Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Inverse design of self-assembling colloidal crystals with omnidirectional photonic bandgaps

Yutao Ma[a] and Andrew L. Ferguson[a,*]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Open colloidal lattices possessing omnidirectional photonic bandgaps in the visible or near-visible regime are attractive optical materials the realization of which has remained elusive. We report the use of an inverse design strategy termed landscape engineering that rationally sculpts the free energy self-assembly landscape using evolutionary algorithms to discover anisotropic patchy colloids capable of spontaneously assembling pyrochlore and cubic diamond lattices possessing complete photonic bandgaps. We validate the designs in computer simulations to demonstrate the defect-free formation of these lattices via a two-stage hierarchical assembly mechanism. Our approach demonstrates a principled strategy for the inverse design of self-assembling colloids for the bottom-up fabrication of desired crystal lattices.

1 Introduction

The self-assembly of colloidal nanoparticles provides a powerful tool for forming many complex structures, such as colloidal aggregates[1–5], multi-shell clusters[6], helical structures[7] and crystals[8–16]. The assembly of open colloidal lattices has drawn particular attention[8–16] because particular structures posses complete photonic bandgaps and are therefore useful as 3D photonic crystals with omnidirectional bandgaps[14,17–19]. The optical properties of a colloidal crystal are dictated by the organization of colloids within the crystal lattice[20]. The size of the colloids, refractive index contrast between the colloids and voids, and specific pattern of refractive index changes due to the packing of colloids dictate the photonic properties of the crystal[20]. Whereas hexagonal close packed (hcp) and face-centered cubic (fcc) lattices are most easily assembled from isotropic colloidal spheres, they do not possess complete photonic bandgaps that forbid passage of photons with particular energies in all directions. It is for this reason that more exotic open lattices such as pyrochlore[14,17,19] and diamond[21,22], and inverse opal[20,23,24] that do possess complete bandgaps have attracted much attention. These crystals have desirable applications in the manipulation of photons in optical wave guiding[25] and in optical computing[26].

Many techniques have been explored to synthesize colloidal crystals by bottom-up self-assembly[8–16]. For example, John and coworkers synthesized single crystalline silicon inverse opal using close-packed silica colloidal spheres as a template[24]. Damman and coworkers assembled colloidal lattices by vertical deposition on curved surfaces and demonstrated that the optical properties of the resulting colloidal crystal could be manipulated by the surface curvature without introducing crystal defects[27]. Crocker and coworkers employed two differently-sized spherical colloids functionalized with complementary DNA oligomers to fabricate "double diamond" (B32) colloidal crystals isomorphic to the NaTl Zintl phase[28]. Pine and coworkers co-assembled tetrahedral colloidal clusters and colloidal spheres using complementary DNA binding to fabricate a colloidal $MgCu_2$ crystal[10]. Grzybowski and coworkers assembled diamond-like colloidal lattices from nearly equally-sized oppositely-charged nanoparticles[15,29]. In the context of the assembly of open crystal structures, patchy colloids functionalized with anisotropic and directional surface interactions have emerged as a promising means for their fabrication by self-assembly[11–13,16,20,23,30]. This class of building blocks is experimentally attractive as they are based on simple spherical colloids that can be flexibly functionalized through anisotropic surface patterning techniques[31–35]. For example, Chen et al.[36] have experimentally demonstrated the self-assembly of triblock patchy colloids into metastructures by a step-wise control of ion concentration in solution, and Morphew et al.[11] have computationally investigated the hierarchical self-assembly of triblock patchy colloids into body-centered cubic and cubic diamond crystals.

A primary challenge in patchy colloid self-assembly is the design of the anisotropic interactions to favor the assembly of the desired crystal lattice. Of particular concern is the existence of competing crystal structures with similar free energies that can frustrate defect-free assembly of the target lattice. For example, the pyrochlore lattice (also known as cubic tetrastack)[19,37] can be viewed as a tetrahedral network of corner-sharing tetrahedral clusters[17], where the tetrahedral clusters occupy the voids of a cubic diamond lattice[10,37]. However, the pyrochlore lattice has

[a] Pritzker School of Molecular Engineering, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA.
[*] Corresponding author: Tel: +1-773-702-5950; E-mail: andrewferguson@uchicago.edu

a closely-related analogue known as hexagonal tetrastack[9,37,38], which differs from the pyrochlore lattice by the orientations of adjacent layers but has similar free energy[9,37,38]. Similarly, cubic diamond (also known as conventional diamond or c-diamond) and hexagonal diamond (also known as h-diamond or Lonsdaleite) are structurally similar crystals with similar free energies that differ in the stacking of subsequent layers[13]. In each case, care must be taken in the patchy particle design to favor one polymorph over the other[9,12,13,37]. As such, a primary concern in our design protocol is to engineer anisotropy into the particle interaction to break the degeneracy between the desired lattice structures and closely related analogues (i.e., pyrochlore vs. hexagonal tetrastack, cubic diamond vs. hexagonal diamond).

A number of inverse design techniques have been proposed for optimizing the interactions between colloidal particles[8,10–12,37,39–48] to favor the formation of the desired target crystal. For example, Truskett and coworkers used inverse design strategies to design engineered isotropic pairwise potentials that favor the formation of various two-dimensional and three-dimensional colloidal crystals[43,45–47]. Torquato and coworkers also employed inverse statistical mechanical strategies to design isotropic potentials that favor various colloidal lattices[42,44,48], as well as anisotropic potentials that favor the formation of various two-dimensional colloidal lattices[39]. Lyubartsev and Laaksonen[49] and Mungan et al.[50] deduced interaction potentials from structural correlation functions. Cohn and Kumar[51] employed linear programming to determine isotropic potentials leading to the desired configuration as ground state. Dijkstra and coworkers employed an isotropic repulsive pairwise potential to favor the formation of a pyrochlore-like colloidal crystal[8]. Escobedo[40] used anisotropic particles and potentials to form different colloidal compounds. Romano and Sciortino employed asymmetric patterning to robustly assemble pyrochlore and disfavor hexagonal tetrastack lattice[37]. Morphew et al.[11] used a basin-hopping optimization method to design the potentials that favor the formation of three-dimensional cubic diamond lattice and BCC lattice via colloidal molecules. Glotzer and coworkers proposed the introduction of angular potentials or charge repulsion to favor cubic over hexagonal diamond[12]. Pine and coworkers designed isotropic DNA-grafted colloidal clusters and singlet colloids to realize a colloidal $MgCu_2$ lattice[10]. In the absence of some means to break the degeneracy between competing polymorphs it is typically necessary to seed the system with a fragment of the desired crystal structure[12] to robustly assemble the target crystal. Failing to break the degeneracy through one or other of these strategies risks the uncontrolled fabrication of hybrid lattices[11].

In this work, we employ a recently developed inverse design protocol termed landscape engineering to systematically discover patchy colloid building blocks capable of spontaneous self-assembly into a pyrochlore lattice and a cubic diamond lattice formed from tetrahedral clusters[52]. The approach iteratively sculpts the free energy surface of the self-assembling colloids using evolutionary algorithms to update the placement and strength of the colloidal patches to stabilize the target lattice over all competing polymorphs. We target pyrochlore and cubic diamond as 3D lattices possessing complete photonic bandgaps that have proven frustratingly elusive to fabrication via self-assembly[11,13]. We show that the colloidal designs predicted by landscape engineering spontaneously nucleate and grow defect-free photonic lattices of the desired crystal polymorphs in a two-stage hierarchical assembly mechanism. Since we conduct inverse design over the free energy surface rather than the potential energy surface, the interaction potentials discovered by landscape engineering are not those that would have been expected by energy minimization or zero-temperature optimization of the target lattice. This demonstrates the importance of incorporating many-body and entropic effects into the particle design.

The anisotropic potentials employed in this work are relatively simple and generic, but may be considered as simplified and idealized models of inter-particle interactions that may be experimentally realized through advanced surface-patterning techniques[10,30–32,35,53–57]. For example, the patchy colloid model considered in this work can be considered as a simplified representation of nanodot-decorated nanoparticles of the sort realized by Bae et al.[32] and Wang et al.[31] through regions of titania or propyl methacrylate whose interaction strengths depend on the specific materials properties. In a similar vein, Zhang et al. used colloidal masks to fabricate anisotropic nanoparticles decorated with nanodots on opposite poles[58]. We might also consider our models to be idealized representations of colloids surface functionalized with localized patches of complementary DNA oligomers with defined sequence and specificity[10,35,55–57]. The Kern-Frenkel model[59] is one of the most popular computational models employed to simulate patchy particle assembly[11,37] and can be considered a simplified model for patchy particles with surface interaction patches deposited via glancing angle deposition[53,54,60]. Accordingly, our computational patchy particle model and others like it are intended as simplified idealizations of experimentally-realizable inter-particle interactions. It is the primary goal of the present work to employ such potentials to expose the fundamental principles governing assembly, provide new insight into the thermodynamic, kinetic, and morphological processes underpinning assembly, and demonstrate a new methodology for the rational design of patchy colloids programmed to self-assemble into desired aggregates. In doing so, we aim to provide new understanding and precepts for the experimental design of self-assembling colloidal lattices. Romano and Sciortino have previously proposed the use of asymmetric Kern-Frenkel type patchy colloids to form pyrochlore lattice[37]. The present work considers a different patchy particle model with defined isotropic surface interactions that may be considered a simplified representation of nanodot-decorated nanoparticles[31,32]. Moreover we design the anisotropic interaction potentials using a systematic and automated inverse design protocol. Accordingly, this work reports a new automated inverse design strategy for the fabrication of desired colloidal lattices, and reduces this to practice in the design of two patchy particle building blocks capable of spontaneously self-assembling pyrochlore and diamond lattices with omnidirectional photonic bandgaps.

2 Methods

2.1 Self-assembling patchy colloid model

2.1.1 Pyrochlore lattice

The pyrochlore lattice can be viewed as a tetrahedral network of corner-sharing tetrahedra. An illustration of the pyrochlore lattice is given in fig. 1a. Every particle (i.e. vertex of tetrahedron) in the pyrochlore lattice exists in a staggered local configuration (fig. 1b) where its six nearest neighbors are rotated by 60° around it. A competing crystal structure with similar free energy is the hexagonal tetrastack lattice in which 75% of particles exist in staggered local configuration and 25% of particles exist in eclipsed local configuration[37] (fig. 1c). In order to favor the pyrochlore lattice against hexagonal tetrastack lattice, we decorate three "B" patches (blue patches) forming an equilateral triangle on the north pole of the central sphere ("A" particle) and three "D" patches (purple patches) forming an equilateral triangle on the south pole. The "D" patches are rotated by 60° degree around the central axis with respect to the "B" patches. As we shall see, the "B" patches serve as the interaction sites for directing individual patchy colloids to form tetrahedral clusters, and "D" patches serve as the interaction sites for directing tetrahedral clusters to form corner-sharing network of tetrahedra while maintaining the staggered local configuration of the vertices of tetrahedra in this network. This model is illustrated in fig. 1d-f. The

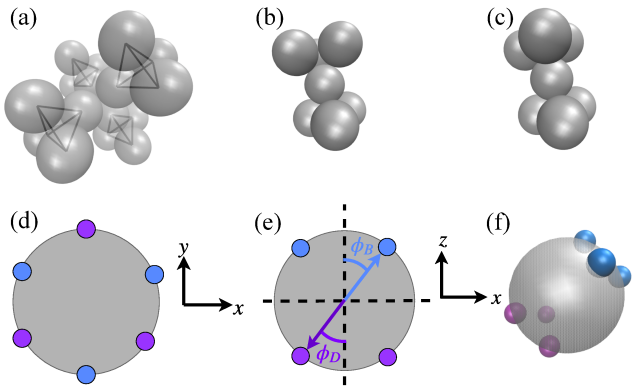

Fig. 1 Pyrochlore lattice. (a) A cell of the pyrochlore lattice. Wireframe tetrahedra connecting the centers of translucent colloids serve as a guide to the eye to illustrate the tetrahedral motif. (b) An illustration of the staggered tetrahedral configuration that forms the fundamental motif of the lattice. The central particle forms a regular tetrahedron with its three nearest neighbors above and similarly with its three nearest neighbors below. The two tetrahedra are rotated 60° relative to one another to form a staggered configuration. (c) An illustration of the eclipsed local configuration. Schematic (d) top-down and (e) side views of the anisotropic patchy colloid building block to be optimized by landscape engineering. The blue "B" patches with interaction strengths ε_B define the vertices of an equilateral triangle on the north pole of the patchy colloid at a polar angle of ϕ_B. The purple "D" patches with interaction strengths ε_D lie at the vertices of an analogous south pole equilateral triangle at a polar angle $\phi_D = \phi_B$ and a relative azimuthal rotation of 60°. The 60° azimuthal rotation between the north and south pole patches energetically favors the staggered tetrahedral configuration (b) over the eclipsed (c). (f) Three dimensional rendering of the patchy colloid where the central particle is made transparent to show the staggered orientation between "B" and "D" patches. Landscape engineering is employed to optimize $\{E_B, \phi_B, E_D, \phi_D\}$ to promote the two-stage hierarchical self-assembly of the pyrochlore lattice.

patches and the central particle are treated as a single rigid body building block, and the interactions within the same rigid body are ignored. The diameters of central particle and the patches are chosen as $\sigma_A = 5\sigma$ and $\sigma_B = \sigma_D = \sigma$. The masses are chosen as $m_A = 125m$ and $m_B = m_D = m$. The interactions between "B"-"B" and "D"-"D" patches on different patchy colloids are modeled by Lennard-Jones potential:

$$U_{\mathrm{LJ}}^{ii}(r) = 4\varepsilon_i \left[\left(\frac{\sigma_i}{r}\right)^{12} - \left(\frac{\sigma_i}{r}\right)^{6} \right] \quad \text{for } i \in \{\mathrm{B,D}\} \tag{1}$$

where ε_i is the well depth, or the "interaction strength", of particle i, and σ_i is its diameter. The interactions between "A"-"X", where "X" $\in$ {"A","B","D"}, and between "B"-"D" particles on different patchy colloids are modeled by Weeks-Chandler-Andersen (WCA) potential[61] to incorporate excluded-volume effects:

$$U_{\mathrm{WCA}}^{ij}(r) = \begin{cases} 4\varepsilon_{ij}\left[\left(\frac{\sigma}{r-\Delta}\right)^{12} - \left(\frac{\sigma}{r-\Delta}\right)^{6}\right] + \varepsilon_{ij} & \text{if } r < 2^{\frac{1}{6}}\sigma + \Delta_{ij} \\ 0 & \text{if } r \geq 2^{\frac{1}{6}}\sigma + \Delta_{ij} \end{cases} \tag{2}$$

$\Delta_{ij} = \frac{\sigma_i + \sigma_j}{2} - 1$ shifts the potential to act on the surfaces of particles i and j. $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ is given by the Lorentz-Berthelot mixing rule. Since there are three "B" patches and three "D" patches on each patchy colloid, we specify the total interaction strength of each of these two species: $E_i = 3\varepsilon_i$ where $i \in \{\mathrm{B,D}\}$, and evenly distribute it among all three patches of each species.

We design the particles to assemble the pyrochlore lattice via a two-step hierarchical assembly mechanism. The "B" patches possess a stronger interaction strength and are activated during the high temperature phase of the assembly process at T_{high} to direct the patchy colloids to assemble into tetrahedral clusters. The "D" patches, which have a weaker interaction strength, are then activated during the temperature cooling process down to T_{low} and direct the assembly of the tetrahedral clusters into the pyrochlore lattice. The two-stage hierarchy in interaction strengths maps to a two-stage hierarchy in structure that has previously been exploited for the fabrication of hierarchically structured materials[11,36,62,63]. In the present work, the high-temperature assembly process can be conceived as producing tetrahedral building blocks from spherical colloids, and the low-temperature assembly process as directing the assembly of the tetrahedra into a pyrochlore lattice. Thermal decoupling between the two levels of the hierarchy is asserted in the relation between E_B and E_D as:

$$E_D = \frac{T_{\mathrm{low}}}{T_{\mathrm{high}}} E_B. \tag{3}$$

Since the "B" and "D" patches both mediate the formation of tetrahedra, their polar angles are related as:

$$\phi_D = \phi_B. \tag{4}$$

As mentioned above, their azimuthal angles are mutually rotated by 60° to favor staggered local configurations (fig. 1d). An illustration of the desired two-stage assembly process is shown in fig. 2.

The "D" patches on the south pole act as the low-temperature

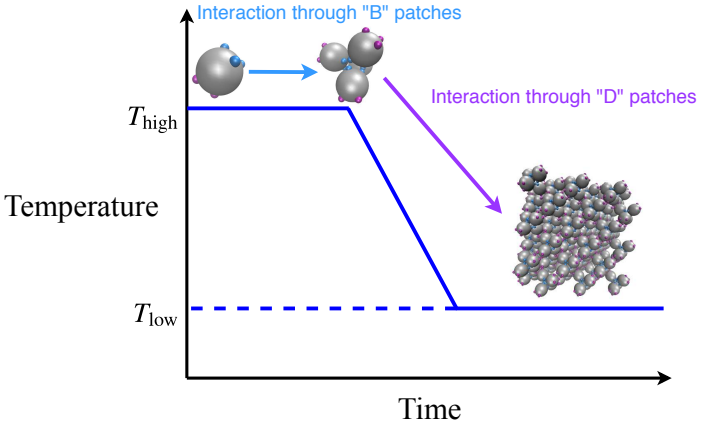

Fig. 2 An illustration of the two-stage hierarchical assembly mechanism for the pyrochlore lattice. During the high temperature phase at T_{high}, the more strongly interacting "B" patches direct the formation of tetrahedral clusters while the more weakly interacting "D" patches are effectively inert. During the cooling process to T_{low}, the "B" patches lock the patchy colloids into the self-assembled tetrahedra and the "D" patches direct the tetrahedral clusters to assemble into the pyrochlore lattice.

counterparts of "B" patches on the north pole and the desired self-assembled tetrahedral motif in both cases is the same. At the high temperature T_{high} we seek to optimize the interaction strength and polar angle of the "B" patches at the north pole $\{E_B, \phi_B\}$ to favor the assembly of tetrahedral clusters relative to all competing structures. We solve this inverse design problem using the landscape engineering approach described below[52]. We then obtain the optimal solution for the "D" patches on the south pole through equations 3 and 4. In doing so we assume that the optimal solution for the "D" patches at the low temperature is identical to that for the "B" patches at the high temperature, with the interaction strength just appropriately scaled by the temperature ratio. The motivation for this equivalence is that the "B" and "D" interfaces are structurally identical, but we do note that the former is formed from constituent monomers whereas the latter is formed from constituent tetrahedra, so the multibody and entropic interactions during the assembly process may differ. Nevertheless, the assumption of this equivalence simplifies the inverse design problem for the pyrochlore lattice by reducing it to a single optimization. As we will show, the assumption proves to be a good one as it leads to the successful assembly of defect-free crystals. As discussed later, in the case of cubic diamond this symmetry is absent and we must independently optimize the two poles of the colloidal building blocks within two separate optimization protocols.

Our computational model for the anisotropic interaction patches is deliberately a simple and generic potential. Experimentally, such directional and specific patches with particular $\{E_B, \phi_B, E_D, \phi_D\}$ might be realized by advanced surface-patterning techniques[10,30–32,35,53–57]. For example, the central colloidal spheres may be functionalized by DNA oligomers with tunable interaction strengths and specificities[35], or patterned with regions of titania or propyl methacrylate whose interaction strengths depend on the specific materials properties[31,32].

We perform our simulation in reduced units, where $\sigma = 1$, $\varepsilon = \varepsilon_A = 1$, and $m = 1$. Using these units, we specify $\sigma_A = 5$, $\sigma_i = \sigma = 1$ for $i \in \{B, D\}$, $m_A = 125$ and $m_i = m = 1$ for $i \in \{B, D\}$. We may define a mapping between our reduced units and real units. For example, consider central particle "A" of diameter $\sigma_A = 5\sigma = 1$ μm and density $\rho_A = 1$ g/cm^3, and an energy scale of $\varepsilon = 1$ k_BT at $T = 298$ K. From these fundamental units, we can derive the temperature in real units as $T = T^* \frac{\varepsilon}{k_B}$ and time in real units as $t = t^* \sigma \sqrt{\frac{m}{\varepsilon}}$, where T^* and t^* are temperature and time in reduced units. In our two-stage assembly process, we use $T^*_{high} = 0.8$ at high temperature and $T^*_{low} = 0.3$ at low temperature, which correspond to $T_{high} = 238.4$ K and $T_{low} = 89.4$ K in real units, respectively. Also, we use a step size of $dt^* = 0.005$ in our simulations, which corresponds to $dt = 0.36$ μs in real units.

2.1.2 Cubic diamond lattice

The fundamental motif of the cubic diamond lattice formed from tetrahedral clusters is dimers of tetrahedra in staggered configurations (fig. 3a)[11,12,37]. (For short, we will henceforward simply refer to this lattice as cubic diamond except when it is unclear to do so.) These staggered tetrahedral dimers come together to form chair-like rings (fig. 3b) within the cubic diamond crystal. A competing structure that shares similar free energy is the hexagonal diamond lattice that contains tetrahedral dimers in both staggered and eclipsed (fig. 3c) configurations. The hexagonal diamond lattice consists of 25% chair-like rings and 75% boat-like rings (fig. 3d)[13]. To favor cubic over hexagonal diamond, we employ a similar patchy colloid design to that for pyrochlore. Three "B" patches with interaction strength ε_B are placed in an equilateral triangle on the north pole at a polar angle of ϕ_B to favor the high temperature formation of tetrahedral clusters (fig. 3e-g). The south pole must be functionalized with two kinds of patches, "D" patches and "E" patches at a polar angle $\phi = \phi_D = \phi_E$, in order to preferentially stabilize the staggered dimer relative to the eclipsed one. The "D" patches (purple) are aziumthally aligned with the "B" patches (blue), and the "E" patches (lime) are rotated by 60°. The "B"-"B" interaction is still modeled by Lennard-Jones potential as in equation 1. The "D"-"E" interaction is also modeled by Lennard-Jones potential with the interaction strength and range given by the Lorentz-Berthelot mixing rules:

$$U^{DE}_{LJ}(r) = 4\varepsilon_{DE}\left[\left(\frac{\sigma_{DE}}{r}\right)^{12} - \left(\frac{\sigma_{DE}}{r}\right)^{6}\right]$$
$$\varepsilon_{DE} = \sqrt{\varepsilon_D \varepsilon_E} \qquad (5)$$
$$\sigma_{DE} = \frac{\sigma_D + \sigma_E}{2}$$

All other interactions are given by WCA potential defined in equation 2. In this way, the "D"-"E" attractive interactions induce the contact dimer formed by two tetrahedra to stabilize a mutual rotation of 60° and favor the staggered dimer over the eclipsed one fig. 3(h). Our use of this design with two patch types in an alternating ring with attractive interactions between unlike patches is motivated by the need to induce a 60° rotation between the two tetrahedra at the dimer interface. There is no clear way to favor

this staggered orientation (fig. 3(a)) over the eclipsed (fig. 3(c)) using only one patch type with attractive interactions regardless of their azimuthal positioning.

The optimal parameters $\{E_B, \phi_B\}$ for the "B" patches will be the same as those for pyrochlore since in both cases the high temperature assembly process into tetrahedral clusters is identical. The remaining goal is to optimize the parameters for "D" and "E" patches $\{E_D, E_E, \phi = \phi_D = \phi_E\}$ through landscape engineering. Without loss of generality, we choose to optimize the parameters for "D" and "E" patches at $T^*_{\text{high}} = 0.8$, and then scale down the interaction strengths to match the low temperature phase $T^*_{\text{low}} = 0.3$ by a factor of $T^*_{\text{low}}/T^*_{\text{high}}$. We specify $\sigma_D = \sigma_E = 1$ and $m_D = m_E = 1$ in reduced units. The mapping between reduced units and real units is defined in the same manner as section 2.1.1.

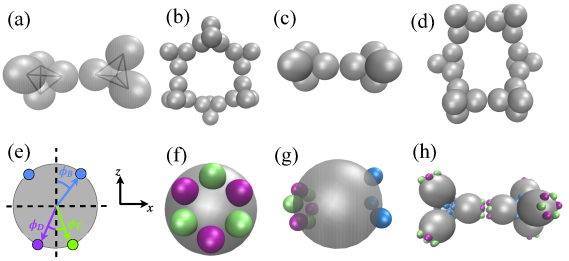

Fig. 3 Cubic diamond lattice. (a) Staggered dimer of tetrahedral clusters. Wireframe tetrahedra connecting the centers of translucent colloids serve as a guide to the eye to illustrate the tetrahedral motif. (b) Chair-like ring of tetrahedral clusters. (c) Eclipsed dimer of tetrahedral clusters. (d) Boat-like ring of tetrahedral clusters. (e) Schematic side view of the anisotropic patchy colloid. The blue "B" patches with interaction strengths ε_B define the vertices of an equilateral triangle on the north pole of the patchy colloid at a polar angle of ϕ_B. The purple "D" and "E" patches with interaction strengths ε_D and ε_E lie at the vertices of a analogous south pole equilateral triangles at a polar angle $\phi = \phi_D = \phi_E$. The "D" patches are azimuthally aligned with the "B" patches and the "E" patches rotated by 60°. The (f) bottom and (g) side views of a three dimensional rendering of the patchy colloid. In (g) the central particle is made transparent to show the relative orientations of the "B", "D", and "E" patch types. (h) Illustration of a staggered dimer formed by two tetrahedral clusters of patchy colloids. Landscape engineering is employed to optimize $\{E_B, \phi_B, E_D, E_E, \phi = \phi_D = \phi_E\}$ to promote the two-stage hierarchical self-assembly of the cubic diamond lattice.

2.2 Landscape engineering

We follow a previously developed optimization procedure, called landscape engineering[52], to optimize the design parameters of patchy colloid. Hereafter we will refer to a set of design parameters as a "candidate". The whole procedure consists of the following steps: Starting from a group of initial candidates, we first conduct Langevin dynamics simulations of the self-assembly of each candidate and use diffusion map nonlinear dimensionality reduction to construct a low-dimensional embedding of the various self-assembled configurations observed over the course of the simulation. Next, on this low-dimensional diffusion map space, we select biasing centers and perform umbrella sampling[64,65] for each candidate in order to enhance the sampling of accessible configurations. From the results of umbrella sampling, for each candidate, we construct its self-assembly free energy surface which contains information about the stability of all accessible

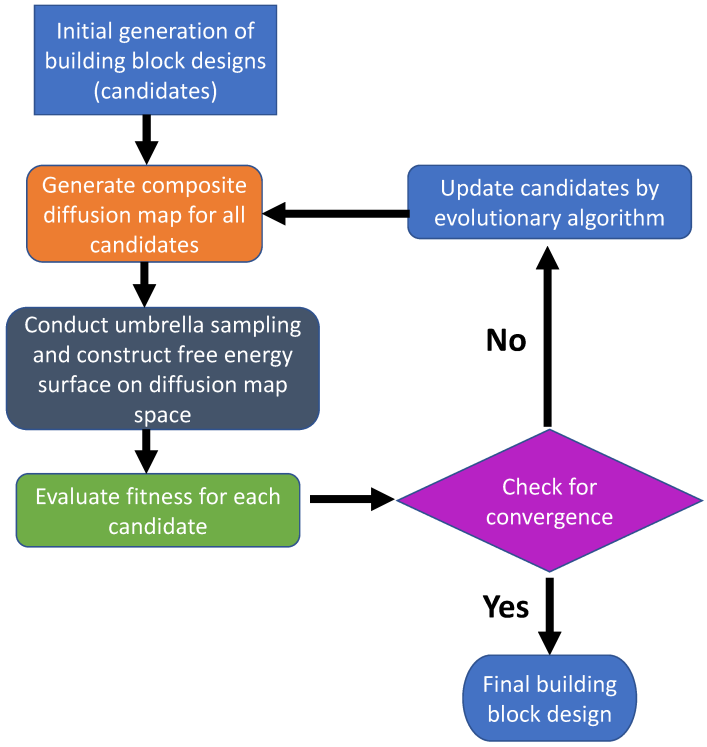

Fig. 4 Flowchart of the landscape engineering inverse design procedure.

configurations. We then locate the target structure on the free energy surface and define a fitness metric based on the relative stability between the target structure and the nearest competitor. Finally, we use an evolutionary algorithm, called Covariance Matrix Adaptation Evolution Strategy (CMA-ES)[66], to propose new and improved candidates based on the fitness values of old candidates. These steps are repeated until the proposed new candidates stabilize around an optimal candidate. A flowchart of this procedure is shown in fig. 4.

As described in the previous section, the parameters being optimized in the patchy colloid models for assembling pyrochlore lattice and cubic diamond lattice are $\{E_B, \phi_B\}$ and $\{E_D, E_E, \phi\}$, respectively. In case of pyrochlore lattice, our goal is to optimize $\{E_B, \phi_B\}$ that favor the formation of tetrahedral cluster at $T^*_{\text{high}} = 0.8$, and then the parameters for the "D" patches are obtained through equations 3 and 4. In this case, we conduct landscape engineering on particles possessing only "B" patches on the central sphere such that we temporarily ignore the "D" patches during this optimization. This approach is valid if the interaction strengths of the "D" patches are sufficiently weak compared to that of the "B" patches to be considered thermally decoupled. After the optimization is complete we place "D" patches on the opposite pole with parameters given by equations 3 and 4.

In case of the cubic diamond lattice, our goal is to optimize $\{E_D, E_E, \phi\}$ to favor dimer formation. During this optimization, we only put "D" and "E" patches on the central sphere and temporarily ignore the "B" patches. This assumption is a warranted if it can be assumed that the "B" and "D","E" interactions are thermally decoupled and the tetrahedra formed by the "B" interactions can be considered to be rock-like building blocks at the

low-temperature at which "D" and "E"-mediated assembly into the cubic diamond lattice proceeds. After the optimization is complete, the "B" patches are added back on the opposite pole with $\{E_B, \phi_B\}$ taken from the pyrochlore optimization, and the interaction strengths of "D" and "E" patches are scaled down from $T^*_{\text{high}} = 0.8$ at which the optimization was conducted to $T^*_{\text{low}} = 0.3$.

We now briefly discuss each step of the landscape engineering procedure (fig. 4).

2.2.1 Langevin dynamics

For each candidate in a group of candidates, we need to estimate its accessible configurations. To do this, we employ Langevin dynamics simulation using HOOMD-blue[67,68]. For each simulation, we initialize the system with 64 randomly placed and oriented patchy colloids in a cubic simulation box with side length $L = 52.52\sigma$. Taking a patchy colloid to be a sphere with radius corresponding to the sum of the radii of the central particle and its surface patches, the corresponding volume fraction of patchy colloids in the system is $\varphi = 0.05$. We use Langevin dynamics integrator with $T^* = 0.8$ and step size $dt^* = 0.005$. We evolve the system for 3.5×10^7 steps and track the cluster formed by one randomly-selected tagged colloid every 3500 steps. This results in a total of 10^4 snapshots per simulation. We perform three independent Langevin dynamics simulations for each candidate.

2.2.2 Diffusion maps

Diffusion maps[69,70] are a widely-used nonlinear dimensionality reduction technique that has previously been applied to study of time evolution of molecular systems[2,71–74]. In the study of self-assembly process, it can provide a dynamically meaningful low-dimensional representation of the assembly process[2]. In our case, the diffusion map embeds the N self-assembled aggregates $\{x_i\}_{i=1}^N$ observed from the molecular simulations onto a low-dimensional manifold. The algorithm starts from constructing the Gaussian kernel matrix based on the pairwise distances between aggregates:

$$A_{ij} = e^{-\frac{d_{ij}^2}{2\epsilon}} \tag{6}$$

where d_{ij} is the pairwise distance between aggregate i and aggregate j, and ϵ is a Gaussian bandwidth. The definition of d_{ij} will be provided below. From this kernel matrix, a stochastic matrix representing the random walk over the data set is defined by row-normalizing the kernel matrix:

$$M_{ij} = \frac{A_{ij}}{\sum_j A_{ij}} \tag{7}$$

where M_{ij} may be interpreted as the probability of hopping from aggregate i to aggregate j in a time step $\Delta t = \epsilon$. The eigenvectors $\{\vec{\psi}_i\}_{i=1}^N$ of the stochastic matrix M are the discrete approximations of the eigenfunctions of the backward Fokker-Planck operator which describes a diffusion process over the data set[69,70]. The eigenfunctions associated with large eigenvalues describe the "slow" modes of the diffusion process, while the eigenfunctions associated with small eigenvalues describe the "fast" modes. The long-time behavior of the system is captured by top few eigenfunctions. Since the matrix M is Markovian, its top eigenvalue is $\lambda_1 = 1$ and the associated eigenvector is the trivial eigenvector $\vec{\psi}_1 = \vec{1}$[75]. The diffusion map nonlinear embedding into a d-dimensional space is achieved by projecting each self-assembled aggregate observed over the course of the simulations $\{x_i\}_{i=1}^N$ into the top d nontrivial eigenvectors:

$$x_i \to \begin{bmatrix} \vec{\psi}_2(i) & \vec{\psi}_3(i) & \dots & \vec{\psi}_{d+1}(i) \end{bmatrix}^T. \tag{8}$$

An appropriate choice of d is defined by a gap in the eigenvalue spectrum[2,52]. In all cases in the present work a gap was identified after the third eigenvalue λ_3 informing two-dimensional embeddings into the two leading non-trivial eigenvectors $\{\vec{\psi}_2, \vec{\psi}_3\}$. The diffusion map embedding can be accelerated by combining landmark selection and Nyström extension to construct diffusion maps over a subset of $M << N$ landmark points and then projecting the remaining $(N - M)$ points into the embedding using an out-of-sample extension[76,77]. The details of this approach are presented in ref. 77, with illustrative applications to molecular systems demonstrating 50-fold speedups with less than 4% error in manifold reconstruction.

The key to construct the diffusion map is to appropriately define the pairwise distance d_{ij} in equation 6. We need a way to compare the structural similarities between aggregates formed by the patchy colloids. To do this, we employ the graph-based approach described in ref. 52, which is a modification of the Isorank algorithm[78]. This graph-based approach transforms the task of comparing structural similarities between aggregates into the task of comparing similarities between the graphs representing the aggregates. It first represents each aggregate by a graph G whose nodes correspond to the colloids within the aggregate and whose edges are weighted by the Euclidean distances between those colloids. In case of two aggregates with different number of colloids (i.e. two graphs with different number of nodes), the algorithm augments the smaller graph with $|N_i - N_j|$ ghost nodes. The algorithm then employs a greedy approach to find the pseudo-optimal alignment between two graphs by seeking an alignment $H_{\min}$ that minimizes the L_1 distance between two graphs G_i and G_j:

$$H_{\min} = \arg\min \frac{\sum_{m,n} |(H^T G_i H)(m,n) - G_j(m,n)|}{|N_j|(|N_j| - 1)} \tag{9}$$

and the "distance" (i.e. structural similarity) between aggregate i and aggregate j is:

$$d_{ij} = \frac{\sum_{m,n} |(H_{\min}^T G_i H_{\min})(m,n) - G_j(m,n)|}{|N_j|(|N_j| - 1)}. \tag{10}$$

Importantly, this graph-based distance measure is invariant to rotation, translation, and particle permutation (i.e., particle relabeling) of the self-assembled aggregates.

During each Langevin dynamics simulation conducted for each candidate in a generation, we record the aggregates comprising a single tagged colloid. Then, we collect together the aggregates sampled from all such simulations for all candidates within a generation. From this group of aggregates, we construct a single composite diffusion map. By generating a composite diffusion map for all candidates in a generation, we obtain a unified low-

dimensional embedding within which to construct and compare free energy surfaces.

2.2.3 Umbrella sampling – hybrid Monte Carlo

Having constructed the low-dimensional diffusion map space, we now need to construct the free energy surface for each candidate in terms of the diffusion map coordinates $\{\vec{\psi}_i\}_{i=2}^{d+1}$. To avoid the possible kinetic trapping due to potentially high free energy barrier, we follow the biased sampling procedure described in ref. 52 that combines umbrella sampling[64,65,79] with hybrid Monte Carlo[80–82] to efficiently sample configurational space by applying biasing potentials within the collective variables (i.e., leading eigenvectors $\{\vec{\psi}_i\}$) determined by diffusion maps. We then reweight these data to estimate the unbiased free energy surfaces governing the self-assembly of each candidate building block. In brief, we tile the d-dimensional diffusion map embedding with harmonic biasing potentials:

$$W(\vec{\psi},\vec{\psi}^*) = \frac{1}{2}(\vec{\psi}-\vec{\psi}^*)^T K(\vec{\psi}-\vec{\psi}^*) \tag{11}$$

where $\vec{\psi}^*$ is the d-dimensional harmonic center and K is a $d \times d$ dimensional diagonal matrix whose elements are the strengths of harmonic potential along each dimension. We conduct an independent biased simulation under each biasing potential and efficiently sample configurational space within the hybrid Monte Carlo (HMC) framework using a NVE integrator to propose a trial move under the unbiased Hamiltonian. The initial translational and angular momenta are drawn from the Maxwell-Boltzmann distribution. The acceptance probability of the trial move from the old state $\{\{q\}^{\text{old}},\{p\}^{\text{old}},\vec{\psi}^{\text{old}}\}$ to the new $\{\{q\}^{\text{new}},\{p\}^{\text{new}},\vec{\psi}^{\text{new}}\}$ is dictated by the Metropolis-Hasting criterion:

$$P_{\text{acc}}(\text{old}\rightarrow\text{new}) = \min\left(1, \frac{e^{-\beta(U(\{q\}^{\text{new}})+K(\{p\}^{\text{new}})+W(\vec{\psi}^{\text{new}},\vec{\psi}^*))}}{e^{-\beta(U(\{q\}^{\text{old}})+K(\{p\}^{\text{old}})+W(\vec{\psi}^{\text{old}},\vec{\psi}^*))}}\right) \tag{12}$$

where $\beta = (k_B T)^{-1}$, $U(\{q\})$ is the potential energy associated with the particle positions $\{q\}$, $K(\{p\})$ is the kinetic energy associated with the particle velocities $\{p\}$, and $W(\vec{\psi},\vec{\psi}^*)$ is the artificial biasing potential defined by equation 11. Importantly, the HMC NVE trial move proposal does not require the calculation of biasing forces on the particles due to the artificial biasing potentials. The diffusion map does not provide an explicit differentiable expression for the collective variables as a function of the particle positions, meaning that analytical expressions for these biasing forces are unavailable. Should these forces be desired to perform, for example, biased molecular dynamics, techniques such as SandCV exist to estimate these forces by approximate interpolation and basis function expansions within the low-dimensional embedding[83], or the CVs themselves could be estimated using artificial neural network approaches such as MESA that provide the necessary derivatives through automatic differentiation[84–86].

We perform umbrella sampling simulations for each candidate around each harmonic biasing potential. In each case the system is initialized from the snapshot that is closest to the center of the harmonic biasing potential in the diffusion map embedding. Next, the aggregate formed by the tagged colloid is frozen, and the system is relaxed using the Fast Inertial Relaxation Engine[87] (FIRE) algorithm until the energy of the system converges within a tolerance of 0.1ε. During the relaxation, only WCA potential is enabled. After the relaxation, the aggregate formed by the tagged colloid is unfrozen and the full Hamiltonian comprising all Lennard-Jones and WCA potentials are enabled. During the first three generations of optimization for tetrahedron, we set the harmonic constant of the biasing potentials to be 2500ε. In later iterations we relax this to 25ε. Each hybrid Monte Carlo loop is conducted at $T^* = 0.8$ and comprises 3500 steps of NVE integration with step size $dt^* = 0.005$. We perform 16000 Monte Carlo loops by equilibrating the system for the first 7000 loops and collecting data for the remaining 9000 loops. All molecular dynamics calculations are performed using HOOMD-blue[67,68]. After conducting umbrella sampling on the diffusion map space, we use BayesWHAM[88] algorithm to reconstruct the maximum a posteriori (MAP) estimate of the unbiased free energy surface for each candidate supported in the basis of the diffusion map collective variables by reweighting the biased umbrella sampling data.

2.2.4 Covariance matrix adaptation – evolution strategy

Having constructed the free energy surface for each candidate in a generation, we then employ an objective function to define their relative fitnesses. The free energy surface is first coarse-grained by its inherent structures[89] by partitioning it into the basins of attraction for local free energy minima detected by steepest descent. The free energy of the inherent structure associated with the target self-assembled aggregate βF_{target} is compared with the lowest free energy inherent structure of a competitor aggregate $\beta F_{\text{competitor}}$. The fitness of each candidate colloidal building block is defined as the free energy gap:

$$\Delta\beta F = \beta F_{\text{target}} - \beta F_{\text{competitor}}. \tag{13}$$

Minimization of this objective function seeks to make the target structure the global free energy minimum on the self-assembly free energy surface and also open up a free energy gap between the nearest metastable competing structure. Although it is a purely thermodynamic objective function, we have previously demonstrated that its optimization can lead to a funneled topography of the free energy landscape centered on the desired aggregate[52]. This topography can carry kinetic benefits in mitigating kinetically-trapped configurations and increasing both the yield and the rate of assembly of the target aggregate[52].

Having evaluated the fitness values for all candidates in a generation, we then propose new candidates by Covariance Matrix Adaptation Evolutionary Strategy[66](CMA-ES), which is a derivative-free algorithm for non-convex optimization problem. By stochastically seeding multiple walkers to probe the local topography based on running estimates of the local covariance matrix, CMA-ES has demonstrated good robustness and convergence rates on a variety of optimization problems and rugged landscapes[66,90]. Based on the fitness values of candidates in generation g, the algorithm first selects top μ candidates. Next, based on these top μ candidates, the algorithm updates the estimate of

covariance matrix C and step size σ. Then, it proposes a new generation $(g+1)$ of P candidates by:

$$\mathbf{x}^{(g+1)} = \langle \mathbf{x}^g \rangle_\mu + \sigma^g \mathcal{N}(0, C^g) \quad (14)$$

where $\mathbf{x}$ is the vector of design parameters characterizing a candidate, and $\langle \mathbf{x}^g \rangle_\mu$ is the mean value of the top μ candidates in generation g. We set $\mu = 3$, so CMA-ES will select top three candidates to propose the next generation. The optimization for the self-assembly of tetrahedral clusters for pyrochlore and cubic diamond proceeds in the two-dimensional design space $\mathbf{x} = [E_B,\ \phi_B]^T$, and for the self-assembly of the tetrahedral clusters into dimers required by the cubic diamond lattice in the three-dimensional space $\mathbf{x} = [E_D,\ E_E,\ \phi = \phi_D = \phi_E]^T$. $\mathcal{N}(0, C^g)$ is a k dimensional multivariate Gaussian random vector with mean 0 and covariance matrix C^g, where k is the dimensionality of the design space. If the standard deviation of each parameter in $\mathbf{x}$ dips lower than 1 k_BT at $T = 298$ K in the interaction strengths and 1° in the polar angle, we declare the CMA-ES to have converged and terminate the optimization. Otherwise, a new generation of candidates is proposed, and we repeat the whole optimization procedure for the new generation: conducting Langevin dynamics simulations for each new candidate, generating composite diffusion map for the new candidates, performing umbrella sampling on the composite diffusion map space, constructing the free energy surfaces and evaluating fitness values for each new candidate.

3 Results

We now proceed to describe our results for the inverse design of patchy colloids by landscape engineering to spontaneously nucleate and grow defect-free photonic lattices by a two-stage hierarchical assembly mechanism.

3.1 Inverse design of self-assembling pyrochlore lattice

3.1.1 Optimization of tetrahedral aggregate formation

We first apply landscape engineering to perform inverse design of patchy colloids to assemble pyrochlore lattice. As described in section 2.1.1, the pyrochlore lattice may be viewed as a tetrahedral network of corner-sharing tetrahedra. The optimization of the anisotropic patchy colloid design in fig. 1 proceeds in the four-dimensional design space $\{E_B, \phi_B, E_D, \phi_D\}$ defining the polar angle and interaction strength of the more strongly interacting "B" patches on the north pole that mediate high-temperature assembly of monomers into tetrahedra at $T^*_{\text{high}} = 0.8$ and the more weakly interacting "D" patches on the south pole that direct assembly of the pre-assembled tetrahedra into the pyrochlore lattice at $T^*_{\text{low}} = 0.3$. By the symmetry of the design, we may first optimize "B" patches at a reduced temperature of $T^* = 0.8$ and then obtain the corresponding parameters for "D" patches by equations 3 and 4. Thus, we optimize $\{E_B, \phi_B\}$ at $T^* = 0.8$ and the target structure is the tetrahedron.

To initialize the optimization, we generate 10 initial candidates from a multivariate Gaussian distribution centered around $(15.41\varepsilon,\ 30.0°)$ with an initial covariance matrix of $C_0 = \text{diag}(5,5)$ (i.e., a diagonal matrix with main diagonal vector (5,5)) and initial step size 1. This relatively large choice of initial covariance matrix and step size was made to favor early exploration of the design space. As detailed in section 2.2, for each candidate in each generation we perform unbiased Langevin dynamics simulations of assembly, construct composite diffusion maps, perform biased umbrella sampling – hybrid Monte Carlo simulations, estimate self-assembly free energy landscapes, and evaluate the relative fitness of each candidate. The evolution of fitness values $\Delta\beta F$ and parameters $\{E_B, \phi_B\}$ as a function of generation is shown in fig. 5. In the 16$^{\text{th}}$ generation, the parameters converge to $E_B = 15.44\varepsilon$ and $\phi_B = 30.44°$ within standard deviations of 1 k_BT at $T = 298$ K and 1°.

To see how the free energy surfaces change across generations, we select the best candidate from each generation and generate a composite diffusion map for all such candidates to provide a common set of collective variables that we can use to compare their free energy surfaces. The result is shown in fig. 6. Here we compare the free energy surfaces of the best candidates in generations 1, 7 and 15. In particular, fig. 6 (d)-(f) show the free energy surfaces of these candidates in the composite diffusion map space, and fig. 6 (a)-(c) show the partition of design space into the Voronoi cells around the candidates in each of these generations. Panels (a)-(c) show that CMA-ES draws the initial distribution of candidates down into the optimum of the fitness landscape in $\Delta\beta F$ over the course of the 16-generation optimization course. Panels (d)-(f) show that the free energy surface is sculpted such that the tetrahedron is preferentially stabilized with respect to all competitors. In the 1$^{\text{st}}$ generation the tetrahedron is the most stable aggregate but the monomer and dimer are also very stable, lying, respectively, just $+1$ k_BT and $+2$ k_BT higher in free energy. The trimer lies at $+4$ k_BT. In the 7$^{\text{th}}$ generation, the stability of the monomer relative to the tetrahedron is decreased to nearly $+3$ k_BT, but that of the dimer and trimer now lie at $+2$ k_BT. Finally in the 15$^{\text{th}}$ generation, the relative stabilities of the dimer and trimer are decreased to $+2.6$ k_BT and $+4$ k_BT, respectively, and the monomer lies at $+2.5$ k_BT, making the tetrahedron at least 2.5 k_BT more stable than all of its competitors. The net effect of the landscape engineering approach can be seen to have maximized the free energy gap (relative stability) between tetrahedron and all competing aggregates.

3.1.2 High-temperature assembly of tetrahedra and comparison with empirical geometric design

Landscape engineering discovers the optimized parameters for the "B" patches of $E_B = 15.44\varepsilon$ and $\phi_B = 30.44°$. We now proceed to verify that this design leads to the self-assembly of tetrahedral aggregates in high yield. We perform four independent unbiased Langevin dynamics simulations at $T^* = 0.8$ for 2×10^6 reduced time units for patchy colloids decorated with "B" patches employing the optimal design parameters. All simulations are initialized with 512 randomly placed and oriented particles in a cubic box of side length $L = 105.54\sigma$, corresponding to a volume fraction of $\varphi = 0.05$. The solid colored lines in fig. 7 show the temporal yield of tetrahedral aggregates as a function of time. Assuming monomers are depleted according to simple first-order kinetics, we can fit an expression for the tetrahedral yield of form $y(t) = b\left(1 - e^{-kt}\right)$,

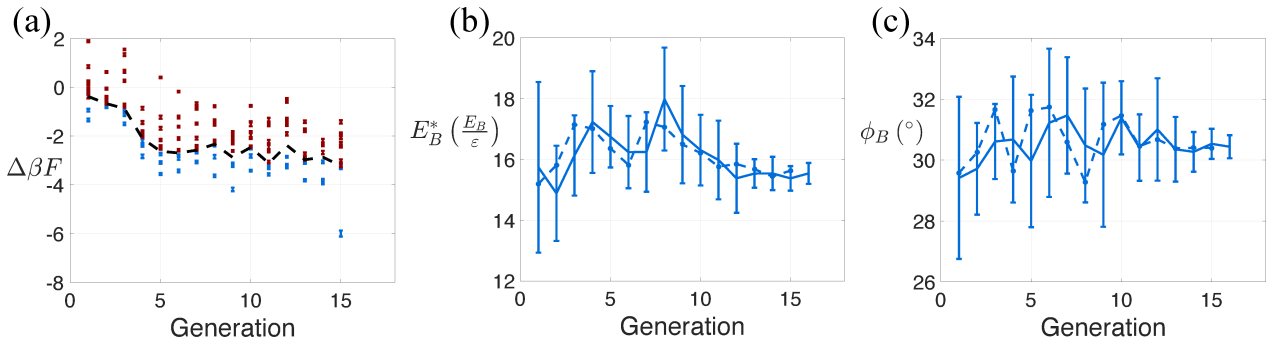

Fig. 5 Landscape engineering of the pyrochlore patchy colloid. (a) The fitness values $\Delta\beta F$ for all candidates in each generation. Error bars are estimated from the standard deviation of the fitness value of each candidate. The blue points correspond to the μ = 3 best candidates selected by CMA-ES in each generation, and the red points to those less fit candidates that are discarded. The black dashed line corresponds to the boundary between them. Evolution of (b) interaction strength E_B and (c) polar angle ϕ_B of the "B" patches as a function of generation. The solid line corresponds to the mean value among all candidates in each generation, and the dashed line corresponds to the mean value of the μ = 3 best candidates in each generation. The optimization converges after 16 generations to $E_B = 15.44\varepsilon$ and $\phi_B = 30.44°$.

where $y(t)$ is the fraction of colloids residing within tetrahedra, t is time, $b = (96.2 \pm 0.3)\%$ is the equilibrium fraction of colloids forming tetrahedral clusters, and $k = (63.4 \pm 6.7)\ \mathrm{s}^{-1}$ is the best-fit first-order rate constant.

It is also instructive to compare the assembly kinetics to that for a patchy colloid design employing the same interaction strength but an empirical patch angle based on the tetrahedral geometry. A polar angle of $\phi_B = 35.26°$ corresponds to the case where the attractive patches point directly towards the neighboring particles within an ideal tetrahedral aggregate. This can be considered the patch angle arising from zero-temperature energy minimization of an isolated tetrahedral cluster[3,91]. (We note that direct application of the landscape engineering approach at $T = 0$ K may present challenges in sampling the configurational energy landscape at absolute zero and would require the use of an alternative sampling technique to molecular dynamics such as simulated annealing or basin hopping.) The dashed lines in fig. 7 present the tetrahedral yield for particles with "B" patches of $E_B = 15.44\varepsilon$ and $\phi_B = 35.26°$. Fitting of the first-order kinetic model yields values of $b = (87.2 \pm 2.3)\%$ and $k = (38.6 \pm 6.0)\ \mathrm{s}^{-1}$, demonstrating that the optimal design discovered by landscape engineering exhibits both higher asymptotic yield and faster assembly kinetics. Analysis of the simulation trajectories shows that the ~5° larger polar angle for the empirical geometric design results in the formation of many clusters larger than tetrahedra. This can be understood as the larger polar angle of the patches enabling promiscuous interactions between the particles comprising a tetrahedral cluster and outsider particles, whereas the smaller polar angle optimized through landscape engineering disfavors the formation of these large aggregates to improve assembly rate and yield.

The landscape engineering optimization was conducted at a volume fraction $\varphi = 0.05$, but it is of interest to assess the robustness of this design in mediating high-yield tetrahedral assembly at other volume fractions. Langevin dynamics simulations conducted at volume fractions over the range φ = 0.025-0.1 reveal the tetrahedral yield to remain at 95% or better for up to two-fold increases and decreases in the volume fraction away from that at which the optimization was conducted. Very high volume fractions risk trapping within kinetically arrested glassy states, whereas very low volume fractions introduce strong entropic driving forces disfavoring assembly. At either of these extremes we anticipate that re-optimization at the volume fraction of interest would be required to maintain high tetrahedral yields.

3.1.3 Two-stage hierarchical assembly of pyrochlore lattice

After obtaining $\{E_B = 15.54\varepsilon, \phi_B = 30.44°\}$ as the optimal design parameters for "B" patches, we obtain the optimal design for the "D" patches according to equations 3 and 4:

$$\begin{aligned} \phi_D &= \phi_B = 30.44° \\ E_D &= \frac{T^*_{\mathrm{low}}}{T^*_{\mathrm{high}}} E_B = \frac{0.3}{0.8} \times 15.54\varepsilon = 5.83\varepsilon. \end{aligned} \tag{15}$$

We then decorate the patchy colloids with the optimal north pole "B" patches and south pole "D" patches to arrive at the final landscape engineering design of the patchy colloids. We validate the capacity of the design to achieve two-stage hierarchical assembly of pyrochlore lattice by locating 512 randomly placed and oriented patchy colloids in a cubic simulation box of side length $L = 105.04\sigma$, corresponding to a volume fraction of $\varphi = 0.05$. The first stage of assembly proceeds by a high-temperature hold at which the system is evolved at $T^*_{\mathrm{high}} = 0.8$ for 2×10^6 reduced time units to allow for the formation of tetrahedral clusters from the colloidal monomers. The second stage of assembly is effected by a two-stage cooling protocol to favor nucleation of the pyrochlore lattice whereby the system is rapidly cooled from $T^*_{\mathrm{high}} = 0.8$ to $T^*_{\mathrm{intermediate}} = 0.5$ for 5×10^5 reduced time units and then slowly cooled from $T^*_{\mathrm{intermediate}} = 0.5$ to $T^*_{\mathrm{low}} = 0.3$ for 1×10^7 reduced time units. Finally, the system is subjected to a low-temperature hold at $T^*_{\mathrm{low}} = 0.3$ for another 5×10^4 reduced time units to gather statistics on the terminal crystal. The evolution of system potential energy and temperature is presented in fig. 8. Nucleation of the pyrochlore lattice occurs during the second slow cooling phase at around $T^* = 0.45$ as indicated by the sudden drop in potential energy corresponding to the latent heat of crystallization.

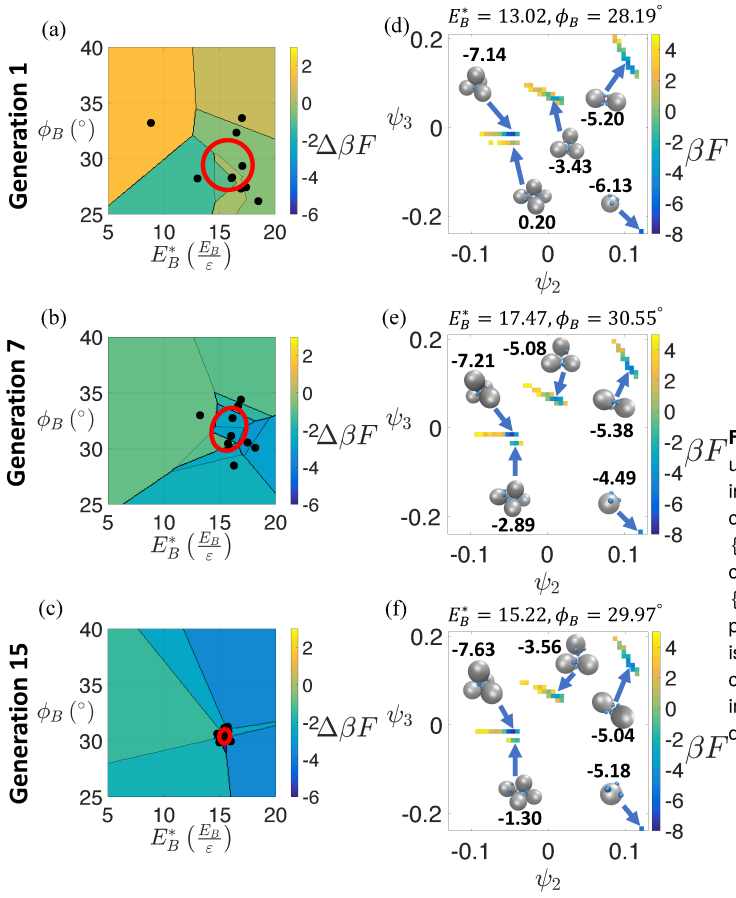

Fig. 6 Landscape engineering sculpting of the self-assembly free energy landscape for tetrahedral cluster formation. (a)-(c) Distribution of candidates within the $\{E_B, \phi_B\}$ design space in generations 1, 7 and 15. The candidates are represented by black dots. The red circle represents the CMA-ES covariance matrix from which the candidates in the current generation are sampled. The top $\mu = 3$ candidates in current generation are then used to update the covariance matrix. For visualization purposes, we partition design space into Voronoi cells around each candidate and color each cell by the fitness $\Delta\beta F$ of the corresponding candidate. (d)-(f) Free energy surfaces of the best candidates in generations 1, 7 and 15 in the composite diffusion map space spanned by the leading two diffusion map collective variables $\{\psi_2, \psi_3\}$. The particular values of $\{E_B, \phi_B\}$ pertaining to each candidate are listed above each panel. Representative aggregates from the local free energy minima are projected onto the low-dimensional embedding. The values of the local free energy minima associated with each aggregate are displayed next to the representative structures.

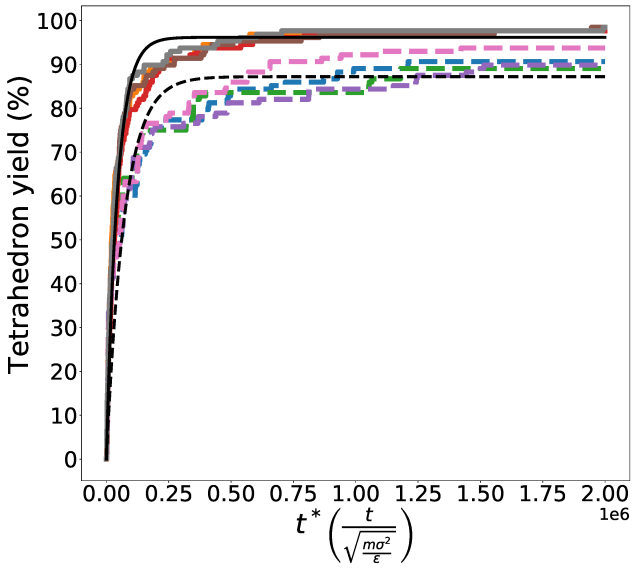

Fig. 7 Yield of tetrahedral clusters as a function of simulation time at $T^* = 0.8$ in unbiased Langevin dynamics simulations. Each colored line corresponds to an independent simulation. The four colored solid lines correspond to patchy colloids decorated with "B" patches of the optimal design $\{E_B = 15.54\varepsilon, \phi_B = 30.44^\circ\}$ deduced by landscape engineering. The four colored dashed lines correspond to "B" patches with $\{E_B = 15.54\varepsilon, \phi_B = 35.26^\circ\}$ employing the same interaction strength but a polar angle corresponding to the zero-temperature energy minimum of an isolated tetrahedral cluster. The black solid and dashed lines are fits of the corresponding data to first-order kinetics. Landscape engineering discovers an improved particle design exhibiting better yield and assembly rate beyond that derived from purely geometric considerations.

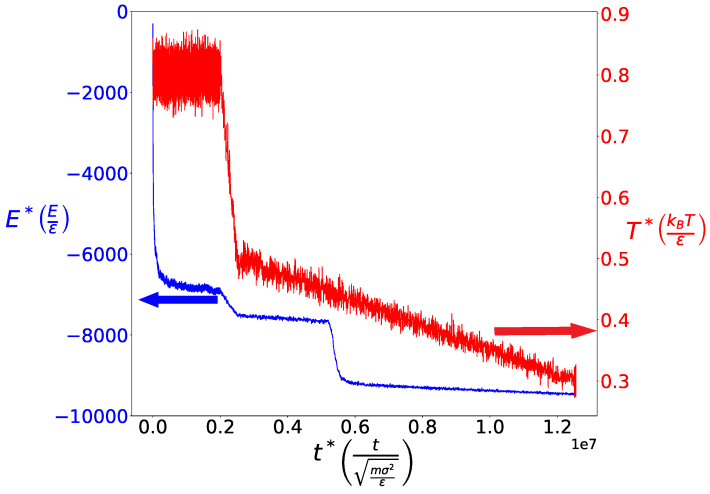

Fig. 8 Evolution of system potential energy and temperature for two-stage hierarchical assembly of pyrochlore lattice. The two horizontal arrows indicate which of the two axes – potential energy or temperature – pertain to each curve on this double y-axis plot.

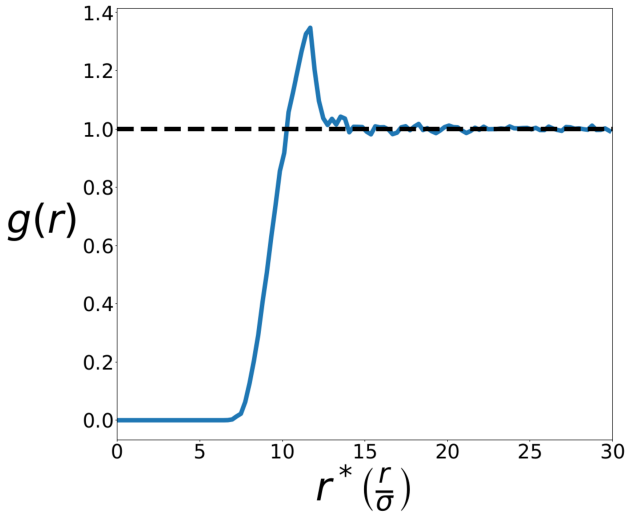

Fig. 9 Radial distribution function between the geometric centers of tetrahedral clusters at the end of the high-temperature assembly stage of pyrochlore lattice.

At the end of the $T^*_{\text{high}} = 0.8$ high-temperature assembly stage the yield of tetrahedral clusters is 97.7% corresponding to the formation of 125 tetrahedral clusters mediated by interactions between the north pole "B" patches. The radial distribution function $g(r)$ between the geometric centers of the tetrahedral clusters demonstrates that they behave effectively as an ideal gas (fig. 9). The small correlation peak at $r^* \approx 12$ indicating the presence of very weak structural correlations between the tetrahedra is largely attributable to the weak interactions between the south pole "D" patches, but numerical simulations in which the "D"-"D" interactions are turned off show that ~20% of the correlation peak can be attributed to effective entropic attractions driven by excluded volume interactions. This validates the design expectation that interactions between "D" patches should be thermally decoupled from that of the "B" patches.

During the slow cooling process we observe nucleation and subsequent growth of a pyrochlore lattice (fig. 10a). We note that sufficiently slow cooling rates are necessary to assure a single nucleation event and production of defect-free crystal. In experimental realizations employing orders of magnitude more colloids than our simulations it can be quite challenging to achieve defect-free crystals. We anticipate that very slow cooling rates, possibly coupled with programmed temperature oscillations to heal defects, may be required to obtain high-fidelity periodic crystal lattices. At the end of the $T^*_{\text{low}} = 0.3$ low-temperature hold, we perform structural characterization of the crystal. To do so we compute the radial distribution function $g(r)$ between the patchy colloid centers of mass (fig. 10b) and Steinhardt bond order parameters[92,93] $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$ (fig. 10c) and $q_4(i)$ (fig. 10d), where "$*$" denotes the complex conjugate. The vector $\vec{q}_l(i)$ is a $(2l+1)$ dimensional vector whose non-normalized elements are:

$$q_{lm}(i) = \frac{1}{N_b(i)}\sum_{k=1}^{N_b(i)} Y_{lm}(\hat{r}_{ik}) = \frac{1}{N_b(i)}\sum_{k=1}^{N_b(i)} Y_{lm}(\theta_{ik},\phi_{ik}) \qquad (16)$$

where $N_b(i)$ is the number of nearest neighbors of particle i, k loops over all such nearest neighbors, $\hat{r}_{ik}$ is the unit displacement vector from particle i to particle k, $\{\theta_{ik},\phi_{ik}\}$ are the polar and azimuthal angles that $\hat{r}_{ik}$ makes with respect to a specific coordinate system, and Y_{lm} are the spherical harmonics. A nearest neighbor is defined as a particle lying within a cutoff distance $d_{\text{cut}} = 6.0\sigma$, where this threshold is calibrated to cover the first peak in the radial distribution function at $r_{\text{peak},1} \approx 5.25\sigma$ (fig. 10b). When computing the inner product $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$, we normalize each vector to have unit 2-norm. It can be shown that the inner product, $\vec{q}^*_l(i)\cdot\vec{q}_l(j)$, between two particles i and j is real and independent of the coordinate system[92,94]. The parameter $q_4(i)$ is defined using the relation[93]:

$$q_l(i) = \sqrt{\frac{4\pi}{2l+1}\sum_{m=-l}^{l}|q_{lm}(i)|^2} \qquad (17)$$

The parameter $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$ defined between nearest neighbor pairs $\{i,j\}$ has been shown to be able to distinguish between pyrochlore lattice and hexagonal tetrastack lattice[37] by aggregating bond angle information between pairs of particles separated by up to three bonds. The parameter $q_4(i)$ provides a more localized structural characterization by averaging over only the nearest neighbors of particle i and providing a way to tell whether a particle exists in a locally staggered or eclipsed configuration. Taken together $g(r)$, $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$, and $q_4(i)$ allow us to determine whether the system adopts the radial and angular order expected for a pure pyrochlore lattice, and whether it contains crystal defects or is a mixture of the pyrochlore and hexagonal tetrastack polymorphs.

The radial distribution function computed over the final snapshot of the simulation at $T^*_{\text{low}} = 0.3$ possesses exactly the characteristic peaks indicative of a pyrochlore lattice (fig. 10b). The distributions of $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$ (fig. 10c) and $q_4(i)$ (fig. 10d) also both possess peaks at the positions expected for the ideal pyrochlore lattice, and lack those expected for the hexagonal tetrastack. Specifically, the pyrochlore lattice will possess only one peak at 0.0123 in $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$ and one peak at 0.375 in $q_4(i)$[37]. In contrast, the hexagonal tetrastack will show two peaks at 0.0123 and –0.5 in $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$ and two peaks at 0.375 and 0.181 in $q_4(i)$[37]. The gray bars in fig. 10c show the distribution of $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$ for all nearest neighbor pairs of patchy colloids. Due to finite-size effects, in some nearest neighbor pairs the constituent colloids lie on the boundary of the crystal structure and do not have exactly six bonded nearest neighbors. These finite-size effects cause the second small peak at $\vec{q}^*_4(i)\cdot\vec{q}_4(j) \approx 0.5$ to emerge. The blue transparent bars show the distribution $\vec{q}^*_4(i)\cdot\vec{q}_4(j)$ among nearest neighbor pairs in which both colloids are constrained to have six bonded nearest neighbors (i.e., crystalline colloids). Here we see that the blue transparent bars are indeed centered around the characteristic peak in ideal infinite pyrochlore lattice and the small shoulder disappears. Similarly, the distribution of $q_4(i)$ in fig. 10d restricted to six-neighbor crystalline colloids possesses a single peak centered on the pyrochlore result. Taken together, fig. 10 demonstrates that the colloids do spontaneously form a defect-free pyrochlore lattice, and not a mixture of the pyrochlore

and hexagonal tetrastack polymorphs.

Having characterized the final structure, we then proceed to compute the band structure of the corresponding periodic crystal using the MIT Photonic Bands (MPB) software[95]. We create an infinite periodic pyrochlore lattice from the primitive lattice vectors $\vec{R}_1 = (0, \frac{a}{2}, \frac{a}{2}), \vec{R}_2 = (\frac{a}{2}, 0, \frac{a}{2}), \vec{R}_3 = (\frac{a}{2}, \frac{a}{2}, 0)$ where a is the lattice constant. The positions of the four basis particles reported in the basis of the primitive lattice vectors (i.e., (l,m,n) denotes a position vector $l\vec{R}_1 + m\vec{R}_2 + n\vec{R}_3$) are given in table 1. The lattice constant a is related to the first peak in radial distribution function (nearest-neighbor distance) $r_{\text{peak},1}$ by $a = \frac{4}{\sqrt{2}} r_{\text{peak},1}$. We estimate the nearest-neighbor distance from the radial distribution function of final configuration (fig. 10b) to be $r_{\text{peak},1} \approx 5.25\sigma$. The radius of the colloidal particles is that of "A" spheres $\sigma_A/2 = 2.5\sigma$ and the dielectric constant is set to $\epsilon_r = 12.0$ corresponding to the value for silicon. The medium is taken to be air. We use a $16\times16\times16$ grid to discretize the primitive unit cell to compute the band structure along the high-symmetry lines in the first Brillouin zone. We verify that our results are converged with respect to the grid spacing. The resulting photonic band structure is shown in Figure 11. The band structure shows the opening of an indirect bandgap between the second and third bands with a width-to-midgap ratio (ratio between the bandgap width and the midgap frequency) of 4.63%.

Basis particle index	Position
1	$\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$
2	$\left(0, \frac{1}{2}, \frac{1}{2}\right)$
3	$\left(\frac{1}{2}, 0, \frac{1}{2}\right)$
4	$\left(\frac{1}{2}, \frac{1}{2}, 0\right)$

Table 1 Positions of basis particles of the pyrochlore lattice in the basis of primitive lattice vectors $\vec{R}_1 = (0, \frac{a}{2}, \frac{a}{2}), \vec{R}_2 = (\frac{a}{2}, 0, \frac{a}{2}), \vec{R}_3 = (\frac{a}{2}, \frac{a}{2}, 0)$ where a is the lattice constant. The tuple (l,m,n) denotes a particle position vector $l\vec{R}_1 + m\vec{R}_2 + n\vec{R}_3$.

3.2 Inverse design of self-assembling cubic diamond lattice

3.2.1 Optimization of staggered dimer formation

Following the success in pyrochlore assembly, we then apply our landscape engineering approach to design a new patchy colloid to assemble cubic diamond lattice via tetrahedral clusters. As described in section 2.1.2 and illustrated in fig. 3, the cubic diamond lattice comprises tetrahedral clusters arranged in staggered dimers. The high-temperature assembly of patchy colloids into tetrahedra proceeds in exactly the same fashion as for pyrochlore, and we adopt $E_B = 15.44\varepsilon$ and $\phi_B = 30.44°$ as the optimal design solution for the "B" patches. The design problem then reduces to optimization of the interaction strengths and polar angles of the south pole "D" and "E" patches to mediate the low-temperature assembly of the pre-assembled tetrahedral aggregates into cubic diamond lattice. We choose to optimize the parameters for "D" and "E" patches at $T^*_{\text{high}} = 0.8$, and then scale down the interaction strengths to match the low temperature phase $T^*_{\text{low}} = 0.3$ by a factor of $T^*_{\text{low}}/T^*_{\text{high}}$. Thus, we optimize $\{E_D, E_E, \phi = \phi_D = \phi_E\}$ at $T^* = 0.8$ and the target structure is a staggered dimer.

We initialize the optimization by generating 10 initial candidates from a multivariate Gaussian distribution centered around $(6.67\varepsilon, 6.67\varepsilon, 26.60°)$ with an initial covariance matrix of $C_0 = \text{diag}(5,5,5)$ and an initial step size 1. The evolution of $\{E_D, E_E, \phi\}$ and the fitness $\Delta\beta F$ over the landscape engineering generations are shown in fig. 12. In the 18$^{\text{th}}$ generation, all parameters have converged to $E_D = 10.02\varepsilon$, $E_E = 11.64\varepsilon$ and $\phi = 26.68°$ within standard deviations of 1 k_BT at $T = 298$ K and 1°.

The distribution of candidates within the design space and free energy surfaces for the best candidates in generations 1, 9 and 17 are presented in fig. 13 to show how landscape engineering changes the design and assembly properties of the building block over the course of the optimization. In fig. 13a-i we partition the design spaces $\{\phi, E_E\}$, $\{\phi, E_D\}$ and $\{E_E, E_D\}$ by the Voronoi cells around the candidates in generations 1, 9 and 17, and we color each Voronoi cell by the fitness value of the corresponding candidate. Here we observe that despite the relatively poor initial guesses for E_D and E_E, CMA-ES was able to efficiently move the mean and shrink the variance of subsequent generations of candidates to converge to the optimum of the $\Delta\beta F$ fitness landscape. In fig. 13j-l we show that the self-assembly free energy surfaces are driven towards a topography in which the staggered dimer is preferentially stabilized relative to all competing aggregates. In the 1$^{\text{st}}$ generation, the monomer is the most stable aggregate lying (-1) k_BT lower in free energy than the dimer. The trimer and tetrahedron are each less stable than the dimer, lying, respectively, +4 k_BT and +8 k_BT higher in free energy. In the 9$^{\text{th}}$ generation, the landscape engineering protocol has successfully rendered the dimer the most stable aggregate on the landscape, with the monomer, trimer, and tetrahedron lying, respectively, +5 k_BT, +5 k_BT, and +6 k_BT higher in free energy. In the 17$^{\text{th}}$ generation, the dimer has been even further stabilized, with the trimer and tetrahedron each lying +6 k_BT higher in free energy, and the monomer rendered completely unstable within the sampling resolution of our calculations.

3.2.2 High-temperature assembly of dimers

We verify the optimal landscape engineering design of $E_D = 10.02\varepsilon$, $E_E = 11.64\varepsilon$ and $\phi = 26.68°$ by performing four unbiased Langevin dynamics simulations at $T^* = 0.8$ for 2×10^6 reduced time units for patchy colloids decorated with "D" and "E" patches. Simulations are initialized with 512 colloidal monomers with random positions and orientations in a cubic simulation box with side length $L = 105.04\sigma$ corresponding to a volume fraction of $\varphi = 0.05$. The yield of staggered dimers as a function of time for the four runs is presented in fig. 14. Fitting the first-order kinetic model for the dimer yield $y(t) = b\left(1 - e^{-kt}\right)$ results in best-fit constants of $k = (459.9 \pm 29.1)\text{s}^{-1}$ and $b = (99.1 \pm 0.1)\%$, demonstrating that this design produces staggered dimers with nearly quantitative yield.

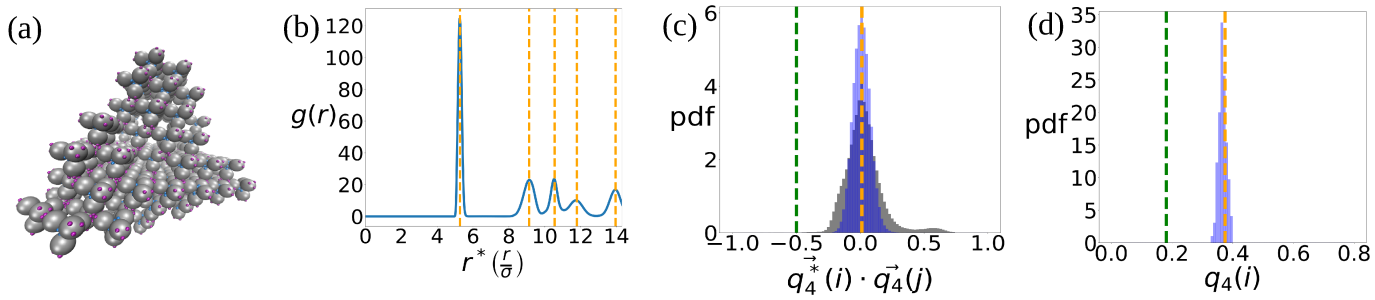

Fig. 10 Structural characterization of the self-assembled pyrochlore lattice. (a) Snapshot of terminal crystal lattice structure. (b) Radial distribution function $g(r)$ between the patchy colloid centers of mass. (c) The distribution of $\vec{q}_4^*(i) \cdot \vec{q}_4(j)$ computed between all patchy colloid nearest-neighbor pairs (gray) and restricted to crystalline colloid pairs (blue) defined as those in which each partner possesses six bonded nearest neighbors. (d) The distribution of $q_4(i)$ for crystalline colloids. In all panels the orange dashed lines represent the expected peak positions for an ideal pyrochlore lattice and the green dashed lines represent those for an ideal hexagonal tetrastack lattice.

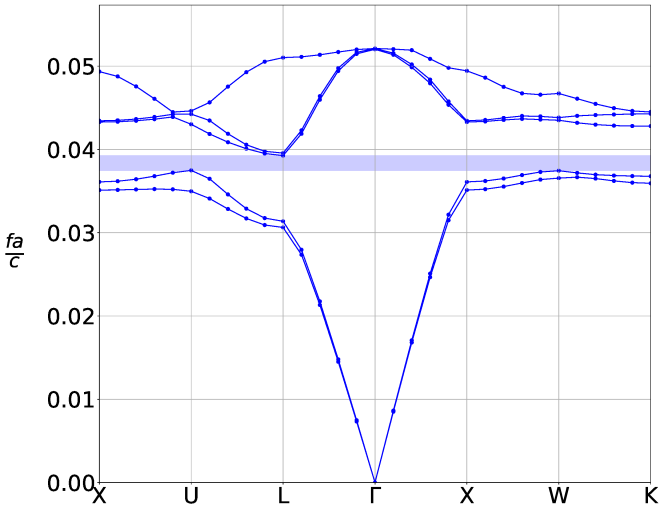

Fig. 11 Photonic band structure of the self-assembled pyrochlore lattice. The y-axis reports the dimensionless frequency $\frac{fa}{c}$, where f is the frequency, a is lattice constant, and c is the speed of light in vacuum. The x-axis labels the corners of the irreducible region of first Brillouin zone in canonical order. For a lattice constant of $a = 2.97$ μm corresponding to our choice of length scale, the bandgap lies within the frequency range of $3.74 < f < 3.94$ THz and wavelength range of $76.2 < \lambda < 80.3$ μm, placing the bandgap in the infrared regime of electromagnetic spectrum. For a lattice constant of $a = 27.0$ nm the bandgap lies within the frequency range of $411 < f < 433$ THz and wavelength range of $692 < \lambda < 729$ nm, placing the bandgap in the visible regime.

3.2.3 Two-stage hierarchical assembly of cubic diamond lattice

Landscape engineering furnished $\{E_D = 10.02\varepsilon, E_E = 11.64\varepsilon, \phi = 26.68°\}$ as the optimal values of design parameters for "D" and "E" patches at $T^*_{\text{high}} = 0.8$. We proportionally scale these interaction strengths by a factor of $T^*_{\text{low}}/T^*_{\text{high}} = 0.3/0.8$ in order to thermally decouple the "D" and "E" interactions from the "B" interactions such that they direct assembly of tetrahedral clusters into cubic diamond lattice at the second, low-temperature stage of assembly. This results in optimal "D" and "E" patch designs of $\{E_D = \frac{T_{\text{low}}}{T_{\text{high}}} 10.02\varepsilon = 3.76\varepsilon$, $E_E = \frac{T_{\text{low}}}{T_{\text{high}}} 11.64\varepsilon = 4.36\varepsilon, \phi = 26.68°\}$. We test our design in simulations of 512 randomly placed and oriented colloids in a cubic simulation box with side length $L = 132\sigma$, corresponding to a volume fraction of $\varphi = 0.025$. We first evolve the system at high temperature $T^*_{\text{high}} = 0.8$ for 2×10^6 reduced time units, then quickly cool the system to $T^*_{\text{intermediate}} = 0.6$ for 5×10^5 reduced time units, then slowly cool the system down to $T^*_{\text{low}} = 0.3$ for 1.5×10^7 reduced time units, and finally equilibrate the system at $T^*_{\text{low}} = 0.3$ for 5×10^4 reduced time units to gather statistics. The evolution of system potential energy and temperature versus simulation time is shown in fig. 15. Nucleation of the cubic diamond lattice occurs at around $T^* = 0.48$ as indicated by the sudden drop in potential energy.

At the termination of the $T^*_{\text{high}} = 0.8$ high-temperature hold the yield of tetrahedral clusters is 95%. The radial distribution function between the geometric centers of tetrahedral cluster demonstrates that they behave as an effective ideal gas with only a small correlation peak due to weak "D" and "E" patch interactions (fig. 16). A snapshot of the structure formed at the end of the $T^*_{\text{low}} = 0.3$ low-temperature hold is presented in fig. 17a. The radial distribution function (fig. 17b), the distribution of $\vec{q}_3^*(i) \cdot \vec{q}_3(j)$ (fig. 17c), and distribution of $\vec{q}_3(i)$ (fig. 17d) between the geometric centers of tetrahedral clusters all show peaks at precisely the expected locations for cubic diamond lattice, and no peaks at the locations for hexagonal diamond lattice. In calculating the Steinhardt bond order parameters, a pair of tetrahedral clusters are defined as nearest neighbors if their geometric centers lie within a cutoff distance $d_{\text{cut}} = 13.0\sigma$ calibrated to cover the first peak in the radial distribution function at $r_{\text{peak},1} \approx 12.05\sigma$ (fig. 17b).

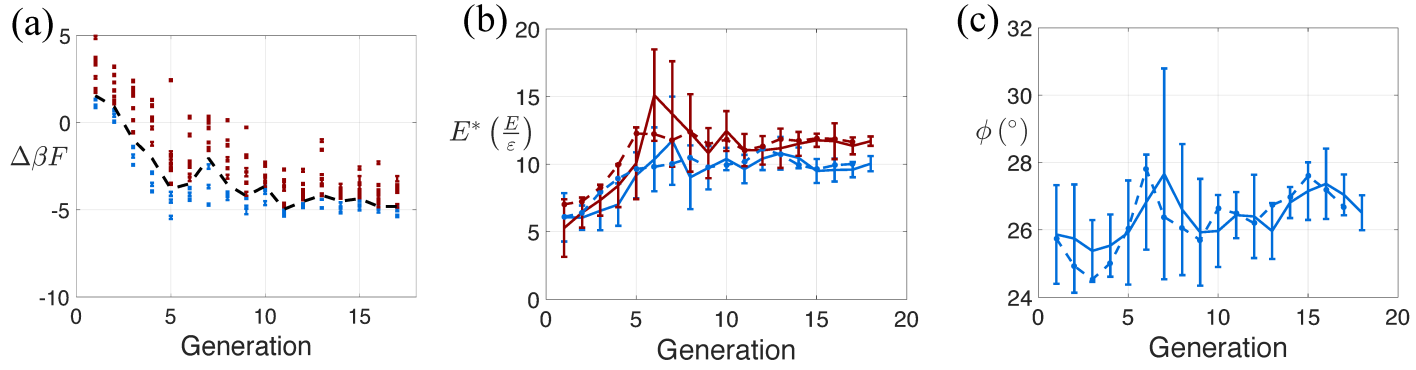

Fig. 12 Landscape engineering of the cubic diamond patchy colloid. (a) The fitness values $\Delta\beta F$ for all candidates in each generation. Error bars are estimated from the standard deviations in the fitness values corresponding to each candidate. The blue points correspond to the μ = 3 best candidates selected by CMA-ES in each generation, and the red points to those less fit candidates that are discarded. The black dashed line corresponds to the boundary between them. Evolution of (b) interaction strengths E_D (blue) and E_E (red) and (c) polar angle $\phi = \phi_D = \phi_E$ as a function of generation. The solid line corresponds to the mean value among all candidates in each generation, and the dashed line corresponds to the mean value of the μ = 3 best candidates in each generation. The optimization converges after 18 generations to $E_D = 10.02\varepsilon, E_E = 11.64\varepsilon$ and $\phi = 26.68°$.

In fig. 17c the gray bars correspond to $\vec{q}_3^*(i)\cdot\vec{q}_3(j)$ computed for all pairs of tetrahedral clusters, and the blue bars correspond to the values computed for pairs of crystalline tetrahedral clusters defined as those in which each cluster has four bonded nearest neighbors. It is clear that the finite-size effect causes the distribution of gray bars to spread out, but the blue bars are centered on the expected peak location for ideal cubic diamond lattice. The value of $q_3(i)$ is the same for the ideal cubic and hexagonal diamond lattices, so this measure possesses no discriminatory power between the two but does show the $q_3(i)$ distribution fot the self-assembled lattice to be located in exactly the expected location (fig. 17d). This structural characterization verifies that the tetrahedral clusters have assembled a defect-free cubic diamond lattice and it is not a mixture of the diamond and hexagonal polymorphs.

The photonic band structure of the assembled cubic diamond lattice is determined by defining an infinite periodic lattice with primitive lattice vectors $\vec{R}_1 = (0,\frac{a}{2},\frac{a}{2}), \vec{R}_2 = (\frac{a}{2},0,\frac{a}{2}), \vec{R}_3 = (\frac{a}{2},\frac{a}{2},0)$, where a is the lattice constant. The eight basis particles within the $\{\vec{R}_1,\vec{R}_2,\vec{R}_3\}$ basis are given in table 2. The lattice constant is related to the first peak in radial distribution function between colloidal geometric centers $r_{\text{peak},1}$ as $a = \frac{4r_{\text{peak},1}}{\sqrt{2}}\frac{2}{2+\sqrt{2-2\cos(109.5°)}}$, where we take $r_{\text{peak},1} \approx 5.25\sigma$ as estimated for pyrochlore. Using the same parameters as for the pyrochlore calculation we employ MPB[95] to obtain the band structure in fig. 18. We observe an indirect bandgap between the second and third bands with a width-to-midgap ratio of 7.47%.

Basis particle index	Position
1	$(\frac{1}{8},\frac{1}{8},\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(-\frac{1}{8},-\frac{1}{8},-\frac{1}{8})$
2	$(\frac{1}{8},\frac{1}{8},\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(\frac{3}{8},-\frac{1}{8},-\frac{1}{8})$
3	$(\frac{1}{8},\frac{1}{8},\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(-\frac{1}{8},\frac{3}{8},-\frac{1}{8})$
4	$(\frac{1}{8},\frac{1}{8},\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(-\frac{1}{8},-\frac{1}{8},\frac{3}{8})$
5	$(-\frac{1}{8},-\frac{1}{8},-\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(\frac{1}{8},\frac{1}{8},\frac{1}{8})$
6	$(-\frac{1}{8},-\frac{1}{8},-\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(-\frac{3}{8},\frac{1}{8},\frac{1}{8})$
7	$(-\frac{1}{8},-\frac{1}{8},-\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(\frac{1}{8},-\frac{3}{8},\frac{1}{8})$
8	$(-\frac{1}{8},-\frac{1}{8},-\frac{1}{8})+\frac{2}{2+\sqrt{2-2\cos(109.5°)}}(\frac{1}{8},\frac{1}{8},-\frac{3}{8})$

Table 2 Positions of basis particles of the cubic diamond lattice of tetrahedral clusters in the basis of primitive lattice vectors $\vec{R}_1 = (0,\frac{a}{2},\frac{a}{2}), \vec{R}_2 = (\frac{a}{2},0,\frac{a}{2}), \vec{R}_3 = (\frac{a}{2},\frac{a}{2},0)$ where a is the lattice constant. A tuple (l,m,n) denotes a particle position vector $l\vec{R}_1 + m\vec{R}_2 + n\vec{R}_3$.

4 Conclusions

In this work we have demonstrated an automated data-driven strategy for the inverse design of colloidal particles capable of spontaneous self-assembly into periodic crystals. This approach, termed landscape engineering, combines molecular simulation, enhanced sampling, and nonlinear dimensionality reduction to efficiently estimate self-assembly free energy landscapes, and the use of evolutionary algorithms to rationally sculpt the topography of the landscape to stabilize desired aggregates by manipulation of the building block design parameters. We demonstrated the technique in the successful design of anisotropic patchy colloids to self-assemble pyrochlore and cubic diamond lattices of tetrahedral clusters as highly sought-after optical materials possessing omnidirectional photonic bandgaps. Our approach presents a principled and constructive means to reverse engineer the optimal building block design. This systematic approach can accelerate design relative to Edisonian trial-and-improvement and avoid traps associated with flawed intuition. The approach can be straightforwardly generalized to arbitrary particle designs and lattice structures, including crystals formed from multiple particle types and even quasicrystals, by identifying the self-assembled aggregates and interfaces to be stabilized and the design variables to be manipulated.

We adopted a relatively simple and generic model of our patchy particles that introduces anisotropy through the precise placement of specific and attractive patches. The interaction potentials we employed (Lennard-Jones and WCA) were deliberately simple in form but sufficient to capture the essential physics of assembly. These patchy particle potentials[3] and similar anisotropic Kern-Frenkel models[11,37] can be considered rude models of the

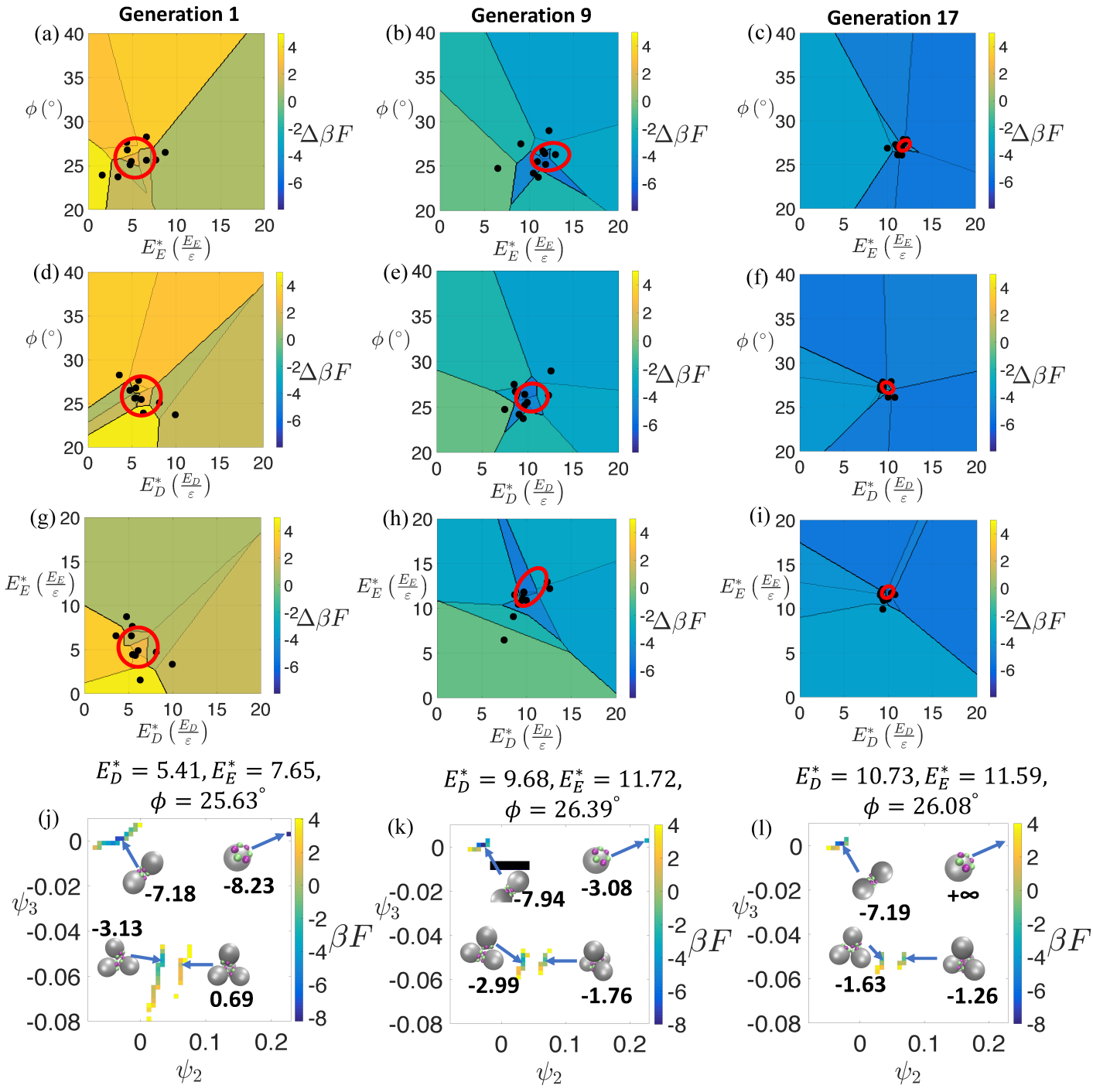

Fig. 13 Landscape engineering sculpting of the self-assembly free energy landscape for the formation of staggered dimers. (a)-(i) Distribution of candidates within the $\{\phi, E_E\}$, $\{\phi, E_D\}$ and $\{E_E, E_D\}$ design space in generations 1, 9 and 17. The candidates are represented by black dots. The red circle represents the CMA-ES covariance matrix from which the candidates in the current generation are sampled. The top $\mu = 3$ candidates in current generation are then used to update the covariance matrix. For visualization purposes, we partition design space into Voronoi cells around each candidate and color each cell by the fitness $\Delta\beta F$ of the corresponding candidate. (j)-(l) Free energy surfaces of the best candidates in generations 1, 9 and 17 in the composite diffusion map space spanned by the leading two diffusion map collective variables $\{\psi_2, \psi_3\}$. The particular values of $\{E_D, E_E, \phi\}$ pertaining to each candidate are listed above each panel. Representative aggregates from the local free energy minima are projected onto the low-dimensional embedding. The values of the local free energy minima associated with each aggregate are displayed next to the representative structures.

anisotropy introduced by current experimental fabrication techniques such as glancing angle deposition[30,53,54], grafting of complementary DNA oligomers[10,35,55–57] and contact layer lithography[32]. In follow-on work it would be of interest to employ more realistic potentials designed to more closely mimic experimentally-realizable interaction potentials and particle designs[10,57], and incorporate the limits of fabrication robustness and precision by considering polydispersity in the building block

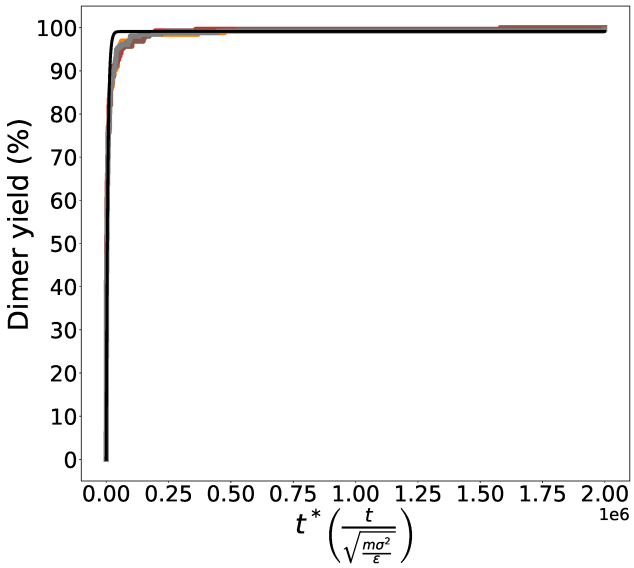

Fig. 14 Yield of staggered tetrahedral dimers as a function of simulation time at $T^* = 0.8$ in unbiased Langevin dynamics simulations. Each colored line corresponds to an independent simulation at the optimal design $\{E_D = 10.02\varepsilon, E_E = 11.64\varepsilon, \phi = 26.68°\}$ deduced by landscape engineering. The solid black line is the fit to first-order kinetics.

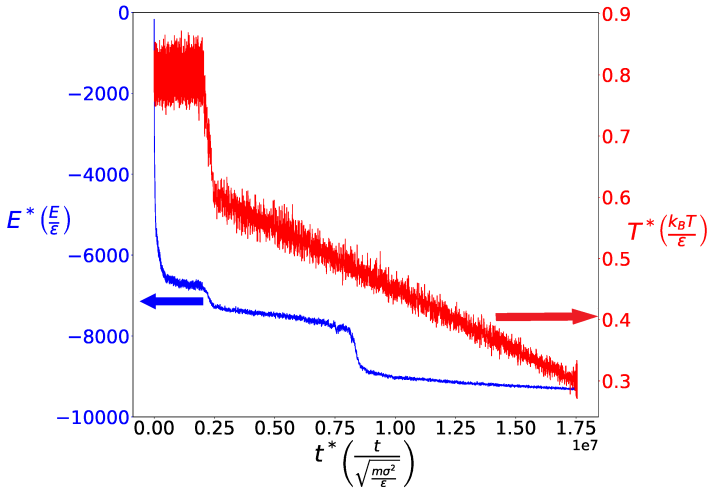

Fig. 15 Evolution of system potential energy and temperature for two-stage hierarchical assembly of cubic diamond lattice. The two horizontal arrows indicate which of the two axes – potential energy or temperature – pertain to each curve on this double y-axis plot.

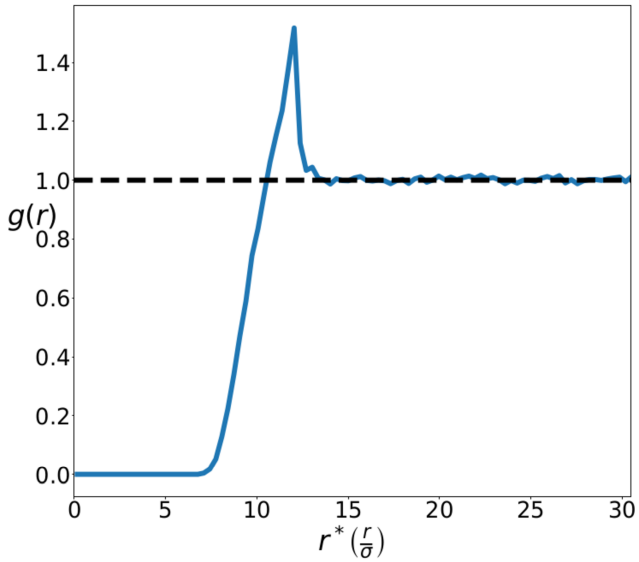

Fig. 16 Radial distribution function between the geometric centers of tetrahedral clusters at the end of the high-temperature assembly stage of cubic diamond.

ensemble[11,37]. Surface charge is another valuable means of controlling assembly[15,29,96–99], and these effects can be modeled through screened Coulomb interactions[99–101]. Also, the rational design strategy used in the current work may be extended to design patchy colloids decorated with nanodots that may form helical structures[7], which are fundamental building blocks for chiral photonic crystals[102]. Possessing omnidirectional photonic bandgaps and the capacity to circularly polarize light, these materials have potential applications as chiral beamsplitters and components of photonic computers[102]. Finally, we envisage extension of the landscape engineering inverse design strategy beyond anisotropic colloids to self-assembling nanoparticles, liquid crystals, or molecules.

Conflicts of Interest

There are no conflicts of interest to declare.

Acknowledgments

We thank Dr. Andrew W. Long for fruitful discussions. This material is based upon work supported by the National Science Foundation under Grant No. DMR-1841800.

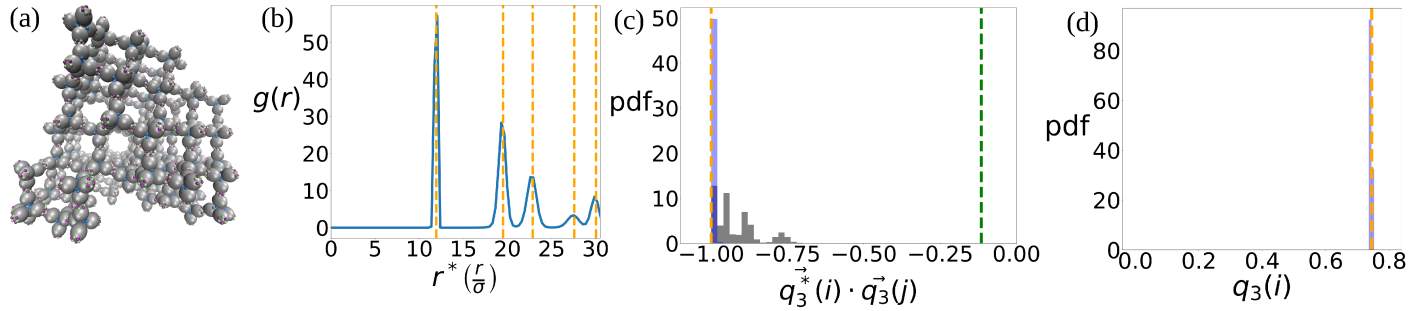

Fig. 17 Structural characterization of the self-assembled cubic diamond lattice. (a) Snapshot of terminal crystal lattice structure. (b) Radial distribution function $g(r)$ between the geometric centers of the tetrahedral cluster. (c) The distribution of $\vec{q}_3^*(i) \cdot \vec{q}_3(j)$ computed between all pairs of tetrahedral clusters (gray) and restricted to pairs of crystalline tetrahedral clusters (blue) defined as those in which each partner has four bonded nearest neighbors. (d) The distribution of $q_3(i)$ for pairs of crystalline tetrahedral clusters. In all panels the orange dashed lines represent the expected peak positions for an ideal cubic diamond lattice and the green dashed lines represent those for an ideal hexagonal diamond lattice. In the last panel the orange and green dashed lines are coincident.

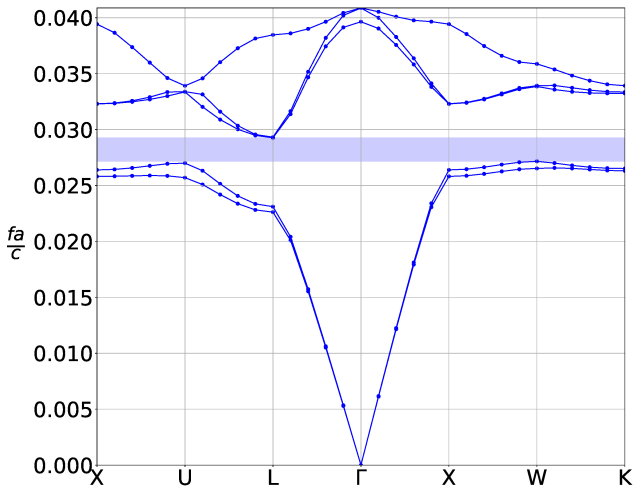

Fig. 18 Photonic band structure of the self-assembled cubic diamond lattice. The y-axis reports the dimensionless frequency $\frac{fa}{c}$, where f is the frequency, a is lattice constant, and c is the speed of light in vacuum. The x-axis labels the corners of the irreducible region of first Brillouin zone in canonical order. For lattice constant of $a = 1.63\ \mu$m corresponding to our choice of length scale, the bandgap lies within the frequency range of $4.95 < f < 5.32$ THz and wavelength range of $56.4 < \lambda < 60.6\ \mu$m, placing the bandgap to be in the infrared regime of electromagnetic spectrum. For a lattice constant of $a = 16.3$ nm the bandgap lies within the frequency range of $495 < f < 532$ THz and wavelength range of $564 < \lambda < 605$ nm, placing the bandgap in the visible regime.

References

1 A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller and P. Wong, *J. Chem. Phys.*, 2007, **127**, 085106.

2 A. W. Long and A. L. Ferguson, *J. Phys. Chem. B*, 2014, **118**, 4228–4244.

3 Z. Zhang and S. C. Glotzer, *Nano Lett.*, 2004, **4**, 1407–1413.

4 E. Lauga and M. P. Brenner, *Phys. Rev. Lett.*, 2004, **93**, 238301.

5 J. T. McGinley, Y. Wang, I. C. Jenkins, T. Sinno and J. C. Crocker, *ACS Nano*, 2015, **9**, 10817–10825.

6 A. J. Williamson, A. W. Wilber, J. P. Doye and A. A. Louis, *Soft Matter*, 2011, **7**, 3423–3431.

7 R. Guo, J. Mao, X.-M. Xie and L.-T. Yan, *Sci. Rep.*, 2014, **4**, 7021.

8 H. Pattabhiraman, G. Avvisati and M. Dijkstra, *Phys. Rev. Lett.*, 2017, **119**, 157401.

9 N. A. Mahynski, L. Rovigatti, C. N. Likos and A. Z. Panagiotopoulos, *ACS Nano*, 2016, **10**, 5459–5467.

10 É. Ducrot, M. He, G.-R. Yi and D. J. Pine, *Nat. Mater.*, 2017, **16**, 652.

11 D. Morphew, J. Shaw, C. Avins and D. Chakrabarti, *ACS Nano*, 2018, **12**, 2355–2364.

12 Z. Zhang, A. S. Keys, T. Chen and S. C. Glotzer, *Langmuir*, 2005, **21**, 11547–11551.

13 F. Romano, E. Sanz and F. Sciortino, *J. Chem. Phys.*, 2011, **134**, 174502.

14 A.-P. Hynninen, J. H. Thijssen, E. C. Vermolen, M. Dijkstra and A. Van Blaaderen, *Nat. Mater.*, 2007, **6**, 202.

15 A. M. Kalsin, M. Fialkowski, M. Paszewski, S. K. Smoukov, K. J. Bishop and B. A. Grzybowski, *Science*, 2006, **312**, 420–424.

16 E. G. Noya, C. Vega, J. P. Doye and A. A. Louis, *J. Chem. Phys.*, 2010, **132**, 234511.

17 A. Garcia-Adeva, *New J. Phys.*, 2006, **8**, 86.

18 M. Maldovan and E. L. Thomas, *Nat. Mater.*, 2004, **3**, 593.

19 T. Ngo, C. Liddell, M. Ghebrebrhan and J. Joannopoulos, *Appl. Phys. Lett.*, 2006, **88**, 241920.

20 L. Gonzalez-Urbina, K. Baert, B. Kolaric, J. Perez-Moreno and K. Clays, *Chem. Rev.*, 2011, **112**, 2268–2285.

21 A. Moroz, *Phys. Rev. B*, 2002, **66**, 115109.

22 M. Maldovan, C. K. Ullal, W. C. Carter and E. L. Thomas, *Nat. Mater*, 2003, **2**, 664.

23 G. I. Waterhouse and M. R. Waterland, *Polyhedron*, 2007, **26**, 356 – 368.

24 A. Blanco, E. Chomski, S. Grabtchak, M. Ibisate, S. John, S. W. Leonard, C. Lopez, F. Meseguer, H. Miguez, J. P. Mondia, G. A. Ozin, O. Toader and H. M. van Driel, *Nature*, 2000, **405**, 437.

25 M. Lončar, T. Doll, J. Vučković and A. Scherer, *J. Lightwave Technol.*, 2000, **18**, 1402.

26 R. Rajan, P. R. Babu and K. Senthilnathan, *Theoretical Foundations and Application of Photonic Crystals*, IntechOpen, 2018, ch. 6, pp. 120–132.

27 B. Kolaric, S. Desprez, F. Brau and P. Damman, *J. Mater. Chem.*, 2012, **22**, 16205–16208.

28 Y. Wang, I. C. Jenkins, J. T. McGinley, T. Sinno and J. C. Crocker, *Nat. Commun.*, 2017, **8**, 14173.

29 K. J. Bishop, N. R. Chevalier and B. A. Grzybowski, *J. Phys. Chem. Lett.*, 2013, **4**, 1507–1511.

30 Q. Chen, S. C. Bae and S. Granick, *Nature*, 2011, **469**, 381.

31 L. Wang, L. Xia, G. Li, S. Ravaine and X. Zhao, *Angew. Chem.*, 2008, **47**, 4725–4728.

32 C. Bae, J. Moon, H. Shin, J. Kim and M. M. Sung, *J. Am. Chem. Soc.*, 2007, **129**, 14232–14239.

33 A. B. Pawar and I. Kretzschmar, *Macromol. Rapid Commun.*, 2010, **31**, 150–168.

34 G.-R. Yi, D. J. Pine and S. Sacanna, *J. Phys.: Condens. Matter*, 2013, **25**, 193101.

35 Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck and D. J. Pine, *Nature*, 2012, **491**, 51.

36 Q. Chen, S. C. Bae and S. Granick, *J. Am. Chem. Soc.*, 2012, **134**, 11080–11083.

37 F. Romano and F. Sciortino, *Nat. Commun.*, 2012, **3**, 975.

38 D. Z. Rocklin and X. Mao, *Soft Matter*, 2014, **10**, 7569–7576.

39 D. Chen, G. Zhang and S. Torquato, *J. Phys. Chem. B*, 2018, **122**, 8462–8468.

40 F. A. Escobedo, *J. Chem. Phys.*, 2017, **147**, 214501.

41 H. Chao and R. A. Riggleman, *Mol. Syst. Des. Eng.*, 2018, **3**, 214–222.

42 M. C. Rechtsman, F. H. Stillinger and S. Torquato, *Phys. Rev. Lett.*, 2005, **95**, 228301.

43 A. Jain, J. R. Errington and T. M. Truskett, *Phys. Rev. X*, 2014, **4**, 031049.

44 É. Marcotte, F. H. Stillinger and S. Torquato, *J. Chem. Phys.*, 2013, **138**, 061101.

45 A. Jain, J. R. Errington and T. M. Truskett, *Soft Matter*, 2013, **9**, 3866–3870.

46 A. Jain, J. A. Bollinger and T. M. Truskett, *AIChE J.*, 2014, **60**, 2732–2740.

47 B. A. Lindquist, R. B. Jadrich and T. M. Truskett, *J. Chem. Phys.*, 2016, **145**, 111101.

48 S. Torquato, *Soft Matter*, 2009, **5**, 1157–1173.

49 A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E*, 1995, **52**, 3730–3737.

50 M. Mungan, C.-H. Sow, S. N. Coppersmith and D. G. Grier, *Phys. Rev. B*, 1998, **58**, 14588–14593.

51 H. Cohn and A. Kumar, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 9570–9575.

52 A. W. Long and A. L. Ferguson, *Mol. Syst. Des. Eng.*, 2018, **3**, 49–65.

53 A. B. Pawar and I. Kretzschmar, *Langmuir*, 2008, **24**, 355–358.

54 A. B. Pawar and I. Kretzschmar, *Langmuir*, 2009, **25**, 9057–9063.

55 W. Liu, M. Tagawa, H. L. Xin, T. Wang, H. Emamy, H. Li,

K. G. Yager, F. W. Starr, A. V. Tkachenko and O. Gang, *Science*, 2016, **351**, 582–586.
56 Y. Wang, Y. Wang, X. Zheng, É. Ducrot, J. S. Yodh, M. Weck and D. J. Pine, *Nat. Commun.*, 2015, **6**, 7253.
57 W. B. Rogers and J. C. Crocker, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, **108**, 15687–15692.
58 G. Zhang, D. Wang and H. Möhwald, *Angewandte Chemie International Edition*, 2005, **44**, 7767–7770.
59 N. Kern and D. Frenkel, *J. Chem. Phys.*, 2003, **118**, 9882–9889.
60 Q. Chen, E. Diesel, J. K. Whitmer, S. C. Bae, E. Luijten and S. Granick, *J. Am. Chem. Soc.*, 2011, **133**, 7725–7727.
61 J. D. Weeks, D. Chandler and H. C. Andersen, *J. Chem. Phys.*, 1971, **54**, 5237–5247.
62 M. N. O'Brien, M. R. Jones and C. A. Mirkin, *Proc. Natl. Acad. Sci. U.S.A.*, 2016, **113**, 11717–11725.
63 I. S. Choi, N. Bowden and G. M. Whitesides, *Angew. Chem.*, 1999, **38**, 3078–3081.
64 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
65 J. Kästner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 932–942.
66 N. Hansen, S. D. Müller and P. Koumoutsakos, *Evol. Comput.*, 2003, **11**, 1–18.
67 J. A. Anderson, C. D. Lorenz and A. Travesset, *J. Comput. Phys.*, 2008, **227**, 5342–5359.
68 J. Glaser, T. D. Nguyen, J. A. Anderson, P. Lui, F. Spiga, J. A. Millan, D. C. Morse and S. C. Glotzer, *Comput. Phys. Commun.*, 2015, **192**, 97–107.
69 R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner and S. W. Zucker, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 7426–7431.
70 B. Nadler, S. Lafon, R. R. Coifman and I. G. Kevrekidis, *Appl. Comput. Harmon. Anal*, 2006, **21**, 113–127.
71 M. A. Rohrdanz, W. Zheng, M. Maggioni and C. Clementi, *J. Chem. Phys.*, 2011, **134**, 124116.
72 L. V. Nedialkova, M. A. Amat, I. G. Kevrekidis and G. Hummer, *J. Chem. Phys.*, 2014, **141**, 09B611_1.
73 A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis and P. G. Debenedetti, *Chem. Phys. Lett.*, 2011, **509**, 1–11.
74 W. Zheng, M. A. Rohrdanz and C. Clementi, *J. Phys. Chem. B*, 2013, **117**, 12769–12776.
75 A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti and I. G. Kevrekidis, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 13597–13602.
76 C. Fowlkes, S. Belongie, F. Chung and J. Malik, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, 214–225.
77 A. W. Long and A. L. Ferguson, *Appl. Comput. Harmon. Anal.*, 2019, **47**, 190–211.
78 R. Singh, J. Xu and B. Berger, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 12763–12768.
79 A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti and I. G. Kevrekidis, *J. Chem. Phys.*, 2011, **134**, 135103.
80 S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, *Phys. Lett. B*, 1987, **195**, 216–222.
81 R. M. Neal, *Handbook of Markov Chain Monte Carlo*, CRC Press, Boca Raton, 2011.
82 M. Gonzalez, E. Sanz, C. McBride, J. Abascal, C. Vega and C. Valeriani, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24913–24919.
83 B. Hashemian, D. Millán and M. Arroyo, *J. Chem. Phys.*, 2013, **139**, 214101.
84 W. Chen and A. L. Ferguson, *J. Comput. Chem.*, 2018, **39**, 2079–2102.
85 W. Chen, A. R. Tan and A. L. Ferguson, *J. Chem. Phys.*, 2018, **149**, 072312.
86 W. Chen, H. Sidky and A. L. Ferguson, *J. Chem. Phys.*, 2019, **150**, 214114.
87 E. Bitzek, P. Koskinen, F. Gähler, M. Moseler and P. Gumbsch, *Phys. Rev. Lett.*, 2006, **97**, 170201.
88 A. L. Ferguson, *J. Comput. Chem.*, 2017, **38**, 1583–1605.
89 N. Nakagawa and M. Peyrard, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 5279–5284.
90 N. Hansen and A. Ostermeier, EUFIT '97, 5th European Congress on Intelligent Techniques and Soft Computing, 1997, pp. 650–654.
91 A. W. Wilber, J. P. K. Doye, A. A. Louis and A. C. F. Lewis, *J. Chem. Phys.*, 2009, **131**, 175102.
92 P. J. Steinhardt, D. R. Nelson and M. Ronchetti, *Phys. Rev. B*, 1983, **28**, 784.
93 W. Lechner and C. Dellago, *J. Chem. Phys.*, 2008, **129**, 114707.
94 P. Rein ten Wolde, M. J. Ruiz-Montero and D. Frenkel, *J. Chem. Phys.*, 1996, **104**, 9932–9947.
95 S. G. Johnson and J. D. Joannopoulos, *Opt. Express*, 2001, **8**, 173–190.
96 P. Bartlett and A. I. Campbell, *Phys. Rev. Lett.*, 2005, **95**, 128302.
97 M. E. Leunissen, C. G. Christova, A.-P. Hynninen, C. P. Royall, A. I. Campbell, A. Imhof, M. Dijkstra, R. Van Roij and A. Van Blaaderen, *Nature*, 2005, **437**, 235.
98 A. M. Kalsin and B. A. Grzybowski, *Nano Lett.*, 2007, **7**, 1018–1021.
99 C. Long, Q.-L. Lei, C.-L. Ren and Y.-Q. Ma, *J. Phys. Chem. B*, 2018, **122**, 3196–3201.
100 S. Glotzer, M. Solomon and N. A. Kotov, *AIChE J.*, 2004, **50**, 2978–2985.
101 J. N. Israelachvili, *Intermolecular and Surface Forces (Third Edition)*, Academic Press, San Diego, 2011.
102 Q.-L. Lei, R. Ni and Y.-Q. Ma, *ACS Nano*, 2018, **12**, 6860–6870.