Inverse Material Design in Colloidal Self-Assembly

by

Yina Geng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2019

Doctoral Committee:

Professor Sharon C. Glotzer, Chair
Assistant Professor Bryan Goldsmith
Professor Nicholas A. Kotov
Assistant Professor Xiaoming Mao
Assistant Professor Greg van Anders, Queen's University

Yina Geng

yinageng@umich.edu

ORCID: 0000-0003-2991-5776

For my parents, my brother, and Huan.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A holy grail in materials science research remains the solution of the inverse design of desired crystal structures, which seeks to identify a set of building block characteristics (shape, interactions) that guarantee the self-assembly of a target colloidal crystal structure. In this work, we solve the inverse design problem using the "Digital Alchemy" framework, which is an extended ensemble method that takes the particle shape as a thermodynamic parameter and produces the optimal shape that maximizes the entropy of a given target structure at a given thermodynamic state point. Combining digital alchemy and machine learning, we build a prediction model to predict crystal structures from building block geometry features.

We carry out the inverse design of particles that can self-assemble target structures due solely to the emergent effects of entropy arising from their shape. We extend the alchemical Monte Carlo algorithm to hundreds of design dimensions, and sample hundreds of millions of shapes to engineer optimal particles shapes for the self-assembly of six target crystal structures, and one as-yet-unknown structure. We extract important shape characteristics and design optimal symmetric shapes that are synthesizable by experiment.

We next predict crystal structures from building block attributes using digital alchemy and machine learning. Our model correctly classifies more than ten thousand polyhedral shapes into 13 different structures with a predictive accuracy above 95% using only two geometric shape measures. We test our model on previously reported colloidal crystal structures for 71 symmetric polyhedra and obtain 92% accuracy.

We also study the inverse design problem focusing on nanoparticle shapes that have already been made in experiment. We determine for eight shape families reported in lab-

oratory synthesis optimal particle shapes for possible target colloidal crystal structures, in an effort to guide experimentalists in choosing optimal building blocks to self-assemble colloidal crystals.

Results from this dissertation demonstrate that the digital alchemy is a generative method to solve the inverse material design problem. It can guide experimentalists to use optimal building blocks to make colloidal materials. We provide a quantitative model to predict crystals solely from the geometry of their building blocks, and prove that entropic colloidal crystals are controlled by surprisingly few parameters. The results give insight into applying data science and machine learning techniques in the material design field.

CHAPTER I

Introduction

1.1 Self-assembly in Colloidal Materials

Self-assembly is the process of individual building blocks spontaneously organizing into ordered structures. Colloidal particles, whose size ranges from 10 nm to several microns, serve as a new building block self-assembling to novel materials. This is achieved because of the recent development in particle synthesis and tunable particle sizes, shapes, surface properties, and interactions [35, 49]. Colloidal structures have applications in drug delivery, disease diagnostic system, energy storage, photonic devices and solar cells [71]. Similar structures with atomic systems and novel structures with no known atomic or molecular analogue have been discovered in colloidal assembly [19, 23, 84]. In particular, colloidal particles interacting with the excluded volume arising from their shape have been observed to self-assemble complex crystals, liquid crystals, and even quasicrystals in compute simulations and experiments [40, 20, 51, 89]. In the hard particle system, the thermodynamic property can be understood through entropy maximization, which is an emergent behavior and dominant when the system becomes crowded [27, 92].

1.2 Inverse Design from Hard Shapes

To design materials with desired properties, previous studies used a 'forward' approach or trial and error, e.g., given a building block with a given set of anisotropies, of what or-

dered structures are possible [24, 1, 20, 30]. However, the inverse material design approach by reverse engineering a target structure to obtain the building block that will give rise to that structure via self-assembly is more efficient and powerful. Here, we carry out the inverse design of particles that can self-assemble target structures due solely to the emergent effects of entropy arising from their shape. To do so, we generalize an extended ensemble approach, "digital alchemy" [93], that treats particle shape parameters as thermodynamic variables, and perform Alchemical Monte Carlo simulations that sample particle shapes in the extended ensemble. By sampling hundreds of millions of shapes, we engineer optimal particle shapes for the self assembly of six target structures known to self-assemble in simulations of hard particles, and one as-yet-unknown structure. This work is shown in Chapter III, and is available online: "Engineering Entropy for the Inverse Design of Colloidal Crystals from Hard Shapes", Yina Geng, Greg van Anders, Paul M. Dodd, Julia Dshemuchadse, Sharon C. Glotzer, *arXiv:1712.02471*.

The gulf between ordered phases predicted to form in computer simulation and what has been so far realized in experiment is narrowing, but is still wide. Prior work shows that synthesized particles are far from optimal "eigenshapes" for target superlattice structures. By optimal "eigenshape", we mean the shape that minimizes the free energy of the structure at a given density. We use digital alchemy to determine for eight families of shapes reported in laboratory synthesis optimal particle shapes for possible target colloidal crystal structures. Within each family we predict optimal building block shapes to obtain several target superlattice structures. This work is shown in Chapter V and will be published.

1.3 Predicting Structures from Hard Shapes

A fundamental challenge for materials researchers is the ability to predict crystal structures solely from knowledge about the constituent atoms, molecules, or particles, without

the need for simulations or experiments. Many researchers have attempted this quest; a well-known example is the early effort of Pauling to predict crystal structures from atoms based solely on their atomic radii [73]. Pauling's rules have since been adopted to program the assembly of DNA-functionalized nanospheres into colloidal crystals isostructural to those Pauling considered for atoms, as well as ones with no atomic counterpart[65, 60, 58, 55]. Prediction of crystals from molecules or anisotropic particles is substantially harder [92, 91, 51]. In 2012, a study of 145 different polyhedrally shaped particles and their entropy stabilized crystals provided sufficient data to discover a correlation between coordination number and isoperimetric quotient (IQ), a measure of the roundness of a particle [20]. The study found that knowledge of the coordination number in the dense fluid (a simple observable in simulations) and the particle IQ allows one to predict whether that fluid of particles will crystallize and, if so, whether it will form a liquid crystal mesophase, a medium-coordination crystal, or a close-packed crystal (including topologically close-packed phases). Despite being only partially predictive, that was the state of the art in 2012 and in the half dozen years since, despite continued discoveries of colloidal crystals that add to the knowledge base. Here we leverage two recent important computational advances, digital alchemy and machine learning to enable a much higher level of predictiveness of colloidal crystal structures from particle shape. This work is shown in Chapter IV, and is available online: "Predicting colloidal crystals from shapes via inverse design and machine learning", Yina Geng, Greg van Anders, Sharon C. Glotzer, *arXiv:1801.06219*.

CHAPTER II

Methods

2.1 Hard Particle Monte Carlo

The fundamental problem of statistical mechanics is to calculate the average value of a quantity of interest in a physical system in thermal equilibrium at temperature T. However, the computation is difficult due to the fact that the total number of microstates of a thermodynamic system is extremely large. Importance sampling allows us to calculate a correct average value by sampling the configuration state point with a relative probability proportional to the Boltzmann factor. Markov chain Monte Carlo, in the Metropolis scheme [64] is a method to do the importance sampling. The system moves from an old (o) configuration state to a new (n) state by making a translation or rotation move of the particle. The system would accept the state change with probability

$$(2.1) \qquad acc(o \rightarrow n) = min(1, exp\{-\beta[U(n) - U(o)]\}),$$

$U(s)$, where $s = \{o, n\}$, is the potential energy of state s, and $\beta = 1/k_B T$.

In the hard particle system, particles do not have attractive or repulsive interactions; they simply can not overlap. The potential energy of interaction between two hard particles P_1 and P_2 is given by

$$(2.2) \qquad U(P_1, P_2) = \begin{cases} \infty & \text{if } P_1 \text{ and } P_2 \text{ overlap;} \\ 0 & \text{otherwise.} \end{cases}$$

4

In the hard particle Monte Carlo, particle moves are rejected if they result in any particle overlaps, and accepted otherwise.

2.2 Digital Alchemy Method

The digital alchemy method treats particle shapes as thermal parameters and allows them to fluctuate in the system. It begins with the extended partition function [93]

$$Z = \sum_{\sigma} e^{-\beta(H - \sum_i \mu_i N \alpha_i - \lambda \Lambda)} \; ,$$
(2.3)

where β is the inverse temperature, μ_i are so-called alchemical potentials that are thermodynamically conjugate to the particle attribute parameters α_i that describe particle shape, N is the number of particles in the system, Λ is the external field that forces the particles to sit in an Einstein crystal with spring constant λ, and the summation is over particle coordinates and orientations and over the space of particle shapes, as in [93]. The combination of $\lambda\Lambda$ serves as the design term. When λ is positive, the system is driven toward particle shape parameters α_i that favor increasing Λ, which allows one to design toward a target structure encoded in Λ. To design purely entropic systems, we model particles with purely hard interactions so that the partition function is a sum over all non-overlapping particle configurations, and the phase space part of the Hamiltonian reduces to kinetic terms, which we can integrate analytically. Hereafter, we set $\mu_i = 0$ to sample shapes without bias. Unbiased shape sampling, coupled with detailed balance, drives randomly chosen initial shapes to converge toward shapes that are thermodynamically optimal (maximizing entropy) for the target structure at a given temperature and density. In the Alchemical Monte Carlo simulation, system moves from state 2 to state 1 by taking particle translation, rotation, or shape moves. The Metropolis [64] criterion derived from detailed balance

condition is

$$(2.4) \qquad \Pi_{2 \to 1} = min\left(1, \frac{(det(I_{\alpha 1}))^{N/2}}{(det(I_{\alpha 2}))^{N/2}} e^{-\beta(U_{\alpha_1} - U_{\alpha_2})}\right),$$

I_{α_s}, where $s = \{1, 2\}$, is the moment of inertia tensor of particle shape with shape parameter α_s, and U_{α_s}, where $s = \{1, 2\}$, is the potential energy of system with particle shapes having parameter α_s. Our method differs with previous extended ensemble methods [93] in that we extend the algorithm to hundreds of design dimensions.

CHAPTER III

Engineering Entropy for the Inverse Design of Colloidal Crystals from Hard Shapes

This chapter is adapted from Ref. [31], and is available online: "Engineering Entropy for the Inverse Design of Colloidal Crystals from Hard Shapes", Yina Geng, Greg van Anders, Paul M. Dodd, Julia Dshemuchadse, Sharon C. Glotzer, *arXiv:1712.02471*.

3.1 Engineering Entropy

Our understanding of entropy has undergone three revolutions since its association with lost heat by Clausius in the 1800s [18]. The first is the discovery by Boltzmann [11] and Gibbs [34] of entropy's central role in statistical mechanics and its colloquial association with disorder. The second is the discovery by Shannon of entropy's central role in information theory as a quantifier of statistical ignorance [83]. The third is the discovery by Onsager [70] in the late 1940's and then by Kirkwood and collaborators [3, 95] in the late 1950's of entropy's seemingly paradoxical implication in ordering hard particles. The systematic study *via* simulation of entropic ordering was pioneered by Frenkel and collaborators in the 1980's, see, e.g. Refs. [27, 28], leading to recent discoveries of a surprisingly large number of possible structurally ordered phases from hard, anisotropically shaped particles [1, 20, 30, 24]. In those works, simulation studies begin with a volume of identical hard particles of fixed shape, and the entropy of the system is maximized to find

thermodynamic equilibrium phases. In 2015 van Anders *et al* [93] introduced a method to start not with a given particle shape, but instead with a target colloidal crystal structure, and, *via* entropy maximization, find a shape within a limited family of shapes that maximizes entropy for that structure at the selected density. That *inverse design* approach – "digital alchemy" – flips the usual idea of entropy optimization in hard particle systems on its head.

In this work, we seek not only to optimize entropy starting from a target structure, but to engineer it so as to inversely design shapes that will self-assemble back into the target structure, and have a good chance of being synthesizable. Engineering entropy is both conceptually and technically difficult because entropy is a globally defined, purely statistical concept. This means there exists no obvious direct, quantitative link between the macroscopic order that emerges from entropy maximization and the microscopic, designable details of a system's components. Moreover, the range of designable attributes of component particle shapes has exploded due to advances in colloidal synthesis, e.g. Refs. [47, 97, 85, 38, 102, 14, 77, 80, 79, 81, 36], and now go well beyond what can be designed by trial and error. In contrast to the design of particle shapes, pairwise interaction potentials (force fields) between atoms or nanoparticles are now routinely designed for simple target structures, and realized in experiment, in cases where potential energy, rather than entropy, dominates [65, 68, 72, 62, 86, 45, 46].

Here, we generalize 'digital alchemy' [93], an extended ensemble approach that treats particle shape parameters as thermodynamic variables, to sample hundreds of millions of particle shapes with no restrictions other than convexity of the particle shape. Whereas an extended ensemble approach was previously applied to ensembles extended in one or two design dimensions [101, 93], here we extend ensembles to treat hundreds of design dimensions, providing a general approach for quantitatively engineering entropy for struc-

ture. We perform Alchemical Monte Carlo (Alch-MC) simulations and engineer optimal particle shapes for the self-assembly of six target structures known to self-assemble in simulations of hard particles. In each case, we then identify key symmetry characteristics of the particle shape necessary for that target structure, symmetrize the optimal particle shape, and by re-running Alch-MC on the symmetrized particle with symmetry restrictions, find an even better (higher entropy) shape that, because of its high symmetry, has the potential to actually be made in the lab. Further, we propose an additional, as-yet-unknown structure and engineer a symmetrized particle shape that forms that structure in simulation. Our approach demonstrates a general, quantitative paradigm for engineering entropy in large design spaces that reflect the diversity of colloids and nanoparticles that can be synthesized using current techniques [47, 97, 85, 38, 102, 14, 77, 80, 79, 81, 36]. Moreover, it opens the possibility of quantitatively engineering entropy for other novel structures or behaviors [21], allows for the discovery of important features determining structural outcomes in self-assembly [32], and can be generalized to systems with enthalpic interactions.

3.2 Digital Alchemy Simulations

We performed Digital Alchemy simulations using the HPMC plugin [5] for HOOMD-Blue [4] in an $NVT\mu$ ensemble at $\mu = 0$. We placed no fewer than 100 particles in a periodic simulation box. The exact number was chosen to be a multiple of the number of particles in the unit cell of the target structure. Particle shapes were initialized randomly with 32 or 64 vertices (arbitrary convex polyhedron case) or with each shape parameter taken as either 0 or 1 as convenient (symmetry-restricted case). Monte Carlo (MC) sweeps involve particle translations, rotations, and collective shape moves for all particles in the system. For each shape move, we (i) either moved a vertex (arbitrary convex polyhedron case) or generated a trial change in shape parameters (symmetry restricted case) (ii) resized

the trial shape to unit volume; (iii) checked if the move induced any particle overlaps, and then (iv) accepted the move based on the Boltzmann factor as described in Ref. [93]. Translation and rotation moves followed standard procedures (see, e.g., Refs. [40, 1, 30, 19, 20, 92, 91]). We compressed the system to packing fraction $\eta = 0.6$, with the spring constant λ fixed to 1000 (where energy is specified in units of $k_{\mathrm{B}}T$, and length units are given in terms of the particle volume). After we reached the target packing fraction, we logarithmically relaxed the spring constant to zero. We then relaxed the system for millions of MC sweeps.

3.3 Two-step Inverse Design

Our inverse design approach proceeds in two steps (see Fig. 3.1). The first step begins with randomly generated, arbitrarily shaped convex polyhedra whose shape evolves during a Monte Carlo (MC) simulation by sampling from particle "shape space" via an extended, "alchemical" ensemble [93]. Unlike a traditional molecular MC simulation in which a system of fixed particle shapes samples configurational states in phase space, in an Alch-MC simulation particles sample not only positions and orientations but also shapes consistent with the target structure, finding thermodynamically optimal shapes. Alch-MC for polyhedra with n vertices explore a $D = 3n - 4$ dimensional parameter space accounting for fixed particle volume and rotational invariance, and produce mathematically irregular but well-defined particle shapes that (i) maximize the entropy of the target structure and (ii) successfully self-assemble the target structure in a MC simulation starting from a disordered fluid. The second step symmetrizes the designed particles to obtain shapes that still easily assemble the target structure, but because of their symmetry can be made today using existing synthesis methods [35, 97, 80, 36]. Depending on the target crystal structure we symmetrized particle shapes either through truncation or through truncation and vertex
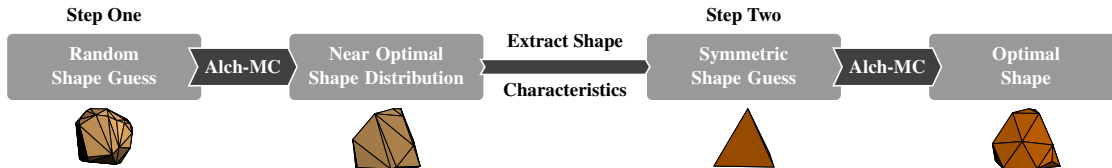
Figure 3.1: Schematic diagram illustrating the inverse design process. In step one, Alchemical Monte Carlo (Alch-MC) starts from a random convex shape and then finds an unsymmetrized optimal shape for the target (here, diamond) structure. Cosine of dihedral angle distribution and PMFT iso-surface of the unsymmetrized optimal shape reveals that it has tetrahedral characteristics. In the second step, fluctuating particle shape alchemical Monte Carlo (Alch-MC) simulation starts from a tetrahedron and finds an optimal symmetrized shape for the diamond structure.

augmentation.

3.3.1   Step One - Designing Unsymmetrized Near Optimal Shapes

We targeted six structures – simple cubic (SC), body-centered cubic (BCC), face-centered cubic (FCC), diamond, β-W, and β-Mn. In Alch-MC, we moved a vertex when we perform a shape change move. After we reached the target packing fraction $\eta = 0.6$ and relaxed the spring constant to zero, we then relaxed the system for 1×10^6 (BCC, FCC, SC, diamond) or 8×10^6 (β-W) or 3.6×10^7(β-Mn) MC sweeps. For each target crystal structure, we performed 20 independent simulations and analyzed the shapes in the final 1.5×10^5 sweeps. For validation we directly computed the free energy [26] of the thermally sampled particle shapes as a function of Alch-MC time (Fig. 3.2) (step one), and verified that after starting from random initial particle shapes, our simulations converged to shapes comprising systems of lower free energy for a given target structure (β-Mn is depicted) than for a geometric ansatz that is a hard, space-filling particle in the shape of Voronoi cells of the target structure and gives a possible candidate to assemble its target structure [82]. Details about the free energy calculation method are in section 3.3.2.
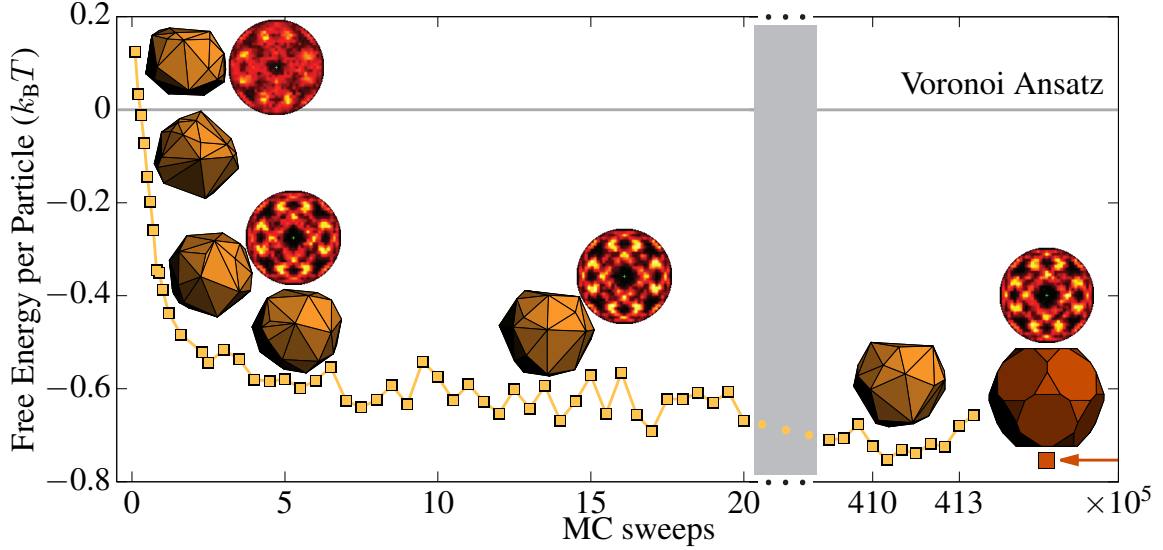
Figure 3.2: Alch-MC for the inverse design of an unsymmetrized thermodynamically optimal hard particle shape to form a target (here, β-Mn) structure. The structure is imposed by an auxiliary design criterion, and detailed balance drives particles to take on shapes (selected shapes are displayed in light yellow) that are favorable for the target structure (indicated by selected bond-order diagrams). Directly computed free energy confirms Alch-MC simulation over $\gtrsim 10^5$ distinct shapes converges to shapes that have lower free energy (by ≈ 0.7 $k_{\mathrm{B}}T$ per particle; numerical errors are smaller than markers) than shapes chosen by Voronoi construction. Desired shape features can be inferred from the equilibrium particle shape distribution and used to create a symmetry-restricted ansatz, which yields a thermodynamically optimal synthesizable shape (shown in dark yellow).

3.3.2 Step Two - Designing Symmetrized Optimal Shapes

In step two, we designed symmetrized optimal shapes that are synthesizable in experiments for the six target crystal structures using Alch-MC simulations. We generated a trial change in shape parameters when we perform a shape move in Alch-MC.

For β-Mn, FCC and BCC, symmetrized, truncated shapes in step two produced lower free energy crystal structures than sampled unsymmetrized polyhedra found in step one. We give detailed results here for the most complex case, β-Mn. For β-Mn, the equilibrium distribution of convex polyhedra shapes resulting from our Alch-MC simulations at packing density $\eta = 0.6$ in step one yields a family of shapes with characteristic cosine of dodecahedral facet angles (distribution peaks at ≈ -0.447 and 0.448, see Fig. 3.3(1), *vs.*

the perfect dodecahedron $\approx \pm 0.447$). Consistent with the particle faceting, potential of mean force and torque (PMFT) calculations [92] for a particle selected from the peak of the shape distribution (Fig. 3.4A) produced isosurfaces with dodecahedral entropic valence [91]. Alch-MC simulation of symmetrized shapes restricted to a one-parameter family of truncated dodecahedra (see Fig. 3.6B) in step two yielded an optimal truncated shape with facet area 0.36 (Fig. 3.3(1)); the peak in facet area differs by less than 3% from the peak observed for the unrestricted shapes (0.37). The optimal shape parameters in step two can be found in Table 3.2. To further validate that the particle shape with manifest dodeca-hedral symmetry is the putative optimal shape, we directly compared the free energy of the target colloidal crystal with the optimal truncated shape and a shape from the peak of the angle distribution of arbitrary convex shapes (Fig. 3.5), and found the symmetric-shape crystal has lower free energy. This result is consistent with our expectation that the free energy landscape of the high-dimensional parameter space of shapes is rough with nearly degenerate minima. For comparison, we also computed the free energy for a packing-based estimate - the Voronoi shape [82]. There are two Voronoi cells in β-Mn, only one of which can self assemble the structure without enthalpic interactions [20]. We computed the free energy for the target structure with the Voronoi shape, and found that our approach produced shapes with lower free energy than the Voronoi shape (Fig. 3.5). Fig. 3.2 shows that Alch-MC converged rapidly to shapes that have lower free energy than the Voronoi ansatz by $\approx 0.7 k_{\mathrm{B}} T$ per particle, and implies the existence of a large space of shapes that are all better than the geometric ansatz. The simulation trajectory shown in Fig. 3.2 explores $\gtrsim 10^6$ shapes that have lower free energy in the target structure than the geometric ansatz. We follow standard conventions and express all (free) energies in units of the thermal energy ($k_{\mathrm{B}} T$). Consistent results were found for BCC (Fig. 3.3(2)) and FCC (Fig. 3.3(3)) target structures. The connection between faceting and the emergence
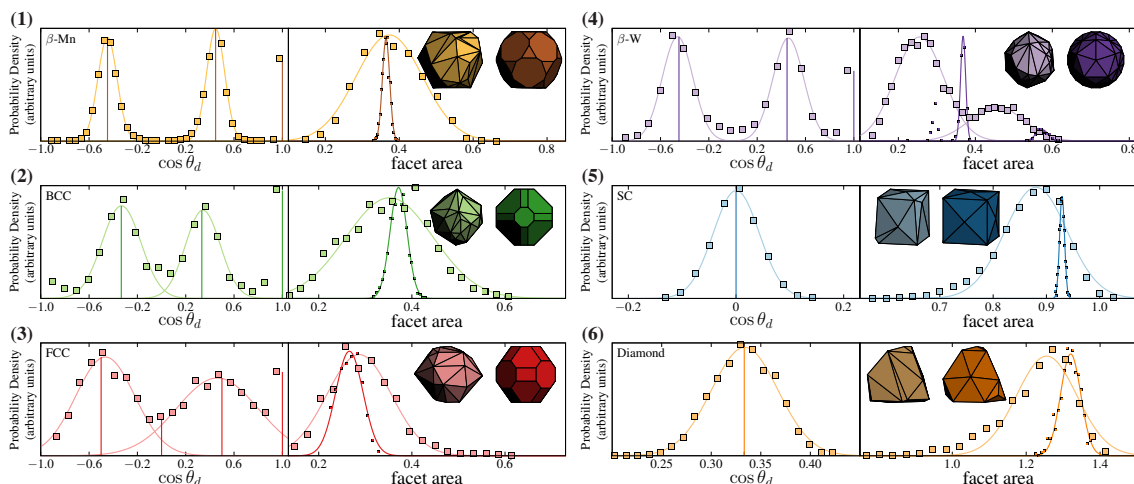
Figure 3.3: Two-step shape alchemical Monte Carlo (Alch-MC) entropic particle-shape optimization for six target structures: β-Mn, BCC, FCC, β-W, SC and Diamond. For each target structure, an initial Alch-MC simulation over 92- or 188-dimensional spaces of convex polyhedra in step one converged to highly faceted modifications of identifiable Platonic, Archimedean, or Catalan solids, obtained by calculation of the equilibrium distribution of the (left, light color, squares) cosine of dihedral angles ($\cos\theta_d$) and (right, light color, squares) facet areas (Gaussian distributions are plotted with solid lines for comparison). We show the mean of the cosine of dihedral angle distributions in Table 3.1. In step two, Alch-MC simulation over symmetry-restricted families of shapes determined a thermodynamically optimal and synthesizable shape (shown in dark color). For each target structure, we calculate the equilibrium distribution of the (left, dark color, vertical line) cosine of dihedral angles ($\cos\theta_d$) and (right, dark color points, with Gaussian distribution fitting) facet areas for symmetrized optimal shapes. The distributions are in arbitrary units. In all cases, representative shapes spontaneously self-assembled target structures in NVT simulations with periodic boundary condition satisfied.

of entropic valence with local structural order is robust (BCC–Fig. 3.4B; FCC–Fig. 3.4C). In the second step we repeat the procedure using symmetric truncated shapes suggested by the shapes observed in the first step. In all cases we obtained lower free energy shapes than the geometric ansatz (β-Mn $-0.753 \pm 0.001\ k_{\mathrm{B}}T$; FCC $-0.907 \pm 0.001\ k_{\mathrm{B}}T$; BCC $-0.334 \pm 0.001\ k_{\mathrm{B}}T$) (see Fig. 3.5).

For β-W, SC and diamond, we found that unsymmetrized polyhedra had lower free energy than symmetrized truncated polyhedra. For these crystals, we implemented step two using symmetrized, truncated, and vertex-augmented polyhedra. We give detailed results

here for the most complex case, β-W. For β-W, Alch-MC simulation of unsymmetrized shapes in step one yielded an equilibrium distribution of convex polyhedra with facet angle distribution peaks at ± 0.458 (Fig. 3.3(4)). Like for β-Mn, this falls near the peaks for dodecahedra, but for β-W the facet area distribution is bimodal, indicating, and confirmed by visual inspection, the existence of two large parallel facets. Faceting is again consistent with emergent entropic valence (Fig. 3.4D) evident in isosurfaces of PMFT measurements [92]. Free energy calculations (Fig. 3.5) confirm that a geometric ansatz shape has 0.433 ± 0.006 $k_{\mathrm{B}}T$ more free energy per particle in the target crystal than a shape at the peak of the distribution of convex shapes. In contrast to the case 1 structures (β-Mn, FCC and BCC), Alch-MC of symmetrized shapes restricted to a two-parameter family of truncated dodecahedra (see Fig. 3.6E top) in step two yielded shapes with lower free energy in the target β-W structure than the geometric ansatz, but higher free energy than for shapes at the peak of the angle distribution of arbitrary convex polyhedra. This finding indicates that the restriction to truncation alone is too severe for β-W. Alch-MC simulation of a refined truncated dodecahedron with vertex-augmented faces (see Fig. 3.6E bottom) converged to a shape with 0.620 ± 0.001 $k_{\mathrm{B}}T$ lower free energy per particle than the geometric ansatz. Truncated and augmented free energy minimizing shapes were also found for SC and diamond (Fig. 3.5) SC -0.704 ± 0.006 $k_{\mathrm{B}}T$; diamond -0.54 ± 0.01 $k_{\mathrm{B}}T$), which again preserve the connection between faceting and entropic valence (SC–Fig. 3.4E; diamond–Fig. 3.4F). Because this facet–valence connection persists, the facet area distributions for SC (Fig. 3.3(5)) and diamond structures (Fig. 3.3(6)) are unimodal due to the simpler local structural motif in those structures compared to β-W where the facet area distribution is bimodal (Fig. 3.3(4)).

Cosine of dihedral angles and facet areas

Unsymmetrized shapes have 32 or 64 vertices. Facets with area $a_f > a_f^*$ (we use
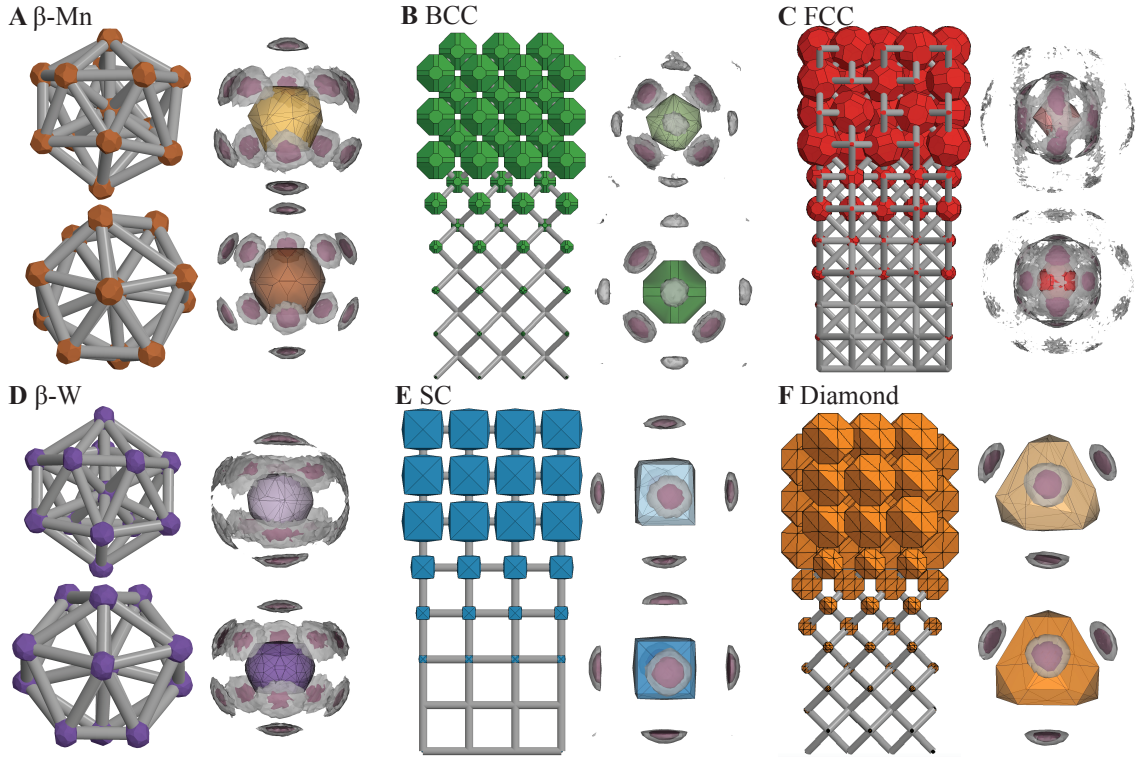
Figure 3.4: Structure and potential of mean force and torque (PMFT) isosurfaces for optimal shapes in six target structures: β-Mn, BCC, FCC, β-W, SC and diamond. Each panel shows structural coordination (global: BCC, FCC, SC, diamond; local: β-Mn, β-W), and PMFT isosurfaces at free energy values of $1.4\ k_\text{B}T$ (light gray) and $0.7\ k_\text{B}T$ (pink) above the minimum value, for an optimal but unsymmetrized convex polyhedron (top) and for an optimal symmetry-restricted polyhedron (bottom). PMFT isosurfaces indicate emergence of particle faceting corresponds with entropic valence localized at particle facets that preferentially align along crystal lattice directions. PMFT isosurfaces for symmetry-restricted polyhedra retain valence–lattice correspondence.

$a_f^* = 0.03$ but our results are not sensitive to changes in a_f^*) were clustered by their normal vector using the DBSCAN [25] scikit-learn module [74]. Clustered facets are represented by area-weighted average normals. We computed aggregate facet areas and the cosine of the angle between all average normals in a polyhedron, which for adjacent facets is just the dihedral angle.

Direct Free Energy Computation

We computed the Helmholtz free energy difference between the target crystal and the Einstein crystal using Frenkel-Ladd thermodynamic integration [26] via the implementa-
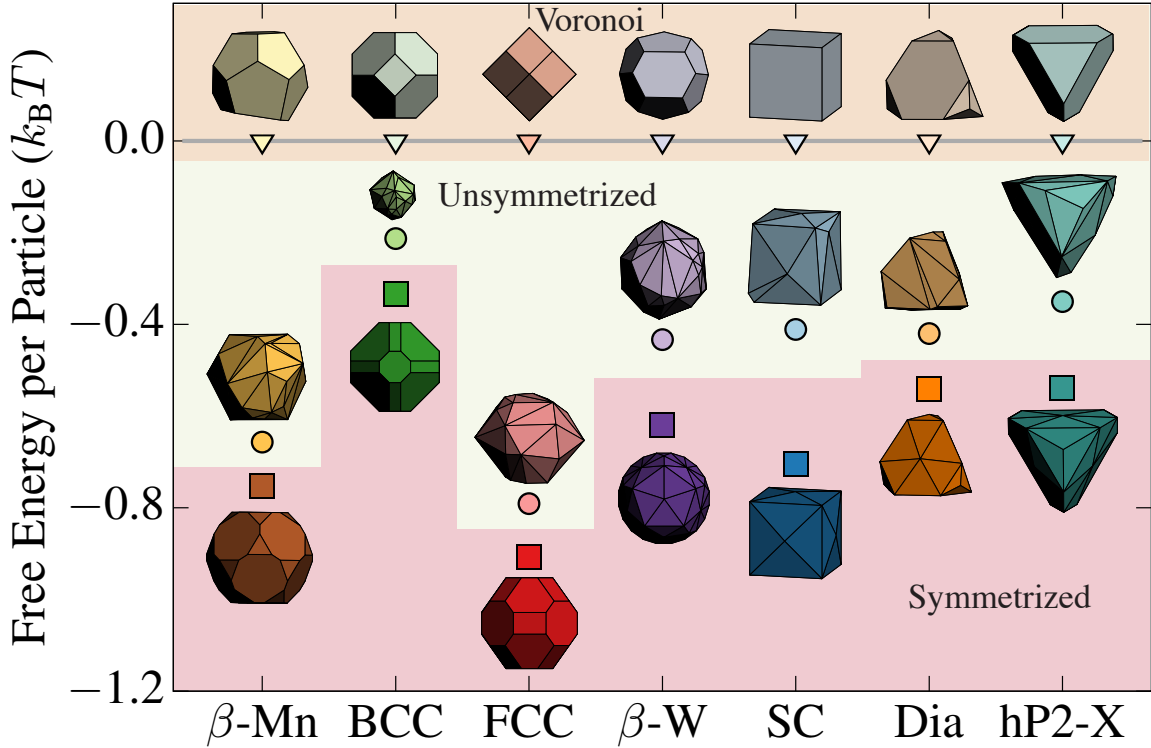
Figure 3.5: Direct free energy comparison of our entropic engineering strategy for seven target structures: β-Mn, BCC, FCC, β-W, SC, diamond and hP2-X. For each structure we calculated the free energy of the target crystal for a shape formed from a geometric ansatz based on the Voronoi decomposition of the structure (triangles). Compared with the Voronoi ansatz, we find that Alch-MC simulation over arbitrary convex polyhedra in step one produces shapes (circles) that spontaneously self-assemble the target structures with higher entropy. Symmetry restricted polyhedra (squares) (truncated polyhedra for β-Mn, BCC and FCC; truncated and vertex-augmented polyhedra for β-W, SC, diamond and hP2-X) inferred from shapes in step one produce putatively thermodynamically optimal particle shapes by maximizing entropy.

tion used in Refs. [39, 82]. We placed approximately 2000 particles in a periodic simulation box; the exact number was chosen to be a multiple of the number of particles in the unit cell of the target structure. For SC, BCC, diamond and hP2-X structures, particles in the assembled structure have orientational order. Einstein crystal positions and orientations were taken directly from the space-filling tessellation. For FCC, β-W, and β-Mn structures, particles in the assembled structure do not show orientational order. To create an appropriate Einstein crystal, we first initialized the simulation at a low packing fraction $\eta = 0.5$, chose particle positions using the target structure, and randomly assigned a set of

Structure	mean of cosine of dihedral angle distribution			
BCC	-1/3	1/3	1	N/A
FCC	-1/2	0	1/2	1
β-Mn	-0.447	0.447	1	N/A
SC	0	1	N/A	N/A
Diamond	1/3	N/A	N/A	N/A
β-W	-0.447	0.447	1	N/A
hP2-X	-0.531	-0.484	0.148	0.617

Table 3.1: Mean of cosine of dihedral angle distribution for unsymmetrized optimal shapes in step one for seven target structures, which used to infer symmetry-restricted shapes in step two.

orientations observed in the assembly. Then we compressed the system to packing fraction $\eta = 0.6$, allowing particles to rotate to resolve overlaps. We computed the alchemical free energy of the target structure [93]. We normalized free energies in all plots by setting the free energy of the target structure with the Voronoi particles to be zero. Negative values of the free energy indicate lower free energy for a given particle than a Voronoi particle shape.

We validated our methodology by performing the free energy calculation described above for truncated tetrahedra in a diamond structure, and checked it reproduces the results reported in Ref. [93] computed via the Bennett acceptance ratio method [9].

Symmetric Shape Constructions in Step Two

We describe the construction of symmetric shape families for inverse engineering optimal particle shapes for each candidate structure. In the Alch-MC in step two, when we perform a shape move, we restrict the shape to change in the constructed symmetric shape families for each target structure. In all cases, after the geometric construction below was carried out, all particle shapes were normalized to unit volume to maintain constant system density.

BCC and FCC

We focus on the spheric triangle group $\Delta_{4,3,2}$[15], which is constructed with three fam-
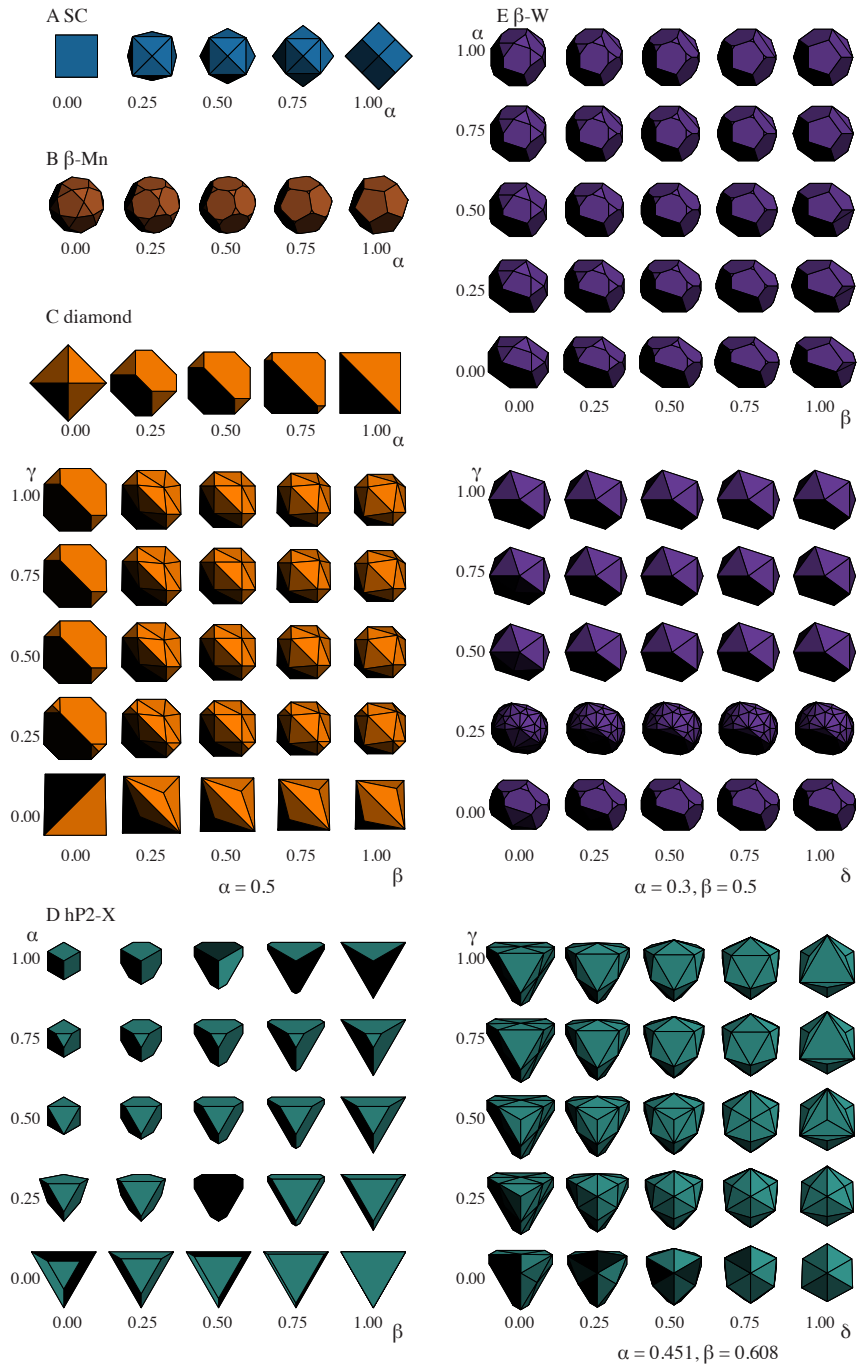
Figure 3.6: Illustration of the geometric constructions used to create symmetric, convex polyhedra for the target structures: SC, β-Mn, diamond, β-W and hP2-X in step two.

Structure	α	β	γ	δ
BCC	0.476	0.194	N/A	N/A
FCC	0.341	0.318	N/A	N/A
β-Mn	0.263	N/A	N/A	N/A
SC	0.130	N/A	N/A	N/A
Diamond	0.392	0.111	0.350	N/A
β-W	0.564	0.486	0.122	0.081
hP2-X	0.451	0.608	0.487	0.043

Table 3.2: Optimal geometric parameters (see section 3.3.2 for parameter descriptions) from Alch-MC simulation for symmetric convex polyhedra in step two to self-assemble seven target structures.

ilies of planes that make up the faces of a rhombic dodecahedron, a cube, and an octahedron. There are truncating planes of two types: type *a* corresponding to the location of the cube faces, and type *c* which correspond to the position of the octahedron faces. We use α to represent the location of the truncating planes of type a and β to represent the location of the truncating planes of type c, both of which are linearly mapped to the interval between 0 and 1. (0, 0) is the cuboctahedron, (0, 1) is the cube, (1, 0) is the octahedron and (1, 1) is the rhombic dodecahedron. For a detailed mathematical construction and images of representative particle shapes see Ref. [15].

β-Mn

We truncate each vertex of a dodecahedron using planes with normals directed along a ray that passes from the geometric center of the dodecahedron through the vertex, truncating by an amount α between 0 and 1. The perfect dodecahedron has $\alpha = 1$, and $\alpha = 0$ when two truncating vertices meet. Representative particles from this shape family are shown in Fig. 3.6B.

SC

We first study the spheric triangle group $\Delta_{3,2,3}$[15] which is constructed with three families of planes that make up the faces of a cube, a tetrahedron, and an octahedron. As in the case of the $\Delta_{4,3,2}$ family, there are two shape parameters, *a* and *c*, which specify the

amount of truncation (or position of the bounding planes). For a detailed mathematical construction and images of representative particle shapes see Ref. [15].

We then study a one-parameter family of shapes formed by adding a vertex at each face center of a cube. Shape parameter $\alpha = 0$ describes a perfect cube and $\alpha = 1$ describes a perfect rhombic dodecahedron. Representative particles from this shape family are shown in Fig. 3.6A.

Diamond

We first study a one-parameter truncated tetrahedron family, where we truncate each vertex of a tetrahedron by an amount α that ranges between a perfect tetrahedron ($\alpha = 1$) and an octahedron ($\alpha = 0$). Representative particles from this shape family are shown in Fig. 3.6C (top).

We then study a three-parameter shape family formed by truncating the vertices of a tetrahedron and adding one more vertex to each face. Each augmenting vertex lies along a ray that passes from the geometric center of the tetrahedron through the center of a face. For this shape family, we again use α to parametrize the amount of truncation on each vertex of the tetrahedron. For the hexagonal faces of a truncated tetrahedron, we use shape parameter β to measure how far the augmenting vertex on the tetrahedron face away from the original face center. We use $\beta = 0$ to indicate that the augmenting vertex lies in the plane of the tetrahedron face (i.e. no augmentation), and $\beta = 1$ to indicate that the distance between the augmenting vertex and center of mass (CM) is twice the distance between the original face center and CM. For the triangular faces of the truncated tetrahedron, we use shape parameter γ to measure the distance between the augmenting vertex and CM. We use $\gamma = 0$ to indicate that the location of the augmenting vertex coincides with the vertex location of a regular (i.e. untruncated) tetrahedron and $\gamma = 1$ to indicate that the augmenting vertex lies at the CM. Representative particles from this shape family are

shown in Fig. 3.6C (bottom).

β-W

We first study a two-parameter family of asymmetrically truncated dodecahedral shapes. We divide the vertices of a dodecahedron into two groups of ten, with one group of vertices on the two parallel faces. Taking the convex hull of the remaining "side" vertices yields a pentagonal antiprism, with pentagonal faces parallel to the top and bottom faces of the dodecahedron. We use α to parametrize the truncation of vertices on the top and bottom faces by truncating planes that lie parallel to the top and bottom faces and are equidistant to the particle CM. We use $\alpha = 1$ to indicate the truncating planes are coplanar with the top and bottom faces of the dodecahedron, and we use $\alpha = 0$ to indicate the truncating planes lie halfway between the top and bottom faces and the pentagonal faces of the antiprism. We use β to parametrize the truncation of the side vertices. The truncations are formed by situating ten equidistant planes that have face normals parallel with passing through the particle CM and each side vertex. We use $\beta = 1$ to indicate no vertex truncation and $\beta = 0$ means two truncated vertices meet. Representative particles from this shape family are shown in Fig. 3.6E (top).

We then study a four-parameter shape family with two parameters describing the vertex truncations as indicated above. The other two parameters describe vertex augmentation. We use γ to parametrize vertex augmentation of the side faces. We use $\gamma = 0$ to indicate the augmenting vertex lies in the plane of the face (i.e. no vertex augmentation) and $\gamma = 1$ to indicate the distance between the augmenting vertex and CM is twice the distance between original face center and CM. We use δ to parametrize the position vertices that augment the top and bottom faces along rays that pass from the center of the particle through the center of the top and bottom faces. We use $\delta = 0$ to indicate the the augmenting vertex lies at the initial pentagon face center and $\delta = 1$ to indicate the augmenting

vertex is at the CM. Representative particles from this shape family are shown in Fig. 3.6E (bottom).

hP2-X

This hypothetical structure is a derivative of the hexagonally close-packed structure (HCP). The structures are very similar from a crystallographic viewpoint: both exhibit space group $P6_3/mmc$ (space group no. 194) and in both of them, one Wyckoff site – $2c$ $1/3, 2/3, 1/4$ – is occupied. However, the ratio of unit cell parameters c/a differs substantially between these two cases. While for the close-packing of spheres $c/a = \sqrt{8/3} \approx 1.633$, this is a free parameter from a symmetry point of view. For the structure used here, we chose a much lower value of $c/a = 0.639$.

This variant of the crystallographically identical unit cell exhibits a different local particle environment, due to the fact that particles can move closer together along the c-direction and therefore are farther apart, relatively, in the a-b-plane. This leads to a coordination number of 8 nearest neighbors – compared with 12 in HCP – and a coordination polyhedron with the shape of a biaugmented hexagonal prism.

We first study a two-parameter family of asymmetrically truncated bipyramid shapes. The dihedral angle between the upper and lower face of the bipyramid is 122.049 degrees. We divide the vertices of a bipyramid into two groups, with one group of vertices at the top and bottom, and another group of three vertices on the side. We use α to parametrize the truncation of the top and bottom vertices. We use $\alpha = 1$ to indicate no vertex truncation, and $\alpha = 0$ to indicate the vertices are truncated to the side vertex position. We use β to parametrize the truncation of the side vertices. The truncations are formed by situating three equidistant planes that have face normals parallel with passing through the particle CM and each side vertex. We use $\beta = 1$ to indicate no vertex truncation and $\beta = 0$ means two truncated vertices meet. Representative particles from this shape family are shown in

Fig. 3.6D (left).

We then study a four-parameter shape family with two parameters describing the vertex truncations as indicated above. The other two parameters describe vertex augmentation. We use γ to parametrize the position vertices that augment the top and bottom faces along rays that pass from the center of the particle through the top and bottom vertices. We use $\gamma = 0$ to indicate the the augmenting vertex lies at the initial top and bottom vertices and $\gamma = 1$ to indicate the augmenting vertex is at the particle CM. We use δ to parametrize vertex augmentation of the side faces. We chose the augmenting vertex be the side face center of shape with $\alpha = 0.435, \beta = 0.52, \gamma = 0.503$. We use $\delta = 0$ to indicate the augmenting vertex lies in the plane of the face (i.e. no vertex augmentation) and $\delta = 1$ to indicate the distance between the augmenting vertex and CM is twice the distance between the original augmenting vertex and CM. Representative particles from this shape family are shown in Fig. 3.6D (right).

3.4   Optimal Shapes for a Novel Structure

We targeted the self-assembly of a hypothetical structure with no known atomic or other equivalent. The structure is a modified version of the hexagonally-close packed (hcp) structure with distorted lattice spacing, so that particles have eight nearest neighbors, see Fig. 3.7A, whereas hcp has 12. We denote this structure as hP2-X. Alch-MC simulations of convex polyhedra with 116 vertex parameters in step one yielded the faceted shape shown in Fig. 3.7B (left). A symmetrized free energy minimizing shape was then found in step two (Fig. 3.7B right). The details about the construction of symmetric shape family in step two can be found in section 3.3.2. We tested that both the unsymmetrized and symmetrized optimal shapes spontaneously self-assembled the target structure from a disordered fluid, with the resulting structure shown in Fig. 3.7C. This demonstrates the inverse design of a
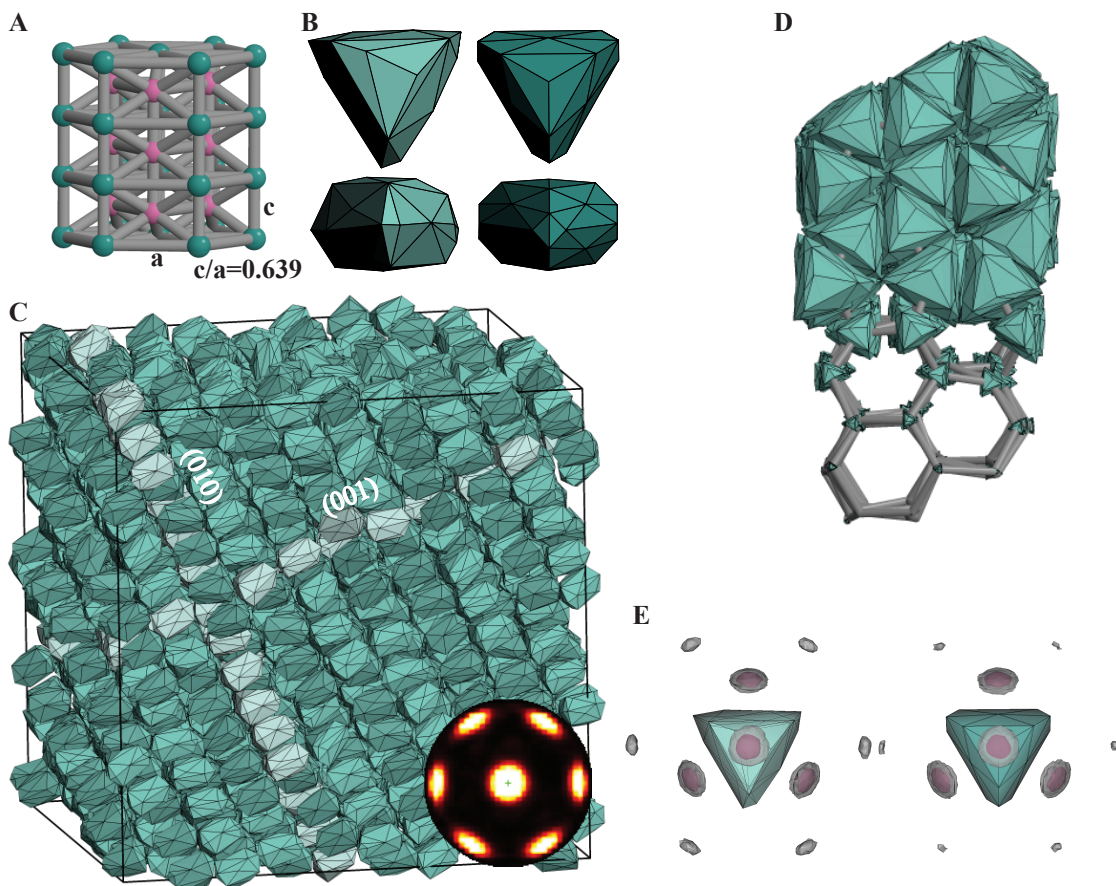
Figure 3.7: Alch-MC design and self-assembly of a previously unreported novel crystal structure with no known atomic equivalent. **(A)** The structure hP2-X is a distorted version of HCP with 8 rather than 12 nearest neighbors. Alch-MC simulation produces a particle (**(B)**) that spontaneously self-assembles the target structure (**(C)**) in simulation. Inset is a bond order diagram of the structure. **(D)** Particle organization relative to lattice directions. **(E)** PMFT isosurface for optimal shapes.

colloidal particle shape to entropically self assemble a previously unknown target structure using only digital alchemy [93].

3.5 Successful Self-assembly of Optimal Shapes

For the structures BCC, FCC, β-Mn, SC, diamond, β-W and HP2-X, we confirmed that the successful self-assembly of each target structure using a geometric ansatz where particles have the shape of the Voronoi cell of the target crystal structure, a near optimal unsymmetrized shape from Alch-MC simulation in step one, and the symmetric optimal shape
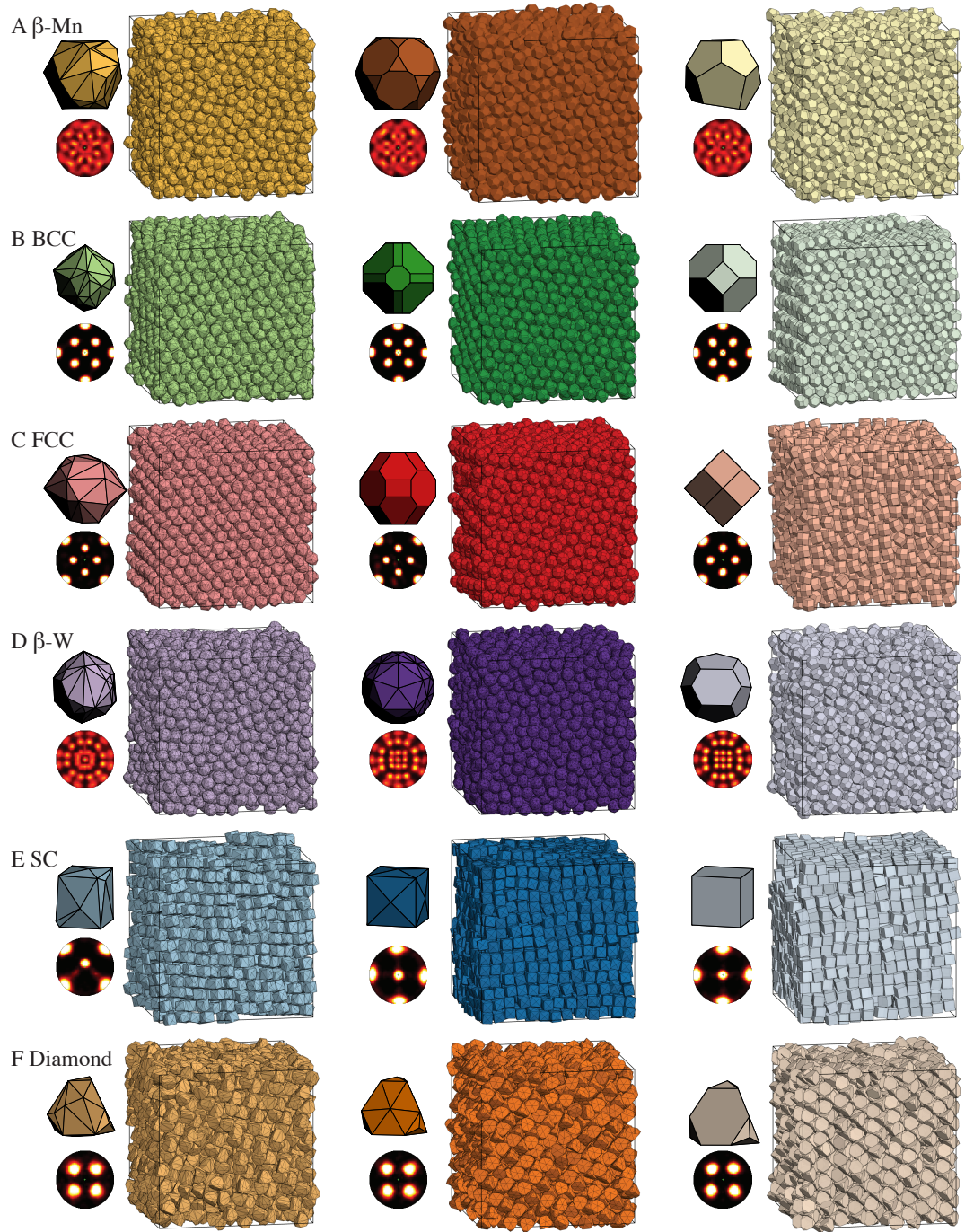
Figure 3.8: Successful self-assembly from disordered fluid. Representative system snapshots indicating the successful self-assembly in NVT MC simulations of near-optimal convex polyhedra obtained from Alch-MC simulation in step one (left column), NVT MC simulation of optimal symmetric convex polyhedra obtained from Alch-MC simulation in step two (center column), and geometric ansatz (right column) for six structures at packing fraction $\eta = 0.6$. Particle images and bond-order diagrams are on the left.

from Alch-MC simulation in step two. See Fig. 3.8 for representative system snapshots. Self-assembly validation was performed using the Hard Particle Monte Carlo (HPMC) [5] plugin for HOOMD-Blue [4]. We simulated in the NVT ensemble with 2197 particles for all structures at packing fraction $\eta = 0.6$, with between 4×10^7 and a maximum of 1×10^9 Monte Carlo sweeps. We identified an assembled crystal structure by computing bond order diagrams and diffraction patterns using particle centroids, following the approach used in Ref. [20].

3.6   Discussion

Particle shape has, in principle, an infinite-dimensional parameter space. Here, for tractability, and motivated by shapes that can be realized using nanoparticle synthesis techniques, we searched for optimal particle shapes over 92- and 188-dimensional parameter spaces of convex shapes, using a precisely defined entropic design criterion. Our method yields not only thermodynamically optimal particle shapes, but also distributions of candidate shapes that provide insight into the sensitivity of target structures to shape features (Fig. 3.3). Emergent entropic valence that is commensurate with the emergence of faceting in an ensemble of arbitrary convex polyhedra, both of which are, in turn, commensurate with local structural coordination, is a strong indication in favor of the hypothesized connection between faceting, emergent directional entropic forces, and structural order [92, 91]. By consistently establishing the connection between the emergence of faceting and entropic valence, our results suggest future work could assume this connection, and either skip our intermediate step of facet characterization by reading particle faceting directly from PMFT measurements, and/or rather than working agnostically, start the Alch-MC shape evolution from a Voronoi cell shape.

CHAPTER IV

Predicting Colloidal Crystals from Shapes via Inverse Design and Machine Learning

This chapter is adapted from Ref. [32], and is available online: "Predicting colloidal crystals from shapes via inverse design and machine learning", Yina Geng, Greg van Anders, Sharon C. Glotzer, *arXiv:1801.06219*.

A fundamental challenge in materials design is linking building block attributes to crystal structure. Addressing this challenge is particularly difficult for systems that exhibit emergent order, such as entropy-stabilized colloidal crystals. Here we combine two recent computational advances to build a prediction model to predict crystal structures from building block geometry features. One advance is the inverse design methodology of digital alchemy [93, 31], a molecular simulation method in which particle attributes (e.g., particle shapes) are treated as thermodynamic variables in a generalized, or extended, thermodynamic ensemble [59, 101]. Digital alchemy simulations produce optimal particle attributes for target thermodynamic phases, including complex colloidal crystals with very large unit cells [31]. The other advance is the application of machine learning methods to materials problems [63, 33, 2, 57, 56, 76]. Machine learning can discover hidden correlations in large datasets, providing clues to the long-sought relationship between building block and structure. By combining these two approaches to thermodynamic systems of hard, polyhedrally shaped particles, we find an empirical but highly predictive model for

28

entropically ordered colloidal crystals from solely geometric measures of their constituent particles. Our model is capable of predicting 13 different entropically stable crystal structures formed by millions of different colloidal polyhedra with 98% fidelity using just two geometric measures of particle shape (see Fig. 4.1 for an illustration of our approach). Though far from a first principles theory, this model can be used immediately to inform experiments and select building blocks for self-assembling nanoparticle superlattices and colloidal crystals.

4.1 Optimal Shapes via Digital Alchemy

To construct the predictive model we performed Alchemical Monte Carlo (Alch-MC) simulations based on the Digital Alchemy framework [93], using an implementation [31] that extends an open-source Monte Carlo plugin, Hard Particle Monte Carlo (HPMC) [5], for the open-source molecular dynamics package HOOMD-Blue [4] to generalized thermodynamic ensembles that include particle shape change. Details about the digital alchemy method is available in section 2.2. We simulated $NVT\mu$ ensembles at constant temperature T, fixed volume V, and alchemical potential $\mu = 0$ for 13 target structures reported previously in entropic self-assembly: simple cubic (SC), body-centered cubic (BCC), $hP2$-X, face-centered cubic (FCC), simple chiral cubic (SCC), hexagonal (HEX-1-0.6), diamond (D), graphite (G), honeycomb (H), body-centered tetragonal (BCT-1-1-2.4), high-pressure Lithium (Li), β-Manganese (β-Mn), and β-Tungsten (β-W). $hP2$-X is a hypothetical structure with no known atomic or other equivalent. The structure is a modified version of the hexagonally-close packed (HCP) structure with distorted lattice spacing, so that particles have eight nearest neighbors, whereas HCP has 12 [31]. The variable μ is conjugate to the shape variable that is allowed to fluctuate in the simulation. We placed a minimum of $N = 100$ particles in a periodic simulation box, with the exact
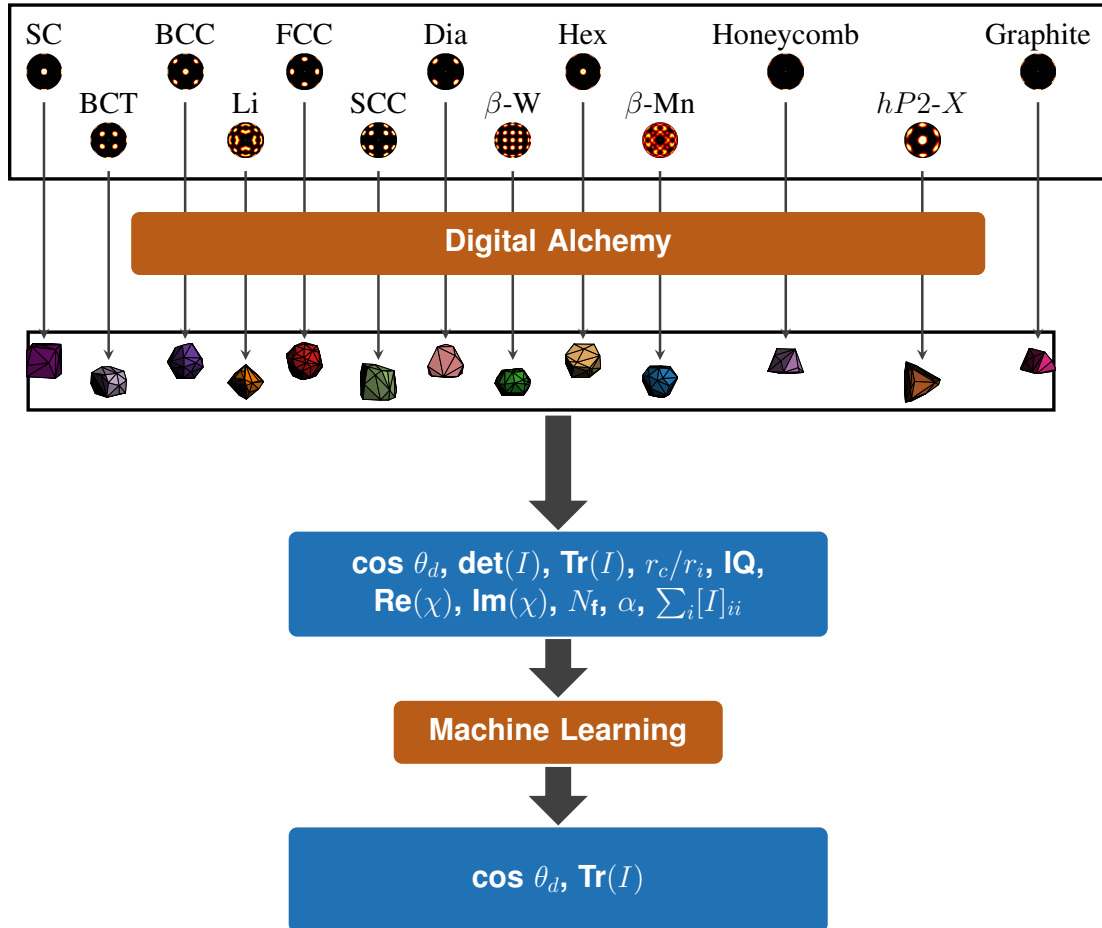
Figure 4.1: We use the Digital Alchemy inverse materials design approach to find optimal and near-optimal hard, convex, colloidal, polyhedral shapes for 13 target structures. We use the Random Forest technique from machine learning to classify shapes. We find that, of 10 measures of shape, two – the dihedral angle ($\cos(\theta_d)$) and the trace of the moment of inertia tensor ($\text{Tr}(I)$) – are sufficient to predict the self-assembly behavior of a shape.

number chosen (for convenience) to be a multiple of the number of particles in the unit cell of one of the 13 target structures. Particle shapes were initialized with as many as 64 vertices randomly generated to create a convex shape. Monte Carlo (MC) sweeps were performed to allow particle translations, rotations, and shape moves via vertex re-location. For each shape move, we (i) moved a vertex, (ii) resized the trial shape to unit volume, (iii) checked if the move induced any particle overlaps, and then (iv) accepted the move based on the Boltzmann factor as described in Ref. [93] and rescaled the trial move domain in the same way as the particle volume overall. Translation and rotation moves followed standard procedures (see, e.g., Refs. [40, 1, 30, 19, 20, 92, 91]).

We slowly compressed the target crystal structure comprised of a randomly generated shape to the target packing fraction, with springs of spring constant 1000 (where energy is specified in units of $k_{\mathrm{B}}T$ and length units are set by the particle size) at each node to maintain the integrity of the structure during the compression phase. In this initial stage, it is highly unlikely that the structure is thermodynamically stable, so fixing the positions of the centers of the particles over an initial set of MC steps allows the system to explore shape space during compression without falling apart. After reaching the target packing fraction, we logarithmically relaxed the spring constant to zero. We then further evolved the system at fixed packing fraction η for $10^6 - 10^7$ MC sweeps, depending on the target crystal structure (Table 4.1). For each target crystal structure, we performed 10 independent simulations and randomly selected and analyzed 100 shapes in the final 5×10^5 sweeps. The output of the Alch-MC simulation procedure for each of the 13 crystal structures is a family of optimal shapes (that is, shapes that minimize the system free energy) with shape measures that fluctuate about some average value.

Target struc-ture	packing fraction η	MC sweeps	Target struc-ture	packing fraction η	MC sweeps
BCC	0.6	1×10^6	graphite	0.65	1×10^6
FCC	0.6	1×10^6	BCT-1-1-2.4	0.7	1×10^6
SC	0.6	1×10^6	Li	0.65	1×10^6
diamond	0.6	1×10^6	SCC	0.7	2×10^6
$hP2$-X	0.6	1×10^6	β-W	0.6	8×10^6
honeycomb	0.65	1×10^6	β-Mn	0.6	3.6×10^7
HEX-1-0.6	0.7	1×10^6			

Table 4.1: In Alch-MC, packing fraction and MC sweeps for 13 target crystal structures.

4.2 Geometric Features Calculation

We calculated 10 measures of shape motivated by prior works studying the relationship between particle geometry and self-assembly behavior.[20] The shape measures calculated for each polyhedral shape are: the cosine of the average dihedral angle $\cos\theta_d$; the number of facets N_f; the determinant of the moment of inertia tensor $\det(I)$; the trace of the moment of inertia tensor $\text{Tr}(I)$; the sum of principal minors of the moment of inertia tensor $\sum_i [I]_{ii}$; the ratio of the circumsphere radius to the insphere radius r_c/r_i; IQ; the asphericity α; and the real and imaginary parts of the chiral parameter χ ($\text{Re}(\chi)$, $\text{Im}(\chi)$). To calculate $\cos\theta_d$, facets with area $a_f > a_f^*$ (we use $a_f^* = 0.02$ but our results are not sensitive to changes in a_f^*) were clustered by their normal vector using DBSCAN [25], a standard clustering algorithm implemented in the open-source Python machine learning package scikit-learn [74]. Clustered facets are represented by area-weighted average normals. We computed the clustered-facets-area-weighted cosine of the angle between average normals of neighboring clustered facets in a polyhedron. The equation is $\cos\theta_d = \sum_{i,j}(s_i + s_j)|\cos(\theta_{d_{ij}})|/\sum_{i,j}(s_i + s_j)$, where s_i is the area of ith clustered facets (sum of component facets), $\cos(\theta_{d_{ij}})$ is cosine of the dihedral angle between clustered facet i and j, and the sum is over all adjacent clustered facets i and j (two

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\text{Re}(\chi)$	$\text{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\text{Tr}(I)$
$\cos\theta_d$	1	0.03	0.02	-0.1	0.01	-0.09	0	0.08	0.02	0.02
r_c/r_i	0.03	1	0.51	-0.23	0.04	-0.06	-0.06	0.3	0.52	0.54
$\det(I)$	0.02	0.51	1	-0.51	0.09	-0.03	-0.07	0.58	1	0.99
IQ	-0.1	-0.23	-0.51	1	-0.02	0.27	-0.31	-0.97	-0.5	-0.5
$\text{Re}(\chi)$	0.01	0.04	0.09	-0.02	1	-0.03	0	0.02	0.08	0.08
$\text{Im}(\chi)$	-0.09	-0.06	-0.03	0.27	-0.03	1	-0.14	-0.22	-0.03	-0.03
N_f	0	-0.06	-0.07	-0.31	0	-0.14	1	0.21	-0.06	-0.06
α	0.08	0.3	0.58	-0.97	0.02	-0.22	0.21	1	0.58	0.57
$\sum_i[I]_{ii}$	0.02	0.52	1	-0.5	0.08	-0.03	-0.06	0.58	1	1
$\text{Tr}(I)$	0.02	0.54	0.99	-0.5	0.08	-0.03	-0.06	0.57	1	1

Table 4.2: Correlation matrix of 10 geometric measures for the β-Mn structure. A cell is blue when the magnitude of the correlation ≤ 0.2, and a cell is orange when the magnitude of the correlation $\geq$ 0.8 (same for Table 4.3 - 4.14).

clustered facets are adjacent if they have two adjacent component facets). The number of facets is the number of clustered facets using DBSCAN. The isoperimetric quotient $IQ = 36\pi v^2/s^3$, where v is the volume and s is the surface area of a polyhedron). The asphericity $\alpha = Rs/3v$, where R is the integrated mean curvature normalized by 4π. Finally, χ is the lowest order measure of the degree of chirality of a molecule [41] and it is zero for achiral molecules. $\chi \propto \sum_{mn} C(234; mn)\rho_{2m}\rho_{3n}\rho^*_{4,m+n}$, where $C(234; mn)$ are the appropriate Clebsch-Gordan coefficients and ρ_{lm} are mass-weighted distance moments with $\rho_{lm} = \sum_{\tau \in T} |\mathbf{r}_\tau|^l Y_{lm}(\theta_\tau, \phi_\tau)$, where the sum is over atoms τ in the molecule T. We calculated χ for a polyhedron by assuming a unit mass atom at each vertex of its vertices.

Table 4.2 reports correlations found for the β-Mn structure between the 10 tested geometric measures applied to 1000 free-energy minimizing shapes randomly selected from among the millions of shapes generated by the Alch-MC simulations. Tables reporting correlations for all other structures studied can be found in Tables 4.3- 4.14.

4.3 Random Forest Method

The geometric feature dataset is to be interrogated using machine learning (ML), with the aim of identifying salient geometric criteria for predicting structures. We utilized the

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\mathrm{Re}(\chi)$	$\mathrm{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\mathrm{Tr}(I)$
$\cos\theta_d$	1	0.02	-0.1	0.2	0.02	0.07	-0.05	-0.2	-0.1	-0.09
r_c/r_i	0.02	1	0.55	-0.44	-0.03	-0.07	-0.05	0.45	0.55	0.56
$\det(I)$	-0.1	0.55	1	-0.91	0.03	0.1	-0.08	0.89	1	1
IQ	0.2	-0.44	-0.91	1	-0.06	-0.13	0.03	-0.96	-0.9	-0.9
$\mathrm{Re}(\chi)$	0.02	-0.03	0.03	-0.06	1	-0.02	-0.03	0.05	0.03	0.03
$\mathrm{Im}(\chi)$	0.07	-0.07	0.1	-0.13	-0.02	1	-0.12	0.11	0.1	0.1
N_f	-0.05	-0.05	-0.08	0.03	-0.03	-0.12	1	-0.03	-0.08	-0.08
α	-0.2	0.45	0.89	-0.96	0.05	0.11	-0.03	1	0.89	0.88
$\sum_i[I]_{ii}$	-0.1	0.55	1	-0.9	0.03	0.1	-0.08	0.89	1	1
$\mathrm{Tr}(I)$	-0.09	0.56	1	-0.9	0.03	0.1	-0.08	0.88	1	1

Table 4.3: Correlation coefficient between 10 geometric measures for the SC structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\mathrm{Re}(\chi)$	$\mathrm{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\mathrm{Tr}(I)$
$\cos\theta_d$	1	0	-0.02	0.02	0.03	0.08	0.09	0.02	-0.02	-0.02
r_c/r_i	0	1	0.41	-0.35	0.04	0	-0.03	0.44	0.41	0.41
$\det(I)$	-0.02	0.41	1	-0.82	0.02	0.03	0.07	0.79	1	1
IQ	0.02	-0.35	-0.82	1	-0.01	0.01	-0.08	-0.95	-0.81	-0.81
$\mathrm{Re}(\chi)$	0.03	0.04	0.02	-0.01	1	0.03	0.02	0.01	0.02	0.02
$\mathrm{Im}(\chi)$	0.08	0	0.03	0.01	0.03	1	0.02	-0.01	0.03	0.02
N_f	0.09	-0.03	0.07	-0.08	0.02	0.02	1	0	0.07	0.07
α	0.02	0.44	0.79	-0.95	0.01	-0.01	0	1	0.79	0.79
$\sum_i[I]_{ii}$	-0.02	0.41	1	-0.81	0.02	0.03	0.07	0.79	1	1
$\mathrm{Tr}(I)$	-0.02	0.41	1	-0.81	0.02	0.02	0.07	0.79	1	1

Table 4.4: Correlation coefficient between 10 geometric measures for the BCC structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\mathrm{Re}(\chi)$	$\mathrm{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\mathrm{Tr}(I)$
$\cos\theta_d$	1	0.06	-0.04	0.05	0.01	0.01	0.39	0.03	-0.04	-0.04
r_c/r_i	0.06	1	0.64	-0.47	0	0.12	-0.19	0.5	0.64	0.65
$\det(I)$	-0.04	0.64	1	-0.88	-0.02	0.15	-0.43	0.88	1	1
IQ	0.05	-0.47	-0.88	1	0.03	-0.17	0.56	-0.98	-0.87	-0.86
$\mathrm{Re}(\chi)$	0.01	0	-0.02	0.03	1	-0.03	0.03	-0.03	-0.02	-0.02
$\mathrm{Im}(\chi)$	0.01	0.12	0.15	-0.17	-0.03	1	-0.08	0.17	0.16	0.16
N_f	0.39	-0.19	-0.43	0.56	0.03	-0.08	1	-0.53	-0.43	-0.42
α	0.03	0.5	0.88	-0.98	-0.03	0.17	-0.53	1	0.87	0.86
$\sum_i[I]_{ii}$	-0.04	0.64	1	-0.87	-0.02	0.16	-0.43	0.87	1	1
$\mathrm{Tr}(I)$	-0.04	0.65	1	-0.86	-0.02	0.16	-0.42	0.86	1	1

Table 4.5: Correlation coefficient between 10 geometric measures for the FCC structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\text{Re}(\chi)$	$\text{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\text{Tr}(I)$
$\cos\theta_d$	1	0.38	0.06	0.26	-0.02	-0.03	0.35	-0.14	-0.12	-0.24
r_c/r_i	0.38	1	0.23	0	-0.02	0.2	-0.01	0.14	0.09	-0.02
$\det(I)$	0.06	0.23	1	-0.65	0.02	-0.2	0.03	0.66	0.95	0.87
IQ	0.26	0	-0.65	1	-0.05	-0.05	0.09	-0.96	-0.61	-0.55
$\text{Re}(\chi)$	-0.02	-0.02	0.02	-0.05	1	0.06	0.02	0.03	0.05	0.06
$\text{Im}(\chi)$	-0.03	0.2	-0.2	-0.05	0.06	1	-0.06	0.13	-0.2	-0.19
N_f	0.35	-0.01	0.03	0.09	0.02	-0.06	1	-0.08	0.05	0.06
α	-0.14	0.14	0.66	-0.96	0.03	0.13	-0.08	1	0.57	0.48
$\sum_i[I]_{ii}$	-0.12	0.09	0.95	-0.61	0.05	-0.2	0.05	0.57	1	0.98
$\text{Tr}(I)$	-0.24	-0.02	0.87	-0.55	0.06	-0.19	0.06	0.48	0.98	1

Table 4.6: Correlation coefficient between 10 geometric measures for the SCC structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\text{Re}(\chi)$	$\text{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\text{Tr}(I)$
$\cos\theta_d$	1	-0.04	-0.05	0.09	0.02	-0.07	0.24	-0.1	-0.05	-0.04
r_c/r_i	-0.04	1	0.61	-0.44	-0.01	0.06	-0.05	0.45	0.6	0.6
$\det(I)$	-0.05	0.61	1	-0.58	0.02	0.13	-0.06	0.57	1	1
IQ	0.09	-0.44	-0.58	1	0	-0.13	0.21	-0.99	-0.55	-0.51
$\text{Re}(\chi)$	0.02	-0.01	0.02	0	1	-0.03	0.02	0	0.02	0.02
$\text{Im}(\chi)$	-0.07	0.06	0.13	-0.13	-0.03	1	0.05	0.09	0.13	0.13
N_f	0.24	-0.05	-0.06	0.21	0.02	0.05	1	-0.22	-0.05	-0.04
α	-0.1	0.45	0.57	-0.99	0	0.09	-0.22	1	0.54	0.51
$\sum_i[I]_{ii}$	-0.05	0.6	1	-0.55	0.02	0.13	-0.05	0.54	1	1
$\text{Tr}(I)$	-0.04	0.6	1	-0.51	0.02	0.13	-0.04	0.51	1	1

Table 4.7: Correlation coefficient between 10 geometric measures for the HEX structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\text{Re}(\chi)$	$\text{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\text{Tr}(I)$
$\cos\theta_d$	1	-0.02	-0.01	-0.04	0	-0.01	-0.07	0.04	-0.01	-0.01
r_c/r_i	-0.02	1	0.42	-0.29	0.03	-0.14	0.05	0.39	0.43	0.43
$\det(I)$	-0.01	0.42	1	-0.72	0.03	0.04	0.01	0.77	1	1
IQ	-0.04	-0.29	-0.72	1	0.02	-0.1	-0.01	-0.97	-0.71	-0.71
$\text{Re}(\chi)$	0	0.03	0.03	0.02	1	-0.06	0.03	0	0.03	0.03
$\text{Im}(\chi)$	-0.01	-0.14	0.04	-0.1	-0.06	1	0.04	0.08	0.04	0.04
N_f	-0.07	0.05	0.01	-0.01	0.03	0.04	1	0	0.01	0.01
α	0.04	0.39	0.77	-0.97	0	0.08	0	1	0.77	0.77
$\sum_i[I]_{ii}$	-0.01	0.43	1	-0.71	0.03	0.04	0.01	0.77	1	1
$\text{Tr}(I)$	-0.01	0.43	1	-0.71	0.03	0.04	0.01	0.77	1	1

Table 4.8: Correlation coefficient between 10 geometric measures for the diamond structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\text{Re}(\chi)$	$\text{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\text{Tr}(I)$
$\cos\theta_d$	1	0.07	0.18	-0.2	0.02	-0.01	0	0.14	0.18	0.18
r_c/r_i	0.07	1	0.33	-0.43	0.01	-0.2	0.03	0.64	0.18	0.13
$\det(I)$	0.18	0.33	1	-0.61	0.02	-0.04	-0.02	0.48	0.96	0.94
IQ	-0.2	-0.43	-0.61	1	0	0.15	0	-0.86	-0.44	-0.39
$\text{Re}(\chi)$	0.02	0.01	0.02	0	1	-0.11	-0.01	0	0.03	0.03
$\text{Im}(\chi)$	-0.01	-0.2	-0.04	0.15	-0.11	1	0.01	-0.2	-0.01	0.01
N_f	0	0.03	-0.02	0	-0.01	0.01	1	0.03	-0.03	-0.04
α	0.14	0.64	0.48	-0.86	0	-0.2	0.03	1	0.27	0.2
$\sum_i[I]_{ii}$	0.18	0.18	0.96	-0.44	0.03	-0.01	-0.03	0.27	1	1
$\text{Tr}(I)$	0.18	0.13	0.94	-0.39	0.03	0.01	-0.04	0.2	1	1

Table 4.9: Correlation coefficient between 10 geometric measures for the graphite structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\text{Re}(\chi)$	$\text{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\text{Tr}(I)$
$\cos\theta_d$	1	0.05	0.03	-0.09	-0.04	-0.02	-0.03	0.09	0.03	0.03
r_c/r_i	0.05	1	0.51	-0.37	-0.01	-0.04	0.02	0.39	0.5	0.49
$\det(I)$	0.03	0.51	1	-0.71	-0.02	-0.05	-0.01	0.72	0.99	0.96
IQ	-0.09	-0.37	-0.71	1	0.02	0.21	0.08	-0.94	-0.63	-0.58
$\text{Re}(\chi)$	-0.04	-0.01	-0.02	0.02	1	0.05	-0.05	-0.02	-0.01	-0.01
$\text{Im}(\chi)$	-0.02	-0.04	-0.05	0.21	0.05	1	0.17	-0.21	-0.02	-0.01
N_f	-0.03	0.02	-0.01	0.08	-0.05	0.17	1	-0.16	0.02	0.03
α	0.09	0.39	0.72	-0.94	-0.02	-0.21	-0.16	1	0.64	0.58
$\sum_i[I]_{ii}$	0.03	0.5	0.99	-0.63	-0.01	-0.02	0.02	0.64	1	1
$\text{Tr}(I)$	0.03	0.49	0.96	-0.58	-0.01	-0.01	0.03	0.58	1	1

Table 4.10: Correlation coefficient between 10 geometric measures for the honeycomb structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\text{Re}(\chi)$	$\text{Im}(\chi)$	N_f	α	$\sum_i[I]_{ii}$	$\text{Tr}(I)$
$\cos\theta_d$	1	0.29	-0.21	0.72	-0.01	0.11	0.7	-0.8	0.1	0.27
r_c/r_i	0.29	1	0.33	0.17	-0.1	0.08	0.12	-0.17	0.47	0.52
$\det(I)$	-0.21	0.33	1	-0.52	-0.01	-0.01	-0.06	0.49	0.94	0.85
IQ	0.72	0.17	-0.52	1	-0.02	0.19	0.38	-0.98	-0.21	-0.03
$\text{Re}(\chi)$	-0.01	-0.1	-0.01	-0.02	1	-0.1	0	0.02	-0.02	-0.03
$\text{Im}(\chi)$	0.11	0.08	-0.01	0.19	-0.1	1	0.02	-0.19	0.05	0.08
N_f	0.7	0.12	-0.06	0.38	0	0.02	1	-0.46	0.12	0.22
α	-0.8	-0.17	0.49	-0.98	0.02	-0.19	-0.46	1	0.17	-0.03
$\sum_i[I]_{ii}$	0.1	0.47	0.94	-0.21	-0.02	0.05	0.12	0.17	1	0.98
$\text{Tr}(I)$	0.27	0.52	0.85	-0.03	-0.03	0.08	0.22	-0.03	0.98	1

Table 4.11: Correlation coefficient between 10 geometric measures for the BCT-1-1-2.4 structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\mathrm{Re}(\chi)$	$\mathrm{Im}(\chi)$	N_f	α	$\sum_i [I]_{ii}$	$\mathrm{Tr}(I)$
$\cos\theta_d$	1	0	0.01	0.06	0.08	0	0.29	-0.02	0.01	0.01
r_c/r_i	0	1	0.53	-0.44	0.03	0.07	-0.13	0.51	0.54	0.54
$\det(I)$	0.01	0.53	1	-0.87	-0.03	0.03	-0.13	0.81	1	1
IQ	0.06	-0.44	-0.87	1	0.04	-0.04	0.14	-0.89	-0.87	-0.87
$\mathrm{Re}(\chi)$	0.08	0.03	-0.03	0.04	1	-0.02	0.09	-0.03	-0.04	-0.04
$\mathrm{Im}(\chi)$	0	0.07	0.03	-0.04	-0.02	1	-0.04	0.1	0.02	0.02
N_f	0.29	-0.13	-0.13	0.14	0.09	-0.04	1	-0.13	-0.14	-0.14
α	-0.02	0.51	0.81	-0.89	-0.03	0.1	-0.13	1	0.8	0.8
$\sum_i [I]_{ii}$	0.01	0.54	1	-0.87	-0.04	0.02	-0.14	0.8	1	1
$\mathrm{Tr}(I)$	0.01	0.54	1	-0.87	-0.04	0.02	-0.14	0.8	1	1

Table 4.12: Correlation coefficient between 10 geometric measures for the Li structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\mathrm{Re}(\chi)$	$\mathrm{Im}(\chi)$	N_f	α	$\sum_i [I]_{ii}$	$\mathrm{Tr}(I)$
$\cos\theta_d$	1	0.14	0.03	0.07	0.03	0.13	-0.12	-0.05	0.05	0.07
r_c/r_i	0.14	1	0.79	-0.7	0.01	0.09	-0.04	0.74	0.8	0.8
$\det(I)$	0.03	0.79	1	-0.87	0.01	0.07	-0.08	0.9	1	1
IQ	0.07	-0.7	-0.87	1	0.01	-0.04	0.12	-0.99	-0.85	-0.84
$\mathrm{Re}(\chi)$	0.03	0.01	0.01	0.01	1	-0.04	0.04	0	0.01	0.01
$\mathrm{Im}(\chi)$	0.13	0.09	0.07	-0.04	-0.04	1	0.07	0.04	0.08	0.08
N_f	-0.12	-0.04	-0.08	0.12	0.04	0.07	1	-0.13	-0.08	-0.08
α	-0.05	0.74	0.9	-0.99	0	0.04	-0.13	1	0.89	0.87
$\sum_i [I]_{ii}$	0.05	0.8	1	-0.85	0.01	0.08	-0.08	0.89	1	1
$\mathrm{Tr}(I)$	0.07	0.8	1	-0.84	0.01	0.08	-0.08	0.87	1	1

Table 4.13: Correlation coefficient between 10 geometric measures for the $hP2$-X structure.

Correlation	$\cos\theta_d$	r_c/r_i	$\det(I)$	IQ	$\mathrm{Re}(\chi)$	$\mathrm{Im}(\chi)$	N_f	α	$\sum_i [I]_{ii}$	$\mathrm{Tr}(I)$
$\cos\theta_d$	1	0.01	0.01	0.19	-0.02	0.01	-0.11	-0.17	0.02	0.03
r_c/r_i	0.01	1	0.58	-0.41	0.05	0.05	-0.01	0.47	0.58	0.57
$\det(I)$	0.01	0.58	1	-0.47	0.05	-0.14	0.08	0.44	1	0.99
IQ	0.19	-0.41	-0.47	1	-0.06	-0.06	-0.15	-0.94	-0.43	-0.4
$\mathrm{Re}(\chi)$	-0.02	0.05	0.05	-0.06	1	-0.01	-0.01	0.07	0.05	0.05
$\mathrm{Im}(\chi)$	0.01	0.05	-0.14	-0.06	-0.01	1	-0.09	0.07	-0.15	-0.16
N_f	-0.11	-0.01	0.08	-0.15	-0.01	-0.09	1	0.09	0.08	0.07
α	-0.17	0.47	0.44	-0.94	0.07	0.07	0.09	1	0.4	0.37
$\sum_i [I]_{ii}$	0.02	0.58	1	-0.43	0.05	-0.15	0.08	0.4	1	1
$\mathrm{Tr}(I)$	0.03	0.57	0.99	-0.4	0.05	-0.16	0.07	0.37	1	1

Table 4.14: Correlation coefficient between 10 geometric measures for the β-W structure.

random forest ML classification technique in scikit-learn. Random forest is an ensemble method that combines a collection of individual decision trees, each trained on a separate bootstrap sample of the input data [13, 12]. The prediction of random forest is the majority vote of the predictions of all trees. The aggregate model is capable of high predictive accuracy and is applicable in high-dimensional problems with highly correlated variables. We use gini impurity [13] as a measure of feature relevance in decision trees. There are two parameters that influence the performance of random forest: number of decision trees $ntree$ and number of features $maxfeature$ considered for each tree split. We used the optimal value for $ntree = 10$ as obtained from a cross-validation analysis [29] and used the recommended $maxfeature = \sqrt{numberoffeatures}$ [74]. Here the 10 shape measures for 1000 shapes in each crystal structure are provided as input features and the 13 crystal structures are produced as output.

4.4 The Prediction Model

Table 4.2 shows that some shape measures are highly correlated, e.g. α and IQ. When strong correlation exists between features, the feature selection method in random forest is unstable and less able to detect the most relevant features [37, 6, 87]. Thus we did not use the feature selection method for random forest in scikit-learn. Instead, we built random forest models with all possible combinations of shape features and computed the 10-fold cross-validation accuracy to see which feature combination gives the best predictive ability. The ML analysis (see Fig. 4.2) of low density data indicates that two shape features – $\mathrm{Tr}(I)$ and $\cos(\theta_d)$ – are the strongest predictors of the shape–structure relationship. The two-parameter model predicts structure from shape with 98% accuracy (using 10-fold cross-validation, see Fig. 4.2 orange point). Fig. 4.3 shows shape distributions for the 13 candidate structures using $\mathrm{Tr}(I)$ and $\cos(\theta_d)$. In the correlation tables of the

ten geometric features (Table 4.2 - 4.14), a cell is blue when the magnitude of the correlation ≤ 0.2, and a cell is orange when the magnitude of the correlation ≥ 0.8. The correlation tables have similar patterns, where $\cos(\theta_d)$ is uncorrelated with all other nine features. The three features $\text{Tr}(I)$, $det(\text{I})$ and $\sum_i [I]_{ii}$ are perfectly correlated with each other, and r_c/r_i, IQ, and α are highly correlated with the three former measures. This indicates that $\cos(\theta_d)$ and $\text{Tr}(I)$ are uncorrelated and the other features can be explained by these two. The remaining three features are the real and imaginary part of the chiral parameter ($\text{Re}(\chi)$, $\text{Im}(\chi)$) and the number of facets N_f. The chiral parameters appear to be unimportant. The number of facets N_f is not important because (1) it is a noisy measure that is based on the number of facets clustered by the DBSCAN algorithm; and (2) it is an integer value and thus is not sufficiently finely resolved among structure.

The self-assembly of shapes that are now understood to be far from optimal [31] has been reported in the literature [1, 20]. To develop a model that also correctly predicts assemblies for highly suboptimal shapes, we run additional simulations at other densities to broaden our data set. We further generate shape distributions at high densities with different number of shape vertices (FCC: number of vertices=32, η=0.65; 32, 0.7; 32, 0.8; 32, 0.9; 32, 0.95; 32, 0.99; 10, 0.6; 12, 0.6; 14, 0.6; 20, 0.6; 50, 0.6; 120, 0.6; BCC: 6, 0.6; 14, 0.64), shape distributions are shown in Fig. 4.5. From this we find that distributions of particle shapes are clustered by geometry, and several structures exhibit multiple free energy basins. The existence of these basins indicates that for those structures multiple distinct particle shapes provide good candidate shapes for self-assembly.

We further note that shape distributions vary in form by structure. For example, G and H structures exhibit a narrow range of $\cos(\theta_d)$ relative to the distributions for SC, SCC, BCC, BCT, and FCC. Conversely SC, SCC, BCC, BCT, and FCC all exhibit a narrow range of $\text{Tr}(I)$ relative to the distributions for G and H. We expect that this relative sensitivity
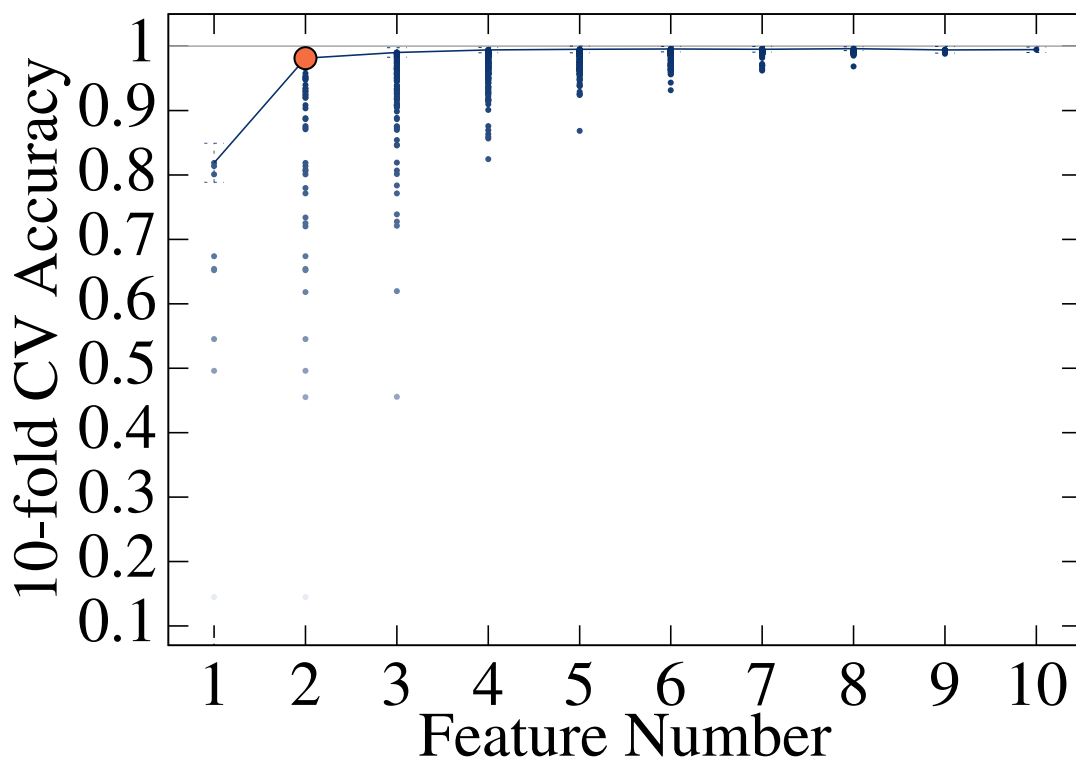
Figure 4.2: From low density shape distributions produced via Alch-MC, we classify shapes using combinations of 10 geometric criteria via the random forest method from machine learning. We build random forest models with different number of geometric features and all possible combinations of features. Each point shows a 10-fold cross-validation (CV) accuracy of a random forest model. The line connects points with best accuracy. We find that two shape features $\cos(\theta_d)$ and $\mathrm{Tr}(I)$ give 98% accuracy (using 10-fold cross-validation) in predicting structures (orange point).

is an important consideration in the synthesis of appropriately shaped particles for self-assembly. Moreover, Fig. 4.4 shows regions that BCC, FCC, β-W, and β-Mn distributions locate tightly together in shape space with some overlap. The existence of this overlap accords with prior work [20] that found some hard polyhedra spontaneously self-assemble more than one structure (e.g. β-Mn and FCC).

Fig. 4.6 shows the two-parameter empirical model, obtained from the Alch-MC simulation data, that maps particle shape to structure. Each color region represents a crystal structure that the random forest model would predict for the specific data point on the
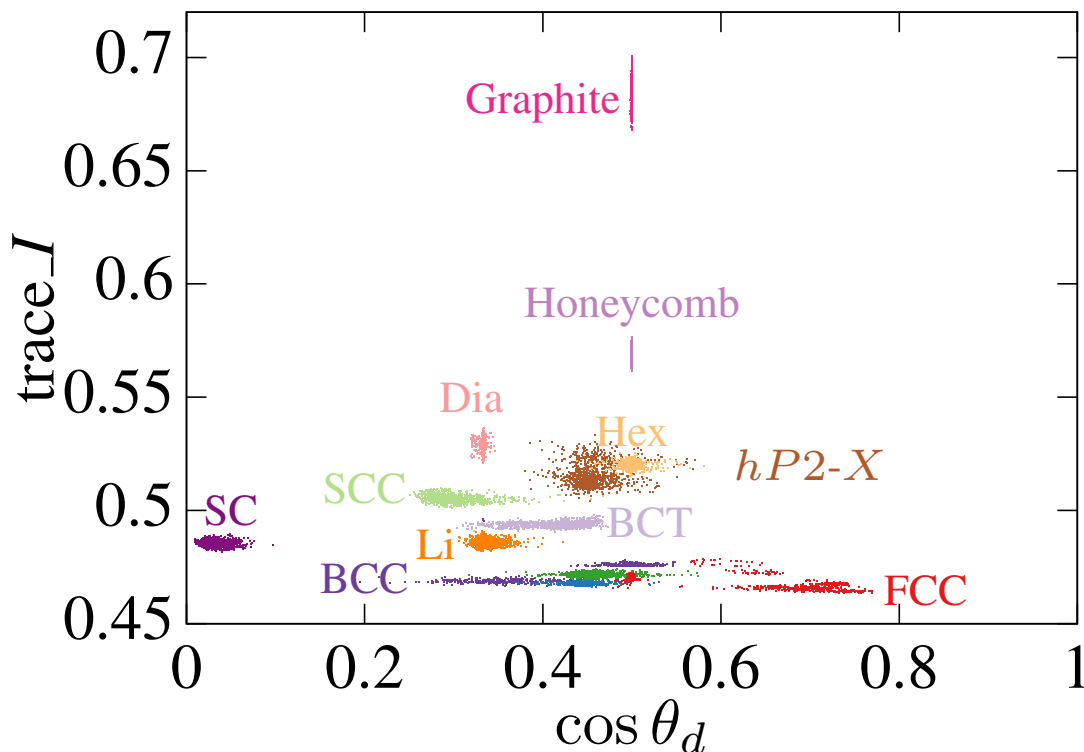
Figure 4.3: Shape distributions from Alch-MC plotted as a function of the two primary shape features (Including both low density and high density shapes). Each mark represents an observed shape, and is colored by its corresponding crystal structure.

$(\cos(\theta_d), \mathrm{Tr}(I))$ space. The dark color means the prediction probability is high, and the light color means the prediction probability of low.

We also extracted thermodynamically optimal shapes from the peaks of the shape distributions in the two salient parameters. Fig. 4.7 shows peak positions, inset particle images show representative particle shapes, and inset bond order diagrams of crystal structures show the spontaneous self-assembly of the optimal shapes from fluid phase.

4.5 Testing the Prediction Model

As a further test of the model's predictive ability, we compared our model's predictions against the previously reported self-assembly behavior [20] of 81 known crystal-forming
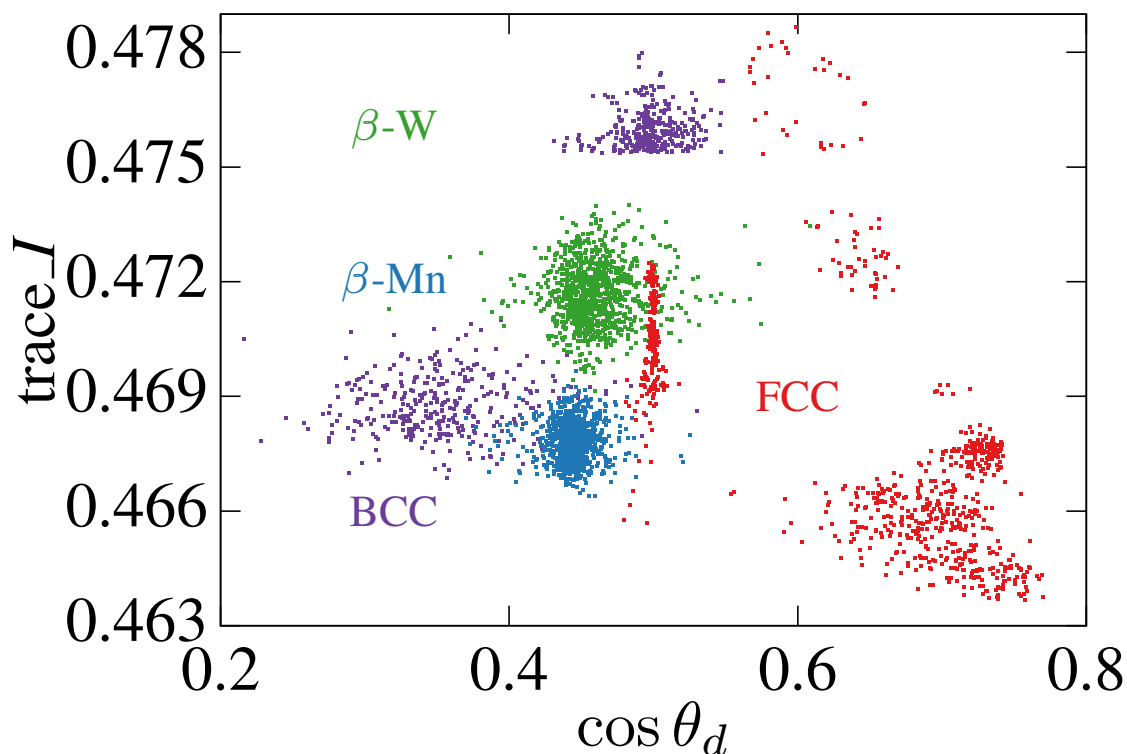
Figure 4.4: Zoom in of the densely distributed structure region in Fig. 4.3.

polyhedra, in which 71 shapes formed target structures included in our set of 13. Contained in that set are the cube [42, 78], truncated cube [98, 42], octahedron [98, 42], truncated octahedron [42], cuboctahedron [42], and rhombic dodecahedron [98], each of whose self-assembled colloidal crystal structures have been shown in experiments to match the simulated structures. In Ref. [20], shapes were self-assembled into crystals without the use of alchemical variables, and thus the shapes are not necessarily the optimal ones for the obtained crystal structure. For that reason, we expect the accuracy of our predictions to be lower than that obtained for optimal shapes. We found that the two-parameter model correctly predicts the crystal structure formed by 65 shapes in the 2012 study with a fidelity of 91.55%. The model is, of course, unable to correctly predict the structure for shapes that formed structures not included in our set of 13, and also failed to predict structures where
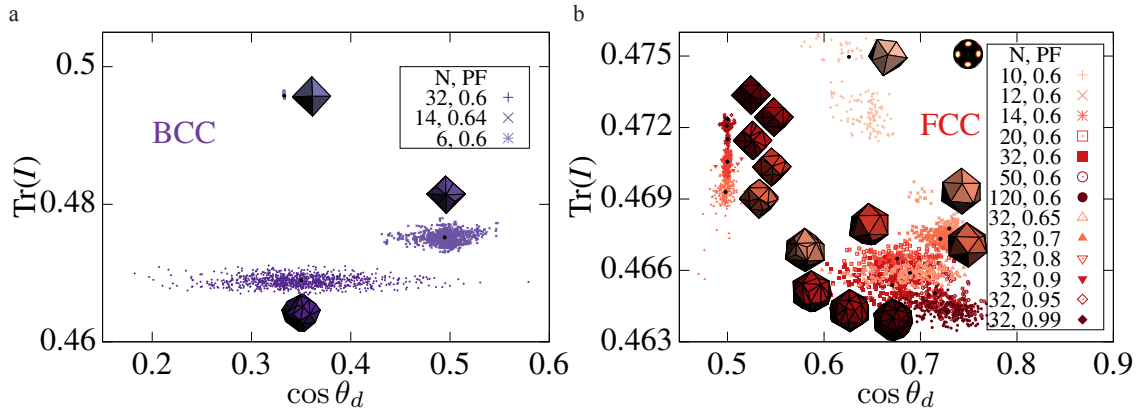
Figure 4.5: (a) Shape distributions for BCC produced via Alch-MC, including suboptimal shapes. (b) Shape distributions for FCC produced via Alch-MC, including suboptimal shapes.

multiple structures were reported. Model predictions and shapes are given in Fig. 4.8.

4.6 Discussion

Our coupling of inverse design and machine learning techniques to create a purely geometric, two-parameter, empirical model that predicts the self assembly of colloidal crystals of convex polyhedra with a fidelity of 98% is a considerable advance over the two-parameter, empirical model presented in Ref. [20]. Nevertheless, there is room for further improvement. First, our model was based on considering only 10 different shape features. It is possible that there are other shape features that could prove better predictors of self-assembly behavior. To facilitate investigations into this possibility, we have made raw shape data available online at `https://deepblue.lib.umich.edu/data/` `concern/generic_works/6q182k84r?locale=en`. (doi:10.7302/Z2T72FN9) We encourage the community to use the data to look for increasingly accurate predictive models. Second, in regions of shape space that are either sparsely populated by our data and so yield poor statistics, or in which different structures are densely clustered, the model may fail to predict the correct structure. We encourage the community to add data to the online data set for additional shapes and structures beyond those considered here. Third,
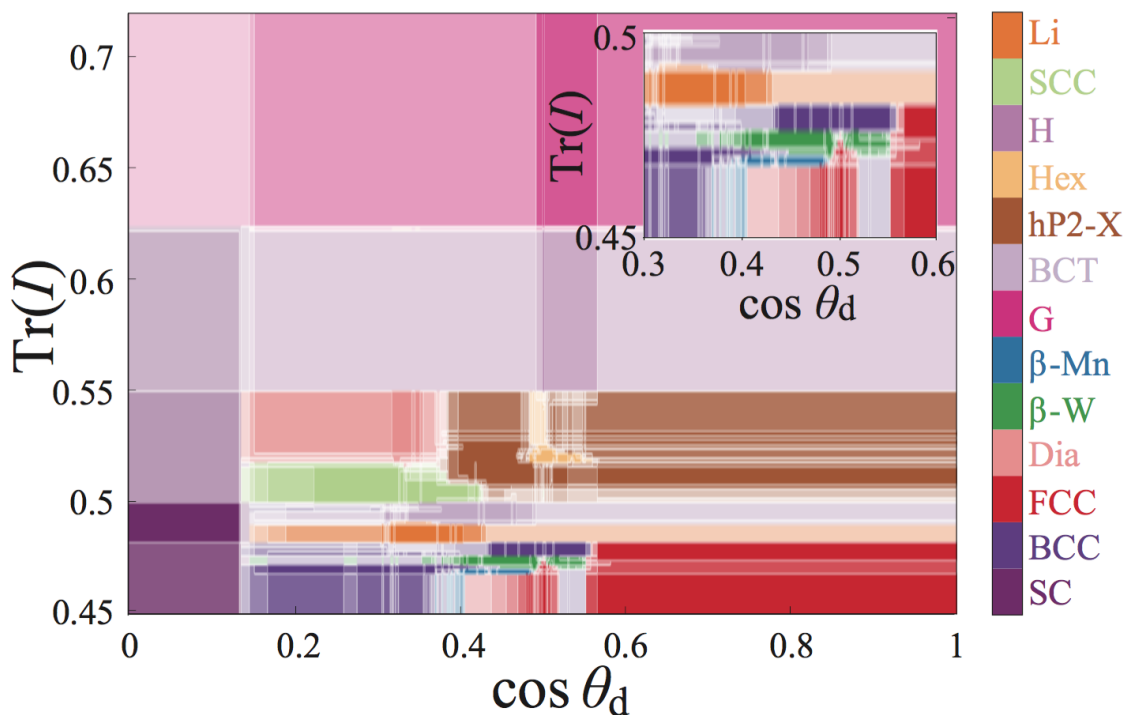
Figure 4.6: The prediction probability from the two-feature random forest model. We train the model using $\cos(\theta_d)$ and $\text{Tr}(I)$ data from the 13000 optimal shapes. The test data is comprised of 67500 $(\cos(\theta_d), \text{Tr}(I))$ combinations distributed evenly on the plot. For each predicted crystal structure class, we divide the prediction probability based on the test data evenly into 10 levels and use the transparency to represent the probability value. Opaque means probability 1 and transparency means probability 0. Inset is the zoom in of the densely distributed structure region.

Digital Alchemy as implemented here does not consider the free energy of the fluid phase, and thus provides no information on the driving force for self assemblies. Such extension could be included in future work.

The fact that as few as two geometric criteria are sufficient to predict structure formation in crystals with as many as 20 (β-Mn) particles in a unit cell suggests that appropriate geometric criteria might be useful for predicting order in colloidal systems with other forces at play, including hydrogen bonding between DNA-programmable colloidal shapes, van der Waals attraction between ligands or between particle cores, depletion interactions, and so on. Alch-MC or other inverse design methods can treat any of these cases to
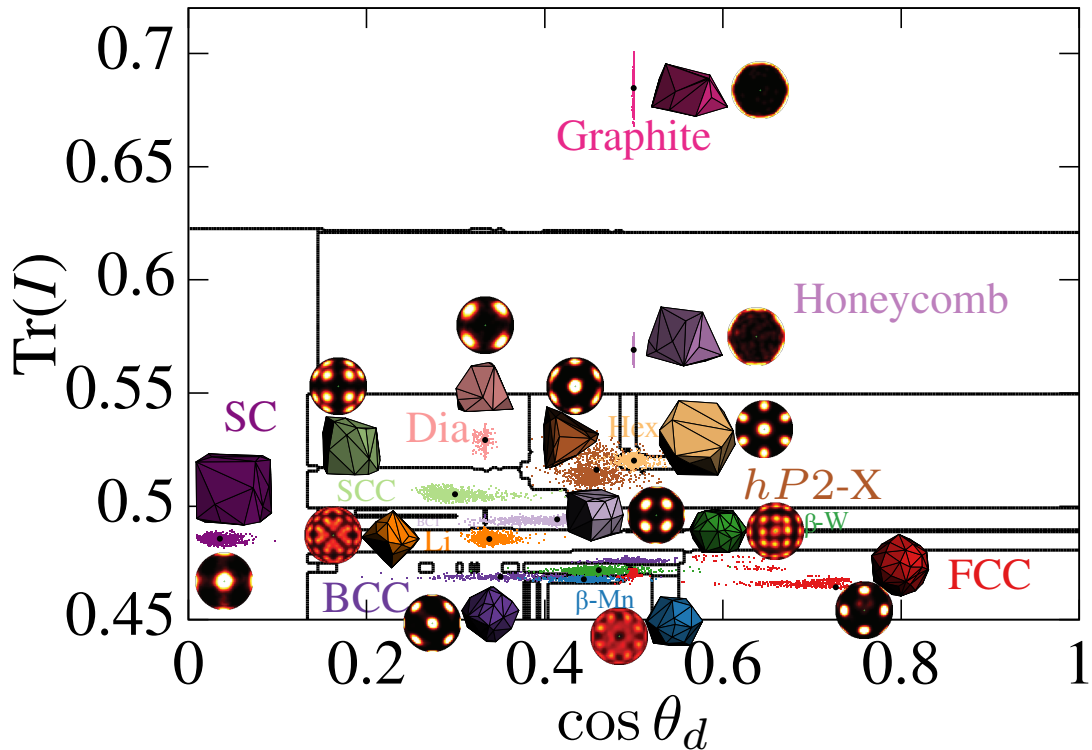
Figure 4.7: Optimal shape representations and self-assembly of optimal shapes. Back lines are the decision boundary of random forest based on ($\cos(\theta_d)$ and $\mathrm{Tr}(I)$). Marks indicate observed particle shapes and are colored by structure (same with Fig. 4.3). Black marks indicate shape distribution peaks corresponding to thermodynamically optimal particle shapes. Example near-optimal particle shapes and the bond order diagrams of crystal structure spontaneously assembled by the near-optimal particle shape are shown.

produce optimized forces as well as shapes. Nevertheless, our predictions should extend to experimental systems of anisotropic colloids and nanoparticles in which entropy plays a major role.[61, 92]

There are two ways in which future work could find results that conflict with ours. We discuss briefly what form those conflicts could take, and what new knowledge would be generated as a result of each conflict.

One possible conflict is if other investigations generate shape distributions that conflict with model predictions in regions of shape space that are sparsely populated by our data.
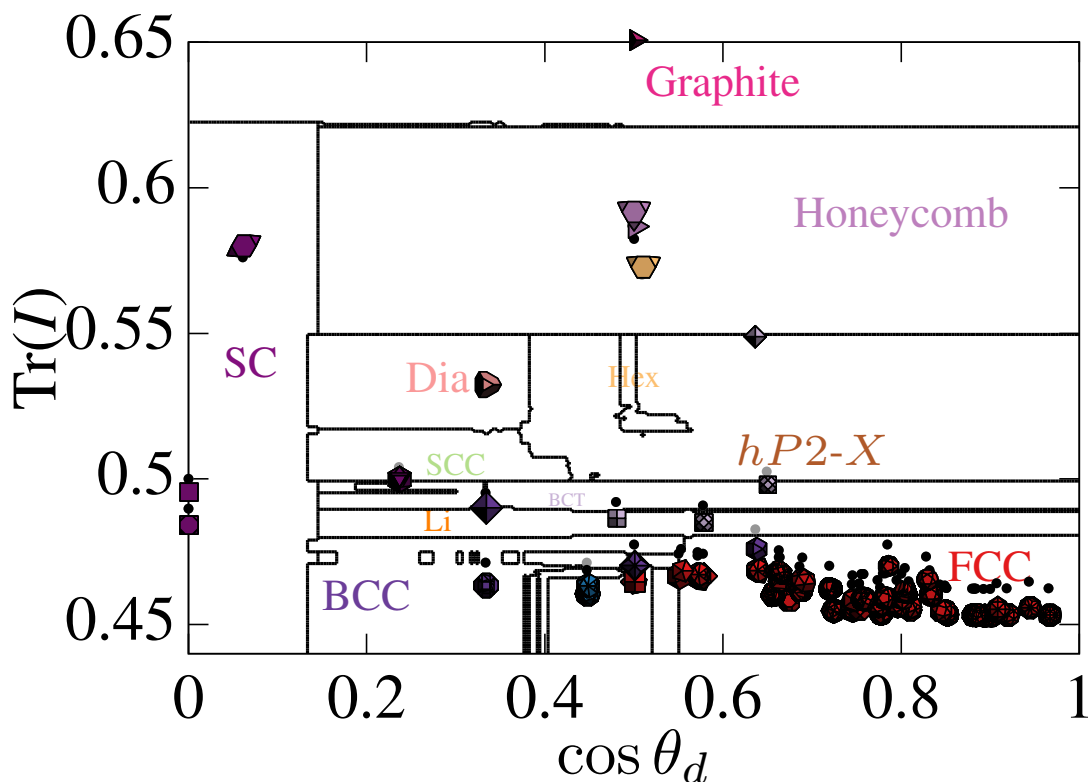
Figure 4.8: Test of empirical model prediction against previously reported dataset describing crystal self-assembly of 71 convex polyhedra. Model correctly predicts crystal structure observed via self-assembly in 65 of 71 cases (91.55%). Black lines are the decision boundary of random forest based on the $\cos(\theta_d)$ and Tr(I) of optimal shapes from Alch-MC. Circle points are the $\cos(\theta_d)$ and Tr(I) of the convex polyhedra in the training data. Black points indicate correct predictions, and gray points indicate incorrect predictions. Convex polyhedra are plotted near their corresponding point and colored by the crystal structure observed from self-assembly in Ref. [20].

Shape space regions that are sparsely populated by our data suggest the possibility that shapes in those regions could correspond to target structures that are not in our candidate set, free energy basins not captured by Alch-MC, or disordered or liquid crystal shape formers. Alch-MC without structural bias, but with shape bias targeting these regions could potentially uncover the existence of new entropically stabilized structures. It is our hope that future work will invalidate our predictions in sparsely sampled regions of shape space and thereby discover new entropically stabilized structures.

Another possible conflict is if other investigations generate shape distributions that con-

flict with model predictions in regions of shape space that are densely populated by our data. To search for shape distributions that conflict with our results in densely populated regions of shape space, one approach would be to generate shape distributions with Alch-MC that includes both structural bias and shape bias that are incommensurate with the model prediction. Such a search would provide a direct means of determining the limitations of our model. As in the case of sparsely populated regions of shape space, though conflicts with our model prediction offer potential positive benefits because they would provide "nearby" shapes that assemble into different crystal structures, which would present opportunities for reconfigurable structures [22].

In the most general space of shapes, our approach only samples near free energy minima. BCC and FCC are unusual structures in the sense that particles that are far from absolute free energy minima will produce those structures. To incorporate this understanding into our model we used more restricted shape spaces that do not have access to the global free energy minimum to find disconnected local free energy minima. In principle it is possible to do this for all structures but we did not do so because we are unaware of highly sub-optimal shapes that produce structures other than BCC and FCC. Hence an arbitrary search for local free energy minima with all possible vertex numbers of all structures with no indication that such minima might exist would not be a responsible use of our computational resources.

CHAPTER V

Optimized Synthesizable Nanoparticle Shapes for Self Assembly

The study of this chapter by Yina Geng, Greg van Anders and Sharon C. Glotzer will be poptimalublished following the completion of this dissertation.

Anisotropic colloids and nanoparticles assemble into a great diversity of colloidal crystals in computer simulation [40, 1, 19, 20, 30, 82]. However, the gulf between the outcomes predicted in simulation [20] and the outcomes realized in experiment [43, 77] is wide. Recent work has shown that many of the ordered structures observed in simulation occur for particles with shapes that are far from those of the "eigenshapes" that minimize the free energy of a system in a given structure [31, 32]. Moreover, the application of machine learning techniques to distributions of particles [32] that result from Digital Alchemy simulations [93, 31] indicate that ordered structures display vastly different amounts of sensitivity to different forms of variation in particle morphologies [32]. Recent work has reported eigenshapes and parametric sensitivities for a large number of structures [32]. However, the reported eigenshapes were computed without regard to potential synthesis methods that could produce them.

Here, we investigate for a variety of structures optimal colloidal particle shapes, computed through inverse design, for families of particles that have been reported in Refs. [50, 7, 104, 53, 94, 88, 43, 17, 44, 54, 75, 100, 16, 103, 67]. These families include shapes

that are synthesizable via vertex truncation, edge truncation, or facet augmentation. We model the nanoparticles as hard shapes and discover thermodynamically optimal shapes within the shape family that minimize the free energy of target crystal structures using Digital Alchemy. Digital Alchemy is an extended ensemble method that treats particle shape parameters as thermodynamic variables, and allows the exploration of shape space in Alchemical Monte Carlo (Alch-MC) simulations to find optimal shapes for the self-assembly of target structures.

Our approach involves two steps. We first identify colloidal crystal structures that self-assemble from a given shape in a shape family. We then fix the crystal structure and determine the optimal shape for that structure within that shape family. By optimal shape, we mean the shape that minimizes the free energy of the structure at a given density. Although we do not optimize for self-assembly explicitly, which would involve optimizing for assembly pathways, we find in practice that most shapes found from Digital Alchemy in fact self-assemble their target structures [31, 32].

5.1   Shape Families Accessible to Experiments

We consider eight shape families accessible to experiments. Below, we describe each family and give examples of shapes from the families made in experiments and reported in the literature. The first six shape families contain shapes that have been studied in simulation [48]. They are predicted to form only simple cubic (SC), body-centered cubic (BCC), face-centered cubic (FCC), diamond, and high-pressure Lithium (Li) structures, depending on shape. For the last two families, we calculate the cosine of dihedral angle $\cos \theta_d$ and the trace of moment of inertia tensor $\text{Tr}(I)$ and predict the preferred structures should be SC and FCC, or run regular MC simulations to predict the preferred structure.

The octahedron to cube shape family via vertex truncation is shown in Fig. 5.1a; the

octahedra has truncation equal to 0 and the cube has truncation equal to 1. This shape family belongs to the spheric triangle group $\Delta_{3,2,3}$[15] which is constructed with three families of planes that make up the faces of a cube, a tetrahedron, and an octahedron. There are two shape parameters, a and c, that specify the amount of truncation (or position of the bounding planes). The shape family we study is the same as the $\Delta_{3,2,3}$ family with $a = c$. The target structures predicted to assemble in this shape family are simple cubic (SC), body-centered cubic (BCC), and face-centered cubic (FCC) [48]. This shape family can be realized in the lab [50, 7, 104, 53, 94, 88, 17, 44, 54, 75, 96]. Chiu *et.al.* used a biomimetic approach to synthesize platinum nanocrystals enclosed by particular facets, where peptide sequences that recognize Pt-100 and Pt-111 planes have been rationally identified and used in directing the shape formation of platinum nanocrystals of typical size 6 nm in a colloidal nanocrystal synthesis process [17] (Fig. 5.2I-a(1)). Liu and coworkers developed a simple method for the direct synthesis of Pd nanocrystals with average size 50 nm in aqueous solution with systematic shape evolution from octahedral to cubic structures through fine control of the amount of KBr introduced [54] (Fig. 5.2I-a(2)). Ho and Huang developed a facile room temperature procedure for the synthesis of Cu_2O nanocrystals capable of a systematic shape evolution from cubic to octahedral structures by varying the amount of reductant added to the reaction mixture [44]. The crystals are generally submicrometer-sized (Fig. 5.2I-a(3)). Tao *et.al.* synthesized this shape family with polyhedra size 80-300 nm using the polyol method [88] (Fig. 5.2I-a(4)). Xia and coworkers used seed-mediated growth to obtain this shape family with size 40 nm - 75 nm [96] (Fig. 5.2I-a(5)). Liu *et.al.* obtain 20 nm palladium octahedra with controlled edge lengths from Pd cubes using a transformation involving oxidative etching and regrowth [53] (Fig. 5.2I-a(6)). Wang *et.al.* obtain this shape family with polyhedra size 5-15 nm by seed-mediated growth method [94] (Fig. 5.2I-a(7)). Zhu and coworkers use an aqueous method to synthesize 11 nm Pd

nanocrystals with shapes in the cube to octahedra family [104] (Fig. 5.2I-a(8)). Bao *et.al.* produced Pd nanocrystals of typical size 10 nm with shapes in the cube to octahedra family using seed-mediated growth [7] (Fig. 5.2I-a(9)). Lim *et.al.* used the seeded growth method to convert Pd nanocrystals from cubes to truncated and regular octahedrons with average size 10-15 nm [50] (Fig. 5.2I-a(10)).

The octahedron to tetrahedron shape family via vertex truncation is shown in Fig. 5.1b; the octahedron has truncation equal to 0 and the tetrahedron has truncation equal to 1. This shape family is the same as the spheric triangle group $\Delta_{3,2,3}$[15] family with $c = 1$. The target structures predicted to assemble in this shape family are BCC, diamond, Lithium (Li) [19]. This shape family has also been realized in experiments [94, 17] (Fig. 5.2I-b(1), Fig. 5.2I-b(2)).

The cube to rhombic dodecahedron shape family via vertex truncation is shown in Fig. 5.1d; the cube has truncation equal to 0 and the rhombic dodecahedron has truncation equal to 1. This shape family belongs to the spheric triangle group $\Delta_{4,2,3}$[15], which is constructed with three families of planes that make up the faces of a rhombic dodecahedron, a cube, and an octahedron. There are truncating planes of two types: type *a* corresponding to the location of the cube faces, and type *c* which correspond to the position of the octahedron faces. The shape family we study is the same with the $\Delta_{4,2,3}$ family with $c = 3$. The target structures predicted to assemble in this shape family are SC and FCC [48]. 110 nm Fe nanocrystals with shape in this shape family were synthesized by means of an electrochemical route [16] (Fig. 5.2I-d).

The vertex and edge truncated cube shape family is shown in Fig. 5.1e; the cube has truncation equal to 0 and the maximum truncation we study is 0.2. This shape family is the same as the spheric triangle group $\Delta_{4,2,3}$[15] family with $c = 4 - a$. The target structure predicted to assemble in this shape family is SC [48]. This shape family has also been

a

e

b

f

Diamond
Li

c

g

d

h

SC
BCC
FCC
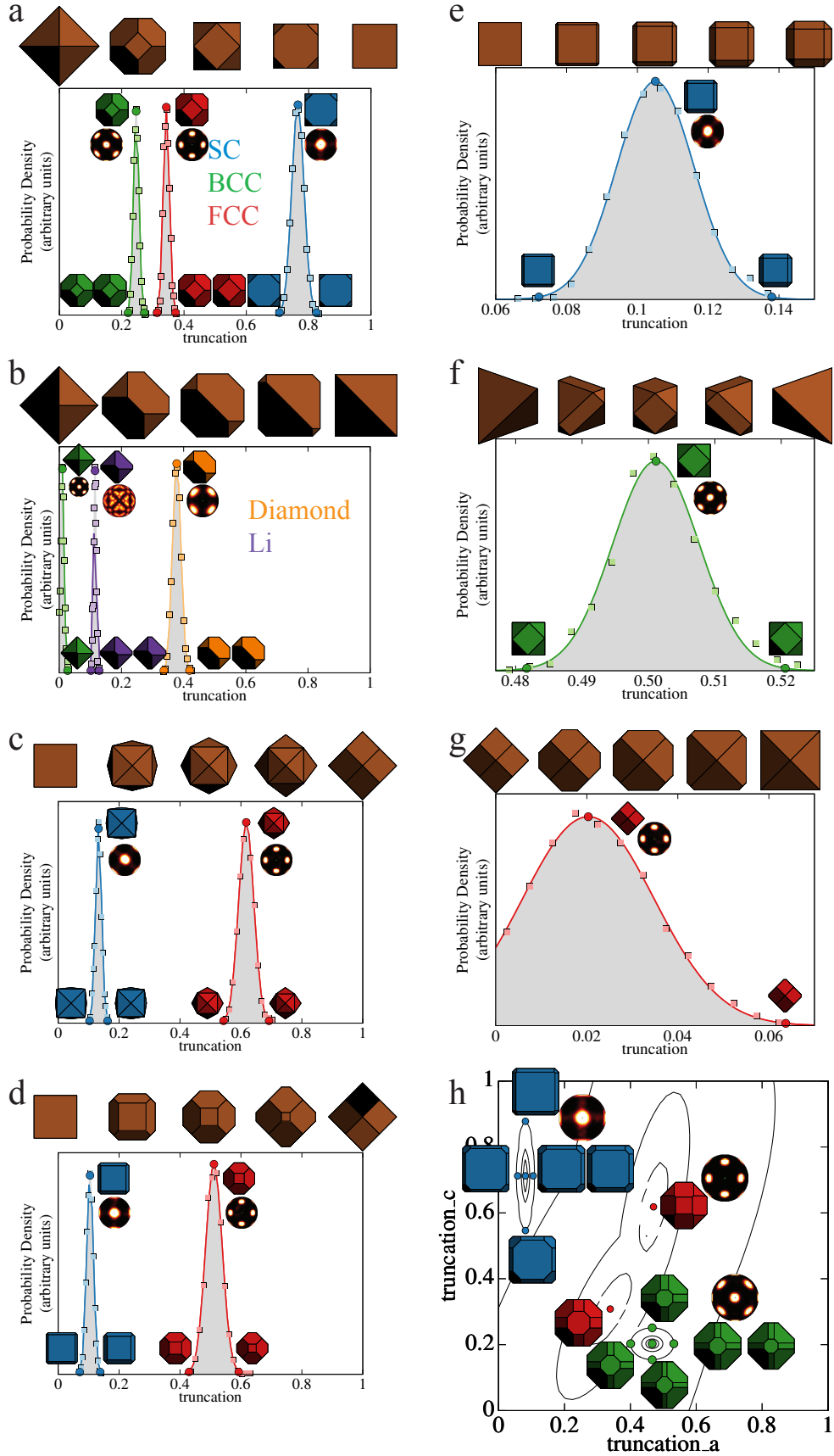
Figure 5.1: We study eight shape families that have been realized in the lab: (a) the cube to octahedron shape family; (b) the octahedron to tetrahedron shape family; (c) the cube to rhombic dodecahedron shape family; (d) the cube to rhombic dodecahedron shape family; (e) the vertex and edge truncated cube shape family; (f) the tetrahedron to cuboctahedron shape family; (g) the rhombic dodecahedron to tetragonal bipyramid shape family; (h) the spheric triangle group $\Delta_{4,2,3}$[15] family. For each target structure in each shape family, we calculated the equilibrium distribution of the shape and then fit the data with a Gaussian distribution. Square points show the histogram of the thermodynamic shape parameters from Alch-MC simulations. The curve shows the Gaussian fit to the histogram. We show the optimal shape (mean of the Gaussian distribution) and two boundary shapes that are three standard deviations away from the optimal value. The shape parameters for the optimal and boundary shapes are presented in Fig. 5.3. Both the optimal shape and the boundary shapes self-assemble their target structure from a fluid phase. For each target structure in each shape family, we show the bond order diagram of the structure assembled by the optimal shape.

realized in the lab [17] (Fig. 5.2I-e).

The tetrahedron to cuboctahedron shape family via vertex and edge truncation is shown in Fig. 5.1f; the tetrahedron has truncation equal to 0 or 1 and the cuboctahedron has truncation equal to 0.5. This shape family is the same as the spheric triangle group $\Delta_{3,2,3}$[15] family with $c = 4 - a$. The target structure predicted to assemble in this shape family is BCC [48]. This shape family has also been realized in experiments [17, 103]. Zheng and coworkers synthesized Au nanocrystals with typical size 45 nm via seed-mediated growth [103] (Fig. 5.2I-f(1)).

The spheric triangle group $\Delta_{4,2,3}$[15] family is shown in Fig. 5.1h. The target structures predict to assemble in this shape family are SC, BCC and FCC [48]. Niu and coworkers used the seed-mediated growth method to synthesize palladium nanocrystals with shapes varying in this family [67] (Fig. 5.2I-h).

The cube to rhombic dodecahedron shape family via facet augmentation is shown in Fig. 5.1c; the cube has truncation equal to 0 and the rhombic dodecahedron has truncation equal to 1. Details about the construction can be found in Ref. [31]. We calculate the

cosine of the dihedral angle $\cos \theta_d$ and the trace of the moment of inertia tensor $\mathrm{Tr}(I)$ for shapes in this family and predict via a machine learning model [32] the preferred structures should be SC and FCC with greater than 90% confidence. This shape family has also been realized in experiment [100]. Through a controlled heteroepitaxial growth process on preformed Au trisoctahedral NC templates, 50 nm nanocrystals with different polyhedral shapes were synthesized (Fig. 5.2I-c).

The rhombic dodecahedron to tetragonal bipyramid shape family via 3-fold vertex truncation is shown in Fig. 5.1g; the rhombic dodecahedron has truncation equal to 0 and the tetragonal bipyramid has truncation equal to 1. The machine learning model [32] predicts structures from this shape family with less than 60% probability, due to most shapes in this family being suboptimal (as we determine later). Instead we run hard particle Monte Carlo [5] simulations for this shape family and find all shapes either assemble FCC or do not assemble. This shape family has also been realized in experiment [16] (Fig. 5.2I-g).

Figure 5.2: **I-a** Cube to octahedra shape family synthesized in experiment. **(1)** TEM and HRTEM images of platinum nanocrystals from cuboctahedra to cube. Scale bars: 2 nm (b), 5 nm (e,h) [17]. **(2)** SEM images of (a1) Pd octahedra, (b1) truncated octahedra, (c1) cuboctahedra, (d1) truncated cubes, and (e1) nanocubes. All scale bars are equal to 100 nm [54]. **(3)** SEM images of the Cu2O nanocrystals: (a) cubes, (b) truncated cubes, (c) cuboctahedra, (d) type I truncated octahedra, (e) type II truncated octahedra, (f) octahedra, (g) short hexapods, and (h) extended hexapods. Scale bar: 1 μm [44]. **(4)** a) A schematic of the nucleation and growth process. b-f) SEM images of cubes, truncated cubes, cuboctahedra, truncated octahedra, and octahedra, respectively (scale bar: 100 nm) [88]. **(5)** SEM images of Ag polyhedrons grown from 40 nm cubic seeds. The inset shows the corresponding 3D model for each type of polyhedron [96]. **(6-1)** (a) Schematic illustration of the five major steps involved in the oxidative etching and regrowth process. (b) Schematic illustrations showing the formation of Pd octahedra with different edge lengths [53]. **(6-2)** TEM images of (a) Pd nanocubes and (b-d) Pd octahedra [53]. **(7)** (a) Schematic illustration showing the formation of Pd octahedrons and tetrahedrons, respectively. (b-d) Typical TEM images of the Pd cuboctahedral seeds, octahedrons, and tetrahedrons. [94]. **(8)** TEM images of Pd nanocrystals [104]. **(9)** TEM images of the Pd polyhedra and the insets show geometrical models of individual nanocrystals [7]. **(10)** A) Schematic illustration of seeded growth of Pd octahedrons with and without truncation at corners from cubic Pd seeds. B) TEM and C) HRTEM images of Pd truncated octahedrons. D) TEM and E) HRTEM images of Pd octahedrons [50]. **I-b** Octahedra to tetrahedra shape family synthesized in experiment. **(1)** TEM and HRTEM images of platinum tetrahedron nanocrystals. Scale bars: 2 nm (c), 5 nm (f,i) [17]. **(2)** see **I-a (7)**. **I-c** Cube to rhombic dodecahedra shape family via facet augmentation synthesized in experiment. SEM images showing the overall morphology of Au@Pd nanocrystals in (A) high and (B) low magnifications and (C) individual nanocrystals in different orientations, with the corresponding geometrical models shown on the right of each SEM image. The scale bar is 50 nm [100]. **I-d** (a-e)Shape transformation of Fe nanocrystals from rhombic dodecahedra to a series of 18-facet polyhedral shapes and finally to cubic [16]. **I-e** TEM and HRTEM images of shape evolution of platinum cube synthesis. Scale bars: 2 nm (c), 5 nm (f,i) [17]. **I-f (1)** TEM images of Au nanocrystals. The scale bar in (a) applies to (b-d). The models at the bottom correspond to those particles in the TEM images labeled with the same number [103]. **(2)** TEM and HRTEM images of platinum tetrahedron nanocrystals. Scale bars: 2 nm (b), 5 nm (e,h) [17]. **I-g** SEM images of Fe nanocrystals and their geometrical model [16]. **I-h (1)** Geometrical models of palladium nanocrystals presented in Figure I-h (2). The 100, 111, and 110 facets are shown in green, blue, and purple, respectively [67]. **(2)** SEM images of polyhedral palladium nanocrystal samples (scale bar: 200 nm) [67]. Note: Permissions to reproduce images need to be approved.

5.2 Inverse Design of Optimal Shapes

5.2.1 Alch-MC Simulations for Eight Shape Families

Following the protocol demonstrated in Ref. [31], we performed Alchemical Monte Carlo (Alch-MC) simulations of symmetric convex polyhedra, with shapes varying within the shape families described in the previous section, in target crystal structures. Alch-MC simulations based on the Digital Alchemy framework [93], using an implementation [31] that extends an open-source Monte Carlo plugin, Hard Particle Monte Carlo (HPMC) [5], for the open-source molecular dynamics package HOOMD-Blue [4] to generalized thermodynamic ensembles that include particle shape change. We simulated $NVT\mu$ ensembles at constant temperature T, fixed volume V, and alchemical potential $\mu = 0$. The variable μ is conjugate to the shape variable that is allowed to fluctuate in the simulation. We placed no fewer than 100 particles in a periodic simulation box. The exact number was chosen to be a multiple of the number of particles in the unit cell of the target structure. Particle shapes were initialized with each shape parameter taken as either 0 or 1. Monte Carlo (MC) sweeps involve particle translations, rotations, and shape moves. For each shape move, we (i) generated a uniform trial shape for all particles in the system with modified shape parameters, (ii) resized the trial shape to unit volume, (iii) checked if the move induced any particle overlaps, and then (iv) accepted the move based on the Boltzmann factor as described in Ref. [93]. Translation and rotation moves followed standard procedures (see, e.g., Refs. [40, 1, 30, 19, 20, 92, 91]). We compressed the system to packing fraction $\eta = 0.6$ (SC, BCC, FCC, diamond) or $\eta = 0.75$ (Li), with the spring constant fixed to 1000 (where energy is specified in units of $k_{\mathrm{B}}T$, and length units are given in terms of the particle volume). After we reached the target packing fraction, we logarithmically relaxed the spring constant. We then relaxed the system for 1×10^6 MC sweeps. For each target crystal structure, we performed at least 60 independent simulations and analyzed the

a

Structure	mean	mean-3σ	mean+3σ
SC	0.77	0.71	0.82
BCC	0.25	0.22	0.27
FCC	0.34	0.31	0.37

b

Structure	mean	mean-3σ	mean+3σ
BCC	0.01	N/A	0.03
diamond	0.38	0.34	0.42
Li	0.11	0.10	0.13

c

Structure	mean	mean-3σ	mean+3σ
SC	0.13	0.10	0.16
FCC	0.62	0.54	0.69

d

Structure	mean	mean-3σ	mean+3σ
SC	0.10	0.07	0.14
FCC	0.51	0.43	0.59

e

Structure	mean	mean-3σ	mean+3σ
SC	0.11	0.07	0.14

f

Structure	mean	mean-3σ	mean+3σ
BCC	0.50	0.48	0.52

g

Structure	mean	mean-3σ	mean+3σ
FCC	0.02	N/A	0.06

h

Structure	mean (a, c)	(a-3σ, c)	(a+3σ, c)	(a, c-3σ)	(a, c+3σ)
SC	(0.08, 0.71)	(0.06, 0.71)	(0.11, 0.71)	(0.08, 0.55)	(0.08, 0.88)
BCC	(0.47, 0.20)	(0.40, 0.20)	(0.53, 0.20)	(0.47, 0.15)	(0.47, 0.25)
FCC	(0.34, 0.31) (0.47, 0.62)	N/A	N/A	N/A	N/A

Figure 5.3: In each shape family a-h, we show the mean shape parameter and the two boundary shape parameters. Both the optimal shape and the boundary shapes self-assemble the target structure. The optimal shape suggests the best shape to assemble the target structure, and the two boundary shapes show the tolerance of a shape deviating from the optimal shape. The boundary shape parameters can also help glide synthesis protocols.

shapes in the final 1.5×10^5 sweeps.

For each target structure in each shape family, we calculated the equilibrium distribution of the shape and then fit the data with a Gaussian distribution. We show the optimal shape and two boundary shapes that are three standard deviations away from the optimal value (see Fig. 5.1). The shape parameters for the optimal and boundary shapes are presented in Fig. 5.3. Both the optimal shape and the boundary shapes self-assemble their target structure from a fluid phase. The optimal shape suggests the best shape to assemble the target structure, and the two boundary shapes show the tolerance of a shape deviating from the optimal shape. The boundary shape parameters can also help decide synthesis protocols.

5.2.2 Direct Free Energy Computation

For SC, BCC and FCC structures, optimal shapes in different shape families assemble the same target structure. We compare the alchemical free energy to get the optimal shape that minimizes the free energy of the target structure. We computed the Helmholtz free energy difference between the target crystal and the Einstein crystal using Frenkel-Ladd thermodynamic integration [26] via the implementation used in Refs. [39, 82]. We placed approximately 2000 particles in a periodic simulation box; the exact number was chosen to be a multiple of the number of particles in the unit cell of the target structure. For SC and BCC structures, particles in the assembled structure have orientational order. Einstein crystal positions and orientations were taken directly from the space-filling tessellation. For FCC, particles in the assembled structure do not show orientational order. To create an appropriate Einstein crystal, we first initialized the simulation at a low packing fraction $\eta = 0.5$, chose particle positions using the target structure, and randomly assigned a set of orientations observed in the assembly. Then we compressed the system to packing fraction $\eta = 0.6$, allowing particles to rotate to resolve overlaps. The Einstein crystal of BCC with octahedron shape can not be compressed to packing fraction $\eta = 0.6$, so we followed the same protocol as for FCC. We computed the alchemical free energy of the target structure assembled by the optimal shape in each shape family [93]. We normalized the free energies in Fig. 5.4 by setting the largest free energy of the target structure to be zero. In Fig. 5.4, we order the optimal shape by the free energy of the system. For each structure, the rightmost optimal shape optimizes the target structure.

5.3 Discussion

The self-assembly of anisotropic nanoparticles into colloidal crystal structures have been shown in experiments [99, 43, 77]. For example, cubes self assemble into SC [43, 77],
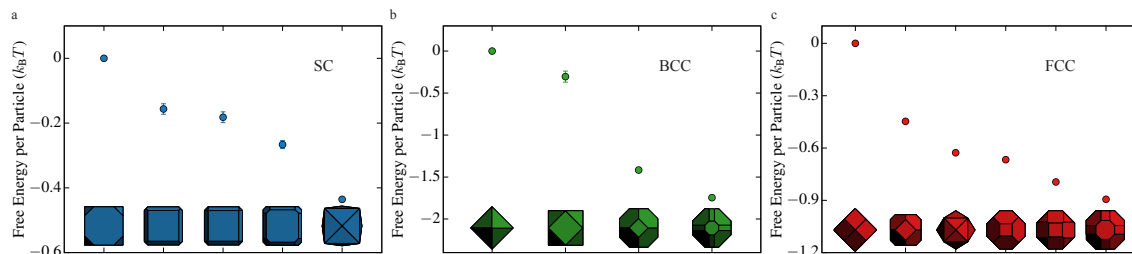
Figure 5.4: (a) Free energy for SC assembled by the optimal shape in each shape family. (b) Free energy for BCC assembled by the optimal shape in each shape family. (c) Free energy for FCC assembled by the optimal shape in each shape family. For each structure, we normalized free energies by setting the largest free energy of the target structure to be zero. We order the optimal shape by the free energy of the system. For each structure, optimal shapes from different shape families can assemble the target structure, and the optimal shape at the furthest right is the best. Error bars are calculated from ten independent free energy computations for each shape (smaller than the data symbols if not shown).

truncated cubes self assemble into SC [99, 43], and truncated octahedra self assemble into BCC [43]. The experiment results match the simulated structures. Here we suggest the best shape to use to self-assemble the target structure. In this work, we used a hard particle model, but future work should include effects of stabilizing ligands. In those cases, the organization of ligands according to the so-called orbifold topological model (OTM) of Travesset controls the structure and stability of the superlattices that can be assembled in experiment [90].

CHAPTER VI

Conclusion and Outlook

6.1 Summary of Results

This dissertation has studied the inverse material design problem. Given a target colloidal crystal structure, we design the optimal building blocks (shapes) that self-assemble the target structure. We also design optimal shapes focusing on colloidal particle shape families have already been made in experiment. Combining the inverse design and random forest – supervised machine learning technique, we build a prediction model to predict the crystal structure from building block geometric features.

In Chapter III, we demonstrate that the digital alchemy method is a general and powerful method to solve the inverse material design problem. It treats the shape parameter as a thermodynamic variable and generates the optimal shape that minimizes the free energy of the target structure system. We design optimal shapes for six simple and complex crystal structures, and one as-yet-unknown structure. We extract important shape characteristics and design optimal symmetric shapes to be synthesizable in experiment.

In Chapter IV, we build a prediction model to predict the crystal structure from building block geometric features. We produce optimal shapes for 13 crystal structures using digital alchemy, and calculate 10 geometric features for each shape. Next, we build a prediction model to classify shapes into structures that the shape optimized for. Our model

achieves 98% prediction accuracy with only two geometric features, and 92% accuracy on the prediction of previously reported structures assembled by 71 symmetric polyhedra. In the future, if we have a new shape, our model helps decide the structure that the shape assembles, without any further simulation or experiment needed.

In Chapter V, we consider shape families that have already been synthesized in experiments and design optimal shapes in those shape families. Our results help experimentalist choose best colloidal particles to synthesize crystal structures and further study their properties.

6.2 Outlook

We successfully apply the digital alchemy method to 13 crystal structures and design optimal shapes that minimize the free energy of the target structure and also spontaneously self-assemble the target structure from fluid phase. Does the method work for all crystal structures? Do we need to modify the design process? Moreover, we design one shape for a target structure. However, some structures have several different Wyckoff positions, $e.g. \beta$-Mn, β-W and AlB$_2$. Should we have one type of particle for one Wyckoff position? Is the structure system with multiple shapes better (having a lower free energy) than the system assembled by one shape? These questions need to be answered in the future.

We build the prediction model using optimal shapes from 13 structures. The limit of our model is that we can not predict other structures outside of the 13 structures. Including more structures and shapes would improve our model's generality and prediction ability. We come up with 10 geometric features based on our experience and literature. It's possible that there are other powerful features we can use to improve the model. In our prediction model, we only consider predicting crystal structures and ignoring disordered or glass phases. How to build a model to predict crystals and other phases? What data can

we generate and use? What shape features we should use? These questions need further study.

We design optimal shapes that can be made experimentally, and help experimentalist choose shapes to synthesize colloidal crystals. However, the gap between what simulation predicts and what experiment can control and achieve is still big. We do not consider polydispersity, which often happens in experiments. This question should be solved in the future.

APPENDIX

APPENDIX A

Evolutionary Algorithm Trial

We first solve the inverse design problem using the evolutionary algorithm, specifically the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). Evolutionary computation has been applied successfully to a wide range of problems, including robotics design [52], crystal structure prediction [69, 10], and granular materials mechanical property optimization [66]. The optimization mechanism of evolutionary computation starts with an initial population of objects, from which those with highest fitness are selected and mutated to create offsprings, comprising the subsequent generation. Here we use evolution strategies (ES) due to their emphasis of manipulation on real domain. Specifically, we use the Covariance Matrix Adaptation Evolution Strategy or CMA-ES which produces a population using a multivariate Gaussian distribution. The key feature of this algorithm is that it uses information from prior generations to deterministically update the mean and covariance matrix. The mean of the distribution is updated to maximize the likelihood of drawing a previously found individual with favorable characteristics (high fitness), and the covariance matrix is updated in order to increase the probability of producing a successful mutation step.

Our design method works as follows: 1) a initial population of guesses (shapes) are created and quenching simulations are performed with particles in each system interacting according to one of these guesses; 2) each system's performance is evaluated by compar-

a

b
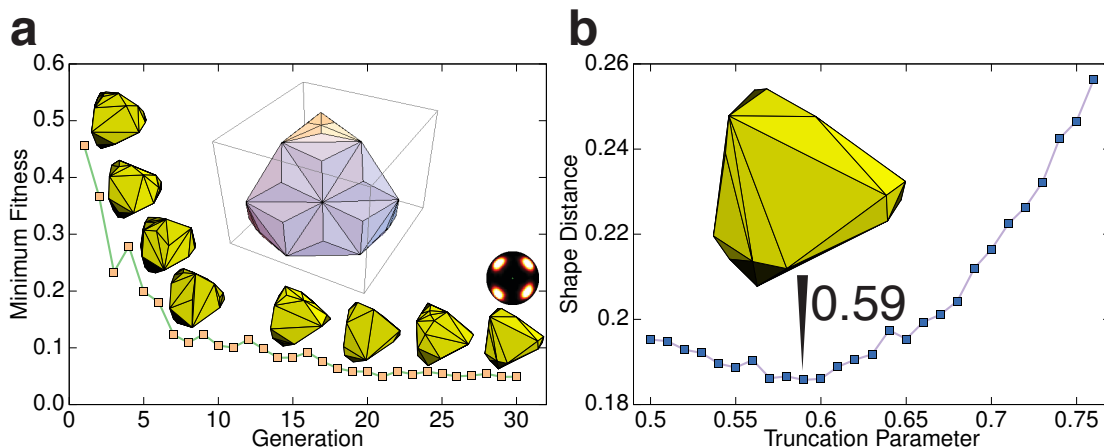
Figure A1: Optimal shapes found for diamond structure, using RDF difference as the fitness function in the CMA-ES. **a**, Fitness (RDF) convergence curve and optimal shapes (yellow) in each generation. The fitness value point is the minimum fitness in each generation in the CMA-ES. The light purple shape is the initial shape. The bond order diagram of diamond crystal assembled by optimal shape is shown. **b**, Shape distance between optimal shape found by minimizing RDF difference and truncated tetrahedra. The optimal shape is closest to truncated tetrahedron with truncation 0.59.

ing the absolute difference between its RDF and that of the target crystal 3) best scoring populations are then combined and mutation is introduced 4) we start back at (1) until algorithmic convergence.

We start with the diamond crystal as a benchmark. Our goal is to find a shape assembling diamond structure. Previous simulations have shown that the Archimedean truncated tetra-hedron shape can self-assemble a diamond structure from the fluid at moderate packing fractions [19]. We start with a truncated tetrahedron but tessellate its shape from 12 to 56 vertices to allow for shape evolution (Fig. A1a). We then use a vector containing the 3-dimensional coordinates of these 56 vertices as the variable to be optimized. Nineteen random candidate shapes are generated from the convex hull of the positions of these points. Next, we replace the particle shape in our previously assembled diamond lattice with each of the candidate shapes, already in their proper orientation, but keep particle positions and orientations fixed, totalizing 19 independent systems that will be simulated

in parallel. For each system of 2048 particles, we run a hard particle Monte Carlo simulation [5] for 6.4e5 sweeps with packing fraction being increased from 0.56 to 0.7. Fitness function we minimize in the CMA-ES is the least square difference between target RDF and the RDF of system with candidate shape. After each simulation and computation of the fitness function, 19 new candidates are created and the process is restarted until convergence is found. Fig. A1a shows the evolution of the fitness function for the best shape in each generation, demonstrating the fast convergence of the algorithm. The final shape (Fig. A1a) self-assembles diamond from liquid phase.

To find what truncated tetrahedron the final shape is closest to, we use shape matching [8] to measure similarity between shapes. The measurement of similarity is preceded by solving for correspondence between points on the two shapes and using the correspondences to estimate an aligning transform. The dissimilarity between the two shapes is computed as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform. We find the final shape is closest to the truncated tetrahedron with truncation 0.59, as shown in Fig. A1b.

However, digital alchemy was recently proposed and it's able to design optimal shapes that minimize the free energy of target structures [93]. The study shows that the truncated tetrahedra with truncation 0.63 is the best shape for diamond structure at packing density 0.6. The evolutionary algorithm method can find a shape to assemble a target structure, but not necessarily the best shape. So we change our design idea to the digital alchemy framework, as shown in previous sections 3.1- 3.6.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Umang Agarwal and Fernando A. Escobedo. Mesophase behaviour of polyhedral particles. *Nat. Mater.*, 10:230–235, 2011.

[2] Ankit Agrawal and Alok Choudhary. Perspective: Materials informatics and big data: Realization of the âĂIJfourth paradigmâĂİ of science in materials science. *APL Materials*, 4(5):053208, 2016.

[3] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27(5):1208–1209, 1957.

[4] Joshua A. Anderson and Sharon C. Glotzer. The Development and Expansion of HOOMD-Blue Through Six Years of GPU Proliferation. 2013. `http://codeblue.umich.edu/hoomd-blue`.

[5] Joshua A Anderson, M Eric Irrgang, and Sharon C Glotzer. Scalable metropolis monte carlo for simulation of hard shapes. *Comp. Phys. Commun.*, 204:21 – 30, 2016.

[6] Kellie J Archer and Ryan V Kimes. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4):2249–2260, 2008.

[7] Shixiong Bao, Xuan Yang, Ming Luo, Shan Zhou, Xue Wang, Zhaoxiong Xie, and Younan Xia. Shape-controlled synthesis of co-free pd nanocrystals with the use of formic acid as a reducing agent. *Chemical Communications*, 52(85):12594–12597, 2016.

[8] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. Technical report, CALIFORNIA UNIV SAN DIEGO LA JOLLA DEPT OF COMPUTER SCIENCE AND ENGINEERING, 2002.

[9] Charles H. Bennett. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comp. Phys.*, 22(2):245–268, 1976.

[10] Emanuela Bianchi, Günther Doppelbauer, Laura Filion, Marjolein Dijkstra, and Gerhard Kahl. Predicting patchy particle crystals: Variable box shape simulations and evolutionary algorithms. *The Journal of Chemical Physics*, 136(21):214102, 2012.

[11] Ludwig Boltzmann. *Vorlesungen über Gastheorie*. Johann Ambrosius Barth, Leipzig, 1896.

[12] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[13] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[14] Ludovico Cademartiri, Kyle J. M. Bishop, Phillip W. Snyder, and Geoffrey A. Ozin. Using shape for self-assembly. *Philos. Trans. R. Soc., A*, 370(1969):2824–2847, 2012.

[15] Elizabeth R. Chen, Daphne Klotsa, Michael Engel, Pablo F. Damasceno, and Sharon C. Glotzer. Complexity in surfaces of densest packings for families of polyhedra. *Phys. Rev. X*, 4:011024, Feb 2014.

[16] Yan-Xin Chen, Sheng-Pei Chen, Zhi-You Zhou, Na Tian, Yan-Xia Jiang, Shi-Gang Sun, Yong Ding, and Zhong Lin Wang. Tuning the shape and catalytic activity of fe nanocrystals from rhombic dodecahedra and tetragonal bipyramids to cubes by electrochemistry. *J. Am. Chem. Soc.*, 131(31):10860–10862, 2009.

[17] Chin-Yi Chiu, Yujing Li, Lingyan Ruan, Xingchen Ye, Christopher B Murray, and Yu Huang. Platinum nanocrystals selectively shaped using facet-specific peptide sequences. *Nature Chemistry*, 3(5):393, 2011.

[18] Rudolph Clausius. *Abhandlungen Über Die Mechanische Wärmetheorie*. Friedrich Vieweg und Sohn, Braunschweig, 1864.

[19] Pablo F. Damasceno, Michael Engel, and Sharon C. Glotzer. Crystalline Assemblies and Densest Packings of a Family of Truncated Tetrahedra and the Role of Directional Entropic Forces. *ACS Nano*, 6(1):609–614, 2012.

[20] Pablo F. Damasceno, Michael Engel, and Sharon C. Glotzer. Predictive Self-Assembly of Polyhedra into Complex Structures. *Science*, 337(6093):453–457, 2012.

[21] Chrisy Xiyu Du, Greg van Anders, and Sharon C. Glotzer. *To appear*, 2018.

[22] Chrisy Xiyu Du, Greg van Anders, Richmond S. Newman, and Sharon C. Glotzer. Shape-driven colloidal crystal–crystal transitions. *Proc. Natl. Acad. Sci. U.S.A.*, 114:E3892–E3899, 2017.

[23] Michael Engel, Pablo F. Damasceno, Carolyn L. Phillips, and Sharon C. Glotzer. Computational Self-Assembly of a One-Component Icosahedral Quasicrystal. *Nat. Mater.*, 14:109–116, 2015.

[24] Fernando A. Escobedo. Engineering entropy in soft matter: the bad, the ugly and the good. *Soft Matter*, 10:8388–8400, 2014.

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

[26] Daan Frenkel. New monte carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres. *J. Chem. Phys.*, 81(7):3188–3193, 1984.

[27] Daan Frenkel. Entropy-driven phase transitions. *Physica A: Statistical Mechanics and its Applications*, 263(1–4):26 – 38, 1999.

[28] Daan Frenkel. Order through entropy. *Nat. Mater.*, 14:9–12, 2015.

[29] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:, 2001.

[30] Anjan P. Gantapara, Joost De Graaf, Ren?? Van Roij, and Marjolein Dijkstra. Phase diagram and structural diversity of a family of truncated cubes: Degenerate close-packed structures and vacancy-rich states. *Physical Review Letters*, 111(1):1–5, 2013.

[31] Yina Geng, Greg van Anders, Paul M. Dodd, Julia Dshemuchadse, and Sharon C. Glotzer. Engineering entropy for the inverse design of colloidal crystals from hard shapes. 2017.

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

[33] Luca M. Ghiringhelli, Jan Vybiral, Sergey V. Levchenko, Claudia Draxl, and Matthias Scheffler. Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.*, 114:105503, Mar 2015.

[34] Josiah Willard Gibbs. *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, New York, 1902.

[35] Sharon C. Glotzer and Michael J. Solomon. Anisotropy of building blocks and their assembly into complex structures. *Nat. Mater.*, 6:557–562, 2007.

[36] Zhe Gong, Theodore Hueckel, Gi-Ra Yi, and Stefano Sacanna. Patchy particles made by colloidal fusion. *Nature*, 550(7675):234–238, oct 2017.

[37] Baptiste Gregorutti, Bertrand Michel, and Philippe Saint-Pierre. Correlation and variable importance in random forests. *Statistics and Computing*, 27(3):659–678, 2017.

[38] Marek Grzelczak, Jorge Perez-Juste, Paul Mulvaney, and Luis M. Liz-Marzan. Shape control in gold nanoparticle synthesis. *Chem. Soc. Rev.*, 37:1783–1791, 2008.

[39] Amir Haji-Akbari, Michael Engel, and Sharon C. Glotzer. Phase diagram of hard tetrahedra. *J. Chem. Phys.*, 135(19):194101, 2011.

[40] Amir Haji-Akbari, Michael Engel, Aaron S. Keys, Xiaoyu Zheng, Rolfe G. Petschek, Peter Palffy-Muhoray, and Sharon C. Glotzer. Disordered, Quasicrystalline and Crystalline Phases of Densely Packed Tetrahedra. *Nature*, 462:773–777, 2009.

[41] A. B. Harris, Randall D. Kamien, and T. C. Lubensky. Molecular chirality and chiral parameters. *Rev. Mod. Phys.*, 71:1745–1757, Oct 1999.

[42] Joel Henzie, Michael Grünwald, Asaph Widmer-Cooper, Phillip L Geissler, and Peidong Yang. Self-assembly of uniform polyhedral silver nanocrystals into densest packings and exotic superlattices. *Nature materials*, 11(2):131, 2012.

[43] Joel Henzie, Michael Grünwald, Asaph Widmer-Cooper, Phillip L. Geissler, and Peidong Yang. Self-assembly of uniform polyhedral silver nanocrystals into densest packings and exotic superlattices. *Nat. Mater.*, 11:131–137, 2012.

[44] Jin-Yi Ho and Michael H Huang. Synthesis of submicrometer-sized cu2o crystals with morphological evolution from cubic to hexapod structures and their comparative photocatalytic activity. *J. Phys. Chem. C*, 113(32):14159–14164, 2009.

[45] Avni Jain, Jeffrey R. Errington, and Thomas M. Truskett. Inverse design of simple pairwise interactions with low-coordinated 3d lattice ground states. *Soft Matter*, 9:3866–3870, 2013.

[46] Avni Jain, Jeffrey R. Errington, and Thomas M. Truskett. Dimensionality and design of isotropic interactions that stabilize honeycomb, square, simple cubic, and diamond lattices. *Phys. Rev. X*, 4:031049, Sep 2014.

[47] Franklin Kim, Stephen Connor, Hyunjoon Song, Tevye Kuykendall, and Peidong Yang. Platonic gold nanocrystals. *Angew. Chem., Int. Ed.*, 43(28):3673–3677, 2004.

[48] Daphne Klotsa, Elizabeth R Chen, Michael Engel, and Sharon C Glotzer. Intermediate crystalline structures of colloids in shape space. *Soft Matter*, sep 2018.

[49] Fan Li, David P. Josephson, and Andreas Stein. Colloidal assembly: The road from particles to colloidal molecules and crystals. *Angew. Chem., Int. Ed.*, 50(2):360–388, 2011.

[50] Byungkwon Lim, Majiong Jiang, Jing Tao, Pedro HC Camargo, Yimei Zhu, and Younan Xia. Shape-controlled synthesis of pd nanocrystals in aqueous solutions. *Advanced Functional Materials*, 19(2):189–200, 2009.

[51] Haixin Lin, Sangmin Lee, Lin Sun, Matthew Spellings, Michael Engel, Sharon C Glotzer, and Chad A Mirkin. Clathrate colloidal crystals. *Science*, 355(6328):931–935, 2017.

[52] Hod Lipson and Jordan B Pollack. Automatic design and manufacture of robotic lifeforms. *Nature*, 406(6799):974, 2000.

[53] Maochang Liu, Yiqun Zheng, Lei Zhang, Liejin Guo, and Younan Xia. Transformation of pd nanocubes into octahedra with controlled sizes by maneuvering the rates of etching and regrowth. *Journal of the American Chemical Society*, 135(32):11752–11755, 2013.

[54] Shu-Ya Liu, Yuan-Ting Shen, Chun-Ya Chiu, Sourav Rej, Po-Heng Lin, Yu-Chi Tsao, and Michael H Huang. Direct synthesis of palladium nanocrystals in aqueous solution with systematic shape evolution. *Langmuir*, 31(23):6538–6545, 2015.

[55] Wenyan Liu, Miho Tagawa, Huolin L Xin, Tong Wang, Hamed Emamy, Huilin Li, Kevin G Yager, Francis W Starr, Alexei V Tkachenko, and Oleg Gang. Diamond family of nanoparticle superlattices. *Science*, 351(6273):582–586, 2016.

[56] Yue Liu, Tianlu Zhao, Wangwei Ju, and Siqi Shi. Materials discovery and design using machine learning. *Journal of Materiomics*, 3(3):159 – 177, 2017.

[57] Andrew W. Long, Carolyn L. Phillips, Eric Jankowksi, and Andrew L. Ferguson. Nonlinear machine learning and design of reconfigurable digital colloids. *Soft Matter*, 12:7119–7135, 2016.

[58] Fang Lu, Kevin G. Yager, Yugang Zhang, Huolin Xin, and Oleg Gang. Superlattices assembled through shape-induced directional binding. *Nat. Commun.*, 6:6912, 2015.

[59] AP Lyubartsev, AA Martsinovski, SV Shevkunov, and PN Vorontsov-Velyaminov. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, 96(3):1776–1783, 1992.

[60] Robert J. Macfarlane, Matthew R. Jones, Andrew J. Senesi, Kaylie L. Young, Byeongdu Lee, Jinsong Wu, and Chad A. Mirkin. Establishing the Design Rules for DNA-Mediated Programmable Colloidal Crystallization. *Angew. Chem., Int. Ed.*, 49(27):4589–4592, 2010.

[61] Vinothan N. Manoharan. Colloidal matter: Packing, geometry, and entropy. *Science*, 349(6251):942, 2015.

[62] Mathew M. Maye, Mudalige Thilak Kumara, Dmytro Nykypanchuk, William B. Sherman, and Oleg Gang. Switching Binary States of Nanoparticle Superlattices and Dimer Clusters by DNA Strands. *Nat. Nano.*, 5:116–120, 2010.

[63] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B*, 89:094104, Mar 2014.

[64] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[65] Chad A. Mirkin, Robert L. Letsinger, Robert C. Mucic, and James J. Storhoff. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature*, 382(6592):607–609, 1996.

[66] Marc Z Miskin and Heinrich M Jaeger. Adapting granular materials through artificial evolution. *Nature materials*, 12(4):326, 2013.

[67] Wenxin Niu, Ling Zhang, and Guobao Xu. Shape-controlled synthesis of single-crystalline palladium nanocrystals. *ACS Nano*, 4(4):1987–1996, 2010.

[68] Dmytro Nykypanchuk, Mathew M. Maye, Daniel van der Lelie, and Oleg Gang. DNA-guided crystallization of colloidal nanoparticles. *Nature*, 451(31):549–552, January 2008.

[69] Artem R Oganov, Andriy O Lyakhov, and Mario Valle. How evolutionary crystal structure prediction worksîŮÿ and why. *Accounts of chemical research*, 44(3):227–237, 2011.

[70] Lars Onsager. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences*, 51(4):627–659, 1949.

[71] Geoffrey A Ozin, Kun Hou, Bettina V Lotsch, Ludovico Cademartiri, Daniel P Puzzo, Francesco Scotognella, Arya Ghadimi, and Jordan Thomson. Nanofabrication by self-assembly. *Materials Today*, 12(5):12–23, 2009.

[72] Sung Yong Park, Abigail K. R. Lytton-Jean, Byeongdu Lee, Steven Weigand, George C. Schatz, and Chad A. Mirkin. DNA-Programmable Nanoparticle Crystallization. *Nature*, 451(7178):553–556, 2008.

[73] Linus Pauling. The principles determining the structure of complex ionic crystals. *J. Am. Chem. Soc.*, 51(4):1010–1026, 1929.

[74] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[75] Maria T Perez Cardenas, Chuncai Kong, Jie He, Samantha Litvin, Melissa L Meyerson, and Zhihong Nie. Immobilized seed-mediated growth of two-dimensional array of metallic nanocrystals with asymmetric shapes. *ACS Nano*, 2018.

[76] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim. Machine Learning and Materials Informatics: Recent Applications and Prospects. 2017.

[77] Laura Rossi, Stefano Sacanna, William T. M. Irvine, Paul M. Chaikin, David J. Pine, and Albert P. Philipse. Cubic crystals from cubic colloids. *Soft Matter*, 7:4139–4142, 2011.

[78] Laura Rossi, Stefano Sacanna, William TM Irvine, Paul M Chaikin, David J Pine, and Albert P Philipse. Cubic crystals from cubic colloids. *Soft Matter*, 7(9):4139–4142, 2011.

[79] Stefano Sacanna, Mark Korpics, Kelvin Rodriguez, Laura Colon-Melendez, Seung-Hyun Kim, David J. Pine, and Gi-Ra Yi. Shaping colloids for self-assembly. *Nat. Commun.*, 4:1688, 2013.

[80] Stefano Sacanna and David J. Pine. Shape-anisotropic colloids: Building blocks for complex assemblies. *Curr. Opin. Colloid Interface Sci.*, 16(2):96 – 105, 2011.

[81] Stefano Sacanna, David J. Pine, and Gi-Ra Yi. Engineering shape: the novel geometries of colloidal self-assembly. *Soft Matter*, 9:8096–8106, 2013.

[82] Benjamin A. Schultz, Pablo F. Damasceno, Michael Engel, and Sharon C. Glotzer. Symmetry Considerations for the Targeted Assembly of Entropically Stabilized Colloidal Crystals via Voronoi Particles. *ACS Nano*, 9(3):2336–2344, 2015.

[83] C.E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, 623–656, 1948.

[84] Elena V. Shevchenko, Dmitri V. Talapin, Nicholas A. Kotov, Stephen O'Brien, and Christopher B. Murray. Structural diversity in binary nanoparticle superlattices. *Nature Letters*, 439:55–59, 2006.

[85] Sara E Skrabalak, Leslie Au, Xingde Li, and Younan Xia. Facile synthesis of ag nanocubes and au nanocages. *Nat. Protoc.*, (2):2182–2190, 2007.

[86] Babji Srinivasan, Thi Vo, Yugang Zhang, Oleg Gang, Sanat Kumar, and Venkat Venkatasubramanian. Designing DNA-grafted particles that self-assemble into desired crystalline structures using the genetic algorithm. *Proc. Nat. Acad. Sci. U.S.A.*, 110(46):18431–18435, 2013.

[87] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1):25, 2007.

[88] Andrea Tao, Prasert Sinsermsuksakul, and Peidong Yang. Polyhedral silver nanocrystals with distinct scattering signatures. *Angewandte Chemie International Edition*, 45(28):4597–4601, 2006.

[89] Andrea R. Tao, Susan Habas, and Peidong Yang. Shape control of colloidal metal nanocrystals. *Small*, 4(3):310–325, 2008.

[90] A Travesset. Topological structure prediction in binary nanoparticle superlattices. *Soft Matter*, 13(1):147–157, 2017.

[91] Greg van Anders, N. Khalid Ahmed, Ross Smith, Michael Engel, and Sharon C. Glotzer. Entropically patchy particles: Engineering valence through shape entropy. *ACS Nano*, 8:931–940, 2014.

[92] Greg van Anders, Daphne Klotsa, N. Khalid Ahmed, Michael Engel, and Sharon C. Glotzer. Understanding shape entropy through local dense packing. *Proc. Natl. Acad. Sci. U.S.A.*, 111:E4812–E4821, 2014.

[93] 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*, 9:9542–9553, 2015.

[94] Yi Wang, Shuifen Xie, Jingyue Liu, Jinho Park, Cheng Zhi Huang, and Younan Xia. Shape-controlled synthesis of palladium nanocrystals: a mechanistic understanding of the evolution from octahedrons to tetrahedrons. *Nano letters*, 13(5):2276–2281, 2013.

[95] W. W. Wood and J. D. Jacobson. Preliminary results from a recalculation of the monte carlo equation of state of hard spheres. *J. Chem. Phys.*, 27(5):1207–1208, 1957.

[96] Xiaohu Xia, Jie Zeng, L Kyle Oetjen, Qingge Li, and Younan Xia. Quantitative analysis of the role played by poly (vinylpyrrolidone) in seed-mediated growth of ag nanocrystals. *Journal of the American Chemical Society*, 134(3):1793–1801, 2012.

[97] Younan Xia, Yujie Xiong, Byungkwon Lim, and SaraâĂĔE. Skrabalak. Shape-controlled synthesis of metal nanocrystals: Simple chemistry meets complex physics? *Angew. Chem., Int. Ed.*, 48(1):60–103, 2009.

[98] Kaylie L Young, Michelle L Personick, Michael Engel, Pablo F Damasceno, Stacey N Barnaby, Reiner Bleher, Tao Li, Sharon C Glotzer, Byeongdu Lee, and Chad A Mirkin. A directional entropic force approach to assemble anisotropic nanoparticles into superlattices. *Angewandte Chemie International Edition*, 52(52):13980–13984, 2013.

[99] Kaylie L. Young, Michelle L. Personick, Michael Engel, Pablo F. Damasceno, Stacey N. Barnaby, Reiner Bleher, Tao Li, Sharon C. Glotzer, Byeongdu Lee, and Chad A. Mirkin. A directional entropic force approach to assemble anisotropic nanoparticles into superlattices. *Angew. Chem., Int. Ed.*, 52:13980–13984, 2013.

[100] Yue Yu, Qingbo Zhang, Bo Liu, and Jim Yang Lee. Synthesis of nanocrystals with variable high-index pd facets through the controlled heteroepitaxial growth of trisoctahedral au templates. *J. Am. Chem. Soc.*, 132(51):18258–18265, 2010.

[101] Junfang Zhang, Ronald Blaak, Emmanuel Trizac, José A Cuesta, and Daan Frenkel. Optimal packing of polydisperse hard-sphere fluids. *J. Chem. Phys.*, 110(11):5318–5324, 1999.

[102] Nan Zhao, Wei Ma, Zhimin Cui, Weiguo Song, Chuanlai Xu, and Mingyuan Gao. Polyhedral maghemite nanocrystals prepared by a flame synthetic method: Preparations, characterizations, and catalytic properties. *ACS Nano*, 3(7):1775–1780, 2009.

[103] Yiqun Zheng, Wenying Liu, Tian Lv, Ming Luo, Hefei Hu, Ping Lu, Sang-Il Choi, Chao Zhang, Jing Tao, Yimei Zhu, et al. Seed-mediated synthesis of gold tetrahedra in high purity and with tunable, well-controlled sizes. *Chemistry–An Asian Journal*, 9(9):2635–2640, 2014.

[104] Cun Zhu, Jie Zeng, Ping Lu, Jingyue Liu, Zhongze Gu, and Younan Xia. Aqueous-phase synthesis of single-crystal pd seeds 3 nm in diameter and their use for the growth of pd nanocrystals with different shapes. *Chemistry-A European Journal*, 19(16):5127–5133, 2013.