

Inverse design of charged colloidal particle interactions for self assembly into specified crystal structures

Cite as: J. Chem. Phys. **151**, 084109 (2019); <https://doi.org/10.1063/1.5111492>

Submitted: 27 May 2019 . Accepted: 06 August 2019 . Published Online: 29 August 2019

Rajneesh Kumar , Gabriele M. Coli, Marjolein Dijkstra , and Srikanth Sastry 



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Compressing physics with an autoencoder: Creating an atomic species representation to improve machine learning models in the chemical sciences](#)

The Journal of Chemical Physics **151**, 084103 (2019); <https://doi.org/10.1063/1.5108803>

[The role of pressure in inverse design for assembly](#)

The Journal of Chemical Physics **151**, 104104 (2019); <https://doi.org/10.1063/1.5112766>

[Multipole-based distance-dependent screening of Coulomb integrals](#)

The Journal of Chemical Physics **151**, 084111 (2019); <https://doi.org/10.1063/1.5111054>





Lock-in Amplifiers

X Zurich Instruments

MFLI

Watch the Video 

Inverse design of charged colloidal particle interactions for self assembly into specified crystal structures

Cite as: J. Chem. Phys. 151, 084109 (2019); doi: 10.1063/1.5111492

Submitted: 27 May 2019 • Accepted: 6 August 2019 •

Published Online: 29 August 2019



Rajneesh Kumar,^{1,a)} Gabriele M. Coli,^{2,b)} Marjolein Dijkstra,^{2,c)} and Srikanth Sastry^{1,d)}

AFFILIATIONS

¹Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru 560064, India

²Soft Condensed Matter, Debye Institute for Nanomaterials Science, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

^{a)}Electronic mail: rajneesh@jncasr.ac.in

^{b)}Electronic mail: g.m.coli@uu.nl

^{c)}Electronic mail: m.dijkstra@uu.nl

^{d)}Electronic mail: sastry@jncasr.ac.in

ABSTRACT

We study the inverse problem of tuning interaction parameters between charged colloidal particles interacting with a hard-core repulsive Yukawa potential, so that they assemble into specified crystal structures. Here, we target the body-centered-cubic (bcc) structure which is only stable in a small region in the phase diagram of charged colloids and is, therefore, challenging to find. In order to achieve this goal, we use the statistical fluctuations in the bond orientational order parameters to tune the interaction parameters for the bcc structure, while initializing the system in the fluid phase, using the Statistical Physics-inspired Inverse Design algorithm. We also find that this optimization algorithm correctly senses the fluid-solid phase boundaries for charged colloids. Finally, we repeat the procedure employing the covariance matrix adaptation-evolution strategy, a cutting edge optimization technique, and compare the relative efficacy of the two methods.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5111492>

I. INTRODUCTION

In the past few years, several inverse methods have emerged that design optimal interactions in such a way that the system spontaneously assembles into a targeted structure.^{1–4} These methods have received considerable attention in materials science^{5–7} and have been successively used to find crystal structures for photonic band-gap applications,⁸ to predict crystals⁹ and protein structures,^{10,11} materials with optimal mechanical and transport properties,¹² and for optimizing the interactions for self assembly.^{13–17} Although many of the methods developed are either based on black-box techniques, in which the algorithm tunes the interaction parameters without taking the statistical nature of the system into account, or are designed *ad hoc* for a particular class of systems, systematic approaches based on a statistical

mechanical formulation, which are general, and allow for application tailored to specific systems of interest, have also been investigated. Recently, several methods have been proposed which take into account the statistical nature of the system in updating the interparticle interaction parameters.^{18–20} Lindquist *et al.*¹⁸ proposed a relative entropy based inverse design approach which tunes the interparticle interaction parameters for the self assembly of crystal structures by exploiting the statistical nature of the system. Later, Adorf *et al.*²⁰ modified this method by carrying out relative entropy minimization directly in Fourier space to target complex crystal structures. These methods have been successfully employed to design isotropic potentials to assemble a wide range of crystal structures.

In the present work, we investigate the efficacy of one such method, the Statistical Physics-inspired Inverse Design

(SP-ID) method developed by Miskin *et al.*²¹ This method considers statistical fluctuations present in the microscopic configurations of the system for tuning the interactions between the particles. Apart from tuning the interparticle interaction parameters, system parameters such as temperature and pressure can also be tuned using this method. In order to design these interactions, we have used a quality function based on bond order parameters (BOP)^{22,23} to rank the generated configurations. The same task can be faced employing numerous optimization techniques like relative entropy based inverse design,^{18–20} (adaptive) simulated annealing,^{24–26} particle swarm optimization,²⁷ and several genetic algorithms.²⁸ To evaluate the effectiveness of the SP-ID algorithm, we compare it with the Covariance Matrix Adaption-Evolutionary Strategy (CMA-ES),^{29–31} which we regard as a state-of-the-art optimization technique for evolutionary computation.

We have chosen a system of colloidal particles as the model for which we wish to design the interactions in order to target a specific crystal structure. The interparticle potential of colloids offers a wide variety of functional forms. It can contain a hard-core term, a dipole-dipole term, a charge dispersion term, a screened-Coulomb (Yukawa) term, and a short-ranged attractive depletion term. More specific designed colloidal particles such as patchy colloids and DNA-functionalized colloids offer even greater diversity of interactions. For all of these, the interaction parameters can be tuned. In the case of Yukawa interactions, the Debye screening length can be adjusted by changing the salt concentration, and the contact value can be tuned by altering the surface charge of the particles. Previous studies have employed what may be termed a *forward method* in which, starting from the microscopic interaction parameters and specified thermodynamic parameters such as temperature and density, one computes the equilibrium properties and finds the stable phase of the system.^{32–34} In the present work, our purpose is to employ an *inverse method*, where the target structure and equilibrium properties are the input from which we wish to design the interparticle interactions for which the particles spontaneously self-assemble into the target structure.

In this study, we considered a charged colloidal system in which particles interact with a hard-core repulsive Yukawa potential. The complete phase diagram of hard-core Yukawa particles is known from earlier studies.^{32–34} Because of the purely repulsive nature of the potential, this system displays only a fluid phase which can freeze into a face-centered-cubic (fcc) or a body-centered-cubic (bcc) crystal phases. The phase diagram of hard-core Yukawa particles shows two triple points where fcc, bcc, and fluid phases coexist. The bcc phase is only stable in a very small region in the phase diagram and, for this reason, constitutes a reasonable test case for the reverse-engineering process. In this work, we search for optimal interparticle interactions which favor the crystallization of the bcc structure. Here, we show three different cases, in which we, respectively, tune one, two, and three parameters. In all cases, we show that both SP-ID and CMA-ES adjust the interparticle interactions that lead to the targeted bcc structure formation.

This paper is organized as follows: In Sec. II, we define the model system studied in this paper and the corresponding phase diagram and bond order parameters to identify the

different phases. In Sec. III, we define the inverse design methods and the form of the quality function used in this study. Results for tuning the interactions to target the bcc structure for all three cases are discussed in Sec. IV. Finally, we summarize our results in Sec. V.

II. MODEL AND SIMULATION METHODS

A. Interaction potential

We consider a hard-core repulsive Yukawa system which represents a standard model for charged colloids. The form of the potential is given by

$$\beta U(r) = \begin{cases} \beta \epsilon \frac{\exp[-\kappa\sigma(r/\sigma-1)]}{r/\sigma} & \text{for } r > \sigma \\ \infty & \text{for } r \leq \sigma \end{cases}, \quad (1)$$

where $\beta\epsilon$ is the contact value of the pair potential expressed in units of $k_B T = 1/\beta$, k_B is the Boltzmann constant, κ is the inverse of the Debye screening length, and σ is the hard-core diameter. In Fig. 1, we show the phase diagram for such a system with $\beta\epsilon = 8$ in the $(1/\kappa\sigma, \eta)$ representation and the $(1/\kappa\sigma, \beta P\sigma^3)$ plane.³⁴ The phase diagram exhibits a stable fluid, bcc, and fcc region, as well as two triple points at which the three phases coexist. Note that we always use scaled variables, a reduced temperature $k_B T/\epsilon$, reduced pressure $\beta P\sigma^3$, and inverse screening length $1/\kappa\sigma$, each of which can be treated as tuning parameter for obtaining the desired behavior.

B. Bond order parameters (BOP)

Every optimization algorithm, including SP-ID and CMA-ES, works on the basis of minimizing a user-defined fitness function. Here, we have used the averaged bond order parameters \bar{q}_l and \bar{w}_l ($l = 6$). The first is computed as follows:^{22,23}

$$\bar{q}_l^{(i)} = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{q}_{lm}^{(i)}|^2 \right]^{1/2}, \quad (2)$$

where

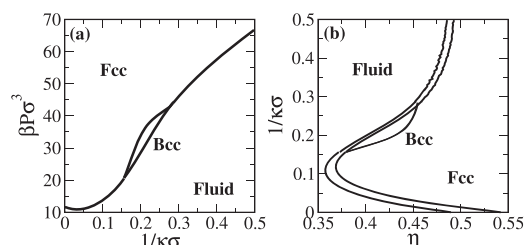


FIG. 1. Phase diagram of a system in which the particles interact via a hard-core repulsive Yukawa pair potential with $\beta\epsilon = 8$ in (a) the $(1/\kappa\sigma, \beta P\sigma^3)$ plane and (b) the $(1/\kappa\sigma, \eta)$ representation. The phase diagram displays a stable fluid, bcc, and fcc phase.

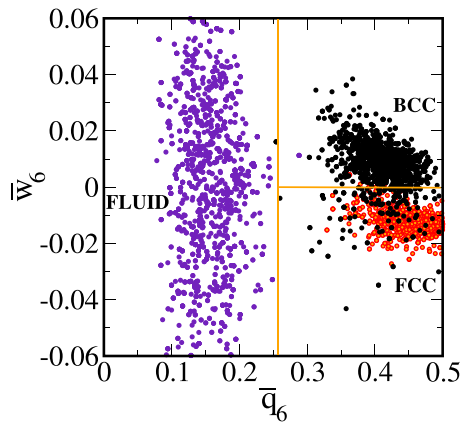


FIG. 2. Scatter plot of the averaged bond order parameters \bar{q}_6 vs \bar{w}_6 for the fluid, bcc, and fcc phase of a system of Yukawa particles with contact value $\beta\epsilon = 8$ at the high-density triple point. Each point corresponds to a single particle. In total, 2000 points were chosen randomly from each structure.

$$\bar{q}_{lm}^{(i)} = \frac{1}{\bar{N}_b(i)} \sum_{j=0}^{\bar{N}_b(i)} q_{lm}^{(j)},$$

$$q_{lm}^{(i)} = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\theta(\mathbf{r}_{ij}), \phi(\mathbf{r}_{ij})).$$

Here, $N_b(i)$ is the number of neighbors of particle i , $\bar{N}_b(i)$ is the number of neighbors including particle i itself, $Y_{lm}(\theta(\mathbf{r}_{ij}), \phi(\mathbf{r}_{ij}))$ denotes the spherical harmonics with \mathbf{r}_{ij} the distance vector from particle i to particle j . The second bond order parameter we used is defined as

$$\bar{w}_l^{(i)} = \frac{\sum_{m_1+m_2+m_3=l} \begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} \bar{q}_{lm_1}^{(i)} \bar{q}_{lm_2}^{(i)} \bar{q}_{lm_3}^{(i)}}{\left(\sum_{m=-l}^l |\bar{q}_{lm}^{(i)}|^2 \right)^{3/2}}.$$

In order to calculate the radius of the first coordination shell for each particle, we employ the solid angle based nearest-neighbor (SANN) algorithm,³⁵ where a nearest neighbor of a particle is identified by attributing a solid angle to each possible neighbor such that the sum of solid angles equals at least 4π .

In Fig. 2, we show the scatter plots of \bar{q}_6 vs \bar{w}_6 for the fluid, bcc, and fcc phases of a system of Yukawa particles with $\beta\epsilon = 8$ at the high-density triple point conditions. We find distinct clouds of points corresponding to the fluid, bcc, and fcc phases, and hence, \bar{q}_6 and \bar{w}_6 can be used to distinguish the three phases.

III. INVERSE DESIGN METHODS

A. Statistical physics-inspired inverse design method

In the SP-ID method, microscopic parameters such as the interparticle pair potential parameters are tuned by exploiting statistical fluctuations in such a way that the system will evolve to those states which correspond to the targeted macroscopic response

of that system.²¹ In this method, the time evolution of the probability distribution of finding the system in configuration x is written as

$$\dot{\rho}(x|\lambda_i) = \rho(x|\lambda_i)[f(x) - \langle f(x) \rangle], \quad (3)$$

where ρ denotes the probability of finding a system in some configuration, λ_i s are the adjustable parameters, and $f(x)$ is a quality function which gives a weight/fitness value to each configuration based on a targeted macroscopic property. Time t is an artificial time that indexes the optimization steps, and $\langle f(x) \rangle$ is the ensemble average over all the configurations for a given parameter set. With straightforward manipulation, the above equation can be recast as equations of motion for λ_i ,²¹

$$\dot{\lambda}_i(t) = \langle \partial_{\lambda_i} \log(\rho) \partial_{\lambda_i} \log(\rho) \rangle^{-1} [f(x) - \langle f(x) \rangle] \partial_{\lambda_i} \log(\rho), \quad (4)$$

where $\langle \dots \rangle$ denotes an ensemble average at a given set of values of λ_i . To integrate Eq. (4), we have used a modified Euler method with a fixed time step of 4.0.

We have built our quality function $f(x)$ in the following way:

$$f(x) = \int dx' \Theta(g(x') \geq g(x)) \rho(x'|\lambda), \quad (5)$$

where $\Theta(a \geq b)$ is the Heaviside function, equal to 1 whenever the inequality in the argument is fulfilled and zero otherwise, and $g(x)$ denotes the fitness function,

$$g(x) = (\bar{q}_6(x) - \bar{q}_6^{\text{target}})^2 + (\bar{w}_6(x) - \bar{w}_6^{\text{target}})^2, \quad (6)$$

with $\bar{q}_6(x) = \sum_i \bar{q}_6^{(i)}/N$ and $\bar{w}_6(x) = \sum_i \bar{w}_6^{(i)}/N$ the averages of the bond order parameters over all the particles in the system, and $\bar{q}_6^{\text{target}}$ and $\bar{w}_6^{\text{target}}$ are the corresponding quantities in the target structure (bcc). Here, the integral over x' represents a sum over a series of n different configurations as obtained from a simulation for a fixed set of parameters ($\kappa\sigma$, $\beta P\sigma^3$, $\beta\epsilon$). The quality function $f(x)$ will have higher values for those configurations whose \bar{q}_6 and \bar{w}_6 values are closer to the target values. More precisely, $f(x)$ equals the probability of having a lower value of $g(x)$ than any other configuration drawn randomly from the equilibrium distribution. To target the bcc structure, we have chosen $\bar{q}_6^{\text{target}} = 0.395$ and $\bar{w}_6^{\text{target}} = 0.011830$ (average values of \bar{q}_6 , \bar{w}_6 for the bcc structure obtained from the scatter plot shown in Fig. 2). For a perfect bcc structure, these values are $\bar{q}_6 = 0.5107$ and $\bar{w}_6 = 0.013161$, but here we have targeted the bond order parameter values for a finite-temperature bcc structure. The choice of the fitness function depends on what the user is interested in, and might include other quantities, like the lattice spacing, radial distribution function, packing fraction, etc., and may further be fine tuned by weighing the different terms [e.g., the terms depending on \bar{q}_6 and \bar{w}_6 in the definition of $g(x)$].

B. Covariance matrix adaptation-evolutionary strategy

In order to implement CMA-ES, we draw n samples from a multivariate Gaussian distribution for each generation whose dimension D corresponds to the number of parameters we wish to tune. Subsequently, we evaluate the fitness function $g(x)$ on the generated samples, and we pick the best k samples. Using the following

equations, we estimate the multivariate Gaussian distribution with mean $\bar{\mu}$ (a D -dimensional vector) and $\Sigma = \sigma^2 \mathbf{C}$ the covariance matrix of the Gaussian distribution for the next generation using

$$\begin{aligned}\mu_i' &= \mu_i + \sum_x w(x)(\lambda_i(x) - \mu_i), \\ q_i' &= (1 - c_1)q_i + c_2\sqrt{\Sigma_{ii}^{-1}}(\mu_j' - \mu_j), \\ p_i' &= (1 - c_3)p_i + c_4(\mu_i' - \mu_i), \\ C_{ij}' &= (1 - c_5 - c_6)C_{ij} \\ &\quad + c_5 \sum_x w(x) \left(\frac{\lambda_i(x) - \mu_i}{\sigma} \frac{\lambda_j(x) - \mu_j}{\sigma} - C_{ij} \right) + c_6 p_i' p_j', \\ \sigma' &= \sigma \exp \left[c_7 \left(\frac{\|q_i'\|}{\langle \|N(0, I)\| \rangle} - 1 \right) \right],\end{aligned}\quad (7)$$

where $\{x\}$ denotes the n samples consisting of multiple configurations calculated for n different parameter sets ($\kappa\sigma$, $\beta P\sigma^3$, $\beta\epsilon$) [denoted by $\lambda_i(x)$ above] in CMA-ES, $w(x)$ is the normalized distribution of weights based on the fitness of the samples, and c_i 's are free parameters. We choose $w(x) \propto \log(k+1) - \log(i)$, where i is the rank index of sample x [$i=1$ for the configuration with the smallest $g(x)$ value] for the best k samples, and set $w(x)=0$ for the rest. \bar{q} and \bar{p} are additional D -dimensional vectors which determine, respectively, the changes in amplitude and directionality of the covariance matrix, and finally, $\langle \|N(0, I)\| \rangle$ is the average length of a vector drawn from a multivariate Gaussian distribution centered in the origin and where the covariance matrix is the identity matrix. In the present work, we use $n=10$ and $k=5$. For the first generation, we initialize \bar{q} and \bar{p} as null vectors. Moreover, since we do not assume any *a priori* correlation between the different tuning parameters, the initial form of the covariance matrix Σ is diagonal. Finally, all the free parameters of the CMA-ES are selected following the recipe in Ref. 31.

Note that there is a substantial difference in the use of the fitness function between the two methods. In SP-ID, we rank different configurations, so $\bar{q}_6(x)$ and $\bar{w}_6(x)$ are bond order parameters computed in a single configuration. The SP-ID method is thus based on the statistical fluctuations in the bond order parameters (Fig. 2) in a simulation at a single set of interaction parameters in order to optimize these values for the desired bcc structure. When using the CMA-ES, we rank samples, so $\bar{q}_6(x)$ and $\bar{w}_6(x)$ are computed as ensemble averages of these bond order parameters over multiple configurations for distinct sets of parameters. The CMA-ES method is thus based on a ranking of the different samples as obtained for different interaction parameter sets in order to optimize the parameter values.

C. Simulation details

In order to evaluate the ensemble averages and to generate distinct configurations for the SP-ID method, we perform constant pressure and constant temperature (NPT) Monte Carlo (MC) simulations on systems consisting of $N=250$ hard-core repulsive Yukawa particles. We initialize the simulations by placing the particles randomly in a cubic simulation box. We equilibrate the system up to 10^5 MC cycles. One Monte Carlo cycle corresponds to N particle moves and one volume move. The particle and volume moves are adjusted in such a way that 45% of the particle moves and 20% of

the volume moves are accepted. We save 10^3 uncorrelated samples and evaluate the ensemble averages in Eq. (4). These uncorrelated samples are used to calculate the new parameters. Once the parameters have been changed, we repeat the whole procedure starting the simulation from the last configuration generated with the previous parameters. We repeat this process until the target structure is reached.

In the CMA-ES algorithm, we evaluate the ensemble averages by performing simulations at different parameter sets at each generation. At every next generation, we take the last configuration of the fittest sample as the starting point for all the new samples.

IV. RESULTS

A. Tuning parameters with the SP-ID method

We start from the case in which we tune only one interparticle interaction parameter, the inverse Debye screening length $1/\kappa\sigma$ of the interaction potential [Eq. (1)] to target the bcc structure. At this stage, reduced pressure $\beta P\sigma^3$ and the contact value $\beta\epsilon$ are kept constant at $\beta P\sigma^3 = 33$ and $\beta\epsilon = 8$. The initial value of $1/\kappa\sigma$ is 0.4, to make sure the system starts from a fluid configuration. Given the form discussed in Sec. III, the quality function gives higher weights to those configurations whose \bar{q}_6 and \bar{w}_6 values are closer to the target values. The probability of finding the system in some configuration x , in the NPT ensemble is given by, $\rho(x|\lambda_i) \propto \exp(-\beta H(x))$. When only one parameter is tuned (1D case), the equation of motion [Eq. (4)] becomes

$$\frac{d}{dt}(\kappa\sigma) = -\text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, \frac{\partial(\beta H)}{\partial(\kappa\sigma)} \right]^{-1} \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, f \right], \quad (8)$$

where $H = U + PV$ is the Hamiltonian of the system.

By solving Eq. (8), a new value of $1/\kappa\sigma$ is obtained and the algorithm keeps on optimizing this interaction parameter until the goal is reached, i.e., $\bar{q}_6 = \bar{q}_6^{\text{target}}$ and $\bar{w}_6 = \bar{w}_6^{\text{target}}$. The path of the parameters is shown in Figs. 3(a) and 3(b) in the $(1/\kappa\sigma, \beta P\sigma^3)$ and $(1/\kappa\sigma, \eta)$ planes. In both, the optimizer correctly tunes $1/\kappa\sigma$ to reach the bcc structure which can also be verified by examining the evolution of the average \bar{q}_6 and \bar{w}_6 values as the simulation proceeds. In Figs. 4(a) and 4(b), we plot the average \bar{q}_6 and \bar{w}_6 values as a function of the simulation time. At the very beginning of the simulation, \bar{q}_6 and \bar{w}_6 values show that the system is in the fluid phase, and as the algorithm optimizes the interactions, there is a sharp transition in both of these values, which exactly happens at the fluid-bcc phase boundary. Once the system reaches the bcc phase, it remains in the bcc phase. We also find that as the system reaches the phase boundaries, there is a sudden change in the slope of the parameter's trajectory as shown in Fig. 3(c). In other words, the optimizer [Eq. (4)] correctly recognizes the phase boundaries present in the phase diagram.

We now analyze the case in which we tune two parameters simultaneously (2D case), one interparticle interaction parameter, the inverse Debye screening length $1/\kappa\sigma$, and one system parameter, reduced pressure $\beta P\sigma^3$, while $\beta\epsilon = 8$ is kept constant. Note that we use σ and $k_B T$ as our unit of length and energy, respectively, which we keep fixed, and hence, $\beta\epsilon$ and $\beta P\sigma^3$ can be varied independently from each other. Here, we initialize the system again in the fluid

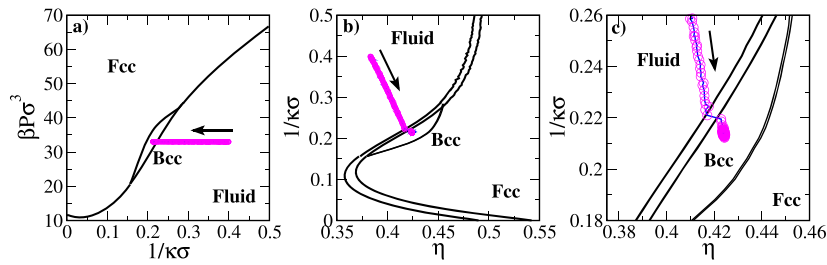


FIG. 3. Evolution of the parameters in (a) the $(1/\kappa\sigma, \beta P\sigma^3)$ plane and (b) the $(1/\kappa\sigma, \eta)$ plane when the system is initialized in the fluid phase at reduced pressure $\beta P\sigma^3 = 33$, inverse Debye screening length $1/\kappa\sigma = 0.4$, and contact value $\beta\epsilon = 8$. $\beta P\sigma^3$ and $\beta\epsilon$ are kept fixed, and only $1/\kappa\sigma$ is tuned. (c) shows an enlarged view of the data near the fluid-bcc coexistence region.

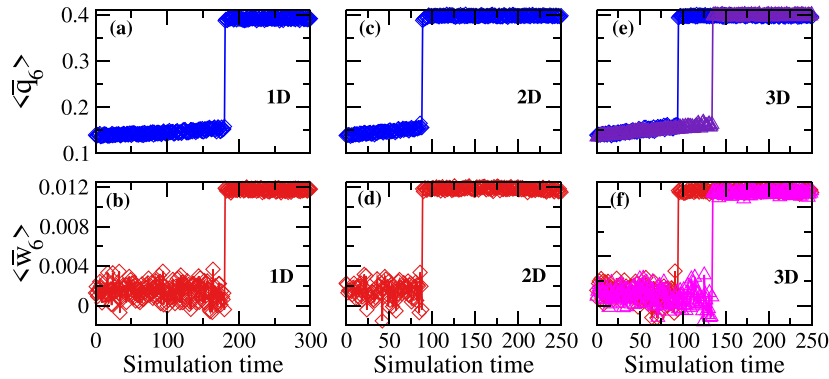


FIG. 4. Evolution of $\langle \bar{q}_6 \rangle$ and $\langle \bar{w}_6 \rangle$ as a function of simulation time during the optimization using the SP-ID algorithm for (a) and (b) the 1D, (c) and (d) the 2D, and (e) and (f) for the 3D case, respectively. The system is initialized in the fluid phase at reduced pressure $\beta P\sigma^3 = 33$, inverse screening length $1/\kappa\sigma = 0.4$, and contact value $\beta\epsilon = 8$. For the 3D case, blue and red colored symbols represent the $\langle \bar{q}_6 \rangle$ and $\langle \bar{w}_6 \rangle$ values when the system is initialized at (i) $1/\kappa\sigma = 0.4$, $\beta P\sigma^3 = 33$, and $\beta\epsilon = 8$, while indigo and magenta colored symbols represent bond order parameter values for the case (ii) $1/\kappa\sigma = 0.4$, $\beta P\sigma^3 = 25$, and $\beta\epsilon = 6$. The simulation time indicates the number of simulations performed at distinct sets of interaction parameters.

phase at $1/\kappa\sigma = 0.4$ and $\beta P\sigma^3 = 33$. The equations of motion [Eq. (4)] for the two parameters become

$$\frac{d}{dt} \begin{bmatrix} \beta P\sigma^3 \\ \kappa\sigma \end{bmatrix} = - \begin{bmatrix} \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, \frac{\partial(\beta H)}{\partial(\beta P\sigma^3)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, \frac{\partial(\beta H)}{\partial(\kappa\sigma)} \right] \\ \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, \frac{\partial(\beta H)}{\partial(\kappa\sigma)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, \frac{\partial(\beta H)}{\partial(\kappa\sigma)} \right] \end{bmatrix}^{-1} \times \begin{bmatrix} \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, f \right] \\ \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, f \right] \end{bmatrix}. \quad (9)$$

In Figs. 5(a) and 5(b), we show the path of the parameters in the $(1/\kappa\sigma, \beta P\sigma^3)$ and $(1/\kappa\sigma, \eta)$ planes. As the simulation time proceeds, SP-ID successfully optimizes both the interaction parameters in such a way that the final structure formed is the bcc crystal. The form of

the quality function is the same as we have used for the one parameter case. Variations of \bar{q}_6 and \bar{w}_6 values also verify the formation of the bcc structure from the fluid phase [Figs. 4(c) and 4(d)]. We find from Fig. 4(c) that the optimizer recognizes the phase boundaries very well as also found for the one parameter case.

Finally, we investigate the case in which we tune three parameters simultaneously (3D case), two interparticle interaction parameters, the inverse Debye screening length $1/\kappa\sigma$ and the contact value $\beta\epsilon$, and one system parameter, reduced pressure $\beta P\sigma^3$. We perform two independent simulations by initializing the system at two different state points in the fluid phase at (i) $1/\kappa\sigma = 0.4$, $\beta P\sigma^3 = 33$, and $\beta\epsilon = 8$; (ii) $1/\kappa\sigma = 0.4$, $\beta P\sigma^3 = 25$, and $\beta\epsilon = 6$ and optimize all three parameters, $\kappa\sigma$, $\beta P\sigma^3$, and $\beta\epsilon$ for the bcc phase. The equations of motion [Eq. (4)] when three parameters are tuned become

$$\frac{d}{dt} \begin{bmatrix} \beta P\sigma^3 \\ \kappa\sigma \\ \beta\epsilon \end{bmatrix} = - \begin{bmatrix} \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, \frac{\partial(\beta H)}{\partial(\beta P\sigma^3)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, \frac{\partial(\beta H)}{\partial(\kappa\sigma)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, \frac{\partial(\beta H)}{\partial(\beta\epsilon)} \right] \\ \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, \frac{\partial(\beta H)}{\partial(\beta P\sigma^3)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, \frac{\partial(\beta H)}{\partial(\kappa\sigma)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, \frac{\partial(\beta H)}{\partial(\beta\epsilon)} \right] \\ \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta\epsilon)}, \frac{\partial(\beta H)}{\partial(\beta P\sigma^3)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta\epsilon)}, \frac{\partial(\beta H)}{\partial(\kappa\sigma)} \right] & \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta\epsilon)}, \frac{\partial(\beta H)}{\partial(\beta\epsilon)} \right] \end{bmatrix}^{-1} \begin{bmatrix} \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta P\sigma^3)}, f \right] \\ \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\kappa\sigma)}, f \right] \\ \text{Cov} \left[\frac{\partial(\beta H)}{\partial(\beta\epsilon)}, f \right] \end{bmatrix}. \quad (10)$$

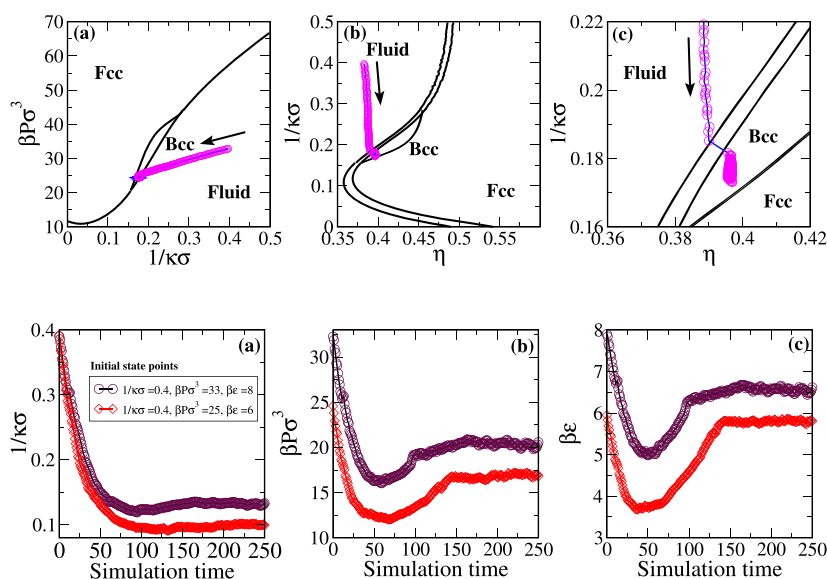


FIG. 5. Evolution of the parameters in (a) the $(1/\kappa\sigma, \beta P\sigma^3)$ plane and (b) the $(1/\kappa\sigma, \eta)$ plane when the system is initialized in the fluid phase at reduced pressure $\beta P\sigma^3 = 33$, inverse Debye screening length $1/\kappa\sigma = 0.4$, and contact value $\beta\epsilon = 8$. $\beta\epsilon$ is kept fixed, and two parameters $1/\kappa\sigma$ and $\beta P\sigma^3$ are tuned. (c) shows an enlarged view of the data near the fluid-bcc coexistence region.

FIG. 6. Evolution of [(a)] $1/\kappa\sigma$, [(b)] $\beta P\sigma^3$, and [(c)] $\beta\epsilon$ as a function of simulation time when the system is initialized in the fluid phase at (i) $\beta P\sigma^3 = 33$, $1/\kappa\sigma = 0.4$ and $\beta\epsilon = 8$ (indigo colored circles), and (ii) $\beta P\sigma^3 = 25$, $1/\kappa\sigma = 0.4$, and $\beta\epsilon = 6$ (red colored diamonds). Three parameters are tuned to target the bcc structure, namely $\beta P\sigma^3$, $1/\kappa\sigma$, and $\beta\epsilon$.

In Fig. 6, we plot the path of the tuned parameter trajectories as a function of simulation time when the system is initialized in the fluid phase and the desired goal is to reach the targeted bcc structure. Initially, all three parameter values decrease while the system is in the fluid phase and once it crosses the phase boundary between the fluid and bcc phase, they start to saturate. As the simulation time proceeds, the optimizer successfully optimizes the parameters in such a way that the final structure becomes the bcc phase. Here we also use the same form of the quality function as we have used earlier for the one and two parameter cases.

To confirm that the final structure is bcc, we also plot the evolution of $\langle q_6 \rangle$ and $\langle \tilde{w}_6 \rangle$ as a function of the simulation time in Figs. 4(e) and 4(f), which indeed confirms that the simulation reaches the optimal bond order parameter values for the bcc phase. Finally, one may enquire what will happen if the targeted structure does not exist in the phase diagram. We find in test runs that at higher temperature $\beta\epsilon = 2$, SP-ID fails to target the bcc structure. This can be seen from the variation of $\langle \tilde{w}_6 \rangle$ and average fitness function $\langle g(x) \rangle$ as a function of the simulation time. To better distinguish different structures quantitatively, one may have to increase the contribution of $\langle \tilde{w}_6 \rangle$ in the fitness function by a constant factor such that both $\langle q_6 \rangle$ and $\langle \tilde{w}_6 \rangle$ contribute equally to the fitness function.

B. Tuning parameters with the CMA-ES method

We now employ the CMA-ES algorithm to analyze all the cases already studied using the SP-ID method, i.e., the tuning of one, two and three parameters, highlighting the differences between the two inverse design optimizers. We use the parameters corresponding to the initial state point employed in the SP-ID algorithm as the initial mean vectors of the multivariate Gaussian distribution in the CMA-ES algorithm. In addition, we start the CMA-ES algorithm with a diagonal covariance matrix with a standard deviation of 10% around its mean value for each parameter. Finally, the initial values of each component of the vectors \vec{q} and \vec{p} are set to zero.

In Fig. 7, we show the results for the one parameter case. The points displayed in Fig. 7(a) represent the mean value of the Gaussian distribution in each generation. At the beginning, the algorithm tries to decrease the fitness function value by decreasing the inverse Debye screening length $1/\kappa\sigma$. Since all the steps point in the same direction, the variance of the Gaussian distribution increases [Fig. 7(b)] and the mean value overtakes the bcc region, ending up in the fcc phase. After this, the algorithm recognizes the right direction, and the subsequent update is in the opposite direction, i.e., the inverse Debye screening length increases. At the same time, the covariance starts to shrink and the mean value of the Gaussian distribution is found in the bcc region for the first time at the 6th generation. From this generation onwards, the updates are in random directions inside the bcc region, leading to a further exponential decrease of the covariance of the Gaussian distribution. At the 12th generation, all the 10 simulations have a $k\sigma$ value for which the structure corresponds to the stable bcc phase.

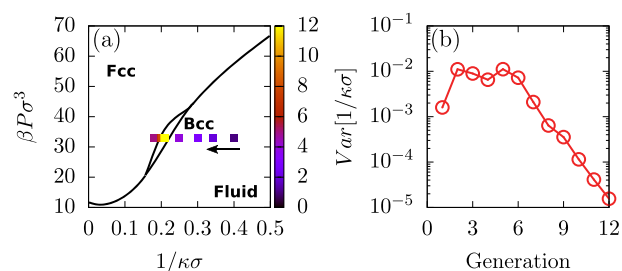


FIG. 7. (a) Evolution of the mean value of the Gaussian distribution for $1/\kappa\sigma$ in the $(1/\kappa\sigma, \beta P\sigma^3)$ plane when the system is initialized in the fluid phase at reduced pressure $\beta P\sigma^3 = 33$, inverse Debye screening length $1/\kappa\sigma = 0.4$, and contact value $\beta\epsilon = 8$. The parameters $\beta\epsilon$ and $\beta P\sigma^3 = 33$ are kept fixed, and $1/\kappa\sigma$ is tuned using the CMA-ES method. (b) The variance of the Gaussian distribution for $1/\kappa\sigma$ at each generation.

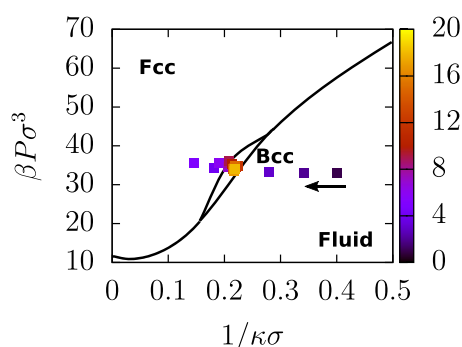


FIG. 8. Evolution of the mean value of the multivariate Gaussian distribution for $1/\kappa\sigma$ and $\beta P\sigma^3$ in the $(1/\kappa\sigma, \beta P\sigma^3)$ plane when the system is initialized in the fluid phase at reduced pressure $\beta P\sigma^3 = 33$, inverse Debye screening length $1/\kappa\sigma = 0.4$, and contact value $\beta\epsilon = 8$. The parameter $\beta\epsilon$ is kept fixed, and $1/\kappa\sigma$ and $\beta P\sigma^3$ are tuned using the CMA-ES method.

The cases in which we tune two or three parameters do not present significantly different behavior of the CMA-ES algorithm with respect to the one parameter case. In the two parameter case, $\bar{\mu}$ enters the bcc region for the first time at the 10th generation, while at the 18th generation all the samples have entered the bcc region, showing how CMA-ES is adversely affected by the dimensionality of the parameter space. The results for the 2D case are shown in Fig. 8. We note that the risk of overshooting is lower when more parameters are varied at the same time, since they all contribute to the increase or decrease of the quality function, and the updates of the mean values of the multivariate Gaussian distribution may therefore not always be in the same direction. This prevents the covariance matrix from growing too fast as already shown. Finally, the results for both the investigated 3D cases are shown in Fig. 9. To confirm that the final structure is bcc, we plot the evolution of $\langle \bar{q}_6 \rangle$ and $\langle \bar{w}_6 \rangle$ as a function of the simulation time in Fig. 10.

These analyses provide benchmarks for how fast the CMS-ES algorithm converges in finding the target structure, as compared to the SP-ID algorithm. Since each generation of the CMA-ES requires n times (10 in this work), the computational effort of SP-ID, a comparison of run lengths in the two methods reveals that their efficiencies are very similar. However, in SP-ID, one has information on

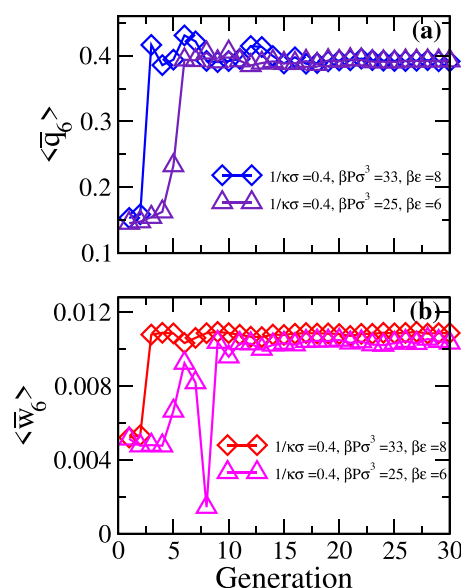


FIG. 10. Evolution of (a) $\langle \bar{q}_6 \rangle$ and (b) $\langle \bar{w}_6 \rangle$ of the sample with the highest value of the fitness at each generation during the 3D optimization with the CMA-ES method for both the investigated cases. We observe a jump of \bar{q}_6 and \bar{w}_6 when, for the first time, at least one of the generated samples is in the bcc region, even if the mean value of the multivariate Gaussian distribution at the same generation does not lie in the same region.

when the phase boundary is crossed. In CMA-ES, the size of the steps from one generation to the other varies, depending on the current form of the covariance matrix. This makes the algorithm explore the landscape in an optimal way, sacrificing information about the phase boundaries. Thus, we conclude that SP-ID and CMA-ES are comparable in their performance, but SP-ID has an edge in retrieving information on the phase boundaries. Tuning the free parameters of CMA-ES (c_i) one can force the algorithm to perform small steps and gain the same information regarding the phase boundaries, but this would result in a huge rise of the computational effort required, which would make CMA-ES an algorithm of poor efficiency and use. Finally, it should be noted that in CMA-ES all the simulations belonging to the same generation can be run simultaneously, which

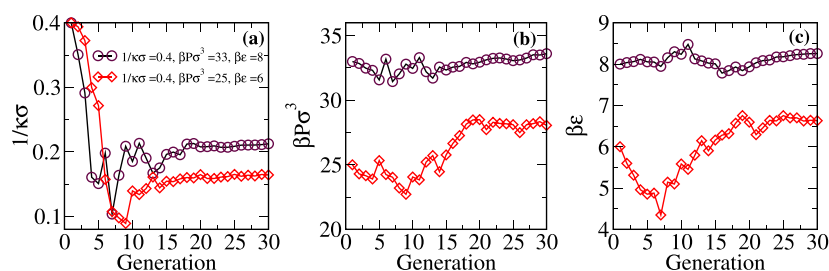


FIG. 9. Evolution of the mean value of the multivariate Gaussian distribution for (a) inverse Debye screening length $1/\kappa\sigma$, (b) reduced pressure $\beta P\sigma^3$, and (c) contact value $\beta\epsilon$ when the system is initialized in the fluid phase at (i) $\beta P\sigma^3 = 33$, $1/\kappa\sigma = 0.4$, and $\beta\epsilon = 8$ (indigo colored circles), and (ii) $\beta P\sigma^3 = 25$, $1/\kappa\sigma = 0.4$, and $\beta\epsilon = 6$ (red colored diamonds). Three parameters are tuned to target the bcc structure, namely $\beta P\sigma^3$, $1/\kappa\sigma$, and $\beta\epsilon$ using the CMA-ES method.

can be a great advantage, depending on the computational resources of the user.

V. CONCLUSIONS

We studied the inverse problem of tuning interaction parameters between charged colloids interacting via a hard-core repulsive Yukawa potential, so that they self-assemble into a targeted crystal structure. We targeted the bcc structure which occupies a narrow region in the phase diagram of the above system and is therefore challenging to find. We showed a comparison between two different optimization algorithms in order to achieve our goal: Statistical Physics-inspired Inverse Design (SP-ID) and Covariance Matrix Adaption-Evolutionary Strategy (CMA-ES). The first makes use of the statistical fluctuations in the bond order parameters to iteratively change the interaction parameters of the system. In addition to effectively tuning the interaction parameters for obtaining the target structure, the SP-ID method correctly identifies the fluid-solid phase boundaries present in the phase diagram. The CMA-ES algorithm generates samples from a multivariate Gaussian distribution at each generation and evaluates the fitness of these samples in order to evolve the interaction parameters of the distribution. The number of generations needed to reach the goal is on average lower in the case of the CMA-ES, and the steps in parameter space are usually larger. This advantage is offset by the need to simulate multiple samples, and we find that the computational effort required in the two methods is comparable. On the other hand, because of the larger step sizes of the parameters, probing phase equilibrium with CMA-ES can be less straightforward than with SP-ID. Most importantly, we showed that both of these inverse methods lead to the targeted bcc structure by tuning the interactions between the particles. Thus, our results demonstrate both methods to be effective search algorithms that may be employed in other design tasks. Although the quality function used here is strictly structural, one may in principle also include quantifiers of dynamics, which may be useful in optimizing the kinetic self-assembly pathways, for example, to account for and exclude glassy dynamics, as well as other kinetic factors of self assembly. Finally, we stress how exploiting thermal fluctuations of a statistical system results in a more effective search in the fitness function space. In the future, it would be useful and fascinating to compare SP-ID to other related inverse design methods.^{18–20}

ACKNOWLEDGMENTS

We gratefully acknowledge the India-Netherlands (DST-NWO) joint project grant for supporting the collaboration between the authors.

REFERENCES

- ¹M. C. Rechtsman, F. H. Stillinger, and S. Torquato, *Phys. Rev. Lett.* **95**, 228301 (2005).
- ²M. C. Rechtsman, F. H. Stillinger, and S. Torquato, *Phys. Rev. Lett.* **97**, 239901 (2006).
- ³M. C. Rechtsman *et al.*, *Phys. Rev. Lett.* **101**, 073902 (2008).
- ⁴M. C. Rechtsman, F. H. Stillinger, and S. Torquato, *Phys. Rev. Lett.* **101**, 085501 (2008).
- ⁵S. Torquato, *Soft Matter* **5**, 1157 (2009).
- ⁶A. Jain, J. A. Bollinger, and T. M. Truskett, *Chem. Eng. J.* **60**, 2732 (2014).
- ⁷J. M. Jaeger, *Soft Matter* **11**, 12 (2015).
- ⁸K. M. Ho, C. T. Chan, and C. M. Soukoulis, *Phys. Rev. Lett.* **65**, 3152 (1990).
- ⁹A. R. Oganov and C. W. Glass, *J. Chem. Phys.* **124**, 244704 (2006).
- ¹⁰B. I. Dahiyat and S. L. Mayo, *Science* **278**, 82 (1997).
- ¹¹B. Kuhlman *et al.*, *Science* **302**, 1364 (2003).
- ¹²S. Hyun and S. Torquato, *J. Mater. Res.* **16**, 280 (2001).
- ¹³A. F. Hannon *et al.*, *ACS Macro Lett.* **2**, 251 (2013).
- ¹⁴J. B. Chang *et al.*, *Nat. Commun.* **5**, 3305 (2014).
- ¹⁵M. C. Rechtsman, F. H. Stillinger, and S. Torquato, *Phys. Rev. E* **73**, 011406 (2006).
- ¹⁶A. F. Hannon *et al.*, *Nano Lett.* **14**, 318 (2014).
- ¹⁷S. Hormoz and M. P. Brenner, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5193 (2011).
- ¹⁸B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, *J. Chem. Phys.* **145**, 111101 (2016).
- ¹⁹R. B. Jadrich, B. A. Lindquist, and T. M. Truskett, *J. Chem. Phys.* **146**, 184103 (2017).
- ²⁰C. S. Adorf, J. Antonaglia, J. Dshemuchadse, and S. C. Glotzer, *J. Chem. Phys.* **149**, 204102 (2018).
- ²¹M. Z. Miskin, G. Khaira, J. J. de Pablo, and H. M. Jaeger, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 34 (2016).
- ²²P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, *Phys. Rev. B* **28**, 784 (1983).
- ²³W. Lechner and C. Dellago, *J. Chem. Phys.* **129**, 114707 (2008).
- ²⁴S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- ²⁵L. Ingber, Global Optimization C-Code, Caltech Alumni Association, Pasadena, CA, 1993.
- ²⁶L. Ingber, preprint [arXiv:cs/0001018](https://arxiv.org/abs/cs/0001018) (2000).
- ²⁷M. A. M. De Oca, T. Stützle, M. Birattari, and M. Dorigo, in *International Workshop on Ant Colony Optimization and Swarm Intelligence* (Springer, 2006), pp. 1–12.
- ²⁸M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, 1998).
- ²⁹N. Hansen and A. Ostermeier, *Evol. Comput.* **9**, 159 (2001).
- ³⁰N. Hansen, S. D. Müller, and P. Koumoutsakos, *Evol. Comput.* **11**, 1 (2003).
- ³¹N. Hansen, *Towards a New Evolutionary Computation* (Springer, 2006), pp. 75–102.
- ³²E. J. Meijer, F. El Azhar, and F. El Azhar, *J. Chem. Phys.* **106**, 4678 (1997).
- ³³F. El Azhar, M. Baus, J. P. Ryckaert, and E. J. Meijer, *J. Chem. Phys.* **112**, 5121 (2000).
- ³⁴A. P. Hynninen and M. Dijkstra, *Phys. Rev. E* **68**, 021407 (2003).
- ³⁵J. A. van Meel, L. Filion, C. Valeriani, and D. Frenkel, *J. Chem. Phys.* **136**, 234107 (2012).