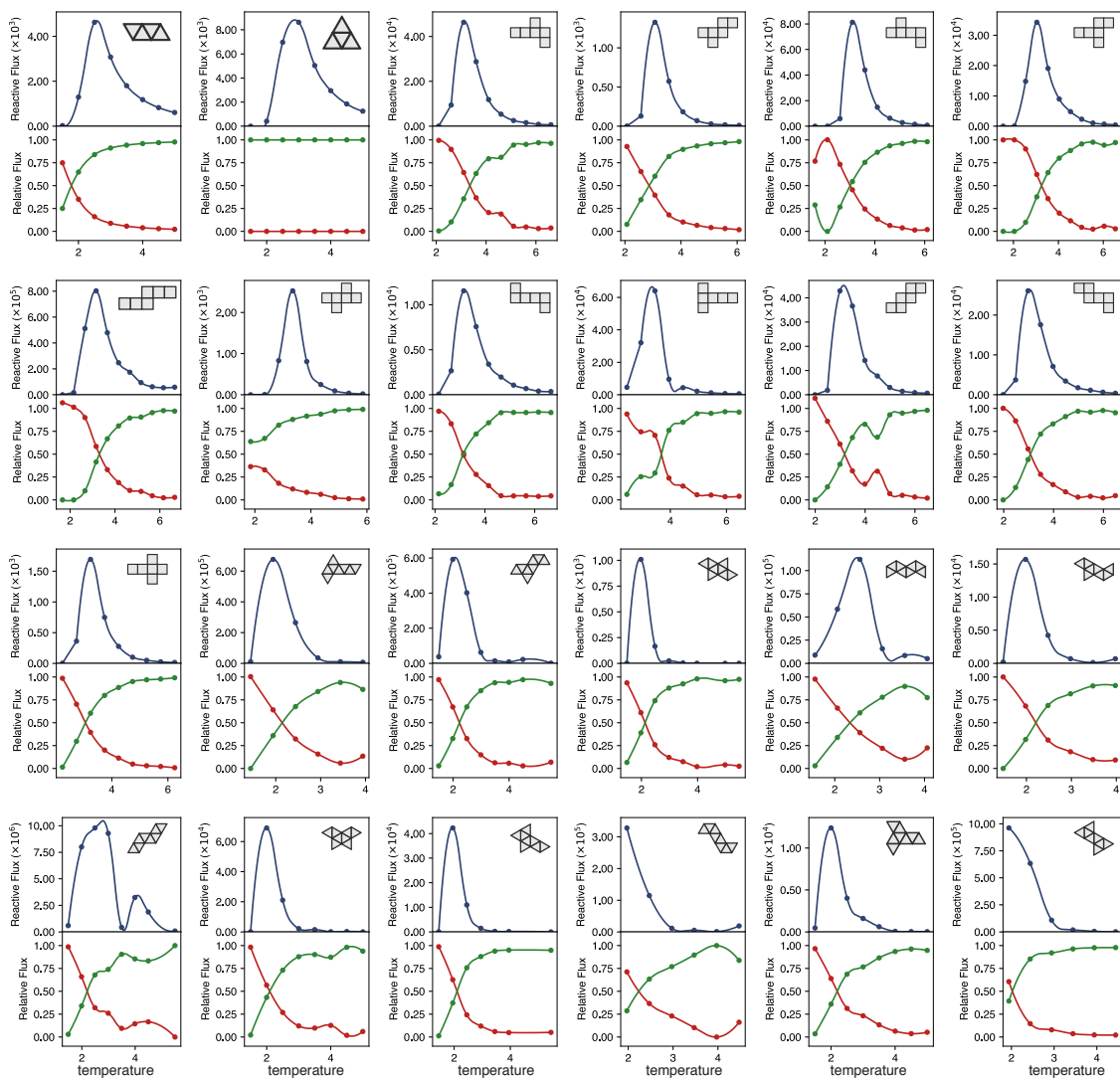
Figure 2.6: Temperature dependence of reactive and relative fluxes for the tetrahedron, cube, and octahedron nets. The net corresponding to each plot is given as an inset. For each plot the blue curve represents the total reactive flux while the red and green curves represent the fraction of the flux corresponding to non-native and native contacts, respectively.
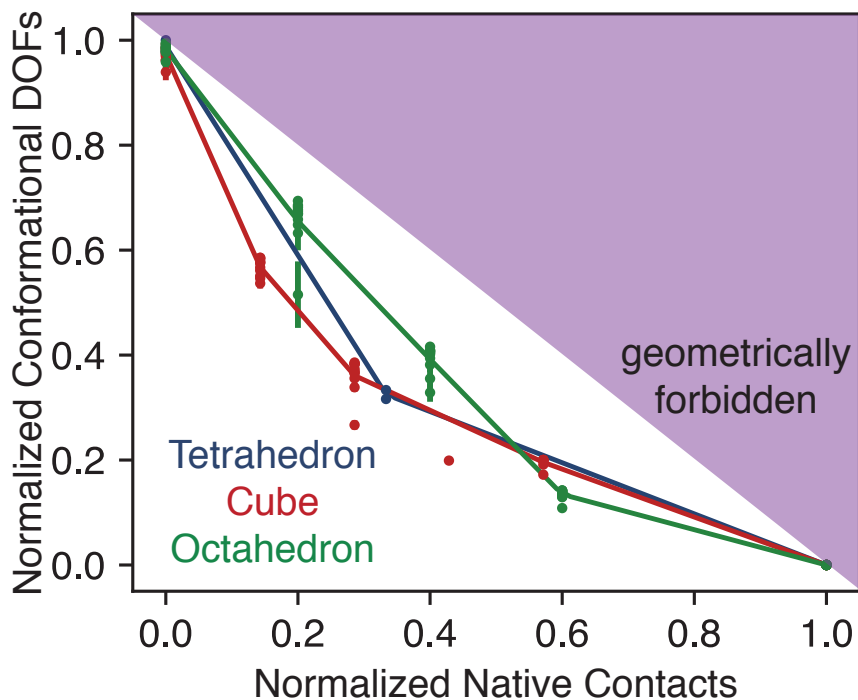
Figure 2.7: Projected pathways computed from the MSM of all of the nets onto two order parameters: native contacts and number of internal degrees of freedom. We find that each net will reduce their degrees of freedom in approximately the same way. The data points are weighted averages of the number of degrees of freedom over all pathways for each given value of Q/Q_{folded}. The lines are drawn to guide the eye. The upper right corner of the plot is a geometrically forbidden region arising from the fact that the net can not gain a bond without losing at least one degree of freedom. The temperature plotted for each net is given by $T_m + 0.5$, T_m is reported for each net in Table 2.1.

To gain an understanding of the mechanisms underlying the observation that native contacts are favored at high temperatures during folding, we calculated the number of degrees of freedom associated with each intermediate.

2.3 Nets Follow a Universal Balance Between Entropy and Enthalpy.

The fact that more compact nets and those with many leaves generally fold with higher yield suggests that nets might fold locally, in a manner that reduces the fewest degrees of freedom, thereby maximizing the conformational entropy along the folding pathways. To test whether this trade-off between maximizing degrees of freedom and forming native contacts occurs at high T, we calculate the number N

of internal degrees of freedom and the number Q of native contacts as a net folds. We do so for all 24 nets of the tetrahedron, cube, and octahedron nets. Fig. 2.7 shows that, remarkably, all nets follow a folding pathway that achieves a narrow balance between reduction of degrees of freedom and gain of potential energy. In practical terms, high temperature folding happens locally such that, at each step of the process, the system strives to maximize its conformational entropy. From this observation, we hypothesize the following mechanism for the folding of general nets at high temperature. Folding should primarily happen: i) via nearby (local) connections – favoring compact nets with many leaves; ii) along one of the optimal trade-off paths – favoring nets with high degeneracy in the number of such optimal paths.
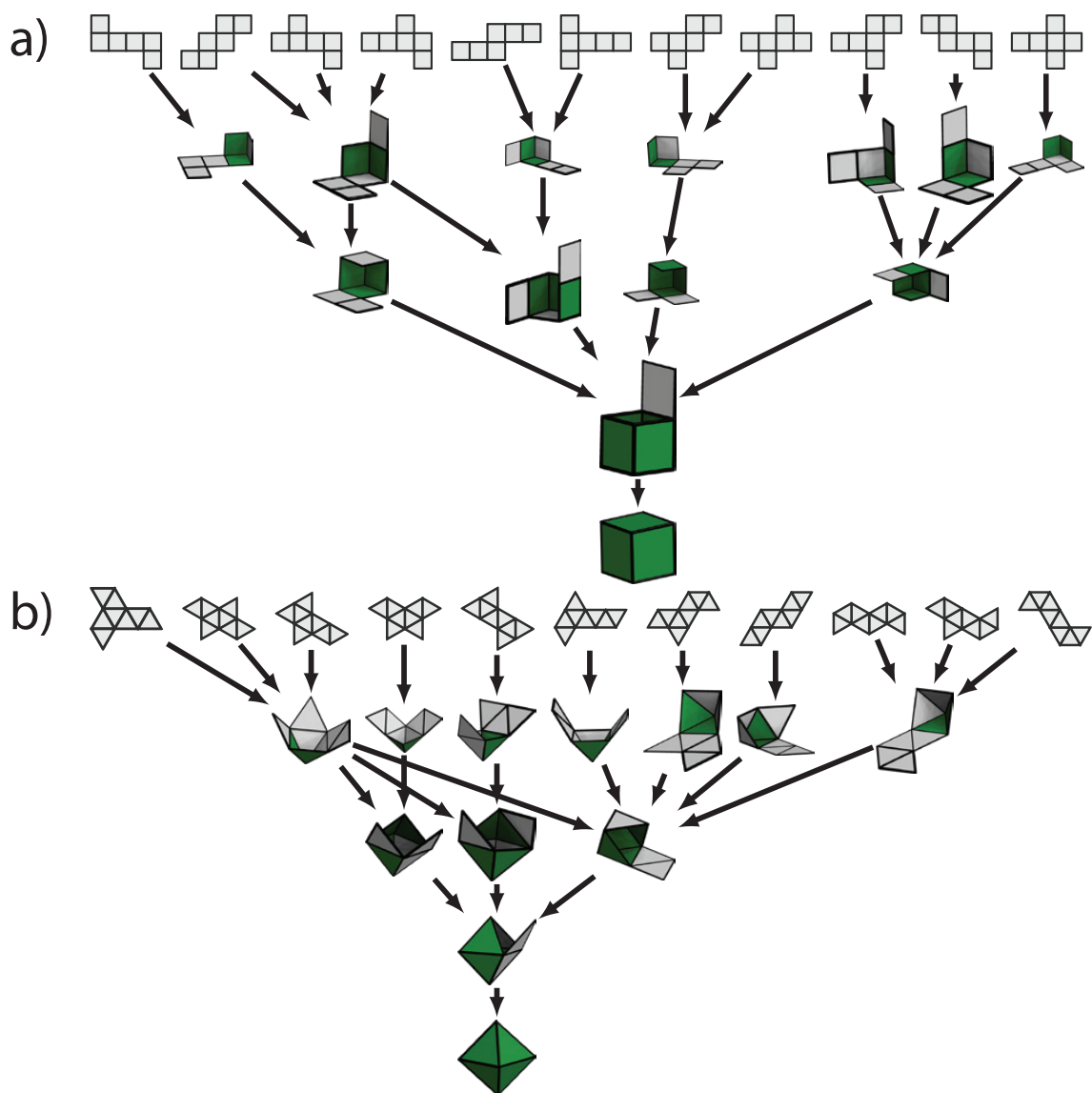
Figure 2.8: Combined dominant pathways for the 11 nets of the cube (a) and the 11 nets of the octahedron (b) calculated at high temperature. In both cases the dominant pathways are sequential (one face folding at a time) and include only native contacts. As in Fig. 2.4, arrows indicate transitions between two states, the arrow thickness being proportional to the probability of observing such a transition.

Using this hypothesis we devised an algorithm to generate high-temperature pathways for the 86,760 nets of the dodecahedron and icosahedron, without the need for a full MSM calculation. We do so by first enumerating candidate bonds that can be made if they are next to each other on the net or in the intermediate and then by
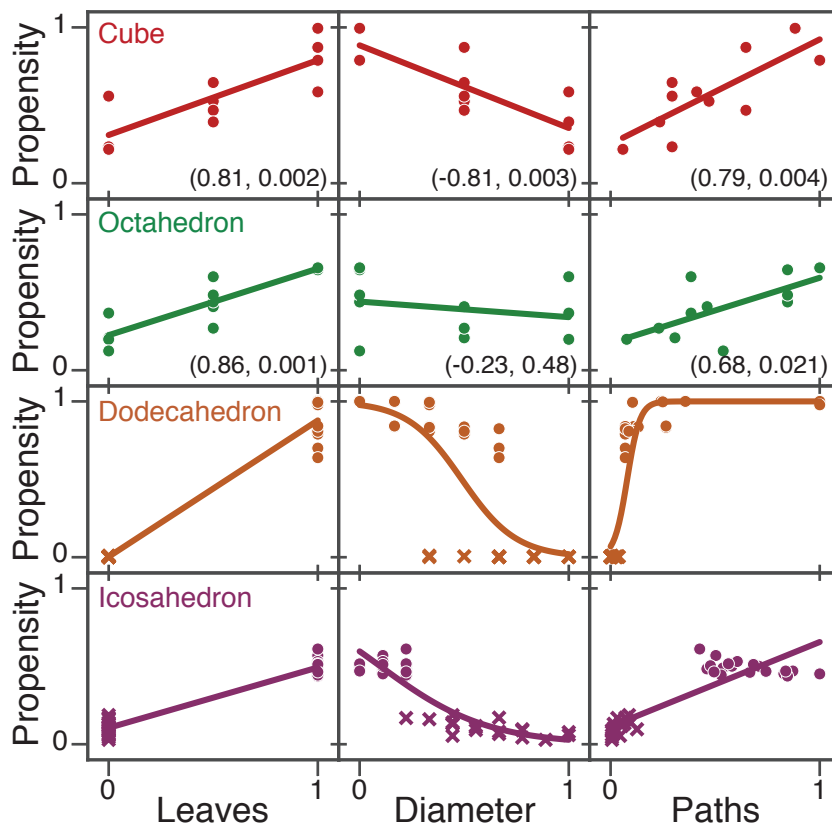
Figure 2.9: Correlation between different geometric and topological quantities calculated for the net and the folding propensity, defined as the average fraction of native contacts formed across all 125 quenching simulations. Each row corresponds to the cube (red), octahedron (green), dodecahedron (orange), and icosahedron (purple) respectively; data for the tetrahedron was omitted because there are only two nets. Pearson coefficients and p-values are reported in the panels for the cube and octahedron, as these data are used to predict the 'good' and 'bad' nets of the dodecahedron and icosahedron. We found that the number of leaves provide the strongest linear correlation for both the cube and octahedron (Pearson coefficients are 0.81 and 0.86, respectively) and so the number of leaves was used to try and predict the nets that would fold with both high and low propensity. In the bottom two rows (dodecahedron and icosahedron), the nets predicted to fold well are plotted with a solid circular marker and the nets predicted to fold poorly are plotted with a solid 'x'. We find that while the number of leaves is a good predictor for the dodecahedron, the icosahedron still folds with relatively low propensity. In fact, none of the icosahedron nets folded into the target structure.

selecting ones that have the largest number of degrees of freedom (see Section 2.5 for more details). Fig. 2.8 shows the combined example pathways followed by all 11 nets for the cube (Fig. 2.8a) and for the octahedron (Fig. 2.8b), illustrating the pathways that maximize degrees of freedom and using only local, native contacts to fold. We

then used these principles to find, among the 86,760 nets those with high and low folding yields (see Fig. 2.10 and 2.11 for specific nets). The corresponding folding propensities are plotted in Fig. 2.9. As expected, polyhedra with many leaves and small diameter show higher propensity for correct folding and those nets also have many high T pathways to the ground state. The correlation between the folding propensity and the leaves may have analogues in other systems as well. The number of leaves is a measure of the amount of local connections that are required to fold from the unfolded state. In the protein folding literature the contact order is a measure of how far away specific native contacts are on the amino acid sequence and it has been shown that low contact order inversely correlates to folding rate[25].

Overall, our observations suggest that folding propensity of a net decreases as the number of faces increase. For instance, while the four-sided tetrahedron folds nearly perfectly, the twenty-sided icosohedron is unable to successfully fold. One exception to this trend is the dodecahedron, which folds with higher probability than the octahedron. While the reason for this exception remains elusive, there are two factors that may play a role. The first factor is the number of degrees of freedom retained by the faces sharing a leaf vertex when the leaf edges form a bond. For instance, octahedron nets can make a bond about the leaf vertex, but due to the unique symmetry the loop, retains one degree of freedom (i.e., the resulting intermediate is not rigid), so the intermediate may still enter a trapped state. In the case of the icosahedron, after folding about leaf vertices, the loop retains two degrees of freedom. In the case of the other three shapes (i.e., tetrahedron, cube, dodecahedron), folding about leaf vertices render the loop rigid. This implies that further constraining the net by increasing the rigidity throughout the folding process is important to sufficiently funnel the net's energy landscape, and may boost the folding probability
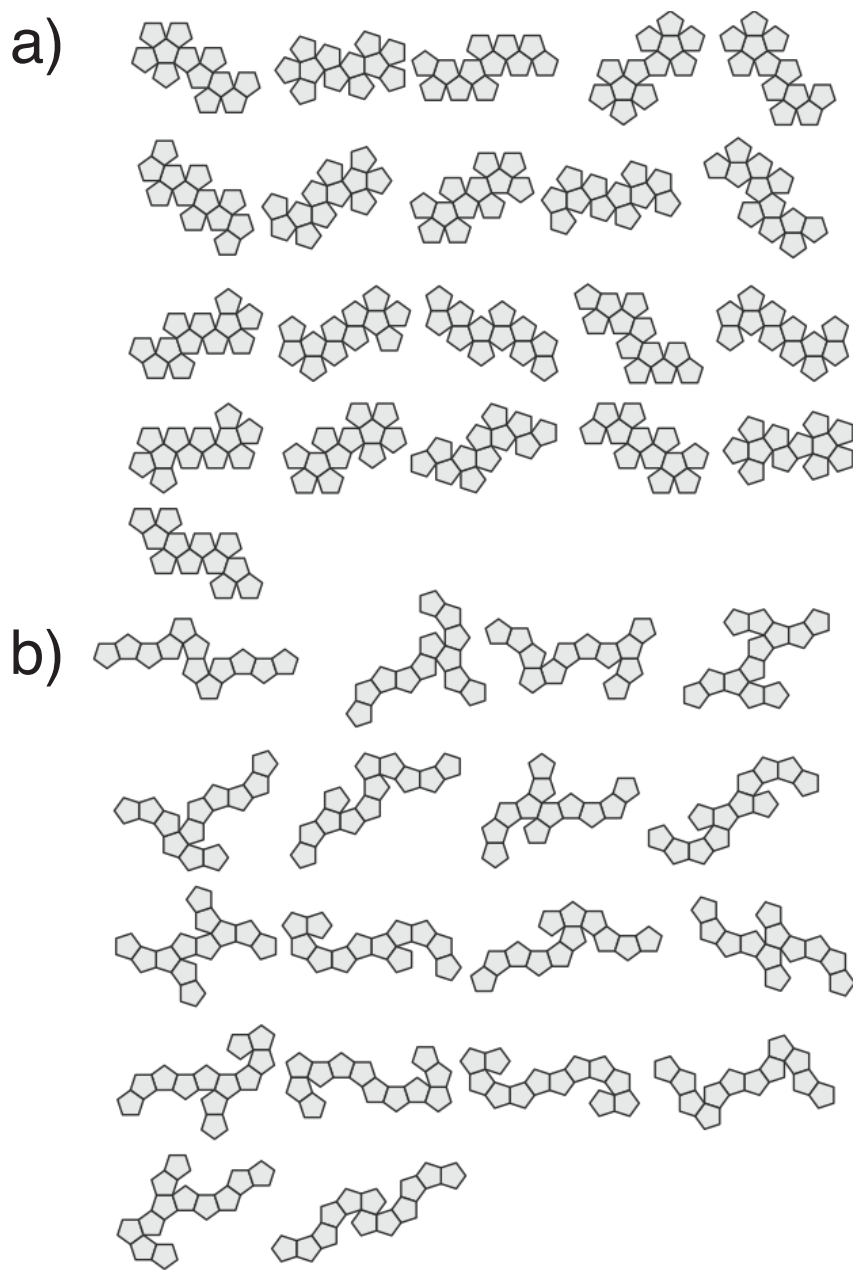
Figure 2.10: Dodecahedron nets used in this study. Dodecahedron nets predicted to be (a) "good" folders and (b) "bad" folders.

for many nets. The second factor is the complexity that arises in trapped states. There are "tetrahedral motifs" on many octahedral nets, and these motifs can fold into full or partial tetrahedra, as seen in the boat conformation mentioned above (see Fig. 2.2a for more examples). For icosahedral nets, there are both tetrahedral and octahedral motifs, and the diversity of the trapped states is further increased
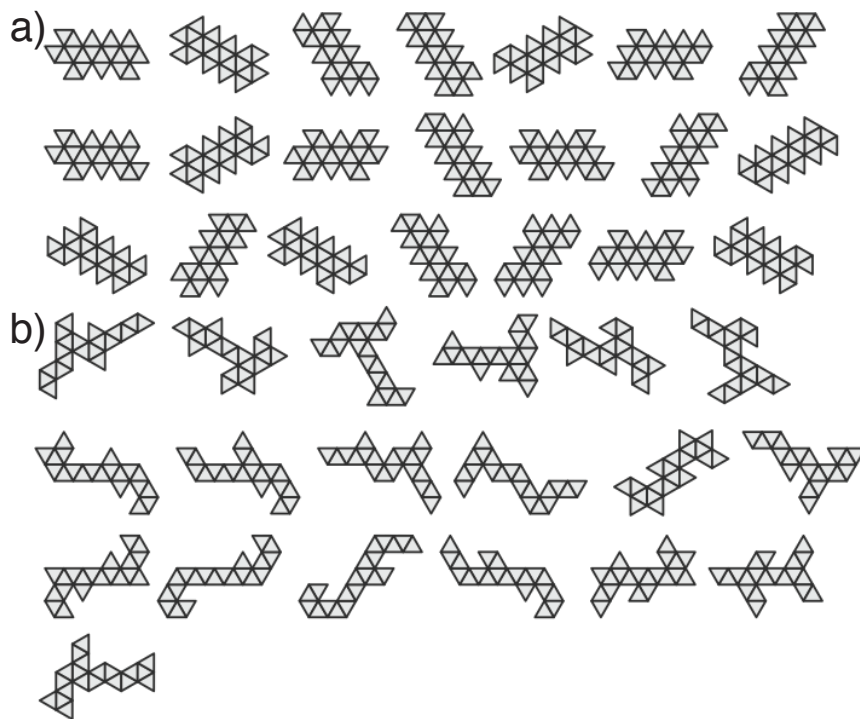
Figure 2.11: Icosahedron nets used in this study. Icosahedron nets predicted to be (a) "good" folders and (b) "bad" folders.

for these nets. This is in sharp contrast to the nets of the other three shapes, for which trapped states typically occur when one face folds on to another face. The contribution of both of these factors to folding should be investigated further in future studies.

2.4   Discussion and Conclusion

The observed preference for native contact pathways at high temperature is not unique to polyhedron nets. Several small proteins have been observed in simulation to fold via native-only contacts when close to their melting temperature[26]. At low T (or high hydrophobicity) the pathways shift to a hydrophobic collapse, in which non-native contacts form followed by further rearrangements leading to the native state [27]. Similarly, colloids assemble via monomeric pathways at high temperature,

forming bonds that are compatible with the overall structure in a equivalent process to native-contact formation[24]. Finally, systems of colloidal sticky spheres prefer to form the same concave, boat-like conformation that we observe for octahedron nets [20].

Our simple model therefore draws connections between the macroscopic irreversible folding of polyhedra [8], assembly of patchy [24] and colloidal particles [20], and the folding of amino acids[28, 26, 27]. The identified trade-off between entropy and enthalpy that dictates high temperature folding provides guiding principles for the assembly of 3-D complex geometries from potentially simpler-to-fabricate 2-D nets. We demonstrated the judicious pathway engineering via the selection of nets with certain characteristics. We found that nets at high temperature fold through pathways that maximize the internal degrees of freedom, regardless of their propensity to fold, and the more compact nets fold with higher propensity. The compactness measures correlated with the number of pathways connecting the unfolded and folded states offering some understanding on why these measures work well. In addition to giving insights into the thermodynamics of folding in naturally occurring systems, our results could also provide a route for the fabrication of anisotropic Brownian shells, paving the way for the self-assembly of complex crystals from nano and colloidal shells [29, 30, 31, 32] capable of encapsulating cargo. We expect these results to impact future experiments on folding of graphene sheets [13], graphene oxide layers [12], or DNA-origami polyhedral nets [33].

2.5   Methods

Enumeration of Polyhedral Nets

We create nets from polyhedra via a process known as edge-unfolding. In edge-unfolding, one cuts along a set of pre-chosen edges, called the cutting tree, of a polyhedron (e.g. a cube) to create a single, contiguous flat 2D sheet of connected (square) faces: a net. We enumerate all of the nets of each polyhedron by generating random weights for the edges of the skeleton-1 graph of the polyhedron on the interval [0,1]. The minimal spanning tree was found using Kruskal's algorithm [34]. We then converted the spanning tree to a net and added it to the database[19] if it did not already exist. We ran this loop for many iterations until we found all the nets for each shape.

Langevin Dynamics and Molecular Dynamics Simulation

Langevin dynamics are employed to model the folding dynamics for each net using HOOMD-Blue [35, 36, 37, 38]. Each face of the net is approximated by a union of spheres acting as a rigid body, with an edge length of 10 spheres. The spheres are arranged in a hexagonal lattice for triangular faces, square lattice for square faces, and a hexagonal approximate for the pentagonal faces. For each simulation the drag coefficient, γ, was set to the inverse of the number of spheres used to create a facet. The spheres in the center of the face interact via a WCA potential shown in blue in Fig. 2.1, while the spheres on the free edges of the net interact via a Lennard-Jones potential; both potentials used ϵ and σ values of 1.0. The rigid facets are tethered together using harmonic springs along the edges using a spring constant of 800 and and equilibrium distance of 1.0.

Quantifying Yield

To quantify the folding yield we ran 125 simulations starting from a high temperature and quenched the temperature to near zero ($T = 0.1$). We then defined the yield as the fraction of simulations that completely folded into the target polyhedron. To distinguish between nets that had very low probability, we defined the folding propensity as the average fraction of native contacts averaged over all of the quenching simulations we ran. If all nets folded perfectly in all runs the folding propensity is one. We linearly quenched 125 systems from $0.1 \leq T \leq T_m + 2.5$, where T_m is the folding temperature defined as the maximum melting temperature (temperature at which the net is unfolded 50% of the time) among all nets for a given target shape. We used two different cooling rates to investigate their influence in the folding yield: $2.5 \times 10^{-6} T/t$ and $2.5 \times 10^{-8} T/t$. The T_m for each shape is listed in Table 2.1.

Markov State Models and Folding Pathway Calculations

Markov State Models (MSM) [21] have been used to study protein folding [22, 23] and virus capsid assembly [39] and can provide a detailed view into the dynamics and thermodynamics of the folding landscape. Each simulation snapshot is classified as a discrete state, and the number of transitions between each state is recorded in a matrix. The dihedral angles completely specify the configuration of a net and are therefore a good set of collective variables. We break the 'up' / 'down' folding degeneracy by keeping track of the dihedral angle on the interval $[0, 2\pi]$. As the simulation is running we also compute the energy between each pair of free edges (edges not part of a hinge) on the net. If the potential energy between two edges is less than $E_{bond} = -5\epsilon$, then the edges are considered to be bonded, and a list of bonded edges is recorded along with the the list of dihedral angles. All intermediates are then

clustered using DBSCAN [40] with a variation of the Manhattan metric, $d(i,j) = \min\limits_{\alpha \in \mathrm{Aut}(G_{\mathrm{net}})} \max |a_{\alpha(i),k} - a_{j,k}|$, which is a way to compare the dihedral angles of different states taking into account the symmetry of the net. Using the bonding information, we turn the MD model into a graph, G_{int}. By looking at the graph automorphisms we can determine if the symmetry could lead to a relabeling of the vertices of the net. If so, clusters returned by the DBSCAN algorithm were merged.

To build the MSM we ran 125 independent NVT simulations for 12.5×10^6 steps and then branched the trajectories using new random seeds. This process was repeated until we obtained a total of 1875 (14×125) trajectories (or, equivalently, 2.3×10^{10} time steps) for each net and each temperature. Bonded edges and dihedral angles for each hinge were computed every 10 time steps and combined to define a state in the MSM described above. The lag time, τ, was found by the standard protocol of identifying the time at which the eigenvalues of the transition probability matrix become constant (Fig. 2.12). We used Transition Path Theory [23, 41, 42] to determine the reactive flux, f_{ij}, of all intermediates. The reactive flux is defined as $f_{ij} = q_j^+ \pi_i P_{ij} q_i^-$, where q_j^+ is the forward committor probability (the probability that the net will fold from state j), π_i is the probability of being in state i, P_{ij} is the probability of transitioning to state j given the system is in state i, q_i^- is the backwards committor probability (the probability that the net is folding, as opposed to unfolding). Finally, the net flux is defined as $f_{ij}^+ = \max\{f_{ij} - f_{ji}, 0\}$. The dominant paths were computed via the "bottle neck" algorithm using the net fluxes [41, 23]. The total reactive flux is defined as the sum of the reactive flux out of the unfolded state, $F = \sum_i f_{ui}$, where u is the unfolded state. The folding rate is then $k_{fold} = F/\tau \pi_r$, where π_r is the probability that the pathway is moving forward.
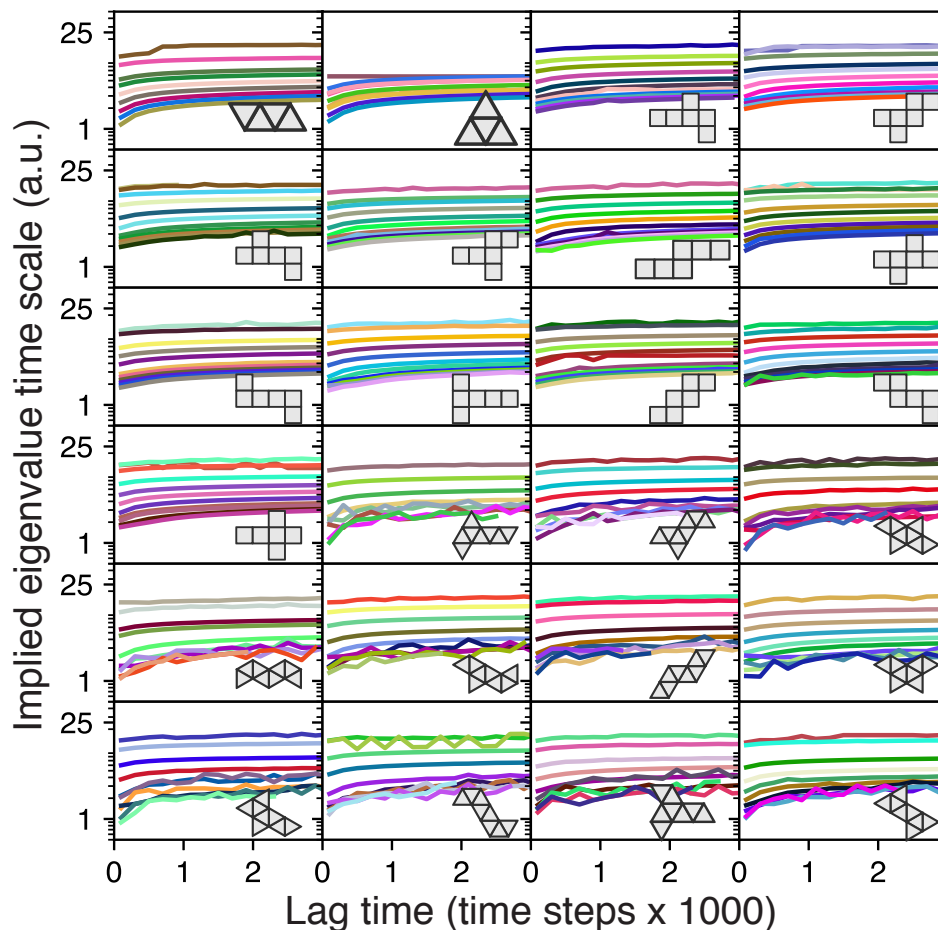
Figure 2.12: Choice of lag time for MSM studies. For each tetrahedron, cube, and octahedron net (shown in the inset), the implied time scales are shown. Each curve shows the first non-trivial eigenvalue, λ_2, of the transition probability matrix as a function of the lag time, τ for a given temperature. All of the eigenvalues flatten out by $\tau = 1000$ time steps. We therefore choose this value as the lag time for each MSM study.

Folding Parameters

The number of native contacts, Q, were calculated by counting the number of edges that were bonded to the correct corresponding edge according to the criteria described above. The non-native contacts were calculated similarly. The diameter is the graph diameter of the face graph of the net. In general the number of degrees of freedom can be difficult to calculate because it can be difficult to deduce which constraints are redundant in the net. In general one can use the pebble game [43],

however if the linkage is not generic or has point group symmetries the pebble game can underestimate the number of degrees of freedom [44, 45]. First we applied the pebble game to each intermediate, then for closed loop motifs we applied a closed loop formula to determine the number of degrees of freedom [46], finally intermediates with high degrees of symmetry were checked by hand, since the pebble game is known to underestimate these cases [44, 45].

Enumerating High Temperature Pathways

An exhaustive search was performed to enumerate high T pathways. Two principles were assumed to be important for the folding pathways at high temperature: local bonds and maximizing number of degrees of freedom. We initialize the algorithm by adding the unfolded state to the queue and creating an empty graph that will contain the pathway information. For each intermediate in the queue, a set of candidate bonds was calculated by finding edges on the intermediate that still needed to be bonded and had a topological distance of one (local). This intermediate was then added to the queue for further processing, and a link between the current state and candidate state is made in the graph. Finally the pathways are taken from the graph and sorted lexicographically by the sequence of degrees of freedom of each intermediate along the pathway. The pathways that have the largest number of degrees of freedom are then returned.

References for Chapter II

[1] G. C. Shephard. Convex polytopes with convex nets. *Mathematical Proceedings of the Cambridge Philosophical Society*, 78(October):389, 1975. ISSN 0305-0041. doi: 10.1017/S0305004100051860.

[2] Wolfram Schlickenrieder. *Nets of Polyhedra*. PhD thesis, Berlin, Technische Universitat, 1997.

[3] Erik D Demaine and Joseph O'Rourke. *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*. Cambridge University Press, New York, NY, USA, 2007. ISBN 0521857570.

[4] Rohan Fernandes and DH Gracias. Self-folding polymeric containers for encapsulation and delivery of drugs. *Advanced drug delivery reviews*, 64(14):1–11, nov 2012. ISSN 1872-8294. doi: 10.1016/j.addr.2012.02.012.

[5] S. Felton, M. Tolley, E. Demaine, D. Rus, and R. Wood. A method for building self-folding machines. *Science*, 345(6197):644–646, 2014. ISSN 0036-8075. doi: 10.1126/science.1252610.

[6] Ying Liu, Julie K. Boyles, Jan Genzer, and Michael D. Dickey. Self-folding of polymer sheets using local light absorption. *Soft Matter*, 8(6):1764, 2012. ISSN 1744-683X. doi: 10.1039/c1sm06564e.

[7] Tae Soup Shim, Shin-Hyun Kim, Chul-Joon Heo, Hwan Chul Jeon, and Seung-Man Yang. Controlled Origami Folding of Hydrogel Bilayers with Sustained Reversibility for Robust Microcarriers. *Angewandte Chemie*, 124(6):1449–1452, feb 2012. ISSN 00448249. doi: 10.1002/ange.201106723.

[8] Shivendra Pandey and Margaret Ewing. Algorithmic design of self-folding polyhedra. *Proceedings of the National Academy of Sciences*, 108(50):19885–90, dec 2011. ISSN 1091-6490. doi: 10.1073/pnas.1110857108.

[9] Kaori Kuribayashi-Shigetomi, Hiroaki Onoe, and Shoji Takeuchi. Cell origami: self-folding of three-dimensional cell-laden microstructures driven by cell traction

force. *PloS one*, 7(12):e51085, jan 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0051085.

[10] VB Shenoy and DH Gracias. Self-folding thin-film materials: From nanopolyhedra to graphene origami. *MRS bulletin*, 37(09):847–854, sep 2012. ISSN 0883-7694. doi: 10.1557/mrs.2012.184.

[11] Ken A Dill, S B Ozkan, T R Weikl, J D Chodera, and V A Voelz. The protein folding problem: when will it be solved? *Current Opinion in Structural Biology*, 17(3):342–6, 2007. ISSN 0959-440X. doi: 10.1016/j.sbi.2007.06.001.

[12] Terry C Shyu, Pablo F Damasceno, Paul M Dodd, Aaron Lamoureux, Lizhi Xu, Matthew Shlian, Max Shtein, Sharon C Glotzer, and Nicholas A Kotov. A kirigami approach to engineering elasticity in nanocomposites through patterned defects. *Nature Materials*, advance on, jun 2015. ISSN 1476-4660.

[13] Melina K. Blees, Arthur W. Barnard, Peter a. Rose, Samantha P. Roberts, Kathryn L. McGill, Pinshane Y. Huang, Alexander R. Ruyack, Joshua W. Kevek, Bryce Kobrin, David a. Muller, and Paul L. McEuen. Graphene kirigami. *Nature*, 2015. ISSN 0028-0836. doi: 10.1038/nature14588.

[14] Daniel M. Sussman, Yigil Cho, Toen Castle, Xingting Gong, Euiyeon Jung, Shu Yang, and Randall D. Kamien. Algorithmic lattice kirigami: A route to pluripotent materials. *Proceedings of the National Academy of Sciences of the United States of America*, 112(24):7449–7453, 2015. ISSN 0027-8424. doi: 10.1073/pnas.1506048112.

[15] Anum Azam, Timothy G Leong, Aasiyeh M Zarafshar, and David H Gracias. Compactness determines the success of cube and octahedron self-assembly. *PloS one*, 4(2):e4451, jan 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0004451.

[16] Ryan Kaplan, Joseph Klobušický, and Shivendra Pandey. Building Polyhedra by Self-Assembly: Theory and Experiment. *Artificial Life*, 31:1–31, 2014. doi: 10.1162/ARTL.

[17] Francis Buekenhout and Monique Parker. The number of nets of the regular convex polytopes in dimension? 4. *Discrete mathematics*, 186(1-3):69–94, 1998.

[18] NAM Araújo, RA da Costa, SN Dorogovtsev, and JFF Mendes. Finding the optimal nets for self-folding kirigami. *arXiv*, 2017.

[19] https://the_real_pdodd@bitbucket.org/the_real_pdodd/polyhedra_nets.git.

[20] Guangnan Meng, Natalie Arkus, Michael P. Brenner, and Vinothan N. Manoharan. The free-energy landscape of clusters of attractive hard spheres. *Science*, 327(5965):560–563, 2010. ISSN 0036-8075. doi: 10.1126/science.1181263.

[21] William C. Swope, Jed W. Pitera, and Frank Suits. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory . *The Journal of Physical Chemistry B*, 108(21):6571–6581, may 2004. ISSN 1520-6106. doi: 10.1021/jp037421y.

[22] Gregory R Bowman and Vijay S Pande. Protein folded states are kinetic hubs. *Proceedings of the National Academy of Sciences*, 107(38):16749–16749, aug 2010. ISSN 0027-8424. doi: 10.1073/pnas.1012332107.

[23] Frank Noé, Christof Schütte, Eric Vanden-Eijnden, Lothar Reich, and Thomas R Weikl. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45):19011–6, nov 2009. ISSN 1091-6490. doi: 10.1073/pnas.0905466106.

[24] Andrew W Long and Andrew L Ferguson. Nonlinear Machine Learning of Patchy Colloid Self-Assembly Pathways and Mechanisms. *The Journal of Physical Chemistry. B*, apr 2014. ISSN 1520-5207. doi: 10.1021/jp500350b.

[25] Richard Bonneau, Ingo Ruczinski, Jerry Tsai, and David Baker. Contact order and ab initio protein structure prediction. 11(8):1937–1944, aug 2002. ISSN 0961-8368 (Print).

[26] Robert B Best, Gerhard Hummer, and William a Eaton. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44):17874–9, 2013. ISSN 1091-6490. doi: 10.1073/pnas.1311599110.

[27] Nicholas D. Socci, José Nelson Onuchic, and Peter G. Wolynes. Protein folding mechanisms and the multidimensional folding funnel. *Proteins: Structure, Function and Genetics*, 32(2):136–158, 1998. ISSN 08873585. doi: 10.1002/(SICI)1097-0134(19980801)32:2¡136::AID-PROT2¿3.0.CO;2-J.

[28] Ken A Dill, Klaus M Fiebig, and Hue Sun Chan. Cooperativity in protein-folding kinetics. *Proceedings of the National Academy of Sciences*, 90(5):1942–1946, 1993.

[29] Sharon C Glotzer and Michael J Solomon. Anisotropy of building blocks and their assembly into complex structures. *Nature Materials*, 6(8):557–562, 2007.

[30] Umang Agarwal and Fernando A Escobedo. Mesophase behaviour of polyhedral particles. *Nature Materials*, 10(3):230–235, 2011.

[31] Joost de Graaf and Liberato Manna. A roadmap for the assembly of polyhedral particles. *Science*, 337(6093):417–418, 2012.

[32] Pablo F Damasceno, Michael Engel, and Sharon C Glotzer. Predictive self-assembly of polyhedra into complex structures. *Science (New York, N.Y.)*, 337 (6093):453–7, jul 2012. ISSN 1095-9203. doi: 10.1126/science.1220869.

[33] Fei Zhang, Shuoxing Jiang, Siyu Wu, Yulin Li, Chengde Mao, Yan Liu, and Hao Yan. Complex wireframe dna origami nanostructures with multi-arm junction vertices. *Nat Nano*, 10(9):779–784, 09 2015.

[34] Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956. ISSN 00029939, 10886826.

[35] HOOMD-blue web page:. *http://codeblue.umich.edu/hoomd-blue*.

[36] Joshua a. Anderson, Chris D. Lorenz, and a. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics*, 227(10):5342–5359, 2008. ISSN 00219991. doi: 10.1016/j.jcp.2008.01.047.

[37] Trung Dac Nguyen, Carolyn L. Phillips, Joshua a. Anderson, and Sharon C. Glotzer. Rigid body constraints realized in massively-parallel molecular dynamics on graphics processing units. *Computer Physics Communications*, 182(11): 2307–2313, 2011. ISSN 00104655. doi: 10.1016/j.cpc.2011.06.005.

[38] Jens Glaser, Trung Dac Nguyen, Joshua A. Anderson, Pak Lui, Filippo Spiga, Jaime A. Millan, David C. Morse, and Sharon C. Glotzer. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications*, 192:97–107, 2015. ISSN 00104655. doi: 10.1016/j.cpc.2015.02.028.

[39] Matthew R. Perkett and Michael F. Hagan. Using markov state models to study self-assembly. *The Journal of Chemical Physics*, 140(21):214101, 2014. doi: 10.1063/1.4878494.

[40] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231, 1996.

[41] Philipp Metzner, C Schütte, and E Vanden-Eijnden. Transition path theory for Markov jump processes. *Multiscale Modeling & Simulation*, 7(3):1192–1219, 2009.

[42] Weinan E and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual review of physical chemistry*, 61:391–420, jan 2010. ISSN 1545-1593. doi: 10.1146/annurev.physchem.040808.090412.

[43] Audrey Lee and Ileana Streinu. Pebble game algorithms and sparse graphs. *Discrete Mathematics*, 308(8):1425–1437, 2008. ISSN 0012365X. doi: 10.1016/j.disc.2007.07.104.

[44] Bernd Schulze and Shin ichi Tanigawa. Linking rigid bodies symmetrically. *European Journal of Combinatorics*, 42(Supplement C):145 – 166, 2014. ISSN 0195-6698. doi: https://doi.org/10.1016/j.ejc.2014.06.002.

[45] Bernd Schulze, Adnan Sljoka, and Walter Whiteley. How does symmetry impact the flexibility of proteins? *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 372(2008):20120041, 02 2014. doi: 10.1098/rsta.2012.0041.

[46] Tiong-Seng Tay. Rigidity of multi-graphs. I. Linking rigid bodies in n-space. *Journal of Combinatorial Theory, Series B*, 36(1):95–112, 1984. ISSN 00958956. doi: 10.1016/0095-8956(84)90016-9.

CHAPTER III

Monte Carlo Sampling Methods for Extended Ensembles

This chapter is a compilation of collaborative works[1, 2, 3]. One of my main roles in each collaboration that I was a part of was to develop new methods, and to help shape the fundamental questions that each study was asking. After developing the Monte Carlo method presented in Section 3.2 and Section 3.3, there was a surge of new problems that my colleagues and I were eager to solve. The first part of this chapter presents a review of these findings, while the second part includes the details of the theory behind the method (Section 3.2), notes on detailed balance for various systems(Section 3.2), and implementation and code design(Section 3.4).

3.1 Background, Motivation and Review

As computers become more powerful and information becomes more freely available, scientists have leveraged new computational tools to design novel materials. The bulk of the effort thus far has been to design new metal alloys that will conform to desired crystal structures, mechanical properties, and optical properties. These results are then cataloged into large databases that can be mined for future studies and applications[4, 5]. The ultimate purpose of these databases is to become giant lookup tables, where the inverse design problem is reduced to a query into the

database. However, this sort of approach is not possible if the desired material is not already catalogued in the database. Furthermore, the challenge for soft matter materials is much more difficult, and exhaustive databases may be impossible to build. When it comes to colloids, polymers, and nanoparticles, the design space is much larger and more complex compared to that of other materials. In addition to the chemical composition of the constituents, one must consider particle shape. Such parameters are said to be dimensions of anisotropy[6]. These dimensions can be hard to quantify, and cannot be exhaustively computed. In addition to the inherent high dimensionality of the design space, many phase diagrams have large regions in the parameter space that are uninteresting (i.e., phases that are undesired), and comparatively few small regions that are interesting. This class of problem requires a much more efficient method to find a solution.

3.2   Digital Alchemy Method [1]

The dimensions of anisotropy, such as shape, surface coverage, and patchiness, are a general constant in most experimental systems. In the digital alchemy model, the anisotropy dimensions are treated as thermodynamic variables and the free energy of the system in this ensemble is minimized to find the equilibrium particle shape for a given colloidal crystal at a given packing fraction. To derive the model, we start with the general Hamiltonian of the system,

$$(3.1) \qquad H(\{\alpha_i\}) = \frac{p^2}{2m} + \frac{1}{2}L^T I^{-1}_{\{\alpha_i\}}L + U_{\{\alpha_i\}}(q, Q).$$

The anisotropy dimensions are denoted by α_i, p is the linear momentum, m is the mass of the particles, L is the rotational momentum, $I_{\{\alpha_i\}}$ is the moment of inertia

---

[1]This work was done in collaboration with Greg van Anders, Daphne Klotsa, Andrew Karas, and Sharon Glotzer and is published in reference [1]

tensor, $U_{\{\alpha_i\}}$ is the potential energy, and q and Q are the positions and orientations of the particles, respectively. The anisotropy dimensions affect the system through both the moment of inertia tensor and the potential energy. We also assume that the anisotropy dimensions are independent of the position and orientation of the particles. The entropy of the system can then be written as

(3.2)
$$S = -k_B \sum_\sigma \pi_\sigma \log \pi_\sigma,$$

where π_σ is the probability that the system is in microstate σ. The equilibrium distribution over the phase space can be found by maximizing the entropy with respect to the distribution. If the anisotropy parameters fluctuate about some mean quantity, $\langle \alpha_i \rangle$, the entropy can be maximized by

(3.3)
$$S = -k_B \sum_\sigma \left[\pi_\sigma \log \pi_\sigma - \beta(\pi_\sigma H - \langle E \rangle - \sum_i \mu_i N(\pi_\sigma \alpha_i - \langle \alpha_i \rangle)) \right],$$

where β and μ_i are Lagrangian multipliers. Thermodynamically, $\beta = 1/k_B T$, and μ_i is the conjugate variable to the anisotropy dimension α_i. The equilibrium distribution is then derived as

(3.4)
$$\pi_\sigma = \frac{1}{Z} e^{-\beta(H - \sum_i \mu_i N \alpha_i)},$$

where Z is the partition function (i.e., normalization constant),

(3.5)
$$Z = \sum_\sigma e^{-\beta(H - \sum_i \mu_i N \alpha_i)}.$$

The complexity of the system prohibits us from solving this problem analytically, and so the sums must be evaluated numerically. These problems are well-suited to be solved using Monte Carlo techniques.

3.3 Detailed Balance for Various Hard Particle Systems

When employing digital alchemy for inverse materials design, one must choose how to encode the target structure to bias the sampling towards shapes that will stabilize the given structure. In this work, we are primarily interested in identifying particles that will assemble into desired crystal structures (e.g., fcc, bcc, diamond, and so on). In general, an effective strategy to accomplish this assembly is to tether each particle in the simulation to a lattice site via a harmonic spring. The Hamiltonian for a system of N identical particles with this constraint becomes

$$(3.6) \qquad H(\{\alpha_i\}) = \frac{p^2}{2m} + \frac{1}{2}L^T I_{\{\alpha_i\}}^{-1} L + U_{\{\alpha_i\}}(q, Q) + E_x(q),$$

where E_x represents an external field that holds each particle to a respective lattice site, and is independent of the shape of the particles. Hard particles are modeled by the following pairwise potential,

$$(3.7) \qquad U_{\{\alpha\}}(q_i, Q_i, q_j, Q_j) = \begin{cases} \infty & \text{particle } i \text{ overlaps with particle } j \\ 0 & \text{otherwise} \end{cases}$$

Substituting Eq. 3.6 and Eq. 3.7 into Eq. 3.5, we obtain (now using continuous coordinates),

$$(3.8) \qquad Z = \int d\alpha \; dp \; dL \; dq \; dQ e^{-\beta(\frac{p^2}{2m} + \frac{1}{2}L^T I_{\{\alpha_i\}}^{-1} L + U_{\{\alpha_i\}}(q,Q) + E_x(q) - \sum_i \mu_i N\alpha_i)}$$

and performing the quadratic integrals over p and L we can simplify,

$$(3.9) \qquad Z \propto \int d\alpha \det(I_{\{\alpha\}})^{N/2} \; dq \; dQ \; e^{-\beta(U_{\{\alpha_i\}}(q,Q) + E_x(q) - \mu N\alpha)}.$$

The proportionality symbol is used as irrelevant multiplicative constants are dropped from the equation. If we implement moves in shape space that hold the center of

mass and orientation of each particle constant, we can further simplify to

$$(3.10) \qquad Z \propto \int d\alpha \det(I_\alpha)^{N/2} \ dq \ dQ e^{-\beta(U_\alpha(q,Q)-\mu N\alpha)}.$$

If we want to guarantee that the Monte Carlo integration converges to the equilibrium distribution described by Eq. 3.10, the system must satisfy detailed balance. Detailed balance states that the probability flux into a given state must be equal to the probability flux out of the state for each neighboring state. Mathematically,

$$(3.11) \qquad P(j,i)\pi_j = P(i,j)\pi_i,$$

where $p(i,j)$ is the transition probability from state i to state j, and π is the stationary distribution of the equilibrium distribution. Equivalently,

$$(3.12) \qquad \frac{P(j,i)}{P(i,j)} = \frac{\pi_i}{\pi_j}$$

$$(3.13) \qquad = \frac{\det(I_{\alpha_i})^{N/2}e^{-\beta(U_{\alpha_i}(q,Q)-\mu N\alpha_i)}}{\det(I_{\alpha_j})^{N/2}e^{-\beta(U_{\alpha_j}(q,Q)-\mu N\alpha_j)}}$$

In this work, we always perform simulations with $\mu = 0$. The Metropolis criterion [7] then becomes

$$(3.14) \qquad P_{acc}(j,i) = \max\{1, \frac{\det(I_{\alpha_i})^{N/2}}{\det(I_{\alpha_j})^{N/2}}e^{-\beta(U_{\alpha_i}-U_{\alpha_j})}\}.$$

Shape Families

Sampling shape parameters from shape families is a relatively trivial task. Given any shape family, the truncation parameters can be mapped to a set of parameters on the interval $\alpha \in [0,1]$ to a subset of all convex polyhedra $\mathcal{C}$. Formally, we state that the shape family $F : [0,1] \times \ldots \times [0,1] \to \mathcal{C}$. If the move is limited to a maximum size, then we have $\alpha_{t+1} = \alpha_t + d\alpha$ where $|d\alpha| \leq d$. To ensure detailed balance is satisfied, we must have $p(d\alpha) = p(-d\alpha)$. If $d\alpha \in [-d,d] \times \ldots \times [-d,d]$, where each dimension

is generated independently, then given $d < \alpha$ and $d < 1 - \alpha$, the symmetry criterion is easily met, even though we generated the move on a cubic trial space.

Convex Shapes with Vertex Translations

Changing shape via different shape families is a suitable strategy to explore shape space in a very constrained way. However, there are times when one wants to break the shackles constraining these families and explore shape space with a relatively free-form move. In a move like this, the vertices of the shapes are selected at random, translated in a randomly selected direction by a randomly selected distance. Finally, the convex hull of the resulting shape is computed. Then the shape is rescaled to preserve volume, and translated to hold the center of mass fixed. Detailed balance for this move is more difficult, and more information must be tracked to ensure an accurate result. Every time we rescale the volume of the particle we must keep track of these scale factors and rescale the step size of the move. In most cases, step sizes of around 0.005 for unit volume shapes will provide a good acceptance ratios. A 2-dimensional example of how this move works is illustrated in Fig. 3.1.

Elastic Shape Deformations [2]

In all of the cases that I have described so far, the shape deformations have been performed at no energy cost. In some experimental systems, this may not be the case. For example, droplets have a deformable shape, but the energy required to make a change to that shape is proportional to the surface tension. One could model a folded protein as a shape that undergoes minor fluctuations about its native state, micelles can undergo mass transfer to change their shape, and of course cells can controllably change their shape for a wide range of reasons. We began attacking this problem

[2]This work was done in collaboration with Luis Y. Rivera-Rivera and Sharon C. Glotzer.
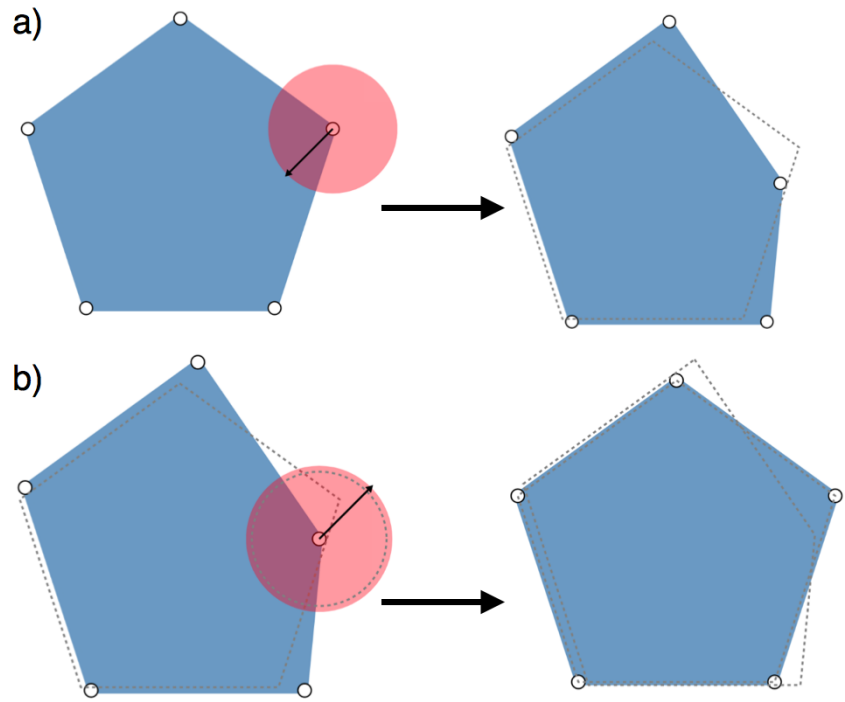
Figure 3.1: a) An example move from a regular pentagon to an irregular pentagon. In each move, a subset of the vertices are chosen (highlighted in red), and that vertex is translated (arrow). We hold the system at constant density by rescaling all vertices such that the volume of the particle remains unchanged. We also translate the particle to keep the center of mass at the origin of the shape. Finally, the volume scaling factors must be tracked so that detailed balance can be satisfied, see (b). b) The reverse of the move shown in (a). The move size must be scaled by the volume factor. The gray circle shows the original search radius, and the red circle shows the valid search radius. In this case, the search radius must be increased such that the original shape is in the trial set of all possible moves.

using the model system of hard polyhedra by asking a particular question. That is, what happens if the shape deformation about an equilibrium shape has some energy cost? Let ξ_o be the initial shape. The shape can undergo a deformation described by a second rank deformation tensor F_1. The deformed shape ξ_1 is then,

$$\xi_1 = F_1\xi_o \tag{3.15}$$

Thus, a sequence of moves in this space can be described by a product of deformation tensors,

$$\xi_n = F_n F_{n-1} \ldots F_1 \xi_o \tag{3.16}$$

$$\xi_n = \left(\prod_{k=1}^{n} F_k\right)^T \xi_o \tag{3.17}$$

$$\xi_n = \bar{F}_k \xi_o \tag{3.18}$$

We denote the $k-$th deformation by F_k, and the total deformation from the equilibrium shape after k deformations by $\bar{F}_k$. In general, the deformation tensor represents a linear mapping from one shape to another, and need not preserve volume[3]. The volume of the deformed shape is related to the reference shape by the relationship $V_1 = \det F_1 V_o$. In our system, we wanted to keep the density constant, so that volume-preserving moves could be generated by ensuring a unit determinant. Finally, the strain energy for the shape deformation is given by

$$W(\xi) = \frac{1}{2} \int S : E dV, \tag{3.19}$$

where $S = \lambda \mathrm{Tr}(E)I + 2\mu E$ is the strain tensor, and $E = (\bar{F}^T \bar{F} - 1)/2$ is the finite strain tensor. Eq. 3.19 can be simplified by the assumptions that, (1) the particles are incompressible, and (2) the material is uniform; then

$$W(\xi) = (\frac{1}{2}\lambda \mathrm{Tr}(E)^2 + \mu E : E)V \tag{3.20}$$

However, we note that the volume change $(V_o/V)^2 \approx 1 + 2\mathrm{Tr}(E) + O(E^2)$, thus if the volume is preserved we assume that $\mathrm{Tr}(E)$ is negligible. Finally, the strain energy for these moves can be calculated by

$$W(\xi) = \mu E : E. \tag{3.21}$$

---

[3]While there is nothing technically wrong with moves that do not preserve volume, when volume changes it can result in some interesting behavior. In fact, in simulations that we ran that included volume change, the particles typically deformed to smaller volume particles due to the increased entropy that the system gains from the increase in free volume.

When generating the moves (deformation tensors F_k), we must take special care to make the moves symmetrically (i.e., $p(F_k) = p(F_k^{-1})$). As we learned from the vertex move, some care must be taken when keeping volume constant, and we also want to avoid rescaling whenever possible. Since the deformations are volume-preserving $(\det(F_k) = 1)$, the set of all possible $\{F_k\}$ forms a group. Another restriction that we wanted to put on the deformation tensor was that it must be irrotational, since we wanted to make sure the shape moves would only change the shape, and not the orientation of the particles. Also, we did not want to observe artificially higher energies that were due to rotations rather than actual shape deformations. These constraints require the deformation tensor to be symmetric. Therefore, F_k is a symmetric matrix, and by the fundamental theorem of linear algebra can be decomposed as $F_k = U\Lambda U^T$. The matrix U is an orthogonal rotation matrix, and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ is a diagonal scaling matrix. Thus, $p(F_k) = p(U, \Lambda)$ if we generate the rotation independently from the scaling matrix, $p(F_k) = p(U)p(\Lambda)$. There are well-known algorithms to generate uniformly random rotations, so we only need to generate the matrix Λ. In the next section, I derive in a general way how to uniformly simulate Λ. If we restrict each $\lambda_i \in [\frac{1}{s}, s]$ for some $s > 1$, the inverse move Λ^{-1} will exist in the trial set, and since the probability density is uniform, detailed balance is guaranteed.

Simulating random numbers on 2-dimensional surfaces in 3 dimensions:
Let f_S be a probability density function defined on the surface S that is given by the equation $h(x, y, z) = k$, where k is a constant. Formally $f_S : S \to \mathbb{R}$ and the integral

taken over the surface must be equal to one,

$$(3.22) \qquad 1 \quad = \quad \iiint\limits_{S} f_S(x, y, z)\, dx\, dy\, dz$$

$$(3.23) \qquad = \quad \iiint\limits_{\substack{(x,y,x):\\ h(x,y,z)=k}} f_S(x, y, z)\, dx\, dy\, dz$$

We see that this is really a surface integral . Let us define R as the shadow region of the surface projected onto the xy-plane. Though this is not always the method we will use for integration, the method outlined here can be easily transferred to other planes. Eq. 3.23 becomes,

$$(3.24) \qquad 1 \quad = \quad \iint\limits_{S} f_S(x, y, z)\, d\sigma$$

$$(3.25) \qquad = \quad \iint\limits_{R} f_S(x, y, z)\frac{|\nabla h|}{|\nabla h \cdot \hat{n}|}\, dA,$$

where $\hat{n}$ is the unit normal to the region R. The term $\frac{|\nabla h|}{|\nabla h \cdot \hat{n}|}$ is a measure of how stretched the surface is above some point in R. We must require that $|\nabla h \cdot \hat{n}| \neq 0$, or that the surface normal is never perpendicular to the plane of integration. Eq. 3.25 tells us how to shift the domain from the surface S to the region R. Defining $f_R : R \to \mathbb{R}$, we have

$$(3.26) \qquad f_R(x, y) = \begin{cases} f_S(x, y, z(x, y))\frac{|\nabla h|}{|\nabla h \cdot \hat{n}|} & (x, y) \in R \\[2ex] 0 & \text{otherwise} \end{cases}$$

where we use the constraint $h(x, y, z) = k$ to solve for z as a function of x and y. This means that if we can sample R according to f_R, then this will produce the distribution f_S on S. Here we assume that we have a way to generate X, Y from a density function $g(x, y)$, and we can use that as a basis to simulate X, Y with density $f(x, y)$. Let c be a constant such that

$$\max \frac{f(x, y)}{g(x, y)} \leq c.$$

Each iteration will be accepted with probability $1/c$, so there will be c itera-

---
Algorithm 1 Accept-Reject Method

1: **procedure** SIMULATE($g(x,y), f(x,y), c$)▷ Simulate random variables X, Y with density $f(x,y)$
2:     **do**
3:         Simulate $X, Y \sim g(x,y)$ and $U \sim \text{uniform}(0,1)$
4:     **while** $U > \frac{f(X,Y)}{cg(X,Y)}$
5:     **return** X, Y
6: **end procedure**

---

tions on average before the procedure accepts a pair (X, Y). In this case $g(x,y)$ can be a uniform distribution on R, therefore $g(x,y) = \frac{1}{\text{Area}(R)}$. Also, $f(x,y) = f_S(x,y,z(x,y))\frac{|\nabla h|}{|\nabla h \cdot \hat{n}|}$. For simplicity we define σ and σ^* as

$$\sigma(x,y) = \frac{|\nabla h|}{|\nabla h \cdot \hat{n}|}$$

$$\sigma^* = \max \sigma(x,y)$$

Then the constant c is given by

$$(3.27) \qquad c \geq \text{Area}(R) \max f_S(x,y,z(x,y))\sigma(x,y).$$

As an example, let us assume $f_S(x,y,z)$ is a uniform distribution. Then, $f_S = \frac{1}{\text{Area}(S)}$. Eq. 3.27 becomes,

$$(3.28) \qquad c = \max \frac{\text{Area}(R)}{\text{Area}(S)}\sigma(x,y)$$

$$(3.29) \qquad = \frac{\text{Area}(R)}{\text{Area}(S)}\sigma^*$$

This means that line 4 in algorithm 1 will become

$$(3.30) \qquad U > \frac{\text{Area}(R)\sigma(X,Y)}{\text{Area}(S)\frac{\text{Area}(R)}{\text{Area}(S)}\sigma^*} = \frac{\sigma(x,y)}{\sigma^*}$$

We see that the efficiency of the algorithm depends on the most stretched part of the surface.

3.4 Implementation and Code Design

Here I briefly discuss the code design, and some of the finer points of code implementation that may prove useful to those interested in modifying the code in the future. This code was implemented as a feature in HOOMD-Blue as part of the Hard Particle Monte Carlo (HPMC) package. Every class in HOOMD belongs to one of three base classes: analyzer, updater, and compute. The analyzers are classes that can operate on the particle data without changing any of it (e.g., writing the particle positions to a file). Updaters are classes that operate on the particle data, and are also allowed to change that data (e.g., translating and rotating particles). Computes are classes that calculate forces or energy for the system. In this case, the alchemical moves change the system, and are therefore classified as updaters. HOOMD's basic design is to make different features modular. This has several advantages when creating software to serve a general purpose. One advantage is that different code paths are independent, and can therefore be tested independently. This feature makes the code much easier to maintain. This modularity also allows the HOOMD infrastructure to be very flexible. Users can easily choose the features they need for a simulation, and it should be guaranteed that each part will work correctly. As HOOMD was originally designed to be a Molecular Dynamics code, it is important to think about these fundamental assumptions and how they can change for Monte Carlo simulations.

In Fig. 3.2a, the main simulation loop is illustrated. On each time step, the analyzer classes run first, and then the updater classes run, and then a special updater called the integrator runs to advance the system one step forward. In Fig. 3.2b, the steps for the alchemical Monte Carlo method are shown. Since the analyzers are purely
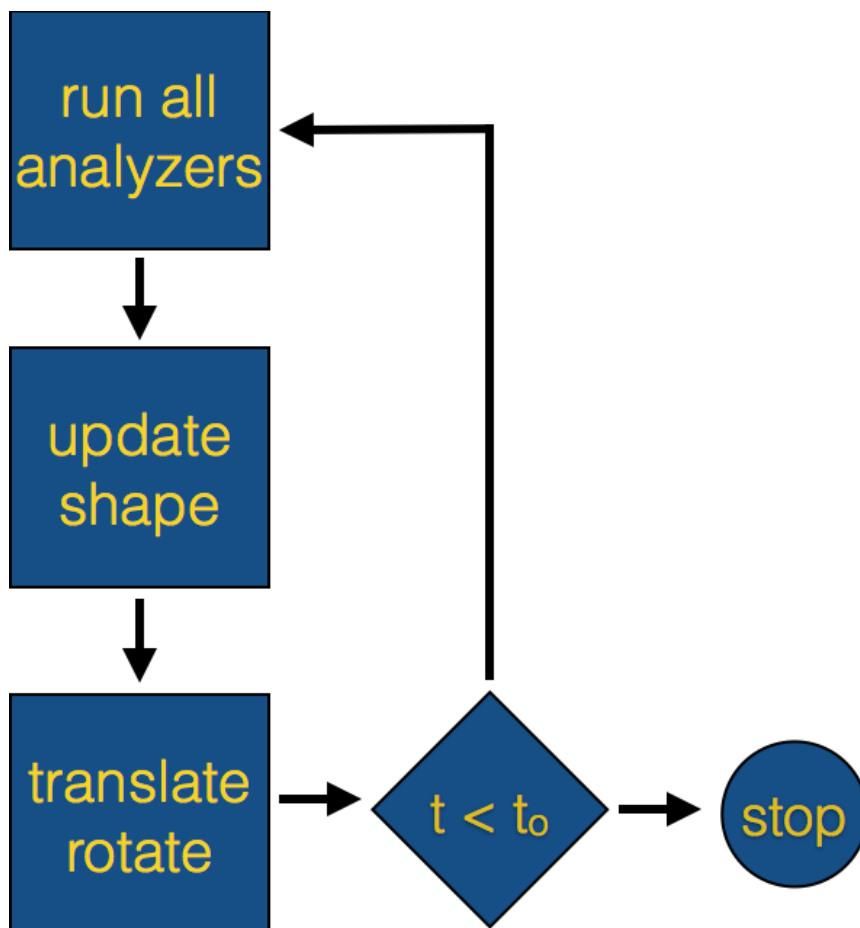
Figure 3.2: The main simulation loop of HOOMD-Blue. a) On each timestep, HOOMD-Blue executes all analyzers, updaters, and computes before integrating the system. Understanding the implementation details is fundamental to understanding the detailed balance for the system as a whole. b) In the specific case of running HPMC with shape moves, those shape moves are always executed before translations and rotations. This breaks strict detailed balance, but for carefully designed moves we can still satisfy the balance condition. Therefore, the long time behavior of the chain will converge to the Boltzmann distribution.

observers of the system, they cannot affect the system. It is important to consider how the shape updater and integrator interact. Since this is always done sequentially and in a fixed order, if we change the shape and then translate and rotate the particles, we must consider whether this can break the detailed balance described above. When considering detailed balance here, we can no longer think about the local move, but rather we must think about the state of the system at each time step. Here is a simple thought experiment that proves that this algorithm *breaks*

strict detailed balance.

Let the state of the system (e.g., shape of the particles, and their positions and orientations) be defined as X. First we make a shape update (that is accepted), and then we make a set of translations and orientations (some of which are accepted), which leaves the system in a new state $\hat{X}$. Now detailed balance is satisfied if $P(\hat{X}|X)P(X) = P(X|\hat{X})P(\hat{X})$. The transition probability can be decomposed into two parts $P(\hat{X}|X) = g(\hat{X}|X)P_{acc}(\hat{X}|X)$, where g is the probability that $\hat{X}$ is attempted given the current state X. P_{acc} is the acceptance probability given by Eq. 3.14. The shape updater will propose a move and accept it independently from translations and rotations $g(\hat{X}|X) = g_\alpha(X'|X)g_{hpmc}(\hat{X}|X')$, where g_α is the shape move probability and g_{hpmc} is the translation and rotation probability. We see that if we want to make the reverse move $(\hat{X} \to X)$, we must first make the reverse shape move. However, the expression $g(X|\hat{X}) = g_\alpha(X''|\hat{X})g_{hpmc}(X|X'')$, supposes that X'' (an original shape with new positions and orientations) now introduces some overlaps, and this probability would then be 0. Therefore we have found a pair of states such that $g(\hat{X}|X) \neq g(X|\hat{X}) = 0$. While this strict form of detailed balance is clearly not satisfied in this implementation, all is not lost. In 1999 Deem et al. [8] showed that this strict detailed balance is not required for the Markov chain to converge to the Boltzmann distribution, so long as the system satisfies the following six properties:

1. $x_{n+1} = Ax_n$

2. $A_{ij} \geq 0 \forall i, j$

3. $\sum_j A_{ij} = 1$

4. $[A_{ij}^m] > 0 \forall i, j$ and some fixed m
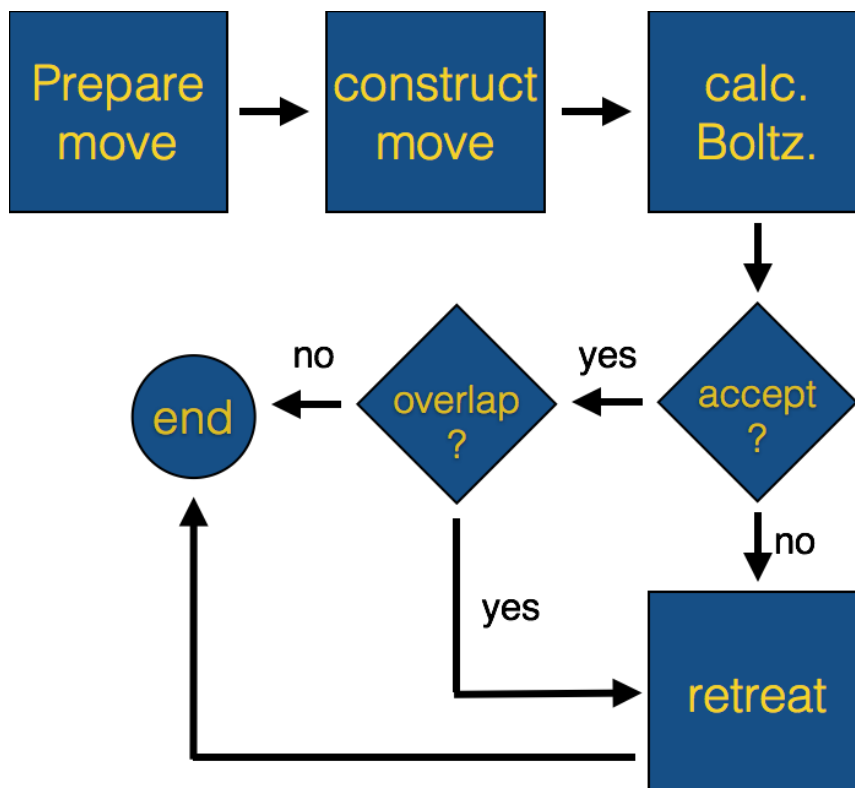
Figure 3.3: A flow diagram of the shape move algorithm. Each move must perform three operations: *prepare*, *construct*, and *retreat*.

5. $A' = (1 - \gamma)A + \gamma I, 0 < \gamma < 1$

6. $Ax^* = x^*$

Properties 1-3 follow directly from the definition of a Markov chain, and are trivial to show. Property 4 states that the chain is ergodic. Property 5 states that null moves are possible and will not affect the chain. Property 6 states that the equilibrium distribution must exist. Some typical cases where Property 6 does not hold are in cyclic chains, where certain states are only accessible on even or odd jumps.

Now that we have discussed the relationship between the alchemical updates and the core HOOMD infrastructure, we can now discuss a few more implementation details that may be interesting to those that are modifying the code or developing similar algorithms. As with most programming challenges, to write sufficiently general code

one needs to abstract the common aspects of the problems into general operations. Fig. 3.3 shows the flow diagram for the shape move algorithm. Specifically, there are three common operations that each move must implement with some guarantee of the internal state: *prepare*, *construct*, and *retreat*.

prepare: should always copy anything that is specific to the move (e.g., shape parameters, step sizes, deformation tensors).

construct: should always leave the internal state of the system as if the move is accepted.

retreat: should always perfectly restore the state of the move as if the last move never happened.

Different Boltzmann factors can be used depending on the ensemble of interest. In this way, we can have many different shape moves and acceptance criteria (Boltzmann factors), all using a single code path.

References for Chapter III

[1] Greg van Anders, Daphne Klotsa, Andrew S. Karas, Paul M. Dodd, and Sharon C. Glotzer. Digital alchemy for materials design: Colloids and beyond. *ACS Nano*, 9(10):9542–9553, 2015. doi: 10.1021/acsnano.5b04181.

[2] Rose K. Cersonsky, Greg van Anders, Paul M. Dodd, and Sharon C. Glotzer. Relevance of packing to colloidal self-assembly. *Proceedings of the National Academy of Sciences*, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1720139115.

[3] Y. Geng, G. van Anders, P. M. Dodd, J. Dshemuchadse, and S. C. Glotzer. Engineering entropy for the inverse design of colloidal crystals from hard shapes. *ArXiv*, 2017.

[4] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013. ISSN 2166532X.

[5] https://materialsdata.nist.gov/.

[6] Sharon C Glotzer and Michael J Solomon. Anisotropy of building blocks and their assembly into complex structures. *Nature Materials*, 6(8):557–562, 2007. ISSN 1476-1122.

[7] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[8] Vasilios I. Manousiouthakis and Michael W. Deem. Strict detailed balance is unnecessary in monte carlo simulation. *The Journal of Chemical Physics*, 110(6): 2753–2756, 1999. doi: 10.1063/1.477973.

CHAPTER IV

Avoiding the Glass Transition using the Alchemical Ensemble

Geometric frustration can present itself in many forms: incommensurate length scales or bond angles, stoichiometric ratios, or lattice geometry. In an age where designer particles, patchy particles, and anisotropic colloids are becoming experimentally feasible to produce, it is necessary to gain a deeper understanding between the constraints that can be experimentally tuned and the resulting bulk assembly behavior. Due to experimental advances, it is no longer necessary to treat certain attributes of a system as constants, which opens new questions about the ways in which dynamic building blocks assemble.

In recent theoretical work, van Anders et al. [1] treated these attributes as variables that fluctuate. The same approach was used to design the first hard particle that assembles into a distorted hcp crystal structure [2]. Beyond materials design, the alchemical ensemble can help to illuminate physical properties and phenomena in the corresponding non-alchemical space [3, 4].

In this work, we use the alchemical ensemble to address the following question: How do shape fluctuations of hard, convex polyhedra affect phase behavior for shapes that would not otherwise assemble into ordered structures? Here we limit the scope to
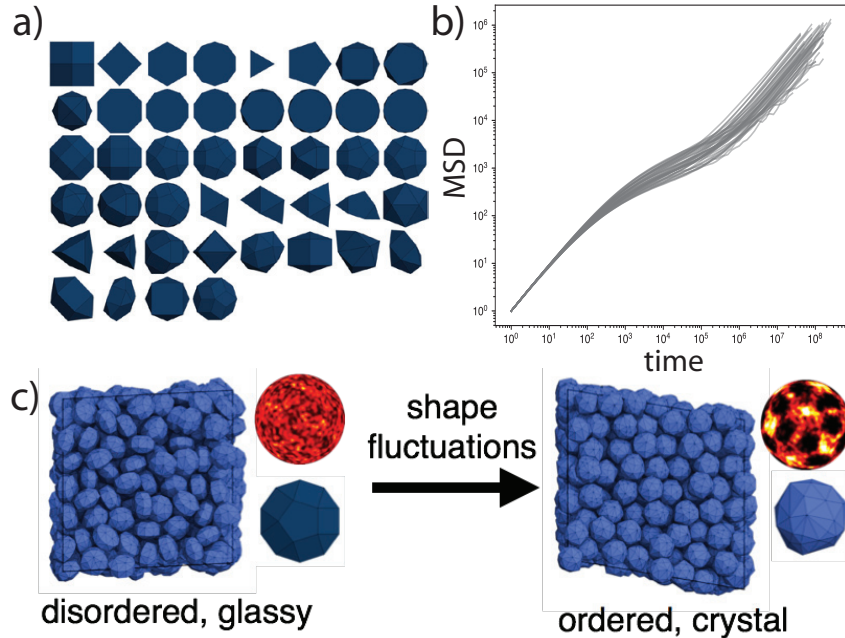
Figure 4.1: a) The 44 shapes that we study in this work. These shapes were previously shown to exhibit glassy behavior [5]. b) The mean squared displacement versus time plotted for each shape shown in (a) near 0.6 packing fraction. Each curve exhibits a plateau at intermediate times, suggesting that the dynamics slow down as the system approaches the glass transition. c) In this work, we first simulate each shape to generate a random glassy state. We then treat the shape as a thermodynamic variable, in which the vertices can move. At the end of this process, we observe that the systems assemble into a variety of structures.

44 hard particles that have been previously shown to form glasses[5]. We show via Monte Carlo simulations in the alchemical ensemble that the geometric frustration of a system is disrupted by shape fluctuations, which then allows the system to assemble into an ordered structure, thereby avoiding the glass transition.

4.1   Results

The shapes studied are shown in Fig. 4.1a. Alch-MC simulations, at a volume fraction of 0.6, produce dense fluids that exhibit a plateau in the mean squared displacement (Fig. 4.1b). This signifies that (1) caging is occurring, and (2) the dynamics slow down near the glass transition [6]. First, we simulate each of the shapes to build a

random glassy state. Next, shapes were allowed to change in a relatively unrestricted way. We performed shape updates where the vertices of the shapes were translated, the convex hull was computed, and the shapes were rescaled to maintain a fixed volume (Fig. 4.1c). The shape deformations do not have energy, aside from the rotational momentum contribution to the partition function. Both initial and final steps were run for 160 million MC sweeps.

In the alchemical ensemble, we observed that the systems underwent a disordered to ordered transition during the majority of trials for each shape. Fig. 4.2a summarizes the frequency of each type of crystal structure that we observe. Somewhat unsurprisingly, we find that fcc/hcp are the crystal structures that predominantly assemble. What is more surprising is that some complex structures, including a15 and u77, assemble for a subset of shapes. We find that independent trials that start from the same initial shape sometimes assembled into distinct crystal structures. For example, independent trials starting from initial shape J38 resulted in fcc, hcp, $\gamma-$brass, and u77 crystals(Fig. 4.2b). We found that of the 44 shapes that we studied at 0.6 packing fraction, 41 have at least one (of five) replicates that assemble into a crystal structure. For the three remaining shapes that did not assemble, we performed simulations at slightly lower packing fractions (i.e., 0.59 and 0.58). These shapes were able to assemble into a crystal structure at these lower densities. An example of each of the observed crystal structures is shown in Fig. 4.2c-h. Moreover, we observed that some systems can undergo more than one phase transition over the course of the simulation. In the case of J07 at a packing fraction of 0.58, the system initially forms a bcc crystal structure, and then proceeds to crystalize into an hcp/fcc crystal. The fact that the same shape evolves in slightly different ways, which ultimately results in the assembly of completely disparate crystal structures
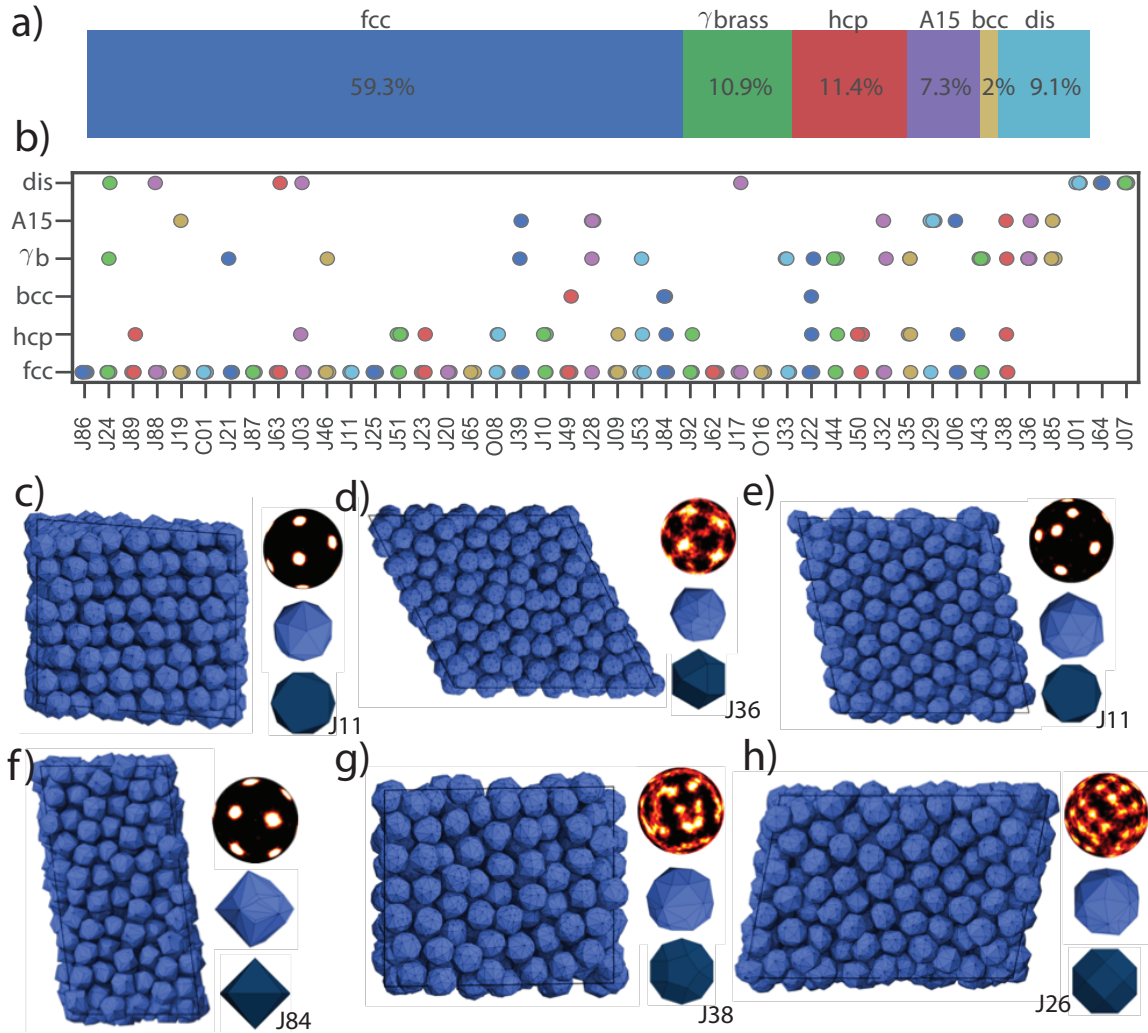
Figure 4.2: a) A percentage breakdown of all of the structures observed to assemble in this study. The majority of the trials assemble into fcc/hcp, however we also observe γ-brass, bcc, and A15/Unknown77. A small fraction of simulations were unable to assemble. There were only three shapes (J01, J64, J07) that did not exhibit any ordering at a packing fraction of 0.6, though these shapes were able to assemble at lower densities. b) A structural breakdown of each of the 44 shapes. Of the shapes that assembled, all shapes (with the exception of J36 and J85) assembled into fcc on one or more trials. c-h) Example systems for each of the crystal structures that we observe. A final snapshot of the system is shown, along with the corresponding bond order diagram, final shape, and initial shape.

leads to several open questions. Which features of a shape lead the system to assemble into one structure over another? Which structure is most stable? Are there multiple ground states in shape space?
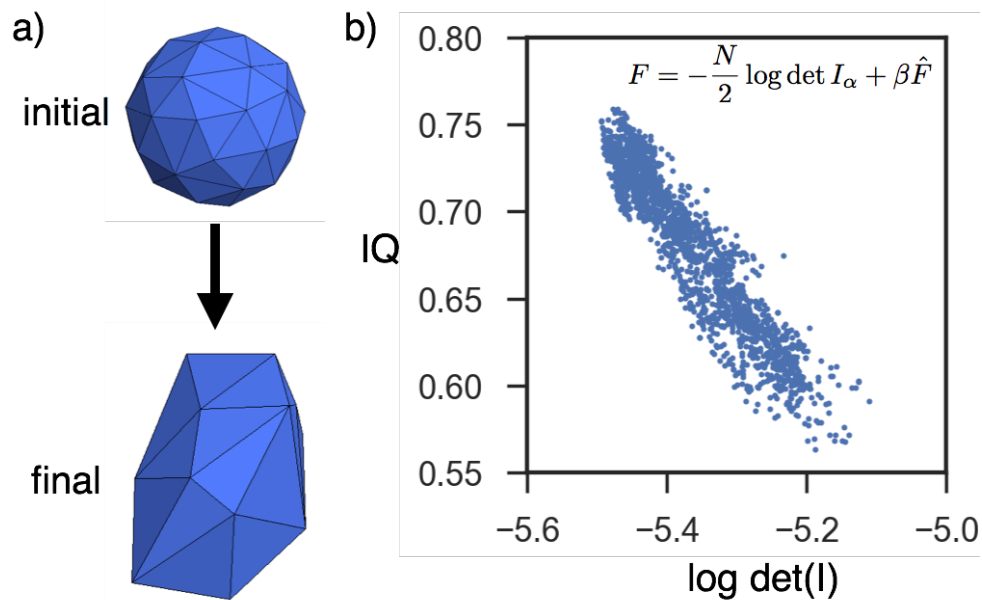
Figure 4.3: a) the initial and final particles for an Alch-MC simulation of a single particle. b) The isoperimetric quotient as a function of the logarithm of the determinant of the moment of inertia tensor. More spherical shapes tend to have a smaller moment of inertia.

Gaining a better understanding of the time evolution of shapes over the course of simulations may shed light on the ways in which systems assemble, and the reasons for structural diversity across trials of the same shape. In Fig. 4.4a, we plot the time evolution of several shape descriptors for J38, one of the shapes that showed the most structural diversity. The calculated shape descriptors include the area-weighted cosine of dihedral angles $(\cos\theta)$ for large facets, the trace of the moment of inertia tensor $(\mathrm{Tr}I)$, and the isoperimetric quotient (IQ). These parameters have been shown to be useful in predicting the crystal structures in previous studies [7, 5]. Details of their calculation can be found in Section 4.3. At the beginning of each simulation, we observe a sharp dip and then a rise in $\cos\theta$. This can be interpreted as either an increase or decrease in the dihedral angles (this parameter has a minimum at 90°). We also observe a significant amount of noise in this parameter, such that it was necessary to plot the moving average with a window size of 50 frames in order
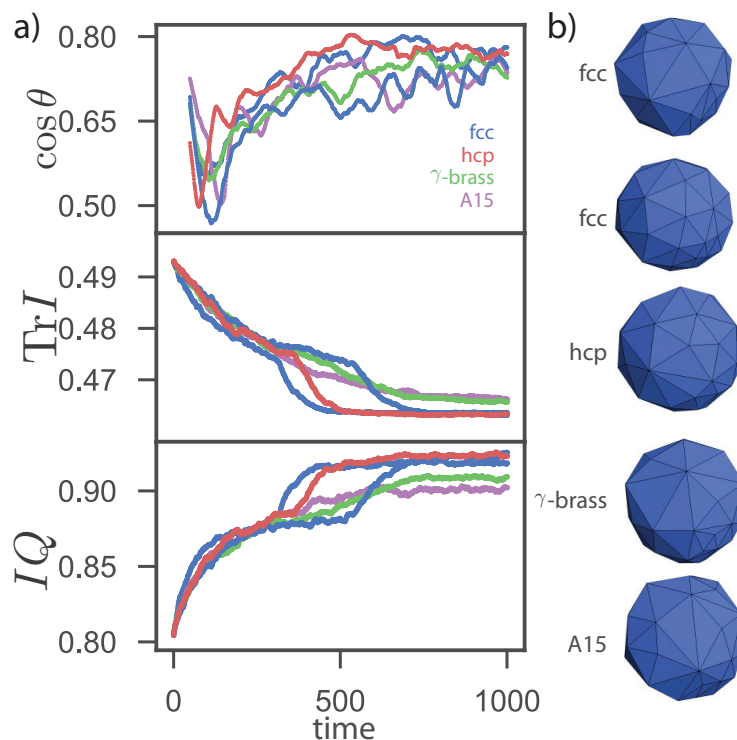
Figure 4.4: a) The time evolution of several shape descriptors are plotted for J38, which is one of the shapes that demonstrated the most structural diversity. (top) The area-weighted cosine of dihedral angles for large facets. This parameter dips sharply at the beginning of each simulation, and then increases. In general, the parameter is fairly noisy over the course of the simulation. (middle) The trace of the moment of inertia tensor. We find that this descriptor decreases throughout the simulation. (bottom) The isoperimetric quotient (IQ) increases throughout the simulation. Together, these data illustrate the widening of the dihedral angles as the shapes become more spherical and less faceted. b) The final shapes for each trial.

to visualize the data. We find that TrI descriptor decreases over the course of the simulation, while the IQ increases. We also find that the simulations that assemble into fcc and hcp exhibit a kink in these parameters at approximate times of 300, 400, and 600. The reason for these kinks remains elusive, but it is possible that they could signify that the shape "fell" into a fcc/hcp basin. We also find that shapes that formed the γ-brass and A15 exhibit increased faceting (lower IQ). Fig. 4.4b shows the final shape obtained from each simulation.

We then related the shape descriptors and the assembled crystal structures for all of the studied shapes to observe any general trends (Fig. 4.5). We did not identify any
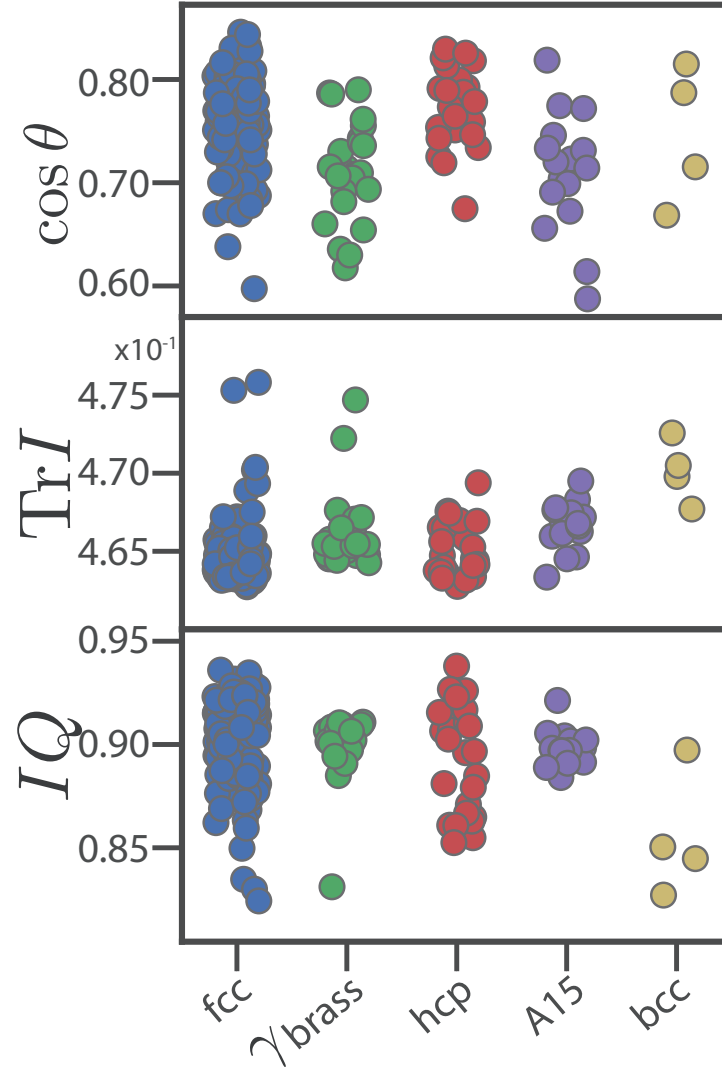
Figure 4.5: (top) The area-weighted cosine of dihedral angles. (middle) The trace of the moment of inertia tensor. (bottom) The isoperimetric quotient aggregated by the crystal structure that we observe. We applied a slight jitter to the x-axis, so that all of the data points can be visualized. We observe that of each of these crystal structures and their shape parameters are nearly identical.

single parameter that could predict the assembled crystal structure. We did observe, however, that the data for γ-brass and A15 seemed to exhibit less noise than the data for the fcc/hcp systems.

Next, we aimed to describe the evolution from the initial shape to its final shape. To accomplish this, we compared the ratio IQ_f/IQ_i, the volume of the symmet-
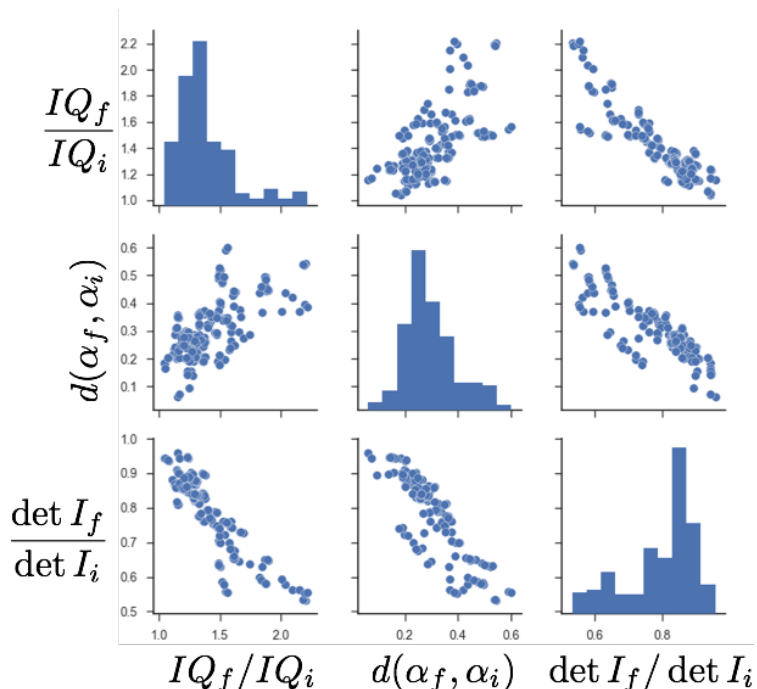
Figure 4.6: A comparison of the initial and final shapes for (a) the IQ, (b) the volume of the symmetric difference, and (c) the determinant of the moment of inertia tensors. These data illustrate that the shapes become more spherical in each trial, and that the shapes that have moved further in shape space are doing so to become more spherical.

ric difference $d(\alpha_i, \alpha_f)$, the ratio of determinant of the moment of inertia tensors $\det I_f / \det I_i$ for initial and final forms in Fig. 4.6. These data illustrate the trend for the shapes to become more spherical for each independent trial, and the shapes tend to become more spherical as they transverse a greater distance in shape space. Importantly the tendency towards increasing sphericity is not inherent in the Alch-MC algorithm. Alch-MC simulations of a single particle in a box tend towards more faceted particles as shown in Fig. 4.3.

Finally, in Fig. 4.7, we find a robust correlation between the coordination number in the ordered state and the coordination number in the glassy state. However, for particles in the glassy state with a low coordination number, we observe that this correlation breaks down. These particles are initially the least spherical. As they become increasingly more spherical, more particles are required to pack efficiently
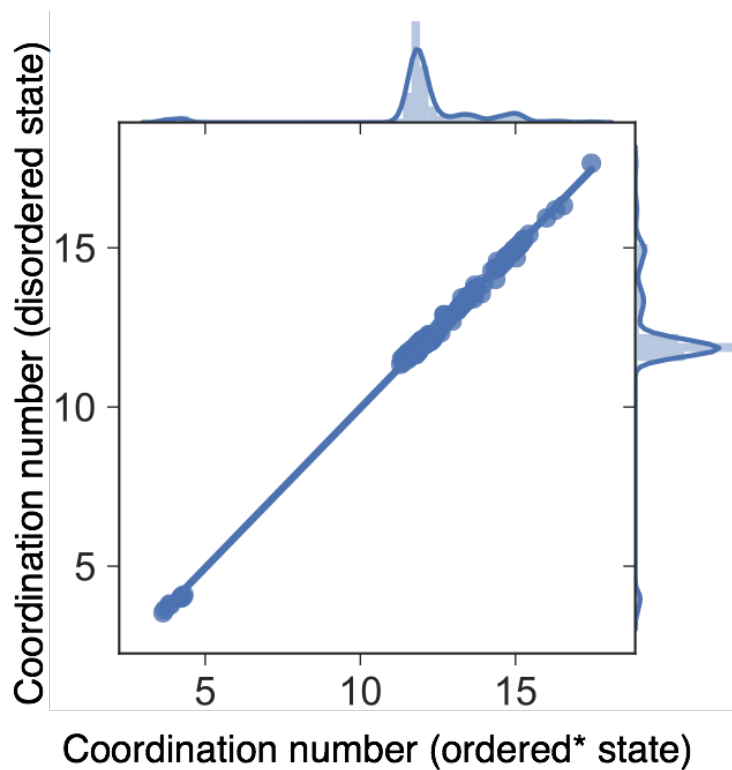
Figure 4.7: The coordination number in the ordered state is strongly positively correlated with the coordination number in the glassy state.

around a given central particle.

4.2 Discussion and Conclusions

In this work, we performed alchemical Monte Carlo simulations of hard particles that do not assemble into any known crystal structures. We found that by allowing the particle's shape to fluctuate during the simulation, the system was able to escape the kinetically-trapped, glassy states and crystallize into ordered structures. We find that fcc/hcp crystal structures are predominantly formed, and the particles become more spherical to minimize free energy. Though some outstanding questions remain, based on this work we can draw several general conclusions, and make some general predictions.

To gain an understanding of the complex underlying energy landscape, the reduction in faceting that the particle undergoes during the simulations is significant. Since these are hard particle systems, the only real driving force in the system is entropy $S = S_t + S_r + S_s$. Contributions to entropy of the system are translational (S_t), rotational (S_r), and shape-related (S_s). The system will tend toward order if at all possible, in an effort to maximize the translational entropy of each particle. Naturally, rotator crystals have a high rotational entropy, thus it is unsurprising that all of the crystals that assemble here are rotator crystals. Finally, the system tends to form structures that also allow for a wide distribution of shapes that are commensurate with a given particular crystal structure. It seems very plausible, therefore, that highly spherical particles represent the optimal set of shapes that allow for the system to maximize each of these contributions. One interpretation of these results as a whole is that faceting can be viewed as a constraint on the system, therefore this constraint must be removed in order to maximize entropy. Those systems that need faceting to assemble (e.g. diamond, simple cubic, quasicrystals) will likely be unstable in this ensemble.

Since all of the systems described here explore the same shape space, any shape could theoretically transform into another belonging to the set. The fact that we observed structural diversity begs the question of whether there exists a single global minimum in this space, or whether the more complex crystal structures are metastable to fcc/hcp.

4.3   Model and Methods

Monte Carlo Simulations

Hard Particle Monte Carlo simulations were employed to model the assembly of each shape that we studied using HOOMD-Blue [8, 9, 10]. Each simulation of 512 particles was broken into three distinct parts. First, each shape was initialized on a cubic lattice, thermalized, and then rapidly compressed to the target packing fraction. After the compression step, we ran each system for 40 million time steps (160 million MC sweeps). During this step, the aspect ratio and shear of the box was allowed to fluctuate, while holding the volume of the box constant. Finally, we ran an additional 40 million time steps where the shapes were allowed to fluctuate according to the move presented in Chapter III. All step sizes (i.e., translation, rotation, shape, and box) were tuned to give an acceptance ratio of 20%.

Shape Descriptors

We calculated five shape descriptors and metrics to quantify different aspects of each shape. Both $\text{Tr}(I)$ and $\det(I)$ are straightforward to calculate from the inertia tensor. The isoperimetric quotient (IQ) is a measure of how spherical a shape is, and is calculated by $IQ = 36\pi v^2/s^3$, where v is the volume of the particle, and s is the surface area. The IQ ratio of a sphere is unity. The cosine of the dihedral angles was calculated by first clustering the normal vectors of the facets. Each cluster was then represented by the average normal vector, which was calculated by the area-weighted average normal vector, n_i. Finally, the dihedral angles of these averaged vectors was computed, and their weighted average was calculated by the following formula,

$$(4.1) \qquad \cos\theta = \frac{\sum_{i,j}(s_i + s_j)|(n_i \cdot n_j)|}{\sum_{i,j}(s_i + s_j)},$$

where the sums are computed over all adjacent faces. We defined the distance metric $d(A, B)$ as the volume of the symmetric difference, where A and B are convex polyhedra. From set theory, it follows that $d(A, B) = V_A + V_B - 2V_{AB}$, where V_A and V_B are the volumes of A and B respectively, and V_{AB} is the volume of the intersection. Intersection polyhedra can be calculated by first identifying a point inside of the intersection. In our case, this was always taken to be the origin as the shapes were translated, so that the centers of mass always rested on the origin. Next, the dual of A and B were computed with respect to the origin and a unit sphere, denoted A^* and B^*. Finally, the intersection was computed by finding the dual of the convex hull of the union of A^* and B^*, $AB = \text{conv}(A^* \bigcup B^*)^*$. Since, all of our studied shapes have fixed unit volume, the distance metric can be simplified to $d(A, B) = 2(1 - V_{AB})$.

References for Chapter IV

[1] Greg van Anders, Daphne Klotsa, Andrew S. Karas, Paul M. Dodd, and Sharon C. Glotzer. Digital alchemy for materials design: Colloids and beyond. *ACS Nano*, 9(10):9542–9553, 2015. doi: 10.1021/acsnano.5b04181.

[2] Y. Geng, G. van Anders, P. M. Dodd, J. Dshemuchadse, and S. C. Glotzer. Engineering entropy for the inverse design of colloidal crystals from hard shapes. *ArXiv*, 2017.

[3] Rose K. Cersonsky, Greg van Anders, Paul M. Dodd, and Sharon C. Glotzer. Relevance of packing to colloidal self-assembly. *Proceedings of the National Academy of Sciences*, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1720139115.

[4] Erin G. Teich, Greg van Anders, and Sharon C. Glotzer. Identity crisis in alchemical space drives the entropic colloidal glass transition. *preprint*, 2018.

[5] Pablo F Damasceno, Michael Engel, and Sharon C Glotzer. Predictive self-assembly of polyhedra into complex structures. *Science (New York, N.Y.)*, 337 (6093):453–7, jul 2012. ISSN 1095-9203. doi: 10.1126/science.1220869.

[6] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. Glass and jamming transitions: From exact results to finite-dimensional descriptions. *Annual Review of Condensed Matter Physics*, 8(1): 265–288, 2017. doi: 10.1146/annurev-conmatphys-031016-025334.

[7] Yina Geng, Greg van Anders, and Sharon C. Glotzer. Predicting colloidal crystals from shapes via inverse design and machine learning. *ArXiv*, 2018.

[8] HOOMD-blue web page:. *http://codeblue.umich.edu/hoomd-blue*.

[9] Joshua a. Anderson, Chris D. Lorenz, and a. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics*, 227(10):5342–5359, 2008. ISSN 00219991. doi: 10.1016/j.jcp.2008.01.047.

[10] Jens Glaser, Trung Dac Nguyen, Joshua A. Anderson, Pak Lui, Filippo Spiga, Jaime A. Millan, David C. Morse, and Sharon C. Glotzer. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications*, 192:97–107, 2015. ISSN 00104655. doi: 10.1016/j.cpc.2015.02.028.

CHAPTER V

Conclusions and Outlook

In this thesis, we have studied two model systems (i.e., nets and hard polyhedra) for inverse materials design, and we have developed novel methods to solve inverse materials design problems.

Self-folding polyhedra offer an intriguing system to study folding on the micro and nanoscales. Throughout this work, these nets are studied using molecular dynamics and Markov state models. By studying the simplest platonic solids (i.e., tetrahedron, cube, and octahedron), we were able to define a set of design rules that could be used to successfully predict the best and worst nets of the dodecahedron and icosahedron, in terms of folding propensity. In general, the set of pathways that each net traverses to reach the folded state is dependent on temperature. At high temperatures the net uses pathways that have only native contacts, while pathways that have non-native contacts are predominantly used at low temperatures. Regardless of the the folding propensity, at high temperatures all of the nets fold along pathways that greedily attempt to maximize the conformational entropy at each step. This model of self-folding is powerful because of its simplicity. The computed pathways are generally very intuitive, making this system a good testbed for emerging folding theories. A primary strength of this work is the level of detail with which we can describe these

folding pathways.

Next, we developed new methods with which to sample extended ensembles using Monte Carlo simulations. Needless to say, shape and geometry are very complex variables in hard particle simulations. We have developed three distinct ways of sampling shape space using shape families, vertex translations, and elastic deformations. For each of these three methods, we were able to prove the detailed balance conditions for the algorithm. We also examined the detailed balance of the entire system, in the context of the HOOMD simulation framework. As new code will always be needed to solve new problems, we outlined the considerations that we used to develop the simulation code in the most general way possible. These new methods make for a novel tool that researchers can utilize to answer questions that would have been impossible to answer previously. By considering shape to be a variable in the simulation, we were able to design a particle for a crystal structure, for which no known particle could assemble previously. Beyond materials design, this method has provided a way to characterize packing in hard particle systems.

Finally, we used the methods developed in Chapter III to study how shapes that do not assemble change to minimize the free energy of the system and avoid the glass transition. We found that these shapes reduce their faceting, which drives a disordered to ordered transition. We found that fcc/hcp was the most common crystal structure that formed from these systems, and that many independent replicates would assemble into several different crystal structures. Understanding the stability of the different crystal structures will help us to comprehend the underlying energy landscape, and can provide some insight into why this rich structural diversity exists.

Outlook

Interaction design for polyhedral nets: In nature proteins can have highly specific interactions that drive the folding process. Understanding how these specific interactions are used to achieve the maximum yield or folding rate could be vital in understanding more complex systems. Using polyhedral nets, one could determine the fewest number of specific interactions the system needs to reliably assemble into a target structure by systematically assigning a specific interaction to each pair of edges. Many open questions remain. How does the number of required interactions scale with the complexity of the nets? How does the number of interactions relate to the folding rate and folding yield? For multi-ground state systems, can these interactions be used to design pathways between ground states?

Molecular/Colloidal Machines: Molecular and colloidal machines offer a promising way to build pluripotent materials. One exciting extension of this work would be to design colloidal machines that can move through self-folding actuations and assemble into multiple ground states. One could design a system to have a programmed response to light, temperature, or electric or magnetic fields to drive the folding process. A two-dimensional example that we modeled was a hinged dissection. When a magnetic field was applied, the linkage folded into a square. When the direction of the field was reversed, the square transformed into a triangle. While this example is very simple, it demonstrates proof of concept. There are still many open questions. How many states can be successfully programmed to assemble into a single system? Are there analogous three-dimensional systems?

To future graduate students:

To close out this thesis (and my journey as a graduate student), I would like to share a favorite story of mine about research. It was told to me by Ron Larson during his Fluid Mechanics course in my first year of graduate school, and has stuck with me ever since. Here I retell the story as it was told to me, to the best of my memory.

In the 1820's, the English botanist Robert Brown was trying answer the question: what is the essence of life? His chosen model system was pollen suspended in a water. Brown observed that the pollen moved with irregular motion when he observed it under a microscope. He hypothesized that this motion was the "essence" that he was searching for, and to test this theory he irradiated the pollen. When he looked under the microscope again, he found that the pollen still moved in the same way. Being a good scientist, Brown wrote his paper describing his methods, results, and conclusions, even though his experiment largely failed. Approximately 80 years later, Einstein wrote a paper mathematically describing this phenomena. However, we still call this random motion "Brownian". This is because Brown did what we all strive to do: *science*, in its purest form – observing the natural world and describing it to the best of our ability.