

Probabilistic inverse design for self-assembling materials

R. B. Jadrich, B. A. Lindquist, and T. M. Truskett

Citation: [The Journal of Chemical Physics](#) **146**, 184103 (2017);

View online: <https://doi.org/10.1063/1.4981796>

View Table of Contents: <http://aip.scitation.org/toc/jcp/146/18>

Published by the [American Institute of Physics](#)

Articles you may be interested in

Communication: Inverse design for self-assembly via on-the-fly optimization

[The Journal of Chemical Physics](#) **145**, 111101 (2016); 10.1063/1.4962754

Perspective: Dissipative particle dynamics

[The Journal of Chemical Physics](#) **146**, 150901 (2017); 10.1063/1.4979514

Perspective: Chemical reactions in ionic liquids monitored through the gas (vacuum)/liquid interface

[The Journal of Chemical Physics](#) **146**, 170901 (2017); 10.1063/1.4982355

Osmotic and diffusio-osmotic flow generation at high solute concentration. II. Molecular dynamics simulations

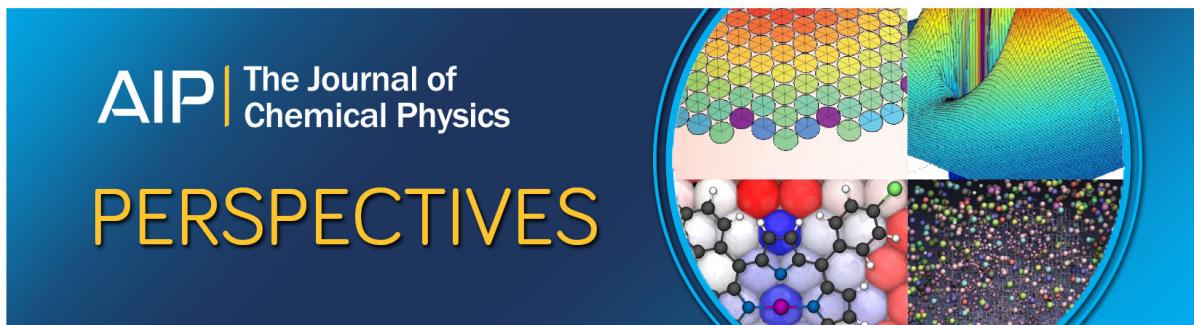
[The Journal of Chemical Physics](#) **146**, 194702 (2017); 10.1063/1.4981794

Evidence for self-organized formation of logarithmic spirals during explosive crystallization of amorphous Ge:Mn layers

[Journal of Applied Physics](#) **121**, 184901 (2017); 10.1063/1.4983068

Stress dependence of the suspended graphene work function: Vacuum Kelvin probe force microscopy and density functional theory

[Applied Physics Letters](#) **110**, 193101 (2017); 10.1063/1.4982931



Probabilistic inverse design for self-assembling materials

R. B. Jadrich, B. A. Lindquist, and T. M. Truskett^{a)}

McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, Texas 78712, USA

(Received 15 February 2017; accepted 3 April 2017; published online 9 May 2017)

One emerging approach for the fabrication of complex architectures on the nanoscale is to utilize particles customized to intrinsically self-assemble into a desired structure. Inverse methods of statistical mechanics have proven particularly effective for the discovery of interparticle interactions suitable for this aim. Here we evaluate the generality and robustness of a recently introduced inverse design strategy [B. A. Lindquist *et al.*, *J. Chem. Phys.* **145**, 111101 (2016)] by applying this simulation-based machine learning method to optimize for interparticle interactions that self-assemble particles into a variety of complex microstructures as follows: cluster fluids, porous mesophases, and crystalline lattices. Using the method, we discover isotropic pair interactions that lead to the self-assembly of each of the desired morphologies, including several types of potentials that were not previously understood to be capable of stabilizing such systems. One such pair potential led to the assembly of the highly asymmetric truncated trihexagonal lattice and another produced a fluid containing spherical voids, or pores, of designed size via purely repulsive interactions. Through these examples, we demonstrate several advantages inherent to this particular design approach including the use of a parametrized functional form for the optimized interparticle interactions, the ability to constrain the range of said parameters, and compatibility of the inverse design strategy with a variety of simulation protocols (e.g., positional restraints). *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4981796>]

I. INTRODUCTION

The engineering of materials with nanoscale structural features is an increasingly relevant design challenge, the difficulty of which grows rapidly with a diminishing structural feature size. One emerging strategy in this area is to tailor interparticle interactions between material building blocks to drive their spontaneous assembly into specified architectures. Working as either a stand alone technique or in tandem with other small scale patterning technologies, like lithography, novel nanostructured morphologies can be realized by self-assembly. Colloids are particularly amenable to such an approach due to the wide variety of experimentally tunable interparticle interactions that are possible, including (approximately) isotropic pair interactions for nearly spherical colloids that depend on factors such as the material composition of the particle and the ligand coating, as well as the quality and composition of the solvent.^{1,2} Directionally specific, short-ranged interactions between “patchy” particles or excluded volume interactions between particles of different shapes are possible as well.^{3,4} Furthermore, various classes of interactions can be controlled in an approximately orthogonal fashion, giving rise to a rich design space for assembly. From colloidal building blocks, numerous self-assembled structures have been observed experimentally. The two-dimensional kagome lattice was assembled via triblock Janus particles, for instance.⁵ Other self-assembled microstructures include tubes, micelles, lamellar sheets, and viral capsid-like structures, to mention a few.^{6,7}

The sheer variety of interactions available for colloids presents an intriguing challenge. How does one rationally select appropriate interactions to drive assembly of a desired structural motif or target a certain property or functionality? For some relatively simple cases, intuition or an empirical method may suffice. However, more complex architectures or non-trivial colloidal interactions necessitate a more systematic approach. In recent years, many inverse design (ID) strategies have emerged to allow for the discovery of interactions that maximize a selected figure of merit in an (ideally) automated fashion.^{8–14} In addition to ID algorithm development, advances in computational power have enabled ID schemes to play an increasingly influential role in the rational design of materials.

In general, ID relies on an optimization toward a well-defined and measurable figure of merit. For self-assembly problems, figures of merit based on structural data are a natural choice, and several ID strategies focused on the structure have been developed within the field of mesoscale modeling. For instance, biomolecular or polymeric coarse-graining techniques aim to reproduce desired structural signatures of an all-atom simulation using a reduced dimensionality model, typically where a single bead represents multiple atoms.^{15–18} We refer to the desired structures (e.g., from the all-atom description in the prior case) to be emulated by the ID interaction potentials as target configurations. For the application of this framework to self-assembly problems, we generate an ensemble of target configurations via a simulation where we apply (often complex and unphysical) many-body constraints to the particles in order to ensure that the desired structural motifs are present. We then systematically determine the ID interactions, i.e., the simplified interactions that best

^{a)}Electronic mail: truskett@che.utexas.edu

reproduce the structure of the target configurations according to the figure of merit. The optimization used for this purpose proceeds in an iterative fashion, where the difference between the data derived from the target simulation and those from the ID interaction simulation informs the update to the potential to be used in the next iteration. The scheme proceeds until either satisfactory convergence is achieved or the desired property is realized. The overall strategy is summarized in Fig. 1.

The techniques that have been developed in the coarse-graining community for discovering pair interactions that will best reproduce the target structure are typically aimed at matching the probability distributions of configurations in the target and ID simulations^{15,19,20} or the forces on the particles in those two simulations.^{15,21} While such methods have enjoyed some use in sophisticated coarse-graining applications, they have not been typically applied to self-assembly problems. In this work, we will primarily minimize the relative entropy (RE), equivalent to the well-known maximum likelihood approach in probability and statistics, in order to carry out our ID calculations.^{19,20} There are many advantages of the RE approach, including a rigorous mathematical framework for the method, well-behaved convergence of the update scheme, and the ease of employing constrained functional forms for the ID interactions. This last point is particularly relevant to colloids, given the many simplified physics-based functional forms that exist for computational modeling of such interparticle interactions.^{1,2} More generally though, constrained functional forms allow one to (1) limit the number of features, (2) impose bounds on parameters, (3) implement a well-defined and smooth cutoff, and (4) easily optimize even if the features of the radial distribution function are much longer-ranged than the interaction.

Of the many possible tunable parameters available, the focus of this paper is on the discovery of isotropic pair interactions that self-assemble into interesting structures, though the maximum likelihood formalism is not restricted to such cases. There are several benefits to discovering pair interactions for a particular self-assembly problem. First, optimization of pair interactions is more computationally tractable than more complex forms. If a pair interaction exists that self-assembles into a desired structure, it can be taken as a proof of realizability with more complex (3-body or higher) interactions as well. Moreover, pair interactions can provide high-level, interpretable insights into the physics necessary to achieve a particular structure. Relevant to the present work, it has been established that

isotropic pair potentials, particularly those possessing competing attractions and repulsions, can self-assemble into a variety of microphase-separated states, such as clusters, columns, and lamellae.^{22–35} Moreover, open lattices, such as diamond in three dimensions and honeycomb in two dimensions, have been assembled from *purely repulsive* isotropic pair potentials.^{10–13,36} The results presented herein not only demonstrate the utility of ID towards self-assembly problems but also help to further explore the limits of assembly via pair interactions.

In this article, we first describe a generally applicable ID methodology for particle self-assembly. The approach is technically simple to implement, particularly for pair interactions. Beginning from a maximum likelihood formulation, we provide an overview of the update scheme, as well as provide additional computational details in Sec. II. In Secs. III–V, we present ID results (in conjunction with a relevant background material) for several classes of particle assemblies: clusters, pores, and crystalline lattices, respectively. Some prior results are recalled for context, but in large the results presented here are new and address various proposed extensions of the prior work encompassing (1) result generality, (2) extensions to different interaction classes, and (3) the ability of the scheme to handle increasing target complexity. Finally, in Sec. VI, we conclude and comment on various extensions of RE minimization towards more complex design goals.

II. COMPUTATIONAL CONSIDERATIONS

The location of each particle i is described by a D -dimensional vector \mathbf{r}_i , and a full configuration of N particles comprises one such vector, $\mathbf{R} \equiv [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N]$. As discussed in the Introduction, a suitable optimization target is required for ID. To this end, creating a target ensemble entails formulating a target probability distribution, $P_{\text{tgt}}(\mathbf{R})$, from which an ensemble of particle configurations can be sampled that embody the desired microstructural organization. In this work, we employ target models motivated by statistical mechanics: $P_{\text{tgt}}(\mathbf{R})$ is the Boltzmann distribution governed by many-body potentials or external fields with some arbitrary notion of a “temperature.” While the interactions used to produce the target distributions are themselves physically unrealistic, they are tailored such that the desired organizational motifs dominate the target ensemble. More details on the constraints employed in each individual case are provided in Secs. III–V.

A. Probabilistic machine learning

For systematic optimization of the dimensionless tunable interaction parameters θ of the ID potential, we exploit maximum-likelihood model fitting, a strategy commonly adopted in probabilistic machine learning and formally equivalent to the relative entropy approach of Shell and co-workers.^{19,20} We provide a brief review of the fundamental principles behind the method below.

The role of maximum-likelihood is to maximize the probability (or equivalently, the log probability) that a classical, particle-based system at thermal equilibrium will sample the same configurations generated by $P_{\text{tgt}}(\mathbf{R})$. Here we consider the case of constant particle number (N), volume (V), and temperature (T), whereby the probability³⁷ for realizing any single

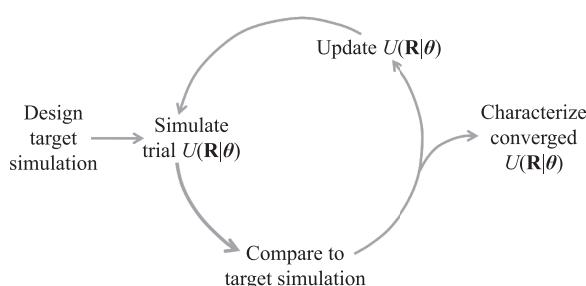


FIG. 1. Schematic representation of an iterative Inverse Design (ID) procedure, where $U(\mathbf{R}|\theta)$ indicates any general interaction potential for a given particle configuration, \mathbf{R} .

configuration, \mathbf{R} , conditioned on θ , is simply the Boltzmann weight

$$P(\mathbf{R}|\theta) \equiv \frac{\exp[-\beta U(\mathbf{R}|\theta)]}{Z(\theta)}, \quad (1)$$

where $U(\mathbf{R}|\theta)$ is the tunable potential energy for configuration \mathbf{R} , and

$$Z(\theta) \equiv \int d\mathbf{R} \exp[-\beta U(\mathbf{R}|\theta)] \quad (2)$$

is the configuration partition function. With these definitions, the quantity we seek to maximize is³⁶

$$\langle \ln P(\mathbf{R}|\theta) \rangle_{P_{\text{tgt}}(\mathbf{R})} = -\langle \beta U(\mathbf{R}|\theta) \rangle_{P_{\text{tgt}}(\mathbf{R})} - \ln Z(\theta). \quad (3)$$

Eq. (3) is effectively a measure of the “overlap” of $P(\mathbf{R}|\theta)$ with $P_{\text{tgt}}(\mathbf{R})$. In fact, maximizing Eq. (3) is equivalent to minimizing the Kullback-Leibler (KL) divergence—an information theory measure of the “distance” between two distributions.³⁸ In the statistical mechanical coarse-graining literature, the RE approach is often formulated in terms of minimizing the KL divergence with respect to θ ^{15,19} and is therefore equivalent to maximum-likelihood learning.

In order to carry out the maximization, we employ a gradient ascent optimization algorithm,

$$\theta^{(k+1)} = \theta^{(k)} + \alpha [\nabla_\theta \langle \ln P(\mathbf{R}|\theta) \rangle_{P_{\text{tgt}}(\mathbf{R})}]_{\theta=\theta^{(k)}}, \quad (4)$$

where k indexes the iteration step, and α is the empirically determined learning rate chosen to yield a stable optimization. Utilizing Eq. (3) along with the definition of the canonical probability distribution (Eq. (1)) and the partition function (Eq. (2)), we find that

$$\begin{aligned} \nabla_\theta \langle \ln P(\mathbf{R}|\theta) \rangle_{P_{\text{tgt}}(\mathbf{R})} &= \langle \nabla_\theta \beta U(\mathbf{R}|\theta) \rangle_{P(\mathbf{R}|\theta)} \\ &\quad - \langle \nabla_\theta \beta U(\mathbf{R}|\theta) \rangle_{P_{\text{tgt}}(\mathbf{R})}. \end{aligned} \quad (5)$$

Iterations are stopped once a sufficiently small gradient has been achieved.

B. Specific strategy for pair interactions

In this paper, we focus exclusively on the case of single component systems with isotropic pair interactions,

$$U(\mathbf{R}|\theta) \equiv \sum_{i=1}^N \sum_{j=i+1}^N u(r_{i,j}|\theta), \quad (6)$$

where $r_{i,j} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between particles i and j . As a consequence, only the isotropic pair correlations contained in $P(\mathbf{R}|\theta)$ or $P_{\text{tgt}}(\mathbf{R})$ are required to compute the averages in Eq. (5); in general, when the averaged quantity depends on fewer degrees of freedom than supported by the full probability distribution, only the reduced (marginalized) probability distribution over those degrees of freedom is required to perform the average.

Therefore, plugging Eq. (6) into Eq. (5) and summing identical terms in the large N limit yields

$$\begin{aligned} \nabla_\theta \langle \ln P(\mathbf{R}|\theta) \rangle_{P_{\text{tgt}}(\mathbf{R})} &= \frac{N^2}{2} \left[\langle \nabla_\theta \beta u(r|\theta) \rangle_{P(r|\theta)} \right. \\ &\quad \left. - \langle \nabla_\theta \beta u(r|\theta) \rangle_{P_{\text{tgt}}(r)} \right], \end{aligned} \quad (7)$$

where $P(r)$ are the isotropic two-particle probability distributions quantifying the probability that a pair of particles is a

radial distance r apart. Formally, $P(r)$ are obtained by integrating the full $P(\mathbf{R})$ over the $N - 2$ excess particle degrees of freedom as well as the orientation degrees of freedom between the remaining particle pairs. Since all particles are identical, any chosen pair of particles yields the same $P(r)$. These reduced probability distributions can be defined in terms of radial distribution functions [$g(r)$] as $P(r) \equiv s(D)r^{D-1}g(r)/V$, where $s(D)$ is the surface area of a unit radius D -dimensional sphere yielding

$$\begin{aligned} \nabla_\theta \langle \ln P(\mathbf{R}|\theta) \rangle_{P_{\text{tgt}}(\mathbf{R})} &= \frac{s(D)}{2} \rho N \int_0^\infty dr r^{D-1} [g(r|\theta) \\ &\quad - g_{\text{tgt}}(r)] \nabla_\theta \beta u(r|\theta), \end{aligned} \quad (8)$$

where $g(r|\theta)$ is the optimized $g(r)$, $g_{\text{tgt}}(r)$ is the target $g(r)$, and ρ is the number density. Thus, the final iterative update framework for *pair interactions* can be succinctly stated as

$$\begin{aligned} \theta^{(k+1)} &= \theta^{(k)} + \alpha \int_0^\infty dr \frac{r^{D-1}}{\sigma^D} [g(r|\theta^{(k)}) \\ &\quad - g_{\text{tgt}}(r)] [\nabla_\theta \beta u(r|\theta)]_{\theta=\theta^{(k)}}, \end{aligned} \quad (9)$$

where σ is a length scale that renders the right hand side, in particular α , dimensionless.

In the limit of optimizing a potential that has been infinitesimally discretized over a finite range in r , the zero-point solution to Eq. (8) is that where $g(r|\theta) = g_{\text{tgt}}(r)$ for all r in the relevant range. Therefore, any other iterative update scheme that converges to an exact matching in $g(r)$ is equivalent to RE minimization under the aforementioned conditions. Iterative Boltzmann Inversion (IBI) is one such heuristic approach that is commonly used in the literature on statistical mechanical coarse-graining.^{15–17} The associated update scheme is

$$u^{(i+1)}(r) \equiv u^{(i)}(r) + \alpha k_B T \ln \left[\frac{g^{(i)}(r)}{g_{\text{tgt}}(r)} \right], \quad (10)$$

where α is again a tunable parameter to control the magnitude of the update. IBI can be technically simpler to implement than unconstrained RE calculations, and therefore, we often perform IBI calculations first to inform our choice for the potential form in the (constrained) RE optimizations.

C. Simulation and analysis

All molecular dynamics (MD) simulations were performed in the canonical ensemble using the GROMACS molecular dynamics package,^{39,40} version 4.6.5 or 5.0.6 using a time step of $dt/\sqrt{\sigma^2 m \beta} \approx 0.001$ (m is the particle mass and $\beta \equiv k_B T$, where k_B is Boltzmann’s constant and T is the temperature) with leapfrog integration. A velocity-rescaling thermostat was used to control T with a characteristic time constant $\tau = 100dt$; the choice of T is arbitrary because the optimization result is the reduced pair interaction, $\beta u(r|\theta)$ which is independent of optimization temperature. Periodic boundary conditions were employed in three dimensions for all simulations, except for the two-dimensional lattices, where boundaries were periodic in the x- and y-directions only. IBI calculations were performed via the Versatile Object-oriented Toolkit for Coarse-graining Applications (VOTCA) package,^{41,42} through its GROMACS interface. Initial configurations corresponded to homogeneous fluids, and initial

guesses for the pair interaction were generally weakly interacting; see Equation (A3) in the Appendix for one example. As expected for an equilibrium statistical mechanics-based approach, the optimization is insensitive to the precise initialization details. Simulation visualizations were created with Visual Molecular Dynamics (VMD).⁴³ More specific details regarding each simulation—such as particle number and the use of positional restraints—are given in the Appendix.

In order to evaluate the success of the ID calculation, we analyze the simulations for fidelity to the targeted microstructure. In some cases, particularly for the lattice assemblies, this can be done by visual inspection. However, to quantify the size of microphase-segregated objects (i.e., clusters and pores), we calculate probability distribution functions to characterize these features. For analyzing the cluster size, we adopt a simple interparticle distance definition of neighboring particles and summarize the statistics in the form of cluster-size distributions (CSDs). Specifically, two particles are considered “connected” if their center-to-center distance does not exceed a cutoff, r_{cut} , or if they are joined via a continuous pathway of neighboring particles.^{44,45} In keeping with the prior work,^{29,46} we set $r_{\text{cut}} = 1.25\sigma$ in Sec. III and $r_{\text{cut}} = 1.33\sigma$ in Sec. IV, where σ is the particle diameter. The CSD [$P(n)$] is normalized such that it provides the probability that a randomly selected cluster comprises n particles. The presence of a *local maximum* in $P(n)$ at n^* occurring in the range $1 \ll n^* \ll N$ indicates a preferred aggregate size and thus the presence of clusters.

For the porous architectures, we first randomly insert test spheres with a diameter of 2σ into each configuration, keeping those that do not overlap with the particles of the configuration (they can overlap with themselves). The inserted test spheres provide a definition of the porous space in the configurations. Discrete void regions are identified by performing a cluster analysis on the test spheres as described above, where overlapping test spheres are taken to be neighbors. The volume of each pore (union volume of the test spheres) is computed via Monte Carlo integration, and the pore size distribution (PSD) is characterized as a function of pore diameter after the pore volumes are equated to the diameter of an equi-volume sphere.

III. CLUSTERS

A. Target simulation

Target configurations of fluids of “ideal” clusters (i.e., fluid-like particle aggregates that are roughly spherical in shape and monodisperse in size) were generated via Monte Carlo simulations of hard-core-like particles that are each assigned to clusters with a specified aggregation number and interact via a many-body potential as described in Ref. 46. In brief, the interaction has two components that stabilize clusters. The first is an energy penalty for each cluster that is quadratic in the deviation of its radius of gyration from the specified size, chosen to be sufficiently strong so as to generate well-defined, spherical clusters, but weak enough to maintain liquid-like, intracluster particle structure and mobility. The second component is an isotropic Yukawa repulsion between the centers of mass of any two clusters which effectively enforces a minimum distance between any two particles

in separate clusters. Empirically, it was found that larger clusters in the target ensemble require larger minimum separations between particles in adjacent clusters to stabilize the inverse design optimization.⁴⁶ In the present study, we focus on clusters that each contain 32 particles, the largest equilibrium aggregates investigated in our prior work.⁴⁶

B. Prior work

The short-ranged attractive, long-ranged repulsive (SALR) isotropic pair interaction type is frequently used in simulation studies of equilibrium cluster fluids—of broad interest due to both their interesting microstructures and the potential implications of clusters for bulk material properties, such as viscosity.^{47–53} This class of pair potentials is believed to provide a reasonable description of charge-stabilized colloids with short-range (e.g., osmotic depletion or hydrophobic) attractions. Typical ranges for the attractive interaction are 1%-10% of the particle diameter. The longer-ranged repulsive component of the potential is generally weak in amplitude and slowly varying, often represented as a (repulsive) Yukawa form.

Instead of performing a tedious forward search for desired clustered morphologies, we utilized ID to discover a class of isotropic pair interactions that assemble particles into highly idealized clusters of a prescribed size that (1) were monodisperse and spherical, (2) had marked spatial and temporal definition, and (3) possessed intra- and inter-cluster liquid-like mobility.⁴⁶ The resultant pair interactions (discovered via IBI) deviated from the traditional SALR model, possessing a broad attractive well with a narrow repulsive hump, as demonstrated in Fig. 2(a). We found that the location of the repulsive barrier is directly related to the cluster size, occurring at sufficiently large r that the barriers do not overlap with the other particles in the cluster, but rather collectively furnish a repulsive corona around it, analogous to the protective intercluster repulsion that was included in the target simulation.⁴⁶ Because (1) the interactions are isotropic and (2) the position of the repulsive

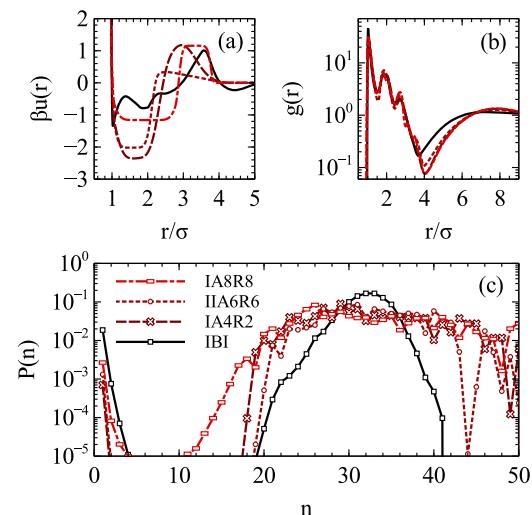


FIG. 2. (a) Dimensionless pair interactions, $\beta u(r)$, discovered via ID (IBI or RE) for equilibrium cluster fluids and their associated (b) radial distribution functions, $g(r)$, and (c) cluster size distributions $P(n)$. The naming conventions and details of potentials obtained via RE (i.e., IA8R8, IIA6R6, and IA4R2) are provided in the Appendix.

barrier and the cluster size is coupled, there is a direct relationship between the cluster diameter and the thickness of the resulting repulsive corona. This relationship explains why target simulations for bigger clusters required larger minimum separations between particles in different clusters to stabilize the ID scheme.

C. Results

From our prior work, several open questions remain. First, the IBI potential contains many secondary features in addition to the primary attractive well and repulsive barrier. We previously postulated that these features were not essential to equilibrium cluster fluid behavior. In the present work, we test this assumption by carrying out optimizations with constrained pair potential forms via RE minimization. Second, in prior work, we compared the clusters generated by our ID potential to those assembled with a more typical SALR model with a shorter-ranged attraction ($\sim 0.01\sigma$). We found that the latter clusters were less “ideal.” However, the SALR potential was chosen for the comparison via a limited forward search through the parameter space, not via a targeted ID optimization, and therefore it does not likely represent the optimal SALR form for an ideal clustering behavior. To remedy this shortcoming here, we constrain the width of the attractive well in the RE calculation to yield an optimized potential that more closely resembles a prototypical SALR interaction.

To test whether particles interacting via a single attractive well and repulsive barrier can form an equilibrium fluid of ideal clusters, we used RE minimization to design potentials with appropriately constrained functional forms of this type (IA4R2, IIA6R6, and IA8R8), the specifics of which—and naming convention for—are provided in the [Appendix](#). The potentials mimic a square well followed by a square shoulder at larger r ; larger numbers in the name indicate more square-like features for the attraction (A) and the repulsion (R), respectively. The Roman numeral indicates the presence (II) or absence (I) of a linear ramp term in the interaction, which lends an asymmetry to the repulsive barrier. Free to vary the range and amplitude of the attractive and repulsive components of the pair potential, the optimization finds interactions which exhibit a broad attractive well in concert with a significant repulsive hump (see Fig. 2(a)). Simulations verify that particles interacting via the resulting pair potentials form clusters with good center-of-mass mobility and internal fluidity. While the constraints imposed by the functional forms employed in the RE calculation necessitate degraded matching between the target and optimized $g(r)$ relative to IBI, $g(r)$ for each system of particles (Fig. 2(b)) shows the characteristic structural features of a strongly clustered fluid: a broad liquid droplet regime followed by an interfacial depletion zone and then low-amplitude oscillatory cluster-cluster correlations. All are similar to $g(r)$ of cluster fluids of potentials designed from IBI—which nearly perfectly reproduces $g_{tgt}(r)$ —but deviations, particularly in the interfacial depletion zone, are evident.

Analysis of the CSD for each system (Fig. 2(c)) indicates that the more restricted interactions were able to generally reproduce the desired microstructure of a clustered fluid with aggregates of order 30 particles. Each potential yields crisp monomer-cluster separation and closely achieves the desired

size, with average cluster sizes of $\langle n \rangle = 32.8, 33.3$, and 32.9 for the IA4R2, IIA6R6, and IA8R8 potentials, respectively. In terms of polydispersity though, a notable increase relative to the IBI potential is evident. This discrepancy may be ascribed to the ability of the IBI optimization to tune the positions in r of the primary attractive well and repulsive barrier independently of their ranges, allowing for greater precision in the repulsive shell surrounding the cluster. Nonetheless, the above results demonstrate that only a single attractive well followed by a repulsive barrier are strictly necessary for the assembly of clusters with a reasonably well-controlled size for potentials of this type.

While this new interaction class seems to provide generally enhanced clustering relative to many previously reported SALR interactions, it is possible that a more traditional SALR interaction form could also lead to more idealized clusters if purposefully engineered via ID to do so. To address this, we have optimized the IIA6R6 potential with the attraction range constrained to 0.25σ . As a consequence of this additional constraint, the intra-cluster droplet structure deviates slightly and the interface region between clusters is less strongly depleted of particles (see the radial distribution functions in Fig. 3(a)). Furthermore, the repulsion optimizes to a more canonical SALR form with a weak, long ranged repulsive ramp, as shown in Fig. 3(b). As evident from the CSDs in Fig. 3(c), clustering is much weaker in the constrained IIA6R6 system when compared to the unconstrained analog. Specifically, an enhanced monomer is found in the former in tandem with a weaker monomer-cluster bimodality. Despite the overall decreased performance with constraining the width of the attractive well, the ID methodology is still able to closely target the desired

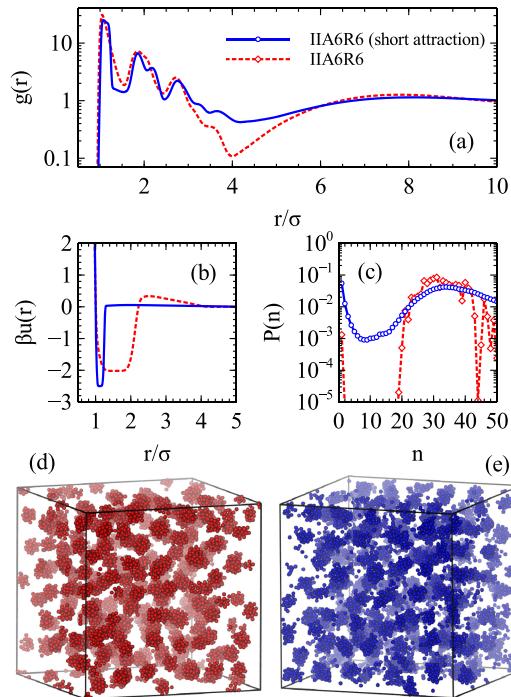


FIG. 3. For an equilibrium clustered fluid target, (a) the radial distribution functions, (b) dimensionless pair interactions, $\beta_u(r)$, (c) their associated CSDs, and representative particle configurations for the IIA6R6 functional form obtained via RE in the absence (d) or presence (e) of a constraint on the breadth of the attractive well (fixed at 0.25σ).

cluster size, achieving a mean aggregate size of $\langle n \rangle = 34.4$, though polydispersity is visibly enhanced. The lower quality of clustering is also evident visually by comparing snapshots of the unconstrained and constrained IIA6R6 systems (Figs. 3(d) and 3(e), respectively).

On the whole, there are several advantages in extending inverse calculations from an unrestricted potential (as in IBI) to a potential with a prescribed functional form. Though the IBI calculations are valuable to discern whether it is at all possible to assemble a desired structural motif with a pair interaction, it is not realistic to expect that every detail present in the unconstrained interaction is realizable in an experimental system. By contrast, the functional forms encoded here possess only two characteristic length scales that are coupled in r . It is conceivable to combine one source of interparticle attraction (e.g., depletion¹ or hydrophobic effect⁵⁴) and one source of repulsion (e.g., electrostatics^{1,55} or sterics¹) in a colloid- or nanoparticle-based to arrive at qualitatively similar interparticle potentials. Ultimately, the discovery of multiple, relatively simple pair interactions that form clusters sets the stage for the use of physics-based models in RE calculations, where physically motivated constraints can be applied to parameters that bear direct relation to experimental conditions.

IV. PORES

A. Target simulation

In Refs. 29 and 30, target ensembles of particles with morphologies consistent with a porous mesophase were fabricated via an MD simulation of a binary hard sphere-like mixture with a large size asymmetry (here, a 4:1 diameter ratio). Large particles—which provided spherical “pore” zones of excluded volume—were fixed to FCC lattice positions in a sea of mobile small particles that constitutes the surrounding matrix. Within this framework, the target ensemble is regarded as the set of small particle configurations generated via the molecular simulation in the presence of the immobile exclusion spheres. Based on our prior work on clusters that demonstrated the importance of constructing a target simulation that can be reproduced by pair interactions,⁴⁶ we placed the large spheres on a lattice with a nearest neighbor distance such that the thickness of the matrix surrounding the pore is commensurate with the pore diameter. By choosing a pore nearest neighbor distance to pore diameter ratio of 1.85, most particles will be positioned such that a repulsive barrier falling near to the pore diameter will function to stabilize multiple pores.²⁹ This is, again, a consequence of the isotropic nature of the pair interactions, such that the repulsive barrier helps to form a spherical corona around the particle center.

B. Prior work

Unlike clustered fluid phases—which have been extensively studied, especially within the context of SALR pair potentials—self-assembly of the conjugate “inverse” cluster phase (also known as the bubble, void, or pore phase) has largely remained elusive, especially in 3D. Intriguing theoretical studies suggested the possibility of pore formation with a simple pair potential comprising competitive attractive and repulsive components.^{32–35} Further support for this notion has

come in the form of 2D lattice models^{56,57} and continuum field theories,⁵⁸ and experimental, quasi-2D interfacial colloid assembly.^{59,60} Prior to the application of ID, however, the closest simulated 3D pore assembly resulted in a single large void within a simulation cell, as specific control over the pore size was not yet possible.³¹ Recently, leveraging ID, we discovered interactions that generate an assembly of spherical pores with a prescribed size.^{29,30} As shown below, the unconstrained IBI optimization yielded an interaction with competing attractive and repulsive components, but also with deviations from the canonical SALR model in a similar fashion to the previously discussed cluster interactions. The previous study also demonstrated the relative insensitivity of the porous assembly to the secondary features in the pair potential: the same constrained functional forms employed in Sec. III C were found to assemble porous structures upon optimization with RE.

Application of the proposed ID methodology not only enabled the discovery of a 3D assembly of pores, but also yielded rich phase behavior akin to that found in block copolymers: by reducing the packing fraction, a phase progression from pores → void columns → bi-continuous → lamellar → particle columns → clusters was discovered,²⁹ validating an array of early theoretical predictions and suggestive simulations using competing interactions.

C. Results

Prior work has shown that competing attractions and repulsions are *not* strictly required in a pair potential to allow particles to form microphase segregated structures: some purely repulsive interactions have been predicted to stabilize clustered phases, for instance.^{61–63} Via application of RE minimization, we discover here a purely repulsive pair potential that assembles particles into a porous mesophase, as can be seen in Fig. 4(a). In Fig. 4(a), lighter spherical-like regions from which particles are absent are interspersed among a fluid of dark blue particles. (The technical implementation of the pair potential constraint applied in the RE optimization is discussed in the Appendix.) Despite the absence of an attractive component, reasonable control over the pore size is achieved; the pores are only slightly smaller than the pores targeted or generated via IBI. Moreover, polydispersity is not significantly altered with respect to the unconstrained (IBI) structures (Fig. 4(b)). The pores are arranged on a BCC lattice; the pore-pore $g(r)$ is shown in Fig. 4(c).

In addition to achieving pore assembly, the purely repulsive pair interaction also exhibits a qualitatively similar block copolymer-like phase progression discovered in the unconstrained system, with all of the same microphase-segregated states (clusters, columns, lamellae) present at some η . An image of the clustered phase is shown in Fig. 4(d); this structural motif emerges near $\eta = 0.07$ —identified as the value of η for which the CSD indicates the strongest preference for clustering (shown in Fig. 4(e)). Compared to the clustered fluids explored in Sec. III, there is a greater degree of monomer and other small aggregates present, which reduces the magnitude of the peak associated with the clusters in the CSD. However, the breadth of the main peak in the CSD indicates that the monodispersity of the clusters is comparable to that seen in Sec. III for the unconstrained (IBI) potential.

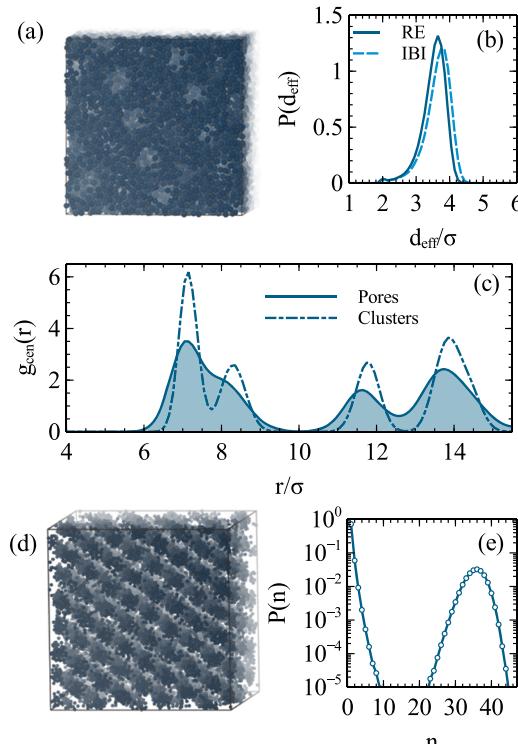


FIG. 4. (a) Representative configuration of a porous assembly generated by the purely repulsive pair interaction obtained via RE, and (b) comparison of the pore size distributions (PSDs) resulting from RE and IBI. (c) The pore-pore and cluster-cluster $g(r)$ with the RE repulsive pair interaction at packing fractions of 0.31 and 0.07, respectively. (d) Representative configuration of the ordered clustered phase from (c), and (e) the CSD at this state point.

Visual inspection of the clustered configuration in Fig. 4(d) reveals that, unlike in Sec. III, the clusters are arranged on a lattice. Interestingly, comparison of the peak positions in the cluster-cluster and pore-pore radial distribution functions in Fig. 4(c) indicates that the microphase-separated objects (either clusters or pores) are positioned in essentially the same crystalline structure: a BCC lattice with an approximately equal nearest neighbor distance. It is interesting that the general pore-cluster symmetry extends to such quantitative details as well.

The ability of a purely repulsive interaction to elicit nearly identical assembly to one with competing attractive and repulsive contributions may seem surprising (see Fig. 5(a)); however, examining the potentials in the Fourier space reveals hidden similarities. Minima in the Fourier transform of a pair interaction, $u(k)$, indicate energetically preferred length scales

for structural correlations, as a minimum in $u(k)$ promotes a corresponding maximum in the structure factor, $S(k)$, at the same wave-vector k .⁶⁴ In this case, the analysis requires extracting the non-core, external contribution to the potential, defined here as $w(r) \equiv 0$ for $r \leq \sigma$ and $w(r) \equiv u(r)$ for $r > \sigma$. The core contribution is omitted because (1) the core amplitude is very large and will otherwise dominate the Fourier transform, and (2) we are interested in how the potential will modify the structure in excess of the geometric constraints imposed via a steeply repulsive core. The Fourier transform of $\beta w(r)$ and $\beta w(k)$ is shown in Fig. 5(b), where it is clear that the low- k behavior of both potentials is rather similar. Both have a strong energetic preference towards intermediate range (IR) wave-vectors, the two lowest of which are at virtually identical k . While the presence of energetically preferred IR wave-vectors in $\beta w(k)$ is not sufficient to guarantee pore formation, if pores are assembled, they should have similar packing correlations and characteristic length scales—as we have found here to be true.

The capability of a purely repulsive interaction to elicit almost identical behavior to an unconstrained counterpart is a testament to the interchangeability of attractions with regions of low force in the interaction and augmented pressure. As can be seen in Fig. 5(a), the repulsive constraint yields plateaus in the pair potential over distances whereby the unconstrained analog has attractions. Such plateaus are even more obvious in the case of crystals; see Sec. V. Coupled with external applied pressure to maintain an appropriate density, these plateaus yield preferential particle-particle packing length scales, akin to attractions, as the rest of the potential acts to force particles apart.

V. CRYSTALS

A. Target simulation

As in Ref. 36, target crystal ensembles were created by tethering *non-interacting* particles to the desired lattice positions via a harmonic constraint in an MD simulation. Relatively strong spring constants were employed (i.e., the “temperature” of the simulation was cold), yielding highly localized crystalline configurations and a sharp coordination shell structure in $g(r)$. Strong localization (1) helps ensure that a stable solution to the optimization exists and (2) mitigates, or removes entirely, any competition from a non-crystalline (fluid) solution to the maximization. However, some fluctuations about the lattice positions are necessary such that $g(r)$ is integrable in the RE update scheme.

B. Prior work

Research on the self-assembly of crystalline lattices from pair potentials has spurred the development of many 1D strategies, most of which employ a ground state (zero-temperature) calculation wherein the target lattice is stabilized relative to an explicit pool of competitor crystals.^{10–13} Neither of these attributes are optimal, as (1) ground state stability does not guarantee any appreciable stability upon heating, and (2) a complete pool of competitor crystals can only be asymptotically approached. As a consequence, post optimization

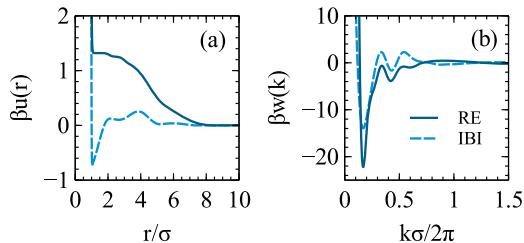


FIG. 5. (a) Pair interactions, $\beta u(r)$, and (b) their Fourier space representations, $\beta w(k)$, optimized with ID for a porous target for the unconstrained (IBI) and purely repulsive interaction (RE) cases. See the main text for the definition of $w(r)$.

simulation of the designed interaction often yields an undesired structural motif that was not included in the competitor pool—necessitating another optimization with an augmented pool. Despite these limitations, such approaches have had notable successes leading to the discovery of purely repulsive pair interactions capable of stabilizing a diverse array of complex, low coordinated crystals in 2D and 3D.

The present methodology utilizes the finite-temperature MD simulations performed at each optimization step to provide an implicit pool of competitors via the self-assembled structures arising during the MD simulations as well as the fluid state that exists prior to crystallization. Once assembled, however, lattices can possess a significant kinetic stability even if they are not thermodynamically stable. Therefore, after self-assembly occurs within the optimization, the resulting potentials from future iterative steps are no longer guaranteed to self-assemble into the desired structure from a fluid. To remedy this, each new optimization step included a heating cycle to melt the crystal, followed by a slow cooling back to the working optimization temperature. Recently, this methodology was employed to discover a variety of purely repulsive pair interactions that each self-assembles particles into a two-dimensional lattice in finite-temperature MD simulations.³⁶ Honeycomb, kagome, square, rectangular, truncated square, and truncated hexagonal crystals were all successfully assembled via ID potentials; the latter two structures had not been realized previously with any pair interaction.

C. Results

Issues of practical realizability aside, we wish to explore the limits of crystal self-assembly within the proposed methodology. For context, we first recall the truncated hexagonal (TH) lattice (shown in Fig. 6(a)) assembled in Ref. 36. While this lattice is composed of identical sites, it is challenging to self-assemble with an isotropic pair interaction given the high distributional inhomogeneity of particles about one another—yielding only two-fold reflection symmetry, as well

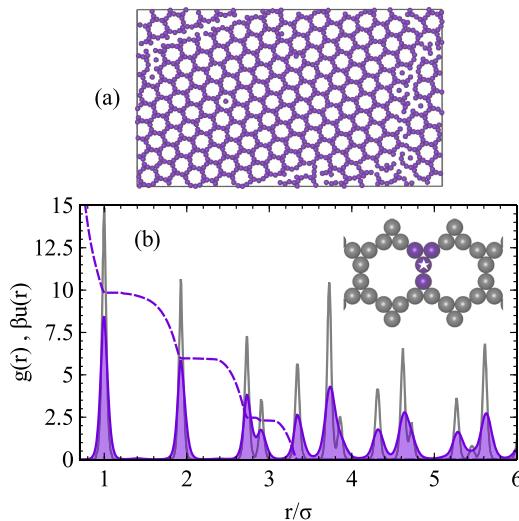


FIG. 6. For the TH lattice, (a) a representative configuration at the end of the ID optimization and (b) the ID optimized pair interaction, $\beta u(r)$ (dashed), and comparison of $g(r)$ given by $\beta u(r)$ (filled) and $g_{tgt}(r)$ (no fill). The inset depicts the ideal TH lattice positions, where the first coordination shell of a selected particle (star) is highlighted in purple.

as a high degree of angular asymmetry, in the first coordination shell (see the inset of Fig. 6(b)). Moreover, there are a variety of competitor phases that are relevant, owing to the complexity and openness of the lattice. In particular, an alternative ID approach based on ground state computations required the presence of 41 distinct structural motifs in the competitor pool to stabilize the TH lattice.⁶⁵ Despite these difficulties inherent to the TH lattice, the ID scheme easily discovered an appropriate repulsive pair interaction for the assembly of the TH lattice; an interaction possessing three broad features was required, spanning five coordination shells (see Fig. 6(b)).

A more complex target lattice is the truncated trihexagonal (TTH) lattice (shown in Fig. 7(a)). The TTH lattice possesses no symmetry whatsoever in the first coordination shell (see the inset of Fig. 7(b)). To our knowledge, no lattice possessing C_1 symmetry with respect to its first coordination has ever been assembled via an isotropic pair interaction. Direct application of the approach outlined in Sec. V B yielded assemblies that, while locally bearing resemblance to structural elements present in the TTH lattice, were malformed and devoid of any TTH unit cells.

Suspecting that slow nucleation kinetics were to blame for the failure to assemble the TTH lattice, we integrated a crystalline seed into the optimization protocol to accelerate self-assembly.^{66,67} The seeded particles are shown in gray in Fig. 7(a). To ensure that the simulated $g(r)$ is smooth and integrable, the seeded particles are not frozen but rather are tethered to lattice positions via a *weak* harmonic constraint. Because the method ultimately requires only the simulated $g(r)$ to update the interparticle interactions, it has the advantage that the optimization scheme remains valid under a variety of simulation protocols (e.g., constraints and fields), so long as the resulting $g(r)$ is integrable over the range of potential. With the crystalline seed, the ID methodology discovered a pair interaction that assembles into TTH. Ultimately, the self-assembly of TTH required that a minimum of 19 coordination shells be encompassed by the pair interaction (Fig. 7(b)). While such

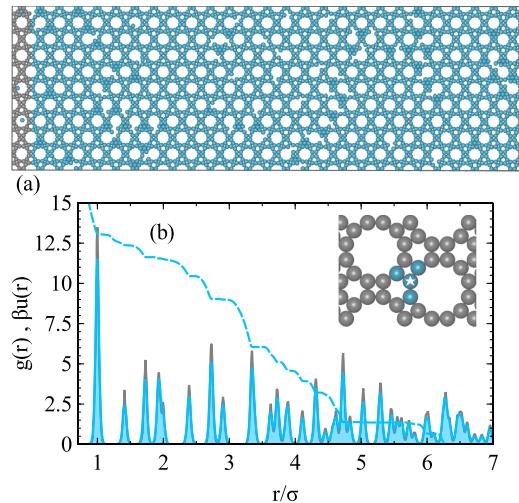


FIG. 7. For the TTH lattice, (a) a representative configuration at the end of the ID optimization where the seeded particles are shown in gray, and (b) the ID optimized pair interaction, $\beta u(r)$ (dashed), and comparison of $g(r)$ given by $\beta u(r)$ (filled) and $g_{tgt}(r)$ (no fill). The inset depicts the ideal TTH lattice positions, where the first coordination shell of a selected particle (star) is highlighted in blue.

an interaction is highly unlikely to be physically realizable, the complexity of potential required for the self-assembly of TTH is a testament to both the difficulty of the target and the robustness of the ID scheme.

The ability to incorporate a seed in the ID scheme is a valuable asset both in terms of accelerating the *in silico* assembly of lattices with slow nucleation kinetics and in relation to the experimental work that may intentionally employ crystalline seeds. However, implications on the thermodynamic stability for the desired lattice in the absence of the seed are unclear. To address this, we generated a small ensemble of unseeded, self-assembled structures by slowly cooling five simulations from a disordered fluid to $T = 0$ with particles interacting via the potential from Fig. 7. The structure shown in Fig. 7 was similarly cooled in the absence of the harmonic constraints on the seed particles. In the absence of a seed during self-assembly, highly malformed structures developed, possessing an overabundance of hexagonal motifs and a significant dearth of dodecagons. The seeded structure, on the other hand, remains as a TTH lattice, even after the harmonic constraints on the seed particles were removed. The malformed assemblies are uniformly higher in energy than the TTH lattice, strongly suggesting that the ground state for this potential, with or without the seed, is TTH.

Despite large structural differences between the seeded and unseeded assemblies, the difference in energy at $T = 0$ is small, with the malformed assemblies only 0.7% higher in energy. With such subtle differences between the correct assembly and the multitude of incorrect alternatives, the role of the seed is clear. In essence, it provides a slight bias towards the desired structure in the vast landscape of malformed configurations that are similar in energy. Besides the computational benefits of seeded ID, this technique is also of direct relevance to the substrate and template-assisted growth whereby the “seed” takes on a literal role.

Because our ID scheme for crystals requires repeated on-the-fly self-assembly of the desired microstructure, slow nucleation kinetics are of particular concern for application

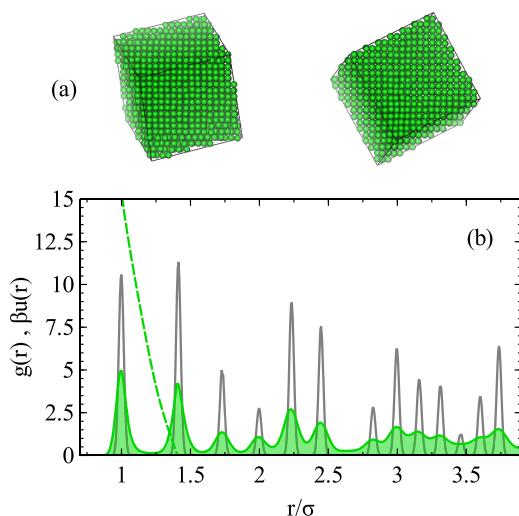


FIG. 8. For the SC lattice, (a) two views of a representative configuration at the end of the ID optimization, and (b) the ID optimized pair interaction, $\beta u(r)$ (dashed), and comparison of $g(r)$ given by $\beta u(r)$ (filled) and $g_{tgt}(r)$ (no fill).

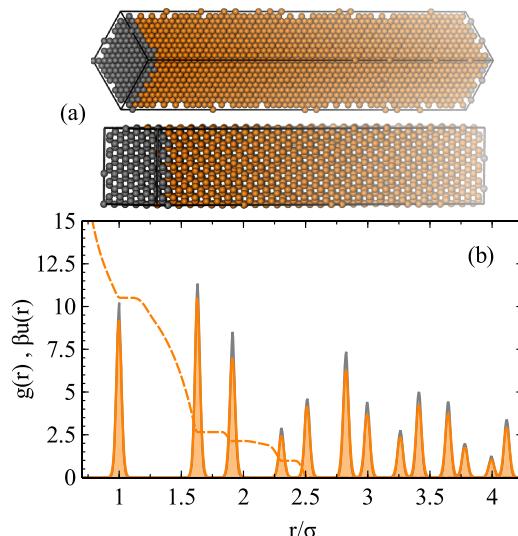


FIG. 9. For the diamond lattice, (a) two views of a representative configuration at the end of the ID optimization where the seeded particles are shown in gray, and (b) the ID optimized pair interaction, $\beta u(r)$ (dashed), and comparison of $g(r)$ given by $\beta u(r)$ (filled) and $g_{tgt}(r)$ (no fill).

to 3D lattices. As dimensionality increases, in general nucleation kinetics becomes slower and more particles are needed to mitigate potential finite size effects. Despite these issues, we discovered via ID a pair interaction that quickly and accurately self-assembles into a simple cubic crystal in the absence of a seed. As shown in Fig. 8, the assembly is almost perfect—only a small defect arises due to incorrect alignment with the simulation box. As with the square lattice (the 2D analogue to simple cubic), the simple cubic lattice is realized via a relatively simple potential comprised of a repulsive core and single shoulder spanning the first and second coordination shells. Another common target in the 3D crystal ID is the diamond lattice which, not surprisingly, we find more difficult to assemble given its relatively open and complex structure. Unseeded ID attempts led to malformed assemblies. Leveraging seeded growth, however, yields a well-assembled diamond lattice with the potential shown in Fig. 9. However, five coordination shells of structure were required to thwart the formation of non-diamond structures bridging the propagating crystalline fronts. Similarly to TTH, the optimized diamond potential did not form the correct lattice in the absence of a seed, in this case forming amorphous glassy states. Using a similar protocol to that employed for TTH above (see the Appendix), we confirmed that the glassy states are $\approx 10\%$ higher in energy than the seeded crystalline morphology, suggestive that slow kinetics for assembly thwart crystallization in the absence of a seed.

VI. CONCLUSIONS

In this article, we outlined a general, statistical mechanics based approach for the ID of interactions resulting in self-assembly and applied the methodology to discover isotropic, pair interactions that formed clusters, pores, and crystalline lattices at equilibrium. These applications emphasize a potential advantage of the method, wherein the discovery of these interactions is recast from a forward search through the

parameter space to the design of an appropriate target simulation to represent a desired structural motif. That is, in forward design studies, parameters must be tuned (often empirically) with the hope of eliciting a desired behavior; however, the number of parameters can be quite large and each can have a complex and non-intuitive effect on the structural properties. By leveraging ID, these challenges are forgone in favor of the challenges inherent to designing a suitable target ensemble. Often, these latter challenges are more amenable to physical intuition as opposed to brute force. For example, in the design of clusters and pores, judicious choice of a single length scale in the problem (the thickness of the repulsive shell around clusters, the pore-pore spacing, respectively) in the target simulations had profound effects on the realizability of each with a pair potential.

Having demonstrated the application of the proposed ID methodology in a variety of contexts, the stage is set to explore various extensions. For instance, of great importance for experimental practicality is the discovery of interactions that exhibit the desired structural motifs over larger ranges in particle concentration. Two directions seem profitable towards this end. The first is an approach borrowed from the machine learning community: flexible hyper-parameters (here, tunable parameters and potential parameters) could be regularized to maximize the model generality to conditions for which it was not trained. For example, penalizing large Akima spline step parameters (effectively penalizing large forces) would inevitably soften the observed shoulders for the lattice interactions, likely improving the density stability range. This would almost certainly come at the cost of diminished temperature stability,⁶⁸ so a compromise would have to be identified. The second more direct option is to leverage multi-state ideas from the biomolecular coarse-graining community. In this scenario, the probabilistic formalism must be extended to maximize the likelihood of realizing target configurations at an array of state points.

Beyond improving the flexibility of designed interactions, it is desirable to extend the approach to leverage anisotropic interactions and to directly target material properties as opposed to self-assembly. In regards to the former, patchy models would serve as a tractable starting point; possible variables include patch location, net attraction or repulsion between patch pairs, and the number of patches. Though widely studied via forward searches,^{3–7} the self-assembly of anisotropic particles has not been heavily investigated via inverse methods, making the area particularly a fertile ground for both the validation of new ID methods as well as discovery of novel anisotropic interactions.^{9,67,69} For the latter, an endless array of properties can be explicitly optimized, including the elastic modulus, viscosity, terminal jamming density upon compaction, single electron hopping conductivity, percolation threshold, and even fractal dimension of a percolated particle network structure. Such developments would nicely complement the self-assembly oriented approach outlined in this article.

Finally, we emphasize that the array of targets employed in this work is far from exhaustive; the only limit pertains to the creativity needed to construct a target ensemble and its realizability with a given potential interaction form. Only out

of convenience did we adopt statistical physics based models for $P_{\text{tgt}}(\mathbf{R})$ (i.e., states are sampled according to a Boltzmann weight with some fictive temperature and many-body potential). This need not be the case. Any means of generating an ensemble of configurations is valid as long as it has convergent large sample statistics (even if $P_{\text{tgt}}(\mathbf{R})$ is not explicitly known for the protocol). For example, random sequential addition (RSA) is a highly studied, non-equilibrium approach to packing spheres. It proceeds via sequential, individual random sphere additions to a volume with rejection if an overlap is generated. This yields a well-defined terminal packing density at which no more spheres can be added. Using ID, one could discover interactions that, at a fixed density and thermodynamic equilibrium, sample similar configurations to those in RSA.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation (Grant No. 1247945) and the Welch Foundation (Grant Nos. F-1696 and F-1848). We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources.

APPENDIX: ADDITIONAL COMPUTATIONAL CONSIDERATIONS

1. Constrained potential forms

In Section III, we leveraged RE minimization and constrained pair interactions to design potentials that assemble a fluid phase of equilibrium clusters, the definitions of which are

$$u_{\text{RE}}^{(i)}(r) \equiv -\epsilon_1 \exp\left[-\left(\frac{r-d-\alpha_1/2}{\alpha_1/2}\right)^{n_1}\right] + \epsilon_2 q_i(\alpha_1, \alpha_2) \exp\left[-\left(\frac{r-d-\alpha_2/2-\alpha_1}{\alpha_2/2}\right)^{n_2}\right], \quad (\text{A1})$$

where

$$q_i(\alpha_1, \alpha_2) \equiv \delta_{i,\text{I}} + \delta_{i,\text{II}} \left[\frac{-r+d+\alpha_1+\alpha_2}{\alpha_2} \right], \quad (\text{A2})$$

where $\delta_{i,j}$ is the Kronecker delta, $[\epsilon_1, \epsilon_2]$ are the [attractive, repulsive] strengths, and $[\alpha_1, \alpha_2]$ are the corresponding ranges. The extra multiplicative factor in the type II potential approximately converts the symmetric repulsive hump into a linear ramp. The potentials also contain two non-optimized exponents, $[n_1, n_2]$, the magnitude of each controls how square-like the attraction and repulsion are, respectively. For convenience, we adopt the compact notation, iAn_1Rn_2 , to convey the interaction type and the exponents employed.

The assembly of purely repulsive porous and crystalline phases in Secs. IV and V, respectively, employed a constrained Akima spline with n knots at locations $\{r_1, r_2, \dots, r_n\}$, where r_n is the cutoff for the potential, and a corresponding set of potential amplitudes $\{u_1, u_2, \dots, u_n\}$, where the Akima spline interpolates through. At each knot, we define a tunable parameter as $\theta_i = u_i - u_{i+1}$ (the difference in amplitude at the current knot relative to the next knot), the entire set of which defines our parameter optimization vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_{n-1}, u_n]$. Adopting this parametrization allows us to easily enforce monotonicity (by demanding all $\theta_i \geq 0$) and a smooth force cutoff (via $\theta_{n-1} = 0$ and $u_n = 0$).

As our starting guess for the Akima spline optimization, we utilize a power law form (weighted by a step-like smoothing function to minimize the force near and at that cutoff) to assign the initial θ based on the set of knot locations corresponding amplitudes extracted via

$$\beta u_0(r) \equiv \frac{1}{2} A \left(\frac{\sigma}{r} \right)^a \left(1 - \tanh \left[k \left(\frac{r}{\sigma} - \frac{r_s}{\sigma} \right) \right] \right), \quad (\text{A3})$$

where A is the dimensionless amplitude, k controls how rapidly the potential dies off, a is the power, r_s is the center of the smoothing function, and σ is the nearest neighbor crystal distance. As in a previous publication, we set $A = 1.8$, $k = 8.9$, $a = 5$, and r_s is set close to, but before, the potential cutoff. Empirically, we find that the optimization is not sensitive to the precise initial guess.

We employed the Akima spline $u(r)$ for the ID of both pore and crystal phases using respective knot spacings of 0.022 and 0.017. For pores, a cutoff of 9.0σ was used, whereas for crystals the cutoff was systematically chosen to include as few coordination shells as possible while still allowing for the correct lattice to form. The specific cutoffs used for truncated hexagonal, truncated trihexagonal, simple cubic, and diamond correspond to 3.49σ , 6.28σ , 1.56σ , and 2.59σ , respectively. For both pores and crystals, the cutoff knot locations are followed by two additional unoptimized knots with parameters set to zero to ensure a smooth vanishing of the force.

2. Additional simulation details

a. Clusters

Here we provide additional cluster simulation details regarding the assembly simulations only; specific details pertaining to the target structure can be found in Ref. 46.

Optimizing clusters required a significant number of iteration steps; however, each iteration is relatively inexpensive due to the low volume fraction and thus minimized number of interacting neighbors. In all cases, 8000 particles were simulated at each iteration and a total of 500 to 1000 iterations were necessary to reach satisfactorily small updates to the pair potential and corresponding structure. Each iteration employed a simulation of order 1×10^6 MD steps. The final $g(r)$ and CSD analysis used another 2×10^6 equilibration steps and 2×10^7 steps to accrue enough statistics to yield a broad temporal sample of the slow cluster exchange—of particular relevance to the CSD calculation.

b. Pores

For the porous structure, converged results were achieved via 40 optimization steps, each comprised of 8×10^5 MD steps. With the optimized potentials in hand, production runs of 11×10^6 steps were carried out, where configurations were sampled from the last 6×10^6 steps to generate statistics for the $g(r)$ and size distributions associated with the pores and clusters. Relative entropy calculations were performed with 5475 particles, and characterization was performed with 18478 particles—chosen so that at the optimization density, the box size was 1.5 times larger than in the optimization to check for finite size effects. More details on the target simulation for the pores can be found in Refs. 29 and 30.

c. Crystals

Here we provide additional crystal simulation details pertaining to both the target and assembly simulations. The 2D truncated hexagonal (TH) and truncated trihexagonal (TTH) crystal simulations contained 1200 and 3240 particles, respectively, with corresponding box dimensions of $37.27\sigma \times 64.55\sigma$ and $40.92\sigma \times 127.59\sigma$. 3D simulations of simple cubic (SC) and diamond (DIA) crystals employed 4096 and 6000 particles, respectively, with simulation box dimensions of $15.98\sigma \times 15.98\sigma \times 15.98\sigma$ and $11.53\sigma \times 11.53\sigma \times 69.19\sigma$. In both the target and seeded ID simulations, harmonic constraints were employed to tether particles to specified crystalline coordinate positions. Specifically, target simulations employed relatively strong spring constants of 2237, 2796, 6408, and $4793 \text{ } k_B T / \sigma^2$ for the TH, TTH, SC, and DIA crystals, respectively, yielding a nearest-neighbor peak in $g(r)$ with a magnitude of order 10. For the ID optimization of the TTH and DIA crystals, we used seeds comprising of 120 and 250 particles, respectively. They were designed to perfectly fill space in the short dimension(s) and provide one unit cell of structure in the long dimension. Like the target simulations, seeds utilized a harmonic constraint to maintain the crystal structure; however, the seed tethers were comparatively weak, possessing spring constants of $40k_B T / \sigma^2$ and $801k_B T / \sigma^2$ for TTH and DIA, respectively. The weaker restoring forces help to minimize the contribution to localization upon crystallization.

The computational demand of the optimization protocol can vary significantly from one crystal to another. For less challenging crystals, like SC and TH, roughly 10–20 optimization steps are sufficient to realize assembly. On the other hand, for more complex targets, like TTH and DIA, of order 100 iterations are required. In addition, each iterative step can vary in computational demand, depending on the nucleation kinetics. The cheapest was SC which used 6×10^5 MD steps starting from a disordered configuration. Next in terms of computational demand was the TH crystal; this used 8×10^6 total MD steps with half of the steps devoted to a slow, linear cooling back to the optimization temperature (T) from a hot initialization ($1.5T$). Finally, for the TTH and DIA crystals, 1.5×10^7 simulation steps were used to incorporate (1) a quick heat and slow cool initialization and (2) a sufficiently long equilibration so as to allow for the crystal to propagate across the long dimension from one seed face to another (periodic image). Specifically, for TTH and D we used corresponding temperature points $[5.0T, 1.8T, T, T]$ at corresponding MD steps $[0, 3 \times 10^5, 1 \times 10^7, 1.5 \times 10^7]$ with linear temperature changes in between. Post-optimization, long equilibration runs (of order 1×10^7 MD steps) for all crystals were performed to calculate $g(r)$ and extract configuration data for snapshots.

To study the effect of the seed in the TTH calculations, we adopted a combined equilibration and cooling protocol to generate the unseeded assemblies. First, we discovered the temperature range in which the optimized interaction crystallized in the absence of a seed. The range for crystallization was empirically found to approximately lie between $0.73T$ and $0.81T$ based off of a sharp change in the potential energy trace with temperature. Using this estimate, we

performed a combined heat, assemble, and equilibrate cycle with linear temperature changes between the sequential temperatures [5.0T, 0.81T, 0.73T, 0.73T] and corresponding simulation steps [$0, 3 \times 10^5, 1.5 \times 10^7, 6 \times 10^7$]. This was performed five times to yield a small ensemble of representative configurations. Each was then slowly annealed according to the specified temperatures [0.73T, 0.0001T, 0.0001T] and molecular dynamics steps [$0, 5.9 \times 10^7, 6 \times 10^7$] with potential energy data sampled from the final 1×10^6 steps. This final annealing protocol was also adopted for calculating the ground state energy of the seeded crystalline assembly after the removal of the harmonic seed constraints.

Similar calculations were performed for the optimized DIA potential in the absence of a seed. Unlike TTH, the DIA potential did not exhibit a well-defined crystallization temperature but rather forms amorphous glass states, and therefore we modified the above annealing protocol to achieve a slow cooling over the large temperature gap that encompasses the glassy transition. First, we performed a cycle with linear temperature changes between the following sequential temperatures [5.0T, 1.25T, 0.71T] and corresponding simulation steps [$0, 3 \times 10^5, 6 \times 10^7$]. Five separate configurations were generated and each slowly annealed according to the specified temperatures [0.71T, 0.0001T, 0.0001T] and molecular dynamics steps [$0, 5.9 \times 10^7, 6 \times 10^7$] with potential energy data sampled from the final 1×10^6 steps. This final protocol was also used to approximate the ground state energy of the seeded DIA crystalline assembly after the removal of the harmonic seed constraints.

- ¹J. Mewis and N. J. Wagner, *Colloidal Suspension Rheology* (Cambridge University Press, 2013).
- ²C. N. Likos, K. A. Vaynberg, H. Löwen, and N. J. Wagner, “Colloidal stabilization by adsorbed gelatin,” *Langmuir* **16**, 4100–4108 (2000).
- ³E. Bianchi, R. Blaak, and C. N. Likos, “Patchy colloids: State of the art and perspectives,” *Phys. Chem. Chem. Phys.* **13**, 6397–6410 (2011).
- ⁴Z. Zhang and S. C. Glotzer, “Self-assembly of patchy particles,” *Nano Lett.* **4**, 1407–1413 (2004).
- ⁵Q. Chen, S. C. Bae, and S. Granick, “Directed self-assembly of a colloidal kagome lattice,” *Nature* **469**, 381–384 (2011).
- ⁶C. H. J. Evers, J. A. Liiken, P. G. Bolhuis, and W. K. Kegel, “Self-assembly of microcapsules via colloidal bond hybridization and anisotropy,” *Nature* **534**, 364–368 (2016).
- ⁷Y. Mai and A. Eisenberg, “Self-assembly of block copolymers,” *Chem. Soc. Rev.* **41**, 5969–5985 (2012).
- ⁸M. Z. Miskin, G. Khaira, J. J. de Pablo, and H. M. Jaeger, “Turning statistical physics models into materials design engines,” *Proc. Natl. Acad. Sci. U. S. A.* **113**, 34–39 (2016).
- ⁹A. F. Hannon, Y. Ding, W. Bai, C. A. Ross, and A. Alexander-Katz, “Optimizing topographical templates for directed self-assembly of block copolymers via inverse design simulations,” *Nano Lett.* **14**, 318–325 (2014).
- ¹⁰S. Torquato, “Inverse optimization techniques for targeted self-assembly,” *Soft Matter* **5**, 1157–1173 (2009).
- ¹¹A. Jain, J. A. Bollinger, and T. M. Truskett, “Inverse methods for material design,” *AIChE J.* **60**, 2732–2740 (2014).
- ¹²E. Marcotte, F. H. Stillinger, and S. Torquato, “Optimized monotonic convex pair potentials stabilize low-coordinated crystals,” *Soft Matter* **7**, 2332–2335 (2011).
- ¹³A. Jain, J. R. Errington, and T. M. Truskett, “Inverse design of simple pairwise interactions with low coordinated 3D lattice ground states,” *Soft Matter* **9**, 3866–3870 (2013).
- ¹⁴S. P. Paradiso, K. T. Delaney, and G. H. Fredrickson, “Swarm intelligence platform for multiblock polymer inverse formulation design,” *ACS Macro Lett.* **5**, 972–976 (2016).
- ¹⁵W. G. Noid, “Perspective: Coarse-grained models for biomolecular systems,” *J. Chem. Phys.* **139**, 090901 (2013).
- ¹⁶J. F. Rudzinski and W. G. Noid, “Coarse-graining entropy, forces, and structures,” *J. Chem. Phys.* **135**, 214101 (2011).
- ¹⁷F. Müller-Plathe, “Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back,” *ChemPhysChem* **3**, 754–769 (2002).
- ¹⁸C.-C. Fu, P. M. Kulkarni, M. S. Shell, and L. G. Leal, “A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties,” *J. Chem. Phys.* **137**, 164106 (2012).
- ¹⁹A. Chaimovich and M. S. Shell, “Coarse-graining errors and numerical optimization using a relative entropy framework,” *J. Chem. Phys.* **134**, 094112 (2011).
- ²⁰M. S. Shell, “The relative entropy is fundamental to multiscale and inverse thermodynamic problems,” *J. Chem. Phys.* **129**, 144108 (2008).
- ²¹S. Izvekov and G. A. Voth, “Multiscale coarse graining of liquid-state systems,” *J. Chem. Phys.* **123**, 134105 (2005).
- ²²D. Andelman, F. Brochard, and J.-F. Joanny, “Phase-transitions in Langmuir monolayers of polar-molecules,” *J. Chem. Phys.* **86**, 3673–3681 (1987).
- ²³A. J. Archer, C. Ionescu, D. Pini, and L. Reatto, “Theory for the phase behaviour of a colloidal fluid with competing interactions,” *J. Phys.: Condens. Matter* **20**, 415106 (2008).
- ²⁴G. F. Kendrick, T. J. Sluckin, and M. J. Grimson, “Macrocrystal phases in charged colloidal suspensions,” *Europ. Phys. Lett.* **6**, 567–572 (1988).
- ²⁵C. B. Muratov, “Theory of domain patterns in systems with long-range interactions of Coulomb type,” *Phys. Rev. E* **66**, 066108 (2002).
- ²⁶M. Seul and D. Andelman, “Domain shapes and patterns—The phenomenology of modulated phases,” *Science* **267**, 476–483 (1995).
- ²⁷Y. Zhuang and P. Charbonneau, “Equilibrium phase behavior of the square-well linear microphase-forming model,” *J. Phys. Chem. B* **120**, 6178–6188 (2016).
- ²⁸Y. Zhuang and P. Charbonneau, “Recent advances in the theory and simulation of model colloidal microphase formers,” *J. Phys. Chem. B* **120**, 7775–7782 (2016).
- ²⁹B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, “Assembly of nothing: Equilibrium fluids with designed structured porosity,” *Soft Matter* **12**, 2663–2667 (2016).
- ³⁰B. A. Lindquist, S. Dutta, R. B. Jadrich, D. J. Milliron, and T. M. Truskett, “Interactions and design rules for assembly of porous colloidal mesophases,” *Soft Matter* **13**, 1335–1343 (2017).
- ³¹A. J. Archer and N. B. Wilding, “Phase behavior of a fluid with competing attractive and repulsive interactions,” *Phys. Rev. E* **76**, 031501 (2007).
- ³²R. P. Sear and W. M. Gelbart, “Microphase separation versus the vapor-liquid transition in systems of spherical particles,” *J. Chem. Phys.* **110**, 4582–4588 (1999).
- ³³A. Ciach, “Universal sequence of ordered structures obtained from mesoscopic description of self-assembly,” *Phys. Rev. E* **78**, 061505 (2008).
- ³⁴A. Ciach, J. Pekalski, and W. T. Gozdz, “Origin of similarity of phase diagrams in amphiphilic and colloidal systems with competing interactions,” *Soft Matter* **9**, 6301–6308 (2013).
- ³⁵H. Shin, G. M. Grason, and C. D. Santangelo, “Mesophases of soft-sphere aggregates,” *Soft Matter* **5**, 3629–3638 (2009).
- ³⁶B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, “Communication: Inverse design for self-assembly via on-the-fly optimization,” *J. Chem. Phys.* **145**, 111101 (2016).
- ³⁷This quantity is actually a probability distribution, though we use the terminology interchangeably without consequence.
- ³⁸S. Kullback and R. A. Leibler, “On information and sufficiency,” *Ann. Math. Stat.* **22**, 79–86 (1951).
- ³⁹D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, “GROMACS: Fast, flexible, and free,” *J. Comput. Chem.* **26**, 1701–1718 (2005).
- ⁴⁰S. Plimpton, “Fast parallel algorithms for short-range molecular dynamics,” *J. Comput. Phys.* **117**, 1–19 (1995).
- ⁴¹V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, “Versatile object-oriented toolkit for coarse-graining applications,” *J. Chem. Theory Comput.* **5**, 3211–3223 (2009).
- ⁴²S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle, and C. Junghans, “Relative entropy and optimization-driven coarse-graining methods in VOTCA,” *PLoS ONE* **10**, e0131754 (2015).
- ⁴³W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graphics* **14**, 33–38 (1996).
- ⁴⁴P. D. Godfrin, N. E. Valadez-Pérez, R. Castañeda-Priego, N. J. Wagner, and Y. Liu, “Generalized phase behavior of cluster formation in colloidal dispersions with competing interactions,” *Soft Matter* **10**, 5061–5071 (2014).

- ⁴⁵J. C. F. Toledano, F. Sciortino, and E. Zaccarelli, "Colloidal systems with competing interactions: From an arrested repulsive cluster phase to a gel," *Soft Matter* **5**, 2390–2398 (2009).
- ⁴⁶R. B. Jadrich, J. A. Bollinger, B. A. Lindquist, and T. M. Truskett, "Equilibrium cluster fluids: Pair interactions via inverse design," *Soft Matter* **11**, 9342–9354 (2015).
- ⁴⁷J. A. Bollinger and T. M. Truskett, "Fluids with competing interactions. I. Decoding the structure factor to detect and characterize self-limited clustering," *J. Chem. Phys.* **145**, 064902 (2016).
- ⁴⁸J. A. Bollinger and T. M. Truskett, "Fluids with competing interactions. II. Validating a free energy model for equilibrium cluster size," *J. Chem. Phys.* **145**, 064903 (2016).
- ⁴⁹R. B. Jadrich, J. A. Bollinger, K. P. Johnston, and T. M. Truskett, "Origin and detection of microstructural clustering in fluids with spatial-range competitive interactions," *Phys. Rev. E* **91**, 042312 (2015).
- ⁵⁰E. Mani, W. Lechner, W. K. Kegel, and P. G. Bolhuis, "Equilibrium and non-equilibrium cluster phases in colloids with competing interactions," *Soft Matter* **10**, 4479–4486 (2014).
- ⁵¹M. B. Sweatman, R. Fartaria, and L. Lue, "Cluster formation in fluids with competing short-range and long-range interactions," *J. Chem. Phys.* **140**, 124508 (2014).
- ⁵²A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf, and P. Schurtenberger, "Equilibrium cluster formation in concentrated protein solutions and colloids," *Nature* **432**, 492–495 (2004).
- ⁵³A. Baumketner and W. Cai, "Equilibrium clusters in suspensions of colloids interacting via potentials with a local minimum," *Condens. Matter Phys.* **19**, 13605 (2016).
- ⁵⁴S. H. Donaldson, C. T. Lee, B. F. Chmelka, and J. N. Israelachvili, "General hydrophobic interaction potential for surfactant/lipid bilayers from direct force measurements between light-modulated bilayers," *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15699–15704 (2011).
- ⁵⁵J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, New York, NY, USA, 2006).
- ⁵⁶J. Pekalski, A. Ciach, and N. G. Almarza, "Periodic ordering of clusters and stripes in a two-dimensional lattice model. I. Ground state, mean-field phase diagram and structure of the disordered phases," *J. Chem. Phys.* **140**, 114701 (2014).
- ⁵⁷N. G. Almarza, J. Pekalski, and A. Ciach, "Periodic ordering of clusters and stripes in a two-dimensional lattice model. II. Results of Monte Carlo simulation," *J. Chem. Phys.* **140**, 164708 (2014).
- ⁵⁸A. J. Archer, "Two-dimensional fluid with competing interactions exhibiting microphase separation: Theory for bulk and interfacial properties," *Phys. Rev. E* **78**, 031402 (2008).
- ⁵⁹J. Ruiz-García, R. Gámez-Corralas, and B. I. Ivlev, "Formation of two-dimensional colloidal voids, soap froths, and clusters," *Phys. Rev. E* **58**, 660–663 (1998).
- ⁶⁰F. Ghezzi and J. Earnshaw, "Formation of meso-structures in colloidal monolayers," *J. Phys.: Condens. Matter* **9**, L517–L523 (1997).
- ⁶¹B. Mladek, D. Gottwald, G. Kahl, M. Neumann, and C. Likos, "Formation of polymorphic cluster phases for a class of models of purely repulsive soft spheres," *Phys. Rev. Lett.* **96**, 045701 (2006).
- ⁶²P. Camp, "Structure and phase behavior of a two-dimensional system with core-softened and long-range repulsive interactions," *Phys. Rev. E* **68**, 061506 (2003).
- ⁶³R. Sear and D. Frenkel, "Continuous freezing in three dimensions," *Phys. Rev. Lett.* **90**, 195701 (2003).
- ⁶⁴C. Likos, N. Hoffmann, H. Lowen, and A. Louis, "Exotic fluids and crystals of soft polymeric colloids," *J. Phys.: Condens. Matter* **14**, 7681–7698 (2002).
- ⁶⁵W. D. Piñeros and T. M. Truskett, "Designing pairwise interactions that stabilize open crystals: Truncated square and truncated hexagonal lattices," *J. Chem. Phys.* **146**, 144501 (2017).
- ⁶⁶G. Zhang, F. H. Stillinger, and S. Torquato, "Probing the limitations of isotropic pair potentials to produce ground-state structural extremes via inverse statistical mechanics," *Phys. Rev. E* **88**, 042309 (2013).
- ⁶⁷W. L. Miller and A. Cacciuto, "Exploiting classical nucleation theory for reverse self-assembly," *J. Chem. Phys.* **133**, 234108 (2010).
- ⁶⁸W. D. Piñeros, M. Baldea, and T. M. Truskett, "Breadth versus depth: Interactions that stabilize particle assemblies to changes in density or temperature," *J. Chem. Phys.* **144**, 084502 (2016).
- ⁶⁹E. Jankowski and S. C. Glotzer, "Screening and designing patchy particles for optimized self-assembly propensity through assembly pathway engineering," *Soft Matter* **8**, 2852–2859 (2012).