Precise determination of pair interactions from pair statistics of many-body systems in and out of equilibrium

Salvatore Torquato [1,2,*] and Haina Wang [3]

[1]*Department of Chemistry, Department of Physics, Princeton Institute of Materials,*
and Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, USA
[2]*School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, New Jersey 08540, USA*
[3]*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA*

The determination of the pair potential $v(\mathbf{r})$ that accurately yields an equilibrium state at positive temperature T with a prescribed pair correlation function $g_2(\mathbf{r})$ or corresponding structure factor $S(\mathbf{k})$ in d-dimensional Euclidean space $\mathbb{R}^d$ is an outstanding inverse statistical mechanics problem with far-reaching implications. Recently, Zhang and Torquato [Phys. Rev. E **101**, 032124 (2020)] conjectured that any realizable $g_2(\mathbf{r})$ or $S(\mathbf{k})$ corresponding to a translationally invariant nonequilibrium system can be attained by a classical equilibrium ensemble involving only (up to) effective pair interactions. Testing this conjecture for nonequilibrium systems as well as for nontrivial equilibrium states requires improved inverse methodologies. We have devised an optimization algorithm to precisely determine effective pair potentials that correspond to pair statistics of general translationally invariant disordered many-body equilibrium or nonequilibrium systems at positive temperatures. This methodology utilizes a parameterized family of pointwise basis functions for the potential function whose initial form is informed by small-, intermediate- and large-distance behaviors dictated by statistical-mechanical theory. Subsequently, a nonlinear optimization technique is utilized to minimize an objective function that incorporates both the target pair correlation function $g_2(\mathbf{r})$ and structure factor $S(\mathbf{k})$ so that the small intermediate- and large-distance correlations are very accurately captured. To illustrate the versatility and power of our methodology, we accurately determine the effective pair interactions of the following four diverse target systems: (1) Lennard-Jones system in the vicinity of its critical point, (2) liquid under the Dzugutov potential, (3) nonequilibrium random sequential addition packing, and (4) a nonequilibrium hyperuniform "cloaked" uniformly randomized lattice. We found that the optimized pair potentials generate corresponding pair statistics that accurately match their corresponding targets with total L_2-norm errors that are an order of magnitude smaller than that of previous methods. The results of our investigation lend further support to the Zhang-Torquato conjecture. Furthermore, our algorithm will enable one to probe systems with identical pair statistics but different higher-body statistics, which will shed light on the well-known degeneracy problem of statistical mechanics.

I. INTRODUCTION

The relationship between the interactions in many-body systems and their corresponding structural and equilibrium or nonequilibrium properties continues to be a subject of great fundamental and practical interest in statistical physics, condensed-matter physics, chemistry, mathematics, and materials science [1–4]. Direct computer simulation of matter at the molecular and colloidal level has generated a long and insightful tradition; see Refs. [5–15] and references therein. The fruitful "forward" approach of statistical mechanics identifies a known substance that possesses scientific and/or technological interest, creates a manageable approximation to the interparticle interactions that operate in that substance, and exploits molecular dynamics or Monte Carlo algorithms to predict nontrivial details concerning the structure, thermodynamics, and kinetic features of the system. Inverse

statistical-mechanical methods [3] allow for a new mode of thinking about the structure and physical properties of condensed phases of matter and are ideally suited for materials discovery by design. In the inverse approach, one attempts to determine a potential function (subject to constraints) that robustly and spontaneously leads to *targeted* many-particle configurations, targeted correlation functions or a targeted set of physical properties over a wide range of conditions [3,16–18]. The targeting of equilibrium crystal states has been particularly successful, including the capacity to design low-coordinated crystals as (zero-temperature) ground states [3,19–24] as well as crystals at positive temperature T [25,26].

Finding the pair potential $v(\mathbf{r})$ that accurately yields a single-component equilibrium disordered state (e.g., simple liquids and polymers) at positive T with a prescribed pair correlation function $g_2(\mathbf{r})$ or corresponding structure factor $S(\mathbf{k})$ in d-dimensional Euclidean space $\mathbb{R}^d$ is an outstanding inverse problem with far-reaching implications. For example, such determinations enable one to devise atomic-based Hamiltonian models that are consistent with the experimental

*torquato@princeton.edu

measurements of $S(\mathbf{k})$ via scattering-intensity studies. In addition, one can use the solutions of such inverse problems to *design*, at will, portions of the equilibrium phase diagram of colloidal systems when *hypothesized* functional forms for the pair statistics are realizable [27–29] by effective pair interactions.

Moreover, Zhang and Torquato [29] recently introduced a theoretical formalism that provides a means to draw equilibrium classical particle configurations from canonical ensembles with one- and two-body interactions that correspond to targeted functional forms for $g_2(\mathbf{r})$ or $S(\mathbf{k})$. This formalism enabled them to devise an efficient algorithm to construct systematically canonical-ensemble particle configurations with such targeted $S(\mathbf{k})$ at all wave numbers, whenever realizable. However, their procedure does not provide the explicit forms of the underlying one- and two-body potentials. In the same study, Zhang and Torquato [29] conjectured that any realizable $g_2(\mathbf{r})$ or $S(\mathbf{k})$ corresponding to a translationally invariant nonequilibrium system can be attained by an equilibrium ensemble involving only (up to) effective pair interactions. Successful solutions to the aforementioned inverse problem would enable one to test this remarkable conjecture and its implications if proved to be true.

While significant progress has been made in ascertaining pair interactions from pair statistics [17,18,30–33], substantial computational challenges remain. Improved methods must be formulated to ensure a highly precise correspondence between the pair potential and the associated pair statistics for both equilibrium and nonequilibrium states of matter. In the case of equilibrium ensembles, Henderson's theorem states that in a classical homogeneous many-body system, the pair potential $v(\mathbf{r})$ that gives rise to a given equilibrium pair correlation function $g_2(\mathbf{r})$ at fixed number density ρ and temperature $T > 0$ is unique up to an additive constant [34]. While this is a powerful uniqueness theorem, in practice, it has been shown that very similar equilibrium pair statistics may correspond to distinctly different pair potentials [35], and hence highly precise pair-correlation information is required to achieve an accurate pair potential function. Furthermore, testing the Zhang-Torquato conjecture for systems out of equilibrium [29] requires computational precision that is beyond currently available techniques. Importantly, no current methods treat situations in which a one-body potential must also be incorporated to stabilize systems with certain long-ranged pair potentials that are required to achieve "incompressible" exotic hyperuniform states [36,37]. Indeed, Buck [16] has noted that in cases where the interaction is long-ranged, previous approaches could yield ambiguous solutions, i.e., different inversion approaches yield different potentials for the same scattering data. Thus, hyperuniform targets demand a completely new and improved inverse methodology.

All previous iterative predictor-corrector methods [30–33], which appear to be the most accurate inverse procedures, begin with an initial discretized (binned) approximation of a trial pair potential, but without considering one-body interactions. The trial pair potential at each binned distance is iteratively updated to attempt to reduce the difference between the target and trial pair statistics. A popular method, called the iterative Boltzmann inversion (IBI) [32], takes the initial trial potential as the potential of mean force and then iteratively

updates the trial potential based on the difference between the potentials of mean force for the target and trial pair correlation functions. A similar fixed-point iteration approach proposed by Lyubartserv *et al.* [31] reaches comparable accuracy to that of IBI [32]. The best currently available scheme is the iterative hypernetted chain inversion (IHNCI) scheme, which was introduced by Levesque, Weiss, and Reatto [30] and refined in Ref. [33]. The IHNCI updates to attempt to match both the target $g_2(\mathbf{r})$ and target $S(\mathbf{k})$ by iterating the Ornstein-Zernike equation using the hypernetted chain approximation. By construction, this procedure can only lead to an approximation to the desired pair potential. Moreover, because all previous methods do not optimize a pair-statistic "distance" functional, they are unable to detect poor agreement between the target and trial pair statistics that may arise as the simulation evolves, leading to increasingly inaccurate corresponding trial potentials.

To improve on previous methods and address the aforementioned new challenges, we have devised an optimization algorithm that is informed by statistical-mechanical theory to find effective one- and two-body potentials yielding positive-T equilibrium states that correspond to pair statistics of general translationally invariant (statistically homogeneous) disordered many-body systems in and out of equilibrium with unprecedented accuracy [38]. Our methodology departs from previous techniques in several significant ways. First, we pose the task as an optimization problem to minimize an objective function that incorporates both the target pair correlation function $g_{2,T}(\mathbf{r})$ and structure factor $S_T(\mathbf{k})$ that would correspond to one- and two-body (pair) potential. Second, unlike previous procedures, our methodology uses a parameterized family of *pointwise basis functions* for the potential function $v(\mathbf{r}; \mathbf{a})$, where $\mathbf{a}$ represents a supervector of parameters (Sec. IV) whose initial form is informed by the small-, intermediate- and large-distance behaviors of the targeted $g_{2,T}(\mathbf{r})$ and $S_T(\mathbf{k})$, as dictated by statistical-mechanical theory. Pointwise potential functions are superior to previously employed binned potentials, since they do not suffer from the accumulation of random errors during a simulation, resulting in more accurate pair interactions. Binned potentials that contain random noise can lead to unrealistic forces in molecular dynamics and result in incorrect supercooled or quenched structures. Third, a nonlinear optimization technique is utilized to minimize the objective function so that the very small, intermediate-, and large-distance correlations are very accurately captured. The reader is referred to Sec. IV for algorithmic details of our inverse methodology.

To illustrate the versatility and power of our methodology, we accurately determine the effective pair interactions of the following four diverse target systems:

(1) two-dimensional (2D) Lennard-Jones system in the vicinity of its critical point

(2) three-dimensional (3D) liquid under the Dzugotov potential [39]

(3) 2D nonequilibrium random sequential addition (RSA) packing near saturation [40,41]

(4) 3D nonequilibrium hyperuniform "cloaked" uniformly randomized lattice (URL) [42].

Figure 1 shows snapshots of all four target configurations. These targets can be problematic for previous procedures to
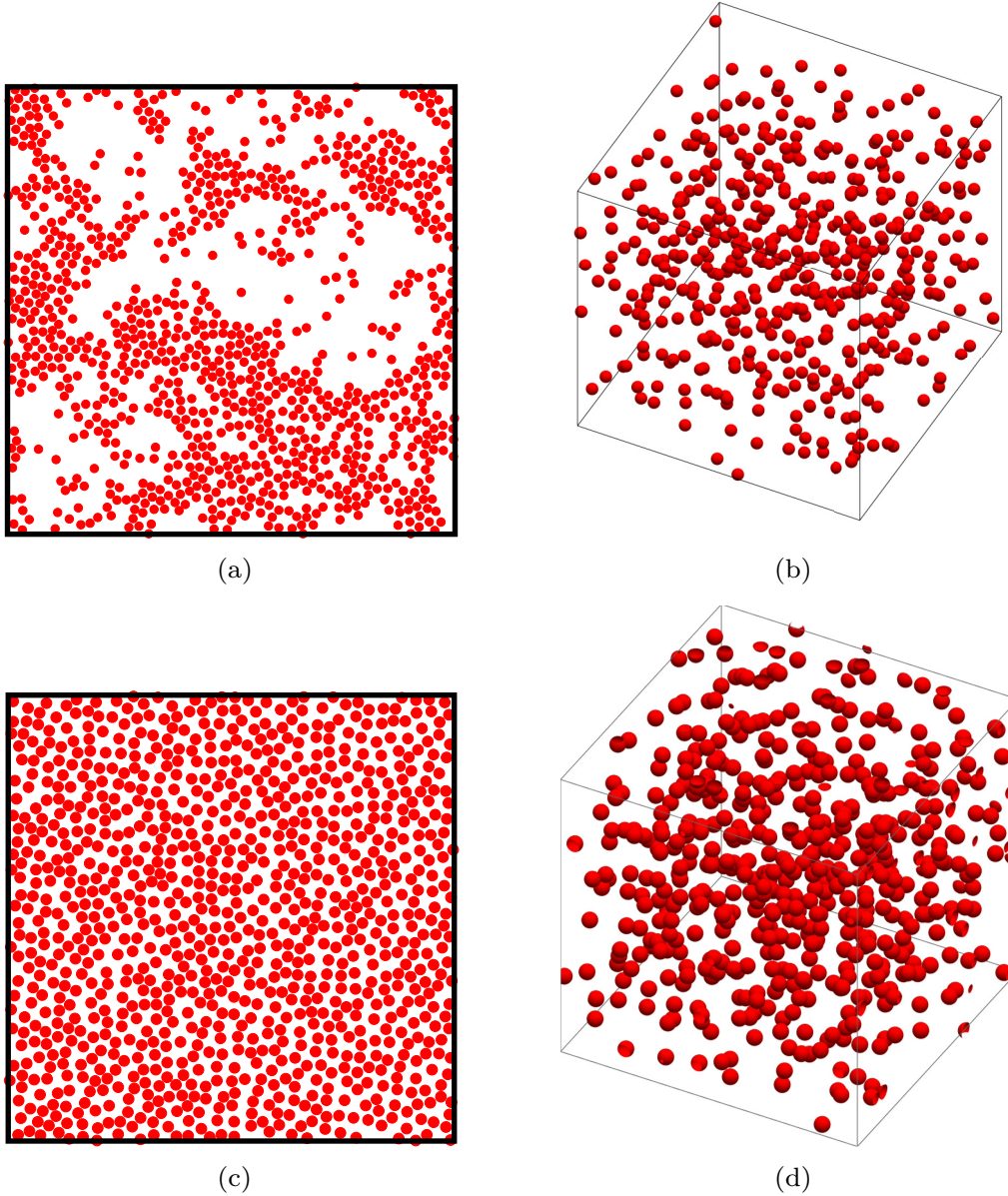
FIG. 1. Snapshots of the configurations corresponding to the four target systems considered in this study. (a) A representative 2D 1000-particle configuration of the target Lennard-Jones system near its critical point. (b) A representative 3D 512-particle configuration of the target liquid under the Dzugutov potential. (c) A portion of a representative 2D 10 000-particle RSA configuration very near the saturation state. Only 1000 particles are displayed. (d) A portion of a representative 3D 2744-particle configuration of a cloaked URL system. Only 512 particles are displayed.

achieve accurate pair potentials for various reasons, as detailed in Sec. V.

To assess the accuracy of inverse methodologies to target pair statistics, we introduce and utilize the following dimensionless L_2-norm error:

$$\mathcal{E} = \sqrt{D_{g_2} + D_S}, \tag{1}$$

where

$$D_{g_2} = \rho \int_{\mathbb{R}^d} [g_{2,T}(\mathbf{r}) - g_{2,F}(\mathbf{r}; \mathbf{a})]^2 \, d\mathbf{r}, \tag{2}$$

$$D_S = \frac{1}{\rho(2\pi)^d} \int_{\mathbb{R}^d} [S_T(\mathbf{k}) - S_F(\mathbf{k}; \mathbf{a})]^2 \, d\mathbf{k}, \tag{3}$$

where $g_{2,F}(\mathbf{r}; \mathbf{a})$ and $S_F(\mathbf{k}; \mathbf{a})$ represent the final pair statistics at the end of the optimization, which depend on the supervector $\mathbf{a}$. It is noteworthy that our algorithm generally provides errors in the final pair statistics for all four targets that are an order of magnitude smaller than previous methods, as we will subsequently show. Importantly, it reaches the precision required to recover the unique potential dictated by Henderson's theorem [34]. Realizing the nonequilibrium pair statistics of RSA and URL systems by corresponding equilibrium systems with effective pair interactions lend further support to the Zhang-Torquato conjecture [29] and shed light on the well-known degeneracy problem [43,44]. Structural, thermodynamic and dynamic properties of the equilibrium systems could also reveal nontrivial information on the nonequilibrium

states, such as the degree to which the latter systems are out of equilibrium, as detailed in Sec. VI.

We begin by providing basic definitions and background in Sec. II. In Sec. III, we describe the asymptotic behavior of the pair potential from large-scale pair correlations for non-hyperuniform and hyperuniform targets, which are dictated by statistical-mechanical theory [37]. Here we also derive a general expression for the small-$|\mathbf{k}|$ behavior of $S(\mathbf{k})$ for systems equilibrated under inverse power-law pair potentials that applies in any space dimension; see Eq. (26). In Sec. IV, we provide a detailed description of our inverse methodology that is generally applicable to statistically homogeneous but anisotropic disordered states. In Sec. V, we apply our methodology on diverse target translationally invariant disordered systems, in and out of equilibrium. We provide concluding remarks in Sec. VI.

II. BASIC DEFINITIONS AND BACKGROUND

A. Pair statistics

We consider many-particle systems in d-dimensional Euclidean space $\mathbb{R}^d$ that are completely statistically characterized by the n-particle probability density functions $\rho_n(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ for all $n \geq 1$ [7]. Our primary interest in this paper is in the one- and two-body statistics. In the case of statistically homogeneous systems, $\rho_1(\mathbf{r}_1) = \rho$ and $\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \rho^2 g_2(\mathbf{r})$, ρ is the number density in the thermodynamic limit, $g_2(\mathbf{r})$ is the pair correlation function, and $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. If the system is also statistically isotropic, then $g_2(\mathbf{r})$ is the radial function $g_2(r)$, where $r = |\mathbf{r}|$. The ensemble-averaged structure factor $S(\mathbf{k})$ is defined as

$$S(\mathbf{k}) = 1 + \rho \tilde{h}(\mathbf{k}), \tag{4}$$

where $h(\mathbf{r}) = g_2(\mathbf{r}) - 1$ is the total correlation function, and $\tilde{h}(\mathbf{k})$ is the Fourier transform of $h(\mathbf{r})$. The Fourier transform of a function $f(\mathbf{r})$ that depends on the vector $\mathbf{r}$ in $\mathbb{R}^d$ is give by

$$\tilde{f}(\mathbf{k}) = \int_{\mathbb{R}^d} f(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) \, d\mathbf{r}, \tag{5}$$

where $\mathbf{k} \cdot \mathbf{r} = \sum_{i=1}^{d} k_i r_i$ is the conventional Euclidean inner product of two real-valued vectors. When it is well defined, the corresponding inverse Fourier transform is given by

$$f(\mathbf{r}) = \left(\frac{1}{2\pi}\right)^d \int_{\mathbb{R}^d} \tilde{f}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) \, d\mathbf{k}. \tag{6}$$

If f is a radial function, i.e., f depends only on the modulus $r = |\mathbf{r}|$ of the vector $\mathbf{r}$, its Fourier transform is given by

$$\tilde{f}(k) = (2\pi)^{\frac{d}{2}} \int_0^\infty r^{d-1} f(r) \frac{J_{(d/2)-1}(kr)}{(kr)^{(d/2)-1}} dr, \tag{7}$$

where $J_\nu(x)$ is the Bessel function of the first kind of order ν. The inverse Fourier transform of $\tilde{f}(k)$ is given by

$$f(r) = \left(\frac{1}{2\pi}\right)^{\frac{d}{2}} \int_0^\infty k^{d-1} \tilde{f}(k) \frac{J_{(d/2)-1}(kr)}{(kr)^{(d/2)-1}} dk. \tag{8}$$

For a single periodic configuration containing number N point particles at positions $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ within a fundamental cell F of a lattice Λ, the *scattering intensity* $\mathcal{I}(\mathbf{k})$ is defined

as

$$\mathcal{I}(\mathbf{k}) = \frac{\left|\sum_{i=1}^{N} e^{-i\mathbf{k} \cdot \mathbf{r}_i}\right|^2}{N}. \tag{9}$$

For an ensemble of periodic configurations of N particles within the fundamental cell F, the ensemble average of the scattering intensity in the infinite-volume limit is directly related to structure factor $S(\mathbf{k})$ by

$$\lim_{N,V_F \to \infty} \langle \mathcal{I}(\mathbf{k}) \rangle = (2\pi)^d \rho \delta(\mathbf{k}) + S(\mathbf{k}), \tag{10}$$

where V_F is the volume of the fundamental cell and δ is the Dirac delta function [37]. In simulations of many-body systems with finite N under periodic boundary conditions, Eq. (9) is used to compute $S(\mathbf{k})$ directly by averaging over configurations.

B. Hyperuniformity and nonhyperuniformity

A *hyperuniform* point configuration in d-dimensional Euclidean space $\mathbb{R}^d$ is characterized by an anomalous suppression of large-scale density fluctuations relative to those in typical disordered systems, such as liquids and structural glasses [36,37]. More precisely, a hyperuniform point pattern is one in which the structure factor $S(\mathbf{k}) = 1 + \rho \tilde{h}(\mathbf{k})$ tends to zero as the wave number $k = |\mathbf{k}|$ tends to zero [36]:

$$\lim_{|\mathbf{k}| \to 0} S(\mathbf{k}) = 0. \tag{11}$$

This hyperuniformity condition implies the following direct-space sum rule:

$$\rho \int_{\mathbb{R}^d} h(\mathbf{r}) \, d\mathbf{r} = -1. \tag{12}$$

An equivalent definition of hyperuniformity is based on the local number variance $\sigma^2(r) = \langle N(R)^2 \rangle - \langle N(R) \rangle^2$ associated with the number $N(R)$ of points within a d-dimensional spherical observation window of radius R, where angular brackets denote an ensemble average. A point pattern in $\mathbb{R}^d$ is hyperuniform if its variance grows in the large-R limit slower than R^d [36].

Consider systems that are characterized by a structure factor with a radial power-law form in the vicinity of the origin,

$$S(\mathbf{k}) \sim |\mathbf{k}|^\alpha \quad \text{for } |\mathbf{k}| \to 0. \tag{13}$$

For hyperuniform systems, the exponent α is positive ($\alpha > 0$) and its value determines three different large-R scaling behaviors of the number variance [36,37,45]:

$$\sigma^2(R) \sim \begin{cases} R^{d-1}, & \alpha > 1 \text{ (class I)} \\ R^{d-1} \ln R, & \alpha = 1 \text{ (class II)} \\ R^{d-\alpha}, & 0 < \alpha < 1 \text{ (class III)}. \end{cases} \tag{14}$$

By contrast, for any *nonhyperuniform* system, the local variance has the following large-R scaling behaviors [46]:

$$\sigma^2(R) \sim \begin{cases} R^d, & \alpha = 0 \text{ (typical nonhyperuniform)} \\ R^{d-\alpha}, & \alpha < 0 \text{ (antihyperuniform)}. \end{cases} \tag{15}$$

For a "typical" nonhyperuniform system, $S(0)$ is bounded. In antihyperuniform systems [37], $S(0)$ is unbounded,

$$\lim_{|\mathbf{k}| \to 0} S(\mathbf{k}) = \infty, \tag{16}$$

and hence are diametrically opposite to hyperuniform systems. Antihyperuniform systems include fractals, systems at thermal critical points (e.g., liquid-vapor and magnetic critical points) [47–51], as well as certain substitution tilings [52]. A *hyposurficial* state is a special nonhyperuniform system that lacks a "surface-area" term in the growth of the number variance and have pair statistics that obeys the following sum rule [36]:

$$\int_0^\infty r^d h(r)\, dr = 0, \tag{17}$$

which implies that they must generally contain both negative and positive correlations. In equilibrium systems, hyposurficial states arise in the supercritical region of the phase diagram. Hyposurficial states have been also shown to arise in computer simulations of phase transitions involving amorphous ices [53].

C. Hyperuniform equilibrium states

Torquato [37] utilized the well-known fluctuation-compressibility relation for single-component equilibrium systems [7]

$$\rho k_B T \kappa_T = S(0) \tag{18}$$

to infer salient conclusions about equilibrium hyperuniform states at $T = 0$ (ground states) and at positive temperatures. Specifically, any ground state ($T = 0$), ordered or disordered, in which the isothermal compressibility κ_T is bounded and positive must be hyperuniform. This is true more generally if the product $T\kappa_T$ tends to a nonnegative constant in the limit $T \to 0$. Moreover, in order to have a hyperuniform system that is in equilibrium at any positive T, κ_T must be identically zero, i.e., the system must be thermodynamically incompressible, at least classically. Such a situation requires an effective pair potential that is unusually long-ranged, as will be specifically shown in Sec. III.

III. ASYMPTOTIC BEHAVIOR OF THE PAIR POTENTIAL FROM LARGE-SCALE PAIR CORRELATIONS

In this section, we describe how to infer the large-distance behavior of the pair potential from targeted large-scale pair correlations for nonhyperuniform and hyperuniform states, as derived by Torquato [37]. We also derive an expression (26) for the small-$|\mathbf{k}|$ behavior of $S(\mathbf{k})$ for systems equilibrated under inverse power-law pair potentials in any space dimension.

For single-component equilibrium fluids, the asymptotic behavior of the direct correlation function $c(\mathbf{r})$, defined via the Ornstein-Zernike integral equation [7], for large $|\mathbf{r}|$ determines the asymptotic behavior of the pair potential $v(\mathbf{r})$, provided that $h^2(\mathbf{r}) \ll |\beta v(\mathbf{r})|$ [54]; specifically, we have

$$c(\mathbf{r}) \sim -\beta v(\mathbf{r}) \quad (|\mathbf{r}| \to \infty), \tag{19}$$

where $\beta = 1/(k_B T)$ and k_B is the Boltzmann constant. While the condition $h^2(\mathbf{r}) \ll |\beta v(\mathbf{r})|$ is violated at a liquid-vapor

critical point [since $h(\mathbf{r}) \sim 1/|\mathbf{r}|^{d-2+\eta}$ in the large-$|\mathbf{r}|$ limit, where $\eta = 1/4$ for $d = 2$ and $\eta \approx 0.036$ for $d = 3$ [51]], it is not violated away from and near the critical point. The latter conclusion follows from the well-known Ornstein-Zernike analysis that in the vicinity of the critical point, the structure factor takes the following Lorentzian form in the infinite-wavelength limit [49,55]:

$$S(\mathbf{k}) = \frac{S(0)}{1 + \xi^2 |\mathbf{k}|^2 + O(k^4)} \quad |\mathbf{k}| \to 0, \tag{20}$$

where ξ is the correlation length. It then immediately follows that the total correlation function has the large-r form

$$h(\mathbf{r}) \sim \frac{\exp(-|\mathbf{r}|/\xi)}{|\mathbf{r}|^{(d-1)/2}} \quad (|\mathbf{r}| \to \infty). \tag{21}$$

At the critical density $\rho = \rho_c$, ξ tends to infinity as the temperature approaches the critical temperature T_c via the scaling relation [49]

$$\xi \propto \left(\frac{T - T_c}{T_c}\right)^{-\nu} \quad T \to T_c^+, \tag{22}$$

where $\nu = 1$ for $d = 2$ and $\nu \approx 0.63$ for $d = 3$ [51]. Thus, in light of (21) and (22), we can conclude (19) applies in the vicinity of a critical point for any bounded correlation length ξ.

The large-$|\mathbf{r}|$ behavior of $c(\mathbf{r})$ or $v(\mathbf{r})$ via (19) can be extracted from the structure factor $S(\mathbf{k})$ using the Fourier representation of the Ornstein-Zernike integral equation [56]:

$$\tilde{c}(\mathbf{k}) = \frac{\tilde{h}(\mathbf{k})}{S(\mathbf{k})} \tag{23}$$

and examining the small-wave-number behavior of $\tilde{c}(\mathbf{k})$, which is the Fourier transform of $c(\mathbf{r})$. If $\tilde{c}(\mathbf{k})$ is analytic at the origin, i.e., the Taylor series about $|\mathbf{k}| = 0$ involves only even powers of the wave number $|\mathbf{k}|$, it immediately follows that $c(\mathbf{r})$ decays exponentially fast or faster [37], and hence via (19), $v(\mathbf{r})$ must decay exponentially fast or faster, i.e., faster than any inverse power law. However, if the pair potential $v(\mathbf{r})$ for large $\mathbf{r}$ is asymptotically given by the following power-law form:

$$v(\mathbf{r}) \sim \frac{E}{|\mathbf{r}|^{d+\zeta}} \quad (|\mathbf{r}| \to \infty), \tag{24}$$

where E is a constant with units of energy that may be positive or negative, $|\mathbf{r}|$ is a dimensionless pair distance and $\zeta > 0$, then according to (19), we have

$$c(\mathbf{r}) \sim -\beta v(\mathbf{r}) = -\frac{\beta E}{|\mathbf{r}|^{d+\zeta}} \quad (|\mathbf{r}| \to \infty). \tag{25}$$

Given a system in which the tail of the potential has the power-law form (24), we can extract the long-wavelength behavior of the structure factor following the asymptotic analysis of the Ornstein-Zernike equation presented in Ref. [37]. The specific behavior of $S(\mathbf{k})$ depends on whether ζ is an odd integer, even integer or fractional. When ζ is an odd integer, the structure factor in the long-wavelength limit consists of nonanalytic and analytic contributions:

$$S(\mathbf{k}) \sim -\frac{E}{k_B T} C_1(\zeta, d) \rho S(0)^2 |\mathbf{k}|^\zeta + f(\mathbf{k}) \quad (|\mathbf{k}| \to 0), \tag{26}$$

where

$$C_1(\zeta, d) = \frac{\pi^{1+d/2}}{2^\zeta \Gamma(1 + \zeta/2)\Gamma[(d+\zeta)/2]\sin(\pi\zeta/2)} \quad (27)$$

and $f(\mathbf{k})$ is an analytic function of $|\mathbf{k}|$ at the origin of the form

$$f(\mathbf{k}) = s_0 + s_2|\mathbf{k}|^2 + s_4|\mathbf{k}|^4 + s_6|\mathbf{k}|^6 + \cdots, \quad (28)$$

whereas $s_0 = S(0)$, independent of the value of the positive exponent ζ, the coefficients s_{2m} for $m \geqslant 1$ depend on ζ and a subset of them may be expressed as even-order moments of the total correlation function $h(\mathbf{r})$. When ζ is an even integer, the asymptotic form (26) no longer applies, but the asymptotic formula now includes a nonanalytic term proportional to $\ln(|\mathbf{k}|)|\mathbf{k}|^\zeta$.

In the case of an equilibrium hyperuniform state whose structure factor is characterized by the power law (13), the Ornstein-Zernike equation (23) implies that $\tilde{c}(\mathbf{k})$ contains a singularity at the origin [37]:

$$\tilde{c}(\mathbf{k}) \sim -|\mathbf{k}|^{-\alpha}, \quad |\mathbf{k}| \to 0. \quad (29)$$

Fourier transformation of Eq. (29) shows that $c(\mathbf{r})$ and hence $v(r)$ have the inverse power-law decay [37]

$$c(\mathbf{r}) \sim -\beta v(\mathbf{r}) \sim -|\mathbf{r}|^{-(d-\alpha)}, \quad |\mathbf{r}| \to \infty. \quad (30)$$

Thus, $v(r)$ is long-ranged in the sense that its volume integral is unbounded. Therefore, to stabilize a classical hyperuniform system at positive T, which is thermodynamically incompressible due to Eq. (18), one must treat it as a system of "like-charged" particles immersed in a rigid "background" of equal and opposite "charge," i.e., the system must have overall charge neutrality [37]. Such a rigid background contribution corresponds to a one-body potential, which is described in detail in Sec. IV D.

IV. NEW INVERSE METHODOLOGY

For present purposes, we consider the canonical ensemble associated with classical many-body systems of N particles in a finite but large region $\Omega \subset \mathbb{R}^d$ of volume V in thermal equilibrium with a heat bath at absolute temperature T. We assume that the particles are subject to a one- and two-body potential, $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r}_1, \mathbf{r}_2)$, respectively. Thus, the configurational potential energy of particles at $\mathbf{r}^N = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ is given by

$$\Phi(\mathbf{r}^N) = \sum_{i=1}^{N} \phi_1(\mathbf{r}_i) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \phi_2(\mathbf{r}_i, \mathbf{r}_j). \quad (31)$$

In order to have a system with a well-defined thermodynamic behavior, we consider ϕ_1 and ϕ_2 that are *stable*, i.e., there exist positive constants $B_1, B_2 \geqslant 0$, such that for all N and all $\mathbf{r}^N \in \mathbb{R}^{Nd}$ [57]

$$\sum_{i=1}^{N} \phi_1(\mathbf{r}_i) \geqslant -NB_1 \quad (32)$$

and

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \phi_2(\mathbf{r}_i, \mathbf{r}_j) \geqslant -NB_2. \quad (33)$$

Ultimately, we are interested in the thermodynamic limit, i.e., the number density $\rho = N/V$ is a fixed constant in the limits $N \to \infty$ and $V \to \infty$. Due to an extension of Henderson's theorem, the one- and two-body potentials $\{\phi_1(\mathbf{r}), \phi_2(\mathbf{r}_i, \mathbf{r}_j)\}$ that give rise to a given set of equilibrium probability density functions $\{\rho_1(\mathbf{r}), \rho_2(\mathbf{r}_i, \mathbf{r}_j)\}$ at fixed positive and finite temperature T are unique up to additive constants [35,58].

In what follows, we describe our inverse methodology to find one- and two-body potentials $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r}_1, \mathbf{r}_2)$ at some positive and finite temperature T such that the corresponding equilibrium one- and two-body probability density functions accurately match the target one- and two-body statistics. Our multi-stage procedure consists of choosing the initial forms of pointwise basis functions for the parameterized potentials, which is informed by statistical-mechanical theory described in Sec. III, and then we iteratively optimize the potential parameters until a convergence criterion is attained. If we do not achieve sufficiently accurate solutions in the first stage, then we choose an alternative set of basis functions by refining the previous choice and iteratively proceed until our convergence criterion is attained, as detailed below.

A. Objective function

Here we describe the objective function that must be minimized for the inverse procedure. We generally assume statistically anisotropic target pair statistics that are vector-dependent, i.e., $g_{2,T}(\mathbf{r})$ and $S_T(\mathbf{k})$, and hence the associated pair potential $\phi_2(\mathbf{r})$ is directionally dependent (anisotropic) and one-body potential ϕ_1 is uniform (constant). Since it has recently been established [35] that methods that target only $g_2(\mathbf{r})$ or only $S(\mathbf{k})$ may generate effective potentials that are distinctly different from the unique pair potential dictated by Henderson's theorem, we consider an objective function that involves a "distance" functional incorporating both functions. Specifically, given the target pair statistics $g_{2,T}(\mathbf{r})$ and $S_T(\mathbf{k})$, we optimize the following objective function $\Psi(\mathbf{a})$ over the "supervector" parameter $\mathbf{a}$:

$$\Psi(\mathbf{a}) = \rho \int_{\mathbb{R}^d} w_{g_2}(\mathbf{r})[g_{2,T}(\mathbf{r}) - g_2(\mathbf{r}; \mathbf{a})]^2 d\mathbf{r}$$
$$+ \frac{1}{\rho(2\pi)^d} \int_{\mathbb{R}^d} w_S(\mathbf{k})[S_T(\mathbf{k}) - S(\mathbf{k}; \mathbf{a})]^2 d\mathbf{k}, \quad (34)$$

where $g_2(\mathbf{r}; \mathbf{a})$ and $S(\mathbf{k}; \mathbf{a})$ correspond to an equilibrated N-particle system under $v(\mathbf{r}; \mathbf{a})$ at temperature T, which can be obtained from Monte Carlo or molecular dynamics simulations under periodic boundary conditions, and $w_{g_2}(\mathbf{r})$ and $w_S(\mathbf{k})$ are weight functions. Here $\mathbf{a}$ is a supervector of parameters, described more precisely below. The weight functions are chosen so that we accurately attain the targeted small-r and small-k behaviors of $g_{2,T}(\mathbf{r})$ and $S_T(\mathbf{k})$, respectively. In this work, we use Gaussians for both w_{g_2} and w_S:

$$w_{g_2}(\mathbf{r}) = \exp\left[-\left(\frac{r}{\sigma_{g_2}}\right)^2\right], \quad (35)$$

$$w_S(\mathbf{k}) = A\exp\left[-\left(\frac{k}{\sigma_S}\right)^2\right], \quad (36)$$

where the constants σ_{g_2} and σ_S regulate the targeted ranges of $g_{2,T}(\mathbf{r})$ and $S_T(\mathbf{k})$, respectively, the constant A is added to

ensure that both terms in Eq. (34) are of the same order of magnitude, $r \equiv |\mathbf{r}|$ and $k \equiv |\mathbf{k}|$.

B. Choosing basis functions for the parameterized pair potential

In what follows, the formulation considers target structures that are both statistically homogeneous and isotropic. Therefore, the one-body potential $\phi_1(\mathbf{r})$ is independent of $\mathbf{r}$ and $\phi_2(\mathbf{r}_i, \mathbf{r}_j)$ is a radial potential function:

$$v(r_{ij}) = v(|\mathbf{r}_i - \mathbf{r}_j|) = \phi_2(\mathbf{r}_i, \mathbf{r}_j), \quad (37)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. The generalization of the methodology to directional or anisotropic pair potentials is described in Sec. IV E. We begin by considering the choice of basis functions for the pair potential $v(r)$ for nonhyperuniform targets, which means the one-body term in Eq. (31) is irrelevant, and so here we set $\phi_1 = 0$. Consider a parameterized isotropic potential function $v(r; \mathbf{a})$ that we decompose into a sum of n smooth pointwise basis functions,

$$v(r; \mathbf{a}) = \varepsilon \sum_{j=1}^{n} f_j(r/D; a_j), \quad (38)$$

where $f_j(r/\sigma; a_j)$ is the jth basis function, a_j is a vector of parameters (generally consisting of multiple parameters), $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is the "supervector" parameter, ε sets the energy scale and D is a characteristic length scale, which in the ensuing discussion is taken to be unity, $D = 1$. The components of a_j are of four types: dimensionless energy scales ε_j, dimensionless distance scales σ_j, dimensionless phases θ_j, as well as dimensionless exponents p_j.

The basis functions are chosen so that they reasonably span all potential functions that could correspond to a targeted $g_{2,T}(r)$ for all r or a targeted $S_T(k)$ for all k under the constraint that the resulting potential function $v(r)$ satisfies the small-r and large-r behaviors dictated by $g_{2,T}(r)$ and $S_T(k)$, as detailed below. For our specific targets, $f_j(r; a_j)$ are chosen from the possible following general forms:

(1) Hard core:

$$f_j(r; a_j) = \begin{cases} \infty, & r \leqslant \sigma_j \\ 0, & r > \sigma_j. \end{cases} \quad (39)$$

(2) Gamma-damped oscillatory form:

$$f_j(r; a_j) = \frac{\varepsilon_j \cos\left(r/\sigma_j^{(1)} + \theta_j\right)}{\Gamma\left(r/\sigma_j^{(2)}\right)}. \quad (40)$$

(3) Exponential-damped oscillatory form:

$$f_j(r; a_j) = \varepsilon_j \cos\left(\frac{r}{\sigma_j^{(1)}} + \theta_j\right) \exp\left[-\left(\frac{r - \sigma_j^{(2)}}{\sigma_j^{(3)}}\right)^M\right]. \quad (41)$$

(4) Yukawa-damped oscillatory form [59]:

$$f_j(r; a_j) = \varepsilon_j \cos\left(\frac{r}{\sigma_j^{(1)}} + \theta_j\right) \exp\left[-\left(\frac{r - \sigma_j^{(2)}}{\sigma_j^{(3)}}\right)^M\right] r^{-p_j}. \quad (42)$$

(5) Power-law-damped oscillatory form:

$$f_j(r; a_j) = \varepsilon_j \cos\left(\frac{r}{\sigma_j^{(1)}} + \theta_j\right) r^{-p_j}. \quad (43)$$

In this work, for simplicity and efficiency, the exponent M in Eqs. (41) and (42) is restricted to be an integer that remains fixed during the optimization process described in Sec. IV C. Of course, other target structures may require the incorporation of other basis functions.

In choosing the basis functions, we ensure that the asymptotic decay rates of all $f_j(r; a_j)$'s are no slower than that of the longest-ranged basis function determined via the large-r asymptotic analysis described in Sec. III. In Eqs. (40)–(43) we include the oscillatory factor $\cos(r/\sigma_j^{(1)} + \theta_j)$ only if the initial form for $v(r)$, which is informed by the Ornstein-Zernike equation (23) for the targeted $g_{2,T}(r)$ and $S_T(k)$, clearly exhibits oscillatory behavior, e.g., if the initial function $v(r)$ passes through the horizontal axis at three or more values of r. If not, we set $\sigma_j^{(1)} = \infty$ and $\theta_j = 0$. We regard a set of basis functions to be a good fit of $v(r)$, if the root mean square of the fit residuals is smaller than 0.01. If there is more than one set of basis functions that satisfy this condition, then we select the one with the lowest Bayesian information criterion (BIC), which is a well-established criterion for model selection that rewards goodness of fit while penalizing an increased number of free parameters [60]. Once the forms of the basis functions f_j are chosen, we then move to the procedure of optimizing the parameter "supervector" $\mathbf{a}$.

The exact large-r behavior of $v(r; \mathbf{a})$ is enforced in (38) to be the large-r behavior of the targeted $c_T(r)$ via relation (19), which is obtained from the targeted structure factor $S_T(k)$ and the Ornstein-Zernike equation (23), as described in Sec. III, provided that the target meets the mild condition described there. This asymptotic form for $v(r; \mathbf{a})$ enables us to choose one or more of the basis functions $f_j(r; a_j)$ to have the same asymptotic form. For example, one can infer from $c(r)$ whether the longest-ranged basis function that obeys the exact large-r behavior is attractive or repulsive, and whether its asymptotic decay is power-law, exponential, or superexponential. To determine the initial small-r behavior of $v(r; \mathbf{a})$, we could use highly accurate but complicated estimates of the bridge diagrams to close the Ornstein-Zernike integral equation [7]. However, such exquisitely high accuracy is not required in the initial form because the entire function $v(r; \mathbf{a})$ is subsequently optimized. To estimate the initial small- and intermediate-r behavior in (38), we simply use the hypernetted-chain (HNC) approximation [7],

$$\beta v_{\text{HNC}}(r) = h_T(r) - c_T(r) - \ln[g_{2,T}(r)]. \quad (44)$$

More specifically, the initial small-r as well as intermediate-r behaviors of the basis functions and their corresponding initial potential parameters a_j's are obtained via a numerical fit on the binned HNC approximation.

C. Optimization algorithm

The second stage of our procedure involves optimizing the supervector parameter $\mathbf{a}$ in $v(r; \mathbf{a})$, such that the objective function $\Psi(\mathbf{a})$ is minimized. For this purpose, we utilize the
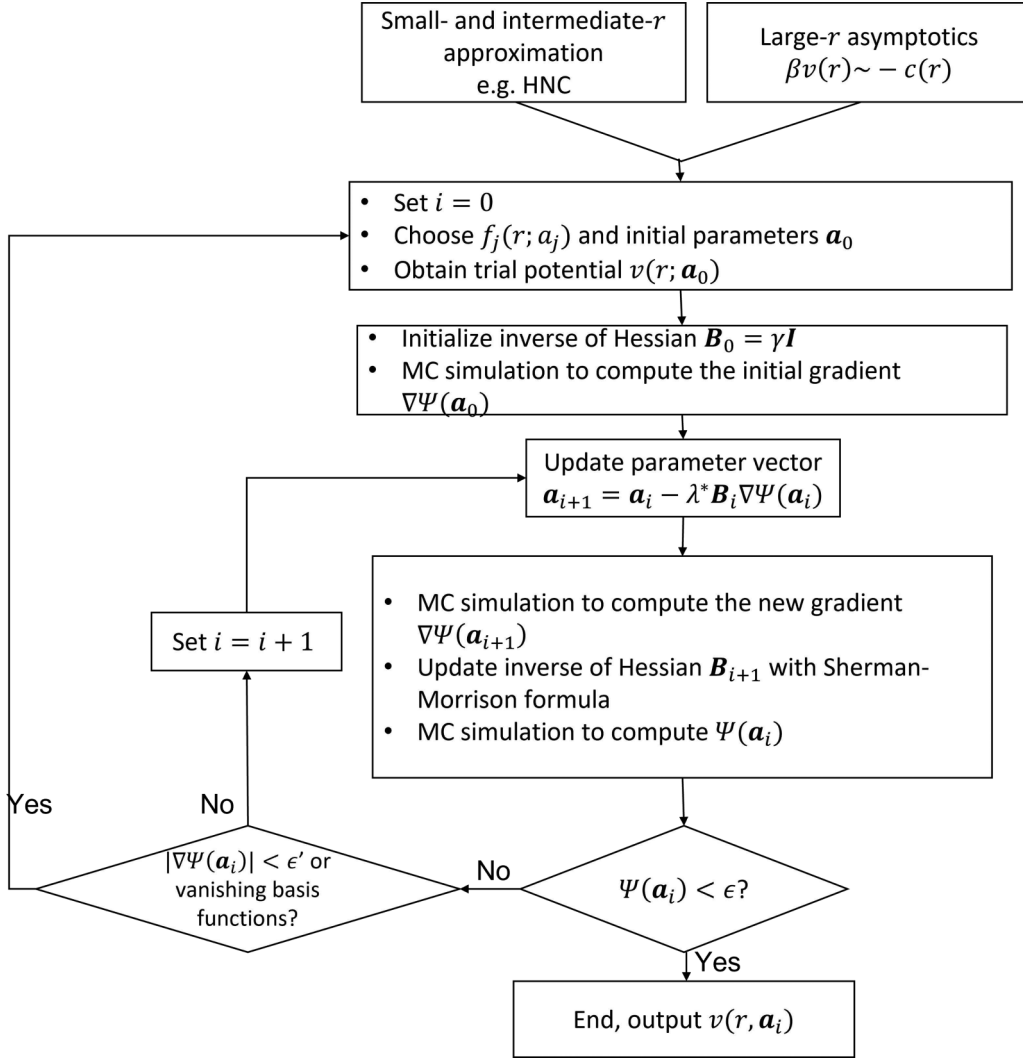
FIG. 2. Flowchart of the optimization algorithm for our inverse procedure, assuming homogeneous target systems.

BFGS algorithm [61], which is a robust nonlinear optimization technique that we have fruitfully employed for many of our previous inverse problems [62–64]. Here we choose to equilibrate systems using Monte Carlo simulations, but, of course, one may also choose to do so via molecular dynamics. Our inverse algorithm consists of the following detailed steps, which are summarized in Fig. 2 as a flowchart:

(1) Set $i = 0$. The initial guess of the inverse Hessian matrix $\mathbf{B}_0$ is taken to be the γI. Here γ is a constant which we set to be 0.1 for all systems in this study, and I is the m by m identity matrix, where $m = |\mathbf{a}|$, i.e., the total number of *scalar* components in $\mathbf{a}$.

(2) To compute the partial derivative of Ψ on the parameter $a^{(k)}$, which is the kth scalar component of $\mathbf{a}$, we choose a finite difference $\delta a > 0$. Monte Carlo simulations are performed under the potentials $v[r; (a^{(1)}, a^{(2)}, \ldots, a^{(k)} - \delta a/2, \ldots, a^{(m)})]$ and $v[r; (a^{(1)}, a^{(2)}, \ldots, a^{(k)} + \delta a/2, \ldots, a^{(m)})]$, respectively, with periodic boundary conditions, to obtain the corresponding pair statistics, namely, the standard binned ensemble-averaged $g_2(r; \mathbf{a})$ and $S(k; \mathbf{a})$ [cf. Eq. (9)] are computed. The objective functions obtained for these two simulations are denoted Ψ_{ik}^-

and Ψ_{ik}^+, respectively. The partial derivative is computed as $(\Psi_{ik}^+ - \Psi_{ik}^-)/\delta a$. The calculations of the partial derivatives are parallelized for computational efficiency. The gradient $\nabla\Psi(\mathbf{a}_i)$ is then simply a column vector containing all the partial derivatives. In practice, it is convenient to use the same value of δa for all parameters. We find that due to simulation errors in Ψ, choosing δa_j to be too small causes large uncertainties in $\nabla\Psi(\mathbf{a}_i)$. A good choice of δa_j is such that the uncertainty of $|\nabla\Psi(\mathbf{a}_i)|$ is less than 10%.

(3) Update the parameters as $\mathbf{a}_{i+1} = \mathbf{a}_i - \lambda^* \mathbf{B}_i \nabla\Psi(\mathbf{a}_i)$, where $\lambda^* = \lambda \min(1, 1/|\mathbf{B}_i \nabla\Psi(\mathbf{a}_i)|)$ and $\lambda > 0$ is a parameter that regulates the "step size" of the updates.

(4) Update the gradient $\nabla\Psi(\mathbf{a}_{i+1})$ with the same procedure as Step 2.

(5) Update the inverse Hessian $\mathbf{B}_{i+1}$ using the Sherman-Morrison formula:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \frac{\left(s_i^T y_i + y_i^T \mathbf{B}_i y_i\right)\left(s_i s_i^T\right)}{\left(s_i^T y_i\right)^2} - \frac{\mathbf{B}_i y_i s_i^T s_i y_i^T \mathbf{B}_i}{s_i^T y_i}, \quad (45)$$

where $\mathbf{y}_i = \nabla\Psi(\mathbf{a}_{i+1}) - \nabla\Psi(\mathbf{a}_i)$ and $\mathbf{s}_i = \mathbf{a}_{i+1} - \mathbf{a}_i$. In practice, due to the uncertainty in $\nabla\Psi(\mathbf{a}_i)$, which causes underestimation of $\mathbf{B}$, we reset the inverse Hessian to $\gamma\mathbf{I}$ if $|\mathbf{s}_i| < 0.05$.

(6) Monte Carlo simulations are performed under the potential $v(r; \mathbf{a}_i)$ with periodic boundary conditions to obtain the corresponding $\Psi(\mathbf{a}_i)$. The optimization iterations stop when $\Psi(\mathbf{a}_i)$ is smaller than some tolerance ϵ.

(7) If $\Psi(\mathbf{a}_i) > \epsilon$, we check whether a different set of basis functions is needed. For example, additional basis functions are required if the magnitude of the gradient $|\nabla\Psi(\mathbf{a}_i)|$ is smaller than some prescribed small value ϵ', because this implies that the BFGS optimization has found a minimum for Ψ, but this minimum does not satisfy the convergence criterion. For simplicity, we choose the additional basis functions $\{f_{n+1}(r; a_{n+1}), \ldots, f_p(r; a_p)\}$ by a numerical fit of the difference between the potentials of mean force for the target and trial pair statistics,

$$\beta \sum_{j=n+1}^{p} f_j(r; a_j) \approx \ln \frac{g_2(r; \mathbf{a}_i)}{g_{2,T}(r)}, \qquad (46)$$

where $p > n$. The forms of the additional basis functions are again chosen from the general forms (39)–(43), and the fitting procedure is exactly analogous to that described in Sec. IV B. On the other hand, if the energy-scale parameters ε_j in some basis functions fall below a certain small value, which we set to be 10^{-3}, then one can remove this basis function for efficiency. In both cases, we choose an alternative set of basis functions and the corresponding parameters according to Sec. IV B. If no modification of the basis functions is needed, we set $i = i + 1$, and repeat steps 3–7.

We note that the parameters $\sigma_{g_2}, \sigma_S, A, \lambda, \delta a, \epsilon$, and ϵ' are fixed throughout the optimization process.

D. Inverse procedure for hyperuniform targets

Here we describe the modification of the inverse procedure presented in Secs. IV B–IV C to target homogeneous systems that correspond to hyperuniform pair statistics in which $S(k) \sim k^{\alpha}$ in the limit $k \to 0$, where $\alpha > 0$. For such targets, one generally requires a long-ranged pair potential with the inverse-power-law form $v(r) \sim 1/r^{d-\alpha}$ for large r. Because the volume integral of $v(r)$ diverges in the thermodynamic limit, a neutralizing one-body potential must be added to maintain stability [37]. Therefore, we utilize the general total potential energy (31) but take the one-body potential $\phi_1(\mathbf{r})$ to be uniform with an isotropic pair potential $v(r)$. The total configurational energy (31) for N particles in a large but finite region Ω is thus given by

$$\Phi(\mathbf{r}^N; \mathbf{a}) = \sum_{i=1}^{N} \phi_1(\mathbf{a}) + \sum_{i<j}^{N} v(r_{ij}; \mathbf{a}), \qquad (47)$$

where $v(r; \mathbf{a})$ is still a sum of basis functions (38) in which the longest-ranged basis function is given by $f_j = a_j / r^{d-\alpha}$. We note that while the formalism proposed by Zhang and Torquato [29] for the realizability of pair statistics of hyperuniform targets implicitly includes a one-body potential, it was not explicitly determined there.

It is noteworthy that particles interacting under the long-ranged pair potential $v(r; \mathbf{a})$ for a hyperuniform target can

be regarded as a generalized Coulombic interaction of "like-charged" particles that are stabilized by a uniform background of equal and opposite charge [37]. Thus, to maintain stability, the one-body potential in a region Ω is taken to be

$$\phi_1(\mathbf{a}) = -\frac{\rho}{2} \int_\Omega v(x; \mathbf{a}) \, d\mathbf{x}, \qquad (48)$$

where $d\mathbf{x}$ is a volume element in Ω and $x = |\mathbf{x}|$. Note that $\phi_1(\mathbf{a})$ is a function that depends on the region Ω. This use of such a background one-body term as a means to provide overall charge neutrality when the two-body potential is long-ranged had been employed to study numerically the one-component plasma [65,66] as well as the Dyson log gas [67]. One can combine the one- and two-body potentials to rewrite Eq. (47) as

$$\Phi(\mathbf{r}^N; \mathbf{a}) = \sum_{i<j}^{N} \left(v(r_{ij}; \mathbf{a}) - \frac{1}{V} \int_\Omega v(x; \mathbf{a}) \, d\mathbf{x} \right)$$

$$= \sum_{i<j}^{N} v_e(r_{ij}; \mathbf{a}). \qquad (49)$$

The summand, which we denote by $v_e(\mathbf{r}; \mathbf{a})$, can be regarded as an effective pair potential which is $v(r; \mathbf{a})$ "screened" by the background.

In order to obtain accurate pair statistics for such long-ranged $v(r; \mathbf{a})$ in the thermodynamic limit, one must simulate the total configurational energy corresponding to an infinitely large system in $\mathbb{R}^d$ that does not impose a cutoff on $v(r; \mathbf{a})$. For this purpose, we set Ω to be a hypercubic simulation box of side length L under periodic boundary conditions and consider its infinite periodic images. To compute the total configurational potential energy for this infinite system, we note that the effective interaction between particle i and all images of particle j is given by

$$v_{e,\text{PBC}}(\mathbf{r}_{ij}; \mathbf{a}) = \sum_{\mathbf{n}} v(\mathbf{r}_{ij} + \mathbf{n}L; \mathbf{a}) - \frac{1}{L^d} \int_{\mathbb{R}^d} v(\mathbf{r}_{ij}; \mathbf{a}) \, d\mathbf{x}, \qquad (50)$$

where $\mathbf{n}$ are sites of the hypercubic lattice $\mathbb{Z}^d$. In going from (49) to (50), we have used the fact that $V = L^d$ and that the integration over the volume elements in all images of the region Ω is equivalent to a volume integral over $\mathbb{R}^d$. Note that due to the summation over lattice sites in Eq. (50), $v_{e,\text{PBC}}(\mathbf{r}_{ij}; \mathbf{a})$ is vector-dependent (anisotropic), even if $v_e(r_{ij}; \mathbf{a})$ is isotropic.

Thus, to find one- and two-body potentials corresponding to hyperuniform target pair statistics, the inverse procedure follows exactly the same steps as those described in Secs. IV B–IV C, except that in the optimization stage (Sec. IV C), the Monte Carlo simulations in steps 2 and 6 are performed under $v_{e,\text{PBC}}(\mathbf{r}_{ij}, \mathbf{a})$, instead of $v(r; \mathbf{a})$. In practice, Eq. (50) can be converted to an absolutely convergent integral and efficiently evaluated using the Ewald summation technique [65,68]. The specific formula for the absolutely convergent integral depends on d and α and the formula in the case $d = 3, \alpha = 2$ will be given in Sec. V D, where we apply our inverse procedure on the 3D cloaked URL.

E. Extension of the methodology to homogeneous anisotropic systems

Our methodology naturally extends to cases where the target system is anisotropic, e.g., nematic liquid crystals. The anisotropy of the system is reflected in the directional dependence of the target pair statistics $g_{2,T}(\mathbf{r})$ and $S_T(\mathbf{k})$. One could compute the anisotropic direct correlation function $c_T(\mathbf{r})$ via the Ornstein-Zernike equation (23) that specifies the large-$|\mathbf{r}|$ behavior of the $v(\mathbf{r};\mathbf{a})$ via $\beta v(\mathbf{r}) \sim -c(\mathbf{r})$ along the relevant directions, which are a set of discretized orientations of $\mathbf{r}$ that are not related by symmetry operations. Note that oftentimes $c(\mathbf{r})$ has orientational symmetries, which reduces the number of relevant directions. Similarly, an initial guess of the small- and intermediate-$|\mathbf{r}|$ behaviors of $v(\mathbf{r};\mathbf{a})$ can be obtained via a numeric fit of the HNC approximation (44).

We note that when choosing the basis functions, one should ensure that all basis functions have the same symmetry as $c(\mathbf{r})$. For example, one could choose the basis functions to be a product of a radial function and an angular function:

$$f_j(\mathbf{r};a_j) = X_j(r;a_{j,X})\Theta_j(\mathbf{u};a_{j,\Theta}), \qquad (51)$$

where $a_{j,X}$ and $a_{j,\Theta}$ are parameter vectors for the radial function and the angular function, respectively, and $a_j = \{a_{j,X}, a_{j,\Theta}\}$. The orientation vector $\mathbf{u}$ is the unit vector in the direction of $\mathbf{r}$. The radial function $X_j(r;\mathbf{a}_{j,X})$ is selected from the general forms (39)–(43) and the angular function $\Theta_j(\mathbf{u};a_{j,\Theta})$ has the orientational symmetry of $c(\mathbf{r})$. For example, one can choose $\Theta_j(\mathbf{u};a_{j,\Theta})$ to be Chebyshev polynomials and spherical harmonics for $d = 2$ and $d = 3$, respectively. The same multistage optimization procedure described in Sec. IV C is then applied to minimize $\Psi(\mathbf{r};\mathbf{a})$, except that the Monte Carlo simulations in steps 2 and 6 now involve anisotropic $g_2(\mathbf{r};\mathbf{a})$ and $S(\mathbf{k};\mathbf{a})$.

V. APPLICATIONS OF THE INVERSE METHODOLOGY TO A DIVERSE SET OF TARGET SYSTEMS

In this section, we present the results of the application of our inverse methodology for various target systems in two and three dimensions, including a 3D Dzugutov liquid [39], a 2D Lennard-Jones fluid in the vicinity of the critical point, as well as pair statistics corresponding to 2D nonequilibrium RSA packing process [40,69,70] and a hyperuniform 3D nonequilibrium cloaked URL [42]. For such diverse target pair statistics, we show that our methodology yields lower values of the L_2-norm error $\mathcal{E}$, defined by (1), than those derived from currently available inverse procedures, including the IBI and IHCNI.

To numerically determine target and trial pair statistics, we performed Monte Carlo simulations to generate ensembles of configurations equilibrated under the corresponding potentials. Our simulations used square or cubic boxes under periodic boundary conditions with $N = 500$ particles. We averaged simulated pair statistics for 1000 configurations to obtain $g_2(r;\mathbf{a})$ and $S(k;\mathbf{a})$. In all cases, we chose the convergence criterion for Ψ to be $\epsilon = 0.002$. We include in the Appendix the implementation details of the methodology in each case, including the specific procedure of choosing the basis functions, as well as the values chosen for

the parameters $\sigma_{g_2}, \sigma_S, A, \lambda, \delta a$, and ϵ' in the optimization algorithm.

A. Lennard-Jones fluid in the vicinity of the critical point

To apply our inverse methodology to systems in the vicinity of the liquid-gas critical point, which is nonhyperuniform, we generated a 2D systems equilibrated under the standard Lennard-Jones 6-12 potential

$$v_{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad (52)$$

and then extracted the corresponding targeted pair statistics $g_{2,T}(r)$ and $S_T(k)$. The distance, density and temperature are made dimensionless as r/σ, $\rho\sigma^2$ and $k_B T/\varepsilon$. Henceforth, we take $\varepsilon = 1$, $\sigma = 1$ and $k_B = 1$. We call $v_{\mathrm{LJ}}(r)$ the *target-generating* potential, which we know is unique according to Henderson's theorem [34]. Importantly, $v_{\mathrm{LJ}}(r)$ was only used to generate the target pair statistics. When applying the inverse algorithm, we treat the target pair statistics simply as given, without regard to the target-generating potential, and hence do not assume the functional form of the potential. On the contrary, the initial guess of the basis functions of any potential is obtained via statistical-mechanical theory, and these are subject to reselection during the optimization procedure; see the Appendix for details. The density and temperature chosen for the target system are $\rho = \rho_c = 0.37$ and $T = 1.1T_c = 0.55$ [71,72]. A snapshot of the target system is shown in Fig. 1(a). This state point was chosen so that the critical scaling of the total correlation function (21) could be observed. We computed the correlation length to be $\xi = 4.83$ and fitted the Lorentzian form (20) for the structure factor $S_T(k)$ in the range $0 < k < 0.5$ to find $S_T(0) = 12.2$, which is clearly nonhyperuniform. We further remark that the effect of critical slowing down [51] is significant for simulations at this state point, requiring a relaxation time is two orders of magnitude larger than that for a typical dense Lennard-Jones liquid.

To apply our methodology described in Sec. IV, we first accurately obtained the small-k behavior for $\tilde{c}_T(k)$ and $S_T(k)$ via the Ornstein-Zernike equation (23) as described in Sec. III, and then performed inverse Fourier transform (8) to determine the large-r behavior of $c_T(r)$, which we found to be an inverse power law $-c_T(r)/\beta \sim v(r) \sim -Er^p$ [Eq. (25)]. Assuming that p takes integer values, we fitted $\tilde{c}_T(k)$ with the form of Eq. (26), but with k^ζ replaced by $\ln(k)k^\zeta$ when ζ is even, from which we robustly determined that $p = 6$, implying $\zeta = 4$. To determine the form of the small-r behavior of $v(r;\mathbf{a})$, we observed that $v_{\mathrm{HNC}}(r)$ is strongly repulsive at small r, which is consistent with the inverse power-law form (43) with $a_j^{(1)} > 0$. Therefore, the initial form of the pair potential was chosen to be

$$v(r;\mathbf{a}) = \varepsilon_1 r^{-p_1} - \varepsilon_2 r^{-6}. \qquad (53)$$

Note that although the form (53) is a generalized LJ potential, it was determined via statistical-mechanical theory from the target pair statistics alone, without knowledge of the functional form of the target-generating potential. We verified that $v(r;\mathbf{a})$ satisfies the condition $h_T^2(r) \ll |\beta v(r;\mathbf{a})|$ for large r, which is required to apply the asymptotic formula (19). The
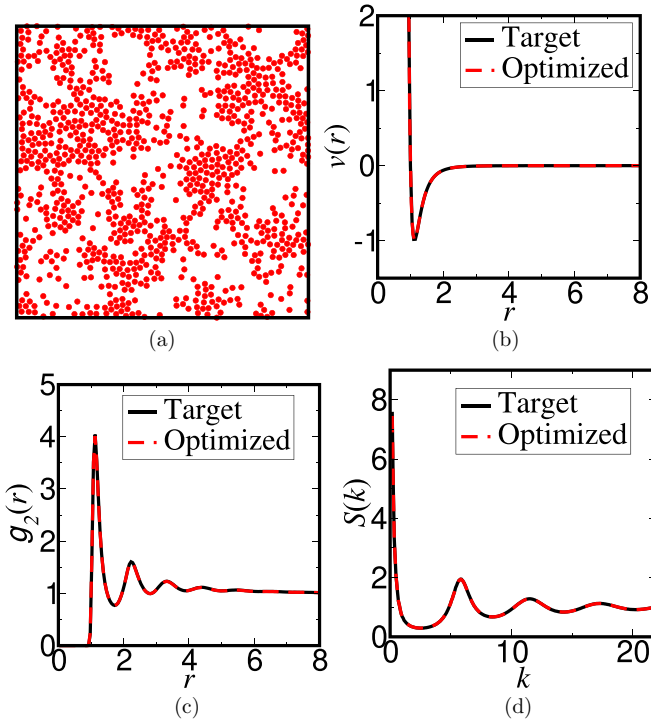
FIG. 3. (a) A snapshot of a 2D configuration 1000-particle system that is equilibrated under the optimized potential [Eq. (53)] for the target LJ fluid in the vicinity of its critical point. (b) The target-generating potential $v_{LJ}(r)$ and the optimized pair potential $v_F(r; \mathbf{a})$. (c) Targeted and optimized pair correlation functions. (d) Targeted and optimized structure factors.

reader is referred to the Appendix for additional implementation details. Minimization of Ψ using the form (53) yielded $\Psi < 0.002$ within one single stage of optimization, i.e., no reselection of the basis function was needed. The optimized parameters are given by $\varepsilon_1 = 3.98$, $p_1 = 11.93$, $\varepsilon_2 = 4.00$, which is in excellent agreement with the target-generating Lennard-Jones potential (52), as shown in Fig. 3(b). One can see that the configuration of the optimized equilibrium system [Fig. 3(a)] resembles that of the near-critical target [Fig. 1(a)] and exhibits an expected large correlation length. As anticipated, there is excellent agreement between the target and optimized pair statistics in both direct and Fourier space [Figs. 3(c)–3(d)]. The individual errors, defined by (2) and (3), are given by $D_{g_2} = 6.0 \times 10^{-4}$ and $D_S = 5.7 \times 10^{-4}$. The total L_2-norm error, defined by (1), is $\mathcal{E} = 0.034$.

We also applied the IBI and the IHNCI procedures to the same target pair statistics. The trial potentials in the IBI procedure failed to converge and gave liquid-gas phase separated states. The IHNCI procedure yielded a potential that is similar to our optimized potential but only in the small-r range $0 \leqslant r \leqslant 1$. However, it does not very accurately capture the large-r behavior of $v_{LJ}(r)$. The errors for the IHNCI procedure are $D_{g_2} = 0.0074$, $D_S = 0.011$, $\mathcal{E} = 0.13$. Note that $\mathcal{E}$ obtained

via IHNCI is an order of magnitude larger than that obtained via our method due to the inaccuracy in the large-r behavior of $v(r)$ as well as the accumulation of simulation errors in the binned potential. We also applied our methodology to LJ fluids at other state points, and find that for systems with long-ranged $g_2(r)$, including dense liquids near freezing, $\mathcal{E}$ obtained via our method is generally an order of magnitude lower than than that obtained via IHNCI. Therefore, our methodology appears to be superior to previous methods in solving these inverse problems for fluids in the critical region and dense liquids.

B. Liquid under the Dzugutov potential

To demonstrate that our inverse methodology is capable of treating systems equilibrated under nonstandard complex interactions, we study a 3D liquid under the Dzugutov potential, which is believed to be a good glass former [39,73,74]. Following Ref. [73], the form of the potential $v_D(r)$ is given by

$$
\begin{aligned}
\frac{v_D(r)}{\varepsilon} &= \frac{A}{\varepsilon}\left[\left(\frac{r}{\sigma}\right)^{-m} - \frac{B}{\varepsilon}\right] \exp\left(\frac{c}{r-a}\right) H\left(\frac{a-r}{\sigma}\right) \\
&\quad + \frac{B}{\varepsilon} \exp\left(\frac{d}{r-b}\right) H\left(\frac{b-r}{\sigma}\right),
\end{aligned}
\tag{54}
$$

where $H(x)$ is the Heaviside step function, A, B are energy parameters in units of ε, m is a dimensionless exponent, a, b, c, d are distance parameters in units of σ. The distance, density, and temperature are made dimensionless as r/σ, $\rho\sigma^3$ and $k_B T/\varepsilon$. Henceforth, we take $\varepsilon = 1$, $\sigma = 1$ and $k_B = 1$. The values of A, B, m, a, b, c, d chosen for this study are given in the Appendix. Figure 4(b) plots the form of the potential, from which it is clear that $v_D(r)$ is piecewise smooth, with oscillations at small r and a cutoff at which the potential exhibits a discontinuous first derivative at the cutoff distance $r = b$. We generated a 3D systems equilibrated under $v_D(r)$ and then extracted the corresponding targeted pair statistics $g_{2,T}(r)$ and $S_T(k)$. The density and temperature chosen for the target system are $\rho = 0.5$ and $T = 1$, which is a dense liquid state. Again, as in Sec. V A, we do not assume the functional form (54) when applying the inverse algorithm, and the only input data to the algorithm are ρ, T, $g_{2,T}(r)$, and $S_T(k)$.

To apply our methodology described in Sec. IV, we first accurately obtained the small-k behavior for $\tilde{c}_T(k)$ from $S(k)$ via the Ornstein-Zernike equation (23) and performed its inverse Fourier transform (8), from which we determined the large-r behavior of $c_T(r) \sim -\beta v(r)$ has an effectively superexponential decay. To obtain an initial guess of $v(r; \mathbf{a})$ for small- and intermediate-r, we observe that $v_{HNC}(r)$ has a strong repulsion for $r < 1$ and is oscillatory in the range $1 < r < 2$. We fitted $v_{HNC}(r)$ with the forms (40)–(42) and found the best fit of $v_{HNC}(r)$ is given by the exponential-damped oscillatory form (41). Thus, we started with an initial $v_(r; \mathbf{a})$ that is a sum of a power-law function and two exponential-damped basis functions. After one iteration of reselecting basis functions,

the final form of the optimized pair potential is given by

$$\frac{v_F(r;\mathbf{a})}{\varepsilon} = \begin{cases} \sum_{i=1}^{2} \frac{\varepsilon_i}{r^{p_i}} + \varepsilon_3 \exp\left[-\left(\frac{r}{\sigma_3^{(1)}}\right)^2\right] \cos\left(\frac{r}{\sigma_3^{(2)}} + \theta_3\right) + \varepsilon_4 \exp\left[-\left(\frac{r}{\sigma_4^{(1)}}\right)^4\right] \cos\left(\frac{r}{\sigma_4^{(2)}} + \theta_4\right) \\ \quad + \sum_{i=5}^{7} \varepsilon_i \cos\left(\frac{r}{\sigma_i} + \theta_i\right), & r \leqslant 2 \\ \sum_{i=1}^{2} \frac{\varepsilon_i}{r^{p_i}} + \varepsilon_3 \exp\left[-\left(\frac{r}{\sigma_3^{(1)}}\right)^2\right] \cos\left(\frac{r}{\sigma_3^{(2)}} + \theta_3\right) + \varepsilon_4 \exp\left[-\left(\frac{r}{\sigma_4^{(1)}}\right)^4\right] \cos\left(\frac{r}{\sigma_4^{(2)}} + \theta_4\right), & r > 2, \end{cases} \tag{55}$$

where the undamped oscillatory functions f_5, f_6, f_7 resulted from the reselection of the basis functions, which fine-tunes $v(r;\mathbf{a})$ in the range $0 \leqslant r \leqslant 2$; see the Appendix for details. Note that while the functional form (55) is different from that of the target-generating potential (54), this is expected, because the latter is not known *a priori* in application of our algorithm. However, Fig. 4(b) shows that the plots for $v_F(v;\mathbf{a})$ and $v_D(r)$ closely agree with one another, indicating that our methodology has precisely recovered the target-generating potential. We verified that $v_F(r;\mathbf{a})$ satisfies the condition $h_T^2(r) \ll |\beta v(r;\mathbf{a})|$ for large r, which is required to apply the asymptotic formula (19). The reader is refereed to the Appendix for additional implementation details. Figures 1(a) and 4(a) show that configurations of the target and optimized systems closely resemble each other. There is also excellent agreement between the target and optimized pair statistics in both direct and Fourier space [Figs. 4(c)–4(d)]. The individual errors, defined by (2) and (3), are given by $D_{g_2} = 0.0010$

and $D_S = 0.0010$. The total L_2-norm error, defined by (1), is $\mathcal{E} = 0.045$. On the other hand, the errors for the IHNCI procedure is $D_{g_2} = 0.0040$, $D_S = 0.0049$, $\mathcal{E} = 0.094$. Note that $\mathcal{E}$ obtained via IHNCI about twice of that obtained via our method due to the accumulation of simulation errors in the binned potential.

C. Equilibrium system corresponding to nonequilibrium random sequential addition pair statistics

Our methodology provides a powerful means to test the Zhang-Torquato conjecture [29] by determining whether classical equilibrium systems under an effective pair interaction can be attained that accurately match the pair statistics of nonequilibrium systems. Probing systems with identical pair statistics but different higher-body statistics will also shed light on the degeneracy problem [43,44].

Here we choose our nonequilibrium target system to be the random sequential addition (RSA) packing process [40,41,69,70,75], which has important applications in protein and polymer deposition models. Starting with an empty but large volume Ω in $\mathbb{R}^d$, the RSA process is produced by randomly, irreversibly, and sequentially placing nonoverlapping (hard) spheres of diameter D into this volume under periodic boundary conditions subject to a nonoverlap constraint. If a new sphere does not overlap with any existing spheres at some time t, it is added to the configuration; otherwise, the attempt is discarded. One can stop the addition process at any t, obtaining RSA configurations with a range of packing fractions $\phi(t)$ up to the infinite-time maximal saturation value $\phi_s = \phi(\infty)$. For identical spheres in 2D, $\phi_s \approx 0.547$ [40,41,70].

It is known that $g_2(r)$'s for RSA processes at the saturation limit possess a logarithmic divergence as $r \to D^+$, independent of dimension [76]. This suggests that the pair potential corresponding to a saturated RSA packing, if it exists, also possesses a corresponding singularity. Therefore, we choose as our target a 2D RSA packing very near, but not exactly at, saturation, with $\phi = 0.534 = 0.976\phi_s$, which avoids treating the aforementioned singularity, but is still a challenging target. We set D to be unity to generate the target RSA system. A snapshot of the target system is given in Fig. 5(a).

The effective pair potential $v(r;\mathbf{a})$ must have a hard core [Eq. (39)] for $r \leqslant 1$ to respect the nonoverlap constraint. To determine the large-r behavior of $v(r;\mathbf{a})$, we used the asymptotic analysis of $\tilde{c}_T(k)$ described in Sec. III and found that that $\beta c_T(r)$ has a superexponential decay rate that is best described by the a Gamma-function form (40). For intermediate values of r, we observed that the HNC approximation contains a minimum at the contact radius $r = 1$ and exhibits apparent oscillations about the horizontal axis. Therefore, starting from
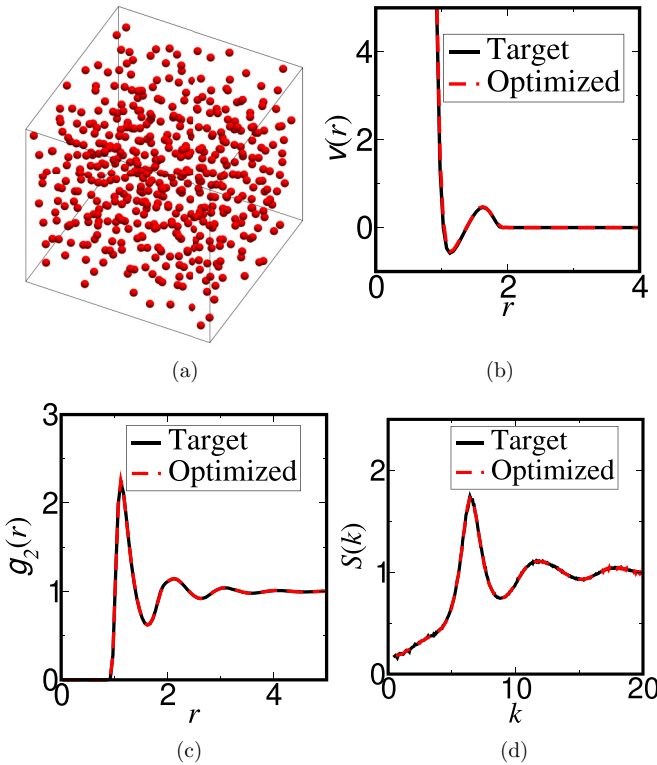

FIG. 4. (a) A snapshot of a 3D configuration 512-particle system that is equilibrated under the optimized potential [Eq. (55)] for the Dzugutov liquid. (b) The target-generating potential $v_D(r)$ (54) and the optimized pair potential (55). (c) Targeted and optimized pair correlation functions. (d) Targeted and optimized structure factors.
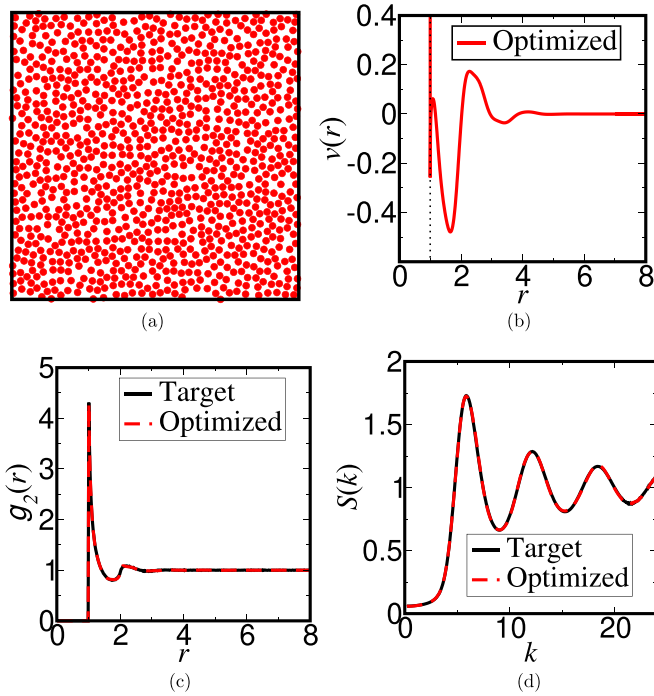
FIG. 5. (a) A snapshot a 2D configuration 1000-particle system that is equilibrated under the optimized potential [Eq. (56)] for the target nonequilibrium RSA packing near its saturation state. (b) Optimized pair potential. (c) Targeted and optimized pair correlation functions. (d) Targeted and optimized structure factors.

an initial potential that is a sum of a hard core, a nonoscillatory Gamma-form function (for the minimum at contact), and two Gamma-damped oscillatory functions, we went through four iterations of optimization and rechoosing basis functions and arrived at the following final form of the parameterized pair potential:

$$v_F(r; \mathbf{a}) = \begin{cases} \infty, & r \leqslant 1 \\ \varepsilon\left[-\sum_{j=1}^{2} \frac{\varepsilon_j}{\Gamma(r/\sigma_j)} + \frac{\sum_{j=3}^{7} \varepsilon_j \cos\left(r/\sigma_j^{(1)} + \theta_j\right)}{\Gamma\left(r/\sigma_3^{(2)}\right)} \right], & r > 1, \end{cases}$$
(56)

which satisfies the condition $h_T^2(r) \ll |\beta v_F(r; \mathbf{a})|$ for large r, which is required to apply the asymptotic formula (19). In our simulations, we let both the energy scale ε and the dimensionless temperature $k_B T/\varepsilon$ to be unity. The optimized parameters are provided in Table IV of the Appendix, which includes other implementations details. First, observe that the configuration of the optimized system [Fig. 5(a)] is visually very similar to the targeted nonequilibrium configuration in Fig. 1(c). Figure 5(b) shows the optimized potential, which possesses a sharp well at the hard-sphere diameter $r = 1$, and another broader well at $r = 1.66$. As shown in Figs. 5(c)–5(d), the optimized pair potential gives pair statistics that agree well with the target nonequilibrium pair statistics. The individual errors are $D_{g_2} = 0.0023$ and $D_S = 0.0015$ and the total L_2-norm error is $\mathcal{E} = 0.062$. This provides a vivid example that demonstrably adds to the growing evidence of the validity of the Zhang-Torquato conjecture [29].

By contrast, the individual errors resulting from the IHNCI procedure are $D_{g_2} = 0.0033$ and $D_S = 0.0060$, while the total L_2-norm errors is $\mathcal{E} = 0.096$, which are slightly larger than those from our method due to the accumulation of random errors. The fact that the IHNCI is relatively accurate for this target is due to the superexponential decay of $g_{2,T}(r)$, which translates to a fast decaying pair potential $v(r)$ via the HNC approximation.

D. Equilibrium system corresponding to nonequilibrium 3D cloaked URL system

In this subsection, we apply our methodology to the challenging task of determining an equilibrium system corresponding to the pair statistics of a nonequilibrium hyperuniform *cloaked uniform randomized lattice (URL) model* [42]. In URL models, each point in a d-dimensional simple cubic lattice $\mathcal{L} = \mathbb{Z}^d$ is displaced by a random vector that is uniformly distributed on a rescaled unit cell $bC = [-b/2, b/2]^d$, where $b > 0$ is a scalar factor and C is a unit cell of the lattice. By definition, the lattice constant is set to be unity. It has been shown that the structure factor for the URL point process contains Bragg peaks that coincide with the unperturbed lattice as well as a diffuse part such that $\lim_{k \to 0} S(k) \sim k^2$ [42,77,78]. Remarkably, Klatt *et al.* showed that the Bragg peaks in the structure factors vanish completely, or become "cloaked," when b takes integer values [42].

We applied our methodology to determine a positive-temperature equilibrium system with one- and two-body interactions that realizes the pair statistics of a 3D cloaked URL model with $\rho = 1$, $b = 1$. Such a classical system must be thermodynamically incompressible; see Sec. II C. The fact that the target system is hyperuniform with $\alpha = 2$ implies that the effective pair potential has the asymptotic form $v(r) \sim 1/r$; see Eq. (30). After fitting the HNC approximation with Eqs. (40)–(43) for the small-r range $0 \leqslant r \leqslant 2$, we determined the initial form of the pair potential to be

$$v(r; \mathbf{a}) = \varepsilon[v_l(r; \mathbf{a}_l) + v_s(r; \mathbf{a}_s)], \quad (57)$$

where

$$v_l(r; \mathbf{a}_l) = \frac{\varepsilon_1}{r} \quad (58)$$

and

$$v_s(r; \mathbf{a}_s) = -\frac{\varepsilon_2 \exp(-r/\sigma_2)}{r^{p_2}} + \varepsilon_3 \exp\left[-\left(\frac{r - \sigma_3^{(1)}}{\sigma_3^{(2)}}\right)^2 \right]$$
(59)

are the long-ranged and short-ranged parts of $v(r; \mathbf{a})$, respectively; see the Appendix for details. We verified that Eq. (57) satisfies the condition $h_T^2(r) \ll |\beta v(r; \mathbf{a})|$ for large r to apply (19). Henceforth, we set both the energy scale ε and the dimensionless temperature $k_B T/\varepsilon$ to be unity. We reiterate that $v_l(r; \mathbf{a}_l)$ can be regarded as a Coulombic interaction between "like-charged" particles. The one-body potential was treated using the procedure described in Sec. IV D. Under periodic boundary conditions with a cubic simulation box of edge length L, the effective interaction between particle i and all

images of particle j (50) is given by

$$v_{e,\text{PBC}}(\mathbf{r}_{ij};\mathbf{a}) = \sum_{\mathbf{n}} \frac{\varepsilon_1}{|\mathbf{r}_{ij}+\mathbf{n}L|} - \frac{\varepsilon_1}{L^3}\int_{\mathbb{R}^3}\frac{1}{x}\,d\mathbf{x} + \sum_{\mathbf{n}} v_s(|\mathbf{r}_{ij}+\mathbf{n}L|;\mathbf{a}_s) - \frac{1}{L^3}\int_{\mathbb{R}^3} v_s(x;\mathbf{a}_s)\,d\mathbf{x}, \tag{60}$$

where $\mathbf{n}$ are sites of the cubic lattice $\mathbb{Z}^3$. Since $v_s(r;\mathbf{a}_s)$ is short-ranged, the third term in (60) is dominated by the minimum image of j and the fourth term of (60) vanishes in the limit of large L. Therefore,

$$v_{e,\text{PBC}}(\mathbf{r}_{ij};\mathbf{a}) = \sum_{\mathbf{n}} \frac{\varepsilon_1}{|\mathbf{r}_{ij}+\mathbf{n}L|} - \frac{\varepsilon_1}{L^3}\int_{\mathbb{R}^3}\frac{1}{x}\,d\mathbf{x} + v_s(r_{ij}), \quad L \to \infty. \tag{61}$$

The first two terms of (61), which are integrals over long-ranged potentials, can be converted to an absolutely convergent integral and efficiently evaluated using the Ewald summation technique [65,68]:

$$\sum_{\mathbf{n}} \frac{\varepsilon_1}{|\mathbf{r}_{ij}+\mathbf{n}L|} - \frac{\varepsilon_1}{L^3}\int_{\mathbb{R}^3}\frac{1}{x}\,d\mathbf{x} = \sum_{\mathbf{n}} \frac{\varepsilon_1}{|\mathbf{r}_{ij}+\mathbf{n}L|} - \frac{\varepsilon_1}{L^3}\int_{\mathbb{R}^3}\frac{1}{|\mathbf{r}_{ij}+\mathbf{x}|}\,d\mathbf{x}$$

$$= \frac{\varepsilon_1}{L}\left(\sum_{\mathbf{n}} \frac{\text{erfc}(\pi^{1/2}|\mathbf{r}_{ij}/L+\mathbf{n}|)}{|\mathbf{r}_{ij}/L+\mathbf{n}|} - 1 + \sum_{\mathbf{n}\neq\mathbf{0}} \frac{1}{\pi n^2}\exp(-\pi n^2)\exp[2i\pi\mathbf{n}(\mathbf{r}_{ij}/L)]\right), \tag{62}$$

where $n = |\mathbf{n}|$. Thus,

$$v_{e,\text{PBC}}(\mathbf{r}_{ij};\mathbf{a}) = \frac{\varepsilon_1}{L}\left(\sum_{\mathbf{n}} \frac{\text{erfc}(\pi^{1/2}|\mathbf{r}_{ij}/L+\mathbf{n}|)}{|\mathbf{r}_{ij}/L+\mathbf{n}|} - 1 + \sum_{\mathbf{n}\neq\mathbf{0}} \frac{1}{\pi\lambda^2}\exp(-\pi\lambda^2)\exp[2i\pi\mathbf{n}(\mathbf{r}_{ij}/L)]\right)$$

$$- \frac{\varepsilon_2\exp(-r/\sigma_2)}{r^{p_2}} + \varepsilon_3\exp\left[-\left(\frac{r-\sigma_3^{(1)}}{\sigma_3^{(2)}}\right)^2\right], \quad L \to \infty, \tag{63}$$

which is the effective potential that we used in Monte Carlo simulations under periodic boundary conditions.

The form of the parameterized potential (57) achieved the desired convergence criterion $\Psi < 0.002$ in one single optimization stage, i.e., no reselection of basis function was needed. The optimized parameters are listed in Table V of the Appendix. Figures 1(d) and 6(a) show the snapshots of the target and the optimized systems, respectively, which are visually indistinguishable. Figure 6(b) plots the short-ranged part of the effective potential, i.e., $v_s(r;\mathbf{a}_s) = v(r;\mathbf{a}) - 0.940/r$ against r. Figures 6(c) and 6(d) show $g_2(r)$ and $S(k)$, respectively, for the target and optimized systems. The pair statistics of the target cloaked URL is in excellent agreement with those of the optimized equilibrium system in both direct and Fourier space. The individual errors are $D_{g_2} = 4.8 \times 10^{-4}$ and $D_S = 7.5 \times 10^{-4}$ and the L_2-norm error is $\mathcal{E} = 0.035$, all of which are remarkably relatively small.

Importantly, the IBI and IHNCI procedures cannot treat hyperuniform targets. First, they do not include one-body potentials. Second, they impose a cutoff on the extent of the pair potential, which is not appropriate for the required long-ranged interactions.

VI. CONCLUSIONS AND DISCUSSION

We have formulated an optimization algorithm to find effective one- and two-body potentials with high precision that correspond to pair statistics of general translationally invariant disordered many-body equilibrium or nonequilibrium systems at positive temperatures. The versatility and power of our inverse methodology to accurately extract the effective pair interactions was demonstrated by consider-

ing four diverse target systems: (1) 3D liquid under the Dzugutov potential, (2) 2D Lennard-Jones system in the vicinity of its critical point, (3) 2D nonequilibrium RSA packing, and (4) a 3D nonequilibrium hyperuniform "cloaked" URL. We showed that the optimized pair potentials generate corresponding pair statistics that accurately match their corresponding targets with total L_2-norm errors that are an order of magnitude smaller than that of previous methods. The results of our investigation lend further support to the Zhang-Torquato conjecture [29], which states that any realizable $g_2(\mathbf{r})$ or $S(\mathbf{k})$ corresponding to a translationally invariant nonequilibrium system can be attained by a classical equilibrium ensemble involving only (up to) effective pair interactions.

The capability of our procedure to precisely determine equilibrium systems corresponding to nonequilibrium pair statistics has important implications. First, our procedure provides an effective means to test the Zhang-Torquato conjecture for structures that span diverse hyperuniform and nonhyperuniform classes [29]. Second, the dynamics leading to a nonequilibrium system that has the same pair statistics as one drawn from an equilibrium ensemble must be reflected in differences in their respective higher-order statistics. Thus, such differences in the higher-order statistics is expected to provide a measure of the degree to which a nonequilibrium system is out of equilibrium relative to its equilibrium counterpart. Such nonequilibrium metrics will be studied in future work. Third, such investigations will enable one to probe systems with identical pair statistics but different higher-body statistics, which is expected to shed light on the well-known degeneracy problem of statistical mechanics [43,44,79]. Fourth, one
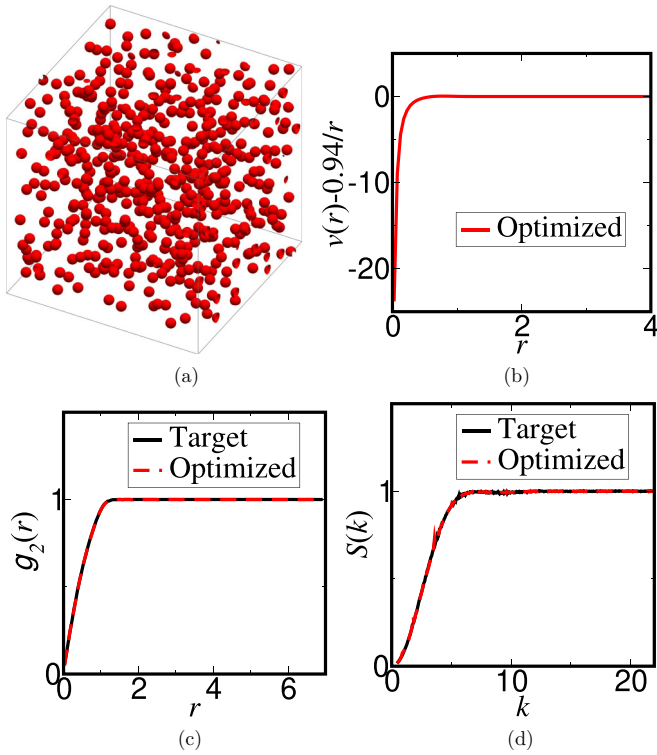
FIG. 6. (a) A portion of a 3D configuration of a 2744-particle system that is equilibrated under the optimized effective one- and two-body potential for the target 3D cloaked URL. Only 512 particles are displayed. (b) Optimized pair potential [Eq. (57)] minus its long-ranged repulsive part $0.940/r$. (c) Targeted and optimized pair correlation functions. (d) Targeted and optimized structure factors.

FIG. 7. Schematic illustration of a statistically homogeneous but anisotropic nematic liquid-crystal configuration.

could explore thermodynamic and dynamic properties of such effectively equivalent equilibrium systems, such as phase behaviors, excess entropies [7,80], and inherent structures [81], which are outstanding problems for future research. Fifth, structural properties of the effectively equivalent equilibrium states, such as nearest-neighbor probability distribution functions, percolation threshold and fluid permeability, enable one to infer these nontrivial attributes of the nonequilibrium states, which are crucial in determining mechanical and electronic properties of materials [82]. Finally, our study enables one to generate tunable nonhyperuniform and hyperuniform materials.

Another promising application of our methodology is to determine the interactions corresponding to numerically investigate the realizability problem, i.e., whether hypothetical functional forms of $g_2(r)$ or $S(k)$ at fixed ρ can be attained by many-particle configurations [27,29,83–86]. This is an outstanding problem in statistical physics, as it has been shown that certain functional forms for pair statistics may not be realizable, even if they meet all the explicitly known realizability conditions [83,85]. Our inverse procedure provides a powerful means to test whether there exist equilibrium many-body systems with up to pair interactions that realize hypothetical target pair statistics at all length scales.

The generalization of our methodology to anisotropic systems described in Sec. IV E can be applied to determine
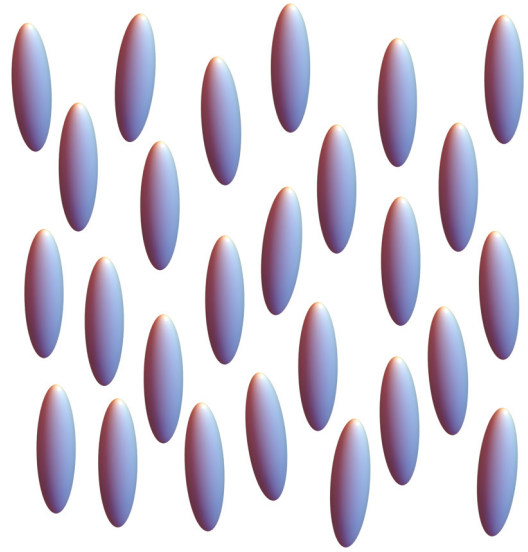
effective one- and two-body potentials that reproduce target pair statistics for liquid crystal phases. Such a concrete example includes the nematic phase of oriented ellipsoids shown in Fig. 7. In this particular case, the shape of the ellipsoids can be inferred from the range of repulsion of the anisotropic effective pair potential $v(\mathbf{r}; \mathbf{a})$. It is also straightforward to extend our methodology to inverse problems for statistically homogeneous multicomponent systems that consist of different species $A, B, \ldots$. Here the goal would be to find equilibrium systems with up to pair potentials that reproduce all the intraspecies and interspecies pair correlation functions $\{g_{AA}(\mathbf{r}), g_{AB}(\mathbf{r}), \ldots, g_{BB}(\mathbf{r}), \ldots\}$ as well as their corresponding structure factors. To accomplish this task, one must choose a different set of basis functions for each species-specific trial pair potential $\{v_{AA}(\mathbf{r}, \mathbf{a}_{AA}), v_{AB}(\mathbf{r}, \mathbf{a}_{AB}), \ldots, v_{BB}(\mathbf{r}; \mathbf{a}_{BB}), \ldots\}$ based on statistical-mechanical theory [87,88]. The algorithm described in Sec. IV C can be then applied to optimize each species-specific pair potential.

A substantially more challenging extension of our methodology is determine the interactions corresponding to targeted pair statistics derived from general statistically inhomogeneous system in which there is a preferred origin in the system, e.g., liquid-gas interfaces, which requires a position-dependent one-body potential $\phi(\mathbf{r})$ and a pair potential that depends on absolute positions $\phi_2(\mathbf{r}_1, \mathbf{r}_2)$. To treat such systems, one could apply the inhomogeneous Ornstein-Zernike equations [7] and the associated closures, such as the Lovett-Mou-Buff-Wertheim equation [89], to obtain an initial guess for the functional forms of both $\phi(\mathbf{r})$ and $\phi_2(\mathbf{r}_1, \mathbf{r}_2)$ [90]. However, it will be significantly more nontrivial to find appropriate pointwise functional forms for the basis functions and to perform the corresponding optimizations for inhomogeneous systems, and so represents an outstanding subject for future research.

We note that machine-learning approaches [91] could be used as an alternative to the BFGS technique for optimization of the potential parameters $\mathbf{a}$ to improve

TABLE I. Fits of $\tilde{c}_T(k)$ for the target near-critical LJ fluid in the range $0 \leqslant k \leqslant 4$.

Form of $v(r)$ at large r	Form of $\tilde{c}(k)$ at small k	m_0	m_2	m_4	m_6	m^*	BIC
$-r^{-5}$	$\sum_{i=0}^{3} m_{2i} k^{2i} + m^* k^3$	2.30	-5.64	-0.292	9.76×10^{-4}	2.44	-252
$-r^{-6}$	$\sum_{i=0}^{3} m_{2i} k^{2i} + m^* \ln(k) k^4$	2.21	-4.44	1.08	0.00862	-0.692	-248
$-r^{-7}$	$\sum_{i=0}^{3} m_{2i} k^{2i} + m^* k^5$	2.14	-4.02	1.08	0.0307	-0.333	-188

computational efficiency. More specifically, the training phase of machine-learning techniques, i.e., generating a sufficient number of [$v(r; \mathbf{a})$, $g_2(r; \mathbf{a})$ and $S(k; \mathbf{a})$] pairs for many different parameter supervectors $\mathbf{a}$, can be completely parallelized. Combining machine-learning techniques with our inverse algorithm is an outstanding problem for future research.

Finally, we remark that while the main focus of the present work concerns classical systems, our procedure allows one to also consider inverse problems for targeted pair statistics of quantum many-body systems, such as those involving boson or fermion statistics [92,93] or superfluidity as exhibited by liquid helium at low temperature [94,95]. One could start with some experimentally measured $g_2(\mathbf{r})$ or $S(\mathbf{k})$ for a quantum fluid and determine if there exist classical many-particle systems interacting with up to pair potentials that produce the same pair statistics, which is often the case. For instance, it is known that the pair statistics of the 1D free-fermion system are equivalent to those of a classical system on a unit circle under a logarithmic interaction at positive T, known as the circular unitary ensemble [96,97]. Another possible problem is to determine (at constant ρ and T) how pair potentials in an initially classical fluid must be modified to keep the same pair distribution as the particles reduce their mass to become increasingly quantized, either as fermions or bosons. Investigating inverse problems for quantum systems is a fascinating area for future research.

APPENDIX: IMPLEMENTATION DETAILS OF THE METHODOLOGY FOR EACH OF THE TARGETS

1. Lennard-Jones fluid in the vicinity of the critical point

To estimate the large-r behavior of $v(r)$, we first fitted the the direct correlation function of the target system $c_T(r)/\beta$ in the range $1.5 \leqslant r \leqslant 5$ with Eqs. (40)–(43), in which the oscillatory factors were eliminated as $c_T(r)$ is not oscillatory about the horizontal axis. The form with the lowest BIC was given by a power law $-c_T(r)/\beta \sim -E r^{-p}$, where $p = 5.9 \pm 1.2$. To more accurately estimate the large-r behavior of $v(r)$, $\tilde{c}_T(k)$ in the range $0 \leqslant k \leqslant 4$ was fitted with the expected small-k functional forms of $\tilde{c}(k)$ associated with $p = 5, 6, 7$, respectively. [We fitted $\tilde{c}_T(k)$ instead of $S_T(k)$,

as the former gave lower fit residuals.] Table I shows the results of the fits along with their corresponding BIC. Although the fit for $p = 5$ yielded the lowest BIC, its corresponding nonanalytic term in $\tilde{c}(k)$ is positive, which is inconsistent with the desired attractive large-r behavior of $v(r)$. Thus, the large-r asymptotic behavior of $v(r)$ is best described by $v(r) \sim -E/r^6$, where E is estimated to be 3.88 via a numeric fit using Eq. (26), but with k^ζ replaced by $\ln(k) k^\zeta$. To choose a functional form for the small-r behavior of $v(r)$, we observed that the HNC approximation in the range $0.9 \leqslant r \leqslant 1$ is strongly repulsive and can be best described by a power-law function (43). Since $v_{\text{HNC}}(r)$ is not oscillatory about the horizontal axis, we eliminated the oscillatory factor $\cos(r/\sigma_j^{(1)} + \theta_j)$. We found that $v_{\text{HNC}}(r)$ in the range $0.9 < r < 10$ can be well fitted by the parameterized trial potential (53), in which the initial values of the parameters are $\varepsilon_1 = 3.90$, $p_1 = 11.51$, $\varepsilon_2 = 3.88$.

In the optimization stage of the inverse procedure, we set the parameters defined in Sec. IV to be $\sigma_{g_2} = 4$, $\sigma_S = 2$, $A = 1$, $\lambda = 1$, $\delta a = 0.1$, $\epsilon = 0.002$, $\epsilon' = 0.02$. The criterion $\Psi < \epsilon$ is achieved within one single stage of optimization. The optimized $v(r; \mathbf{a})$ gives $\varepsilon_1 = 3.98$, $p_1 = 11.93$, $\varepsilon_2 = 4.00$.

2. Liquid under the Dzugutov potential

The target-generating Dzugutov potential is given by (54), where the parameters used in this work are listed in Table II. To apply our methodology, we first note that $S_T(k)$ in the range $0 \leqslant k \leqslant 1.5$ is fitted accurately by $0.169 + 0.0251 k^2 + 0 k^4$, which suggests $v(r)$ at large r has exponential or superexponential decay. We accurately obtained the small-k behavior for $\tilde{c}_T(k)$ and performed inverse Fourier transform (8), from which we determined that $-c_T(r)/\beta \sim v(r)$ has an effectively superexponential decay at large r.

To obtain an initial guess of $v(r; \mathbf{a})$ for small- and intermediate-r, we observe that $v_{\text{HNC}}(r)$ has a strong repulsion for $r < 1$ and is oscillatory in the range $1 < r < 2$. We fitted $v_{\text{HNC}}(r)$ in the range $0 \leqslant r \leqslant 5$ with the forms (40)–(42) and found the fit with the lowest BIC is given by the exponential-damped oscillatory form (41). Thus, we started with an initial $v_{(r; \mathbf{a})}$ that is a sum of a power-law function

TABLE II. Parameters of the target-generating Dzugutov potential [Eq. (54)].

A	5.92	b	1.94
B	1.28	c	1.1
m	16	d	0.27
a	1.87		

TABLE III. Optimized parameters of the pair potential (55) for the 3D Dzugutov liquid.

ε_1	3.743	θ_4	0.7713
p_1	11.108	ε_5	-1.072
ε_2	0.7342	σ_5	0.1619
p_2	16.641	θ_5	4.455
ε_3	45.126	ε_6	2.893
$\sigma_3^{(1)}$	0.7738	σ_6	0.2416
$\sigma_3^{(2)}$	0.1227	θ_6	1.349
θ_3	0.1501	ε_7	6.017
ε_4	24.704	σ_7	1.036
$\sigma_4^{(1)}$	0.1096	θ_7	3.193
$\sigma_4^{(2)}$	0.1686		

TABLE IV. Optimized parameters of the RSA effective pair potential (56).

ε_1	1.200×10^{62}	ϵ_5	0.5765
σ_1	0.0200	$\sigma_5^{(1)}$	0.223
ε_2	7.610	θ_5	0.5056
σ_2	0.214	ε_6	-0.2232
ε_3	0.05790	$\sigma_6^{(1)}$	0.213
$\sigma_3^{(1)}$	0.0912	θ_6	1.352
θ_3	1.918	ε_7	0.5458
ε_4	0.3947	$\sigma_7^{(1)}$	0.292
$\sigma_4^{(1)}$	0.190	θ_7	4.395
θ_4	1.329	$\sigma_3^{(2)}$	0.697

and two exponential-damped basis functions with $M = 2$ and 4, respectively. This combination of M values achieved the lowest BIC among all combinations:

$$v(r;\mathbf{a}) = \frac{\varepsilon_1}{r^{p_1}} + \varepsilon_2 \exp\left[-\left(\frac{r}{\sigma_2(1)}\right)^2\right]\cos\left(\frac{r}{\sigma_2^{(2)}} + \theta_2\right)$$
$$+ \varepsilon_3 \exp\left[-\left(\frac{r}{\sigma_3^{(1)}}\right)^4\right]\cos\left(\frac{r}{\sigma_3^{(2)}} + \theta_3\right). \quad \text{(A1)}$$

In the optimization stage of the inverse procedure, we used the parameters $\sigma_{g_2} = 4, \sigma_S = 2, A = 1, \lambda = 0.05, \delta a = 0.1, \epsilon = 0.002, \epsilon' = 0.02$. The optimization using the initial set of basis functions (A1) stalled at $\Psi = 0.011$, which is larger than the tolerance ϵ. Thus, we added basis functions via a fit of the difference between target and trial potentials of main force (46). The additional basis functions include another power-law function to accurately capture the repulsive interactions, as well as three undamped oscillatory functions in the range $0 \leqslant r \leqslant 2$ to capture the asymmetric peaks and valleys in $g_{2,T}(r)$ in that range. Upon re-optimization of the parameters, the potential form (55) achieved the desired error tolerance $\Psi < 0.002$. The optimized parameters are listed in Table III.

3. Equilibrium system corresponding to nonequilibrium random sequential addition pair statistics

To choose an initial set of basis functions, we first observe that the corresponding $\tilde{c}_T(k)$ in the range $0 \leqslant k \leqslant 3$ can be excellently fitted by the analytic function $-24.4 + 1.51k^2 + 0.00668k^4$. This implies that $v(r)$ at large r has exponential or superexponential decay. Thus, we fitted $v_{HNC}(r)$ in the range $1 < r < 8$ with Eqs. (40)–(42) and found that the fit with lowest BIC was achieved by a sum of two oscillatory Gamma functions and one nonoscillatory Gamma function. Our initial choice of the parameterized potential is given by

$$v(r;\mathbf{a}) = \begin{cases} \infty, & r \leqslant 1 \\ -\frac{\varepsilon_1}{\Gamma(r/\sigma_1)} + \frac{\sum_{j=2}^{3}\varepsilon_j \cos\left(r/\sigma_j^{(1)} + \theta_j\right)}{\Gamma(r/\sigma_2^{(2)})}, & r > 1. \end{cases} \quad \text{(A2)}$$

In the optimization stage of the inverse procedure, we used the parameters $\sigma_{g_2} = 4, \sigma_S = 2, A = 1, \lambda = 0.5, \delta a = 0.1, \epsilon = 0.002, \epsilon' = 0.02$. The optimization using the initial set of basis functions (A2) stalled at $\Psi = 0.01$, which is larger

than the tolerance ϵ. We observed that this was due to two reasons: (a) the optimized $g_2(1^+;\mathbf{a})$ was lower than the target $g_{2,T}(1^+)$, and (b) the maximum in $g_{2,T}(r)$ at $r = 2.2$ was not well reproduced: the optimized $g_2(r;\mathbf{a})$ had a maximum at $r = 2.05$ instead. In order to fix these discrepancies, in the second stage, we included in $v(r;\mathbf{a})$ one more oscillatory and one more nonoscillatory Gamma functions,

$$v(r;\mathbf{a}) = \begin{cases} \infty, & r \leqslant 1 \\ -\sum_{j=1}^{2}\frac{\varepsilon_j}{\Gamma(r/\sigma_j)} + \frac{\sum_{m=3}^{5}\varepsilon_j \cos\left(r/\sigma_j^{(1)} + \theta_j\right)}{\Gamma(r/\sigma_3^{(2)})}, & r > 1. \end{cases} \quad \text{(A3)}$$

Reoptimizing the parameters in Eq. (A3) still did not reach the desired convergence criterion, but this time solely due to the reason (b) described above. Therefore, we iteratively added Gamma-damped oscillatory functions and optimized the parameters, until $\Psi < 0.002$ was achieved. The final form of the parameterized pair potential is given by Eq. (56) and the optimized parameters are listed in Table IV.

4. Equilibrium system corresponding to nonequilibrium 3D cloaked URL system

The fact that the target is hyperuniform with $\alpha = 2$ implies that the pair potential has the asymptotic form $v(r) \sim 1/r$ [Eq. (30)]. To obtain an initial form of the small- and intermediate-r behaviors, we fitted $v_{HNC}(r)$ in the range $0 \leqslant r \leqslant 2$ with Eqs. (40)–(43), where the oscillatory factors were eliminated as $v_{HNC}(r)$ is not oscillatory about the horizontal axis. The form with lowest BIC was achieved by Eq. (57).

During the optimization step, we used the parameters $\sigma_{g_2} = 4, \sigma_S = 2, A = 1, \lambda = 2, \delta a = 0.05, \epsilon = 0.002$ and $\epsilon' = 0.02$. The criterion $\Psi < \epsilon$ was achieved in one single stage of optimization. The optimized parameters are listed in Table V.

TABLE V. Optimized parameters of the 3D cloaked URL effective pair potential (57).

ε_1	0.940	ε_3	0.0910
ε_2	1.57	$\sigma_3^{(1)}$	0.650
σ_2	0.195	$\sigma_3^{(2)}$	0.292
p_2	0.790		

[1] K. Van Workum and J. F. Douglas, Symmetry, equivalence, and molecular self-assembly, Phys. Rev. E **73**, 031502 (2006).

[2] M. C. Rechtsman, F. H. Stillinger, and S. Torquato, Designed isotropic potentials via inverse methods for self-assembly, Phys. Rev. E **73**, 011406 (2006); Erratum: Designed isotropic potentials via inverse methods for self-assembly [Phys. Rev. E 73, 011406 (2006)], **75**, 019902 (2007).

[3] S. Torquato, Inverse optimization techniques for targeted self-assembly, Soft Matter **5**, 1157 (2009).

[4] H. Cohn and A. Kumar, Optimality and uniqueness of the leech lattice among lattices, Ann. Math. **170**, 1003 (2009).

[5] D. Frenkel and J. F. Maguire, Molecular dynamics study of the dynamical properties of an assembly of infinitely thin hard rods, Mol. Phys. **49**, 503 (1983).

[6] F. H. Stillinger and T. A. Weber, Computer simulation of local order in condensed phases of silicon, Phys. Rev. B **31**, 5262 (1985).

[7] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, New York, 1986).

[8] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).

[9] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, New York, 1996).

[10] M. Dijkstra, R. Roij, and R. Evans, Direct Simulation of the Phase Behavior of Binary Hard-Sphere Mixtures: Test of the Depletion Potential Description, Phys. Rev. Lett. **82**, 117 (1999).

[11] S. K. Nath, F. A. Escobedo, and J. J. de Pablo, On the simulation of vapor–liquid equilibria for alkanes, J. Chem. Phys. **108**, 9905 (1998).

[12] M. Watzlawek C. N. Likos, and H. Lowën, Phase Diagram of Star Polymer Solutions, Phys. Rev. Lett. **82**, 5289 (1999).

[13] F. Romano, E. Sanz, and F. Sciortino, Crystallization of tetrahedral patchy particles *in silico*, J. Chem. Phys. **134**, 174502 (2011).

[14] I. Saika-Voivod, F. Smallenburg, and F. Sciortino, Understanding tetrahedral liquids through patchy colloids, J. Chem. Phys. **139**, 234901 (2013).

[15] A. A. Bertolazzo, A. Kumar, C. Chakravarty, and V. Molinero, Water-like anomalies and phase behavior of a pair potential that stabilizes diamond. J. Phys. Chem. B **120**, 1649 (2016).

[16] U. Buck, Inversion of molecular scattering data, Rev. Mod. Phys. **46**, 369 (1974).

[17] R. B. Jadrich, B. A. Lindquist, and T. M. Truskett, Probabilistic inverse design for self-assembling materials, J. Chem. Phys. **146**, 184103 (2017).

[18] Z. M. Sherman, M. P. Howard, B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Inverse methods for design of soft materials, J. Chem. Phys. **152**, 140902 (2020).

[19] M. C. Rechtsman, F. H. Stillinger, and S. Torquato, Optimized Interactions for Targeted Self-Assembly: Application to Honeycomb Lattice, Phys. Rev. Lett. **95**, 228301 (2005); Erratum: Optimized Interactions for Targeted Self-Assembly: Application to a Honeycomb Lattice [Phys. Rev. Lett. 95, 228301 (2005)], **97**, 239901 (2006).

[20] M. C. Rechtsman, F. H. Stillinger, and S. Torquato, Synthetic diamond and wurtzite structures self-assemble with isotropic pair interactions, Phys. Rev. E **75**, 031403 (2007).

[21] É. Marcotte, F. H. Stillinger, and S. Torquato, Designed diamond ground state via optimized isotropic monotonic pair potentials, J. Chem. Phys. **138**, 061101 (2013).

[22] G. Zhang, F. H. Stillinger, and S. Torquato, Probing the limitations of isotropic pair potentials to produce ground-state structural extremes via inverse statistical mechanics, Phys. Rev. E **88**, 042309 (2013).

[23] A. Jain, J. R. Errington, and T. M. Truskett, Inverse design of simple pairwise interactions with low-coordinated 3D lattice ground states, Soft Matter **9**, 3866 (2013).

[24] A. Jain, J. A. Bollinger, and T. M Truskett, Inverse methods for material design, AIChE J. **60**, 2732 (2014).

[25] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Communication: Inverse design for self-assembly via on-the-fly optimization, J. Chem. Phys. **145**, 111101 (2016).

[26] D. Chen, G. Zhang, and S. Torquato, Inverse design of colloidal crystals via optimized patchy interactions, J. Phys. Chem. B **122**, 8462 (2018).

[27] O. U. Uche, F. H. Stillinger, and S. Torquato, On the realizability of pair correlation functions, Physica A **360**, 21 (2006).

[28] T. Kuna, J. L. Lebowitz, and E. R. Speer, Necessary and sufficient conditions for realizability of point processes, Ann. Appl. Probab. **21**, 1253 (2011).

[29] G. Zhang and S. Torquato, Realizable hyperuniform and non-hyperuniform particle configurations with targeted spectral functions via effective pair interactions, Phys. Rev. E **101**, 032124 (2020).

[30] D. Levesque, J. J. Weis, and L. Reatto, Pair Interaction from Structural Data for Dense Classical Liquids, Phys. Rev. Lett. **54**, 451 (1985).

[31] A. P. Lyubartsev and A. Laaksonen, Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach, Phys. Rev. E **52**, 3730 (1995).

[32] A. K. Soper, Empirical potential Monte Carlo simulation of fluid structure, Chem. Phys. **202**, 295 (1996).

[33] M. Heinen, Calculating particle pair potentials from fluid-state pair correlations: Iterative Ornstein-Zernike inversion, J. Comput. Chem. **39**, 1531 (2018).

[34] R. L. Henderson, A uniqueness theorem for fluid pair correlation functions, Phys. Lett. A **49**, 197 (1974).

[35] H. Wang, F. H. Stillinger, and S. Torquato, Sensitivity of pair statistics on pair potentials in many-body systems, J. Chem. Phys. **153**, 124106 (2020).

[36] S. Torquato and F. H. Stillinger, Local density fluctuations, hyperuniform systems, and order metrics, Phys. Rev. E **68**, 041113 (2003).

[37] S. Torquato, Hyperuniform states of matter, Phys. Rep. **745**, 1 (2018).

[38] Since our study focuses on pure, single-component homogeneous systems, it excludes coexisting-phase states and systems with an interface.

[39] J. P. K. Doye, D. J. Wales, F. H. M. Zetterling, and M. Dzugutov, The favored cluster structures of model glass formers, J. Chem. Phys. **118**, 2792 (2003).

[40] J. Feder, Random sequential adsorption, J. Theor. Biol. **87**, 237 (1980).

[41] G. Zhang and S. Torquato, Precise algorithm to generate random sequential addition of hard hyperspheres at saturation, Phys. Rev. E **88**, 053312 (2013).

[42] M. A. Klatt, J. Kim, and S. Torquato, Cloaking the underlying long-range order of randomly perturbed lattices, Phys. Rev. E **101**, 032118 (2020).

[43] Y. Jiao, F. H. Stillinger, and S. Torquato, Geometrical ambiguity of pair statistics: Point configurations, Phys. Rev. E **81**, 011105 (2010).

[44] F. H. Stillinger and S. Torquato, Structural degeneracy in pair distance distributions, J. Chem. Phys. **150**, 204125 (2019).

[45] C. E. Zachary and S. Torquato, Hyperuniformity in point patterns and two-phase heterogeneous media, J. Stat. Mech.: Theory Exp. (2009) P12015.

[46] S. Torquato, Structural characterization of many-particle systems on approach to hyperuniform states, Phys. Rev. E **103**, 052126 (2021).

[47] B. Widom, Equation of state in the neighborhood of the critical point, J. Chem. Phys. **43**, 3898 (1965).

[48] L. P. Kadanoff, Scaling laws for Ising models near T_c, Phys. Phys. Fiz. **2**, 263 (1966).

[49] M. E. Fisher, The theory of equilibrium critical phenomena, Rep. Prog. Phys. **30**, 615 (1967).

[50] K. G. Wilson and J. Kogut, The renormalization group and the ϵ expansion, Phys. Rep. **12**, 75 (1974).

[51] J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Oxford University Press, Oxford, 1992).

[52] E. C. Oğuz, J. E. S. Socolar, P. J. Steinhardt, and S. Torquato, Hyperuniformity and anti-hyperuniformity in one-dimensional substitution tilings, Acta Cryst. A **A75**, 3 (2019).

[53] F. Martelli, S. Torquato, N. Giovambattista, and R. Car, Large-Scale Structure and Hyperuniformity of Amorphous Ices, Phys. Rev. Lett. **119**, 136002 (2017).

[54] G. Stell, Fluids with long-range forces: Toward a simple analytic theory, in *Statistical Mechanics, Part A*, edited by B. J. Berne (Plenum Press, New York, 1977), pp. 47–82.

[55] M. E. Fisher, Correlation functions and the critical region of simple fluids, J. Math. Phys. **5**, 944 (1964).

[56] L. S. Ornstein and F. Zernike, Accidental deviations of density and opalescence at the critical point of a single substance, Proc. Akad. Sci. (Amsterdam) **17**, 793 (1914).

[57] D. Ruelle, Superstable interactions in classical statistical mechanics, Commun. Math. Phys. **18**, 127 (1970).

[58] J. T. Chayes and L Chayes, On the validity of the inverse conjecture in classical density functional theory, J. Stat. Phys. **36**, 471 (1984).

[59] H. Yukawa, On the interaction of elementary particles. I, Prog. Theor. Phys. Suppl. **1**, 1 (1955).

[60] G. Schwarz, Estimating the dimension of a model, Ann. Stat. **6**, 461 (1978).

[61] D. C. Liu and J. Nocedal, On the limited memory BFGS method for large scale optimization, Math. Program. **45**, 503 (1989).

[62] G. Zhang, F. H. Stillinger, and S. Torquato, Ground states of stealthy hyperuniform potentials: I. Entropically favored configurations, Phys. Rev. E **92**, 022119 (2015).

[63] O. U. Uche, F. H. Stillinger, and S. Torquato, Constraints on collective density variables: Two dimensions, Phys. Rev. E **70**, 046122 (2004).

[64] R. D. Batten, F. H. Stillinger, and S. Torquato, Classical disordered ground states: Super-ideal gases, and stealth and equi-luminous materials, J. Appl. Phys. **104**, 033504 (2008).

[65] J. P. Hansen, Statistical mechanics of dense ionized matter. I. Equilibrium properties of the classical one-component plasma, Phys. Rev. A **8**, 3096 (1973).

[66] R. C. Gann, S. Chakravarty, and G. V. Chester, Monte Carlo simulation of the classical two-dimensional one-component plasma, Phys. Rev. B **20**, 326 (1979).

[67] F. J. Dyson, Statistical theory of the energy levels of complex systems. I, J. Math. Phys. **3**, 140 (1962).

[68] P. P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, Ann. Phys. **369**, 253 (1921).

[69] B. Widom, Random sequential addition of hard spheres to a volume, J. Chem. Phys. **44**, 3888 (1966).

[70] S. Torquato and F. H. Stillinger, Exactly solvable disordered sphere-packing model in arbitrary-dimensional Euclidean spaces, Phys. Rev. E **73**, 031106 (2006).

[71] B. Smit and D. Frenkel, Vapor-liquid equilibria of the two-dimensional Lennard-Jones fluid(s), J. Chem. Phys. **94**, 5663 (1991).

[72] Y.-W. Li and M. P. Ciamarra, Phase behavior of Lennard-Jones particles in two dimensions, Phys. Rev. E **102**, 062101 (2020).

[73] M. Dzugutov, Glass formation in a simple monatomic liquid with icosahedral inherent local order, Phys. Rev. A **46**, R2984 (1992).

[74] É. Marcotte, F. H. Stillinger, and S. Torquato, Nonequilibrium static growing length scales in supercooled liquids on approaching the glass transition, J. Chem. Phys. **138**, 12A508 (2013).

[75] D. W. Cooper, Random-sequential-packing simulations in three dimensions for spheres, Phys. Rev. A **38**, 522 (1988).

[76] Y. Pomeau, Some asymptotic estimates in the random parking problem, J. Phys. A: Math. Gen. **13**, L193 (1980).

[77] A. Gabrielli and S. Torquato, Voronoi and void statistics for superhomogeneous point processes, Phys. Rev. E **70**, 041105 (2004).

[78] J. Kim and S. Torquato, Effect of imperfections on the hyperuniformity of many-body systems, Phys. Rev. B **97**, 054105 (2018).

[79] O. Gereben, L. Pusztai, and A. Baranyai, Calculation of the three-particle contribution to the configurational entropy for two different models of amorphous Si, Phys. Rev. B **49**, 13251 (1994).

[80] D. M. Nicholson, C. Y. Gao, M. T. McDonnell, C. C. Sluss, and D. J. Keffer, Entropy pair functional theory: Direct entropy evaluation spanning phase transitions, Entropy **23**, 234 (2021).

[81] F. H. Stillinger and T. A. Weber, Dynamics of structural transitions in liquids, Phys. Rev. A **28**, 2408 (1983).

[82] S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties* (Springer-Verlag, New York, 2002).

[83] M. Yamada, Geometrical study of the pair distribution function in the many-body problem, Prog. Theor. Phys. **25**, 579 (1961).

[84] O. Costin and J. Lebowitz, On the construction of particle distributions with specified single and pair densities, J. Phys. Chem. B **108**, 19614 (2004).

[85] S. Torquato and F. H. Stillinger, New conjectural lower bounds on the optimal density of sphere packings, Experimental Math. **15**, 307 (2006).

[86] T. Kuna, J. L. Lebowitz, and E. R. Speer, Realizability of point processes, J. Stat. Phys. **129**, 417 (2007).

[87] D. Henderson, F. F. Abraham, and J. A. Barker, The Ornstein-Zernike equation for a fluid in contact with a surface, Mol. Phys. **31**, 1291 (1976).

[88] E. Schöll-Paschinger and G. Kahl, Self-consistent Ornstein-Zernike approximation for a binary symmetric fluid mixture, J. Chem. Phys. **118**, 7414 (2003).

[89] R. Lovett, C. Y. Mou, and F. P. Buff, The structure of the liquid-vapor interface, J. Chem. Phys. **65**, 570 (1976).

[90] J. M. Brader, Structural precursor to freezing: An integral equation study, J. Chem. Phys. **128**, 104503 (2008).

[91] A. Moradzadeh and N. R. Aluru, Transfer-learning-based coarse-graining method for simple fluids: Toward deep inverse liquid-state theory, J. Phys. Chem. Lett. **10**, 1242 (2019).

[92] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).

[93] S. M Girvin and K. Yang, *Modern Condensed Matter Physics* (Cambridge University Press, Cambridge, 2019).

[94] L. Landau, Theory of the superfluidity of helium II, Phys. Rev. **60**, 356 (1941).

[95] R. P. Feynman and M. Cohen, Energy spectrum of the excitations in liquid helium, Phys. Rev. **102**, 1189 (1956).

[96] B. Sutherland, Quantum many-body problem in one dimension: Ground state, J. Math. Phys. **12**, 246 (1971).

[97] S. Torquato, A. Scardicchio, and C. E. Zachary, Point processes in arbitrary dimension from Fermionic gases, random matrix theory, and number theory, J. Stat. Mech.: Theory Exp. (2008) P11019.