Thesis Proposal: Designing an information-driven approach for targeted colloidal self-assembly

Shannon Moran

February, 2018

This paper is submitted in partial fulfillment of the University of Michigan Chemical Engineering Department Preliminary Exam requirements.

Committee:

Prof. Sharon Glotzer (Advisor)
Prof. Ronald Larson
Prof. Robert Ziff
Prof. Xiaoming Mao (Cognate: Physics)

Note: Total length must be less than 15 pages of text. Includes figures, excludes title page, list of references, and CV.

Contents

List of Figures

Preliminary Exam: Project Summary

Paragraph 1: Motivation

Paragraph 2: Where there are openings in the conversation about the role of information in self-assembly

Paragraph 3: 2 sentences each on the projects proposed

Paragraph 4: Concluding thoughts on future work

1 Introduction and motivation - 1 pg

When we think about the major challenges facing materials science, we are fundamentally faced with this idea of inversely designing materials. That is, I decide that I want to create a material that behaves your sweat-wicking shirt under one condition, stiffens under another, and when given a particular stimulus can reconfigure its structure. Currently, if I wanted to make a material like that for you, I'd naively take materials that have each of those properties and figure out how I could get them to work together. Or, I'd look for novel materials that have properties close to those of the material I want to make. We would call this designing the material.

This is inefficient. In the inverse design problem, we take the properties we want and create the materials that will give us those properties. Machine learning and materials science are coming together on the active front of this research. However, being able to predict, even perfectly, what can be made from existing materials by definition limits us to the set of materials that currently exist. This is a known challenge in materials science– how do we probabilistically explore phase space outside of phase space where we have data? While intriguing in its own right, that is not the topic of the thesis proposed here.

Instead, we might think about this from a fundamental physics point of view. If we want to make complex materials that have embedded stimuli responses, or assemble into a specific target structure, we must give the building blocks of such complex materials some amount of direction. We can think about this amount of direction as an amount of *information*.

This is not to say that we are looking to have building blocks act as storage devices, as in [**?**]. In that work, each building block is a cluster of multiple particles in whose arrangement can be stored a "high density" of information.

Similarly, in a recent proposal between our group and those of Marke Bathe (MIT), Mawgwi Bawendi (MIT), and Oleg Gang (Columbia), we proposed a biosynthetic, high-density storage structure composed of DNA nanocubes (Figure 6). Within these nanocubes, information could be stored in the different dies intercolated into the frame of the cube, into quantum dots placed into the frames, or even in the shape of the frames themselves. If we them move a level higher, we can imagine storing additional information in the order of these nanocubes relative to one another.

However, using these and solutions like them for high-density storage requires us being able to write, read, and store information into these formats. Fundamentally, these three challenges are predicated upon the ability to specifically place blocks where they need to go ("write"). Current methods include sonic and laser tweezers (manual), specific DNA interactions (energetic), or incremental addition (kinetic). How do we compare between these methods, though?

Here, I propose that the ability of building blocks to form a target structure can be distilled down into a concept of *information*.

This is not a new concept. In our group, we are comfortable with the concept that a target structure is the result of a minimization of free energy. In systems devoid of inter-particle forces, this then reduces down to a maximization of entropy.

Statistical mechanical "entropy" shares its name with information "entropy" in communications theory. While this directly came about because of the form similarity between the two, much energy since has been devoted to developing frameworks connecting the two. Jaynes,

in the 1950s, spent two long articles trying to reconcile the two. Books, and multiple articles, have been dedicated to explaining why these concepts are similar.

While much time has been spent developing the theory, very little time has been spent directly leveraging this concept for embedding information in systems governed by statistical mechanical ensembles– such as colloidal-scale self-assembly.

Key line from Simons proposal: "A coherent framework of thermodynamic and non-equilibrium processes seen through information theoretic eyes could lead to new theories for encoding information in matter– which would allow for the design of novel materials and novel material behavioral control."

New outline:

- Self-assembly in materials science can lead to a variety of structures, complexity
- The future of materials science relies upon inverse design
- Inverse design requires understanding how the building blocks of a material inform its overall structure
- Thinking more simply about tailored self-assembly... we want to direct the behavior of a system
- We can think of this as giving the system some amount of information: binding preferences, kinetics, etc
- Lots of work has gone into trying to understand how to measure and then most efficiently provide systems with that direction (will detail in the background section)
- You know what does this really well, though? Proteins in nature
- Proteins conformationally change while binding; lots of complexity
- Here we use a simpler system of folding nets, which actually has much less complexity
- The overarching challenge is: what is the most efficient way of assembling a given structure?
- Such a question could motivate a career, so we will further limit this scope in the coming pages.

2 Literature Background - 2-3pg

2.1 What is information?

Let's first look at information in the context of communications theory.

The *information*, I, we get from an event happening is given by:

$$I(p) = -\log(p)_b \tag{1}$$

where p is the probability of an event happening and b is the base. Base 2 is commonly used in information theory, and forms the unit of information. For instance, the unit of base 2 information is a bit, base 3 are trits, base 10 are Harleys, and base e are nats.

In 195X, Claude Shannon also extended this concept by introducing the concept of *information entropy*. In this context, entropy is the average (expected) amount of information gained from a given event. Specifically, for an event with n different outcomes this can be

written as:

$$\text{Entropy} = \sum_{i=1}^{n} p_i \log p_i \tag{2}$$

For a discrete random variable X with $p(x)$, the entropy can be written as:

$$H(X) = \sum_{x} p(x) \log p(x) \tag{3}$$

Entropy does not range from 0 to 1. The range is set based on the number of possible outcomes n, i.e. $- \leq \text{Entropy} \leq \log(n)$. Entropy is equal to 0 (minimum entropy) when one of the probabilities is 1 and the rest are 0's. Entropy is $\log(n)$ (maximum entropy) when all the probailities have equal values of 1/n.

In the case of designing specific outcomes for an event, then, we want to minimize the entropy along each leg of the pathway leading to an event. Put another way, we want to maximize the probability that the event will proceed down the pathway we want it to.

However, the concept of "information" in this context is then counter-intuitive. Information in communication refers to how likely an event is. When a rare event happens, we gain more "information" from that event. However, in the context of designing specific outcomes, we are not looking to read out bits of information once an event has happened. We are looking to design the likelihood of an event occurring.

In the words of MIT professor Cèsar Hidalgo, "It is hard for us humans to separate information from meaning because we cannot help interpreting messages." We face the same problem here– by saying that a pathway has more information than another, we are implicitly saying that it is a rarer event than a lower-information pathway.

Counter-intuitively, in designing pathways for self-assembly, then, we are looking to design minimum-information pathways. **However, in aligning with our intuition from self-assembly, this means we are looking to maximize the entropy of an assembly pathway.**

However, we can use the concept of *mutual information* in defining how much information is stored in an interaction in an intuitive manner. (See notes on the Brenner paper below.) Mutual information $I(X;Y)$ is a global measure of interaction specificity in systems with many distinct species. It quantifies how predictive the identity of a lock x_i is to the identity of key y_i found bound to it.

2.1.1 How does this tie into statistical mechanics?

2.2 What is information, in the context of self-assembly?

Let's first look at a paper from the Brenner group, the "Information capacity of specific interactions' [**?**]. Their main thesis is that specific binding interactions have energetics that allow binding to occur with measurable probability. Thus, we can measure the relative information in different types of binding. This is more in line with the communication theory view of information (rare events giving more information) than it is with the materials view

of information, in which high information events imply high probability of a desired event happening.

Our group, and many others in the materials community, are looking to engineering materials to control their structures, behaviors, etc. A common method of engineering these materials is by tailoring the interactions between their components through chemistry, shape, etc. By understanding how much *assembly information* can be contained in these interactions, we can:

1. Compare the efficacy of different types of interactions in delivering desired behavior(s)
2. Theoretically predict the efficacy of new types of interactions

Let's take the example of a lock and key system. ADD SECTION ON BRENNER PAPER FROM LIT REVIEW LAST YEAR

Any of Jacobs' papers that talk about this?: uses connectivity graphs

2.3 Where have folks applied directed self-assembly? Why do we care about it?

Glotzer, Kotov - self-assembly

Mirkin - experimental, DNA-directed

Kamien - kirigami

Glotzer, Desmaine - folding

Frenkel, Jacobs - pathway design

Wales - pathway designs, disconnectivity graphs

2.4 Already proposed work: Semiconductor Synthetic Biology

Include work done on the NSF grant proposal: Using DNA-mediated assembly to store information in nanoparticle arrays.

Specifically, this is an example of 1b): addressable complexity, then trying to engineer how to get the particles to where they should go in the most energetic and information/complexity-efficient manner possible.

2.5 Review of the literature

Let's take a simple system and use it to illustrate the state of the art in the theory for engineering self-assembly behavior.

Let us say that we have the simple target 2D array shown in Figure 1.

If we assume all the particles in our system are uniform, we simply care about making sure particles assemble into the target structure. As in any type of assembly, this implies

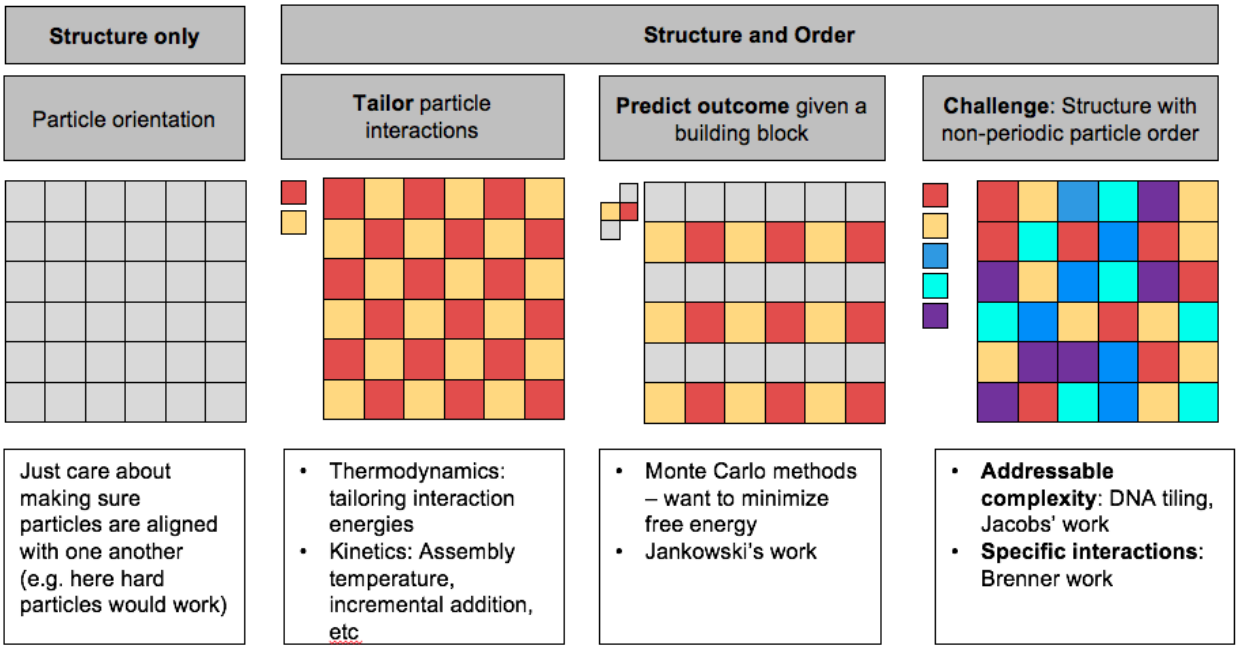

Figure 1: Demonstration of the state of the art in the literature

a minimization of free energy. In the example of hard squares above, minimization of free energy corresponds to a maximization of entropy.

- Archimedean tilings
- Glotzer work on assembling target structures

[?]: Can treat shape as giving particles an effective entropic patchiness

[?]: Develop a metric of self-assembly complexity, flow chart in Figure 3. Archimedean tilings (ATs) Here, we report the minimal set of interactions needed to self-assemble experimentally accessible ATs from regular polygons, mimicking nanoplates assembled into crystalline monolayers (Figure 1). We show through Monte Carlo simulations the self-assem- bly of these tilings by exploiting entropic and enthalpic interactions encoded in the shape of the polygons. We arrive at a design strategy for patchy polygon particles that is accessible to current experimental techniques and present the minimal set of design rules for each AT. We report that four ATs, namely, the (63), (36), (44), and (3.122) tilings, can be assembled solely with hard interactions, highlighting the role of directional entro- pic forces39,40 that arise from the particle shape.

After selecting the building blocks, the design pro- cess examined the constituent polygonal building blocks and alters the interaction complexity by chan- ging the specificity of interactions. The four models ranked in terms of specificity are hard, symmetric patches, shape-specific patches, and edge-specific patches

Manually adds complexity in the amount of specificity: Initially, we test if entropic interactions are sufficient to self-assemble each crystalline structure. If the infinite pressure ground state (hard particle) won't assemble the target structure, add attractive interactions. If the crystalline structure for a mixture of building blocks does not contain the alternating building block property, it is necessary to use edge-specific interactions.

As we look to make more complex materials, though, we also want to specifically order the particles within a structure. We can approach this through tailoring the thermodynamics and/or the kinetics.

1. Glotzer, Kotov work on patchy particles for getting different arrays/structures
2. BUBBA work
3. DNA tilings, DNA-tailored assembly
4. need to find some other citations here?

1. [**?**, **?**, **?**]
2. [**?**, **?**, **?**]
3. [**?**, **?**, **?**]

[**?**]: Despite the apparent simplicity in particle geometry, the combination of shape-induced entropic and edge-specific energetic effects directs the formation and stabilization of unconventional long-range ordered assemblies not attainable otherwise.

[**?**]: can create crystals using DNA functionalization [**?**]: Colloidal crystallization can be pro- grammed using building blocks consisting of a nano- particle core and DNA bonds to form materials with controlled crystal symmetry, lattice parameters, stoichiometry, and dimensionality. Can tailor DNA length and shape to get different crystal structures

DNA programmable assembly. The sequence-specific binding property of DNA can be applied to direct the assembly behavior of colloidal particles at the nanoscale. This is a powerful strategy to control nanomaterial assembly because it allows tuning of the interparticle interaction in highly specific ways. For example, attaching DNA linkers with self-complementary sequences to particles directs them to maximize the number of nearest neighbors, resulting in fcc arrangement. On the other hand, particles with non-self-complementary linkers, which only allow A-B contacts, assemble into bcc or CsCl-type arrangement by maximizing the number of A-B contacts around the first coordination shell [**?**].This technique has recently been applied to anisotropic particles and opened possibilities to access diverse crystal structures in nanoscale by employing the DNA programmability as well as geometrical properties of the anisotropic shapes (25, 26).Despite its powerful potential, studies regarding the assemblies of DNA-coated anisotropic colloids have until recently (Sangmin's paper) been limited to simple crystal structures with small unit cells (27, 28).

To test the interactions we've tailored above, we also need to be able to test the assembly properties of these building blocks. We want to see what structure they will assemble into to minimize their free energy. Our initial instinct would be to run these simulations to steady state or equilibrium. However, systems with complex interactions such as these can get themselves caught in metastable free energy local minima that MC methods may not have large enough energy fluctuations to escape from. This is where thermodynamics and kinetics really come to play together

- BUBBA work
- Wales work (disconnectivity graphs) - more of the thermo
- Kinetic pathway design - Frenkel, Jacobs
- Connectivity graphs to show the order of assembly
- Pathway assembly sampling
- swear I read something earlier today about metastable/kinetic traps

Real-world:

- Protein assembly
- Complex DNA tiling, assembly - Mirken, Winfree, Rutherford, check semisynbio sources

Theory perspective:

- Addressable complexity
- Efficiency of specific interactions
- Getting partition functions– hard to find all the possible states; folks are trying to tackle this with information-theoretic approaches
- Related to the above, even in 2D assembling structures given this level of specificity is a drag, and is not a simple model system (interactions between all the different particles, kinetics of the assembly, etc)
- Figure 6 summarizes these challenges nicely
- Need to find a simpler system

Folding nets are a pretty well-defined system. We need a system with limited degrees of freedom to establish these concepts.

- Paul's paper (Glotzer)
- Demaine folding paper, and any additional citations
- Origami papers

If we didn't care about the ordering of the A and B particles on the lattice, we know that we could leave the particles as hard particles with only volume exclusion, and at proper densities the system would self-assemble into a square lattice to maximize entropy [CITE]. However, we care about the order of species within the lattice. Our goal is to give the A and B particles some interaction that allows them to kinetically and thermodynamically reliably assemble that target structure.

Have specific interactions between the particles

with interaction energies ϵ_{AB}, ϵ_{BA}, ϵ_{AA}, and ϵ_{BB}

Information as a measure of the likelihood of a particular configuration being preferred. In 2015, a review article in *Nature Physics* (which has since been cited 255 times) reviewed the state of the art on applying information theoretic entropy– i.e. Shannon entropy– as a way of understanding non-equilibrium thermodynamics [**?**]. In this work, they investigate information entropy as a placeholder for non-equilibrium entropy production. This entropy production gives an overall likelihood of a configuration (one which minimizes the non-equilibrium free energy of a system while maximizing the non-equilibrium entropy). However, this method applies to the overall structure, or the overall likelihood of a structure being the preferred structure.

"Addressable complexity" seeks to engineer pathways for particular particles to reach their destination. Low free energies of a target structure, however, do not guarantee efficient assembly. There are a number of ways addressing this problem in the literature. One way of forcing systems into assembly is to design a free energy landscape that minimizes such meta-stable traps [**?**]. Competition between degenerate structures of equivalent potential energy was reported for clusters of six attractive spherical colloids, where symmetry breaking leads to higher rotational entropy of the less symmetric conformation, resulting in lower free energy [**?**].

Taken from [**?**]: A well-known example of addressable complexity– that is, specific binding–

can be found in "one-pot" DNA self-assembly of DNA tiles, which use the hybridization of complementary DNA sequences to construct complex structures consisting of hundreds of subunits from a single soup of monomers [**?**] (5). Simulation results have shown that such one-pot self-assembly can succeed with highly simplified model subunits that lack the molecular details of DNA tiles, suggesting that similar design strategies should be widely applicable [**?**] (6). In the work by Jacobs *et al*, they had particles with designed interactions between one another. They represented the target bonds by a graph, G. However, this model is based on the assumption that "designed interactions in the target structure are typically much stronger than any incidental associations between sub-units that should not be connected in the final assembly". This is a fine assumption for their proof of concept, but is not valid in real-world system. As a concrete example, protein-folding is perhaps the most well-explored biological system that assembles due to specific interactions [**?**]. However, one of the major challenges to solving the protein-folding problem are competing "cross-talk" interactions [CITATIONS NEEDED; chaperoned folding and assembly Chakrabarty 2017].

In later work, Jacobs et al addressed this oversight and accounted for incidental interactions in addition to designed interactions [**?**].

Low energies may not guarantee efficient assembly compartmentalized, multi-stage assembly grannemana and baserga 2004

Talk outline: 1. self-assembly kinetics can be rationally designed– leverage thermodynamics 2. evolution has already selected for optimal assembly pathways in complex biomolecules

Specific binding interactions can be tailored to lead to target structures. However, there is a delicate balance of specificity required. On the over-specified side, we have bonds that are specific to their intended neighbor with probability 1.0. On the under-specified side, we have non-specific interaction patches that will bond to any other patch with probability $1/n$, where n is the total number of patches in the system.

In work from our group, Eric Jankowski sought to generate energy-minimizing configurations for such patchy particles [**?**] in a process he called "bottom-up building block analysis", or BUBBA. Cluster Monto Carlo (cMC) and LAcMC methods are relatively poor methods for finding potential energy minima formed from patchy particles with disparate interaction energies due to their tendency to become trapped in metastable configurations as well as the low degeneracy of potential energy-minimizing configurations (Q). BUBBA effectively searches a subset of the configuration space for energy-minimizing configurations. Jankowski predicted that that BUBBA would be useful for evaluating many different particles for self-assembly "propensity".

Partition functions encode all the thermodynamics of a system, but for most systems of practical importance they cannot cannot be calculated exactly. This is due to many indistinguishable degenerate states. In the cases where small numbers of distinguishable configurations comprise a majority of a partition functions' weight, as is the case for systems at low temperatures and for many anisotropic building blocks with disparate interactions, BUBBA is a particulalry effective method for generating partition functions that have been heretofore inaccessible. This allows us to ask "What structures are thermodynamically favored for this building block at any temperature?" to be answered independently of assembly kinetics. [**?**]

Both thermodynamic and kinetic barriers to assembling target structures.

Key problem, taken from [**?**]: Self-assembly holds promise for creating new materials and devices because of its inherent parallelism, allowing many building blocks to simultaneously organize using preprogrammed interactions. An important trend in nanoparticle and colloid science is the synthesis of particles with unusual shapes and/or directional (??patchy??) interactions, whose anisotropy allows, in principle, assemblies of unprecedented complexity. However, patchy particles are more prone to long relaxation times during thermodynamically driven assembly, and there is no a priori way of predicting which particles might be good assembly candidates. Here we demonstrate a new conceptual approach to predict this information using sequences of intermediate clusters that appear during assembly. **Unfortunately, when an equilibrium solution or simulation of patchy particles fails to generate an ordered pattern it is not always obvious whether the culprit is thermodynamics or kinetics.** Recently there have been studies that attempt to quantify kinetic trapping through fluctuation-dissipation ratios (21,22) and through the interplay between specific and nonspecific interactions (3,5,23) but these methods do not provide predictive capabilities for thermodynamically stable structures. The fact that both thermodynamics and kinetics can prevent a system of particles from self-assembling is particularly trou- blesome for experimentalists that search parameter space via trial-and-error because experiments that fail to assemble do not provide information about how assembly might be improved.

We are ultimately searching for rational design of building blocks optimized for self-assembly that focuses on assembly pathway engineering: identifying the traps that occur as a system assembles so they may be circumvented. As systems self-assemble we hypothesize that the thermodynamically stable intermediate clusters that arise hold information about their ability to order. These sequences of intermediate clusters are assembly pathways and we propose a methodical analysis of them to predict the degree to which a system of building blocks will assemble a target pattern, which we refer to as the building block?s assembly propensity for the pattern. We foresee assembly pathway engineering proceeding as a collaboration among structural identification, kinetic measurements, and the assembly pathway analysis described here. [**?**]

2.6 Motivation 1: DNA assembly, DNA tiles, DNA cubes

Taken from intro of [**?**]:
The observation by Ke et al. [Science 338, 1177 (2012)] that large numbers of short, predesigned DNA strands can assemble into three-dimensional target structures came as a great surprise, as no colloidal self-assembling system has ever achieved the same degree of complexity. That failure seemed easy to rationalize: the larger the number of distinct building blocks, the higher the expected error rate for self-assembly. The experiments of Ke et al. have disproved this argument. Here, we report Monte Carlo simulations of the self-assembly of a DNA brick cube, comprising approximately 1000 types of DNA strand, using a simple model. We model the DNA strands as lattice tetrahedra with attractive patches, the interaction strengths of which are computed using a standard thermodynamic model. We find that, within a narrow temperature window, the target structure assembles with high probability. Our simulations suggest that misassembly is disfavored because of a slow nucleation step. As our model incorporates no aspect of DNA other than its binding properties, these simulations suggest that, with proper design of the building blocks, other systems, such as colloids, may also assemble into truly complex structures.

Need to include Winfree, Rutherford papers in here.

Include work from NSF proposal.

2.7 Motivation 2: Protein folding

Steal references.

3 Description of proposed research (7-8 pg, 2-3 pg per aim)

"Information" are those factors that impact the yield and kinetics of self-assembly (thermodynamics of the free energy landscape and the kinetics of the path to get to a target structure from a given starting point).

Specifically, as we look to both understand how nature governs self-assembly into target structures, we need a language to understand this.' Nature is very good at already picking an optimized route through a free energy landscape [**?**].

3.1 What I want to accomplish in my thesis

1) Can we figure out which bonds/interactions are the most critical in an assembly process or an assembly pathway? We intuitively know that in Need transition state or pathway sampling methods (or might actually be able to get this just from Paul?s data? That would be sick)

2) With this metric, can we then define and minimize an information efficiency, e.g. the amount of bond specificity we need across the system to get a given success rate of assembly? E.g. is it more efficient 3) Given all this, can we then use machine learning to determine a priori the ?most efficient? level of specificity for self-assembling a target structure? How to do pathway design is kind of an open question We can find feature correlations, like Paul did There?s also an approach called ?Computable Information Density? published by some colleagues (Chaikin) on the arxiv last August Basic idea is that you can (1) somehow represent your system as an array of information which you can (2) run though a compression algorithm and (3) the ?information? is just the length of that compressed information Would be really interesting to see if I could extend that idea to the features of an assembly system? in this case, nets? and

1) Define a measure of pathway information.

- We already have ways of measuring how good a particular bond is
- Are there particular bonds/connections that are the most important to get correct to enable forming the desired final structure?
-

2) Use that measure of pathway information to design ideal pre-cursors for target structures.

3) Attempt to use machine learning to predict ideal pre-cursors for given target structures.

- [**?**]: Nonlinear Machine Learning of Patchy Colloid Self-Assembly Pathways and Mechanisms out of the Furguson group

4) Why is it important we find the "most important" pathway points? from [**?**]
How then can self-folding origami be folded with a minimal number of actuators? A lesson can be drawn from similar glassy landscape search problems in models of protein folding (e.g., Levinthal?s paradox [17, 19, 20, 41]) and related NP-hard satisfiability (SAT) problems [21, 42] that vary from the Traveling Salesman Problem to Sudoku [43]. A common element in these satisfiability problems is that random seeding of the search for the global minimum leads to repeated backtracking after reaching local minima, both in the context of computer algorithms (as the DPLL algorithm for k-SAT [21]) or for physical dynamics (as in protein folding) [42]. However, careful seeding of the search - e.g., if the right boxes are filled in first in Sudoku [43] or if the right parts of the protein are folded first - can greatly reduce or even eliminate backtracking [21] before reaching the global minimum. Correct seeding is even more critical for origami since folding is assumed to happen at ?zero temperature? (e.g., without any noise or fluctuations). As a result, the structure cannot backtrack out of a local minimum as in the case of non-zero temperature SAT problems [42].

This reference also has a really good introduction section relating origami and self-assembly [**?**].

3.2 Active shapes work

Relation to proposed thesis topic: If we define information broadly as any quality of a building block that impacts the emergent behavior of a system of those particles (in this case, force direction and shape), then we can argue this work is looking at a few aspects of information in active systems.

3.3 SemiSynBio Work

From intro: In Aim 1, we will investigate monomeric block formation, exploring the self-assembly of arbitrary geometric DNA objects with incorporated optical elements that can be manufactured as information carriers, while allowing for superstructure formation through DNA-sequence barcoding. We will explore static assembly of 1D arrays of such DNA nanoparticles integrated with Memory Blocks (DNAMB) for encoding bitstream information that can be read out by fluorescence and electron microscopy. In Aim 2, we will explore 2D and 3D assembly, investigating techniques to algorithmically assemble and read out digital 2D and 3D information using optical and tomographic methods. In Aim 3, we will use molecular decision computing to assemble distinct, alternative lattices based on specific external signals. These results will offer the ability to encode and decode arbitrary datasets in ultra-dense molecular hard-drives, with environmental sensing and recording.

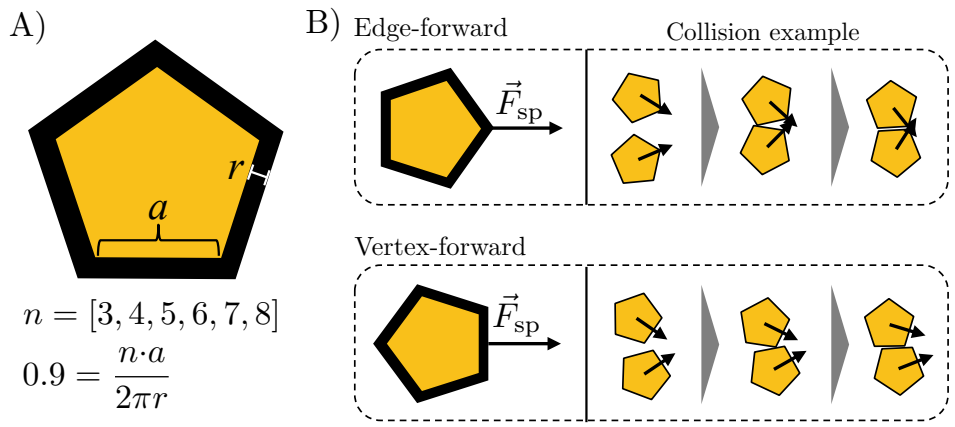

Figure 2: **Model system**: (1) We use rounded shapes of constant S-C ratio, where the corners are rounded by a WCA potential, to ensure we can distinguish between shape steric (anisotropic) effects and isotropic behavior; (2) Force can be applied either perpendicular to the face or directed out a corner

(Qs) should all particles be the same size?– can't be, the way it's set up; all particles have drag of an equivalent disk; should we neglect noise? what does that mean for these simulations?

Aim 2. Dense, programmable molecular memory in 2D and 3D bit module lattice assemblies Overview & Rationale. Nanoparticle self-assembly depends on a balance of interaction forces, entropic effects, and system kinetics37,53,69-73. We can leverage these properties to direct self-assembly of shaped DNA nanoparticle into 2D and 3D arrays by controlling the position and valency overhangs that provide connectivity between DNA nanoparticles. Wireframe structure of DNA particle is highly suitable for encapsulation of memory blocks (e.g. Au NP, QDs, fluorescent dyes) and creation of DNAMB, a pixel in 2D or 3D arrays. To achieve information storage capabilities, it is required to investigate how the connectivity properties of DNAMB can be translated into their designed arrangement in the information-storing arrays. To self-assemble these systems into 2D and 3D ultra-dense data blocks, we will investigate the minimum interaction specificity needed to direct self-assembly into high fidelity ordered 2D and 3D arrays. We will also explore information retrieval from these arrays in 2D and 3D within pixels consisting of 1x1, 2x2, 4x4, etc., nanoparticle block arrays. In addition, we will establish methods for generating robust memory arrays that can preserve information under extreme conditions.

Proposed research, Aim 1: Computationally, Glotzer and colleagues will develop a bit module interaction model to study the role of the DNA linkages on DNA cage self-assembly. Specifically, previous work on modeling solid particles with DNA-facilitated attraction37 will be extended to model the DNA cages that will be experimentally made by Gang, and it will consider realistic features of nanoparticle systems82,83. With this model in place, we can then extend the framework of digital alchemy, which treats particle properties as a thermodynamic variable, to particle interactions (here, DNA linkages)54. In this way, we can inversely design ideal DNA cages (e.g. shape, patchy interactions) that will robustly assemble a target structure. We will seek to balance site specificity without being overly unique?that is, design the highest information interactions that will allow for the minimum amount of linkage specificity for directing self-assembly84,85. This computational framework for the inverse design of bit packages that will assemble a given structure will enable high-

throughput screening of particles of interest and serve as the basis for complex hierarchical structure and array assembly in the remainder of Aims 2 and 3.

Sub-Aim 2.2. Hierarchical assembly logic for higher-dimensional information storage Overview. In Sub-aim 2.1, we explored approaches to assembling DNA frames into target 2D and 3D assemblies. Next, we precisely order ?bit modules??that is, DNA cages carrying functional particles? into arrays of discrete information. Toward this end, we design modules that carry the minimal information needed to reach target arrangements through a combination of particle anisotropy and DNA linkers. DNA computing groups have previously used DNA linkers to self-assemble complex 2D patterns6 and 3D shapes9. Here we extending these approaches to realize hierarchical 3D nanoparticle assembly design so that pixelated images act as dense data storage units. We will explore several complementary approaches to hierarchical assembly engineering, including sequential nanoparticle addition and ?one-pot assembly?, each of which will be explored together with inverse computational design of self-assembly pathways and particle geometries to achieve a robust assembly of designed arrays. Using the optical characterization strategies from Sub-aim 2.1, we will decode the information encoded in the structure, and probe sources of error and information loss in the self-assembly and read-out processes.

Proposed research. Assembly of encoded 3D arrays can be approached in two strategies, or a combination thereof (Fig. 6). In an entirely ?one-pot? assembly of a 3D array, all modules are linked with a large binding sequence set that has been fully computationally defined. Such an approach requires an enormous number of unique binding sequences, and even if fully defined, can run into high error rates when considering the assembly and packing of large (as compared with molecular assembly) and charged modules and/or materials. A second approach using sequential binding based on module groups of similar binding layouts requires less sequence diversity and can be automated using robotic liquid handling. However, this is vastly more process- and time-intensive than one-pot assembly. This approach represents hierarchical assembly, whereby 1D structures (?strings?) would be formed from the modules, 2D planes formed from the 1D libraries, and finally 3D encoded arrays from stacking of selected planes. An optimal assembly process that balances fully-encoded organization with direct addition of binding components would offer a hybrid approach of hierarchical assembly with sequential addition of groupings of computationally defined structures. Each of these strategies will be explored in this aim, using a combination of high-throughput, structure-based computational modeling and experiments.

The Glotzer group will extend their digital alchemy framework to probe diverse DNA linkage sequences and conjugation designs to realize specific, targeted inter-particle interactions. In addition, they will explore the roles of these interactions on the kinetics of array assembly to enable pathway design [**?**] into desired arrays while avoiding undesirable ?side products?. In this way, we will explore computationally the interplay between the two extremes of one-pot and sequential assembly, and identify which combinations provide lowest assembly error while minimizing both assembly time and the number of required unique binding sequences. The Gang group will employ a home-built robotic system for automatic synthesis and assembly of DNAMB; that will allow establishing practical methods for creation of large number of diverse blocks required for the hierarchical assembly. While such approaches have been applied to molecular systems, they have not yet been realized for DNA frames integrating inorganic NP. To implement complementary pathway design strategies, Gang will fabricate DNA frames with thermally differentiated inter-vertex hybridizations to promote highly specific assembly path during thermally-driven self- assembly. For example, DNAMB strings will be assembled at higher temperatures, and planar and 3D arrays assembled at

lower and lowest temperatures, respectively. We will use SAXS and tomography methods to reveal the pathway-controlled assembly process. The computational design of frames and pathways will be performed jointly between the Bathe, Glotzer, and Gang labs.

3.4 Paper 2 - Defining information as it pertains to colloidal systems - 2.5 pg

3.5 Paper 3 - Applying that definition of information to nets (folding systems) - 2.5 pg

Goal

Background

Hypothesis

Methods to use

- Molecular dynamics simulations
- Pathway sampling: forward flux, transition states, etc (check Chrissy and Simon's notes from this)

3.6 Paper 4 - Using machine learning to design information-rich starting structures: 1 pg

4 Time table

See Figure 7 for key tasks and milestones through 2020, based on the projects outlined in the above sections.

5 Conclusions and potential impact

Long, long-term goal: Instead of simply observing emergent behavior as an outcome of collective motion of individuals, we could instead engineer such behavior as a quantifiable outcome of the interaction of an information-rich network of agents.

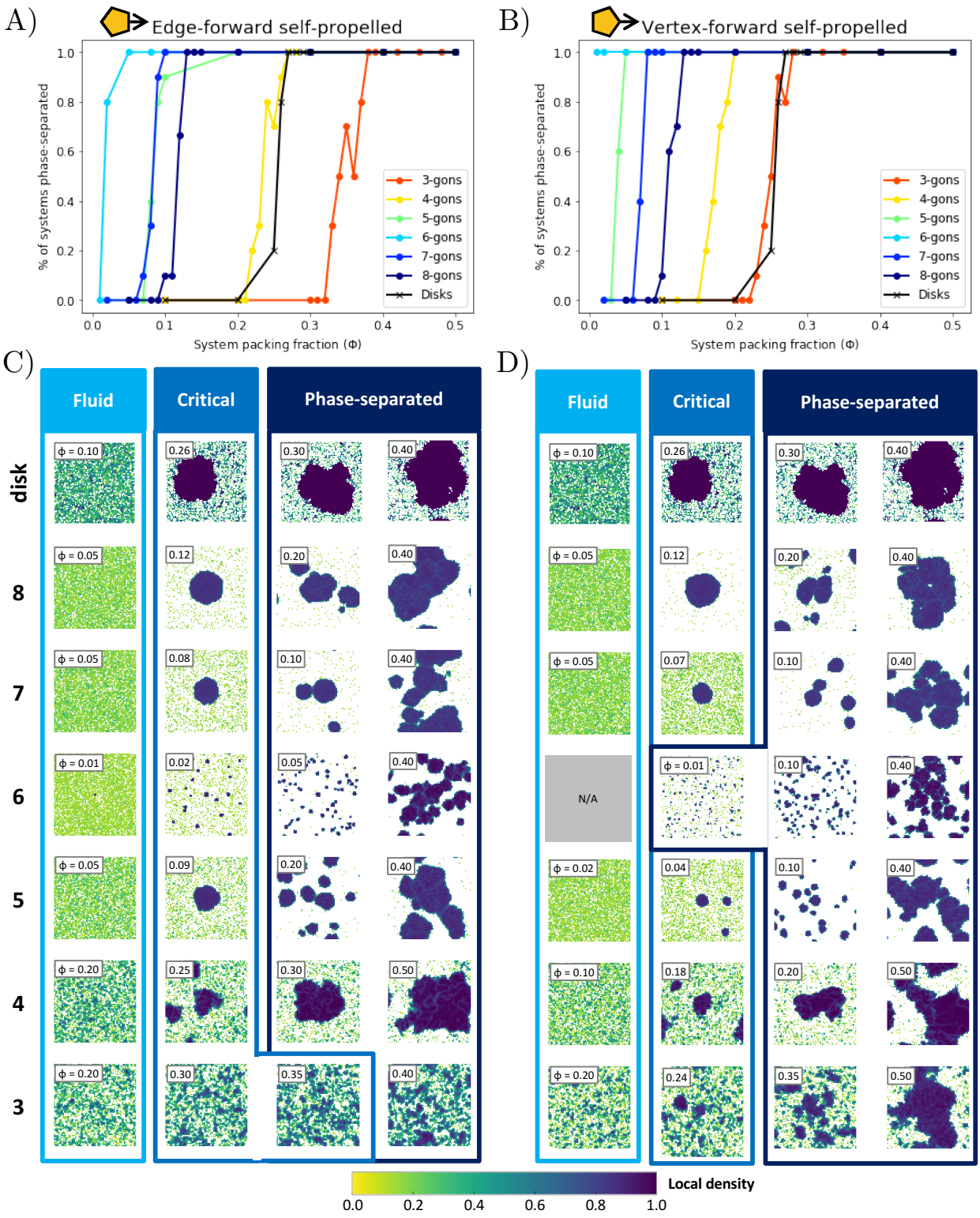

Figure 3: **Critical density and nucleation behavior**: (A) Average domain size (cluster size? grain size?) versus time is different for disks versus shapes, and also depends on force director. (B) Critical density, the density at which SOME DEF OF CLUSTERING OCCURS, depends on both shape and direction of force director.

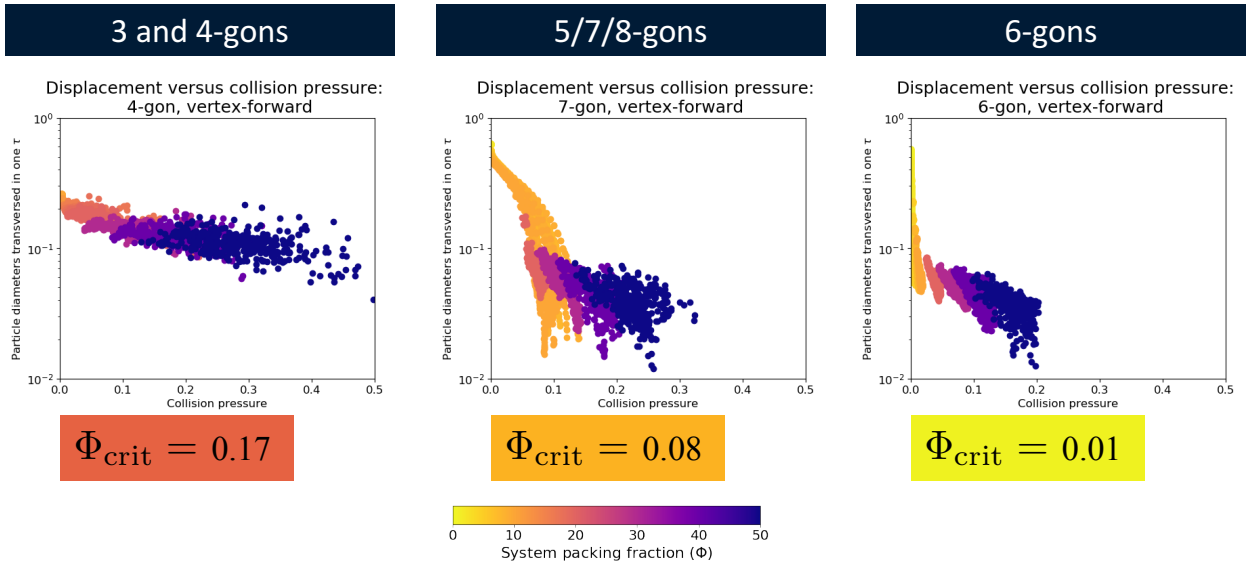

Figure 4: **Collision efficiency**: Something with pressure? Not sure how to use this yet, but feel like there's something here...

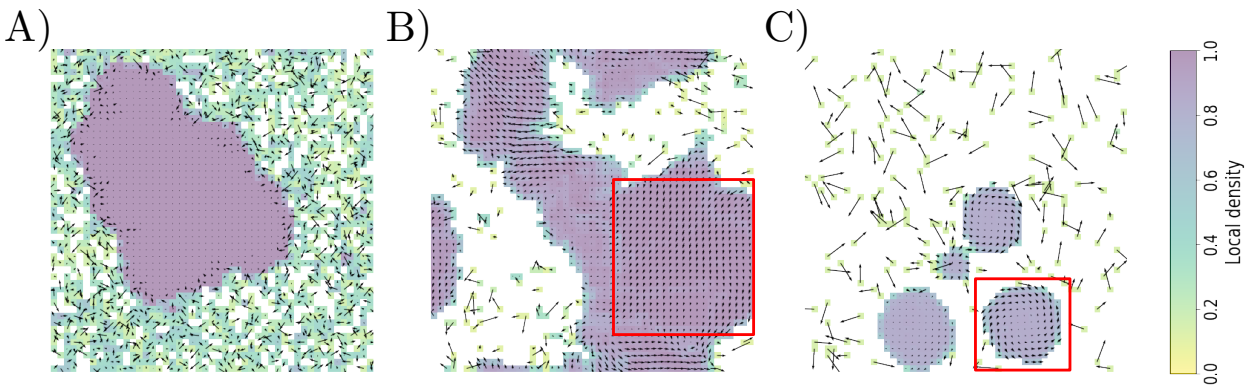

Figure 5: **Displacement fields**: In contrast with disks, clusters are able to convert translational forces into rotation (highlighted in red boxes). Clusters of disks can?t sustain translational or rotational motion? clusters of shape can (this was off-hand noted in Suma et al)

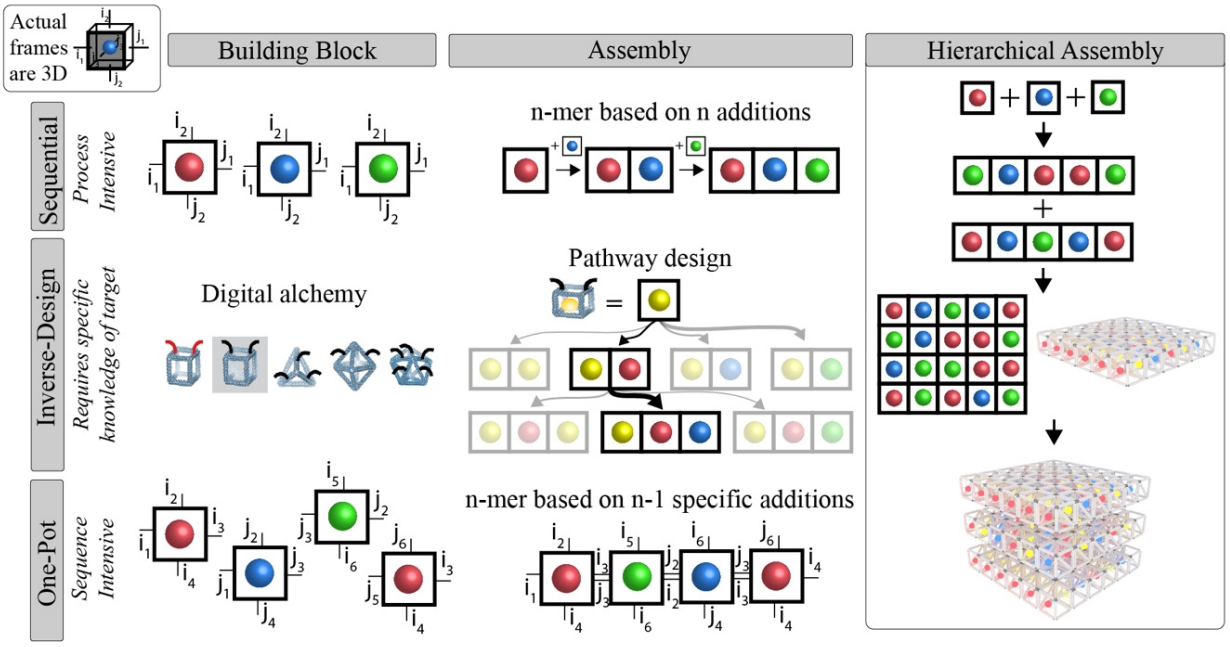

Figure 6: (Lifted from SemiSynBio Proposal) Strategies for building information encoded 1D, 2D and 3D arrays. Sequential operations are very deterministic and can be carried out by automated robotic equipment, but even so are heavily process intensive and require many individual assembly steps. One-pot systems can be fully computationally defined, though in practice are heavily sequence intensive and be subjected to errors more readily than in molecular- scale systems when accounting for kinetic and thermodynamics of packing larger objects and materials. A hierarchical assembly methodology offers a hybrid approach of both strategies, where a sequential addition of structures preformed in a one-pot setup provide the desired 3D material organization.

			2017				2018				2019				2020	
Project	**Status**	**Description**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**
Active shapes	100%	Data collection														
		Analysis														
		Writing														
		Submit for publication														
Defining information		Object 1														
		Object 2														
		Object 3														
		Object 4														
Applying information		Object 1														
		Object 2														
		Object 3														
		Object 4														
Machine learning		Object 1														
		Object 2														
		Object 3														
		Object 4														
Thesis		Data meeting														
		Defense														

Figure 7: Key milestones and tasks from Preliminary Exam through target defense date.