Machine Learning for Automatic Structure Analysis and Experimental Design

by

Matthew Spellings

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

Doctoral Committee:

Professor Sharon C Glotzer, Chair
Assistant Professor Bryan Goldsmith
Professor Ronald G Larson
Assistant Professor Xiaoming Mao

Matthew Spellings

mspells@umich.edu
ORCID iD: 0000-0002-4061-4299

Table of Contents

List of Figures

List of Tables

Abstract

Self-assembly is a ubiquitous process that holds great promise for the design and engineering of new materials and systems. While chemistry today is largely based on combining a few building blocks into molecules with desirable properties, recent advances in colloidal and nanoscale self-assembly have allowed us to move beyond the periodic table of elements to design building blocks with attributes tailored to their desired applications. With this great flexibility, however, comes a cost: both experiments and models of these systems are often laden with many tunable parameters, frustrating analysis and engineering efforts. Furthermore, it is often not known whether only a subset of these parameters is important or if observed behaviors depend on the confluence of multiple variables. In this work, we show examples of complex design spaces for colloidal and nanoscale self-assembly—including systems of far-from-equilibrium, anisotropic particles. We further show how machine learning can be applied to two major problems involved in studying self-assembly *in silico*: analyzing three-dimensional structure and engineering building blocks with many design variables.

Chapter 1

Introduction to Self-Assembly Studies

1.1 Basic Self-Assembly

The world is full of self-organizing phenomena. For objects ranging in length scale from atomic to galactic, we often base our understanding of matter on its spatial configuration. Indeed, one of the key tenets of materials science is the idea that the structure of a material plays an enormous role in its final properties. One alluring idea of engineering materials at the nanoscale is through bottom-up design of self-assembling particles: by carefully crafting the building blocks of a system, we can control how they come together to form larger-scale structures. Creating materials at this scale holds great promise: in contrast to the building blocks available in nature—the elements of the periodic table—we have enormous freedom in how we design colloidal and nanoscale building blocks. However, there are also significant challenges with engineering these systems. At these length scales, the bulk behavior of particles often depends on a complex interplay of interactions, many of which are difficult or impossible to measure or reliably control in experiment. Much of our understanding in this realm is guided by thermodynamics and statistical mechanics. Computational models of systems of these particles are vital for not only dissecting which features of building blocks are important to obtain the observed behavior, but also to virtually screen the broad design parameter space available to experimentalists for their particles. For these reasons, computational experiments have become an important component of the scientific study of colloidal- and nanoscale self-assembly.

Some of the earliest computational studies of self-assembly focused on systems of hard spheres, where particles do not interact in any way except that they are not allowed to overlap. These were studied both through Monte Carlo methods[1] (MC) and *via* molecular dynamics[2] (MD) in the 1950s. Briefly, Metropolis Monte Carlo methods attempt to sample the equilibrium probability distribution of a system by iteratively proposing a change to the microscopic state of the system—for example, moving a particle by a small amount—and accepting or rejecting the change randomly in order to generate the correct statistical mechanical ensemble. In contrast, molecular dynamics methods compute forces on particles and integrate their motion over time. Each method has its own set of benefits and limitations for solving the problems we encounter in the field. Not long after the studies of hard spheres, square well[3] and Lennard-Jones[4] type interactions were used in MC and MD simulations, allowing researchers to capture many of the effects of interparticle attraction.

The decades since these first studies have seen many things drastically change, while others have remained remarkably constant. Computers have become orders of magnitude more powerful and codes immensely more complex, but the Lennard-Jones interaction has remained as a workhorse of coarse-grained modeling, in realms from materials science[5] to biology[6]. Rather than performing a few simulations of a few hundred particles for a study, researchers now routinely run hundreds of simulations, each with thousands to millions of particles. Performing simulations on this scale brings forth a new set of challenges in terms of data management and analysis, and dealing with these challenges is one of the key focuses of my work.

1.2 Shapes and Other Complex Interactions

It seems logical that more complex building blocks would be able to exhibit more complex behavior. But how can we intelligently modulate the complexity of our particles? One strategy is to formulate more complex isotropic pair potentials. In contrast to Lennard-Jones interactions, these may include many wells or other features[7, 8]. However, with this complexity comes the ability to form a wide variety of simple and complex phases by tuning relatively few parameters[9, 10, 11, 12, 13].

Another way to encode complexity in particle interactions is through anisotropy[14]: particles can be made with different interactions on each side[15] or with particular shapes, commonly polyhedra[16]. The results from these studies are especially interesting: although they have no interactions except excluded volume, researchers have found that hard polyhedral-shaped particles can form a variety of structures[17], including a quasicrystal[18]. This research has opened an entire avenue of study into the behavior of hard polyhedral particles, including systematically studying the effects of shape on structure[17, 19] as well as other material properties, such as photonic activity[20].

Recent experimental work has found several systems where both particle shape and soft or sticky interactions between the surfaces of particles cooperate to determine the final assembly behavior. This could be as simple as the grafting density of molecules on the surface of particles affecting the assembled crystal structure[21]. More elaborate methods involve coating particles in particularly-designed double-stranded or single-stranded DNA, which often allows researchers to dictate symmetry and specificity of the resulting local environment of particles[22, 23]. These strategies hold great promise for the "inverse design" problem of formulating building blocks based on desired structural properties or other attributes.

1.3 Summary: Complexity in Self-Assembly

We have seen that colloidal- and nanoscale particle design can be extremely complex, even when systems are governed by relatively simple rules. Particles can be changed in numerous ways, any of which could have a drastic impact on how they assemble, if they assemble at all. The situation becomes even more complex when systems are driven away from equilibrium, removing equilibrium statistical mechanics from the list of useful conceptual tools. To illustrate one simple change that drastically impacts assembly behavior, I present some of my work on two-dimensional, active systems in Chapter 2. These particles have shape and are rotationally-driven, which makes it difficult to make detailed predictions about their steady-state behavior based on statistical mechanics. Nevertheless, mixtures of clockwise- and counterclockwise-spinning particles segregate into similarly-rotating phases in a process very similar to equilibrium spinodal decomposition of a binary mixture[24]. One of the things I studied is how these particles behave under confinement within a membrane, which turns out to be determined by a confluence of static and dynamic factors.

The second main topic of my work focuses on applying machine learning to three-dimensional structures that we find in self-assembling systems. While our ability to easily perform many experiments in a high-throughput manner is one of the great strengths of modeling and computational experiments, often the difficult structural analysis step—particularly when complex structures are involved—relies on manual effort. Defects and occlusion can also conspire to make it especially difficult to determine how a system has ordered in three dimensions. Ideally we would develop and use reliable characterization algorithms that are powerful enough to automatically analyze these systems, but it is difficult to create flexible and robust numerical descriptions of three-dimensional structure. I present an approach using machine learning (ML) to generate algorithms that can reliably identify complex structures in Chapter 3. At its core, ML simply consists of formulating models that characterize our observations—in this case, linking characteristics of local particle environments to the structures they form. The ML methods I describe here are currently being used by my colleagues to automatically identify structures for a variety of scientific purposes. Other applications of ML to adjacent scientific fields could include predicting material properties[25], accelerating simulations[26], or directly designing molecules[27, 28]. In Chapter 4, I show how the previously-described ML algorithms can be extended to incrementally analyze the phase behavior of a study and perform new experiments in promising regions of parameter space, making higher-dimensional studies involving more tunable parameters feasible.

Chapter 2

Nonequilibrium Self-Assembly

2.1 Shape Control and Compartmentalization in Active Colloidal Cells

Note: this section was originally published in the Proceedings of the National Academy of Sciences in 2015[29] and is reproduced with minor modifications below.

Active matter describes particulate systems with the characteristic that each "particle" (agent) converts energy into motion [30, 31]. Active matter covers a range of length scales that include molecular motors in the cytoskeleton [32, 33, 34], swimming bacteria [35, 36, 37], driven colloids [38, 39], flocks of birds and fish [40, 41, 42, 43], and people and vehicles in motion [44]. Over the last decade, studies of active matter have demonstrated behavior not seen in equilibrium systems, including giant number fluctuations [45, 46], emergent attraction and superdiffusion [47, 48, 49], clustering [50, 51], swarming [52, 53, 54, 55, 56], and self-assembled motifs [57, 58]. These systems provide interesting theoretical and engineering challenges as well as opportunities to explore and target novel behaviors that proceed outside of thermodynamic equilibrium.

Of particular interest are systems found in nature or inspired by natural phenomena. Biological systems usually operate in confined regions of space—think of intracellular space, interfaces and membranes, and the crowding of cells near surfaces. The role of hydrodynamics in confinement has been studied for biological swimmers, such as bacteria and sperm, showing accumulation at the walls [59, 60, 61] and upstream swimming along surfaces [62] or in a spiral vortex [63, 64, 65]. Attraction to walls has also been reported in the absence of hydrodynamics for disks [66, 67], spheres [68] and dumb-bell swimmers [69]. But while these examples study the behavior under the influence of hard boundaries, biological swimmers typically interact with soft boundaries, such as membranes and biofilms. Another design variable is the possibility that the boundary itself is active, as in the surface of a bacterium covered with flagellae or, as demonstrated recently, active nematic vesicles [70].

In this work, we propose and investigate an active matter system under flexible, active confinement. We call this system an active colloidal cell. Our realization of an active colloidal cell consists of independent particles, called spinners [24], that translate and rotate in two dimensions and are constrained within a finite area by a flexible boundary that is also built from spinners. Each spinner has a gear-like geometry, which consists of a large central disk and four smaller satellite disks (Fig. 2.1a). Similar gear-shaped rigid aggregates of self-propelled particles have been formed experimentally [71]. Spinners are freely mobile in the cell interior. On the cellular boundary, spinners are connected to one another by a flexible chain of beads attached by finitely extensible springs. Both the interior and the boundary spinners can be subject to a clockwise or counterclockwise driving torque, which makes them active.

Rotationally-driven particles can synchronize and self-organize [72, 73] in the absence [24] and in the presence [74, 75, 76] of hydrodynamic interactions. Crystallization has recently been observed in rotating magnetic Janus colloids [77] and fast-moving bacteria [78]. Spinners in the interior of the cell resemble molecular motors that push themselves forward on their neighbors and, thus, sustain convective dynamics. The effect of the boundary spinners is similar to that found in the cilia of living tissues, which stir nearby fluid. Our results demonstrate that a natural consequence of the activity present in the colloidal cell is control over both its external shape and internal structure. We report compartmentalization into regions of clockwise and counterclockwise spinners—a behavior which is affected by, and can be controlled via, properties of the enclosing boundary configuration as previously suggested [79]. Transitions in the internal structure of the colloidal cell occur as its radius increases, and as the composition of the interior spinners and the patterning of the boundary are varied.

A previous study of spinners in bulk [24] showed phase separation into clockwise- and counterclockwise domains. Cates and collaborators [35, 80, 81] have suggested that phase separation is a generic consequence of local energy input in an otherwise equilibrium system. Here and in the study of bulk spinners we demonstrate phase separation due to local rotational, rather than translational, energy input. We obtain our results using a particulate, microscopic model (Fig. 2.1c) as well as a continuum model (Fig. 2.1b). This allows us to conclude that the phenomena we observe are robust with respect to details of the model.

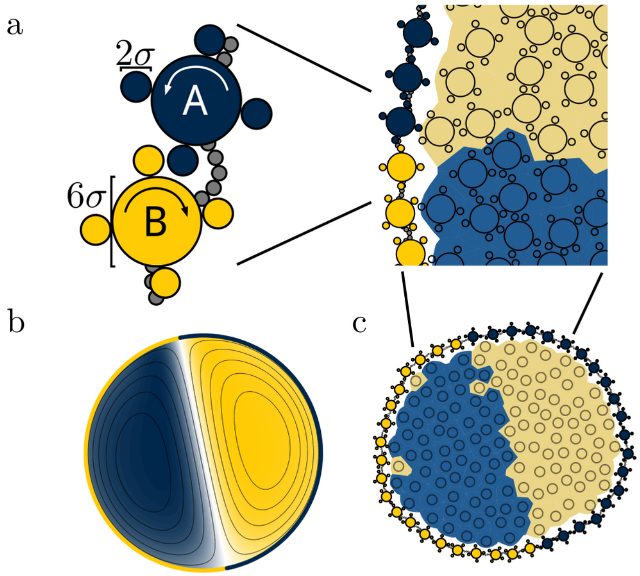


Figure 2.1: **Schematic of the confined spinner models.** (a) Our active colloidal cell is made up of spinners driven counterclockwise (blue) or clockwise (yellow). Boundary spinners are connected by a flexible bead-spring chain (gray). We compare the behavior of a continuum model (b) to a microscopic model (c). The compartmentalization of interior spinners is visualized by coloring the Voronoi tessellation in the microscopic model.

In this study we use two models to study the behavior of an active colloidal cell, illustrated in Figure 2.1. The *microscopic model* describes spinners as individual particles and simulates their motion using Langevin dynamics. It resolves the behavior of individual spinners but does not include hydrodynamic effects. In contrast, the *continuum model* describes the spinner system as a viscous binary fluid, which is governed by an incompressible Navier-Stokes equation coupled to a Cahn-Hilliard equation. Both models are described in detail in the Materials and Methods section below. Note that the microscopic model was introduced in earlier work using Brownian dynamics [24] and is extended here to include boundaries.

2.1.1 Results

Shape Control from Active Confinement

We first study the behavior of colloidal cells with passive (non-driven) spinners in the interior and active spinners on the boundary. We use the microscopic model while varying two parameters: the number of boundary segments n and the driving torque on the boundary τ. As shown in Fig. 2.2a and SI Movie 1, the effect of the active boundary is a deformation of the cell shape. The shape deformation follows the symmetry of the boundary pattern (horizontal axis in the figure) and becomes more prominent as the driving torque τ increases (vertical axis). Buckling occurs at places on the boundary where the direction of the driving torque switches. In particular, we observe inward buckling when two adjacent spinners on the boundary push interior spinners away from the space between them. Similarly, we find outward buckling when the boundary spinners pull interior spinners toward the space between them. Colloidal cells with active spinners in the interior display similar, but less well-pronounced behavior.

In order to understand the deformation of the active colloidal cell, we analyze the velocity field of the passive spinners in the interior while fixing the geometry of the boundary. After reaching a steady-state, we observe that the flow field has developed regions of counterclockwise and clockwise convection, which we visualize using the vorticity field w and the resulting streamlines in Fig. 2.2b. We apply the same color scheme for vorticity (blue for counterclockwise and yellow for clockwise) in the continuum model as for the rotation of individual spinners in the microscopic model. Note that counterclockwise (clockwise) flow is exclusively in contact with a clockwise (counterclockwise) rotating boundary.

We now use the continuum model to study the deformation of the colloidal cell. The active boundary drives convective flow along the cell wall. Because the passive interior spinners are transported fastest along the interface, they collect at places where they turn to flow inward (*i.e.* where the boundary changes from yellow to blue traveling counterclockwise) inducing a positive pressure on the boundary. The result is a higher pressure and outward buckling of the boundary. In contrast, at the other junction, passive spinners are transported away rapidly when they approach the boundary from the center of the cell. Thus, the boundary buckles inwards at places where the boundary activity changes from blue to yellow traveling counterclockwise. The magnitude of the driving torque τ affects the strength of the pressure difference and thus the anisotropy of the cell shape.

To complete the comparison with the microscopic model, we release the boundary in the continuum model, adjusting its geometry based on the stresses acting on it from the interior fluid. We observe in Fig. 2.2c that the cell shape readily adjusts to a shape predicted by the microscopic simulations, confirming that the observed buckling is independent of using a microscopic (particulate) or a continuum model.

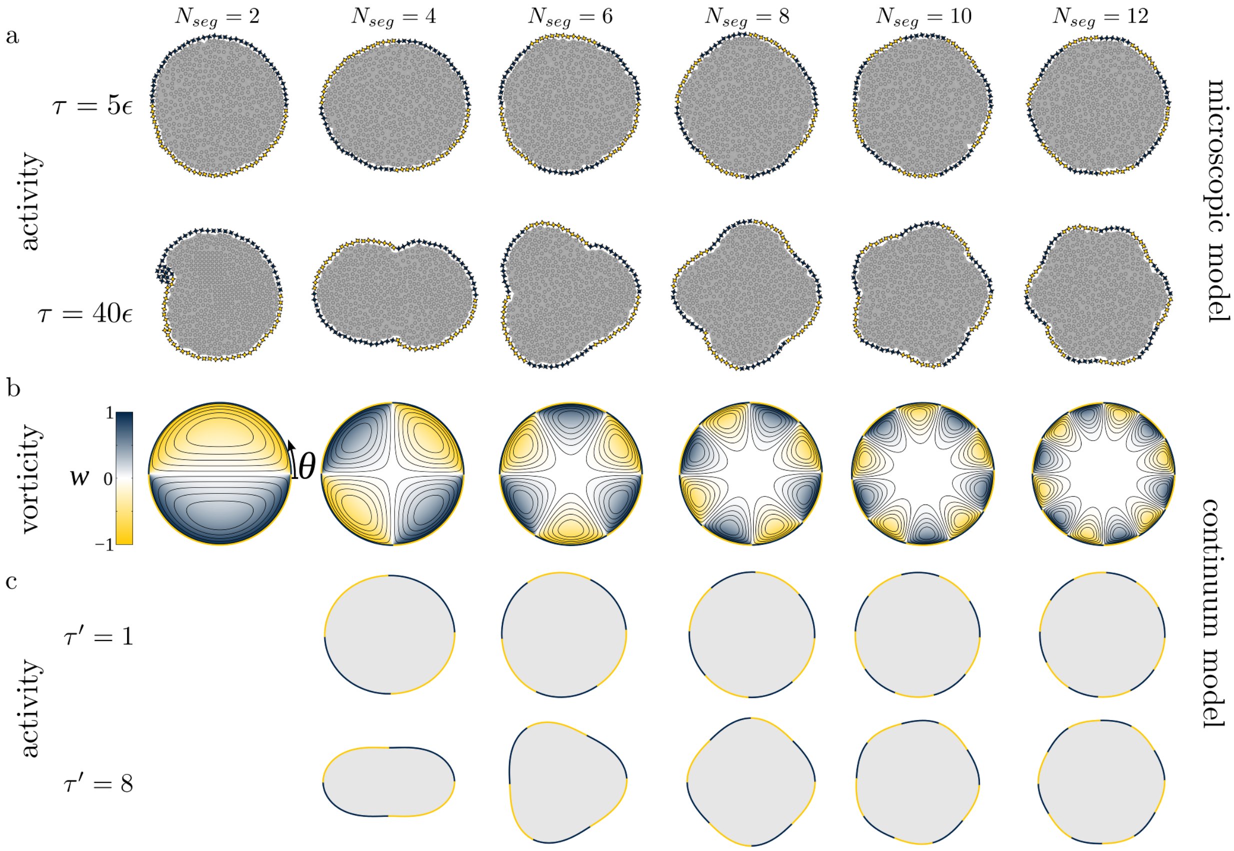


Figure 2.2: **Cellular shape control for active boundaries with passive interior.** (a) Simulations in the microscopic model with 512 interior spinners and 80 boundary spinners reveal a symmetric buckling of the colloidal cell. The buckling is suppressed for low driving torque τ (top row) but appears if the driving torque is sufficiently high (bottom row). We can control the symmetry by changing the number of alternately-driven segments on the active boundary, varied horizontally. (b) Simulations in the continuum model with fixed boundaries reveal convective flows of the interior spinners, which we visualize via the vorticity field w. (c) Simulations in the continuum model with free boundaries confirm the shape changes observed in the microscopic model. Again, we show results for two different levels of activity τ' for frictional damping $\gamma' = 0.1$ and boundary tension $\kappa = 80$ in a cell of size $R = 20$. The case of $n = 1$ exhibits a cusp-like singularity that cannot be captured by the perturbation analysis used to compute the cell shape (see Materials and Methods).

Compartmentalization

We next consider what happens if active spinners are confined within active boundaries. It is known that without confinement phase separation through a spinodal decomposition-like process eventually results in complete demixing of clockwise- and counterclockwise-driven spinners [24]. As we will see, the presence of an active boundary still allows phase separation, but also induces a preference of oppositely-driven interior and boundary spinners to be in contact near the boundary. We term this behavior compartmentalization. The presence of the active boundary can lead to more complex phase behavior than that found in the bulk system. The size of the colloidal cell also plays an important role for compartmentalization.

We systematically vary the two composition ratios, *i.e.* the fraction of clockwise- to counterclockwise-driven spinners in the interior (horizontal direction in Fig. 2.3) as well as the fraction on the boundary (vertical direction). We find perfect agreement between the microscopic model and the continuum model. The precise geometry of compartmentalization changes under variation of the composition ratios. We distinguish three cases:

(1) Core-shell. All boundary spinners are driven in the same direction and the domain interface forms a circle concentric to the boundary. We call this the core-shell configuration. The core-shell configuration maximizes the contact between interior spinners and boundary spinners of the same type.

(2) Inverted Janus. For equal ratios of clockwise- to counterclockwise-driven spinners on the boundary and in the interior the domain interface is a straight line. We call this the inverted Janus configuration. The inverted Janus configuration maximizes the contact between unlike spinners at the cell boundary, as shown in SI Movie 2.

(3) Bubble-crescent. Intermediate to the extreme cases of core-shell and inverted Janus is the bubble-crescent configuration. In the bubble-crescent configuration one of the spinner species attempts to minimize its area into a circular domain while simultaneously avoiding contact with the boundary.

Compartmentalization can be understood as the result of competition between two effects. The observation of spinodal decomposition in the bulk system [24] suggests that like-driven spinners in the cell interior and boundary prefer to be in contact. We call this behavior the *boundary preference.* In addition, at an interface between clockwise- and counterclockwise-driven spinners, spinners develop a velocity profile flowing parallel to the interface. We call this behavior the *interface preference.*

Both boundary preference and interface preference are satisfied for a single-component active boundary, which explains the geometry of the core-shell case. In the case of a Janus

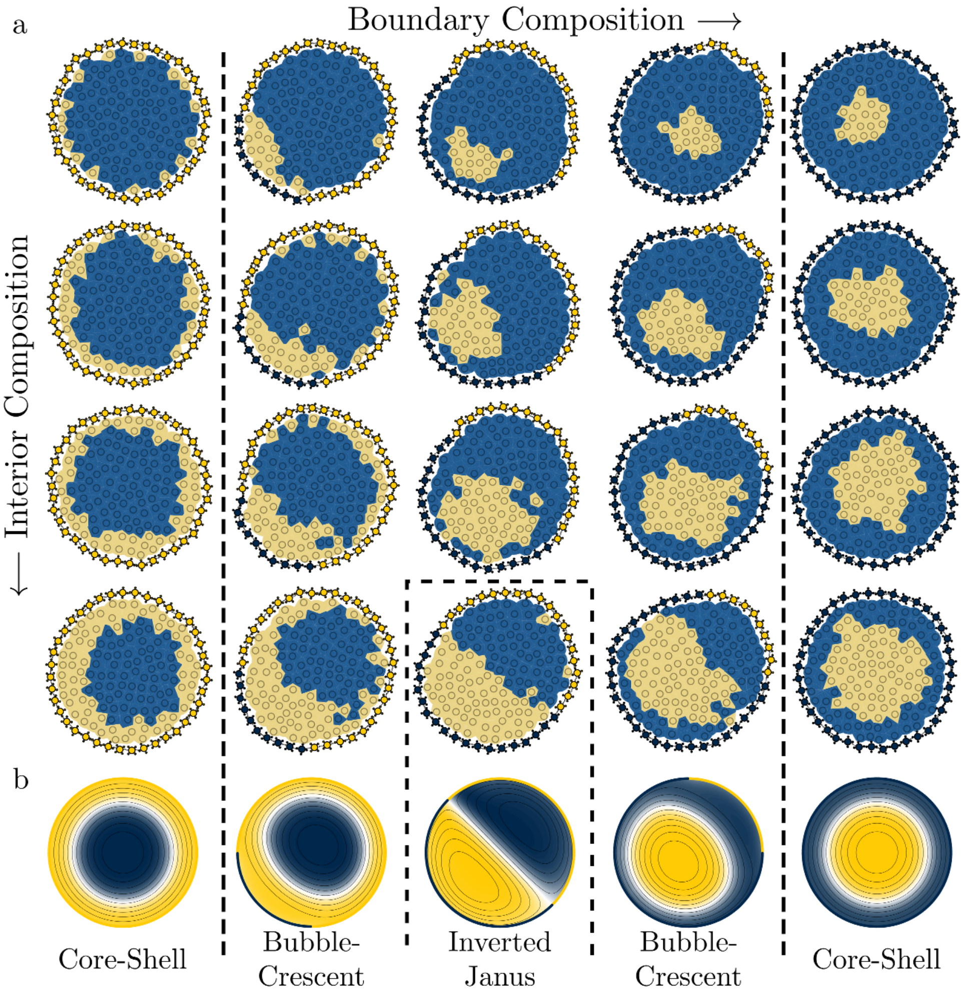


Figure 2.3: **Compartmentalization of a colloidal cell with active boundary and active interior.** (a) A grid of representative snapshots of active colloidal cells with varying boundary (horizontal direction) and interior (vertical direction) composition in the microscopic model. 128 interior spinners are enclosed by a boundary of 40 spinners. All spinners are active. We observe the core-shell and the inverted Janus configurations where the contact between like and unlike spinners, respectively, is maximized. The bubble-crescent configuration interpolates between these two extrema. (b) We confirm the steady state behavior in the continuum model with $\tau' = 8$ and $\gamma' = 0.1$ for cell of size $R = 10$ with different patterns of boundary activity. The configurations in (b) correspond to the bottom row of (a).

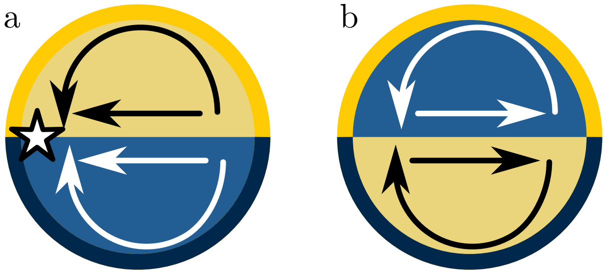


Figure 2.4: **Explanation for the inverted Janus configuration.** We compare the flows in the (non-observed) Janus configuration (a) to the (observed) inverted Janus configuration (b). In the inverted Janus configuration the flow maintains two circular vortices, while in the non-inverted Janus configuration the flow would converge to a singular point ($\star$).

boundary, boundary preference and interface preference work against each other and competition results. This can be understood from the schematic in Fig. 2.4. If the interior spinners were to phase separate into a regular Janus pattern (*i.e.* maximizing the contact between like spinners at the boundary) then the flows induced on the interior spinners by both the boundary and the interface between the two interior phases would converge to a single point, causing the entire colloidal cell to jam (Fig. 2.4a). In fact, if the boundary activity of an inverted Janus cell is instantaneously swapped to put the cell into a Janus configuration, the cell first jams and mixes before demixing into the inverted Janus configuration once more, as shown in SI Movie 3. By creating an inverted Janus configuration, the domain interface stabilizes a flow of spinners in the opposite direction of the pressure gradient imposed by the active boundaries and the circular flow of the spinners in the interior can be maintained (Fig. 2.4b). This behavior is similar to the cell sorting model reported in Ref. [66], where a mixture of self-propelled soft disks in confinement move towards the walls of the container. Finally, the bubble-crescent case is an intermediate case. Spinners are slowed down when they enter a region of unlike boundary contact resulting in the formation of the crescent. The boundary preference causes a layer of whichever species is dominant to form a wetting layer in contact with the cell boundary of the same species. In the perfectly balanced inverted Janus case, neither species is dominant and the time-averaged width of the wetting layer approaches zero.

Quantifying Intracellular Order and Finite-size Effects

To quantify the geometry of the domains and to distinguish the three types of colloidal cell compartmentalization, we construct a family of cellular order parameters m_n for nonnegative integral n corresponding to observed symmetries within the cell. Each spinner is assigned a fictitious charge c_j of $+1$ or -1 depending on whether it is being driven clockwise or counterclockwise. The order parameter m_n is defined as

$$m_n = \frac{1}{m_n^0} \left| \sum_j c_j r_j e^{in\theta_j} \right| , \tag{2.1}$$

where (r_j, θ_j) is the position of the interior spinner j in polar coordinates with the origin at the center of mass of the colloidal cell, and m_n^0 a normalization factor.

The order parameter m_0 is designed to be maximal when one species moves to the outside of the cell, *i.e.* for the core-shell configuration. The order parameters m_n, $n > 0$ are maximal for systems that phase separate into radial sectors with n-fold symmetry, which is the case for the n-fold alternating boundary of Eq. 2.3 For example, m_0 measures radial asymmetry and m_1 measures dipolar order.

We find that the type of intracellular compartmentalization not only depends on the interior and boundary spinner compositions but also on the size of the colloidal cell (SI Movie 4). We simulate cells with an even composition of spinners at the boundary and in the interior. It is apparent from the order parameter histograms in Fig. 2.5a that small cells do not order well due to the stronger influence of noise. As the number of interior spinners increases to $N = 128$, the effect of noise decreases and the colloidal cell approaches the inverted Janus configuration. A further increase of N induces a transition of the Janus configuration into the bubble-crescent configuration by a spontaneous breaking of the mirror symmetry. Interestingly, the symmetry breaking is not persistent. As time progresses, the colloidal cell switches reversibly between a clockwise bubble and a counter-clockwise bubble (Fig. 2.5b). We can explain the symmetry breaking as a kinetic effect. A colloidal cell can gain net rotation due to an imbalance in the number of interior spinners of each type touching the boundary of the same type. This behavior is self-reinforcing: once more clockwise-driven interior spinners are in contact with clockwise-driven boundary spinners, the cell boundary will begin to rotate clockwise, which brings it into contact with more clockwise-driven interior spinners. In small cells the boundary is able to switch rapidly between the two bubble-crescent configurations due to their size. As the cell size increases, however, fluctuations become less capable of inverting the bubble-crescent configuration and one species persists as the bubble. In the limit of infinite cell size, we expect the symmetry breaking to become irreversible. When the boundary spinners are pinned in place, the colloidal cell does not form the bubble-crescent pattern but remains in the inverted Janus configuration, as shown in SI Movie 5.

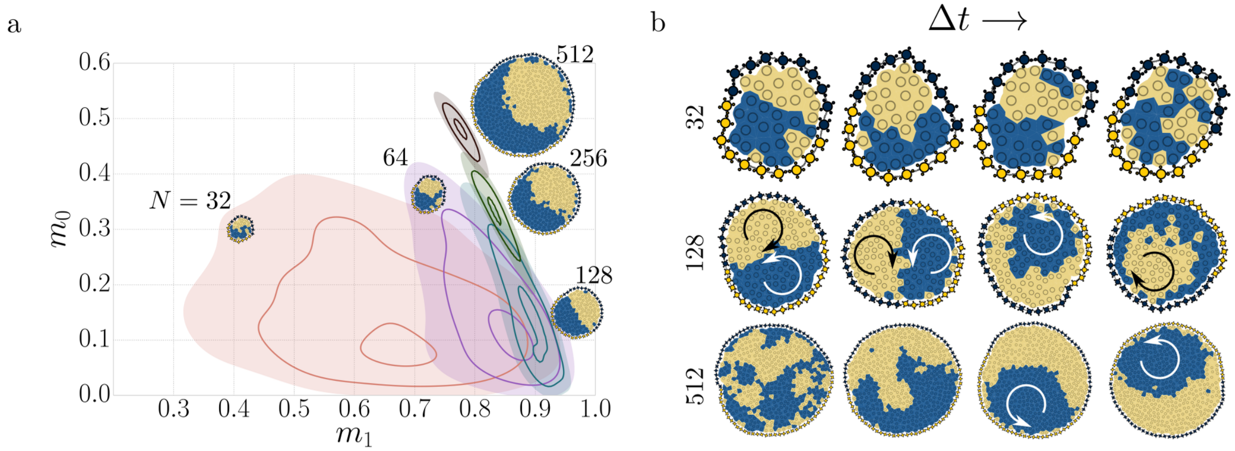


Figure 2.5: **Order parameter histograms of common intracellular compartmentalizations.** We use the cellular order parameters m_0 and m_1 to quantify core-shell and Janus behavior in simulations of the microscopic model. The order parameters are normalized such that $m_0 = 1$ and $m_1 = 1$ for the perfectly ordered configurations. (a) The interior composition and the boundary composition are kept even. As the size of the cell grows from 32 (salmon), 64 (violet), 128 (teal), 256 (green), and 512 (brown) interior spinners, the cell configuration transitions from inverted Janus to bubble-crescent. For very small cells, thermal fluctuations smear out the histogram significantly. (b) Transient behavior in cells of 32 (top row), 128 (middle row), and 512 (lower row) interior spinners. Medium-sized cells exhibit periodic migrations where spinners of one type travel along the segment of the same type in the boundary. For large cells, this migration becomes self-reinforcing due to the boundary rotation it induces, causing a spontaneous symmetry-breaking.

The continuum model also captures the qualitative dependence on the cell size observed in the microscopic model. For a 50:50 mixture of active spinners confined within a 50:50 active boundary, we observe a destabilization of the inverted Janus configuration and the concomitant formation of the bubble-crescent configuration upon increasing the size of the cell from $R = 10$ to $R = 20$ (Fig. 2.6) in agreement with the result of the microscopic model. These results can be seen in SI Movie 6 and SI Movie 7.

2.1.2 Discussion and Conclusion

We have introduced the active colloidal cell as a simple model for the study of an active matter system under confinement. The confinement is itself active and soft, allowing cells to be flexible in geometry. The main effect of the active boundary is the stirring of flow in the interior of the cell, which is observed in both a microscopic model without hydrodynamics and a continuum model. In both models, confinement results in a competition between complete phase separation with a straight interface, favored by spinners in the bulk [24], and deviating behavior imposed by the boundary conditions. There is an optimal size of the colloidal cell (in our model around $N = 128$), which is large enough such that strong thermal fluctuations are suppressed, but not too large for the system to still be influenced by the activity of the boundary. A colloidal cell of this size conjures up the image of a soft, miniature robot that can change its shape, vary its internal patterning with two species of spinners, and even exhibit a bistable equilibrium that switches back and forth.

Although we have taken just an early first theoretical step on two simple model systems, it is tempting to consider the possibility of exploiting the novel behaviors we report in the development of colloidal machines—integrated systems of colloids able to carry out functions. Recent experimental reports of functional bionic nanoparticle assemblies [82] and colloidal actuators [83] demonstrate prototypical functions of converting light into energy and converting energy into mechanical work, respectively—both functions critical for a colloidal machine. The behaviors of predictable compartmentalization, shape control, and switchability demonstrated in our model system of colloidal spinners provide additional, machine-like functionality. We expect such machines could be made from anisotropic colloids exhibiting e.g. catalytically propelled or magnetic field induced motion [84].

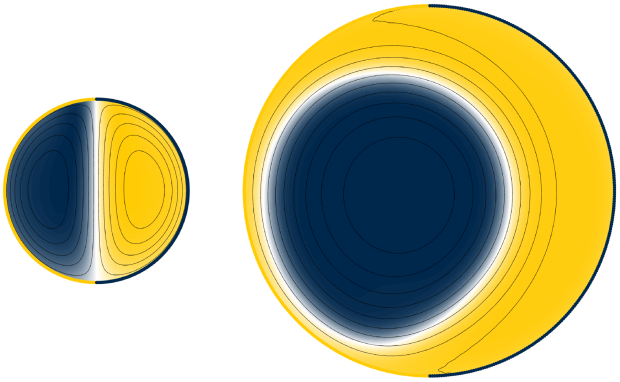


Figure 2.6: **Active Interior, Active Boundary.** Snapshots of the steady-state composition φ for two different cell sizes $R = 10$ and $R = 20$. The cells contain 50:50 mixtures of clockwise- and counterclockwise-rotating spinners within an active boundary comprising two equally sized, counter-rotating domains. Increasing the size of the cell results in the destabilization of the inverted Janus configuration and the formation of the bubble-crescent configuration. Here, the strength of active rotation is $\tau' = 8$; the strength of frictional damping is $\gamma' = 0.1$.

2.1.3 Microscopic Model

In the microscopic model, spinners are rigid bodies consisting of four peripheral disks of radius σ symmetrically arranged about a central disk of radius 3σ (Fig. 2.1a). The system is governed by the Langevin equation for translation,

$$m\frac{\partial \vec{v}_i}{\partial t} = \vec{F}_i - \gamma \vec{v}_i + \vec{F}_i^R, \tag{2.2}$$

where m is the mass and $\vec{v}_i$ is the translational velocity of each of the disks that comprise the i-th spinner. If the spinner is active then its rigid body is driven by an external driving torque $\tau_i = \pm\tau$ of constant magnitude, with positive sign for counterclockwise rotation (‘A’, blue) and negative sign for clockwise rotation (‘B’, yellow). In a real system, this torque could be due to the four peripheral disks being self-propelled particles, oriented symmetrically to impose a net torque but no net force on the spinner as a whole, much like the particles described in [71]. The torque on the particles would then be balanced by a torque on the stationary substrate. Spinners are hard particles that interact via a repulsive contact potential, resulting in internal forces $\vec{F}_i$. Translational and rotational kinetic energy is dissipated through the translational drag force $-\gamma\vec{v}_i$ applied to each constituent disk of a spinner. Noise is included via Gaussian random forces $\vec{F}_i^R = \sqrt{2\gamma k_B T}R(t)$ that model a heat bath at temperature T. Here $R(t)$ are normalized zero-mean white-noise Gaussian processes, which ensure thermodynamic equilibrium in the absence of the externally applied torques. Note that in contrast to earlier work [24] we apply drag and random forces to each constituent disk of the spinner separately, which means we do not have to specify a separate Langevin equation for rotation. Because the random forces are not applied pairwise, the thermostat is non-momentum-conserving. Therefore, our microscopic model explicitly does not include hydrodynamics.

The boundary is modeled in two steps. First, we connect the ends of a linear chain of disks with radius $\sigma/2$ that interact with their adjacent neighbors via a finitely extensible nonlinear elastic (FENE) potential $U(r) = -\frac{1}{2}kr_0^2 \ln\left(1 - (r/r_0)^2\right)$ with spring constant k, divergence length r_0, and separation distance r. Second, we rigidly attach a spinner to every tenth boundary disk, so that the boundary spinners can rotate freely without colliding if the boundary is sufficiently stretched. We can choose to make the boundary spinners active by driving them rotationally and vary the patterning of the boundary by constructing it from segments of equal driving torque. To describe the patterning we introduce the boundary activity function $f(\theta) \in [-1, 1]$, where $\theta \in [0, 2\pi)$ is an angle that describes the position at the boundary (if formed into a circle). $f = 1$ indicates a counterclockwise driving torque while $f = -1$ indicates a clockwise driving torque. Simple examples are the uniform boundary $f(\theta) = 1$ (all boundary spinners are driven counterclockwise), the Janus boundary $f(\theta) = 1 - 2H(\theta - \pi)$ (half counterclockwise and half clockwise), and the n-fold alternating boundary

$$f(\theta) = 1 + 2\sum_{j=1}^{2n-1}(-1)^j H(\theta - j\pi/n), \tag{2.3}$$

where H is the Heaviside step function.

Langevin Dynamics simulations are performed on graphic processing units (GPUs) with HOOMD-blue [85, 86, 87] for colloidal cells with between 16 and 512 spinners in the interior. The contact between spinners is a Weeks-Chandler-Andersen potential [88] with parameter ϵ shifted to the surface of each disk such that its range is a small fraction of the disk diameter, thereby approximating hard shapes. Throughout the paper we report results for $\gamma = 1\sqrt{m\epsilon}/\sigma$, $r_0 = 1.5\sigma$, $k = 20\epsilon/\sigma^2$, and thermal noise $k_BT = 1\epsilon$ using σ as the length unit and ϵ as the energy unit. These parameter choices correspond to the overdamped, diffusive limit. Active systems are often described by overdamped equations of motion where inertia is neglected[40, 46, 48]. Although inertia is incorporated in our model, we confirmed that it is not crucial for any of the observed behavior. Additional studies of the role of convection in bulk systems of spinners is presented in Ref. [89]

2.1.4 Continuum Model

In the continuum model, the spinner dynamics is described by coupling the Cahn-Hilliard phase field equation to a Navier-Stokes equation with an active term representing the rotational driving torque. Previously, a continuum model was used to describe separation of translationally-driven particles into high- and low-density phases, much like vapor-liquid or vapor-solid coexistence in single-component equilibrium systems [49, 90, 91]. Here, instead, we model separation into clockwise- and counterclockwise-driven domains, analogous to equilibrium phase separation of a binary mixture of immiscible fluids as reported in [24] for our microscopic model. Boundary effects are taken into account via the choice of boundary conditions for the equations. Depending on the presence or absence of activity at the boundary and in the interior we distinguish various cases. Here we present the governing equations for the most general situation and refer to the Materials and Methods section for details and derivations.

To describe the binary fluid of actively rotating spinners, we start with the Cahn-Hilliard equation for the fraction of clockwise- or counterclockwise-driven spinners within a fluid volume, φ. The Cahn-Hilliard equation can be written in non-dimensional form as

$$\frac{d\varphi}{dt} = \nabla^2(-\varphi + \varphi^3 - \nabla^2\varphi). \tag{2.4}$$

The two-dimensional fluid is modeled as a generalization of an incompressible, Newtonian fluid governed by the (non-dimensional) Navier-Stokes equations [92, 93],

$$\mathrm{Re}\frac{d\vec{v}}{dt} = -\nabla p + \nabla^2\vec{v} - \gamma'\vec{v} + \mathrm{Ca}^{-1}\mu\nabla\varphi + \tau'\nabla \times (\varphi\vec{e}_z), \tag{2.5}$$

$$0 = \nabla \cdot \vec{v}, \tag{2.6}$$

where $\vec{v}$ is the fluid velocity, Re is a Reynolds number, Ca is a capillary number, γ' is a translational drag coefficient present in the microscopic model, and τ' measures the strength of the rotational driving torque. The form of the rotating driving implies that the torque density is proportional to the local composition φ [94]. For simplicity, we neglect the effects of fluid inertia as well as that of capillary-like forces acting normal to the fluid-fluid interface (*i.e.*, $\mathrm{Re} \to 0$ and $\mathrm{Ca}^{-1} \to 0$). These contributions are expected to be unimportant for the relatively small fluid domains described here. With these simplifications, convective flows are driven only by forces due to active rotation directed parallel to the interface separating the counter-rotating domains, $\tau'\nabla \times (\varphi\vec{e}_z)$.

To solve for the two-dimensional velocity field, it is convenient to introduce the stream function ψ where $\vec{v} = \nabla \times (\psi \vec{e}_z)$. As shown in the SI, the momentum equation (Eq. 2.5) can then be recast in terms of the stream function,

$$0 = \nabla^4 \psi - \gamma' \nabla^2 \psi + \tau' \nabla^2 \varphi. \tag{2.7}$$

Importantly, the dimensionless coefficients τ' and γ' characterizing the strength of active rotation and frictional drag in the continuum model are directly analogous (in an order-of-magnitude sense) to the parameters τ and γ in the microscopic model. Together with an impermeable boundary with spatially varying stress, Eqs. 2.4 and 2.7 govern the dynamics of the composition φ and stream function ψ of the fluid. We integrate these continuum equations using COMSOL Multiphysics version 4.4.

The fluid is confined within an impermeable passive or active boundary, entering as boundary conditions for the continuum equations. The driving torques applied to the boundary spinners are equal to those driving the rotation of the interior spinners. This scenario can be approximated by the two boundary conditions $\psi = 0$ and $\nabla \times \vec{v} = -\tau'(f(\theta) - \varphi)$, where $f(\theta) \in [-1, 1]$ is the boundary activity. Furthermore, in the microscopic model, the active boundary is not fixed in place but is free to rotate relative to the stationary surroundings (*e.g.*, an underlying substrate) To describe this effect in the continuum model, we fix the shape of the boundary to a circle of radius R but allow for its rotational motion with an angular velocity Ω. In the low Reynolds number limit, the use of a rotating reference frame does not affect the equations of motion with the exception of the frictional damping term in the Navier-Stokes equation. Finally, the activity-induced flows create non-uniform stresses normal to the boundary that result in its deformation. To model shape changes of the active colloidal cell, we assume that the normal component of the stress at the boundary is balanced by a surface tension-like force, which is proportional to the local curvature of the interface.

Governing Equations

To describe the binary fluid of actively rotating spinners, we start with the convective Cahn-Hilliard equation for the compositional order parameter φ,

$$\frac{\partial \varphi}{\partial t} + \vec{v} \cdot \nabla \varphi = M \nabla^2 \mu, \tag{2.8}$$

where $\vec{v}$ is the fluid velocity, M is a mobility coefficient, and μ is the chemical potential. For simplicity, we assume the chemical potential is of the form

$$\mu = -r\varphi + \lambda\varphi^3 - K\nabla^2\varphi, \tag{2.9}$$

where r, λ, and K are positive coefficients. Physically, these coefficients determine the thickness, $(K/r)^{1/2}$, of the interface separating two equilibrium phases with composition $\varphi_{\text{eq}} = \pm(r/\lambda)^{1/2}$.

We further assume that the fluid is incompressible, Newtonian, and symmetric such that the bulk properties of the two phases are equal—in particular, the density ρ and viscosity η. The two components of the fluid are driven to rotate in opposite directions by a torque density $a\varphi\vec{e}_z$, which is proportional to the order parameter φ and to a constant a that describes the magnitude of rotation. Under these conditions, conservation of mass and momentum imply that

$$0 = \nabla \cdot \vec{v}, \tag{2.10}$$

$$\rho\frac{d\vec{v}}{dt} = -\nabla p + \eta\nabla^2\vec{v} - b\vec{v} + \mu\nabla\varphi + a\nabla \times (\varphi\vec{e}_z), \tag{2.11}$$

where the additional terms in Eq. 2.11 describe (i) frictional drag against the stationary surroundings, $-b\vec{v}$, (ii) capillary-like forces acting normal to the fluid-fluid interface, $\mu\nabla\varphi$, and (iii) forces due to active rotation directed parallel to the interface, $a\nabla \times (\varphi\vec{e}_z)$.

Non-Dimensionalization

At this point, it is convenient to non-dimensionalize the governing equations using characteristic scales for the interfacial thickness $(K/r)^{1/2}$, the time of demixing K/Mr^2, and the equilibrium composition $(r/\lambda)^{1/2}$. In these dimensionless units, Eqs. 2.8 and 2.11 become

$$\frac{d\varphi}{dt} = \nabla^2\left(-\varphi + \varphi^3 - \nabla^2\varphi\right), \tag{2.12}$$

$$\text{Re}\frac{d\vec{v}}{dt} = -\nabla p + \nabla^2\vec{v} - \gamma'\vec{v} + \text{Ca}^{-1}\mu\nabla\varphi + \tau'\nabla \times (\varphi\vec{e}_z), \tag{2.13}$$

where $\text{Re} = \rho Mr/\eta$ is a Reynolds number, $\text{Ca} = M\lambda\eta/K$ is a capillary number, and the dimensionless coefficients τ' and γ' characterize the strength of active rotation and frictional drag, respectively. Here, we focus exclusively on the low Reynolds number limit ($\text{Re} \to 0$) and neglect capillary forces ($\text{Ca}^{-1} \to 0$) such that fluid flow is driven solely by the active rotation of the particles. Systems with passive fluid interiors are described by setting $\tau' = 0$.

To solve for the two-dimensional velocity field, it is convenient to introduce the stream function ψ where $\vec{v} = \nabla \times (\psi \vec{e}_z)$ such that Eq. 2.13 becomes

$$0 = \nabla^4 \psi - \gamma' \nabla^2 \psi + \tau' \nabla^2 \varphi. \tag{2.14}$$

For such two-dimensional flows, the fluid vorticity (in the z-direction), $w = \nabla \times \vec{v}$, can be related to the stream function as $w = -\nabla^2 \psi$. Together, Eqs. 2.12 and 2.14 govern the dynamics of the composition φ and the flow field.

Passive Boundary

In the microscopic model, a passive boundary refers to that formed by passive spinners that are otherwise free to rotate and translate subject to the constraints of their connectivity. In the continuum model, the passive boundary is described by a circle of radius R with no flow normal to the boundary and no stress tangent to the boundary

$$\vec{n} \cdot \vec{v} = 0, \tag{2.15}$$

$$\vec{n} \cdot \vec{\tau} \cdot \vec{t} = 0. \tag{2.16}$$

Here, $\vec{n}$ and $\vec{t}$ are the unit vectors normal and tangent to the boundary (with the convention $\vec{n} \times \vec{t} = \vec{e}_z$), and $\vec{\tau} = \nabla \vec{v} + (\nabla \vec{v})^T$ is the viscous stress tensor. In terms of the stream function and the vorticity, these conditions imply

$$\psi = 0, \tag{2.17}$$

$$w = 0, \tag{2.18}$$

everywhere along the circular boundary.

Similarly, for systems with active interiors, we require two boundary conditions to fully specify the composition field φ. First, there is no flux normal to the boundary

$$\vec{n} \cdot \nabla \mu = 0. \tag{2.19}$$

We also require a "wetting" condition that determines the effective contact angle between the counter rotating fluid phases and the bounding surface. For the symmetric fluids described here, this effective contact angle should be $\pi/2$ such that

$$\vec{n} \cdot \nabla \varphi = 0. \tag{2.20}$$

Subject to these boundary conditions, Eqs. 2.12 and 2.14 are solved numerically using the commercial finite element solver COMSOL.

Active Boundary

In the microscopic model, an active boundary refers to that formed by active spinners which are driven to rotate in either direction with a constant torque. These boundary spinners can induce stresses tangent to the boundary that drive flows of the interior fluid, which may be active or passive. In the continuum model, the active boundary is described by a circle of radius R with no flow normal to the boundary and a tangential stress related to the local composition φ and the boundary activity $f(\theta)$ as

$$\psi = 0, \tag{2.21}$$

$$w = -\tau'(f(\theta) - \varphi), \tag{2.22}$$

where $f(\theta) \in [-1, 1]$ with $f = 1$ for counterclockwise-rotating boundary spinners and $f = -1$ for clockwise boundary spinners. Here, the driving torques applied to the boundary spinners are assumed equal to those driving the rotation of the interior spinners. Note that the vorticity w at the boundary (here equal to the tangent stress) depends on the difference between the spinner composition in the fluid interior and that at the boundary. An interface between like rotating particles results in no net stress (*i.e.*, when $f = \varphi$); the largest stresses occur at interfaces between counter-rotating particles (*i.e.*, when $f = -\varphi$). Systems with passive fluid interiors can be described by setting $\varphi = 0$ in Eq. 2.22.

To model the boundary activity used in the microscopic model, the function $f(\theta)$ was chosen as

$$f(\theta) = \tanh\left(\frac{\cos(n\theta)}{n\delta}\right), \tag{2.23}$$

where n determines the number of domains on the boundary, and δ is a length characterizing the width of the transition from one domain to the next (here, $\delta = 0.01$). The resulting flows for passive interiors with $n = 1$ to $n = 5$ are shown in Fig. 2.2b.

Finally, it is important to note that the active boundary is not fixed in place but is free to rotate relative to the stationary surroundings (*e.g.*, an underlying substrate). To describe this effect in the continuum model, we fix the shape of the boundary to a circle of radius R but allow for its rotational motion with an angular velocity Ω. We adopt a rotating frame of reference which is fixed to the boundary and participates in its motion. In the low Reynolds number limit, the use of a rotating reference frame does not affect the equations of motion with the exception of the frictional damping term in the Navier-Stokes equation, which describes the resistance to motion relative to the stationary surroundings. In the rotating reference frame, the stream function Eq. 2.14 becomes

$$0 = \nabla^4 \psi - \gamma' (\nabla^2 \psi + 2\Omega) + \tau' \nabla^2 \varphi. \tag{2.24}$$

It is further assumed that the net torque T acting on the surroundings is identically zero at all times (otherwise, the system would accelerate or decelerate its rotational motion). This condition implies that

$$\vec{T} = \gamma' \int_S \vec{r} \times (\vec{v} - \vec{\Omega} \times \vec{r}) dS = 0, \tag{2.25}$$

where the integral is carried out over the entire fluid domain S. For a circular domain of radius R, the angular velocity Ω is therefore

$$\Omega = \frac{2}{\pi R^4} \int_S (\vec{r} \times \vec{v}) dS. \tag{2.26}$$

This integral constraint must be solved at each time step to describe the rotation of the cell. Note that such complications are necessary only in describing the most general case of actively rotating fluids confined by an active boundary.

Shape Change

The activity-induced flows create non-uniform stresses normal to the flexible boundary that can result in its deformation. To describe these deformations, we assume that the normal component of the stress at the boundary is balanced by a surface tension-like force, which is proportional to the local curvature $\mathcal{H}$ of the interface

$$- p + \tau_{nn} = \kappa \mathcal{H}, \tag{2.27}$$

where $\tau_{nn} = \vec{n} \cdot \vec{\tau} \cdot \vec{n}$ is the normal component of the viscous stress at the boundary, and κ is the surface tension of the boundary (in units of $\eta M r^{3/2}/K^{1/2}$). Provided that forces due to surface tension are large compared to those due to boundary activity (*i.e.*, $\kappa \gg R\tau'$), deformations in the shape of the boundary will be small. Under these conditions, we can use the stress computed for the circular boundary to approximate changes in cell shape, $r(\theta)$, which is computed as

$$\mathcal{H}(\theta) = \frac{r^2 + 2r_\theta^2 - r r_{\theta\theta}}{(r^2 + r_\theta^2)^{3/2}}, \tag{2.28}$$

where $\mathcal{H}(\theta)$ is the local curvature specified by Eq. 2.27, and the subscripts denote differentiation with respect to θ.

2.1.5 Connecting the Continuum and Microscopic Models

In the microscopic model, the key parameters such as the driving torque τ, the frictional drag γ, and the elasticity of the boundary κ are expressed using natural microscopic scales for length σ, time $\sigma(m/\epsilon)^{1/2}$, and energy $\epsilon = k_B T$. Here, we connect these characteristic scales to those used in non-dimensionalizing the continuum model. The characteristic length in the continuum model is taken to be the thickness of the interface separating the two counter-rotating phases, $(K/r)^{1/2}$; this length should be comparable to the size of the particles such that $(K/r)^{1/2} \sim \sigma$. The characteristic time used in the continuum model is that of unmixing, K/Mr^2, which should be comparable to the time required for a particle to diffuse one particle diameter—that is, $K/Mr^2 \sim \sigma(m/\epsilon)^{1/2}$. Finally, the characteristic energy scale used in the continuum model is taken to be ηMr. Approximating the fluid viscosity as $\eta \sim (m\epsilon/\sigma^2)^{1/2}$ [95] and applying the two relations above, this energy scale becomes, $\eta Mr \sim \epsilon$.

To summarize, the characteristic scales used in the continuum model are—to within an order of magnitude—the same as those used in the microscopic model. Consequently, the parameter values used in each of the models—though not exactly equivalent—should be directly comparable to one another.This is reflected by the use of common notation, *i.e.* τ and τ' for the driving torque in the microscopic model and the continuum model, respectively, as well as γ and γ' for the frictional drag.

We summarize the characteristic scales used in the respective models:

	Microscopic	Continuum
length	σ	$(K/r)^{1/2}$
time	$\sigma(m/\epsilon)^{1/2}$	K/Mr^2
energy	ϵ	ηMr

Table 2.1: Characteristic unit scales for the two types of models employed to study active coarsening behavior.

2.2 Coarsening Dynamics of Binary Liquids with Active Rotation

Note: this section was originally published in Soft Matter in 2015[89] and is reproduced with minor modifications below. M. S. and Sharon C Glotzer designed and implemented the microscopic model, while Syeda Sabrina and Kyle J Bishop designed and implemented the continuum model.

The distributed conversion of energy into motion within ensembles of many self-propelled units can lead to complex collective behaviors operating outside the constraints of thermodynamic equilibrium [31]. Well-studied examples of such active matter include migrating organisms[30], the cell cytoskeleton[96, 97], driven granular materials [98, 99], and self-phoretic colloids[38, 100, 101, 102, 39, 51]. In many of these systems, the activity of the individual units can lead to phase separation and coexistence even in the absence of attractive interactions. This behavior is clearly illustrated by simple physical models such as that of active Brownian particles (ABPs), in which hard spheres move at a constant speed in a direction subject to rotational Brownian motion [38]. ABPs are known to phase separate in 2D [46, 48, 49] and 3D [91, 103] due to a kinetic trapping mechanism, whereby particles incident on the surface of a condensed phase are "trapped" by other incoming particles[48]. More generally, activity-induced phase separation of self-propelled "particles" is expected whenever the average particle velocity decreases sufficiently rapidly with particle density[35, 55]. This basic mechanism is believed to underlie phase separation in such disparate systems as mussel beds [104], bacterial colonies [80], and active colloids [51].

Activity-induced phase separation has also been observed in systems of rotating particles, in which otherwise identical, gear-like disks are driven to rotate in opposite directions[24]. Active rotation induces effective interactions between the particles that can cause their segregation into counter-rotating, fluid and crystalline domains [24]. Similar behaviors have been observed in simulations of spherical particles rotating within a fluid, where viscous shear forces couple the rotational motions of neighboring particles [105, 76, 106]. Experimentally, there exists several promising mechanisms by which to rotate colloidal components using magnetic fields [72, 107], electrokinetic flows [108], self-phoretic motions [109, 110, 71], or circularly polarized light [111]. Recent studies have shown that single component systems of magnetically rotated colloids[77] or self-rotating bacteria[78] can segregate into high and low density phases driven primarily by attractive dipolar or hydrodynamic interactions, respectively.

Importantly, the unmixing of actively rotating particles is accompanied by steady convective flows that originate along the interface separating the counter-rotating domains [24]. Under appropriate conditions, these activity-driven flows are expected to influence the coarsening of actively rotating fluids and may lead to new types of dynamically-organized structures. By contrast, the coarsening of linearly self-propelled particles such as ABPs exhibit strong similarities to that of passive liquids[91], in which self-similar domains of size R grow in time as $R \propto t^{1/3}$. Thus, while different forms of microscopic activity can lead to similar mesoscale behaviors (e.g., phase separation), these differences may contribute to qualitatively different dynamical behaviors at the macroscale (e.g., coarsening dynamics).

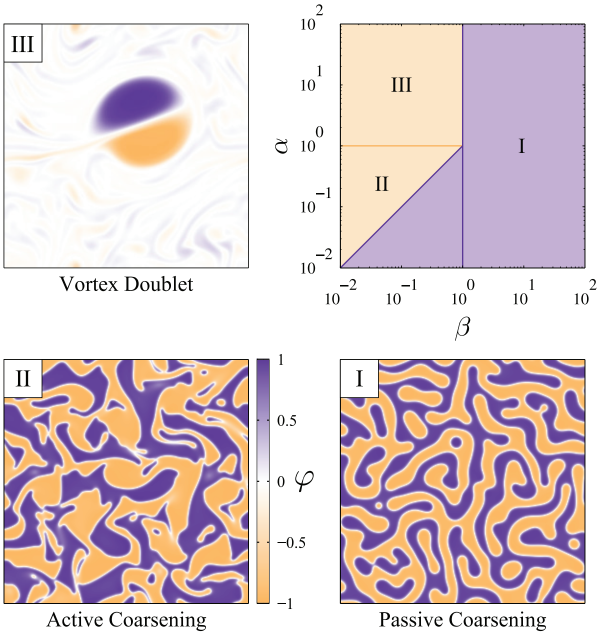


Figure 2.7: Phase diagram in the $\alpha\beta$-plane illustrating the different dynamical regimes (top right). Here, α measures the strength of active rotation, while β measures that of frictional damping against the stationary surroundings (see text for details). One representative snapshot of the compositional order parameter φ is shown for each regime.

Here, we investigate the dynamics of liquid-liquid unmixing in a binary fluid subject to active rotation using a phenomenological, phase-field model based on the convective Cahn-Hilliard equation[92, 93]. In this description, the rotational actuation of the two components in opposite directions introduces additional forces within the Navier-Stokes equations governing fluid motion. Depending on the strength of active rotation and that of frictional interactions with the stationary surroundings, we observe three distinct dynamical regimes as summarized in Figure 2.7. For strong frictional damping, coarsening of the counter-rotating domains is identical to that of a passive fluid without active rotation (in agreement with previous particle-based simulations[24]). By contrast, when frictional forces are relaxed, the system exhibits new dynamical behaviors such as "active coarsening" driven by convective flows induced by the rotation of the particles as well as the emergence of self-propelled "vortex doublets". We use numerical simulation along with scaling arguments to characterize the system's dynamics within each flow regime. Additionally, we show that many of the qualitative behaviors identified by the continuum model can also be found in discrete, particle-based simulations of actively rotating liquids. These results highlight opportunities for achieving complex dissipative structures by directing collective excitations within active matter.

2.2.1 Model Dynamics

The coarsening dynamics of binary liquids with active rotation is studied using two models: a continuum, hydrodynamic model and a microscopic, particle-based model. The former is an extension of previous phase field models[92, 93] for two phase flow that accounts for the active rotation of the fluid components. The microscopic model[24, 29] describes the Langevin dynamics of hard, gear-like particles, which are driven to rotate in opposite directions. Here, we limit our investigation to two-dimensional systems; however, both models can be readily extended to three-dimensions.

Continuum Model

In the continuum approach, we consider a binary fluid in which the local composition is characterized by an order parameter φ governed by the convective Cahn-Hilliard equation[92, 93],

$$\frac{\partial \varphi}{\partial t} + \nabla \cdot (\varphi \mathbf{v}) = M \nabla^2 \mu, \tag{2.29}$$

where $\mathbf{v}$ is the fluid velocity, M is a mobility coefficient, and μ is the chemical potential. Physically, the composition $\varphi(x, y, t)$ describes the relative amount of counter-clockwise-rotating components ($\varphi > 0$) and clockwise-rotating components ($\varphi < 0$) at a given point in space and time. For simplicity, we assume the chemical potential is of the form

$$\mu = -r\varphi + \lambda\varphi^3 - K\nabla^2\varphi, \tag{2.30}$$

where r, λ, and K are positive coefficients. These coefficients determine the thickness $(K/r)^{1/2}$ of the interface separating two phases of composition $\varphi = \pm(r/\lambda)^{1/2}$. We emphasize that this simple model does not attempt to explain the *origins* of phase separation driven by active rotation. Instead, we assume phase separation *a priori* and focus on the role of activity on the dynamics with which these rotating phases coarsen in time.

To describe the activity-driven flows, we further assume that the fluid is incompressible, Newtonian, and "symmetric" such that the bulk properties of the two phases are equal—in particular, the density ρ and viscosity η. Under these conditions, conservation of mass and momentum imply that

$$\nabla \cdot \mathbf{v} = 0, \tag{2.31}$$

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla p + \eta \nabla^2 \mathbf{v} + \mu \nabla \varphi + \nabla \times (\varphi \boldsymbol{\tau}) - b\mathbf{v}. \tag{2.32}$$

In addition to the usual pressure and viscous forces present in the Navier-Stokes equation, equation (2.32) incorporates forces due to (i) capillarity[92], (ii) active rotation, and (iii) frictional drag, respectively. In particular, we consider that the two components of the fluid are driven to rotate in opposite directions by a torque density $\varphi\boldsymbol{\tau}$, which is proportional to the order parameter φ and to a constant vector $\boldsymbol{\tau}$ that describes the magnitude and direction of rotation. These local torques combined with spatial variations in the composition give rise to forces that act parallel to the interface separating the counter-rotating phases [94]. In our 2D simulations, the fluid moves in the xy-plane with active rotation in the z-direction ($\boldsymbol{\tau} = \tau \mathbf{e}_z$). Physically, the system can be thought to represent an ensemble of active particles moving and rotating above a planar substrate as is often the case in experimental realizations of active matter in 2D. To account for interactions between the particles and the underlying substrate, we include a frictional force in equation (2.32) characterized by a constant friction coefficient b .

At this point, it is convenient to non-dimensionalize the governing equations using characteristic scales for the interfacial thickness $(K/r)^{1/2}$, the time of unmixing K/Mr^2, the equilibrium composition $(r/\lambda)^{1/2}$, and the chemical potential $(r^3/\lambda)^{1/2}$. In dimensionless units, equations (2.29) and (2.32) reduce to

$$\frac{\partial \varphi}{\partial t} + \mathbf{v} \cdot \nabla \varphi = \nabla^2 \left(-\varphi + \varphi^3 - \nabla^2 \varphi \right), \tag{2.33}$$

$$Re \frac{d\mathbf{v}}{dt} = -\nabla p + \nabla^2 \mathbf{v} + Ca^{-1} \mu \nabla \varphi + \alpha \nabla \times (\varphi \mathbf{e}_z) - \beta \mathbf{v}, \tag{2.34}$$

where $Re = \rho M r/\eta$ is a is a Reynolds number, $Ca = M\lambda\eta/K$ is a capillary number, and the dimensionless coefficients α and β characterize the strength of active rotation and frictional drag, respectively. In this paper, we focus exclusively on the low Reynolds number limit ($Re \to 0$) and neglect capillary forces ($Ca^{-1} \to 0$) such that fluid flow is driven solely by the active rotation of the fluid components.

The governing equations (2.33) and (2.34) are solved numerically on a square domain ($L\times L$) with periodic boundaries using a semi-implicit Fourier spectral method[112] for different values of the parameters α and β (see ESI for details).† Initially, the system is prepared in a homogeneous state, in which the composition at each point is assigned a random value drawn uniformly from the interval $[-0.1, 0.1]$. Depending on the strength of active rotation α and frictional drag β, this model exhibits a variety of different coarsening mechanisms ranging from passive, diffusive coarsening to active coarsening and the emergence of “vortex doublets” (Fig. 2.7).

Microscopic Model

To confirm the generality of active coarsening in rotating fluids, we study an analogous particle-based system whereby collections of hard, gear-shaped “spinners” are driven to rotate in opposite directions by an applied torque [24, 29]. Each spinner contains five circular disks of radius σ fixed symmetrically about a central disk of radius 3σ. The dynamics of these composite particles is governed by the following Langevin equation for the velocity of the ith disk

$$m\frac{d\mathbf{v}_i}{dt} = \mathbf{F}_i - \gamma\mathbf{v}_i + \mathbf{F}_i^R, \tag{2.35}$$

where m is the mass of each disk, $\mathbf{F}_i$ and $\mathbf{F}_i^R$ represent deterministic and stochastic forces, and γ is a frictional drag coefficient. The deterministic forces $\mathbf{F}_i$ contain both active and passive contributions. First, all spinners are driven to rotate by a constant torque $\tau_i = \pm\tau$ with equal numbers rotating in each direction. Additionally, spinners interact both through a repulsive contact potential and through a short ranged attraction between like-rotating spinners. The latter is included to ensure phase separation even in the absence of active rotation by analogy to the continuum model, although it was not considered in previous works[24]. The stochastic force, $\mathbf{F}_i^R = \sqrt{2\gamma k_B T}\mathbf{X}(t)$, ensures that the system approaches thermal equilibrium at temperature T in the absence of active rotation. Langevin dynamics simulations were performed on graphic processing units (GPUs) with the HOOMD-blue[85, 86] software package for $16,384$ spinners in the system (see ESI for details).†

Although a rigorous connection between the microscopic and continuum models is lacking and outside the scope of this paper, we use order-of-magnitude reasoning to identify dimensionless parameters α' and β' in the microscopic model that are analogous to α and β in the continuum model (see ESI for details).† Specifically, $\alpha' = \tau/k_BT$ measures the strength of active rotation relative to the thermal energy, whereas $\beta' = \gamma\sigma/\sqrt{mk_BT}$ measures the strength of frictional damping. The connection between these parameters and those of the continuum model are discussed further in the Supporting Information.† Below, all results from the microscopic model are presented in dimensionless form using characteristic scales σ, $\sigma(m/k_BT)^{1/2}$, and k_BT for length, time, and energy, respectively.

2.2.2 Results and Discussion

We first use the continuum model to map out three qualitatively distinct parameter regimes as summarized in Figure 2.7. We discuss each regime in turn and provide detailed scaling arguments to explain the behaviors observed in the simulations. Building on insights from the continuum model, we reproduce many—though not all—of the qualitative coarsening behaviors using the microscopic model.

Strong Damping ($\beta \gg 1$)

In the presence of strong frictional damping ($\beta \gg 1$), the coarsening dynamics of the active fluid is independent of the strength of active rotation (i.e., of α) and identical to that of a passive fluid, for which $\alpha = 0$ (Fig. 2.8). This "passive coarsening" regime has been studied extensively using the Cahn-Hilliard equation in the absence of fluid flow [113]. At short times ($t \ll 1$), the initially homogeneous fluid undergoes an instability characterized by a wavenumber $k = 2^{-1/2}$, which grows in time at a rate $1/4$ until the formation of bulk domains with composition $\varphi \approx \pm 1$ separated by an interfacial region of unit thickness. At longer times ($t \gg 1$), these domains grow in size as $R \sim t^{1/3}$ due to small composition gradients ($\Delta\varphi \sim R^{-1}$) which drive diffusive fluxes ($j \sim \Delta\varphi/R$) that act to grow the domains ($dR/dt \sim j$) and reduce the curvature of the interface (Fig. 2.8a). Here, the domain size R is defined as the first root of the radial pair correlation function[112] unless otherwise stated (see ESI for details).†

In this regime, the active rotation of the fluid drives convective flows along the the interface separating the counter-rotating domains (Fig. 2.8b). The dominant terms of equation (2.34) are $\alpha\nabla \times (\varphi\mathbf{e}_z) \approx \beta\mathbf{v}$, such that forces due to active rotation are everywhere balanced

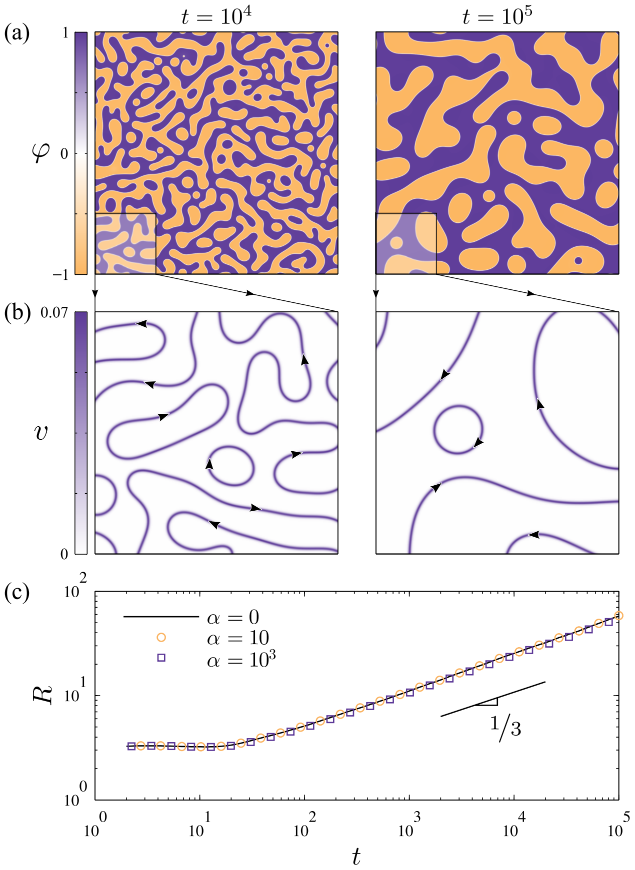


Figure 2.8: **Passive coarsening, $\beta \gg 1$.** (a) Compositional order parameter $\varphi(x, y)$ at times $t = 10^4$ and 10^5 for parameters $\alpha = 10$ and $\beta = 10^2$; the size of simulation cell is $L = 1024$. (b) Velocity field $v(x, y)$ corresponding to the insets in (a). Arrows show the direction of fluid flow. (c) Domain size R as a function of time for $\beta = 10^2$ and $\alpha = 0$, 10, and 10^3; here, R is defined as the first zero of the radial pair correlation function, $g(R) = 0$ [112]. See ESI for the corresponding movie.†

by frictional drag. As a result, all flows are directed perpendicular to gradients in the order parameter, and the effects of convective transport are negligible (i.e., $\mathbf{v} \cdot \nabla\varphi \approx 0$ in equation (2.33)). Consequently, the domain size R increases as $R \sim t^{1/3}$ independent of both α and β in quantitative agreement with passive diffusive coarsening (Fig. 2.8c).

Weak Damping & Weak Rotation ($\beta \ll 1$, $\alpha \ll 1$)

For weak frictional damping ($\beta \ll 1$), flows due to active rotation are no longer confined to the interface but rather extend into the bulk domains to influence the dynamics of unmixing (Fig. 2.9). The morphology of the growing domains (Fig. 2.9a) is visibly different from that due to passive coarsening: arrays of counter-rotating vortices (Fig. 2.9b) create thin filaments that break-up and merge with the larger domains. The characteristic domain size R increases faster with time than expected by diffusive coarsening alone (Fig. 2.9c).

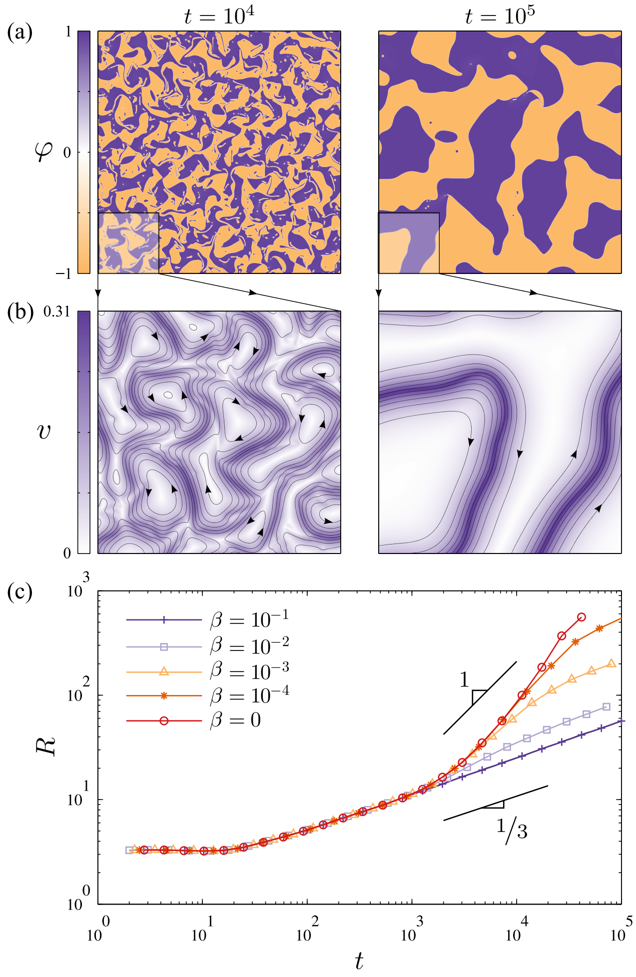


Figure 2.9: **Active coarsening, $\beta \ll 1$ and $\alpha \ll 1$.** (a) Order parameter $\varphi(x, y)$ at times $t = 10^4$ and 10^5 for parameters $\alpha = 10^{-2}$ and $\beta = 10^{-3}$; the size of simulation cell is $L = 2048$. (b) Velocity field $v(x, y)$ corresponding to the insets in (a). Arrowheads show the direction of fluid flow. (c) Domain size R as a function of time for $\alpha = 10^{-2}$ and different amounts of frictional damping β. See ESI for the corresponding movie.†

To understand these qualitative observations in more detail, consider that rotation within the bulk creates an interfacial stress of order α in a direction tangent to the interface. This active stress is balanced by the viscous stress U/ℓ, where U is a characteristic velocity, and ℓ is a length scale over which the velocity falls to zero. For small domains ($R \ll \beta^{-1/2}$), velocity gradients extend throughout the bulk such that $\ell \sim R$ and $U \sim \alpha R$. As the domains grow larger ($R \gg \beta^{-1/2}$), the velocity decays exponentially with distance from the interface over a length $\beta^{-1/2}$ due to frictional drag; the velocity approaches a constant value $U \sim \alpha\beta^{-1/2}$ (Fig. 2.9b).

Using these estimates for the fluid velocity, we introduce a Péclet number, $Pe = \ell U$, which characterizes the relative importance of convective and diffusive transport on the coarsening of the domains.[3] For small Péclet number ($Pe \ll 1$), activity-driven flows do not affect the coarsening dynamics, which is analogous to that of a passive fluid. Using the above estimates for the fluid velocity, this condition implies that small domains, $R \ll \alpha^{-1/2}$, are unaffected by active rotation. By contrast, domains that grow larger than a critical size, $R^* \sim \alpha^{-1/2}$, induce flow velocities capable of influencing the coarsening dynamics. This effect is illustrated in Fig. 2.9c which shows that the domain size $R(t)$ follows that of the passive fluid for $R < 10$ when $\alpha = 10^{-2}$. Beyond the critical size R^*, coarsening accelerates due to convection driven by the rotating fluid. In this regime, domain growth is expected to scale as $dR/dt \sim U\Delta\varphi \sim \alpha$, where $\Delta\varphi \sim R^{-1}$ is the magnitude of curvature-induced variations in composition.

Eventually, however, the rate of coarsening slows as the domain size R grows larger than the length $\ell \sim \beta^{-1/2}$, which characterizes the decay of velocity with distance from the interface. Under these conditions ($R \gg \beta^{-1/2}$), flows are increasingly confined within a thin interfacial region and no longer influence the rate-limiting process of diffusion throughout the bulk domain. As a result, the domain growth returns to the diffusive scaling, $R \propto t^{1/3}$, at long times (Fig. 2.9c).

To summarize, "active coarsening" occurs when (i) the Péclet number is large, and (ii) activity-driven flows extend throughout the bulk domains. These conditions are satisfied provided that the domain size is in the range $\alpha^{-1/2} \ll R \ll \beta^{-1/2}$. This dynamical regime is denoted by region II of the phase diagram in Fig. 2.7.

Zero Damping & Weak Rotation ($\beta \to 0$, $\alpha \ll 1$)

To better understand the "active coarsening" regime, we examine the limit of zero frictional damping, $\beta \to 0$, such that interfacial stresses due to active rotation are propagated by viscosity throughout the bulk domains—regardless of their size (Fig. 2.10). The domain structure is no longer characterized by a single length scale in contrast to the self-similar structures formed by "passive coarsening". Instead, we observe a spectrum of different length

---

[3] In dimensional units, the Péclet number takes the more familiar form of $Pe = \ell U/Mr$ where Mr is identified as the diffusivity.

scales spanning a finite range from R_{min} to R_{max}.

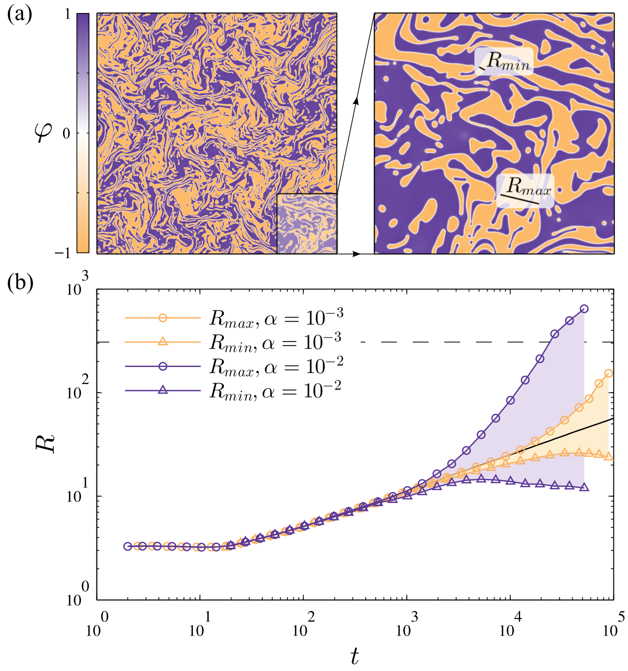


Figure 2.10: **Active coarsening,** $\beta \to 0$ **and** $\alpha \ll 1$. (a) Order parameter $\varphi(x, y)$ at time $t = 2 \times 10^4$ for $\alpha = 10^{-2}$ and no frictional damping, $\beta = 0$; the size of simulation cell is $L = 2048$. The right image shows a magnified view highlighting the two characteristic lengths, R_{min} and R_{max}, described in the text. (b) Length scales, R_{min} and R_{max}, as a function of time for $\alpha = 10^{-2}$ and $\alpha = 10^{-3}$ in the absence of frictional damping, $\beta = 0$. The dashed black curve corresponds to $0.15L$ beyond which the finite size of the simulation domain becomes important; the solid black curve shows the domain size evolution for a passive fluid. See ESI for the corresponding movie.†

The lower bound can be estimated as $R_{min} \sim A/C$, where A is the total area, and C is the length of interface separating the bulk domains. Physically, R_{min} describes the width of the filamentous structures that are repeatedly drawn from the edges of the larger rotating domains. Because these structures are shaped by convective flows, their size must be sufficiently large to achieve Péclet numbers of order unity—that is, $Pe = \alpha R_{min}^2 \sim 1$ such that $R_{min} \sim \alpha^{-1/2}$. This scaling result is supported by numerical simulations (Fig. 2.10b), which reveal that R_{min} remains roughly constant throughout the coarsening process.

The larger length scale R_{max} is evaluated like R above as the first root of the pair correlation function. Physically, the composition at two points separated by distances less than R_{max} are positively correlated; however, the strength of these correlations is considerably less than those observed for passive coarsening owning to heterogeneity within these larger domains.† R_{max} grows roughly linearly in time until *ca.* $0.1L$, beyond which the finite size of the simulation cell begins to significantly influence the systems' dynamics.†

In contrast to systems with frictional damping, which ultimately phase separate into bulk domains of arbitrary size, the multi-scale structures that arise in the zero-friction limit appear to avoid macroscopic phase separation indefinitely. Instead, active rotation continuously stretches and folds the growing domains in an effort to "mix" the fluid while it stubbornly attempts to "unmix". Competition between these two processes cause the formation of the smaller structures of order R_{min}, which appear to persist indefinitely (barring finite size effects).

Zero Damping & Strong Rotation ($\beta \to 0$, $\alpha \gg 1$)

As the magnitude of active rotation is increased such that $\alpha \gg 1$, the system transitions to a new dynamical regime characterized by the nucleation of localized vortices that move, interact, and combine within an otherwise homogeneous fluid (Fig. 2.11). Ultimately, a single pair of counter-rotating vortices—a "vortex doublet"—emerges and propels itself autonomously throughout the domain, thereby mixing the fluid and preventing further phase separation.

In this regime, activity-driven flows begin to shape the dynamics of the composition *prior* to the formation of the bulk phases. At these early times, the characteristic fluid velocity scales as $U \sim \alpha\Delta\varphi$, where $\Delta\varphi$ characterizes the magnitude of composition variations over a unit length (corresponding to the size k^{-1} of the fastest growing mode). As above, convection begins to compete with diffusive transport when the Péclet number is of order unity, $Pe \sim \alpha\Delta\varphi \sim 1$. For strong rotation, activity-driven flows become significant even for partial phase separation—that is, when $\Delta\varphi \sim \alpha^{-1} \ll 1$.

Importantly, these convective flows have the potential to inhibit the further unmixing of the two fluid components. In the absence of active rotation, fluid unmixing proceeds exponentially as $\Delta\varphi \propto \exp(t/4)$. To inhibit phase separation, the shear rate in the fluid must exceed the rate of unmixing. Partial phase separation results in activity-driven flows with shear rates of order $\alpha\Delta\varphi$. Thus, when the extent of unmixing reaches a critical value—namely, $\Delta\varphi > \alpha^{-1}$—the resulting flows will act to oppose further unmixing.

Nevertheless, the spatial heterogeneity of the shear field allows for the nucleation of small vortices in locations with lower shear rates. As vortices form, they create regions of low shear (but high vorticity) in their interior that allow for further phase separation. At the same time, these vortices induce high shear rates in the surrounding fluid, which inhibits unmixing

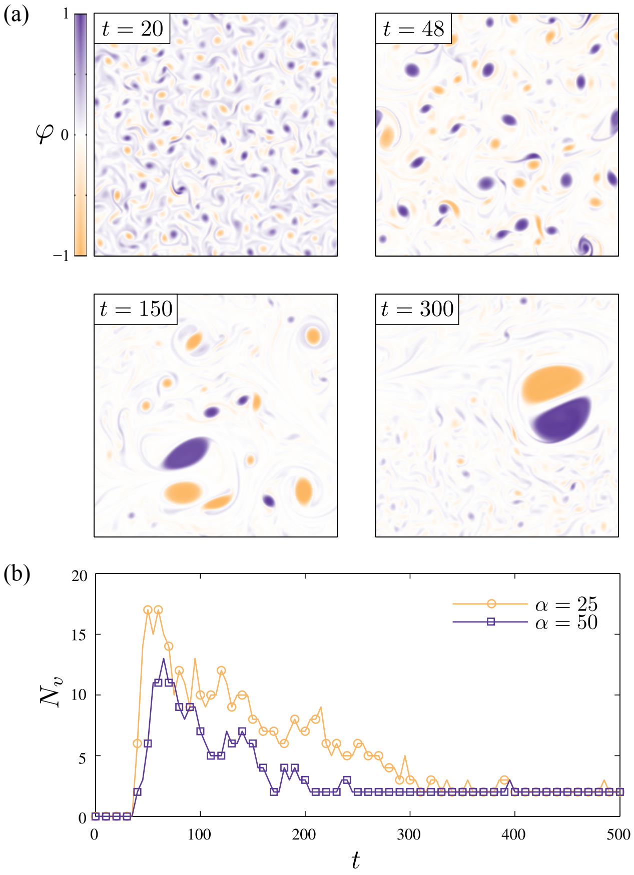


Figure 2.11: **Vortex doublet, $\beta \to 0$ and $\alpha \gg 1$.** (a) Order parameter $\varphi(x, y)$ at times $t = 20$, 48, 150, and 300 for parameters $\alpha = 50$ and $\beta = 0$; the size of simulation cell is $L = 256$. See ESI for the corresponding movie.† (b) Number of vortices as a function of time for $\beta = 0$ and $\alpha = 25$ and 50. Here, a vortex is identified as a localized region in which the composition is $|\varphi| \geq 0.3$ (see ESI for details).†

therein. As a vortex strengthens, the composition in its interior approaches $\varphi \sim \pm 1$, while that of the exterior remains largely homogeneous with $|\varphi| \ll 1$.

Following the initial nucleation phase, vortices move and deform in the swirling flows induced by their neighbors. Some are destroyed by strong shear due to larger neighbors; others grow and merge to form larger and more powerful vortices. In this way, the number of vortices decreases in time until only two large, counter-rotating vortices remain (Fig. 2.11). Together, these vortices form a stable, self-propelled vortex doublet of size R that swims about the domain with velocity $U \sim \alpha R$. The doublet creates a velocity disturbance that decays as $\alpha R^3/r^2$ with distance r from its center.† Consequently, a single vortex doublet can create shear rates of order unity (that necessary for mixing) at distances of $R\alpha^{1/3}$. This result is consistent with the simulation results shown in Figure 2.11, in which a single vortex doublet effectively "mixes" a region *ca.* four times as large as itself with $\alpha = 50$.

Results of the Microscopic Model

The key insight suggested by the continuum model is that phase separation in actively rotating liquids can drive convective flows that feedback into the system and direct the dynamic evolution of the growing phases. This result is further supported by microscopic simulations of actively rotating particles (Fig. 2.12). In particular, we studied the collective dynamics of $16,384$ spinners subject to moderate driving torques ($\alpha' = 0.25$) and different levels of frictional damping. Under these conditions, spinners unmix to form domains of like-rotating particles (Fig. 2.12a,b) that grow steadily in time. Here, the size R' of the growing domains is quantified as the first root of the *integral* of the pair correlation function. Note that this measure is different from that used in the continuum model and is chosen for its decreased sensitivity to statistical fluctuations.

For relatively strong damping ($\beta' = 1$), the domain size increases as $R' \propto t'^{1/3}$ (Fig. 2.12a,c) in agreement with the continuum model (Fig. 2.8) and with previous microscopic results [24]. By contrast, as the damping parameter is decreased, convective flows emerge and accelerate the rate of coarsening (Fig. 2.12b). Consistent with the continuum model, the growth exponent increases from 1/3 toward 1 as the damping parameter β' is reduced. The exploration of smaller damping parameters and/or longer simulations times in the microscopic model was found to be computationally prohibitive. Therefore, it is unclear if the microscopic model will approach a scaling exponent of *ca.* 1 in the limit as $\beta' \to 0$ or if it will return to a scaling exponent of 1/3 in the limit of long times.

To explore the possibility of "vortex doublets" in the microscopic model, we increased the driving torque to $\alpha' = 2.5$ under conditions of weak damping ($\beta' = 0.01$). Consistent with the continuum model, the increased rotational activity of the particles was sufficient to inhibit the unmixing of the spinners; however, we did not observe the nucleation of localized vortices from homogeneous initial conditions. Furthermore, we applied the same driving torque to an initially phase-separated system under conditions of weak damping ($\alpha' = 10$ and $\beta' = 0.01$). The active rotation of the particles resulted in their complete mixing, which suggests that the absence of vortex doublets in the microscopic model is not the result of a nucleation barrier.

The discrepancies between the two models under conditions of high torque likely arise from a failure of the continuum model to account for the microscopic effects of active rotation

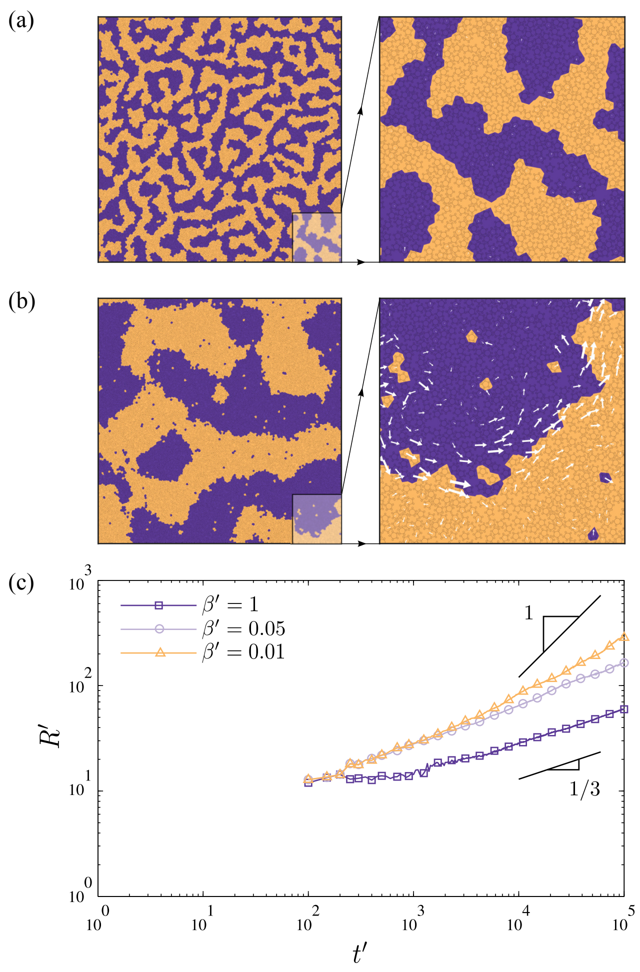


Figure 2.12: **Microscopic model.** (a,b) Representative snapshots of a 50:50 mixture of 16,384 spinners driven to rotate in clockwise (orange) and counterclockwise (purple) directions at time $t' = 10^5$. The visualization of the particles by their Voronoi tessellation is shown in the right images along with the fluid flows. The image in (a) show "passive coarsening" for parameters $\alpha' = 0.25$ and $\beta' = 1$ corresponding to strong frictional damping. Image in (b) show "active coarsening" for parameters $\alpha' = 0.25$ and $\beta' = 0.01$ corresponding to weak damping. White arrows show the direction of fluid flows. (c) Domain size R' as a function of time for $\alpha' = 0.25$ and different amounts of frictional damping $\beta' = 1$, 0.05, and 0.01; this plot is analogous to that in Fig. 2.9c. Here, R' is defined as the first zero of the integral of the pair correlation function; see ESI for the corresponding movie.†

on fluid-fluid phase separation. In the microscopic model, the driving torque α' sets an energy scale, which must be significantly less than that of attractive interparticle interactions to achieve microscopic phase separation. In the high-torque simulations, these two energy scales are comparable such that active rotation prohibits any and all phase separation. In such systems, the strong torques required for the formation of vortex structures lead instead to the complete mixing of the binary fluid. Additional theoretical work is required to account for the microscopic details of the particle-based simulations (e.g., particle shape) within an accurate hydrodynamic description. Nevertheless, it remains likely that the vortex structures observed in the hydrodynamic model could also be realized in analogous microscopic models provided the driving force for phase separation is sufficiently strong.

2.2.3 Conclusions

To summarize, we presented a phenomenological, continuum model for studying the dynamics of phase separation in binary liquids with active rotation. Convective flows induced by the fluid activity result in accelerated coarsening as compared to spinodal decomposition in passive liquids. The transition from passive to active coarsening is determined primarily by the strength of active rotation and that of frictional interactions with the stationary surroundings. In addition to active coarsening, the continuum model also predicts the formation of self-propelled vortex doublets under conditions of strong rotation and weak frictional damping. These dissipative structures emerge spontaneously from the competition between fluid mixing via active rotation and fluid unmixing due to interparticle interactions. Many of the trends observed in the continuum model such as accelerated coarsening are also reproduced by microscopic kinetic simulations of counter-rotating particle mixtures. Further work is needed to develop a more rigorous connection between such microscopic models and the continuum hydrodynamics of actively rotating fluids. We are currently developing experimental models of counter-rotating particle mixtures to explore and elucidate the collective dynamics that emerge from active rotation. Such nonequilibrium materials have the potential to achieve life-like properties such the abilities to adapt, reconfigure, and repair dynamically in response to environmental stimuli.

Acknowledgments

This work was supported as part of the Center for Bio-Inspired Energy Science, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0000989.

Chapter 3

Machine Learning for Crystal Structure

Note: much of this chapter was originally published in AIChE Journal in 2018[114] and is reproduced with modifications below.

Machine learning (ML) is a rapidly growing field of study that uses data-driven approaches to perform a wide variety of tasks that are often difficult to design algorithms for by hand. Within the last few years, there has been an explosion of interest in ML as great progress has been made in image classification[115], object detection[116, 117], image enhancement[118, 119, 120, 121] and modification[122, 123, 124], and even playing games[125, 126, 127, 128]. Advancements in the field have been driven not only by novel algorithms and software, but also by parallel and specialized hardware development. In this chapter, I discuss strategies I have developed to apply these exciting new tools and methods to the problem of three-dimensional structure analysis.

3.1 Representations in Machine Learning

One of the key components of applying machine learning to new problems lies in finding an appropriate *representation* of the data. Typically these involve computing numerical "fingerprints" of each observation which can then be fed into one of many machine learning methods. The best sorts of representations are highly sensitive to the phenomena we are interested in studying (like microscopic structure), insensitive to phenomena that we are not interested in (such as thermal noise), efficient to compute, low in number of external parameters, and dense in information. For example, the current workhorse representation for image tasks involves utilizing the Fourier-domain representation of images through convolutional neural networks[115], which enables learning of translation-invariant features directly within a neural network. This representation is particularly useful because it can be inverted, allowing neural networks that have been trained for a task to be run "in reverse," to generate characteristic samples for an output they have learned[129].

There are a few broad classes of representations in common use, depending on the choice of problem and, especially, the ML algorithm to be applied. One of the most prevalent early approaches was the "bag of features:" for the application of interest, create a list of all the things we know how to compute that may have some correlation—however weak—to the attributes we would like to model. This list of descriptors would often be pared down based on validation data through feature selection algorithms before creating a model *via* linear regression, simple neural networks, or similar methods. The "bag of features" method can be powerful if we already know much about how the underlying system works, but it can be difficult to apply to very generic problems. This is because the features we use encode our own biases and knowledge of how the system behaves into the representation, and if we truly knew everything about the behavior of the system we would be less likely to use ML to model it.

Another broad approach to representation is through *kernel methods*. Kernel-based methods operate on similarities or distances defined between pairs of observations, rather than a fixed, high-dimensional "feature space" as in the "bag of features" approach. These can be useful in the many cases when it is difficult to construct fixed representations for observations, as in many types of data that are structured like graphs. However, it can sometimes be difficult to transfer the learned information about a dataset to new observations, or to incrementally add new observations to the model. For example, dimensionality reduction problems are often constructed in terms of minimizing some function of the coordinate assignments of each vertex in a weighted graph. The coordinates for a set of points can be computed simultaneously, but it is difficult with many methods to efficiently compute the coordinates of a new observation. This makes it necessary to keep a library of "anchor" observations—which could be the entire dataset, or some subset of the original data—as a reference to apply the method to new data.

A third approach to representation involves learning a useful numerical description from the data at training time. For example, the layers in convolutional neural networks (CNNs) commonly used for image tasks learn a localized "filter" function that is applied to a neighborhood of a few pixels and produces a new image. As CNNs are trained, lower layers in the network typically learn to identify edges and corners, while higher levels could be tuned to identify higher-level concepts, like faces or other objects[130, 131]. In a similar manner, ML models related to language problems often use word embeddings, that map words to vectors in some continuous "embedding space." These vectors can reveal intuitive analogies; for example, if $E(w)$ is the embedding of a word, it is possible for a network to learn that $E(\text{king}) - E(\text{man}) + E(\text{woman}) \approx E(\text{queen})$[132, 133].

3.2 Representing Crystal Structures

Many representations have previously been proposed or used to apply machine learning or automatic analysis to problems involving crystal structures. For decades, people have used the Steinhardt order parameters[134] to describe the ordering of various structures, from crystalline solids[135] to liquid water[136]. These scalar quantities measure the development of *bond order*—the arrangement of nearest-neighbor bonds between particles in a symmetrical fashion—by essentially computing the power spectrum of the spherical harmonic decomposition at a particular spherical harmonic degree ℓ:

$$Q_\ell(i) = \sqrt{\frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{m=\ell} \left| \frac{1}{N_{n,i}} \sum_{j=1}^{N_{n,i}} Y_\ell^m(\vec{r}_{ij}) \right|^2} \tag{3.1}$$

Looking at the power spectrum rather than particular spherical harmonic Y_ℓ^m values is required in this case to make the representation invariant with respect to rigid rotations of the local environment. While the Steinhardt order parameters have been successfully used for many applications, they are not an ideal representation. For example, the standard version presented in Equation 3.1 often has difficulty between distinguishing the two common close-packed sphere packings, $cF4$-Cu[1] (commonly known as FCC) and $hP2$-Mg (commonly known as HCP)[137]. Furthermore, two different sets of particle interactions that yield the same assembled structure can exhibit substantial differences in the distribution of Steinhardt order parameters[137], limiting the transferability of any machine learning model or other algorithm based on these values.

More recent work has revealed a plethora of potential data representations for self-assembling systems, each with its own strengths and weaknesses[138, 139]. In the study

---

[1] Here we identify structure types by a two-part description: the Pearson symbol—identifying the Bravais lattice and number of particles in the unit cell—and a compound that is known to commonly form that structure.

of glassy dynamics, representations based on near-neighbor distances and angles have been successfully applied to formulate a structural basis for the glass transition[140, 141, 142, 143]. For crystal structures, methods have been devised to utilize the radial distribution function[11], kernel methods based on graph representations of particle neighborhoods[144, 145, 146], and particular sets of highly symmetric environments[147]. Even more recently, specialized neural network architectures and encoding schemes have been proposed for point clouds[148, 149, 150] and graphs[151], both of which can be mapped to the local environment of self-assembling particles in a straightforward manner.

A visual representation of crystal structure that has often been used to identify complex structures is the bond-orientational order diagram, or BOOD[152, 153, 9, 17, 12]. In a BOOD, the nearest-neighbor bonds of all particles are translated to the origin and projected onto the surface of a sphere, as shown in Figure 3.1.

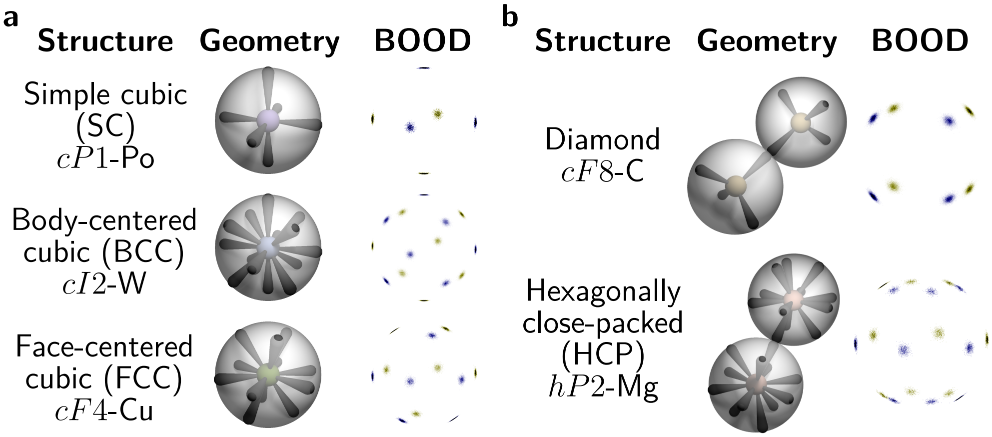


Figure 3.1: Generation of Bond-Orientational Order Diagrams (BOODs) from crystal structures. Nearest-neighbor bonds for all particles are projected onto the surface of a sphere. Bonds on the reader-facing half of the sphere have a blue corona, while back-facing bonds have a yellow corona. (a) Structures with only one local environment have BOODs that are easy to identify directly from the configuration of a single particle's neighbors. (b) BOODs of structures with multiple orientations of local environments are the superposition of all orientations and environment types.

While the BOOD is a useful way to quickly identify many structures at a glance, it is accompanied by limitations that make it difficult to apply as an order parameter for simulation data. The primary limitation is that the BOOD is a simple global projection of all of the bonds in the system. This means that, if multiple structures or grains are present in the same system, the resulting BOOD is the superposition of more than one ideal

BOOD. Often, crystals with grain boundaries will orient themselves in a particular way, which can lead to BOODs that are still highly symmetric, but ultimately misleading without deeper investigation. Two examples of these types of structures are in Figure 3.2, which shows configurations of self-assembled particles mostly arranged in the $cF4$-Cu structure, but with various commonly-occurring defects. Another problem is that, when attempting to use BOOD fingerprints for structure identification in an automated manner, the points would first have to be brought into a common orientation with reference BOODs. This process, known in the computer vision community as the "registration" problem, can be computationally costly and requires a prebuilt library of reference structures for comparison.

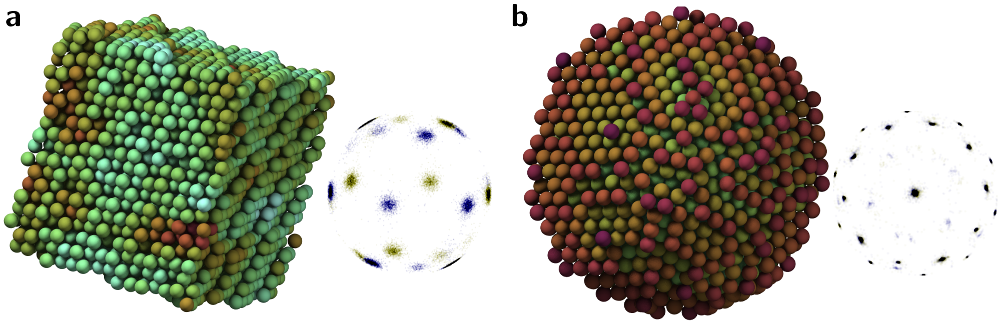


Figure 3.2: Self-assembled structures of close-packed spheres and their BOODs. Particles are colored by local density. (a) Hard spheres commonly form a face-centered cubic close packed structure, $cF4$-Cu, with stacking faults. This causes the BOOD to have a similar pattern to $hP2$-Mg, a structural competitor found under similar circumstances. (b) Here, attractive particles have assembled a structure that is predominantly $cF4$-Cu, but with defects that cause the BOOD to exhibit approximate 5-fold symmetry.

3.3 Local Bond-Orientational Order Diagrams

In this work we use descriptors based on spherical harmonics of nearest-neighbor bonds[114], much like the Steinhardt order parameters. However, instead of creating a rotationally-invariant combination of spherical harmonics, we first orient the environment of each particle according to a local measure: the inertia tensor of the environment. This method gives meaning to the position of neighboring particles in this reference frame while simultaneously making the description rotationally invariant[2]. In this manner, we can directly utilize indi-

---

[2] Instead of building rotational invariance into the descriptors, it is possible to replicate training data many times, each with a different rotations applied to them; this process, called *data augmentation* in the

vidual spherical harmonics Y_ℓ^m instead of only rotationally invariant combinations of them, yielding a much richer description of particle neighborhoods.

In greater detail, we pick some number of nearest neighbors, N_n. For every reference particle i in the system, we compute its N_n spatially nearest neighbors and compute the vector $\vec{r}_{ij}$ from the reference particle to its neighbor j. We calculate the inertia tensor of this set of points, defined as

$$\bar{I}(i, N_n) = \sum_{j=1}^{N_n} (\vec{r}_{ij} \cdot \vec{r}_{ij})\bar{1} - \vec{r}_{ij} \otimes \vec{r}_{ij}. \tag{3.2}$$

We then find an orientation that diagonalizes the inertia tensor by computing its eigenvalues and eigenvectors, which correspond to moments of inertia and principal axes of inertia, respectively. We assign the eigenvector with the largest eigenvalue to the z direction, the second-largest to the y direction, and the smallest to the x direction. After repeating this diagonalization procedure for each particle in the system, we are left with a set of bonds for every particle, oriented in a common way regardless of the orientation of the crystal the particles are forming. If multiple crystalline grains are present in the system, the signal of defects is proportional to the number of particles participating in the defect, rather than the volume of the crystal in each grain as in a standard, global BOOD. Because these descriptions create point clouds much like standard BOODS, but are based on purely local information, we call them *local BOODs*. We show several local BOODs for different neighborhood sizes in Figure 3.3.

As with any representation, local BOODs have their own set of caveats related to their usage. First, the appearance of the local BOOD is strongly related not only to the crystal structure, but also the neighborhood size N_n we choose. In particular, if the structure happens to be nearly cubically symmetric for the number of neighbors we pick, then the moments of inertia will be approximately equal and the orientation vectors we find will be random, making us unable to even distinguish between a fluid and the crystal. To solve this problem, we compute the local BOODs for a range of neighborhood sizes. While fluids will exhibit noise for all neighborhood sizes, crystals with cubic symmetry will be noisy only for particular neighborhood sizes and will have well-behaved local BOODs for others. A second issue in the use of local BOODs lies in choosing a particular number of *nearest* neighbors. Because crystals usually have particles arranged in concentric shells of equal distances, for

---

ML community, depends on the ML model to learn rotational invariance from the data and likely requires larger amounts of training data and more complicated models.

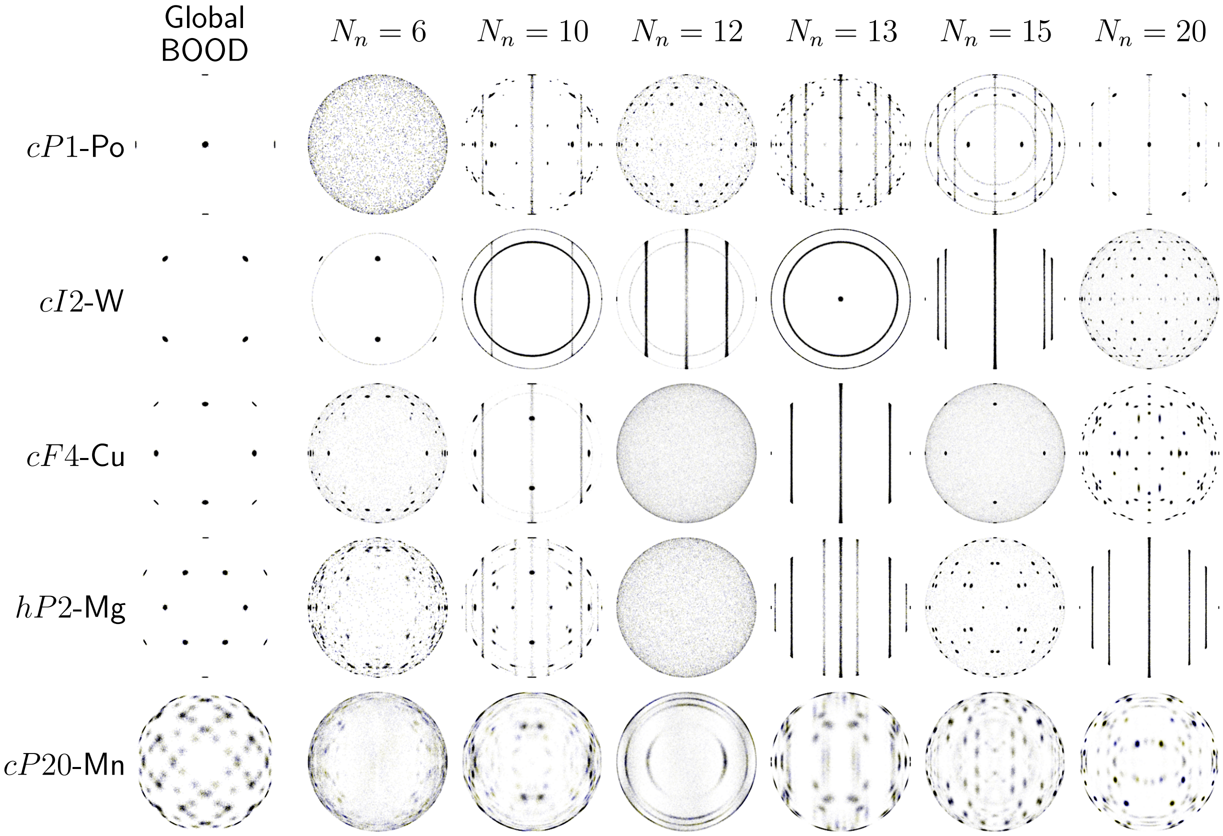


Figure 3.3: Global and local BOODs for various structures (rows) and neighborhood sizes (columns).

real crystals with thermal noise our algorithm essentially randomly picks neighbors from these shells in most cases. For example, the nearest 8 neighbors in a $cI2$-W (BCC) structure are arranged like the vertices of a cube. If we attempt to find the nearest 2 neighbors of the reference particle in the center of the cube, we may find those two neighbors along a cube edge, on a diagonal across a face of the cube, or a diagonal across the body of the cube. This problem can be solved by sufficient averaging to sample the distribution of these configurations. Alternatively, supervised learning algorithms would be able to naturally learn that the three distinct appearances in the above example correspond to the same structure.

To convert local BOODs—which are essentially a density map on the surface of the sphere—into a numerical representation, we utilize the spherical harmonic transformation. By summing the spherical harmonics for a given particle's local BOOD over its neighbors, we can find particular Y_ℓ^m configurations that constructively interfere in a way corresponding to a given symmetric arrangement.

3.3.1 Comparison to Steinhardt Order Parameters for Two Simple Structures

A natural first question is how well local environment spherical harmonics perform compared to the Steinhardt order parameters. We compare the performance of these two methods on a problem that is notoriously difficult for the Steinhardt order parameters: distinguishing two close-packed arrangements of spheres, $cF4$-Cu and $hP2$-Mg. We create thermalized configurations by replicating each unit cell to at least 4,000 particles and applying Gaussian noise (with the ratio of noise to nearest-neighbor distance $\frac{\sigma}{|r_{NN}|} = 0.05$). We compute the Steinhardt order parameter Q_6 using the 12 nearest neighbors for each particle with freud[154]. We define the *neighbor-averaged* local environment spherical harmonics as $\bar{Y}_\ell^m(i, N_n)$:

$$\bar{Y}_\ell^m(i, N_n) = \frac{1}{N_n} \left| \sum_{j=1}^{N_n} Y_\ell^m(\theta_{ij}, \phi_{ij}) \right| . \tag{3.3}$$

We formulate a logistic regression model to map sets of $\bar{Y}_\ell^m(i, N_n)$ (for 4 to 12 neighbors and spherical harmonic degree from 1 to 6) onto a predicted probability of being one structure or the other using scikit-learn[155]. Half the data were reserved for evaluation of the logistic regression model. As can be seen in histograms of the identification results shown in Figure 3.4, the local environment spherical harmonics are significantly better at distinguishing between the two structures than the Steinhardt Q_6. This is because the local environment spherical harmonics are a much higher-dimensional, richer description of the neighborhood around each particle.

3.4 Structure Identification using Local Environment Spherical Harmonics

We next analyze structures in a phase diagram that was identified by hand in 2015[12], using machine learning[114]. That paper studied the self-assembly behavior of particles interacting *via* an isotropic pair potential with two parameters, k and ϕ, defined as

$$V(r)/\epsilon_0 = \frac{1}{r^{15}} + \frac{1}{r^3} \cos\left(k(r - 1.25) - \phi\right) . \tag{3.4}$$

This interaction was truncated at the third maximum, with a shift to zero and smoothing applied to the force at the cutoff distance to create a short-range, attractive potential. Systems

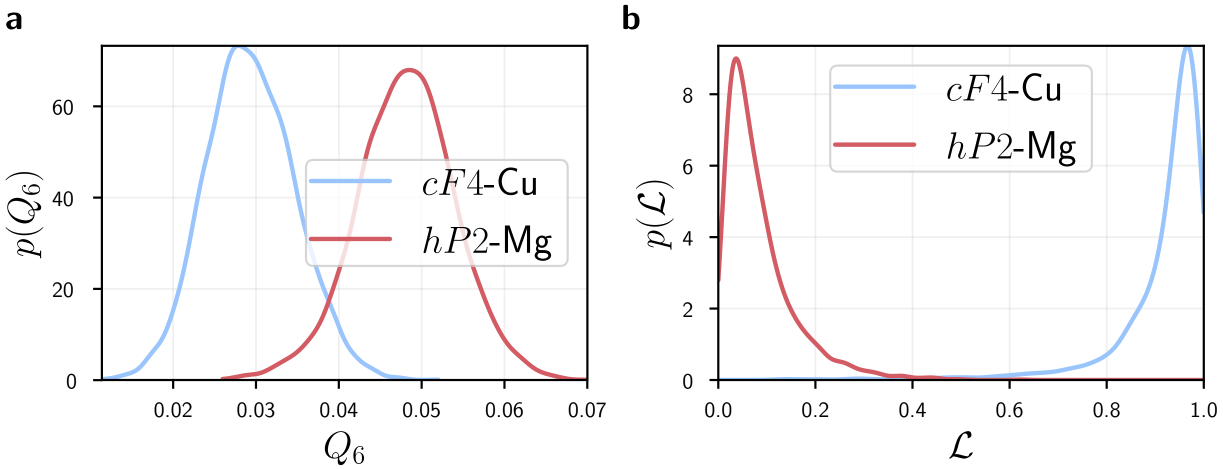


Figure 3.4: Probability distributions of (a) the Steinhardt order parameter Q_6 and (b) the probability under a logistic regression model for particles to be in the $cF4$-Cu structure using local environment spherical harmonics as in Equation 3.3. While the Steinhardt order parameters have difficulty distinguishing between the two structures for some observations, the local environment spherical harmonics are much richer in information, enabling easy classification.

of particles were cooled from a high temperature slowly, causing a liquid droplet to form and usually crystallize into an ordered structure. The phase diagram and unit cells of many structures, as identified by manual inspection, are shown in Figure 3.5. Within the clathrate region of the phase diagram, three different structures are found: clathrates I ($cP54$-K_4Si_{23}), II ($cF160$-$Na_{24}Si_{136}$), and IV ($hP47$-$Li_{14.7}Mg_{36.8}Cu_{21.5}Ga_{66}$). The clathrate structures often occur in mixed configurations within the same sample, but there is a bias toward $cP54$-K_4Si_{23} at low k, $cF160$-$Na_{24}Si_{136}$ at high k, and $hP47$-$Li_{14.7}Mg_{36.8}Cu_{21.5}Ga_{66}$ in between. The quasicrystal region contains two types of icosahedral quasicrystal—distinguished by their density as a low-density and intermediate-density—as well as a high-density, periodic icosahedral quasicrystal approximant that exhibits similar structural motifs to the true quasicrystal structures.

To study these data with ML, we use two main approaches: unsupervised learning and supervised learning. Unsupervised learning is so-called because, for each piece of training data, the algorithm is not given the desired model output. These models are often used for clustering (placing points into distinct, similar “buckets”) or projection (placing points on a continuous space such that nearby points are similar to each other). In contrast, supervised learning models are trained with both the model input and the desired model output for each

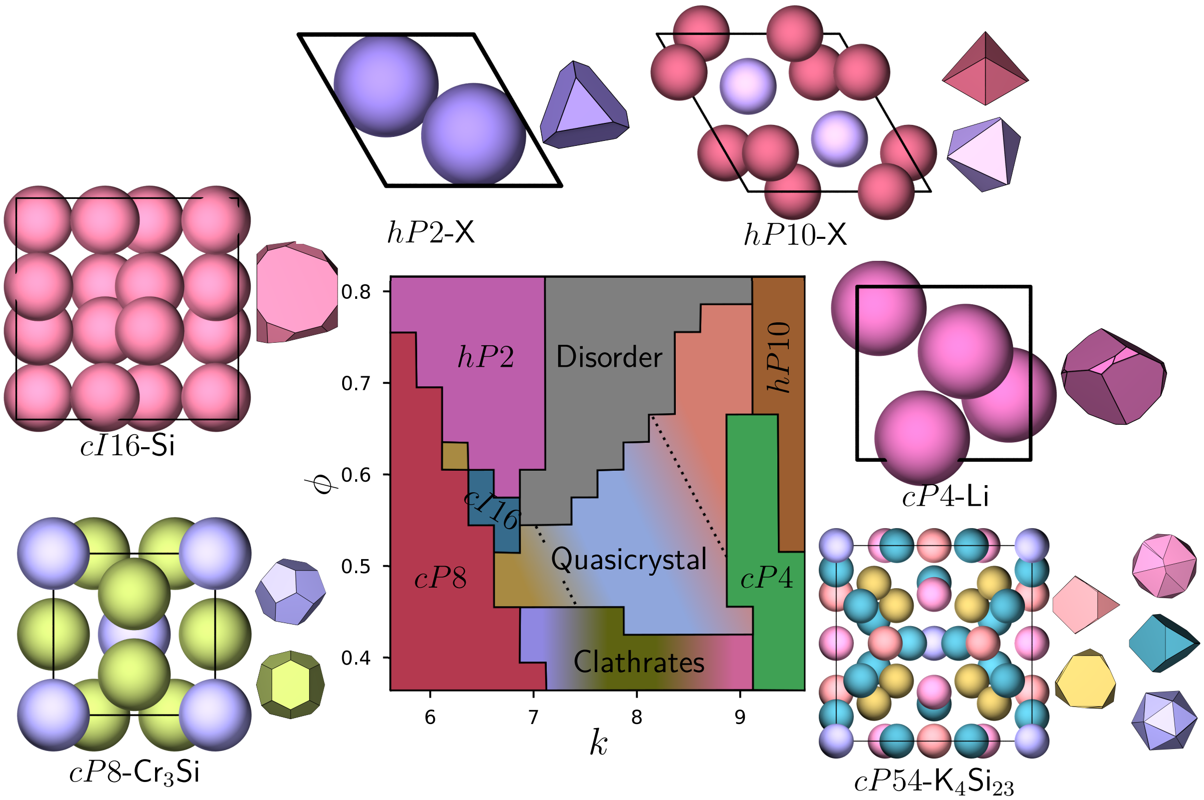


Figure 3.5: Parameter space and assembled structures for a two-well oscillatory pair potential[12]. Unit cells and Voronoi polyhedra corresponding to each type of local environment are displayed. The regions that form clathrates form one of three clathrate structures, which are often mixed within the same simulation box; in this figure we only show the unit cell for clathrate I, $cP54$-K_4Si_{23}.

point. These are often used for classification (assigning points to a category) and regression (associating points with one or more continuous values).

3.4.1 Unsupervised Learning

We first analyze these simulation results with ML *via* Gaussian mixture models (GMMs)—a form of unsupervised learning—available in the scikit-learn python package[155]. GMMs create a probability distribution function for a set of observations using one or more multivariate Gaussian distributions. Unsupervised learning methods—like GMMs—are a good choice for simulation data that have not yet been analyzed, for at that point we do not know which structures will be formed. The expectation-maximization (EM) algorithm used to train GMMs here finds local, not global, optima as a function of the model parameters. A common trick in training these models is to perform multiple rounds of training with random initializations of the model parameters before performing EM, but regardless the method benefits from having compact, well-averaged data. We make the data from this dataset—which could be expressed as approximately 1,000 simulations of 4,000 particles each with spherical harmonic vector dimension of 1,000—more manageable for the model in a few ways. First, we use the *globally-averaged* local environment spherical harmonics for each system, calculated as

$$\bar{\bar{Y}}_\ell^m = \frac{1}{N_p N_n} \left| \sum_{i=1}^{N_p} \sum_{j=1}^{N_n} Y_\ell^m(\theta_{ij}, \phi_{ij}) \right| \tag{3.5}$$

where θ_{ij} and ϕ_{ij} correspond to the spherical coordinates of particle i's neighbor j using the principal axes of inertia as described in Section 3.3, N_p is the number of particles in the system, and N_n the number of neighbors chosen (here, ranging from 4 to 7). We then apply principal component analysis (PCA)[156] to further reduce the dimensionality of the data. PCA should ideally capture our intuition that there would be a greater distance in descriptor space between different structures than between two examples of the same structure under thermal noise.

We then train GMMs on the PCA projection of the data. To choose the optimal number of components, we use the Bayesian information criterion[157], which measures the quality of the GMM fit while penalizing GMMs with many free parameters to avoid formulating overly complex models. It is possible that clusters—which would ideally correspond to distinct structures—are not perfect Gaussians in the descriptor space. In that case, multiple Gaussian components may be used to fit the data from a single structure. To fix this, we perform a step of greedily merging GMM components into clusters according to the likelihood of observations being confused between two clusters[158]. This yields an entropy curve that typically has a sharp, upward-facing elbow; the appropriate cluster merging map lies somewhere around the curved region of the elbow. The results of this merging procedure—as well as three phase diagrams during the cluster merging process—are shown in Figure 3.6.

The unsupervised learning process yields a cluster identity (or a distribution of cluster likelihoods) for each simulation snapshot, but not a direct structure assignment. A characteristic example from each cluster can be used to identify the cluster in its entirety, reducing the number of systems to manually identify from over 1,000 to 15 or fewer in this case. For this phase diagram, the GMMs agree remarkably well with the manually-identified results,

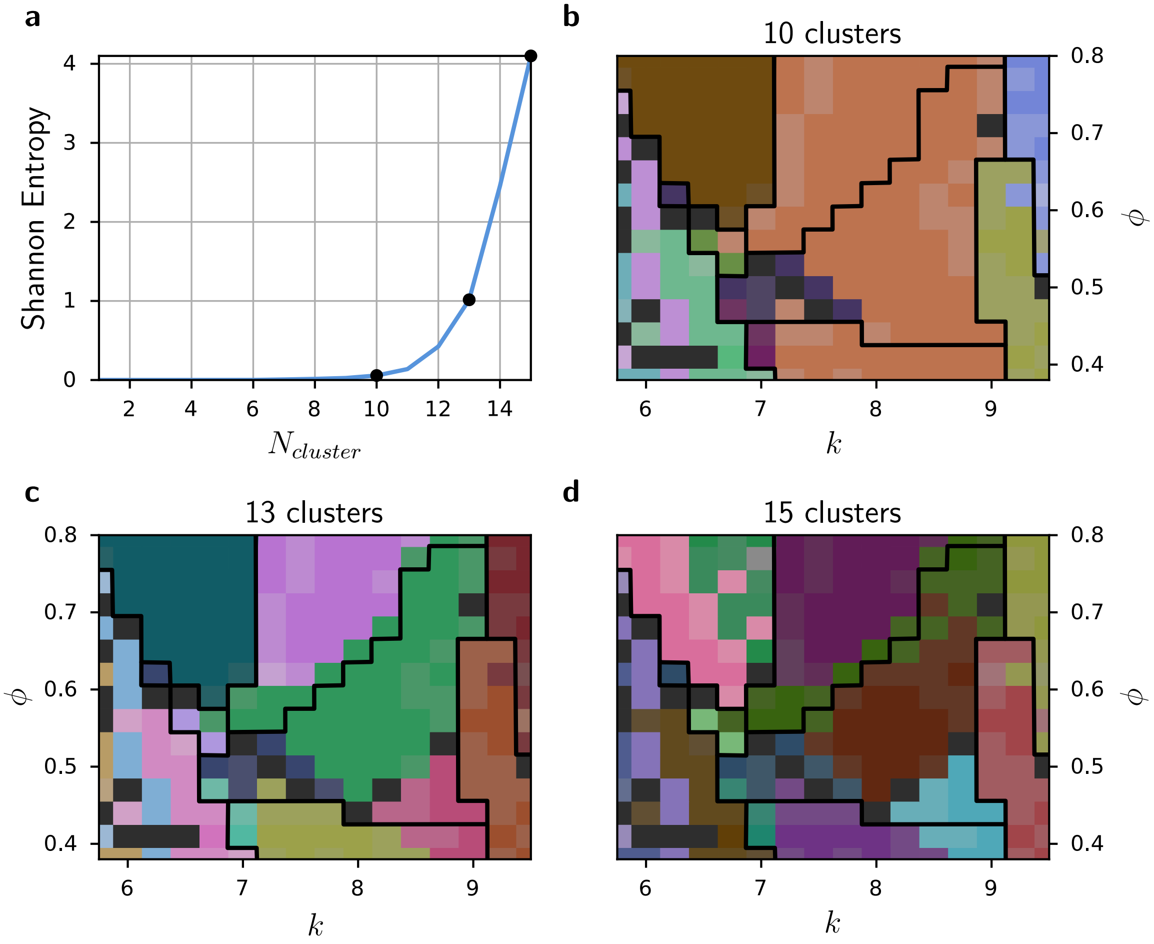


Figure 3.6: Results of training Gaussian Mixture Models (GMMs)[114] on a dataset forming several structures, including an icosahedral quasicrystal[12]. Phase boundaries found by hand in the original study are indicated by black lines. Gray regions indicate ambiguous points with multiple cluster assignments among the independent replicas. (a) Shannon entropy of the observations as GMM components are merged from 15 distinct clusters to 1. (b-d) Phase diagrams, colored by the clustering scheme, as GMM components are merged into clusters for 10, 13, and 15 clusters.

even down to the presence of multiple distinct types of clathrate and quasicrystal structures.

3.4.2 Supervised Learning

If the structural behavior of a system is already known, then supervised learning can be utilized. Supervised classification algorithms are generally easier to apply than unsupervised clustering algorithms because there are multiple conflicting ideas of what may constitute a "cluster"[159]. For the quasicrystal phase diagram dataset, we can train a simple feed-forward artificial neural network (ANN)[3] to classify per-particle neighbor-averaged spherical harmonics, $\bar{Y}_\ell^m$, as in Equation 3.3. We use one simulation frame of data for each of the crystal structures we are interested in as training data, including the various distinct types of clathrate and icosahedral quasicrystal structures found by hand. We also select four systems from the disordered region of the phase diagram to be able to identify when particles do not form any globally-ordered structure. We then generate a phase diagram by identifying the most prevalent structure prediction for all particles in each simulated system, as shown in Figure 3.7.

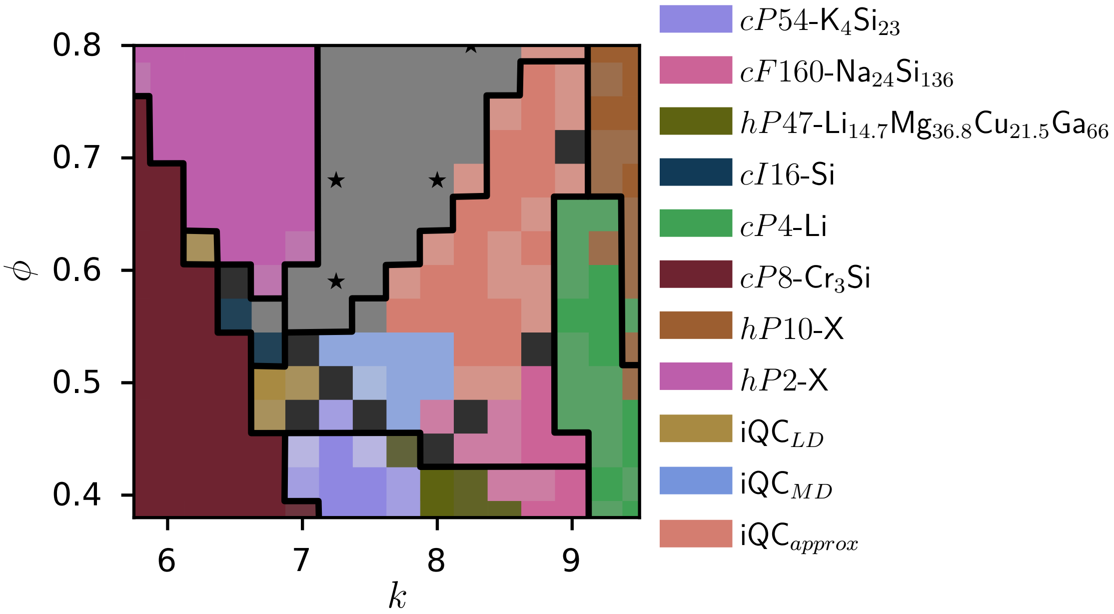


Figure 3.7: Phase diagram of the icosahedral quasicrystal dataset, generated *via* supervised learning. Black lines indicate phase boundaries as identified by hand[12]. Coordinates in parameter space are colored by the most prevalent structure identified among its independent statistical replicas. Dark gray regions did not have a clear majority among the replicas. The four stars in the disordered region indicate parameters which were used as training data to formulate the "disordered" structure class. The "iQC" regions indicate the icosahedral quasicrystals or their high-density, periodic approximant.

While applying supervised learning algorithms for structure identification seems straightforward, there are several details above and beyond the typical overfitting concerns of ML that require careful attention. The first concern lies in crystalline data generation. While

---

[3] Artificial neural network models were produced with the python library Keras [160].

it should be possible to use any structure, as long as it is at least slightly thermalized in order to sample the distribution of neighbor shell placements as described in Section 3.3, trained models may yield poor results when thermal noise is increased or the models are used to identify real thermalized systems. As an example, consider the schematic in Figure 3.8. Suppose that Figure 3.8(a) illustrates the distribution of two structures in structure space, trained with small Gaussian noise applied to particle positions for unit cells replicated a few thousand times. The ML models we train will formulate some sort of decision surface, like one of the dashed lines present in the diagram. Many types of ML model care primarily about classification rate and would rank each of these surfaces similarly. Now suppose that increasing the noise (or performing a simulation using interactions that may change the structure of the noise entirely) yields a different distribution in descriptor space, as shown in 3.8(b). For this reason, it may be desirable to use less well-ordered structures as training data. This may be as simple as applying sufficiently large random noise to a perfectly well-ordered structure, but any workflows using this method instead of real structural examples should carefully validate classification results by hand.

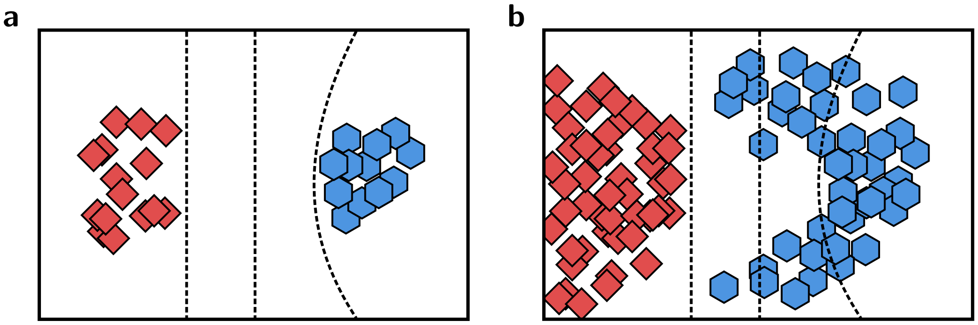


Figure 3.8: Schematic illustrating the possible impact of structural order on ML models. ML models trained on high-quality crystals (a) may perform poorly on less well-ordered variants of the same structure (b)

A related question concerns how to best deal with disordered structures. While some ML methods—like Bayesian neural networks[161]—can give an estimate of the uncertainty of their predictions and could thereby be able to distinguish disordered structures given only ordered structures as training data, most methods are not so equipped and would need to also be given training data if disordered systems are to be classified. Another important consideration is the distinction between local and global ordering: it is possible for a system to exhibit some sort of ordered signal locally without that order propagating through the whole sample. We have some evidence that this can be the case for the local environment spherical harmonics: consider the GMM results of Figure 3.6(d). On the border between the disordered phase and the icosahedral region, a distinct GMM component is found that lies within the globally ordered icosahedral region as well as the globally disordered part of the phase diagram. We may expect this to be due to the particles forming locally organized motifs that are unable to fully propagate through the system under the given assembly conditions.

3.5 Other Applications

3.5.1 Dimensionality Reduction

Dimensionality reduction techniques, broadly speaking, are methods to transform observations in some input space into a lower dimensional space. This transformation could be a linear, simple projection—as is the case with PCA, for example—or nonlinear. Nonlinear dimensionality reduction techniques often work on a graph of nearest-neighbor distances[162, 163, 164], which allow them to work even when we only want to use a distance kernel instead of a fixed observational space. These nonlinear dimensionality reduction techniques try to cause neighbors that are similar to each other in observational space to be near each other in the projected space. That said, the results of nonlinear dimensionality reduction are typically quite sensitive to hyperparameters[165] and the quantity of input data, making them often more useful for qualitative, rather than quantitative, comparisons.

Understanding Observation Configurations in Descriptor Space

Dimensionality reduction can be used to give an idea of the topology of observations in descriptor space. While the details of shapes of individual clusters may vary depending on the details of the methods and hyperparameters used, key signatures like number of individual clusters are often robust. For example, we show the projection of the globally-averaged local environment spherical harmonics for the icosahedral quasicrystal dataset using two methods, t-SNE[162] and UMAP[164], in Figure 3.9.

We color each observation by the most common structure as found by an ANN over all particles in the system, as shown in Figure 3.7. Both dimensionality reduction methods very clearly show that there are distinct sets of observations for the simple periodic structures. However, the more complex structures—the clathrates and quasicrystals—are joined to the

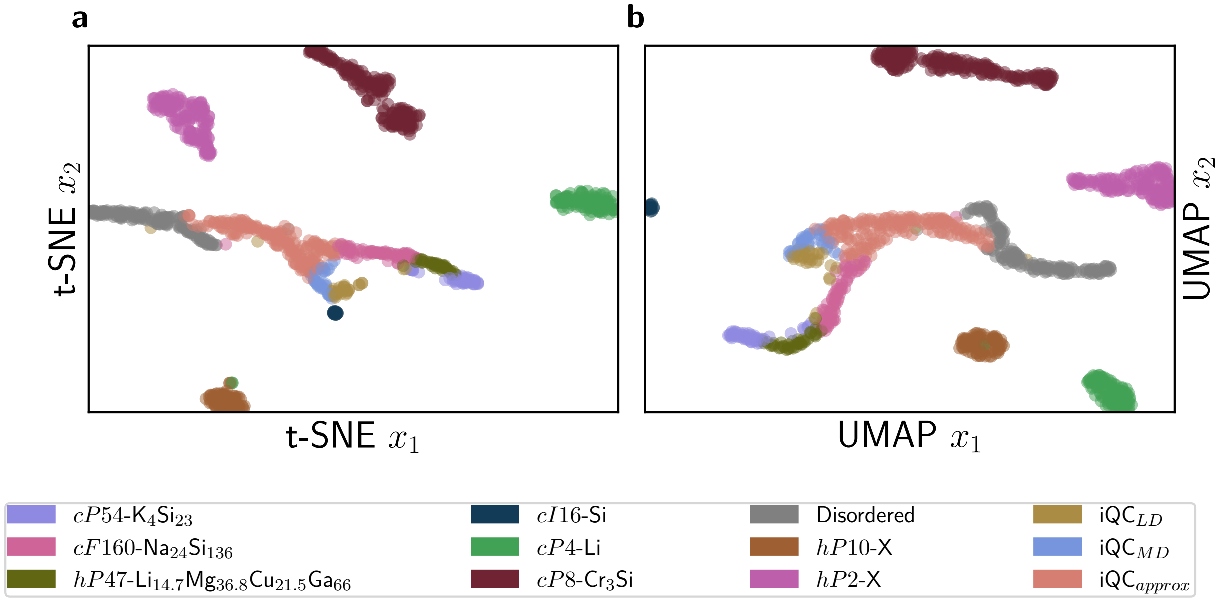


Figure 3.9: Nonlinear dimensionality reduction of globally-averaged local environment spherical harmonics using two methods, t-SNE (a) and UMAP (b). Observations are colored according to ANN analysis as in Section 3.4.2.

globally disordered systems *via* the high-density quasicrystal approximant. From this observation, we may expect common structural signatures to be found in the approximant, the $cF160$-Na$_{24}$Si$_{136}$ structure, and at least some of the disordered systems.

Studying Phase Transitions

Dimensionality reduction can also provide insight into dynamical processes, like crystallization. We perform simple simulations of Weeks-Chandler-Andersen[88] spheres in MD as we compress them from a thermalized fluid. Above a certain packing fraction, the spheres spontaneously crystallize into the $cF4$-Cu structure. We perform UMAP dimensionality reduction[164] on a 16-dimensional PCA projection of the neighbor-averaged local spherical harmonics of each particle (Equation 3.3) for 4 to 12 neighbors and spherical harmonic degree ℓ from 1 to 12 and map the UMAP value to a color for each particle. To account for the neighbor-shell placement problem described in Section3.3, we average each particle's spherical harmonics over 64 noisy versions of the positions, with normal-distributed noise ($\sigma_{noise} = 5 \cdot 10^{-2}\sigma_{WCA}$). We show snapshots of the system as crystallization occurs in Figure 3.10.

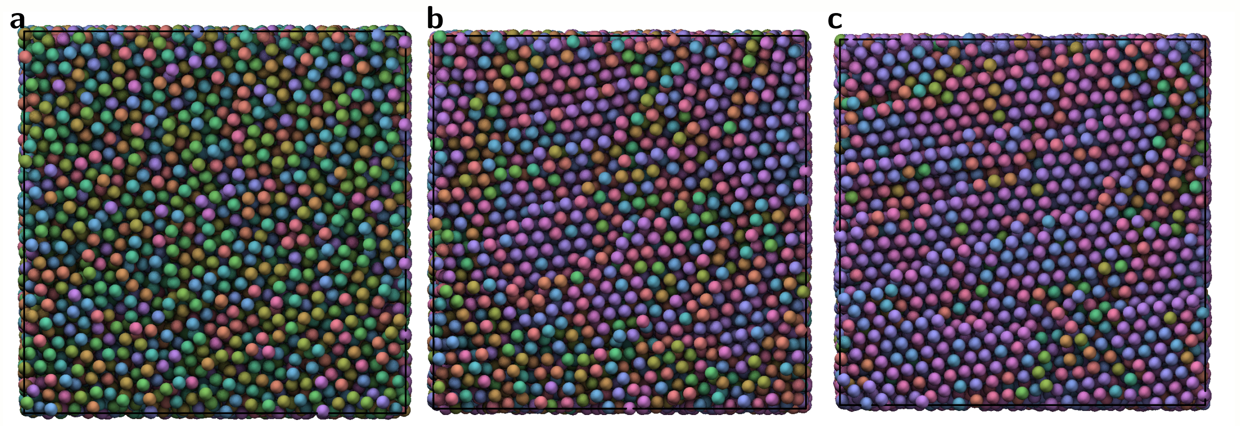


Figure 3.10: WCA spheres crystallizing into a $cF4$-Cu structure as they are compressed. Particles are colored by an order parameter derived *via* UMAP dimensionality reduction. The system is at a lower density in (a), intermediate density in (b), and higher density in (c).

3.5.2 Nucleation and Growth

One exciting direction of study enabled by this work is the characterization of structures and phenomena that were previously considered too complex to be tractable. For example, studying nucleation and growth of complex crystal structures with multiple local environments using the Steinhardt order parameters may involve difficult guesswork, but we can derive order parameters from observations using ML and local environment spherical harmonics. We observe the nucleation and growth of the $tP30$-CrFe structure, simulated using an oscillatory pair potential similar to the one in Equation 3.4, as the system is cooled below its crystallization temperature. We train an ANN to distinguish between the pre-nucleation fluid and a fully crystallized system of 16,384 particles and then apply the ANN to identify solid regions within a much larger simulation of 100,000 particles. As shown in Figure 3.11, the ANN is able to identify the regions that form crystals. It would be straightforward to use methods like these to study the nucleation behavior of this crystal *via* umbrella sampling[166, 167] or transition path sampling[168], for example.

3.6 Other Structural Representations

To aid in reproducibility of analysis methods, we created the *pythia* library, which contains many methods to generate static representations of structure. Providing a common interface

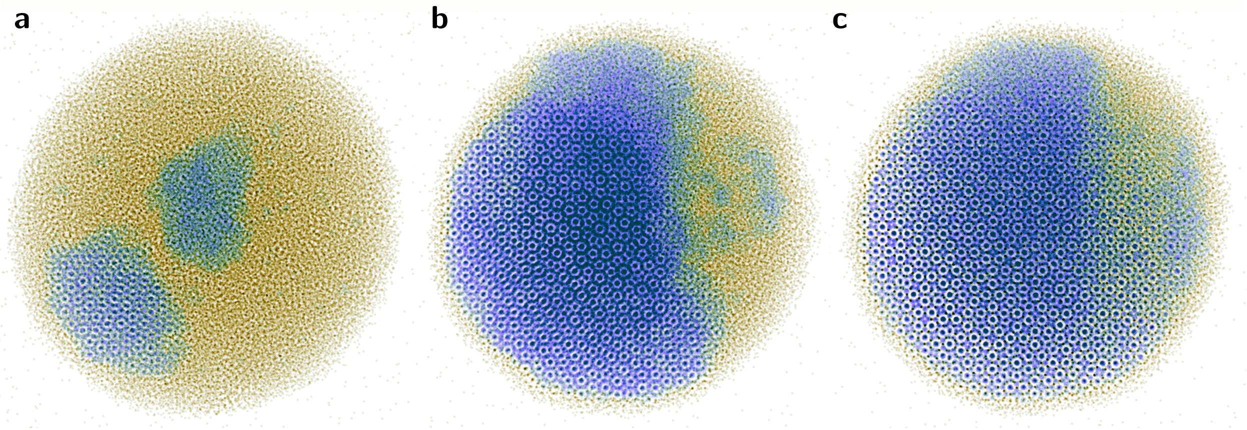


Figure 3.11: Projected density map of particles interacting *via* an oscillatory pair potential to form $tP30$-CrFe. Solid-like particles are blue and fluid-like particles are yellow. Parts (a)–(c) mark the forward progression of time.

for the various types of numerical fingerprints makes it easier to compare representations and choose the most suitable representation for the task at hand.

3.6.1 Nearest-Neighbor Distance

One of the simplest rotationally-invariant representations one could imagine is a set of nearest-neighbor distances; that is, for particle i,

$$[|\vec{r}_{i,1}|, |\vec{r}_{i,2}|, ..., |\vec{r}_{i,j}|]. \tag{3.6}$$

One way to fix permutation invariance in this description is to sort neighbors according to their distance from particle i. To create a scale-invariant description of structure, the distances can be expressed as ratios relative to the shortest distance $|\vec{r}_{i,1}|$. This type of description is appealing for its simplicity and a similar form has been used to study crystallization[169, 170]. Of course, this representation can be too simplistic for many tasks: for example, the twelve nearest neighbors of $cF4$-Cu and $hP2$-Mg are all at exactly the same distance from a central particle, yielding identical representations in this scheme. To get an idea of how these descriptors perform for various structures, we utilize UMAP dimensionality reduction on an 8-dimensional PCA projection, as in Section 3.5.1. We create spatially-averaged descriptors by computing the minimal spanning tree of the weighted graph of neighbors, where the edge between particles i and j is the distance between them. We then randomly select some number of bonds in this graph to clip and average the descriptors over each connected component of the graph. The resulting UMAP projection of the observations is shown in Figure 3.12(a). As expected, some pairs of structures—like $cF4$-Cu and $hP2$-Mg—are placed in the same cluster because they have similar or identical representations.

3.6.2 Bispectrum Coefficients

Another likely useful type of description for crystal structures is the bispectrum[171, 172]. The bispectrum can be thought of as the two-dimensional correlation function of spherical harmonic degree—in contrast to the Steinhardt order parameters, which are like a power spectrum for a given spherical harmonic degree. We again perform UMAP dimensionality reduction as in Section 3.5.1 and show the results in Figure 3.12(b). In contrast with the neighbor distance vectors, the bispectrum descriptors are able to distinguish between $cF4$-Cu and $hP2$-Mg due to the information encoded about angles between bonds of neighboring particles.

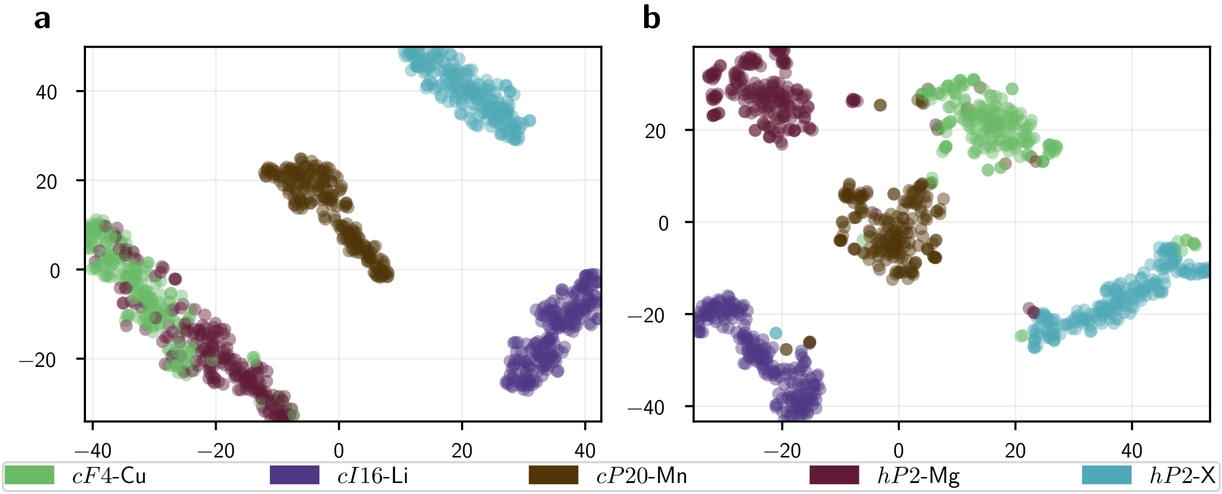


Figure 3.12: UMAP embeddings of five crystal structures. (a) Embedding of structures using normalized nearest-neighbor distance vectors. (b) Embedding of structures using bispectrum spherical harmonic fingerprints.

Chapter 4

Intelligently Designing Experiments using Machine Learning

As we have seen, colloidal and nanoscale particles have a wide variety of tunable properties, each of which may influence their assembly[14]. This abundance of tunable parameters often leads to simulation studies that probe the effects not only of external thermodynamic degrees of freedom such as temperature and pressure, but also internal aspects of the particles themselves like shape and stickiness. Increased dimensionality of studies plagues self-assembly simulations in at least two ways. First, as the dimensionality increases we expect the number of simulations that must be analyzed to increase significantly, generating large amounts of data and practically requiring automated analysis to be able to sort through, as in Chapter 3. Second, the typically-employed strategy of sampling points in a regular grid scales poorly as dimensionality increases. Often, the interesting regions of self-assembly phase diagrams occur in a few restricted spaces which punctuate larger swathes of more boring behavior. Placing points on a grid may cause a large fraction of the points to fall within these "boring" regions—a problem that only becomes worse as dimensionality increases further. Finally, it is difficult to extend a grid-based sampling strategy in a systematic way to account for the information learned in successive rounds of simulation—that is, there is no immediately clear best way to determine what the extents and number of points in each dimension should be. Methods for overcoming this problem have been suggested in the past[173], but have not reached widespread adoption. Intuitively, this problem is related to the idea in the machine learning literature of *active learning*[174, 175, 176, 177, 178]: given some set of choices and knowledge about previous choices and outcomes, active learning seeks to find the overall optimal strategy with the fewest number of choices. In this work, we formulate tools to intelligently explore the structural landscape of a high-dimensional simulation parameter space. This consists of two key parts: automatically characterizing structural observations and autonomously sampling parameter space in a systematic manner.

4.1 Characterizing Structural Observations

We use GMMs to estimate the probability density function (PDF) of our simulation results in the structure space formed by the local neighborhood spherical harmonic descriptors, much like in previous work[114]. To reduce the data to a more manageable dimensionality, we first project the observed spherical harmonic descriptors from 819 to 64 dimensions using Principal Component Analysis (PCA) [156]. We optimize the number of components of the GMM using the Bayesian Information Criterion [157], which measures the fit of the GMM to the observed data while penalizing GMMs with large numbers of free parameters. In this way, we create the best estimate of the PDF in structure space with few tunable parameters.

After we fit a GMM to our structural observations, we use it to compute how *interesting* simulation results are. For a set of structural descriptors x, we use $-\log(< p_{GMM}(x) >)$—where the average is taken over all observed descriptors for a given set of parameters—as an estimate of the novelty of a given structure. In this way, structures that are uncommon in the input data set will be assigned high novelty, while structures that are prevalent in a phase diagram will have low novelty. We note that the meaning of the individual Gaussian components does not matter, since we are only using the GMMs to compute a probability density function in the high-dimensional structure space; in other words, as long as we find a good fit for the density function, we do not care which observations are assigned to which Gaussian components. This is in contrast to a common use case of GMMs wherein distinct components are expected to correspond to different classes of observations.

We formulate an experiment emulating the results of a very simple phase diagram study. We generate a series of GMMs trained on two structures ($cI2$-W and $cF4$-Cu) and then gradually add examples of a third structure (the simple cubic structure type $cP1$-Po) to the set of training observations. We plot the *interest* of the newly introduced structure, as well as that of simple and more complex reference structures (here: the simple hexagonal structure type $hP1$-$Ca_{0.15}Sn_{0.85}$ and the clathrate-I structure type $cP54$-K_4Si_{23}, respectively) that are never given as training data. Figure 4.1 shows that, as more examples of $cP1$-Po are included in the training set, the *interest* of that structure rapidly decreases, while the two structures that are not given as observations remain *interesting*. Using the methods described in the next section, we would selectively bias new simulations toward regions forming these more interesting structures.

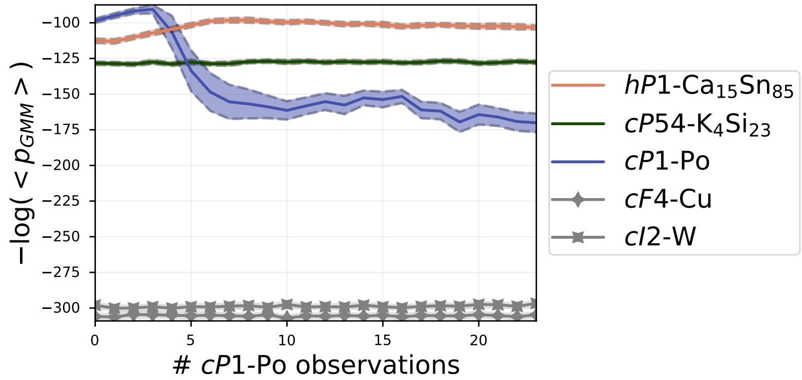


Figure 4.1: Interest of structures after training a GMM on many samples of $cF4$-Cu and $cI2$-W and varying numbers of samples of $cP1$-Po.

4.2 Optimization of Structural Interest for Phase Diagrams

Before applying parameter exploration to an unknown system, we apply it to easier-to-understand test systems of randomly-generated phase diagrams. We randomly select structures from the Crystallography Open Database [179, 180, 181, 182] (after filtering out organic compounds and duplicates) and assign each structure to one of 16–64 reference points inside an artificial parameter space to produce simpler and more complex phase diagrams. Reference points are generated in a parameter space of dimensionality D by iteratively randomly placing points within a randomly selected simplex ($D + 1$ points) from an initial set of size $2 \cdot D + 1$ to a final size of 16–64 points. This yields a clustered distribution of reference points and a broad distribution of volumes over phases in parameter space. We assign the structure of any given point in parameter space to be that of the reference point closest to it, corresponding to the Voronoi tessellation of parameter space given the set of random reference points.

We compare four different exploration algorithms to optimize the interest of the found structure. The first performs Gaussian Process (GP) optimization of expected improvement through the Python package GPyOpt [183]. Second, we use a stochastic method inspired by the LIPO algorithm [184], which we call Lipschitz optimism-pessimism (*LOP*). This samples an interpolated estimate for test points randomly between two extrema: an "optimistic" extreme that prioritizes exploration of unknown parameter space based on an estimate of the Lipschitz constant (the maximum rate of change of the structure interest with respect to the parameter values) and a "pessimistic" extreme that prioritizes exploitation of points in parameter space known to be good (using a constant, nearest-neighbor estimate). Third, maximum distance (*MD*) places new points as far as possible from any previous observations. The MD strategy can be thought of as a purely exploration-based approach. Fourth, random search (*rand*) explores new points uniformly and randomly throughout parameter space.

In Figure 4.2, we plot the standard deviation of the observed probabilities of each structure, as well as the number of structures found, as a function of the number of "simulations" that have been performed for each method. An ideal algorithm would perfectly distribute simulations among the structures found, causing the standard deviation value to rapidly approach 0. The ideal algorithm would also quickly find all structures, causing the number of structures found to immediately approach its final value. We find that GP optimization works well after exploring many points—at least for the structure type variance metric—but it remains stuck for a rather long initialization time sampling closely around a few points rather than exploring the whole input space available to it. This could be due to the objective function changing at every iteration as new GMMs are fit to the data and more information about the structures present in the phase diagram is uncovered. The LOP algorithm narrowly outperforms the pure exploration (MD and rand) algorithms, with an advantage that narrows as the dimensionality increases.

4.3 Model System: the Binary Oscillatory Pair Potential (OPP)

For this work we study self-assembly results from the binary Oscillatory Pair Potential (OPP), where the pairwise interaction energy between particles of type α and β is given as

$$U_{\alpha\beta}/\epsilon_0 = \frac{1}{r^{15}} + \frac{1}{r^3}\cos\left(k_{\alpha\beta}(r-1) + \phi_{\alpha\beta}\right). \tag{4.1}$$

The potential is truncated after the second attractive well, *i.e.*, at the second maximum if the value at the first local minimum is lower than the second local maximum, and otherwise at the third local maximum. Systems are slowly cooled from a thermalized initial configuration

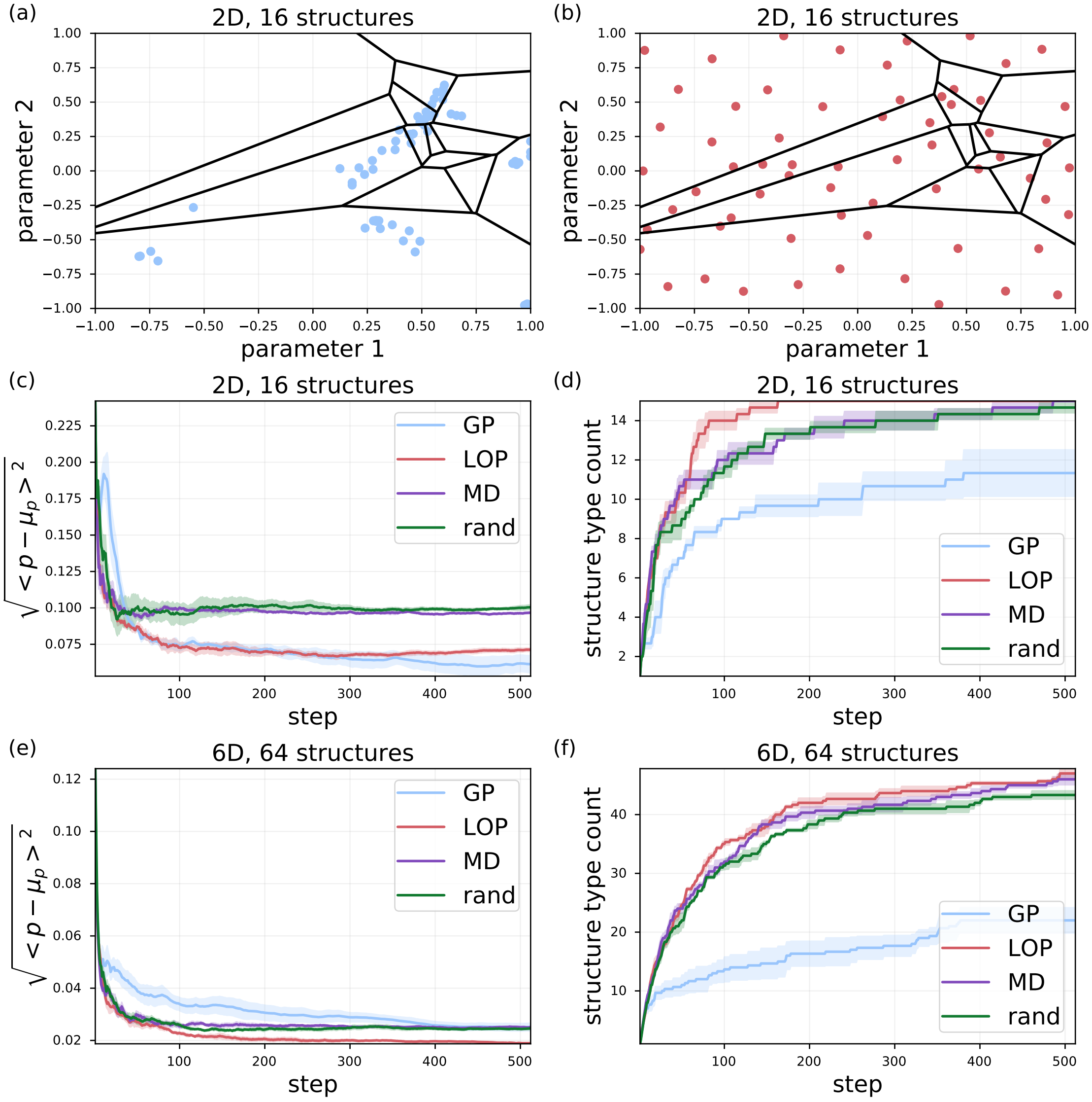


Figure 4.2: Parameter exploration results for randomly-generated phase diagrams with four different optimization methods: Gaussian process, Lipschitz optimism/pessimism, maximum distance, and random. All error bars indicate one standard error of the mean over three replicas. (a-b) Example exploration progress after 64 exploration steps for (a) GP and (b) LOP algorithms. Phase boundaries are indicated by lines and explored points by dots. (c-f) Structure fraction variance and number of structures found for each algorithm in (c-d) two-dimensional and (e-f) six-dimensional phase diagrams.

at $k_BT = 2\epsilon_0$ to $k_BT = 0.05\epsilon_0$ over $50,000\tau_0$, where $\tau_0 = \sqrt{\frac{m_0\sigma_0^2}{\epsilon_0}}$ is the basic time scale derived from the basic mass, distance, and energy scales.

For interactions between particles of type A and B, our parameter space consists of seven variables: $(k_{\alpha\beta}, \phi_{\alpha\beta})$ for the functional forms of the OPPs representing A–A, A–B, and B–B interactions, as well as another parameter controlling the stoichiometry of A and B. To be able to distinguish structures that differ only in the assignment of particle types, we use as descriptors for a particle of type α the concatenation of the neighbor-averaged spherical harmonics ignoring types and the neighbor-averaged spherical harmonics ignoring all particles of type $\beta \neq \alpha$.

4.3.1 Optimization Results

We use three of the optimization algorithms discussed above (*GP*, *LOP*, and *rand*) to simulate over 3,000 parameter points each and plot the number of unique multicomponent structures found. Structures were identified by an artificial neural network. The neural network was primarily trained on trained on replicated unit cells of ideal structures with varying Gaussian noise (from $0.05\sigma_0$ to $0.125\sigma_0$, where σ_0 is the unit length scale of the simulations). Single-component and binary structure types were included, as well as single-component structures with random binary assignment of types by setting a random selection of 25% and 50% of the particles to be of type B. For structures without a well-defined three-dimensional unit cell, the network was trained on selected examples from the simulation data. This procedure yielded approximately 1 million particle spherical harmonic vectors as training data. The simple feedforward neural network was trained to associate each particle to the crystal structure it came from, as in previous work[114]. The number of distinct structures found by each optimization algorithm is plotted in Figure 4.3. While random exploration still performs quite well compared to Gaussian process optimization, our LOP algorithm outperforms both methods by a narrow margin after an initialization period.

We identify the structures found during simulation for the development and tuning of these algorithms—over 35,000 simulations—using supervised and unsupervised learning [114] to find distinct structures. The actual structure identification—*i.e.*, their association with known atomic compounds—was conducted manually.

4.3.2 Previously Known Structures

Many crystal structures found *via* the binary OPP are known from materials on the atomic scale. This is not surprising, as the functional form of the potential is inspired by Friedel oscillations and was originally chosen to model the interactions in intermetallic compounds[185,

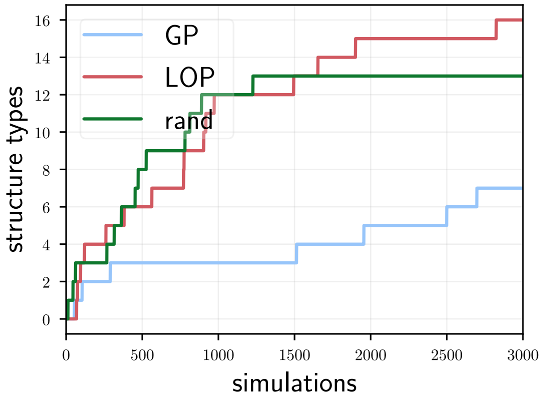


Figure 4.3: Exploration results for the binary system of OPPs, using three different optimization methods: Gaussian process, Lipschitz optimism-pessimism, and random.

186, 12]. The two-component structures that were found are listed in Tab. 4.1.

The structure types found range in complexity from the simplest binary crystals to those with dozens of particles in the unit cell. Snapshots of the unit cells are shown in Figure 4.4.

4.3.3 New Structures

We present four new periodic structure types—*i.e.*, ones that do not have atomic-scale equivalents—found by parameter exploration here. Some are variations or distortions of well-known structures. For example, the $oC8\text{-}A_4B_4$ structure has space group $Cmcm$ and is a simple orthorhombic, distorted version of the well-known $cF8$-NaCl (rock salt) structure. Similarly, the $tI4\text{-}A_2B_2$ structure has space group $I\bar{4}m2$ and is a distorted version of $tI4$-(Ag,Ga)Te. It can also be thought of as a two-component coloring of $cI16$-Si.

The $tP5\text{-AB}_4$ structure has a space group of $P4/mmm$ and can be seen as an AB stacking of layers of individual A particles on a square lattice with square motifs of four B particles. This structure and its BOOD are shown in Figure 4.5(a)–(b). The $cP40\text{-}A_{20}B_{20}$ structure has a significantly more complicated unit cell and has space group $P4_132$. All particles have tetrahedral coordination. A snapshot of the structure, as well as a BOOD, are shown in Figure 4.5(c)–(d).

We also find a two-component icosahedral quasicrystal in the binary OPP—to our knowledge, the first example in the literature of a self-assembled binary icosahedral quasicrystal structure. The quasicrystal has an average coordination of over 12 after removing particles on the surface of the crystallized droplet. This is distinct from the icosahedral quasicrystals

Table 4.1: List of binary OPP structures also found on the atomistic scale.

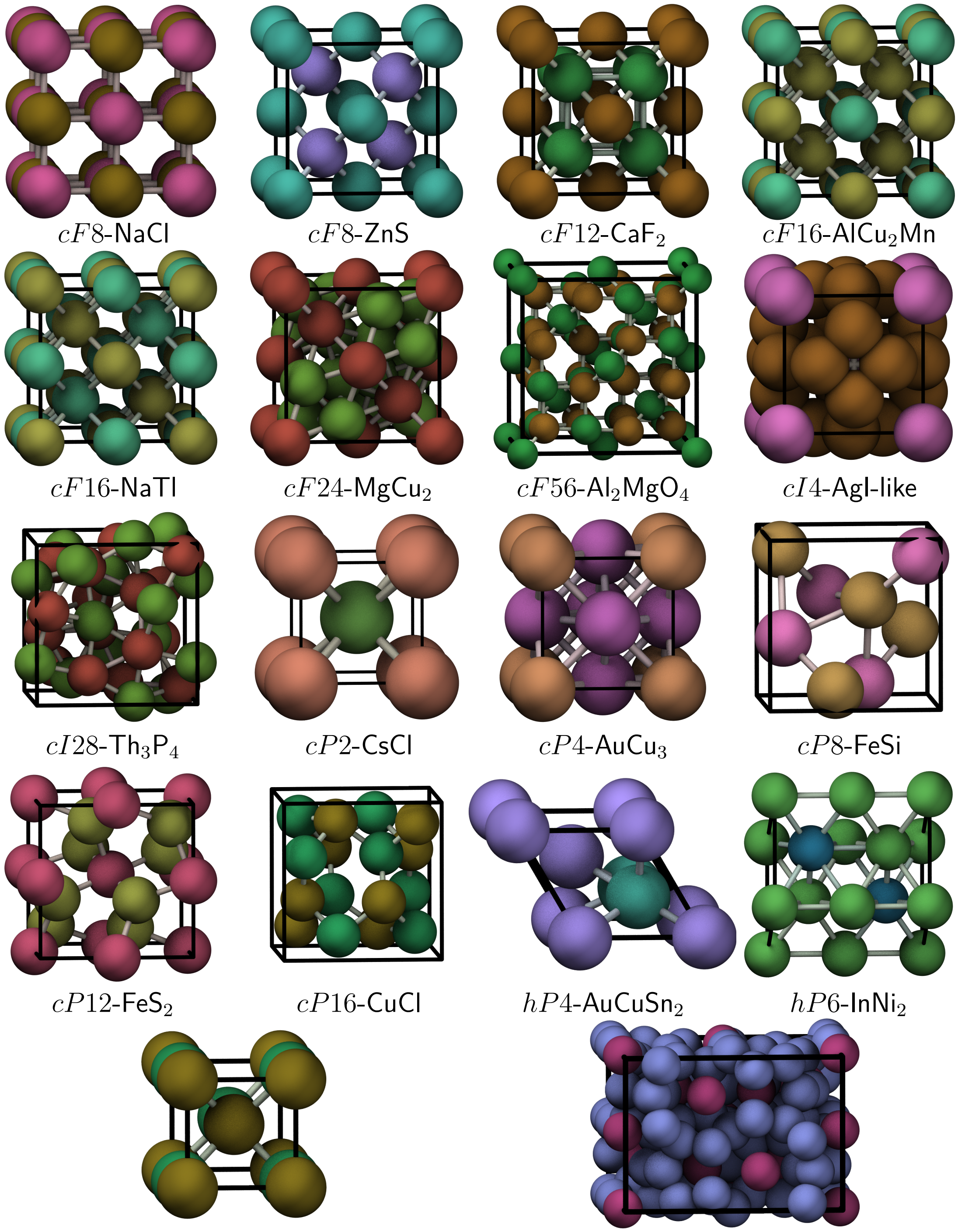
Structure name	**Common name/details**
$cF8$-NaCl	rock salt
$cF8$-ZnS	zinc blende / sphalerite
$cF12$-$\mathrm{CaF_2}$	fluorite
$cF16$-$\mathrm{AlCu_2Mn}$	Heusler phase
$cF16$-NaTl	
$cF24$-$\mathrm{MgCu_2}$	(Laves phase)
$cF56$-$\mathrm{Al_2MgO_4}$	spinel
$cI4$-AgI-like	(Ag sites at $\approx 1/8$ occupancy)
$cI28$-$\mathrm{Th_3P_4}$	
$cP4$-$\mathrm{AuCu_3}$	auricupride
$cP12$-$\mathrm{FeS_2}$	pyrite
$cP16$-CuCl	
$cP2$-CsCl	
$cP8$-FeSi	naquite
$hP4$-$\mathrm{AuCuSn_2}$	
$hP6$-$\mathrm{InNi_2}$	
$oS104$-$\mathrm{Rb_{12}Na_4Ge_{88}}$	
$tI4$-GeAs	

found previously using a similar but single-component interaction potential[12], which had average coordination numbers of roughly 4 and 7. An example assembled binary quasicrystal structure, BOOD, and coordination number histogram are shown in Figure 4.6.

4.4 Conclusion

High-dimensional models are ubiquitous in science and engineering; often one of the most alluring aspects of computational modeling is the ability to screen combinations of parameters much more quickly than would be feasible in physical experiments. Here we have shown that virtual experiments with many parameters can be improved further by intelligently incorporating optimization methods into the experimental design. We have coupled these optimization methods to GMMs, which model our observations in a high-dimensional structure space. Even though we ran only a few thousand simulations in this seven-dimensional parameter space—leaving a more exhaustive exploration of the system for a later study—we were able to find several structures that have not been reported before.

Moving human analysis out of the tight inner loop of the scientific process lets us perform



$cF8$-NaCl $cF8$-ZnS $cF12$-$\mathrm{CaF_2}$ $cF16$-$\mathrm{AlCu_2Mn}$

$cF16$-NaTl $cF24$-$\mathrm{MgCu_2}$ $cF56$-$\mathrm{Al_2MgO_4}$ $cI4$-AgI-like

$cI28$-$\mathrm{Th_3P_4}$ $cP2$-CsCl $cP4$-$\mathrm{AuCu_3}$ $cP8$-FeSi

$cP12$-$\mathrm{FeS_2}$ $cP16$-CuCl $hP4$-$\mathrm{AuCuSn_2}$ $hP6$-$\mathrm{InNi_2}$

$tI4$-GeAs $oS104$-$\mathrm{Rb_{12}Na_4Ge_{88}}$

Figure 4.4: Unit cells of structures found in the binary OPP system that are equivalent to structures found in nature, found *via* parameter exploration.

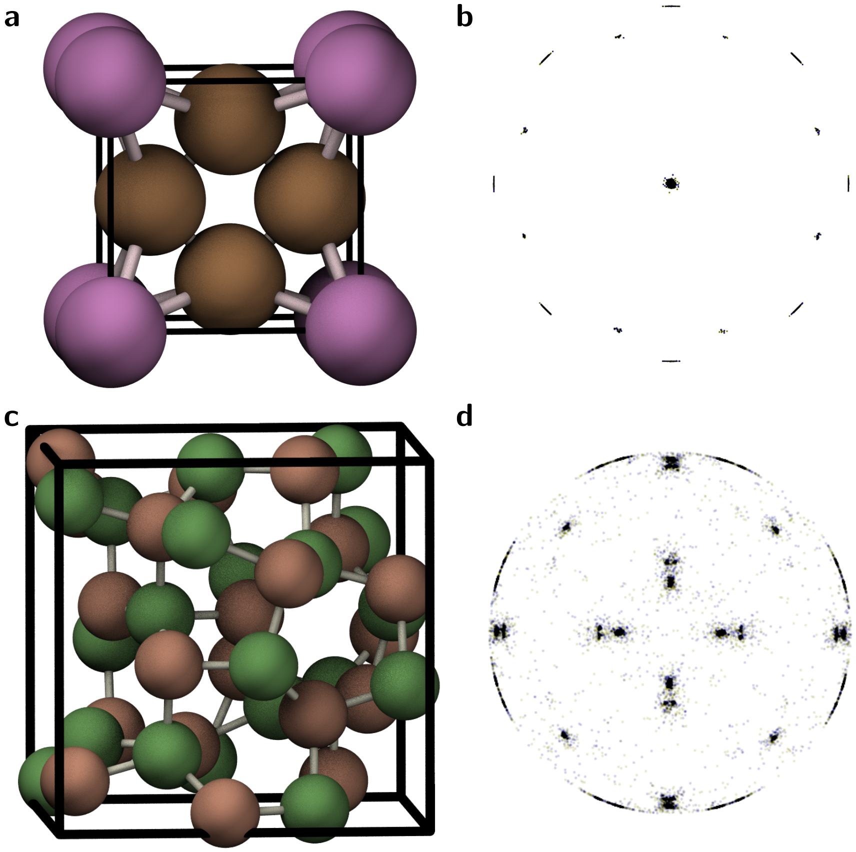


Figure 4.5: Above: snapshot (a) and bond-orientational order diagram (b) of the $tP5$-AB$_4$ structure. Below: snapshot (c) and bond-orientational order diagram (d) of the $cP40$-$A_{20}B_{20}$ structure.

simulations on much larger spaces and feasibly begin to explore systems with many parameters. As parallel processing power improves, we anticipate that fully automated workflows like the one presented here will become more prevalent, enabling more nuanced studies of more powerful models. In summary, we hope that methods like these will help make high-dimensional studies like this one become tractable for the scientific community at large.

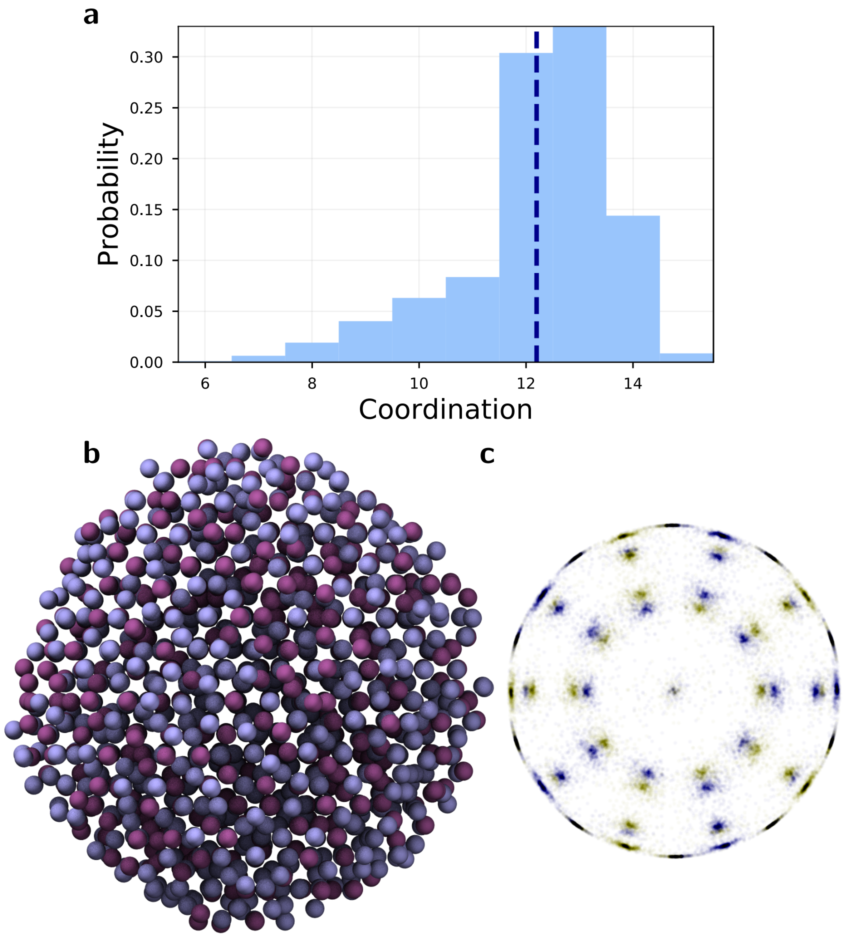


Figure 4.6: Dense icosahedral quasicrystal found using the binary oscillatory pair potential. (a) Histogram of coordination number of particles internal to the crystallized droplet. The mean coordination number of approximately 12.2 is indicated by a vertical dashed line. (b) Snapshot of a droplet, with particles colored by their type. (c) Bond-orientational order diagram (BOOD) displaying the icosahedral arrangement of local environments within the crystal.

Chapter 5

Outlook

Overall, I believe that this is an extremely exciting time to be involved in the field of molecular modeling. Within the last few decades, developments in algorithms, scientific methods, and high-performance hardware have driven great improvements in the quality of models we can reasonably simulate and the quantity of data we can generate. These developments, paired with the rise of data science as a paradigm, hold enormous potential for our field and practical applications in the areas of physics, chemistry, materials science, and biology. I believe that machine learning techniques will play a key role in allowing us to formulate and analyze higher-dimensional, more accurate models and I hope that my contributions have helped the field make some of the first of many steps toward this future.

In summary, we have seen several examples of complex structures and behavior originating from simple interactions. In some cases—either due to model complexity or simply the sheer number of simulations performed—it is easy for our experimental analysis and planning capacity to become overwhelmed. However, by utilizing machine learning methods—as shown in Chapter 3—we can create powerful and flexible automated analysis methods that can scale our analytical ability to be on par with our simulation capability. In Chapter 4, we show that we can integrate this type of analysis directly into the experimental design loop in order to autonomously study the structures possible in a parameter space.

In the future, I expect machine learning-based methods to become a staple in our field. As we have seen in the literature, various types of building blocks—be they hard particles with shape, particles interacting with isotropic pair potentials, or polymers—can self-assemble into the same set of structures. I think that machine learning could play a crucial role in identifying the similarities and differences in the assembly behavior of these systems *via* the insight it awards us on a microscopic, stochastic level. Using ML, I think we can begin to pick away at the problem of determining which attributes cause disparate building blocks to assemble into the same structure. ML has already exhibited great power in finding a structural basis for the glass transition[142, 141, 143], a problem that many expected to be difficult or impossible. Similarly, ML may be able to help us discern useful rules governing nonequilibrium self-assembly in other contexts, as in the case of the rotationally-driven particles presented in Chapter 2. In the longer term, I expect integrated analysis and expedited simulation—to intelligently record interesting behavior only when it is occurring and to dynamically guide simulations toward configurations of interest—to become a dominant paradigm in simulation. I think this evolution will be driven by more flexible programming languages and simulation frameworks that could support these types of workflows. I expect that ML-based models will play an ever-increasing role in these types of algorithms and look forward to the advancements in the field over the next several years.

Appendices

Appendix A

Discrete Element Method

Note: this chapter was originally published in the Journal of Computational Physics in 2017[187] and is reproduced with minor modifications below.

A.1 Introduction

The impact of particle shape on the self-assembly of systems of colloidal- and nanoscale particles is receiving ever-increasing attention.[16] Hard particle simulations are the most straightforward way to determine the impact of particle shape on assembly and have been highly successful in elucidating the phase behavior of anisotropic particles.[17, 188] Monte Carlo (MC) methods are ideal for probing the equilibrium behavior of such systems and can be implemented efficiently on modern highly parallel architectures.[189, 190] However, studying nonequilibrium behavior often requires dynamical, rather than stochastic, simulation methods. While hard-particle Monte Carlo methods like those used in the previous studies can accurately predict the equilibrium behavior of shapes, it is more difficult to directly study nonequilibrium behaviors like crystal nucleation[191] and active matter.[58, 29] Event-driven molecular dynamics[192, 193, 194, 195] (EDMD) is one such dynamical method, but it can be difficult to parallelize or to extend for arbitrary shapes and can slow down at the moderate to high densities of interest in many self-assembly studies. As such, there are currently no EDMD codes publicly available for performing dynamical simulations of particles with arbitrary shape.

Treating the issue of shape in MD is hardly a new concept; beyond the previously mentioned techniques, it is also common to see coarse-grained models of shapes constructed from stacked spheres. Such stacked sphere models consist rigid bodies of spheres in a more- or less-principled way to create particles with shape.[196, 197] However, this leaves particles with a rough surface which could introduce artifacts in assembly behavior when compared to the faceted particles often created in experiments. Particles simulated via the method described in this manuscript have smooth faces, yielding a higher-fidelity model of particle shape.

The Discrete Element Method (DEM) has been used extensively by the granular materials community to study dynamics of anisotropic, frictional particle systems.[198, 199, 200, 201, 202, 203, 204] This method models interactions between particles as interactions between the minimal set of lower-dimensional geometric features needed to capture the effects of particles' shapes. DEM is also a natural method to implement as a force field in a classical molecular dynamics (MD) framework, which lends itself to hardware acceleration. Programs using graphics processing units (GPUs) can achieve order-of-magnitude speedups over single-CPU programs, but only if they are made to take full advantage of the parallel nature of the GPU. Here we present an adaptation of DEM to run on GPUs within the HOOMD-Blue[85] MD framework. This sets our DEM implementation in a different class of algorithms than existing codes such as LIGGGHTS, where particle shape is constructed by stacking spheres. Furthermore, our implementation lacks the frictional contact force present in traditional DEM implementations by design - we have redesigned the algorithm to be appropriate for the kinds of forces present in colloidal and nanoscale materials, rather than those typical in granular matter. To our knowledge, our implementation is the first to bring DEM into the domain of classical MD. The method is intended purely for capturing steric repulsive forces between particles - such as those that have been ligated, charge-screened, or otherwise functionalized to only interact via short-ranged forces. Our method can be supplemented with additional colloidal or nanoscale forces typically used in coarse-grained MD particle simulations to account for intermediate or long-range forces like dipoles or charged surfaces.[205] This method is most directly applicable for people who would like to study the behavior of coarse-grained faceted colloidal- and nanoscale particles with higher shape fidelity than can be achieved with rigid bodies of spheres.

Finally, while it is tempting to draw comparisons between this and existing methodologies such as MC and EDMD, we have made no attempts at rigorous thermodynamic comparisons between various methods due to the inherent complexity of such comparisons. Here we have simply set forth the methodology, and leave it to the user to decide if DEM is the appropriate choice for their coarse-grained models. The implementation has been freely-available beginning with HOOMD-Blue v2.0, allowing them to easily and efficiently study

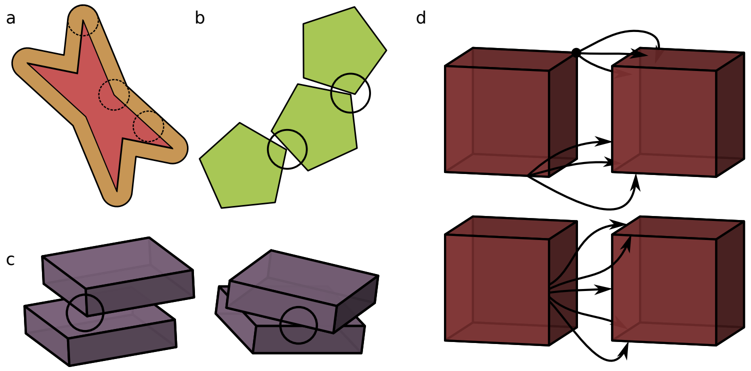


Figure A.1: The DEM model. (a) Shapes are represented by a polygon or polyhedron (inner shape), rounded out by a disk or sphere (dotted circles) to give the outer shape. The inner shape can be concave or convex. (b) In two dimensions, a repulsive contact force is evaluated between the nearest point on all pairs of vertices and edges. (c) In three dimensions, the contact force is evaluated between all pairs of vertices and faces and all pairs of edges and edges. (d) Because interactions are always evaluated between some features, some can be overcounted: when two aligned cubes come together, eight vertices will interact with three faces each, while four edges will interact with five neighboring edges and four other edges will interact with three neighboring edges each.

the equilibrium and non-equilibrium behavior of faceted particles.

A.2 Model

Some assumptions must be made about the simulated particles for the method described here to be useful in classical molecular dynamics. First, we assume that the polytopal shapes have been rounded by a sphere or disk of a given radius, as shown in Figure A.1(a). Rounding arises from the isotropic, conservative interaction we compute between interacting points and prevents discontinuities in the force arising from corners of shapes. The optimal rounding radius to use depends strongly on the geometry of the particles—like the presence of a shell of surface ligands around otherwise sharply faceted shapes—and the quantities that are being studied. For studies of nanoparticle and colloidal assembly, we later show that the small amount of rounding required by our DEM algorithm has no discernible effect on self-assembly behavior for rounding radii of more than 10% of the edge length, depending on the shapes involved and the phenomena under investigation. Second, we assume that particles interact only through short-range, repulsive forces. These assumptions allow us to approximate interactions between the volumes of particles i and j as interactions between the nearest points of lower-dimensional features (faces, edges, or vertices) of the particles using simple point-point interactions, without integrating over the surface or volume of the particles. In general, we consider a minimal set of interactions between features that will prevent two polyhedral shapes from overlapping as they collide. To prevent particle overlap in 2D, it is sufficient to consider interactions between all pairs of vertices and edges between the two particles, as shown in Figure A.1(b). In 3D, simulating polyhedral volumes requires calculating interactions between vertex-face pairs, and all edge-edge pairs as in Figure A.1(c).

Often in MD and MC simulations of anisotropic particles, a single "particle" is built from several spheres, joined together rigidly[196, 206]; interactions are calculated between all pairs of spheres, but translations and rotations are applied to the body as a whole[207]. Ideally, to create an anisotropic potential from a shape, we would integrate an isotropic potential over the surface or volume of a pair of shapes. Within DEM, particles are instead built up out of the geometric features of a two- or three-dimensional polytope: vertices (V), edges (E), and faces (F). The functions $V(\cdot)$, $F(\cdot)$, and $E(\cdot)$ yield the coordinates of the vertices, faces, and edges of their argument, respectively, and r^*_{ab} returns the distance between the nearest points of two features with types a and b. We then define the potential energy U_{ij} between particles i and j in three dimensions using a point particle potential $\mathcal{U}$ as

$$U_{ij}^{3D} = \sum_{\substack{E_i \in E(i) \\ E_j \in E(j)}} \mathcal{U}(r^*_{EE}(E_i, E_j)) + \sum_{\substack{V_i \in V(i) \\ F_j \in F(j)}} \mathcal{U}(r^*_{VF}(V_i, F_j)) + \sum_{\substack{V_j \in V(j) \\ F_i \in F(i)}} \mathcal{U}(r^*_{VF}(V_j, F_i)) \tag{A.1}$$

By considering vertex-face and edge-edge interactions, one or more potential interactions U_{ij}^{3D} will become nonzero as two polyhedra come into contact, as illustrated in Figure A.1. In two dimensions these features are reduced to checks between vertices and edges only:

$$U_{ij}^{2D} = \sum_{\substack{V_i \in V(i) \\ E_j \in E(j)}} \mathcal{U}(r^*_{VE}(V_i, E_j)) + \sum_{\substack{V_j \in V(j) \\ E_i \in E(i)}} \mathcal{U}(r^*_{VE}(V_j, E_i)) \tag{A.2}$$

Similarly to three dimensions, in two dimensions vertex-edge computations are sufficient to make one or more terms in U_{ij}^{2D} become positive as two polygons contact each other. The nearest points given by r^*_{ab} can be found using standard point-line, line-line, and point-plane formulae. Forces are computed using the derivative of this potential and torques are based on the interaction point on each particle.

Because we are simulating nanoscale and colloidal systems, we choose a conservative pair potential $\mathcal{U}$ that is representative of the interactions of such materials and well vetted within the community. A truncated and shifted version of the Lennard-Jones (LJ) potential, the Weeks-Chandler-Andersen (WCA) potential,[208] creates a steep, purely repulsive force from the particle surface with a rounding radius of $\frac{1}{2}\sigma_{ij}$:

$$U_{\mathrm{ij}}^{LJ}(r) = 4\epsilon_{\mathrm{ij}} \left[\left(\frac{\sigma_{\mathrm{ij}}}{r} \right)^{12} - \left(\frac{\sigma_{\mathrm{ij}}}{r} \right)^{6} \right] \tag{A.3}$$

$$U_{\mathrm{ij}}^{WCA}(r) = \begin{cases} U_{\mathrm{ij}}^{LJ}(r) - U_{\mathrm{ij}}^{LJ}(r_{\mathrm{cut}}^{WCA}) & r < r_{\mathrm{cut}}^{WCA} \\ 0 & r \geqslant r_{\mathrm{cut}}^{WCA} \end{cases} \tag{A.4}$$

where $r_{\mathrm{cut}}^{WCA} = 2^{\frac{1}{6}}\sigma_{ij}$.

When initializing particles on a lattice, e.g. prior to thermalization, one may encounter collisions of perfectly parallel edges when the lattice spacing is small. These collisions introduce a numerical instability for the molecular dynamics integrator: the points of interaction fluctuate at every timestep between the endpoints of each edge, yielding an unstable torque that changes sign at every timestep. To alleviate this issue while still only using point interactions, when two edges are sufficiently close to parallel the interaction point is taken to be the midpoint of overlap between the two edges.

Another common occurrence during simulation is for features to be "overcounted"; features that are shared among two separate faces (edges and/or vertices) are counted individually for polygonal face. This results from the fact that each polyhedron is decomposed into a set of polygonal faces (containing vertices and edges), and that the components of each face are subsequently checked against all other faces. A concrete example of this overcounting is shown in Figure A.1(d). If two cubes are touching perfectly face to face, they will have an interaction strength 66 times as large as a single vertex-face interaction: eight vertices are interacting with three faces each, four edges are interacting with five edges each, and four edges are interacting with three edges each. If the single vertex-vertex interaction had a strength of $1k_BT$, then the $1k_BT$ isosurface for the now $66k_BT$ interaction would have moved out by 10% of the rounding radius of the shape. This could lead to energetic "bumps" in the interaction: while the cutoff radius is not affected, the interaction is increased by a multiplicative factor according to the geometry of the two interacting sites, causing equipotential lines to expand slightly around vertices and edges. This effect should not matter to the extent that the potential used is a good approximation of a "hard" force field. Regardless of the geometry, we note that due to the cutoff in the WCA potential, it is impossible for the rounding radius to be increased by more than a factor of $2^{1/6} \approx 1.12$ with this overcounting effect.

A.3 Algorithm

The total force, torque, and potential energy for a given particle is the sum of the force, torque, and potential energy contributions between it and its neighbors. We evaluate these contributions from each particle's features independently by splitting the features among different GPU threads, then summing them efficiently in shared memory. In simulations constrained to only move in two dimensions (2D), two CUDA threads are assigned to each vertex of particle i, as shown in Figure A.2. The first thread assigned to a given vertex calculates and sums the force, torque, and potential energy contributions between that vertex and the nearest point to that vertex on each edge of each neighboring particle j. The second thread assigned to a given vertex calculates and sums the force, torque, and potential energy contributions between the nearest point on the edge beginning at that vertex (travelling counterclockwise) in particle i to each vertex in each neighboring particle j.

In fully three-dimensional (3D) simulations, two CUDA threads are assigned to each vertex of particle i and one thread is assigned to each edge of particle i, as shown in Figure A.2. The first vertex thread calculates the interaction between that vertex and the nearest point to each face in each neighboring particle j. The second vertex thread calculates the interaction between that vertex in each neighboring particle j and the nearest point of each face of particle i. The edge thread calculates the interaction between the nearest point on its edge of particle i to each edge of each neighboring particle j.

A.4 Results

A.4.1 Energy Conservation

To perform rigorous thermodynamic calculations using this method, we must first ensure that the NVE integrator conserves energy when combined with our force algorithm. We analyze both the short-term and long-term energy conservation of our model.[209] We use reduced units ε_0 for energy, σ_0 for length, m_0 for mass, and $\tau_0 = \sqrt{\frac{m_0 \sigma_0^2}{\varepsilon_0}}$ for time. We run NVE simulations of a fluid of squares (with edge length $4.24\sigma_0$) in 2D and tetrahedra (with

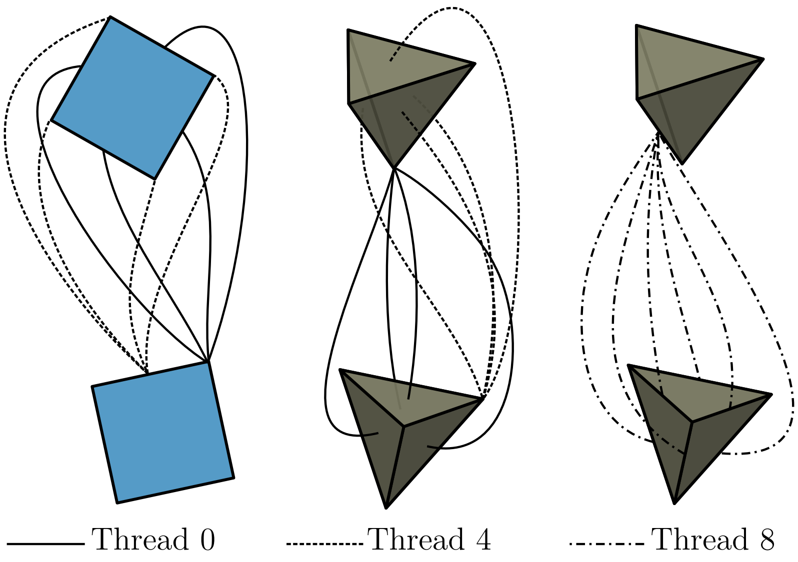


Figure A.2: Thread decomposition for the GPU. In 2D (left), each vertex is assigned a thread and each edge is assigned a thread; in 3D (center, right), each vertex is assigned two threads and each edge is assigned one thread.

edge length $8.49\sigma_0$) in 3D which have been thermalized at temperature $T^* = 1$ in reduced units.

To measure the short-term energy conservation, we compute the standard deviation of the total energy per particle $\sigma(E/N)$, recorded at a high frequency over a short NVE simulation of duration $10\tau_0$. We measure $\sigma(E/N)$ for systems using both single and double precision floating point arithmetic with a varying integration timestep size δt and present the results in Figure A.3.

For small δt, rounding errors saturate the accuracy of the method, causing the energy conservation to plateau. These rounding errors appear at a much smaller value of δt when using double- rather than single-precision floating point arithmetic. At large δt, error is introduced through the coarse time step as $\frac{\Delta E}{<E>} \sim \Delta t^2$ for the second-order integrator we use (a standard Velocity-Verlet algorithm with extensions for anisotropic particles[210], implemented as *md.integrate.nve* in HOOMD-Blue), increasing the energy deviation. For the remaining tests, which are performed in single precision only, we choose $\delta t = 0.01\tau_0$ as the timestep size to balance energy conservation and simulation speed.

To study the long-term energy conservation, we calculate the drift of the total energy per particle, $\frac{1}{N}\left(E(t) - E(0)\right)$, over long simulations. For squares, we achieve an energy drift of $1.36(6) \cdot 10^{-5}$ $\Delta E/(N/\varepsilon_0/\tau_0)$ and for tetrahedra we obtain $6.33(2) \cdot 10^{-4}$ $\Delta E/(N/\varepsilon_0/\tau_0)$.

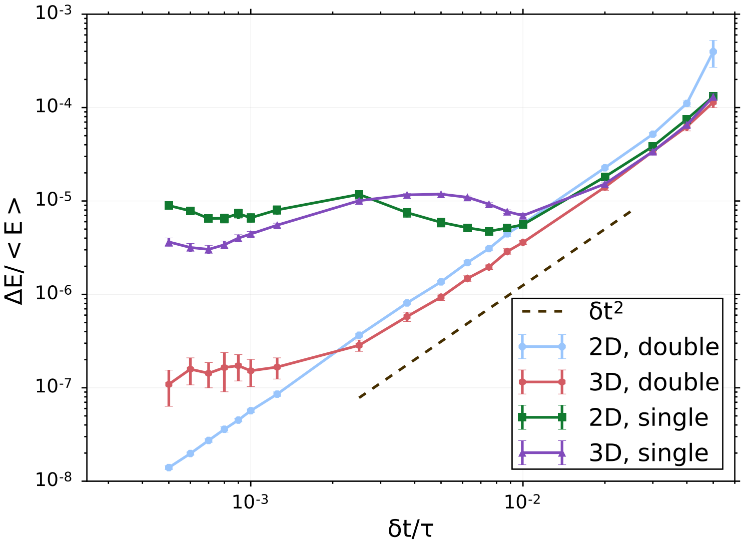


Figure A.3: Magnitude of total energy fluctuations for squares (2D) and tetrahedra (3D) in constant energy simulations for short times as a function of timestep size δt. Error bars indicate two standard errors of the mean for the energy fluctuations.

The energy drifts are significantly higher than those reported for isotropic particles[211], $1.3(1) \cdot 10^{-7}\ \Delta E/(N/\varepsilon_0/\tau_0)$ over the same period of $10\tau_0$, likely due to the new rotational degrees of freedom and the approximation of expressing the energy between features as the potential evaluated between their closest points. In practice, we find these energy drifts to be acceptable for the coarse-grained simulations at which this method is targeted.

A.4.2 Performance

We evaluate the speed benefit of our GPU parallelization scheme using an NVIDIA Quadro M6000 relative to both 18 cores and a single core of an Intel E5-2680V2 CPU for dense fluids of several systems with shapes of varying complexity in three dimensions: spheres, triangular plates, cubes, and icosahedra. The anisotropic shapes are modeled with the DEM potential, and spheres are modeled with a central WCA interaction, as in Equation A.4. As shown in Figure A.4, we achieve speedups of 15-75 times on the GPU and 10-27 times using domain decomposition, depending on the particle shape and system size. For many shapes the relative speed saturates at system sizes of a few thousand particles. Importantly, Figure A.4 also shows little to no speed up at small particle numbers on the GPU over the CPU for isotropic particles. This is due to the fact that threads are being under-utilized on the GPU for the isotropic particles; however, as can be seen in Figure A.4, the DEM algorithm saturates the GPU even for relatively modest numbers of particles. We note that the apparent jump in icosahedron performance for both GPU and parallel CPU implementations is due to a decrease of speed on the single-core CPU, likely due to memory locality effects, rather than an increase in GPU speed. Typical absolute performance numbers, in intensive units of particle-timesteps per second (PTPS)—that is, the number of MD timesteps completed per second multiplied by the number of particles in the system—are reported in Table A.1. In contrast, spheres are only just beginning to saturate the GPU at 65,000 particles. This finding demonstrates that the feature-based DEM parallelization scheme allows users to take advantage of GPU performance even for relatively small systems. Put another way, this method allows one to simulate the dynamics of a few thousand micron-sized particles for seconds of simulated time in a day on the GPU, rather than a week or more on a single CPU core.

Shape	N	$PTPS_{CPU}$	$PTPS_{GPU}$	$PTPS_{CPU,parallel}$
Sphere	256	$3.25(2) \cdot 10^6$	$3.3(3) \cdot 10^6$	N/A
Sphere	4096	$3.137(1) \cdot 10^6$	$4.64(7) \cdot 10^7$	$2.932(1) \cdot 10^7$
Sphere	65536	$3.07(2) \cdot 10^6$	$1.586(8) \cdot 10^8$	$3.901(5) \cdot 10^7$
Cube	256	$6.50(5) \cdot 10^3$	$1.85(4) \cdot 10^5$	N/A
Cube	4096	$6.49(7) \cdot 10^3$	$3.85(5) \cdot 10^5$	$7.81(8) \cdot 10^4$
Cube	65536	$5.48(9) \cdot 10^3$	$3.49(5) \cdot 10^5$	$9.04(1) \cdot 10^4$
Icosahedron	256	$1.83(2) \cdot 10^3$	$5.9(1) \cdot 10^4$	N/A
Icosahedron	4096	$1.76(1) \cdot 10^3$	$8.1(2) \cdot 10^4$	$2.21(3) \cdot 10^4$
Icosahedron	65536	$9.23(6) \cdot 10^2$	$6.88(9) \cdot 10^4$	$2.452(9) \cdot 10^4$

Table A.1: Absolute performance of CPU, GPU, and domain-decomposition parallel CPU MPI implementations, in particle-timesteps per second (PTPS). Numbers in parentheses show two standard errors of the mean in the least significant digit. Parallel CPU results are not available for small systems due to the domain decomposition scheme used[212].

A.4.3 Assembly

We simulate the self-assembly of shapes into crystals to compare our results to those of hard particle colloidal crystals obtained *via* MC simulations[17]. We simulate 1,728 square bipyramids, cubes, and icosahedra in the NPT ensemble after thermalization at a dilute gas over a range of pressures, as shown in Figure A.5. Mass and moments of inertia are calculated based on the volume of the Minkowski sum of the polyhedron and a sphere of rounding radius

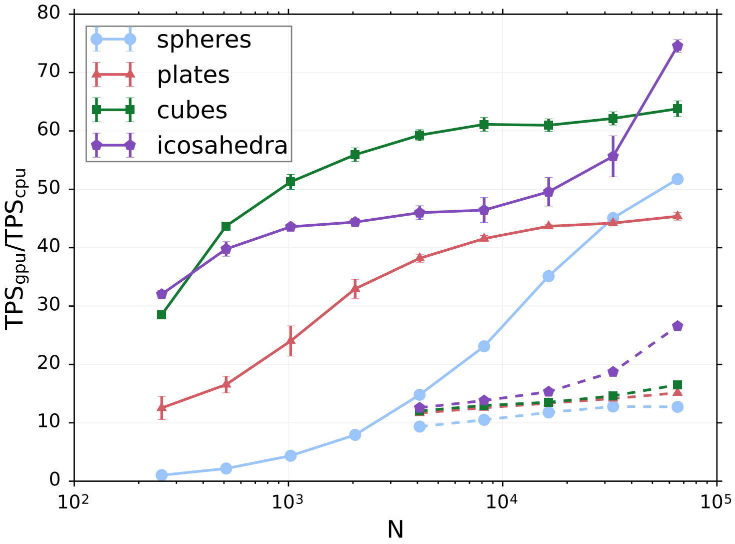


Figure A.4: Simulation speedup for different three-dimensional shapes by using a GPU (solid lines) and 18 CPU cores in parallel (dashed lines) relative to a single CPU core. Error bars indicate two standard errors of the mean.

$0.5\sigma_0$ and the inertia tensor of the inner polyhedron without rounding, respectively, using a density of $1\frac{m_0}{\sigma_0^3}$. We use a standard Martyna-Tobias-Klein NPT thermostat (available as *md.integrate.npt* in HOOMD-Blue) applied to both translational and rotational degrees of freedom[213, 214, 210] and the barostat computes the pressure from the virial tensor, just as in standard MD[215]. The time constants τ_T for the thermostat and τ_P for the barostat are each 1 τ_0 and the thermal energy for the thermostat is set to 1 $\frac{k_B T}{\varepsilon_0}$. We generate systems with a range of pressures and select the samples with the cleanest assembly behavior to show here. Numerical details of the simulations are presented in Table A.2. Similarly to MC results from hard particles with perfectly sharp corners, we find that the wide, flat bipyramids with a height of $\frac{1}{\sqrt{2}}$ relative to their equatorial edge length form a nematic phase, cubes form a simple cubic crystal, and icosahedra form a face-centered cubic crystal. Additionally, Figure A.5 contains particle bond-orientational order diagrams (BOODs) inset at right. The BOODs show a histogram of particle positions, at a fixed distance, averaged over every particle in the system and projected onto the surface of a sphere; the sphere is then “unrolled” into a map projection to show the entire diagram. The BOOD gives a quantitative image of the local orientation and environment of each particle relative to its neighbors. Additional information on BOODs, and their role in structure classification, can be found in the references.[9, 17] At the lower right, a cone represents preferred nearest-neighbor positions for surrounding

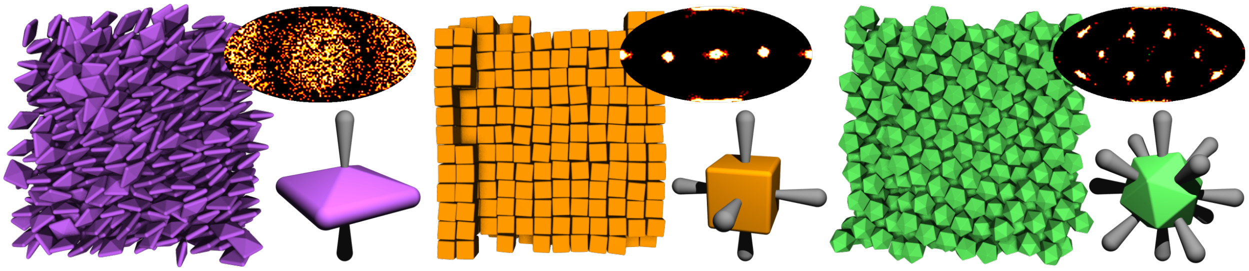


Figure A.5: Self-assembled crystal structures in systems of square bipyramids, cubes, and icosahedra. Bond-orientational order diagrams depicting a global histogram of neighboring particle positions are in the upper-right of each snapshot. Particles with cones indicating preferred nearest-neighbor positions are located to the lower-right of each snapshot.

particles, which was extracted from the BOOD.

Shape	r_{round}/r_{edge}	m/m_0	$I_{xx}/m_0\sigma_0^2$	$I_{yy}/m_0\sigma_0^2$	$I_{zz}/m_0\sigma_0^2$	$P/\frac{k_BT}{\sigma_0^3}$	t/tau_0
Bipyramid	$\frac{1}{6\sqrt{2}} \approx 0.118$	14.3	18.6	18.6	35.0	0.375	$3.6 \cdot 10^4$
Cube	$\frac{1}{6\sqrt{2}} \approx 0.118$	339	1296	1296	1296	0.1	$4.0 \cdot 10^4$
Icosahedron	$\frac{1}{12} \approx 0.0833$	644	4441	4441	4441	0.05	$6.3 \cdot 10^4$

Table A.2: Table of simulation parameters used to produce the hard particle assemblies shown in Figure A.5. Parameters include number of particles N, rounding radius compared to edge length r_{round}/r_{edge}, particle mass m, components of inertia tensor $I_{\alpha\beta}$, simulation pressure P, thermostat and barostat time constants τ_T and τ_P, and simulated time t in the NPT ensemble before creating the snapshots shown in the figure. The edge length used for the bipyramid is that of each of the four edges on its equator.

A.5 Summary

Here we have described a method to simulate conservative interactions of purely repulsive, rounded polytopes in molecular dynamics simulations. Because evaluation of this potential is more intensive than that of a point particle, we are able to more efficiently utilize graphics processing units for smaller system sizes with this potential than with isotropic potentials. Although the simulated shapes are still rounded, this method affords less opportunity for unphysical interdigitation than when rigid bodies of tangent or overlapping disks or spheres are used because the particle surface is smooth rather than rough.

In the future, there are many potentially useful optimizations that could be applied to this method. Currently, the contact point search and pair potential evaluation happen within the same GPU kernel; in three dimensions in particular, this leads to large register usage and low GPU occupancy. By splitting the contact search and force evaluation into two separate steps, some of this inefficiency could be avoided. In general, the search for contact points could also be improved through the use of shape-local cell lists[4], octrees[216], or other spatial data structures.

In summary, the DEM-MD method presented here enables dynamical simulations of faceted particles. With the particular integrators used, the method as implemented is most relevant to the simulation of nanoparticles and colloids in solvents where particle motion is dominated by inertia. Our implementation is available as of version 2.0 of HOOMD-Blue[1], along with Brownian and Langevin thermostats for modeling non-inertial regimes. Because the interactions are conservative, this method is useful for computation of thermodynamic quantities. Having real dynamical information enables a more direct mapping to studies of nonequilibrium processes such as crystallization, vitrification, jamming, and self-assembly of active matter than Monte Carlo simulations.

---

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

Bibliography

[1] N. Metropolis et al. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. ISSN: 00219606. DOI: `doi:10.1063/1.1699114`.

[2] B. J. Alder and T. E. Wainwright. “Phase Transition for a Hard Sphere System”. In: *The Journal of Chemical Physics* 27.5 (Nov. 1957), pp. 1208–1209. ISSN: 0021-9606, 1089-7690. DOI: `10.1063/1.1743957`.

[3] B. J. Alder and T. E. Wainwright. “Studies in Molecular Dynamics. I. General Method”. In: *The Journal of Chemical Physics* 31.2 (Aug. 1959), pp. 459–466. ISSN: 0021-9606. DOI: `10.1063/1.1730376`.

[4] L. Verlet. “Computer ”Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. In: *Physical Review* 159.1 (July 1967), pp. 98–103. DOI: `10.1103/PhysRev.159.98`.

[5] H. Heinz et al. “Accurate Simulation of Surfaces and Interfaces of Face-Centered Cubic Metals Using 12-6 and 9-6 Lennard-Jones Potentials”. In: *The Journal of Physical Chemistry C* 112.44 (Nov. 2008), pp. 17281–17290. ISSN: 1932-7447. DOI: `10.1021/jp801931d`.

[6] M. Karplus and G. A. Petsko. “Molecular Dynamics Simulations in Biology”. en. In: *Nature* 347.6294 (Oct. 1990), pp. 631–639. ISSN: 1476-4687. DOI: `10.1038/347631a0`.

[7] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett. “Communication: Inverse Design for Self-Assembly via on-the-Fly Optimization”. In: *The Journal of Chemical Physics* 145.11 (Sept. 2016), p. 111101. ISSN: 0021-9606. DOI: `10.1063/1.4962754`.

[8] C. S. Adorf et al. “Inverse Design of Simple Pair Potentials for the Self-Assembly of Complex Structures”. In: *arXiv:1711.04153 [cond-mat]* (Nov. 2017). arXiv: `1711.04153 [cond-mat]`.

[9] J. Roth and A. R. Denton. “Solid-Phase Structures of the Dzugutov Pair Potential”. In: *Physical Review E* 61.6 (June 2000), pp. 6845–6857. DOI: `10.1103/PhysRevE.61.6845`.

[10] M. Engel and H.-R. Trebin. “Self-Assembly of Monatomic Complex Crystals and Quasicrystals with a Double-Well Interaction Potential”. In: *Physical Review Letters* 98.22 (June 2007), p. 225505. DOI: `10.1103/PhysRevLett.98.225505`.

[11] C. L. Phillips and G. A. Voth. "Discovering Crystals Using Shape Matching and Machine Learning". In: *Soft Matter* 9.35 (2013), pp. 8552–8568. DOI: `10.1039/C3SM51449H`.

[12] M. Engel et al. "Computational Self-Assembly of a One-Component Icosahedral Quasicrystal". In: *Nature Materials* 14.1 (2015), pp. 109–116. DOI: `10.1038/nmat4152`.

[13] P. F. Damasceno, S. C. Glotzer, and M. Engel. "Non-Close-Packed Three-Dimensional Quasicrystals". en. In: *Journal of Physics: Condensed Matter* 29.23 (2017), p. 234005. ISSN: 0953-8984. DOI: `10.1088/1361-648X/aa6cc1`.

[14] S. C. Glotzer and M. J. Solomon. "Anisotropy of Building Blocks and Their Assembly into Complex Structures". In: *Nature Materials* 6.7 (Aug. 2007), pp. 557–562. ISSN: 1476-1122, 1476-4660. DOI: `10.1038/nmat1949`.

[15] S. Jiang et al. "Janus Particle Synthesis and Assembly". In: *Advanced Materials* 22.10 (2010), pp. 1060–1071. ISSN: 09359648. DOI: `10.1002/adma.200904094`.

[16] J. Henzie et al. "Self-Assembly of Uniform Polyhedral Silver Nanocrystals into Densest Packings and Exotic Superlattices". In: *Nature Materials* 11.2 (Feb. 2012), pp. 131–137. ISSN: 1476-1122. DOI: `10.1038/nmat3178`.

[17] P. F. Damasceno, M. Engel, and S. C. Glotzer. "Predictive Self-Assembly of Polyhedra into Complex Structures". In: *Science (New York, N.Y.)* 337.6093 (July 2012), pp. 453–457. ISSN: 1095-9203. DOI: `10.1126/science.1220869`.

[18] A. Haji-Akbari et al. "Disordered, Quasicrystalline and Crystalline Phases of Densely Packed Tetrahedra". en. In: *Nature* 462.7274 (Dec. 2009), pp. 773–777. ISSN: 0028-0836. DOI: `10.1038/nature08641`.

[19] D. Klotsa et al. "Intermediate Crystalline Structures of Colloids in Shape Space". en. In: *Soft Matter* (Aug. 2018). ISSN: 1744-6848. DOI: `10.1039/C8SM01573B`.

[20] R. K. Cersonsky et al. "Pressure-Tunable Photonic Band Gaps in an Entropic Colloidal Crystal". In: *arXiv:1803.06388 [cond-mat]* (Mar. 2018). arXiv: `1803.06388 [cond-mat]`.

[21] X. Ye et al. "Competition of Shape and Interaction Patchiness for Self-Assembling Nanoplates". en. In: *Nature Chemistry* 5.6 (June 2013), pp. 466–473. ISSN: 1755-4349. DOI: `10.1038/nchem.1651`.

[22] H. Lin et al. "Clathrate Colloidal Crystals". In: *Science* 355.6328 (Mar. 2017), pp. 931–935. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aal3919`.

[23] F. Lu et al. "Self-Assembly of DNA-Functionalized Cubes and Octahedra". In: (2019). in preparation.

[24] N. H. P. Nguyen et al. "Emergent Collective Phenomena in a Mixture of Hard Shapes through Active Rotation". In: *Physical Review Letters* 112.7 (Feb. 2014), p. 75701. ISSN: 00319007. DOI: `10.1103/PhysRevLett.112.075701`.

[25] G. Pilania et al. “Accelerating Materials Property Predictions Using Machine Learning”. en. In: *Scientific Reports* 3 (Sept. 2013), p. 2810. ISSN: 2045-2322. DOI: `10.1038/srep02810`.

[26] K. T. Schütt et al. “How to Represent Crystal Structures for Machine Learning: Towards Fast Prediction of Electronic Properties”. In: *Physical Review B* 89.20 (May 2014), p. 205118. DOI: `10.1103/PhysRevB.89.205118`.

[27] A. Kadurin et al. *druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico.* EN. https://pubs.acs.org/doi/abs/10.1021/acs.molpharmaceut.7b00346. Research-Article. Aug. 2017. DOI: `10.1021/acs.molpharmaceut.7b00346`.

[28] B. Sanchez-Lengeling and A. Aspuru-Guzik. “Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering”. en. In: *Science* 361.6400 (July 2018), pp. 360–365. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aat2663`.

[29] M. Spellings et al. “Shape Control and Compartmentalization in Active Colloidal Cells”. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.34 (Aug. 2015), E4642–50. ISSN: 1091-6490. DOI: `10.1073/pnas.1513361112`.

[30] T. Vicsek and A. Zafeiris. “Collective Motion”. In: *Physics Reports* 517.3-4 (Aug. 2012), pp. 71–140. ISSN: 03701573. DOI: `10.1016/j.physrep.2012.03.004`.

[31] M. C. Marchetti et al. “Hydrodynamics of Soft Active Matter”. In: *Reviews of Modern Physics* 85.3 (July 2013), pp. 1143–1189. ISSN: 00346861. DOI: `10.1103/RevModPhys.85.1143`.

[32] S. M. Khan et al. “Active Actin Gels”. In: *Communicative and Integrative Biology* 5.1 (2012), pp. 39–42. ISSN: 19420889. DOI: `10.4161/cib.5.1.18286`.

[33] S. Wang and P. G. Wolynes. “Active Contractility in Actomyosin Networks”. In: *Proceedings of the National Academy of Sciences* 109.17 (2012), pp. 6446–6451. DOI: `10.1073/pnas.1204205109`.

[34] T. Gao et al. “Multiscale Polar Theory of Microtubule and Motor-Protein Assemblies”. In: *Phys. Rev. Lett.* 114.4 (Jan. 2015), p. 048101. DOI: `10.1103/PhysRevLett.114.048101`.

[35] J. Tailleur and M. E. Cates. “Statistical Mechanics of Interacting Run-and-Tumble Bacteria”. In: *Physical Review Letters* 100.21 (2008), p. 218103. ISSN: 00319007. DOI: `10.1103/PhysRevLett.100.218103`.

[36] N. C. Damton et al. “Dynamics of Bacterial Swarming”. In: *Biophysical Journal* 98.10 (2010), pp. 2082–2090. ISSN: 00063495. DOI: `10.1016/j.bpj.2010.01.053`.

[37] H. H. Wensink et al. “Meso-Scale Turbulence in Living Fluids”. In: *Proceedings of the National Academy of Sciences* 109.36 (2012), pp. 14308–14313. ISSN: 0027-8424. DOI: `10.1073/pnas.1202032109`.

[38] J. R. Howse et al. “Self-Motile Colloidal Particles: From Directed Propulsion to Random Walk”. In: *Phys. Rev. Lett.* 99 (2007), p. 048102. ISSN: 00319007. DOI: `10.1103/PhysRevLett.99.048102`.

[39] J. Palacci et al. “Living Crystals of Light-Activated Colloidal Surfers”. In: *Science* 339.6122 (Mar. 2013), pp. 936–940. ISSN: 1095-9203. DOI: `10.1126/science.1230020`.

[40] T. Vicsek et al. “Novel Type of Phase Transition in a System of Self-Driven Particles”. In: *Physical Review Letters* 75.6 (Aug. 1995), pp. 1226–1229. DOI: `10.1103/PhysRevLett.75.1226`.

[41] C. András, H. E. Stanley, and V. Tamás. “Spontaneously Ordered Motion of Self-Propelled Particles”. In: *Journal of Physics A: Mathematical and General* 30.5 (1997), p. 1375.

[42] a. Czirok, M. Vicsek, and T. Vicsek. “Collective Motion of Organisms in Three Dimensions”. In: *Physica A: Statistical Mechanics and its Applications* 264.1-2 (1999), p. 3. ISSN: 03784371. DOI: `10.1016/S0378-4371(98)00468-3`.

[43] A. Cavagna and I. Giardina. “Bird Flocks as Condensed Matter”. In: *Annual Review of Condensed Matter Physics* 5.1 (Mar. 2014), pp. 183–207. ISSN: 1947-5454. DOI: `10.1146/annurev-conmatphys-031113-133834`.

[44] D. Helbing. “Traffic and Related Self-Driven Many-Particle Systems”. In: *Reviews of Modern Physics* 73.4 (Dec. 2001), pp. 1067–1141. DOI: `10.1103/RevModPhys.73.1067`.

[45] S. Ramaswamy, R. A. Simha, and J. Toner. “Active Nematics on a Substrate: Giant Number Fluctuations and Long-Time Tails”. In: *Europhysics Letters* 62.2 (Apr. 2002), pp. 196–202. ISSN: 0295-5075. DOI: `10.1209/epl/i2003-00346-7`.

[46] Y. Fily and M. C. Marchetti. “Athermal Phase Separation of Self-Propelled Particles with No Alignment”. In: *Physical Review Letters* 108.23 (2012), p. 235702. ISSN: 00319007. DOI: `10.1103/PhysRevLett.108.235702`.

[47] J. Bialké, T. Speck, and H. Löwen. “Crystallization in a Dense Suspension of Self-Propelled Particles”. In: *Physical Review Letters* 108.16 (Apr. 2012), p. 168301. ISSN: 00319007. DOI: `10.1103/PhysRevLett.108.168301`.

[48] G. S. Redner, M. F. Hagan, and A. Baskaran. “Structure and Dynamics of a Phase-Separating Active Colloidal Fluid”. In: *Physical Review Letters* 110.5 (Jan. 2013), p. 55701. ISSN: 00319007. DOI: `10.1103/PhysRevLett.110.055701`.

[49] J. Stenhammar et al. “A Continuum Theory of Phase Separation Kinetics for Active Brownian Particles”. In: *Physical Review Letters* 0.14 (2013), pp. 1–5. DOI: `10.1103/PhysRevLett.111.145702`.

[50] F. Peruani, A. Deutsch, and M. Bär. “Nonequilibrium Clustering of Self-Propelled Rods”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 74.3 (2006), p. 30904. ISSN: 15393755. DOI: `10.1103/PhysRevE.74.030904`.

[51] I. Buttinoni et al. "Dynamical Clustering and Phase Separation in Suspensions of Self-Propelled Colloidal Particles". In: *Physical Review Letters* 110.23 (June 2013), p. 238301. ISSN: 00319007. DOI: `10.1103/PhysRevLett.110.238301`.

[52] M. R. D'Orsogna et al. "Self-Propelled Particles with Soft-Core Interactions: Patterns, Stability, and Collapse". In: *Physical Review Letters* 96.10 (Mar. 2006), p. 104302. ISSN: 00319007. DOI: `10.1103/PhysRevLett.96.104302`.

[53] D. Grossman, I. S. Aranson, and E. Ben Jacob. "Emergence of Agent Swarm Migration and Vortex Formation through Inelastic Collisions". In: *New Journal of Physics* 10.2 (2008), p. 23036. ISSN: 13672630. DOI: `10.1088/1367-2630/10/2/023036`.

[54] N. H. P. Nguyen, E. Jankowski, and S. C. Glotzer. "Thermal and Athermal Three-Dimensional Swarms of Self-Propelled Particles". In: *Physical Review E* 86.1 (July 2012), p. 11136. ISSN: 1539-3755. DOI: `10.1103/PhysRevE.86.011136`.

[55] M. E. Cates and J. Tailleur. "When Are Active Brownian Particles and Run-and-Tumble Particles Equivalent? Consequences for Motility-Induced Phase Separation". In: *Europhysics Letters* 101.2 (2013), p. 20010. ISSN: 0295-5075. DOI: `10.1209/0295-5075/101/20010`.

[56] J. A. Cohen and R. Golestanian. "Emergent Cometlike Swarming of Optically Driven Thermally Active Colloids". In: *Phys. Rev. Lett.* 112.6 (Feb. 2014), p. 68302. DOI: `10.1103/PhysRevLett.112.068302`.

[57] R. Soto and R. Golestanian. "Self-Assembly of Catalytically Active Colloidal Molecules: Tailoring Activity through Surface Chemistry". In: *Phys. Rev. Lett.* 112.6 (Feb. 2014), p. 68301. DOI: `10.1103/PhysRevLett.112.068301`.

[58] H. H. Wensink et al. "Controlling Active Self-Assembly through Broken Particle-Shape Symmetry". In: *Physical Review E* 89.1 (Jan. 2014), p. 10302. ISSN: 15393755. DOI: `10.1103/PhysRevE.89.010302`.

[59] V. Rothschild. "Non-Random Distribution of Bull Spermatozoa in a Drop of Sperm Suspension". en. In: *Nature* 198.4886 (June 1963), pp. 1221–1222. ISSN: 1476-4687. DOI: `10.1038/1981221a0`.

[60] S. Chilukuri, C. H. Collins, and P. T. Underhill. "Impact of External Flow on the Dynamics of Swimming Microorganisms near Surfaces." In: *Journal of physics. Condensed matter : an Institute of Physics journal* 26.11 (Mar. 2014), p. 115101. ISSN: 1361-648X. DOI: `10.1088/0953-8984/26/11/115101`.

[61] J. Elgeti and G. Gompper. "Run-and-Tumble Dynamics of Self-Propelled Particles in Confinement". In: *EPL (Europhysics Letters)* 109.5 (2015), p. 58003.

[62] J. Hill et al. "Hydrodynamic Surface Interactions Enable Escherichia Coli to Seek Efficient Routes to Swim Upstream". In: *Physical Review Letters* 98.6 (Feb. 2007), p. 68101. ISSN: 00319007. DOI: `10.1103/PhysRevLett.98.068101`.

[63] H. Wioland et al. “Confinement Stabilizes a Bacterial Suspension into a Spiral Vortex”. In: *Physical Review Letters* 110.26 (June 2013), p. 268102. ISSN: 00319007. DOI: `10.1103/PhysRevLett.110.268102`.

[64] E. Lushi, H. Wioland, and R. E. Goldstein. “Fluid Flows Created by Swimming Bacteria Drive Self-Organization in Confined Suspensions”. In: *Proceedings of the National Academy of Sciences* (2014), p. 201405698.

[65] A. Cheng, H. Tsang, and E. Kanso. “Circularly Confined Microswimmers Exhibit Multiple Global Patterns”. In: *Physical Review E* 91.4 (Apr. 2015), p. 043008. DOI: `10.1103/PhysRevE.91.043008`.

[66] X. Yang, M. L. Manning, and M. C. Marchetti. “Aggregation and Segregation of Confined Active Particles”. In: *Soft Matter* 10.34 (2014), pp. 6477–6484. ISSN: 1744-6848. DOI: `10.1039/c4sm00927d`.

[67] Y. Fily, A. Baskaran, and M. F. Hagan. “Dynamics of Self-Propelled Particles under Strong Confinement”. In: *Soft Matter* 10.30 (2014), pp. 5609–5617. DOI: `10.1039/C4SM00975D`.

[68] J. Elgeti and G. Gompper. “Wall Accumulation of Self-Propelled Spheres”. In: *EPL (Europhysics Letters)* 101.4 (2013), p. 48003. ISSN: 0295-5075. DOI: `10.1209/0295-5075/101/48003`.

[69] J. P. Hernandez-Ortiz, C. G. Stoltz, and M. D. Graham. “Transport and Collective Dynamics in Suspensions of Confined Swimming Particles”. In: *Physical Review Letters* 95.20 (Nov. 2005), p. 204501. ISSN: 00319007. DOI: `10.1103/PhysRevLett.95.204501`.

[70] F. C. Keber et al. “Topology and Dynamics of Active Nematic Vesicles.” In: *Science (New York, N.Y.)* 345.6201 (Sept. 2014), pp. 1135–1139. ISSN: 1095-9203. DOI: `10.1126/science.1254784`.

[71] S. Ebbens et al. “Self-Assembled Autonomous Runners and Tumblers”. In: *Phys. Rev. E* 82 (2010), 015304(R). ISSN: 15393755. DOI: `10.1103/PhysRevE.82.015304`.

[72] B. A. Grzybowski, H. A. Stone, and G. M. Whitesides. “Dynamic Self-Assembly of Magnetized, Millimetre-Sized Objects Rotating at a Liquid-Air Interface”. In: *Nature* 405.6790 (2000), pp. 1033–1036. ISSN: 1476-4687. DOI: `10.1038/35016528`.

[73] J. Yan et al. “Linking Synchronization to Self-Assembly Using Magnetic Janus Colloids.” In: *Nature* 491.7425 (Nov. 2012), pp. 578–81. ISSN: 1476-4687. DOI: `10.1038/nature11619`.

[74] S. Martin et al. “Direct Observation of Hydrodynamic Rotation-Translation Coupling between Two Colloidal Spheres”. eng. In: *Physical Review Letters* 97.24 (Dec. 2006), p. 248301. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.97.248301`.

[75] N. Uchida and R. Golestanian. “Synchronization and Collective Dynamics in a Carpet of Microfluidic Rotors”. In: *Physical Review Letters* 104.17 (Apr. 2010), p. 178103. ISSN: 00319007. DOI: `10.1103/PhysRevLett.104.178103`.

[76] K. Yeo, E. Lushi, and P. M. Vlahovska. “Emergent Collective Dynamics of Hydrodynamically Coupled Micro-Rotors”. In: *arXiv preprint arXiv:1410.2878* (2014).

[77] J. Yan, S. C. Bae, and S. Granick. “Rotating Crystals of Magnetic Janus Colloids”. In: *Soft Matter* 11.1 (2015), pp. 147–153. ISSN: 1744-683X. DOI: `10.1039/C4SM01962H`.

[78] A. P. Petroff, X.-L. Wu, and A. Libchaber. “Fast-Moving Bacteria Self-Organize into Active Two-Dimensional Crystals of Rotating Cells”. In: *Phys. Rev. Lett.* 114.15 (Apr. 2015), p. 158102. DOI: `10.1103/PhysRevLett.114.158102`.

[79] A. J. Liu et al. “Opportunities in Theoretical and Computational Polymeric Materials and Soft Matter”. In: *Soft Matter* 11.12 (2015), pp. 2326–2332. ISSN: 1744-683X. DOI: `10.1039/C4SM02344G`.

[80] M. E. Cates et al. “Arrested Phase Separation in Reproducing Bacteria Creates a Generic Route to Pattern Formation.” In: *Proceedings of the National Academy of Sciences of the United States of America* 107.26 (2010), pp. 11715–11720. ISSN: 0027-8424. DOI: `10.1073/pnas.1001994107`.

[81] J. Schwarz-Linek et al. “Phase Separation and Rotor Self-Assembly in Active Particle Suspensions.” In: *Proceedings of the National Academy of Sciences of the United States of America* 109.11 (Mar. 2012), pp. 4052–7. ISSN: 1091-6490. DOI: `10.1073/pnas.1116334109`.

[82] J. I. Park et al. “Terminal Supraparticle Assemblies from Similarly Charged Protein Molecules and Nanoparticles.” In: *Nature communications* 5.May (2014), p. 3593. ISSN: 2041-1723. DOI: `10.1038/ncomms4593`.

[83] A. A. Shah et al. “Actuation of Shape-Memory Colloidal Fibres of Janus Ellipsoids”. en. In: *Nature Materials* 14.1 (Jan. 2015), pp. 117–124. ISSN: 1476-4660. DOI: `10.1038/nmat4111`.

[84] W. F. Paxton et al. “Catalytic Nanomotors: Autonomous Movement of Striped Nanorods”. In: *Journal of the American Chemical Society* 126.41 (2004), pp. 13424–13431. ISSN: 00027863. DOI: `10.1021/ja047697z`.

[85] J. A. Anderson, C. D. Lorenz, and A. Travesset. “General Purpose Molecular Dynamics Simulations Fully Implemented on Graphics Processing Units”. In: *Journal of Computational Physics* 227.10 (May 2008), pp. 5342–5359. ISSN: 0021-9991. DOI: `10.1016/j.jcp.2008.01.047`.

[86] T. D. Nguyen et al. “Rigid Body Constraints Realized in Massively-Parallel Molecular Dynamics on Graphics Processing Units”. In: *Comput. Phys. Commun.* 182.11 (Nov. 2011), pp. 2307–2313. ISSN: 00104655. DOI: `10.1016/j.cpc.2011.06.005`.

[87] J. A. Anderson and J. Glaser. *HOOMD-Blue*.

[88] J. D. Weeks and D. Chandler. “Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids”. In: *Journal of Chemical Physics* 54.12 (1971), p. 5237. ISSN: 00219606. DOI: `10.1063/1.1674820`.

[89] S. Sabrina et al. “Coarsening Dynamics of Binary Liquids with Active Rotation”. In: *Soft Matter* 11.43 (2015), pp. 8409–8416. ISSN: 1744-683X, 1744-6848. DOI: `10.1039/C5SM01753J`.

[90] T. Speck et al. “Effective Cahn-Hilliard Equation for the Phase Separation of Active Brownian Particles”. In: *Physical Review Letters* 112.21 (May 2014), p. 218304. ISSN: 10797114. DOI: `10.1103/PhysRevLett.112.218304`.

[91] R. Wittkowski et al. “Scalar Φ^4 Field Theory for Active-Particle Phase Separation”. In: *Nature Communications* 5 (2014). DOI: `doi:10.1038/ncomms5351`.

[92] D. Jasnow and J. Vinals. “Coarse-Grained Description of Thermo-Capillary Flow”. In: *Phys. Fluids* 8.3 (1996), pp. 660–669. ISSN: 10706631. DOI: `10.1063/1.868851`.

[93] D. Anderson, G. McFadden, and a. a Wheeler. “Diffuse-Interface Methods in Fluid Mechanics”. In: *Annu. Rev. Fluid Mech.* 30.1 (Jan. 1998), pp. 139–165. ISSN: 0066-4189. DOI: `10.1146/annurev.fluid.30.1.139`.

[94] R. E. Rosensweig. “Continuum Equations for Magnetic and Dielectric Fluids with Internal Rotations”. In: *Journal of Chemical Physics* 121.3 (July 2004), pp. 1228–1242. ISSN: 00219606. DOI: `10.1063/1.1755660`.

[95] R. García-Rojo, S. Luding, and J. J. Brey. “Transport Coefficients for Dense Hard-Disk Systems”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 74.6 (2006), pp. 1–11. ISSN: 15393755. DOI: `10.1103/PhysRevE.74.061305`.

[96] J.-F. Joanny and J. Prost. “Active Gels as a Description of the Actin-Myosin Cytoskeleton”. In: *HFSP J.* 3.2 (Jan. 2009), pp. 94–104. ISSN: 1955-2068. DOI: `10.2976/1.3054712`.

[97] Y. Sumino et al. “Large-Scale Vortex Lattice Emerging from Collectively Moving Microtubules”. In: *Nature* 483.7390 (2012), pp. 448–452. ISSN: 0028-0836. DOI: `10.1038/nature10874`.

[98] M. V. Sapozhnikov et al. “Dynamic Self-Assembly and Patterns in Electrostatically Driven Granular Media.” In: *Phys. Rev. Lett.* 90.11 (2003), p. 114301. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.90.114301`.

[99] V. Narayan, S. Ramaswamy, and N. Menon. “Long-Lived Giant Number Fluctuations in a Swarming Granular Nematic”. In: *Science* 317.5834 (2007), pp. 105–108. DOI: `10.1126/science.1140414`.

[100] M. Ibele, T. E. Mallouk, and A. Sen. “Schooling Behavior of Light-Powered Autonomous Micromotors in Water”. In: *Angew. Chem. Int. Ed.* 48.18 (Jan. 2009), pp. 3308–3312. ISSN: 1521-3773. DOI: `10.1002/anie.200804704`.

[101] D. Kagan, S. Balasubramanian, and J. Wang. “Chemically Triggered Swarming of Gold Microparticles”. In: *Angew. Chem. Int. Ed.* 50 (2011), pp. 503–506. ISSN: 14337851. DOI: `10.1002/anie.201005078`.

[102] I. Theurkauff et al. “Dynamic Clustering in Active Colloidal Suspensions with Chemical Signaling”. In: *Phys. Rev. Lett.* 108.26 (2012), p. 268303. DOI: `10.1103/PhysRevLett.108.268303`.

[103] A. Wysocki, R. G. Winkler, and G. Gompper. “Cooperative Motion of Active Brownian Spheres in Three-Dimensional Dense Suspensions”. In: *EPL (Europhysics Letters)* 105.4 (2014), p. 48004.

[104] Q.-X. Liu et al. “Phase Separation Explains a New Class of Self-Organized Spatial Patterns in Ecological Systems”. In: *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013), pp. 11905–11910. ISSN: 1091-6490. DOI: `10.1073/pnas.1222339110`.

[105] Y. Goto and H. Tanaka. “Purely Hydrodynamic Ordering of Rotating Disks at a Finite Reynolds Number”. en. In: *Nature Communications* 6 (Jan. 2015), p. 5994. ISSN: 2041-1723. DOI: `10.1038/ncomms6994`.

[106] K. Yeo, E. Lushi, and P. M. Vlahovska. “Collective Dynamics in a Binary Mixture of Hydrodynamically Coupled Microrotors”. In: *Physical Review Letters* 114.18 (May 2015), p. 188301. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.114.188301`.

[107] P. Tierno et al. “Controlled Swimming in Confined Fluids of Magnetically Actuated Colloidal Rotors”. In: *Physical Review Letters* 101.21 (Nov. 2008), p. 218304. ISSN: 00319007. DOI: `10.1103/PhysRevLett.101.218304`.

[108] A. Boymelgreen et al. “Spinning Janus Doublets Driven in Uniform AC Electric Fields”. In: *Phys. Rev. E* 89.1 (2014), p. 011003. ISSN: 15393755. DOI: `10.1103/PhysRevE.89.011003`.

[109] L. Qin et al. “Rational Design and Synthesis of Catalytically Driven Nanorotors”. In: *J. Am. Chem. Soc.* 129.48 (2007), pp. 14870–14871. ISSN: 00027863. DOI: `10.1021/ja0772391`.

[110] Y. Wang et al. “Dynamic Interactions between Fast Microscale Rotors”. In: *J. Am. Chem. Soc.* 131.29 (July 2009), pp. 9926–9927. ISSN: 1520-5126. DOI: `10.1021/ja904827j`.

[111] Z. Cheng, P. M. Chaikin, and T. G. Mason. “Light Streak Tracking of Optically Trapped Thin Microdisks”. In: *Phys. Rev. Lett.* 89.10 (2002), p. 108303. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.89.108303`.

[112] J. Zhu et al. “Coarsening Kinetics from a Variable-Mobility Cahn-Hilliard Equation: Application of a Semi-Implicit Fourier Spectral Method”. In: *Phys. Rev. E* 60.4 (Oct. 1999), pp. 3564–3572. ISSN: 1063-651X.

[113] A. J. Bray. “Theory of Phase-Ordering Kinetics”. In: *Adv. Phys.* 51.2 (2002), pp. 481–587.

[114] M. Spellings and S. C. Glotzer. “Machine Learning for Crystal Identification and Discovery”. In: *AIChE Journal* 64.6 (2018), pp. 2198–2206. DOI: `10.1002/aic.16157`.

[115] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105.

[116] D. Maturana and S. Scherer. “VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition”. In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Sept. 2015, pp. 922–928. DOI: `10.1109/IROS.2015.7353481`.

[117] Y. Zhou and O. Tuzel. “VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection”. In: *arXiv:1711.06396 [cs]* (Nov. 2017). arXiv: `1711.06396 [cs]`.

[118] R. Zhang, P. Isola, and A. A. Efros. “Colorful Image Colorization”. In: *arXiv:1603.08511 [cs]* (Mar. 2016). arXiv: `1603.08511 [cs]`.

[119] X. Jia, H. Chang, and T. Tuytelaars. “Super-Resolution with Deep Adaptive Image Resampling”. In: *arXiv:1712.06463 [cs]* (Dec. 2017). arXiv: `1712.06463 [cs]`.

[120] R. Zhang et al. “Real-Time User-Guided Image Colorization with Learned Deep Priors”. In: *arXiv:1705.02999 [cs]* (May 2017). arXiv: `1705.02999 [cs]`.

[121] H. S. Mousavi, T. Guo, and V. Monga. “Deep Image Super Resolution via Natural Image Priors”. In: *arXiv:1802.02721 [cs]* (Feb. 2018). arXiv: `1802.02721 [cs]`.

[122] L. A. Gatys, A. S. Ecker, and M. Bethge. “A Neural Algorithm of Artistic Style”. In: *arXiv:1508.06576 [cs, q-bio]* (Aug. 2015). arXiv: `1508.06576 [cs, q-bio]`.

[123] D. Ulyanov et al. “Texture Networks: Feed-Forward Synthesis of Textures and Stylized Images”. In: *arXiv:1603.03417 [cs]* (Mar. 2016). arXiv: `1603.03417 [cs]`.

[124] Y. Jing et al. “Neural Style Transfer: A Review”. In: *arXiv:1705.04058 [cs, eess, stat]* (May 2017). arXiv: `1705.04058 [cs, eess, stat]`.

[125] V. Mnih et al. “Playing Atari with Deep Reinforcement Learning”. In: *arXiv:1312.5602 [cs]* (Dec. 2013). arXiv: `1312.5602 [cs]`.

[126] D. Silver et al. “Mastering the Game of Go with Deep Neural Networks and Tree Search”. en. In: *Nature* 529.7587 (Jan. 2016), pp. 484–489. ISSN: 1476-4687. DOI: `10.1038/nature16961`.

[127] D. Silver et al. “Mastering the Game of Go without Human Knowledge”. en. In: *Nature* 550.7676 (Oct. 2017), p. 354. ISSN: 1476-4687. DOI: `10.1038/nature24270`.

[128] O. Vinyals et al. “Starcraft II: A New Challenge for Reinforcement Learning”. In: *arXiv:1708.04782 [cs]* (Aug. 2017). arXiv: `1708.04782 [cs]`.

[129] I. Goodfellow et al. “Generative Adversarial Networks”. In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680.

[130] D. Erhan et al. “Visualizing Higher-Layer Features of a Deep Network”. In: (2009).

[131] C. Olah, A. Mordvintsev, and L. Schubert. “Feature Visualization”. en. In: *Distill* 2.11 (Nov. 2017), e7. ISSN: 2476-0757. DOI: `10.23915/distill.00007`.

[132] T. Mikolov et al. “Efficient Estimation of Word Representations in Vector Space”. In: *arXiv:1301.3781 [cs]* (Jan. 2013). arXiv: `1301.3781 [cs]`.

[133] T. Mikolov et al. “Distributed Representations of Words and Phrases and Their Compositionality”. In: *Advances in Neural Information Processing Systems*. 2013, pp. 3111–3119.

[134] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti. “Bond-Orientational Order in Liquids and Glasses”. In: *Physical Review B* 28.2 (July 1983), pp. 784–805. DOI: `10.1103/PhysRevB.28.784`.

[135] U. Gasser et al. “Real-Space Imaging of Nucleation and Growth in Colloidal Crystallization”. en. In: *Science* 292.5515 (Apr. 2001), pp. 258–262. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1058457`.

[136] Z. Yan et al. “Structural Order for One-Scale and Two-Scale Potentials”. In: *Physical Review Letters* 95.13 (Sept. 2005), p. 130604. DOI: `10.1103/PhysRevLett.95.130604`.

[137] W. Lechner and C. Dellago. “Accurate Determination of Crystal Structures Based on Averaged Local Bond Order Parameters”. In: *The Journal of Chemical Physics* 129.11 (Sept. 2008), p. 114707. ISSN: 0021-9606, 1089-7690. DOI: `10.1063/1.2977970`.

[138] A. S. Keys, C. R. Iacovella, and S. C. Glotzer. “Characterizing Structure through Shape Matching and Applications to Self-Assembly”. In: *Annual Review of Condensed Matter Physics* 2.1 (2011), pp. 263–285. DOI: `10.1146/annurev-conmatphys-062910-140526`.

[139] A. S. Keys, C. R. Iacovella, and S. C. Glotzer. “Characterizing Complex Particle Morphologies through Shape Matching: Descriptors, Applications, and Algorithms”. In: *Journal of Computational Physics* 230.17 (July 2011), pp. 6438–6463. ISSN: 0021-9991. DOI: `10.1016/j.jcp.2011.04.017`.

[140] E. D. Cubuk et al. “Identifying Structural Flow Defects in Disordered Solids Using Machine-Learning Methods”. In: *Physical Review Letters* 114.10 (Mar. 2015), p. 108001. DOI: `10.1103/PhysRevLett.114.108001`.

[141] S. S. Schoenholz et al. “A Structural Approach to Relaxation in Glassy Liquids”. en. In: *Nature Physics* 12.5 (May 2016), pp. 469–471. ISSN: 1745-2481. DOI: `10.1038/nphys3644`.

[142] E. D. Cubuk et al. “Structural Properties of Defects in Glassy Liquids”. In: *The Journal of Physical Chemistry B* 120.26 (July 2016), pp. 6139–6146. ISSN: 1520-6106. DOI: `10.1021/acs.jpcb.6b02144`.

[143] S. S. Schoenholz et al. “Relationship between Local Structure and Relaxation in Out-of-Equilibrium Glassy Systems”. en. In: *Proceedings of the National Academy of Sciences* 114.2 (Oct. 2017), pp. 263–267. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.1610204114`.

[144] A. L. Ferguson et al. "Nonlinear Dimensionality Reduction in Molecular Simulation: The Diffusion Map Approach". In: *Chemical Physics Letters* 509.1–3 (June 2011), pp. 1–11. ISSN: 0009-2614. DOI: `10.1016/j.cplett.2011.04.066`.

[145] A. W. Long and A. L. Ferguson. "Nonlinear Machine Learning of Patchy Colloid Self-Assembly Pathways and Mechanisms". In: *The Journal of Physical Chemistry B* 118.15 (Apr. 2014), pp. 4228–4244. ISSN: 1520-6106. DOI: `10.1021/jp500350b`.

[146] W. F. Reinhart et al. "Machine Learning for Autonomous Crystal Structure Identification". en. In: *Soft Matter* (June 2017). ISSN: 1744-6848. DOI: `10.1039/C7SM00957G`.

[147] N. E. R. Zimmermann et al. "Assessing Local Structure Motifs Using Order Parameters for Motif Recognition, Interstitial Identification, and Diffusion Path Characterization". English. In: *Frontiers in Materials* 4 (2017). ISSN: 2296-8016. DOI: `10.3389/fmats.2017.00034`.

[148] S. Ravanbakhsh, J. Schneider, and B. Poczos. "Deep Learning with Sets and Point Clouds". In: (Nov. 2016).

[149] N. Thomas et al. "Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds". In: *arXiv:1802.08219 [cs]* (Feb. 2018). arXiv: `1802.08219 [cs]`.

[150] T. S. Cohen et al. "Spherical CNNs". In: *arXiv:1801.10130 [cs, stat]* (Jan. 2018). arXiv: `1801.10130 [cs, stat]`.

[151] R. Kondor et al. "Covariant Compositional Networks for Learning Graphs". In: *arXiv:1801.02144 [cs]* (Jan. 2018). arXiv: `1801.02144 [cs]`.

[152] M. Dzugutov. "Formation of a Dodecagonal Quasicrystalline Phase in a Simple Monatomic Liquid". In: *Physical Review Letters* 70 (1993), pp. 2924–2927. DOI: `10.1103/PhysRevLett.70.2924`.

[153] J. W. Roth, R. Schilling, and H.-R. Trebin. "Nucleation of Quasicrystals by Rapid Cooling of a Binary Melt: A Molecular-Dynamics Study". In: *Physical Review B* 51.22 (June 1995), pp. 15833–15840. DOI: `10.1103/PhysRevB.51.15833`.

[154] E. S. Harper et al. *Freud*. Zenodo. Nov. 2016. DOI: `10.5281/zenodo.166564`.

[155] F. Pedregosa et al. "Scikit-Learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[156] K. Pearson. "On Lines and Planes of Closest Fit to Systems of Points in Space". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), pp. 559–572.

[157] G. Schwarz. "Estimating the Dimension of a Model". EN. In: *The Annals of Statistics* 6.2 (Mar. 1978), pp. 461–464. ISSN: 0090-5364, 2168-8966. DOI: `10.1214/aos/1176344136`.

[158] J.-P. Baudry et al. "Combining Mixture Components for Clustering". en. In: *Journal of Computational and Graphical Statistics* 19.2 (Jan. 2010), pp. 332–353. ISSN: 1061-8600, 1537-2715. DOI: `10.1198/jcgs.2010.08111`.

[159] C. Hennig. “Methods for Merging Gaussian Mixture Components”. en. In: *Advances in Data Analysis and Classification* 4.1 (Apr. 2010), pp. 3–34. ISSN: 1862-5347, 1862-5355. DOI: `10.1007/s11634-010-0058-3`.

[160] F. Chollet. *Keras*. GitHub, 2015.

[161] R. M. Neal. “Bayesian Learning via Stochastic Dynamics”. In: *Advances in Neural Information Processing Systems 5*. Ed. by S. J. Hanson, J. D. Cowan, and C. L. Giles. Morgan-Kaufmann, 1993, pp. 475–482.

[162] L. van der Maaten and G. Hinton. “Visualizing Data Using T-SNE”. In: *Journal of Machine Learning Research* 9.Nov (2008), pp. 2579–2605. ISSN: ISSN 1533-7928.

[163] J. Tang et al. “Visualizing Large-Scale and High-Dimensional Data”. In: *Proceedings of the 25th International Conference on World Wide Web - WWW ’16* (2016), pp. 287–297. DOI: `10.1145/2872427.2883041`. arXiv: `1602.00370`.

[164] L. McInnes, J. Healy, and J. Melville. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. In: *arXiv:1802.03426 [cs, stat]* (Feb. 2018). arXiv: `1802.03426 [cs, stat]`.

[165] M. Wattenberg, F. Viégas, and I. Johnson. “How to Use T-SNE Effectively”. In: *Distill* (2016). DOI: `10.23915/distill.00002`.

[166] B. Roux. “The Calculation of the Potential of Mean Force Using Computer Simulations”. In: *Computer Physics Communications* 91.1-3 (1995), pp. 275–282. ISSN: 00104655. DOI: `10.1016/0010-4655(95)00053-I`.

[167] A. L. Ferguson. “BayesWHAM: A Bayesian Approach for Free Energy Estimation, Reweighting, and Uncertainty Quantification in the Weighted Histogram Analysis Method”. en. In: *Journal of Computational Chemistry* 38.18 (July 2017), pp. 1583–1605. ISSN: 1096-987X. DOI: `10.1002/jcc.24800`.

[168] C. Dellago et al. “Transition Path Sampling and the Calculation of Rate Constants”. In: *The Journal of Chemical Physics* 108.5 (Feb. 1998), pp. 1964–1977. ISSN: 0021-9606. DOI: `10.1063/1.475562`.

[169] R. B. Jadrich, B. A. Lindquist, and T. M. Truskett. “Unsupervised Machine Learning for Detection of Phase Transitions in Off-Lattice Systems I. Foundations”. In: *arXiv:1808.00084 [cond-mat, physics:physics]* (July 2018). arXiv: `1808.00084 [cond-mat, physics:physics]`.

[170] R. B. Jadrich et al. “Unsupervised Machine Learning for Detection of Phase Transitions in Off-Lattice Systems II. Applications”. In: *arXiv:1808.00083 [cond-mat, physics:physics]* (July 2018). arXiv: `1808.00083 [cond-mat, physics:physics]`.

[171] R. Kondor. “A Novel Set of Rotationally and Translationally Invariant Features for Images Based on the Non-Commutative Bispectrum”. In: *arXiv:cs/0701127* (Jan. 2007). arXiv: `cs/0701127`.

[172] R. Kakarala. “The Bispectrum as a Source of Phase-Sensitive Invariants for Fourier Descriptors: A Group-Theoretic Approach”. In: *arXiv:0902.0196 [math]* (Feb. 2009). arXiv: `0902.0196 [math]`.

[173] C. L. Phillips. “A Learning Heuristic for Space Mapping and Searching Self-Organizing Systems Using Adaptive Mesh Refinement”. In: *Journal of Computational Physics* 272 (Sept. 2014), pp. 799–813. ISSN: 0021-9991. DOI: `10.1016/j.jcp.2014.05.001`.

[174] N. Kutsuna et al. “Active Learning Framework with Iterative Clustering for Bioimage Classification”. In: *Nature Communications* 3 (Aug. 2012), p. 1032. ISSN: 2041-1723. DOI: `10.1038/ncomms2030`.

[175] D. Reker and G. Schneider. “Active-Learning Strategies in Computer-Assisted Drug Discovery”. In: *Drug Discovery Today* 20.4 (Apr. 2015), pp. 458–465. ISSN: 1359-6446. DOI: `10.1016/j.drudis.2014.12.004`.

[176] P. Bachman, A. Sordoni, and A. Trischler. “Learning Algorithms for Active Learning”. In: *arXiv:1708.00088 [cs]* (July 2017). arXiv: `1708.00088 [cs]`.

[177] E. V. Podryabinkin and A. V. Shapeev. “Active Learning of Linearly Parametrized Interatomic Potentials”. In: *Computational Materials Science* 140.Supplement C (Dec. 2017), pp. 171–180. ISSN: 0927-0256. DOI: `10.1016/j.commatsci.2017.08.031`.

[178] C. Dai, I. R. Bruss, and S. C. Glotzer. “Efficient Phase Diagram Sampling by Active Learning”. In: *arXiv:1803.03296 [cond-mat, physics:physics]* (Mar. 2018). arXiv: `1803.03296 [cond-mat, physics:physics]`.

[179] R. T. Downs and M. Hall-Wallace. “The American Mineralogist Crystal Structure Database”. In: *American Mineralogist* 88 (2003), pp. 247–250.

[180] S. Gražulis et al. “Crystallography Open Database – an Open-Access Collection of Crystal Structures”. In: *Journal of Applied Crystallography* 42.4 (Aug. 2009), pp. 726–729. DOI: `10.1107/S0021889809016690`.

[181] S. Gražulis et al. “Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration”. In: *Nucleic Acids Research* 40.D1 (2012), pp. D420–D427. DOI: `10.1093/nar/gkr900`.

[182] A. Merkys et al. “COD::CIF::Parser: An Error-Correcting CIF Parser for the Perl Language”. In: *Journal of Applied Crystallography* 49.1 (Feb. 2016). DOI: `10.1107/S1600576715022396`.

[183] The GPyOpt authors. *GPyOpt: A Bayesian Optimization Framework in Python.* 2016.

[184] C. Malherbe and N. Vayatis. “Global Optimization of Lipschitz Functions”. In: *arXiv:1703.02628 [stat]* (Mar. 2017). arXiv: `1703.02628 [stat]`.

[185] R. Egger and H. Grabert. “Friedel Oscillations for Interacting Fermions in One Dimension”. en. In: *Physical Review Letters* 75.19 (Nov. 1995), pp. 3505–3508. ISSN: 0031-9007, 1079-7114. DOI: `10.1103/PhysRevLett.75.3505`.

[186] M. Mihalkovič and C. L. Henley. "Empirical Oscillating Potentials for Alloys from Ab Initio Fits and the Prediction of Quasicrystal-Related Structures in the Al-Cu-Sc System". In: *Physical Review B* 85.9 (Mar. 2012), p. 092102. DOI: `10.1103/PhysRevB.85.092102`.

[187] M. Spellings et al. "GPU Accelerated Discrete Element Method (DEM) Molecular Dynamics for Conservative, Faceted Particle Simulations". In: *Journal of Computational Physics* (2017). ISSN: 0021-9991. DOI: `10.1016/j.jcp.2017.01.014`.

[188] K. L. Young et al. "A Directional Entropic Force Approach to Assemble Anisotropic Nanoparticles into Superlattices". In: *Angewandte Chemie International Edition* (Nov. 2013), pp. 13980–13984. ISSN: 14337851. DOI: `10.1002/anie.201306009`.

[189] J. A. Anderson et al. "Massively Parallel Monte Carlo for Many-Particle Simulations on GPUs". In: *Journal of Computational Physics* 254 (2013), pp. 27–38. ISSN: 0021-9991. DOI: `10.1016/j.jcp.2013.07.023`.

[190] J. A. Anderson, M. E. Irrgang, and S. C. Glotzer. "Scalable Metropolis Monte Carlo for Simulation of Hard Shapes". In: *Computer Physics Communications* 204 (2016), pp. 21–30. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2016.02.024`.

[191] R. Ni et al. "Glassy Dynamics, Spinodal Fluctuations, and the Kinetic Limit of Nucleation in Suspensions of Colloidal Hard Rods". In: *Physical Review Letters* 105.8 (Aug. 2010), p. 088302. ISSN: 1079-7114. DOI: `10.1103/PhysRevLett.105.088302`.

[192] M. N. Bannerman, R. Sargant, and L. Lue. "DynamO: A Free O(N) General Event-Driven Molecular Dynamics Simulator". In: *Journal of Computational Chemistry* 32.15 (Nov. 2011), pp. 3329–38. ISSN: 1096-987X. DOI: `10.1002/jcc.21915`.

[193] M. Marín. "Event-Driven Hard-Particle Molecular Dynamics Using Bulk-Synchronous Parallelism". In: *Computer Physics Communications* 102.1–3 (1997), pp. 81–96. ISSN: 0010-4655. DOI: `10.1016/S0010-4655(97)00011-8`.

[194] S. Miller and S. Luding. "Event-Driven Molecular Dynamics in Parallel". In: *Journal of Computational Physics* 193.1 (Jan. 2004), pp. 306–316. ISSN: 00219991. DOI: `10.1016/j.jcp.2003.08.009`.

[195] F. Smallenburg et al. "Vacancy-Stabilized Crystalline Order in Hard Cubes". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.44 (Oct. 2012), pp. 17886–90. ISSN: 1091-6490. DOI: `10.1073/pnas.1211784109`.

[196] B. S. John, A. Stroock, and F. A. Escobedo. "Cubatic Liquid-Crystalline Behavior in a System of Hard Cuboids". In: *The Journal of Chemical Physics* 120.19 (2004), pp. 9383–9389. DOI: `http://dx.doi.org/10.1063/1.1711594`.

[197] C. L. Phillips et al. "Optimal Filling of Shapes". In: *Physical Review Letters* 108.19 (May 2012), p. 198304. DOI: `10.1103/PhysRevLett.108.198304`.

[198] J. Ghaboussi and R. Barbosa. "Three-Dimensional Discrete Element Method for Granular Materials". In: *International Journal for Numerical and Analytical Methods in Geomechanics* 14.7 (1990), pp. 451–472. ISSN: 1096-9853. DOI: `10.1002/nag.1610140702`.

[199] S. Galindo-Torres et al. "Molecular Dynamics Simulation of Complex Particles in Three Dimensions and the Study of Friction Due to Nonconvexity". In: *Physical Review E* 79.6 (2009), p. 060301. DOI: `10.1103/PhysRevE.79.060301`.

[200] F. Alonso-Marroquín and Y. Wang. "An Efficient Algorithm for Granular Dynamics Simulations with Complex-Shaped Objects". In: *Granular Matter* 11.5 (July 2009), pp. 317–329. ISSN: 1434-5021. DOI: `10.1007/s10035-009-0139-1`.

[201] J. Wang et al. "Particle Shape Effects in Discrete Element Modelling of Cohesive Angular Particles". In: *Granular Matter* 13.1 (Oct. 2010), pp. 1–12. ISSN: 1434-5021. DOI: `10.1007/s10035-010-0217-4`.

[202] S. Mack et al. "Experimental Validation of Polyhedral Discrete Element Model". In: *Powder Technology* 214.3 (Dec. 2011), pp. 431–442. ISSN: 00325910. DOI: `10.1016/j.powtec.2011.08.043`.

[203] P. Langston, J. Ai, and H.-S. Yu. "Simple Shear in 3D DEM Polyhedral Particles and in a Simplified 2D Continuum Model". In: *Granular Matter* (May 2013), pp. 13–15. ISSN: 1434-5021. DOI: `10.1007/s10035-013-0421-0`.

[204] M. Boton et al. "Quasistatic Rheology and Microstructural Description of Sheared Granular Materials Composed of Platy Particles". In: *Physical Review E* 87.3 (Mar. 2013), p. 032206. DOI: `10.1103/PhysRevE.87.032206`.

[205] K. J. M. Bishop et al. "Nanoscale Forces and Their Uses in Self-Assembly". en. In: *Small* 5.14 (July 2009), pp. 1600–1630. ISSN: 1613-6829. DOI: `10.1002/smll.200900358`.

[206] T. D. Nguyen, E. Jankowski, and S. C. Glotzer. "Self-Assembly and Reconfigurability of Shape-Shifting Particles". In: *ACS Nano* 5.11 (2011), pp. 8892–8903. DOI: `10.1021/nn203067y`.

[207] A. Rahman and F. H. Stillinger. "Molecular Dynamics Study of Liquid Water". In: *The Journal of Chemical Physics* 55.7 (1971), pp. 3336–3359. DOI: `10.1063/1.1676585`.

[208] D. Chandler, J. Weeks, and H. Andersen. "Van Der Waals Picture of Liquids, Solids, and Phase Transformations". In: *Science* 220 (1983), pp. 787–794.

[209] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Oxford university press, 1989.

[210] H. Kamberaj, R. J. Low, and M. P. Neal. "Time Reversible and Symplectic Integrators for Molecular Dynamics Simulations of Rigid Molecules". In: *The Journal of Chemical Physics* 122 (2005), p. 224114.

[211] J. A. Anderson and S. C. Glotzer. “The Development and Expansion of HOOMD-Blue through Six Years of GPU Proliferation”. In: *arXiv:1308.5587 [physics]* (Aug. 2013). arXiv: `1308.5587 [physics]`.

[212] J. Glaser et al. “Strong Scaling of General-Purpose Molecular Dynamics Simulations on GPUs”. In: *Computer Physics Communications* 192 (July 2015), pp. 97–107. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2015.02.028`.

[213] G. J. Martyna, D. J. Tobias, and M. L. Klein. “Constant Pressure Molecular Dynamics Algorithms”. In: *The Journal of Chemical Physics* 101.5 (1994), p. 4177. DOI: `10.1063/1.467468`.

[214] J. Cao and G. J. Martyna. “Adiabatic Path Integral Molecular Dynamics Methods. II. Algorithms”. In: *The Journal of Chemical Physics* 104.5 (1996), p. 2028. DOI: `10.1063/1.470959`.

[215] M. E. Tuckerman et al. “A Liouville-Operator Derived Measure-Preserving Integrator for Molecular Dynamics Simulations in the Isothermal–isobaric Ensemble”. In: *Journal of Physics A: Mathematical and General* 39 (2006), pp. 5629–5651. DOI: `10.1088/0305-4470/39/19/S18`.

[216] D. Meagher. “Geometric Modeling Using Octree Encoding”. In: *Computer Graphics and Image Processing* 19 (1982), pp. 129–147. DOI: `10.1016/0146-664X(82)90104-6`.