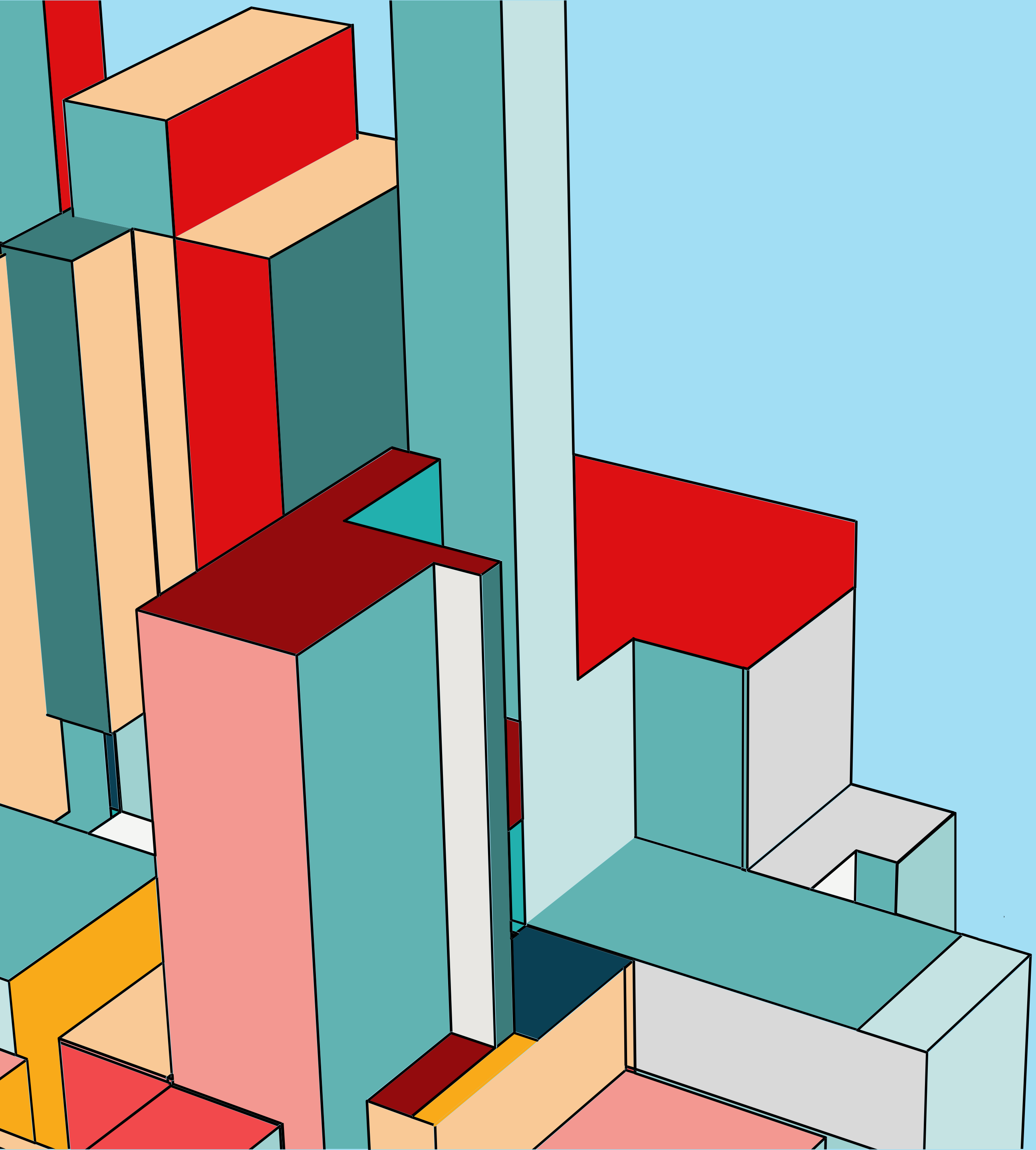


PRINCETON UNIVERSITY
DEPARTMENT OF CHEMISTRY

DESIGNER PAIR STATISTICS FOR MANY-BODY SYSTEMS WITH NOVEL PROPERTIES

Haina Wang
Salvatore Torquato*

Muri Webinar, 3/15/2024

Wang and Torquato, JCP (2024).

ACKNOWLEDGEMENTS

Salvatore Torquato

Complex Materials Theory Group

Inverse statistical mechanics:
Assemble structures through designed interactions

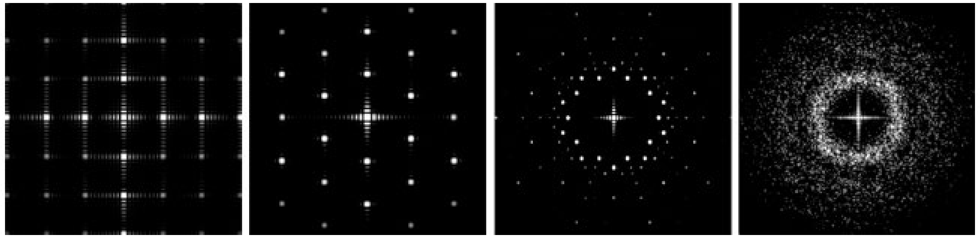
Microscopic interactions → (Stat. mech.) → Macroscopic structures and properties

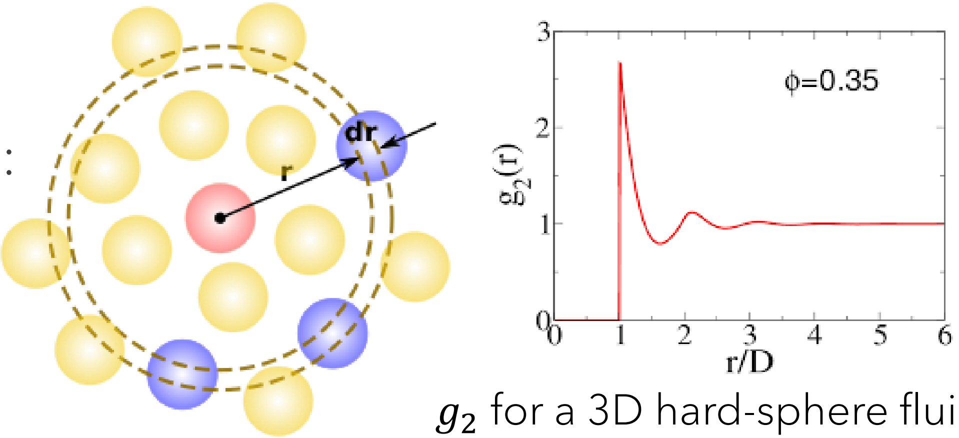
Microscopic interactions ← (Inverse stat. mech.) ← Macroscopic structures and properties

The pair correlation function $g_2(r)$ is a scaled histogram of pair distances.

- Computationally simple
- Experimentally available via scattering
- Contains information e.g., pressure

"Doppelganger" of $g_2(r)$ in Fourier space:
The structure factor $S(k) = 1 + \rho\tilde{h}(k)$



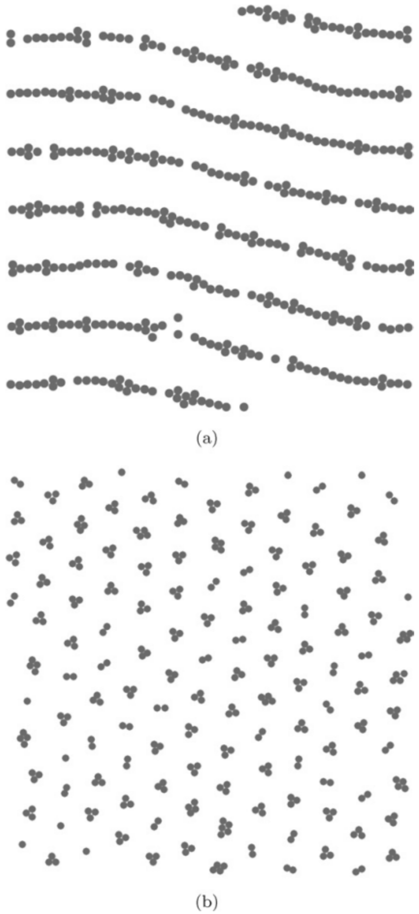


g_2 for a 3D hard-sphere fluid.

Torquato, *Random Heterogeneous Materials* (2002).

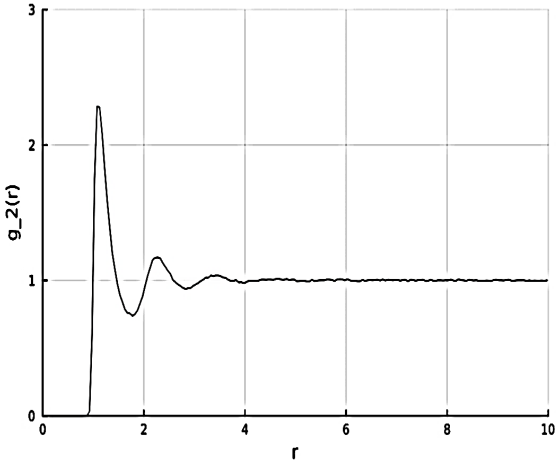
Motivation 1: Knowledge of analytical pair functions is limited

- There exist only a small number of disordered many-particle systems whose pair functions are known analytically over their entire parameter regime.
  - Determinantal point processes, e.g., free Fermi gases
  - 1D equilibrium hard rods
  - “Ghost” RSA across dimensions
  - 2D and 3D unit-step function g_2
- Expanding this database enables the analytical exploration of equilibrium and nonequilibrium properties and **materials design**.



Torquato, *Soft Matter* (2009).

Motivation 2:
The realizability problem



Can any mathematical function be the $g_2(r)$ for some many-body system?

Known necessary (but insufficient) conditions for realizability:

- $g_2(r) \geq 0$ for all r.
- $S(k) \geq 0$ for all k.
- The Yamada condition
- An uncountable number of necessary and sufficient conditions are generally required

We need **highly accurate numerical algorithms** to probe the realizability.

Torquato and Stillinger, *Expt. Math.* (2006)

Motivation 3:
Design large-scale correlations

- Hyperuniform state: $\lim_{k\to 0} S(k) = 0$. Usually, one has a power law $S(k) \sim k^{\alpha}$ at small k.
- The exponent α determines the large-*R* scaling behaviors of the number variance.

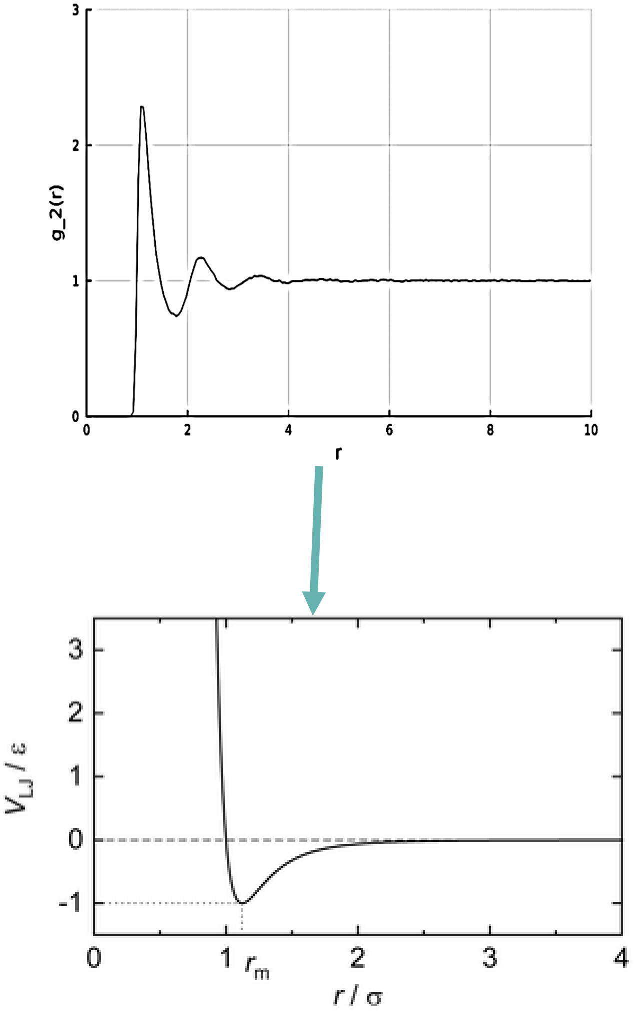
$$\sigma^2(R) \sim \begin{cases} R^{d-1}, & \alpha > 1 \text{ (class I)}, \\ R^{d-1} \ln R, & \alpha = 1 \text{ (class II)}, \\ R^{d-\alpha}, & 0 < \alpha < 1 \text{ (class III)}. \end{cases} \qquad \sigma^2(R) \sim \begin{cases} R^{d}, & \alpha = 0 \text{ (typical nonhyperuniform)}, \\ R^{d-\alpha},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

- We consider a broad range of functional forms that span across a wide spectrum of hyperuniform and nonhyperuniform classes.

Torquato and Stillinger, PRE (2003)
Torquato, PRE (2021)

Functional form	d	$g_2(r)$	$S(k)$	α
Gaussian	1, 2, 3	$1-\exp\left(-\pi r^2\right)$ (3)	$1-\exp\left(-\frac{k^2}{4\pi}\right)$ (4)	2
Generalized OCP[48]	3	$1-\exp\left(-\frac{4\pi}{3}r^3\right)$ (5)	$1-{}_1F_4\left(1;\frac{1}{3},\frac{2}{3},\frac{5}{6},\frac{7}{6};-\frac{k^6}{20\,736\pi^2}\right)+\frac{2\pi\,\mathrm{ber}_{\frac{2}{3}}\left(\frac{(k^2)^{3/4}}{3\sqrt{\pi}}\right)}{3\sqrt{3}}+\frac{2\pi\,\mathrm{ber}_{-\frac{2}{3}}\left(\frac{(k^2)^{3/4}}{3\sqrt{\pi}}\right)}{3\sqrt{3}}$ (6)	2
Fourier dual of OCP	1	$1-\frac{1}{\pi r^2+1}$ (7)	$1-\exp\left(-k/\pi\right)$ (8)	1
Fourier dual of OCP[48]	3	$1-{}_1F_4\left(1;\frac{1}{3},\frac{2}{3},\frac{5}{6},\frac{7}{6};-\frac{1}{324}\pi^4r^6\right)+\frac{2\pi\,\mathrm{ber}_{\frac{2}{3}}\left(\frac{2}{3}\sqrt{2\pi}\left(r^2\right)^{3/4}\right)}{3\sqrt{3}}+\frac{2\pi\,\mathrm{ber}_{-\frac{2}{3}}\left(\frac{2}{3}\sqrt{2\pi}\left(r^2\right)^{3/4}\right)}{3\sqrt{3}}$ (9)	$1-\exp\left(-\frac{2}{3(2\pi)^2}k^3\right)$ (10)	3
Hyperbolic secant $g_2(r)$	1	$1-\mathrm{sech}(\pi r)$ (11)	$1-\mathrm{sech}(k/2)$ (12)	2
Hyperbolic secant $g_2(r)$[49]	2	$1-\mathrm{sech}(2\sqrt{\pi G}r)$ (13)	No closed analytical form	2
Hyperbolic secant $g_2(r)$	3	$1-\mathrm{sech}\left(\frac{\pi^{4/3}}{2^{1/3}}r\right)$ (14)	$1-\frac{2^{2/3}\sqrt[3]{\pi}\tanh\left(\frac{k}{2^{2/3}\sqrt[3]{\pi}}\right)\mathrm{sech}\left(\frac{k}{2^{2/3}\sqrt[3]{\pi}}\right)}{k}$ (15)	2
Gaussian-damped polynomial	2	$1-\frac{1}{3}e^{-\pi r^2}\left(4\pi^2r^4-8\pi r^2+3\right)$ (16)	$1-\frac{e^{-\frac{k^2}{4\pi}}\left(k^4-8\pi k^2+12\pi^2\right)}{12\pi^2}$ (17)	2
Hermite–Gaussian[50]	1	$1-\lambda\left(4r^4-12r^2+3\right)e^{-\frac{r^2}{2}}$ (18)	$1-\lambda\sqrt{2\pi}\left(-4k^4+12k^2-3\right)e^{-\frac{k^2}{2}}$ (19)	0
Hyposurficial[51]	3	$1+\frac{\exp\left(-r^*\right)}{4\pi}-\frac{\exp\left(-r^*\right)\sin\left(r^*\right)}{r^*}$ (20)	$\frac{6k^{*8}+12k^{*6}+19k^{*4}+24k^{*2}+16}{6(k^{*2}+1)^2(k^{*2}-2k^*+2)(k^{*2}+2k^*+2)}$ (21)	0
Ghost RSA[52]	1	$\frac{2\Theta(2r-1)}{2-(1-r)\Theta(2-2r)}$ (22)	$1-4\cos(k)\left[\mathrm{Ci}\left(\frac{3k}{2}\right)+\mathrm{Ci}(2k)\right]-4\sin(k)\left[\mathrm{Si}\left(\frac{3k}{2}\right)+\mathrm{Si}(2k)\right]-\frac{2\sin(k)}{k}$ (23)	0
Ghost RSA[53]	2	$\frac{2\Theta(r^*-1)}{2-\frac{2}{\pi}\left[\cos^{-1}\left(\frac{r^*}{2}\right)-\frac{r}{2}\sqrt{1-\frac{r^{*2}}{4}}\right]\Theta(2-r^*)}$ (24)	No closed analytical form	0
Antihyperuniform[54]	1, 2, 3	$\Theta(r/D-1)\left(\frac{A}{(r/D)^{d-\frac{1}{2}}}+1\right)$ (25)	See Appendix	$-\frac{1}{2}$

Inverse algorithm



As suggested in [Zhang and Torquato, PRE (2020)] , we employ inverse algorithms that determines **classical equilibrium states with up to pair interactions $v(r)$** at positive temperature T that precisely match targeted forms for both $g_2(r)$ and $S(k)$.

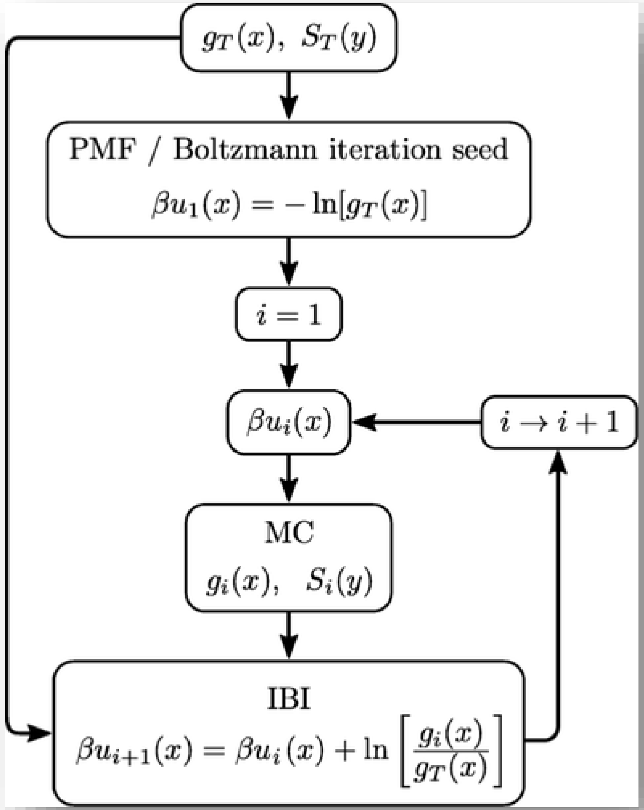
Popular previous methods: Iterative Boltzmann inversion (IBI)

Target pair statistics

Initial guess potential
(It's exact at $\rho = 0$)

Simulated pair statistics

Update rule



Next iteration

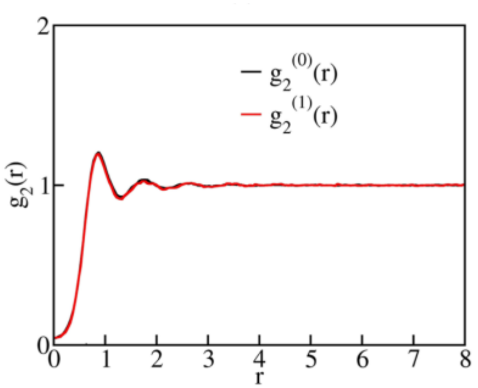
If $g_2(r)$ is too large, increase energy at r so that there are fewer particle pairs with distance r.

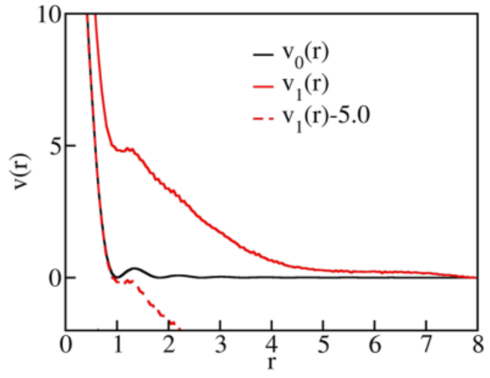
Soper, *Chem. Phys.* (1996)

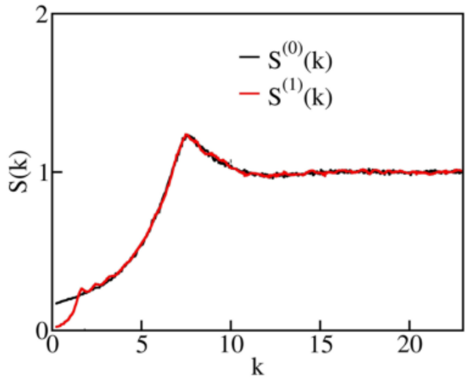
But IBI is not precise enough for our goal…

TLDR: The solution is not unique even if it should be.

- **Henderson's theorem**: Given ρ and $T > 0$, the pair potential that generates a target $g_2(r)$ at equilibrium is unique up to an additive constant.
- In practice, very different potentials can give very similar $g_2(r)$.
- … because information in the tail of $g_2(r)$ is truncated or buried in the noise.







Solution: Target both $g_2(r)$ and $S(k)$.

Wang, Stillinger and Torquato, JCP (2020)

Toward a more effective inverse algorithm

Challenge	Solution
Solution is not unique	Define the objective. $$\Psi(\mathbf{a}) = \rho \int_{\mathbb{R}^d} w_{g_2}(\mathbf{r}) \left(g_{2,T}(\mathbf{r}) - g_2(\mathbf{r};\mathbf{a})\right)^2 d\mathbf{r} + \frac{1}{\rho(2\pi)^d} \int_{\mathbb{R}^d} w_S(\mathbf{k}) \left(S_T(\mathbf{k}) - S(k;\mathbf{a})\right)^2 d\mathbf{k},$$
Unclear optimization objective	
Sensitive to simulation noise	Use parametrized smooth analytical potential functions. $$v(r,\boldsymbol{a}) = \varepsilon \sum_{j=1}^{n} f_j(r; a_j)$$
Too model-dependent	Use a more general updating rule, e.g., BFGS, rather than the difference in PMF. •To compute $\nabla\Psi(\boldsymbol{a})$, simulations are performed at perturbed $\boldsymbol{a}_j$'s and partial derivative are computed via finite differences. •$2n$ simulations are needed for n potential parameters.

Torquato and Wang, PRE (2022)

Toward a more effective inverse algorithm

- **Automatic differentiation** can reduce gradient computation from $O(n)$ to $O(1)$! But it needs a **deterministic** objective function, unlike our **stochastic** $\Psi(\boldsymbol{a})$.
- To use AD, we need a deterministic function $\psi(\boldsymbol{a}')$ to approximate $\Psi(\boldsymbol{a}_i)$ for $\boldsymbol{a}'$ in the vicinity of $\boldsymbol{a}_i$.

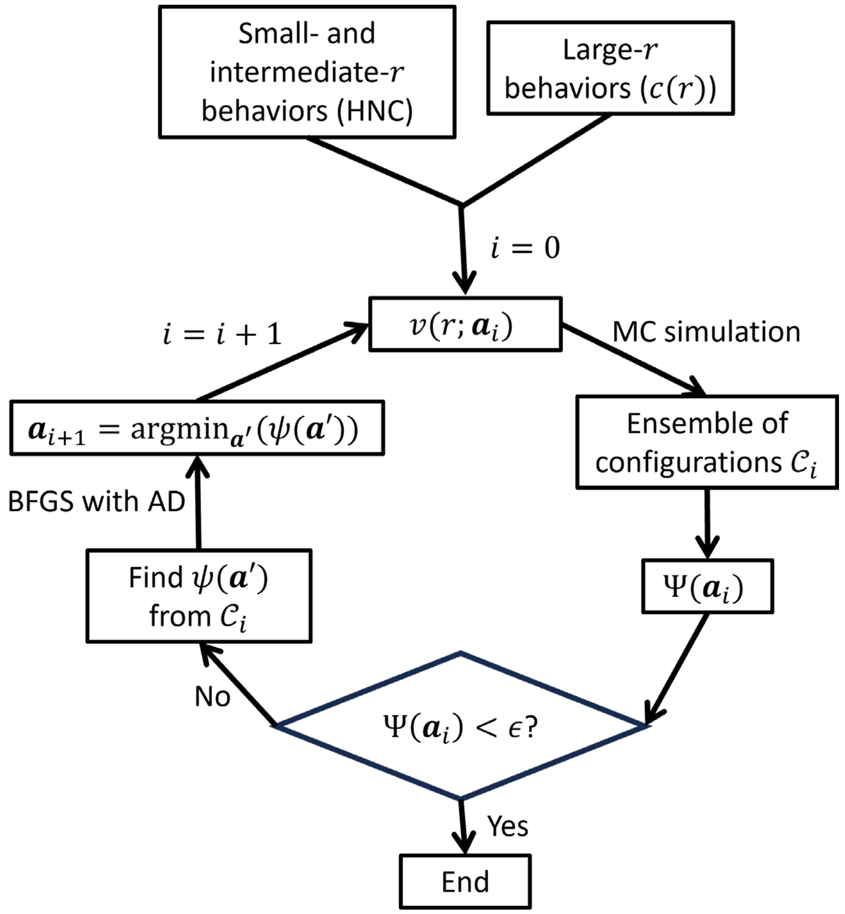
$$\Psi(\mathbf{a}) = \rho \int_{\mathbb{R}^d} w_{g_2}(\mathbf{r}) \left(g_{2,T}(\mathbf{r}) - g_2(\mathbf{r};\mathbf{a})\right)^2 d\mathbf{r} + \frac{1}{\rho(2\pi)^d} \int_{\mathbb{R}^d} w_S(\mathbf{k}) \left(S_T(\mathbf{k}) - S(k;\mathbf{a})\right)^2 d\mathbf{k},$$

- We use the **canonical ensemble reweighing technique** [Norgaard et al. Biophys J. (2008)]:

$$\varphi(x;\mathbf{a}') = \frac{\sum_{\mathbf{r}^N \in \mathcal{C}_i} \varphi(x;\mathbf{r}^N) \exp[-(\Phi(\mathbf{r}^N;\mathbf{a}') - \Phi(\mathbf{r}^N;\mathbf{a}_i))/T]}{\sum_{\mathbf{r}^N \in \mathcal{C}_i} \exp[-(\Phi(\mathbf{r}^N;\mathbf{a}') - \Phi(\mathbf{r}^N;\mathbf{a}_i))/T]},$$

where $\varphi(x)$ is $g_2(r)$ or $S(k)$.

- The approximant $\psi(\boldsymbol{a}')$ is obtained by inserting $\varphi(x;\boldsymbol{a}')$ into the objective function.



Hyperuniform equilibrium states

- To achieve equilibrium hyperuniform systems at positive T, which are thermodynamically incompressible, one requires **long-ranged pair interaction** in the large-r limit [Torquato, Phys. Rep. (2018)]

$$v(r) \sim \begin{cases} r^{-(d-\alpha)}, & d \neq \alpha \\ -\ln(r), & d = \alpha. \end{cases}$$

- …as well as a **background one-body potential** to maintain charge neutrality.
- Example: Dyson log gas, in which 1D particles interact via the log potential and are immersed in a uniform background of the opposite "charge" [Dyson, *J. Math. Phys.* (1962)].

Results for designed pair correlation functions

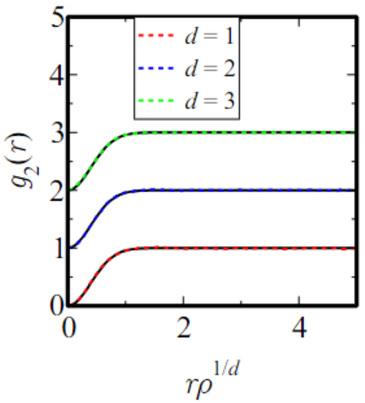
- For fluids away from the critical point, convergence can be achieved within **5h** on an Intel 2.8 GHz CPU with 15 cores and 4GB memory per core.
- <24h for critical-point states.

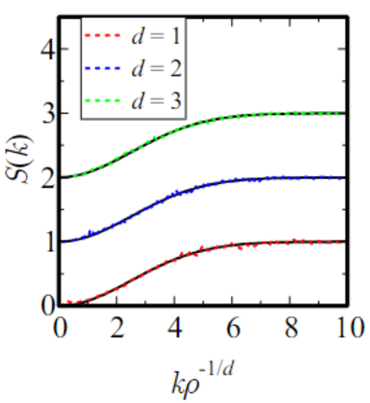
Gaussian pair functions

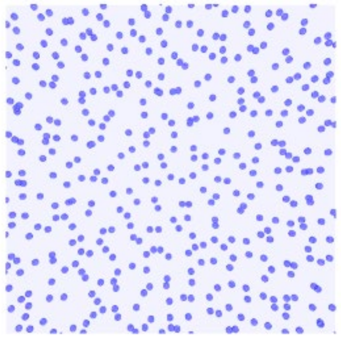
Gaussian pair statistics

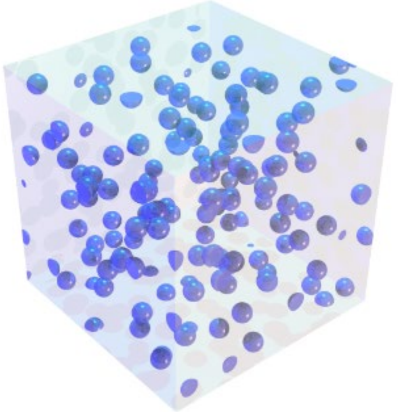
- HU with $\alpha = 2$.
- Incorporate **soft-core short-ranged repulsions** in macromolecules and are commonly used to model **polymer systems**.
- Self-similar under Fourier transforms.
- Related to Gaussian-core model [Stillinger, JCP (1976)]

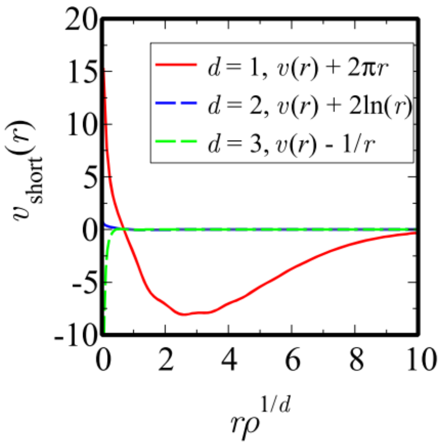
- We have realized them in 1, 2 and 3 dimensions.
- For the 2D state, it is known that the effective pair potential is exactly given by $v(r) = -2\ln(r)$. Our inverse procedure has accurately recovered this potential.





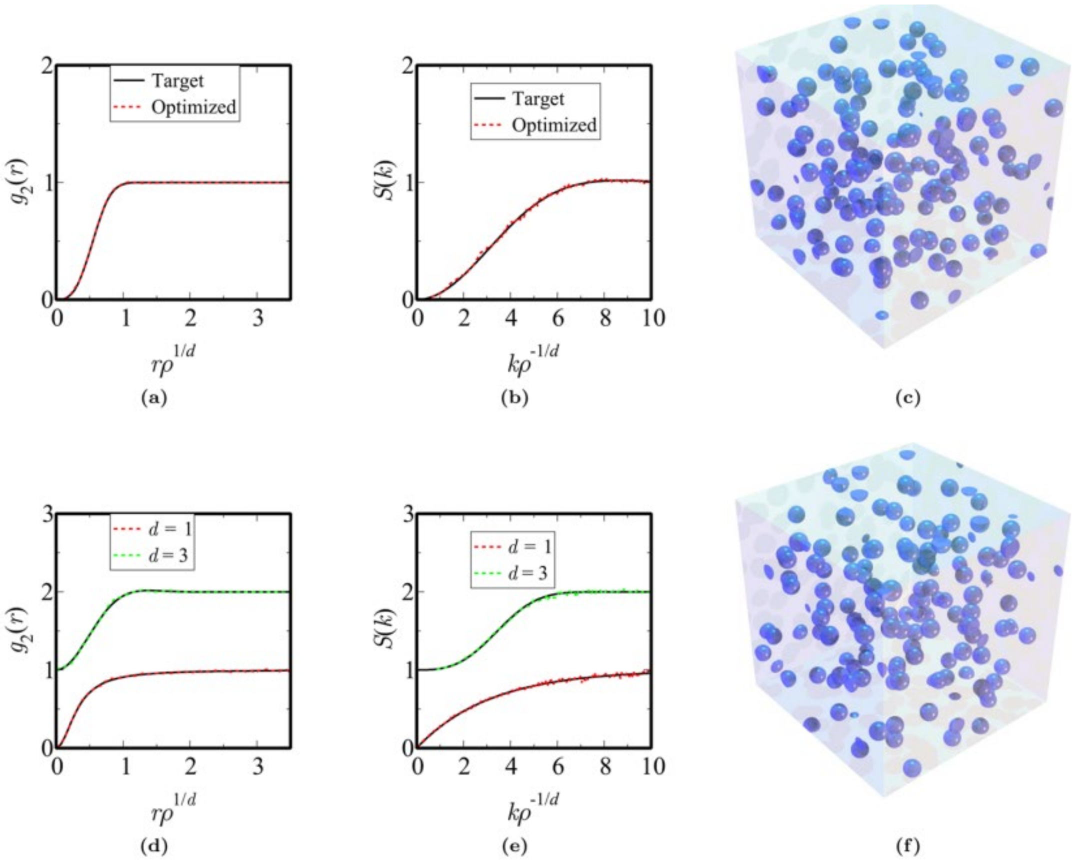






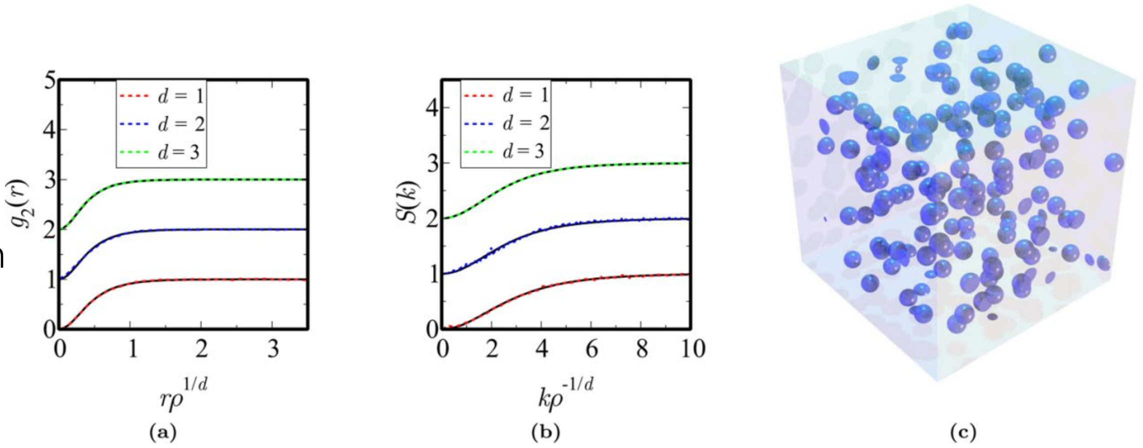
One-component plasmas and their Fourier duals

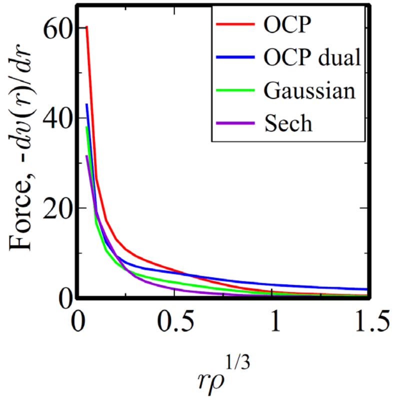
- OCP
  - HU with $\alpha = 2$.
  - Represent identical point charges interacting via the Coulomb potential and immersed in a rigid uniform background of opposite charge.
  - $g_2(r) = 1 - \exp(-v_1(r))$
- Fourier dual of OCP
  - HU with $\alpha = d$
  - $S(k) = 1 - \exp(-v_1\left(\frac{k}{2\pi}\right))$
- The 3D Fourier dual of OCP is the first known equilibrium system with $\alpha > 2$.
  - Large $\boldsymbol{\alpha}$ are desirable for optical purposes because they tend to be effectively transparent for a wide range of wave numbers compared to nonhyperuniform ones.



Hyperbolic secant function $g_2(r)$

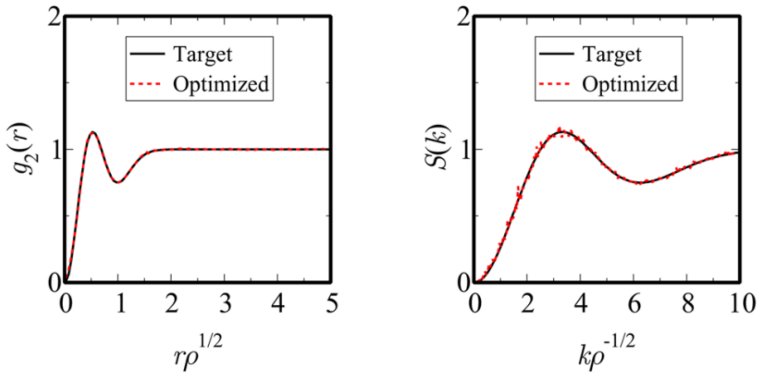
- Hyperbolic secant pair functions
  - HU with $\alpha = 2$
  - "**Leptokurtic**": heavier tailed than Gaussian
  - Describe polymers with softer repulsive interactions than the Gaussian core
- The 3D configuration contains more clusters with four or more closely spaced particles with pair distances $r \leq 0.4\rho^{-\frac{1}{3}}$ compared to the Gaussian case.
- The effective potential is the least repulsive at small r among the 3D hyperuniform model.



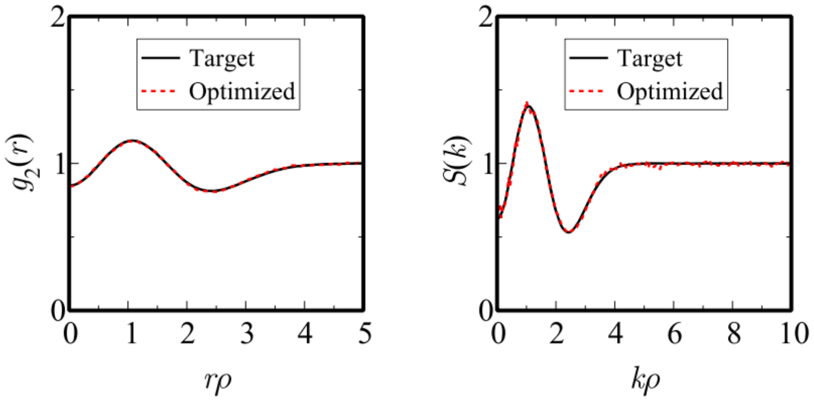


Gaussian-damped polynomial and Hermite–Gaussian pair statistics

- They mimic pair statistics for dense polymer systems, which often exhibit oscillations due to the coordination shells.
- The pair functions are **self-similar under Fourier transforms** up to scaling.
  - Enables one to exploit **duality relations** in the phase diagrams. [Torquato and Stillinger, PRL (2008)]
  - Their low-density and high-density ground-state structures under the effective potentials can be mapped to one another.
- The effective potentials are oscillatory and may be realized by **"sticky" macromolecules**, such as DNA-coated nanoparticles.



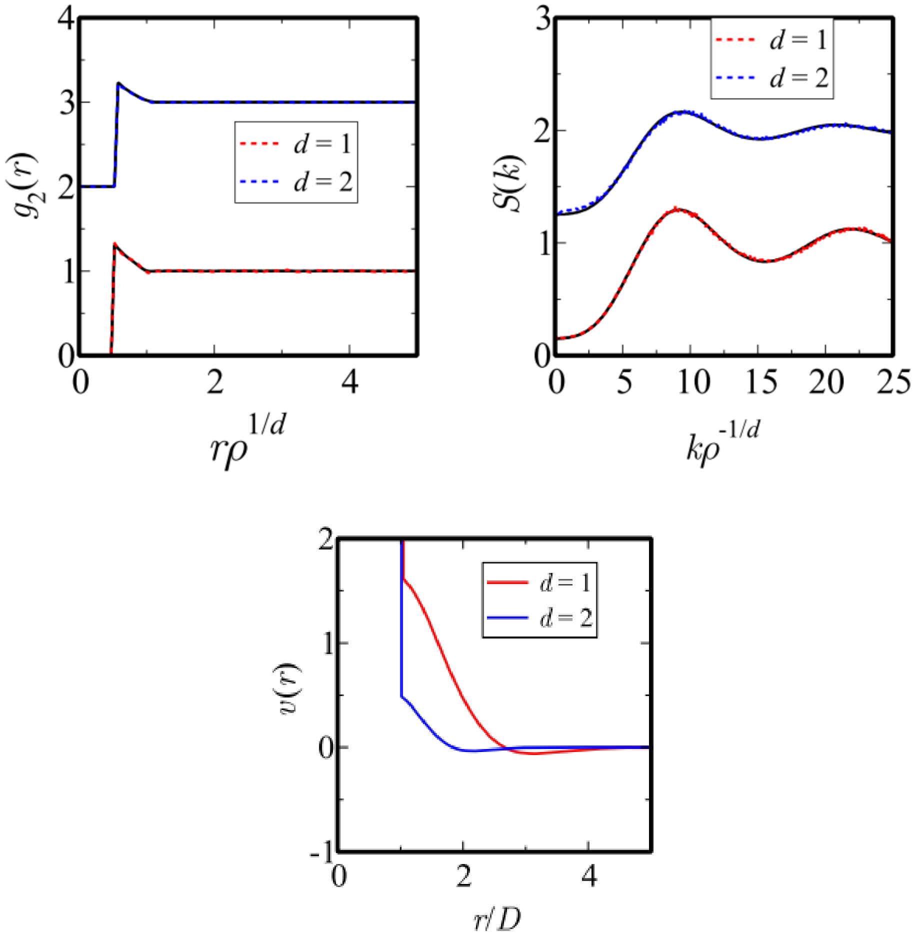
(a) (b)



(a) (b)

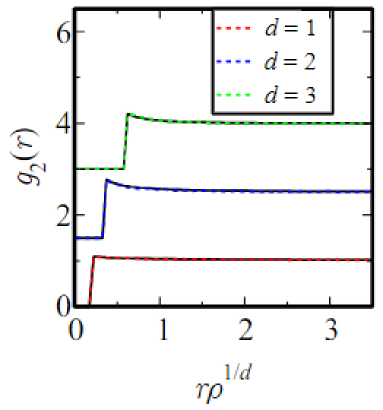
Ghost random sequential addition

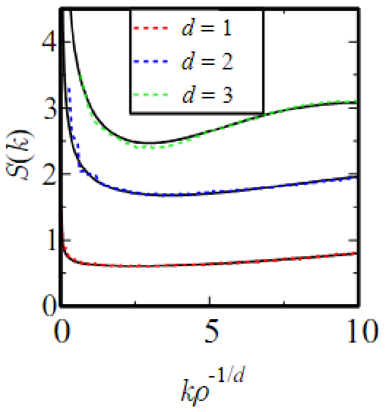
- Ghost RSA
  - NonHU. **The only known model for which all n-particle correlation functions are exactly solvable**.
  - Spherical particles of diameter D are added according to a translationally invariant Poisson process.
  - A sphere is retained iff it does not overlap with any test sphere added in the time interval $(0, t)$.
- 3D ghost RSA pair statistics at the terminal packing fraction $\phi_c = 1/2^d$ has been shown to be realizable [Wang, Stillinger and Torquato, JCP (2023)].
- We realized the 1D and 2D cases here.
- Soft repulsion is more significant for the 1D case than the 2D case, reflecting the decorrelation principle.

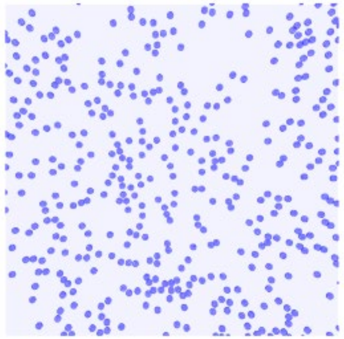


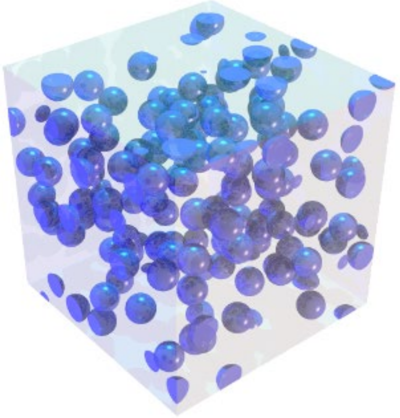
Designed critical point

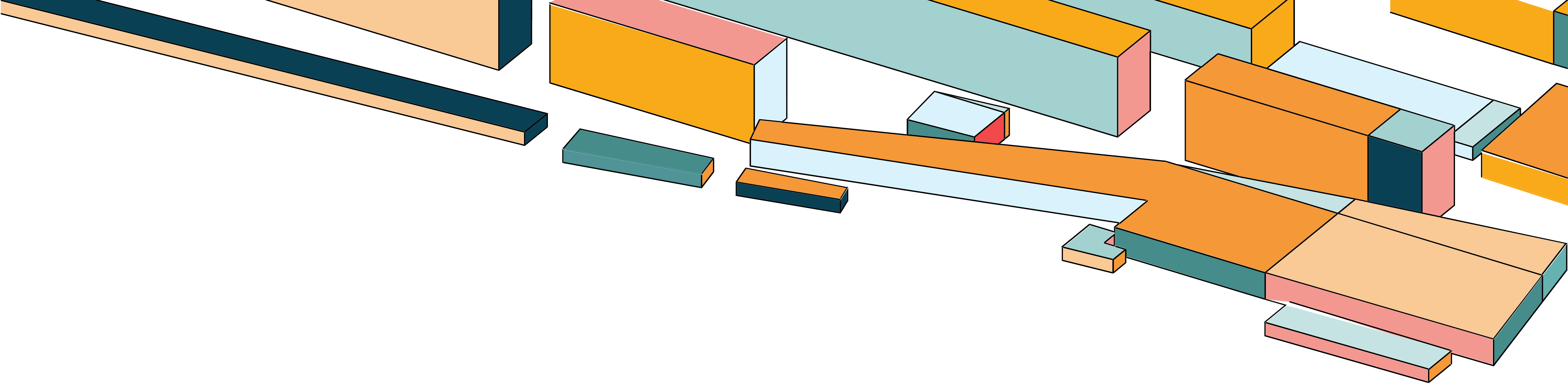
- We target the antihyperuniform pair statistics
  - $S(k) \sim k^{-1/2}$
  - $g_2(r) = \Theta\left(\frac{r}{D} - 1\right)\left(\frac{A}{\left(\frac{r}{D}\right)^{d-\frac{1}{2}}} + 1\right)$
  - Correspond to **critical points out of the Ising universality class**.
- We have realized these pair statistics in 1, 2 and 3 dimensions with some chosen A and ρ.
- Note the large clusters and holes characteristic of critical-point fluids!







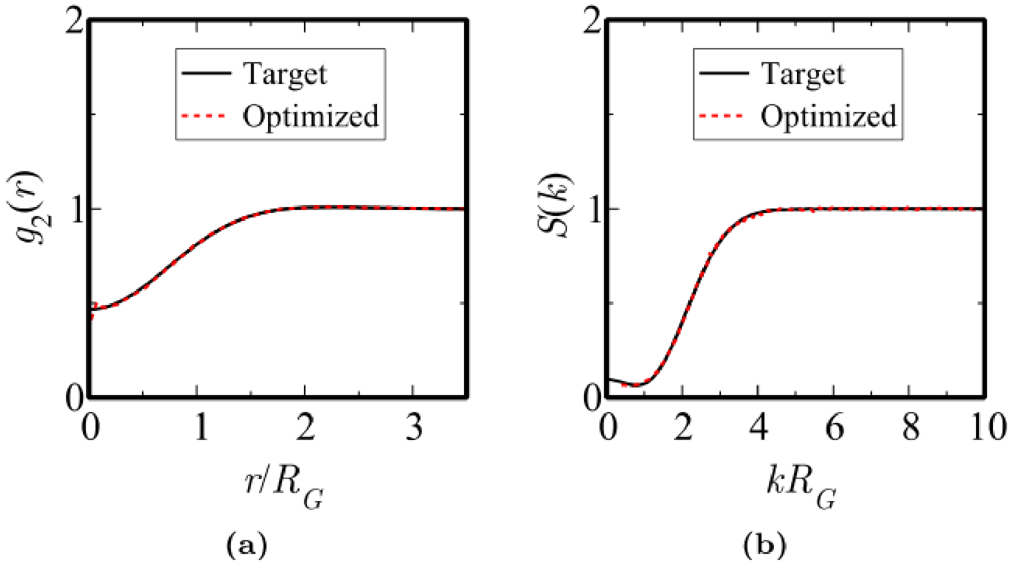




Applications

Predict experiment feasibility

- We obtain pair statistics from a polyethylene model [Yatsenko et al. PRL (2004)] with **parameters out of the experimented range**.
- We find that these pair statistics are realizable by an effective potential.
- Thus, they are probably achievable in experiments, e.g., via polyethylene molecules with a smaller number of monomers per chain than those considered by Yatsenko et al.



Controlled mass transport

- To demonstrate that our analytical pair functions enable one to achieve **a wide range of structural order and physical properties**, we compute the translational order metric τ and the self-diffusion coefficient D for the 3D states.

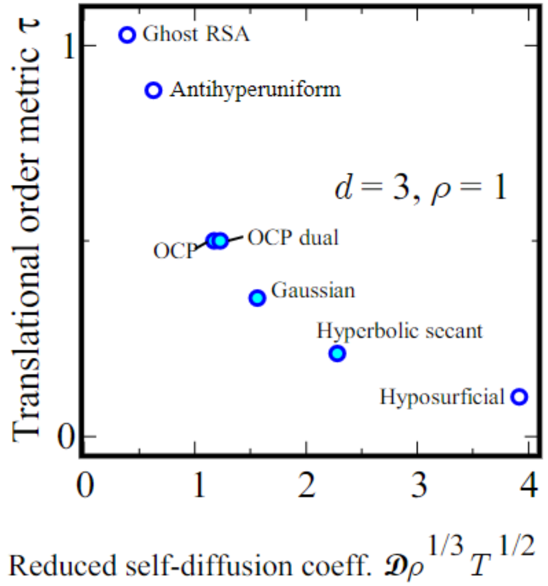
$$\tau = \rho \int h^2(r)d\mathbf{r} = \frac{\rho}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{h}^2(k)d\mathbf{k},$$

$$\mathscr{D}\rho^{1/3}T^{1/2} = 0.37(-s_2)^{-2/3},$$

where

$$s_2 = -(\rho/2) \int_{\mathbb{R}^d} \{g_2(r)\ln[g_2(r)] - g_2(r) + 1\}d\mathbf{r}$$

- **τ is inversely related to D.**
- Small τ → Less hindrance of particle motion → large D
- Useful in controlled drug delivery and diffusion-controlled catalysis



Database of realizable pair statistics

- Realizability via equilibrium classical states does not exclude the possibility of realizing the same pair statistics via quantum-mechanical equilibrium systems or even nonequilibrium systems.
- Our realizable pair functions can be used as basis functions to create new designs at unit density

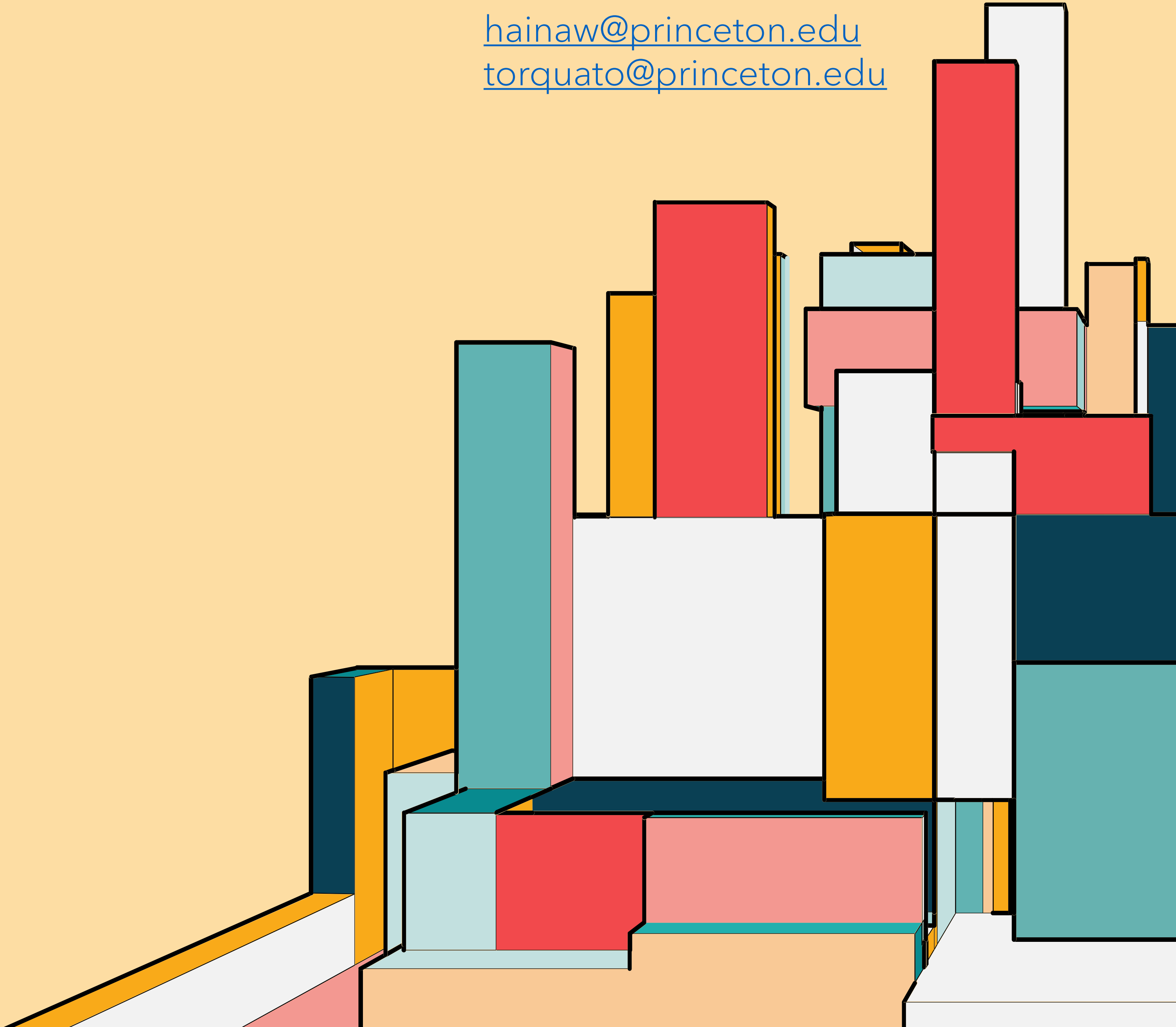
$$\overset{*}{g}_2(r) = \Sigma_{i=1}^{n} \lambda_i g_{2,i}(r), \quad 0 \leq \lambda_i \leq 1, \Sigma_i^n \lambda_i = 1.$$

- If a given $g_2(r)$ is realizable at $\rho\ = 1$, then it is also realizable at all lower densities because the realizable density range for a given $g_2(r)$ is always a continuous interval [Stillinger and Torquato, *Mol. Phys.* (2005)].

Conclusions

- Our inverse methodology enables us to precisely determine the unique potential corresponding to given pair statistics.
- Our method is expected to facilitate the self-assembly of soft-matter systems with tunable properties.
- Further support the Zhang-Torquato conjecture:
  - any realizable set of pair statistics, whether from a nonequilibrium or equilibrium system, can be achieved by equilibrium systems involving up to two-body interactions

hainaw@princeton.edu
torquato@princeton.edu



THANKS!

Q&A