

Advancing Interatomic Force Prediction with Machine Learning: Accelerated Methods for Extracting Force Fields from Experimental Scattering Measurements

Brennon L. Shanks, Harry W. Sullivan, Michael P. Hoepfner
University of Utah, Department of Chemical Engineering

A New Use for Neutron Diffraction?

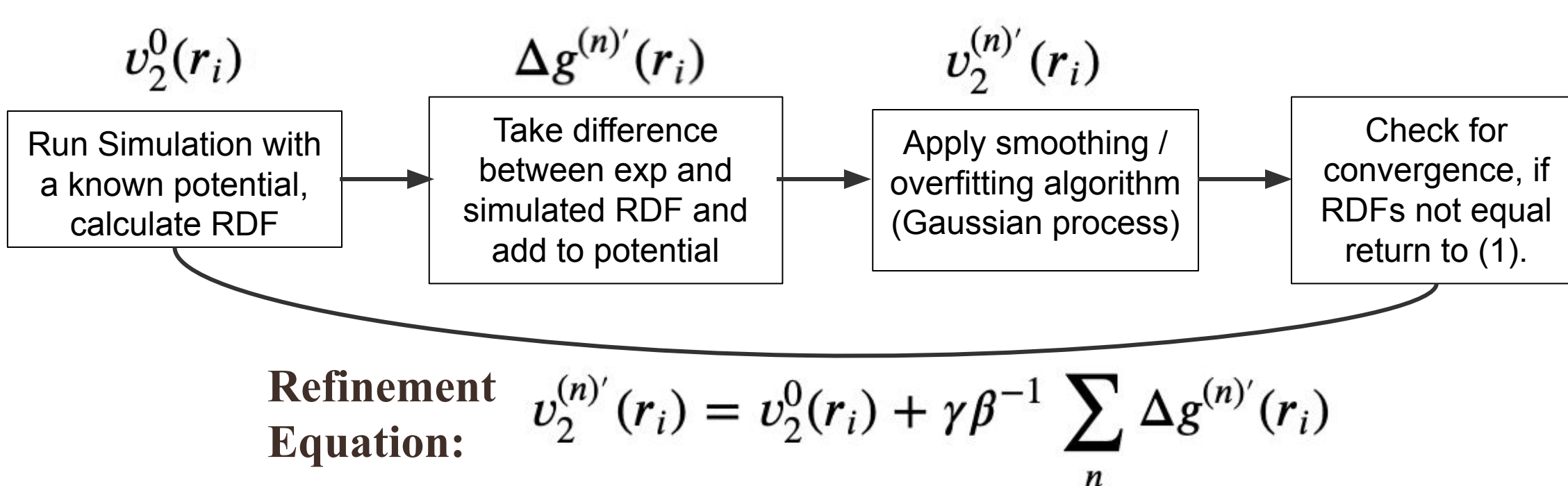
Calculating potential energy or force field parameters from structure data is called the statistical mechanical inverse problem.

Inverse methods learn force parameters from experimental data collected at the same length scale as the molecular model (<0.1 Å).

Accurate inverse methods could improve our understanding of self-assembly, fluid structure, and interatomic interactions.

Structure Optimized Potential Refinement (SOPR)

SOPR is a numerical Henderson inverse theorem algorithm designed to generate transferable potentials from experimental scattering data.

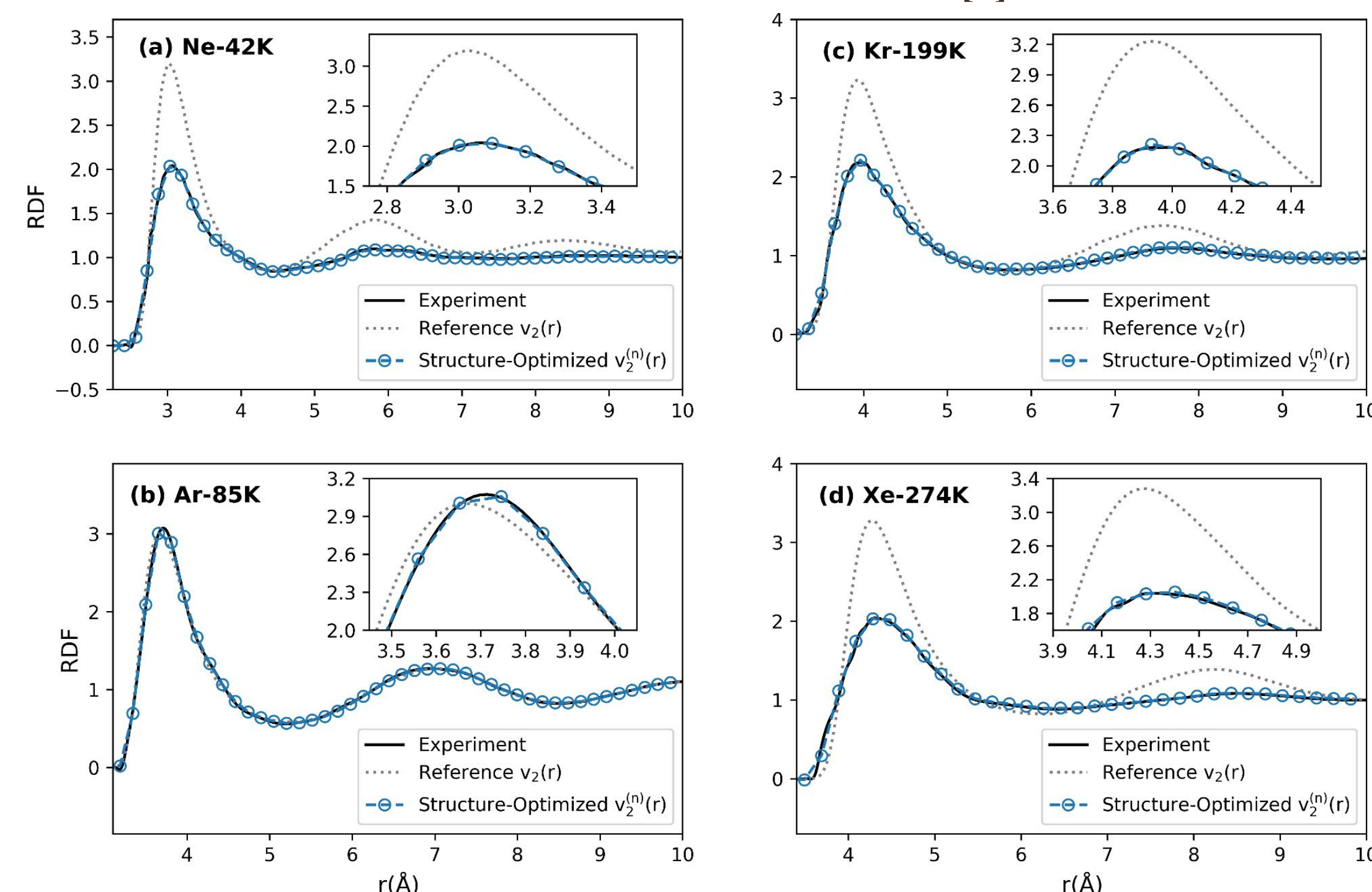


SOPR Potentials for the Noble Gas Series

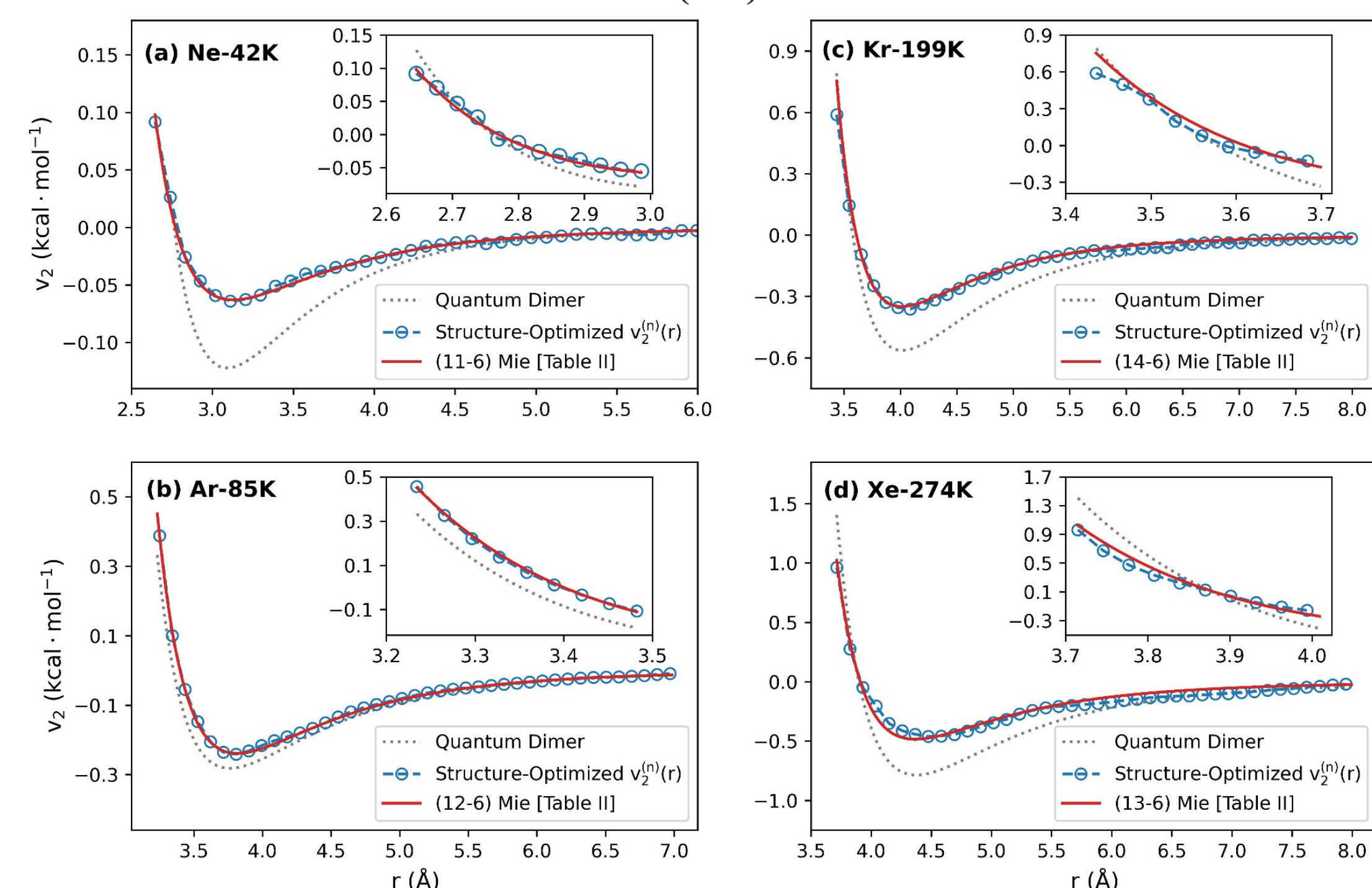
We explore whether structure-optimized potentials can predict fluid structure and other thermodynamic properties simultaneously.

SOPR was tested on the noble gas series Ne, Ar, Kr, and Xe.

Radial Distribution Functions [1]

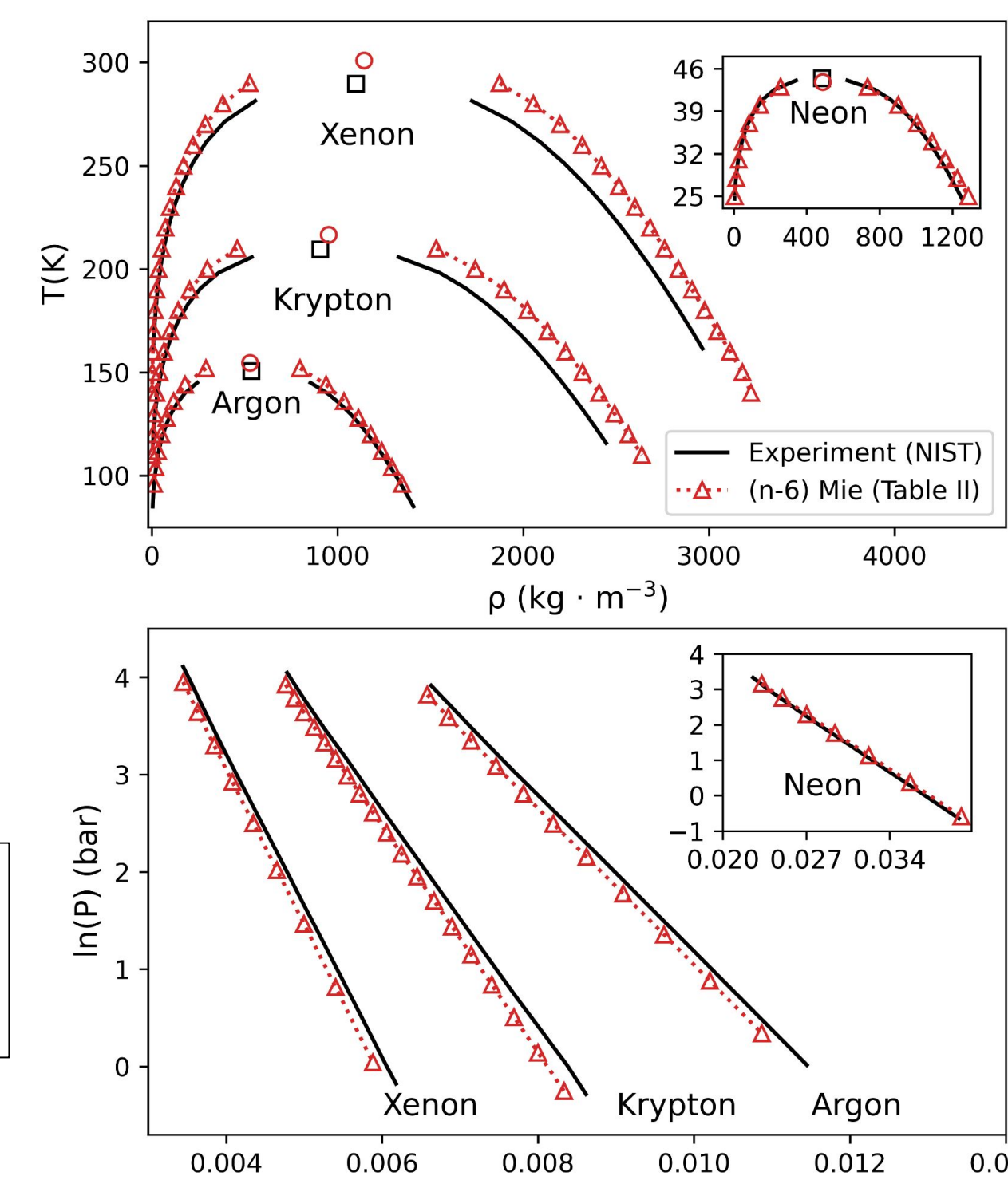


SOPR Potentials + (n-6) Mie Potential Fit



Transferability to Vapor-Liquid Equilibria (VLE)

VLE Predictions from SOPR Potentials



SOPR potentials can predict vapor-liquid phase envelopes and critical points with <5% error.

Repulsive exponents consistent with independent VLE FFs.

First evidence that Henderson Inverse techniques can be used to build force fields for real fluids.

Ensemble structures are sensitive to details of interatomic force fields.

Motivates continued development of SOPR for molecules/mixtures.

First demonstration of transferable potentials from structure in real fluids.

Problem: Currently, we don't know how reliable / robust these predictions are, particularly for diffraction measurements with uncertainty.

Uncertainty Quantification with Bayesian Analysis

Bayesian optimization provides accurate uncertainty quantification to help understand the credibility of predictions.

$$\text{Bayes Theorem: } p(\theta|\mathcal{Y}) = \frac{p(\mathcal{Y}|\theta)p(\theta)}{p(\mathcal{Y})}$$

Problem: Bayesian optimization is computationally infeasible for molecular simulations because it requires ~100,000 model calls.

Local Gaussian Process Surrogate Models

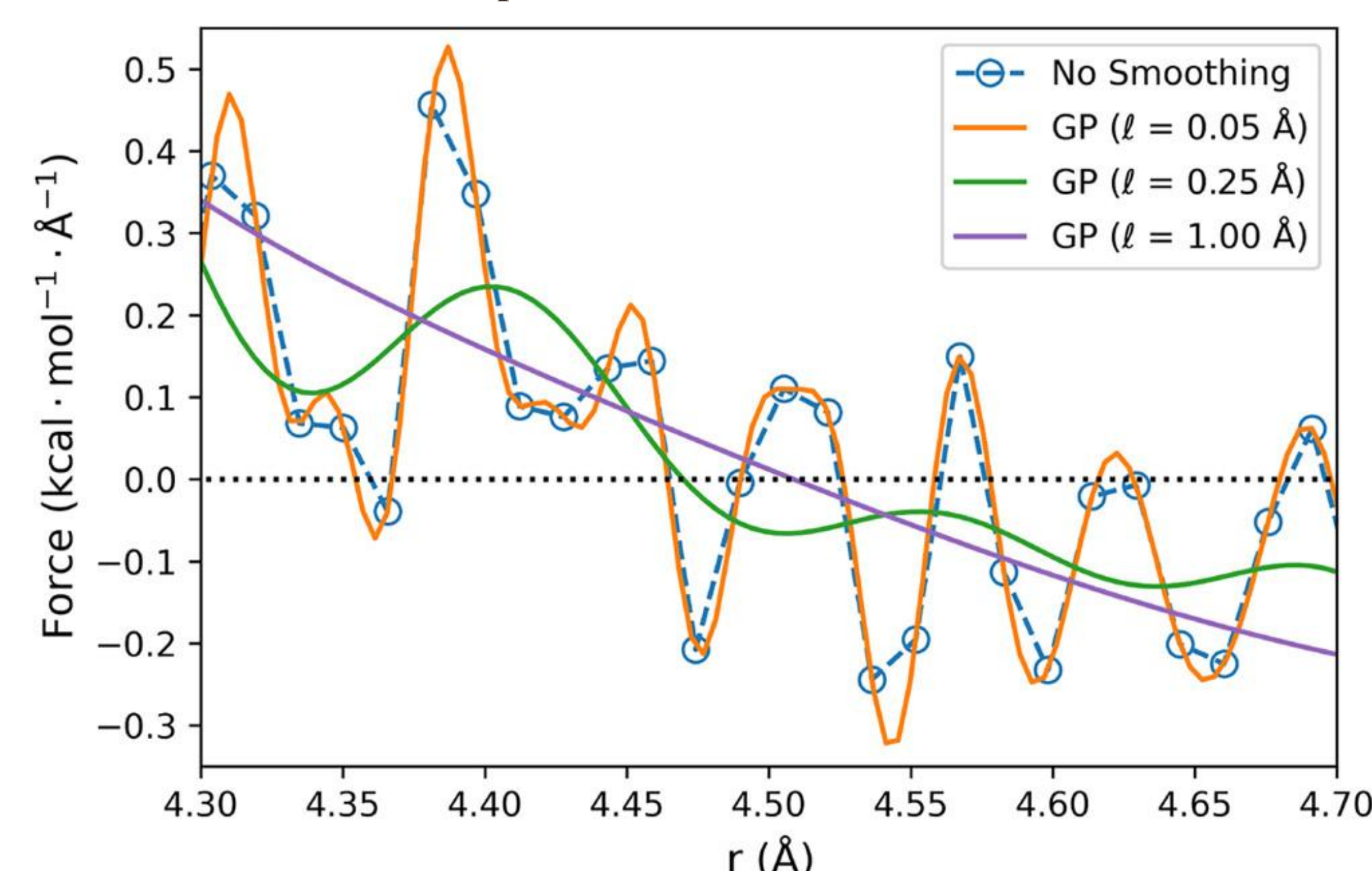
Gaussian processes are non-parametric, Bayesian function approximators that can model complex functions without overfitting.

We apply a local Gaussian process surrogate model to speed up Bayesian optimization of molecular simulation force fields (~5,000x speed-up).

The GP estimates a structure factor from a given set of FF parameters.

$$S^{*l}(q_k) = \mathbb{E}[GP_k(\mathbf{x}^{*l})] = \mathbf{k}_{\mathbf{x}^{*l}, \hat{\mathbf{x}}'} [\mathbf{k}_{\hat{\mathbf{x}}', \hat{\mathbf{x}}'} + \sigma_{noise}^2 \mathbf{I}]^{-1} \hat{\mathbf{Y}}_k'$$

Example of a Gaussian Process [2]



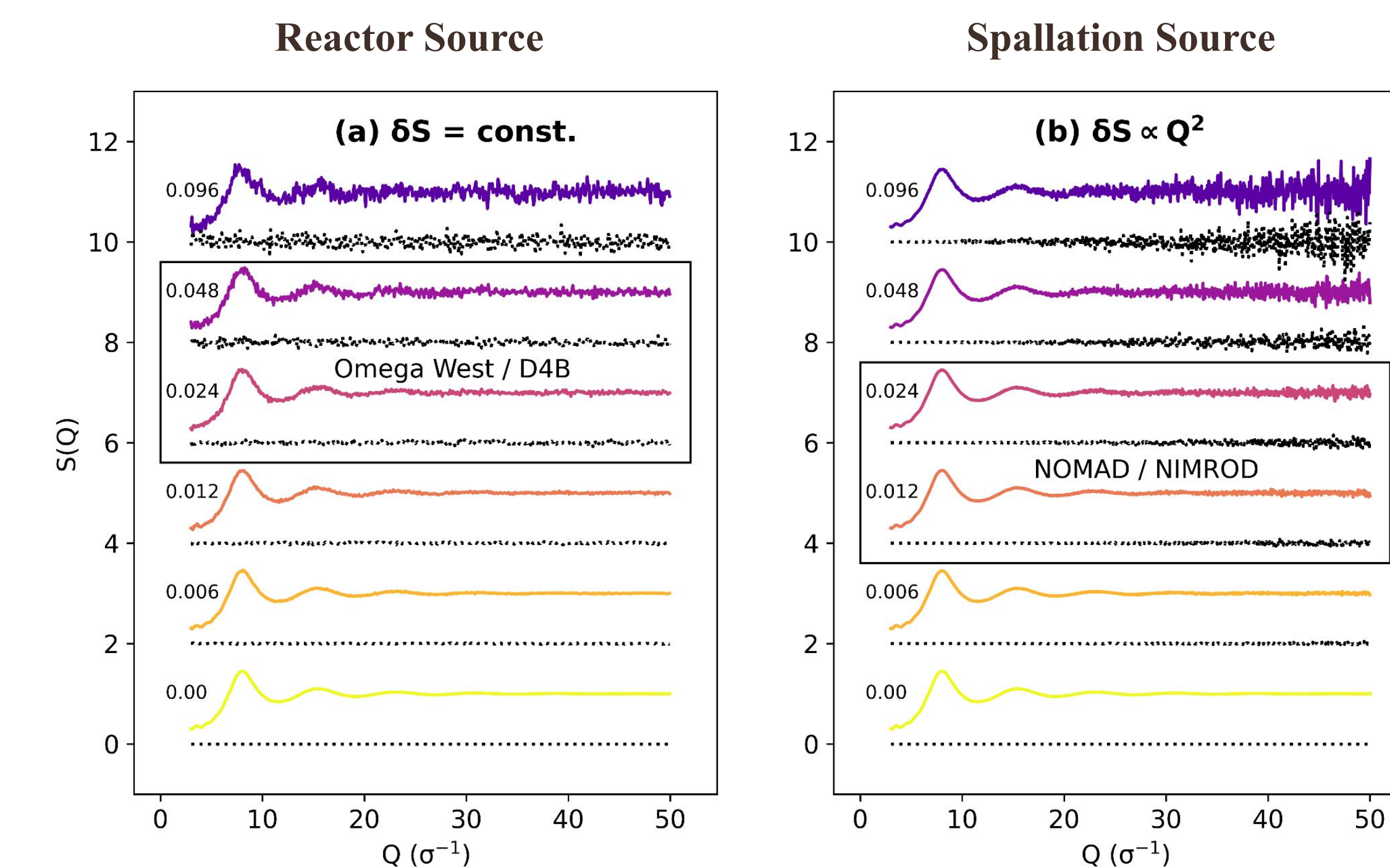
Can We Reconstruct Force Fields from Scattering?

We don't fully understand if/and/or/how uncertainty in experimental structure prevents the prediction of accurate force field parameters.

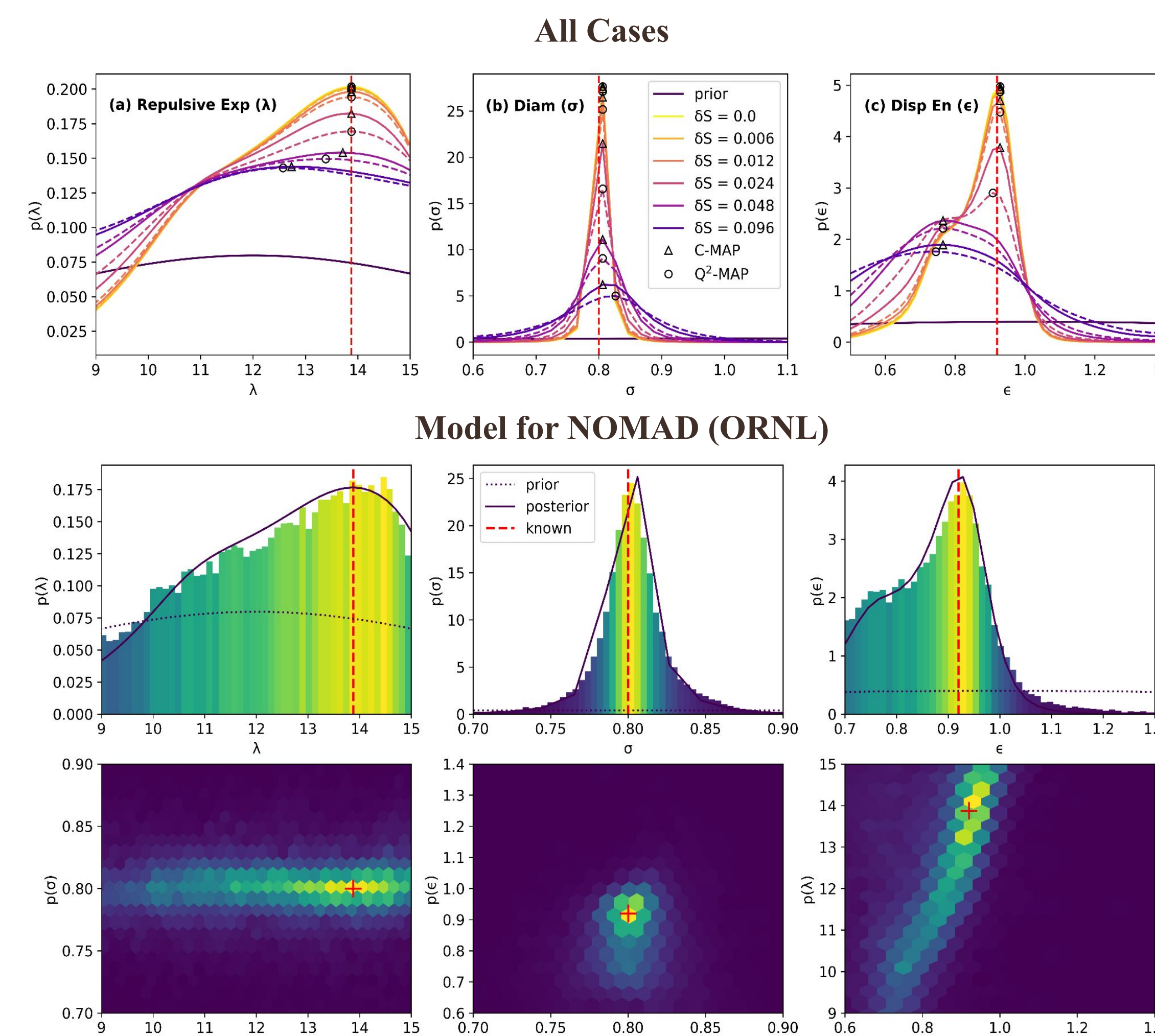
We first study this question in a model system using Bayesian optimization paired with our rapid and reliable Gaussian Process surrogate model.

Bayesian Analysis of the (n-6) Mie Force Field

Structure factors were generated for a (n-6) Mie fluid model based on existing experimental neutron instrument uncertainty.



Bayesian optimization of the (n-6) Mie force field parameters was performed given these structure factors as the "experimental" target data.



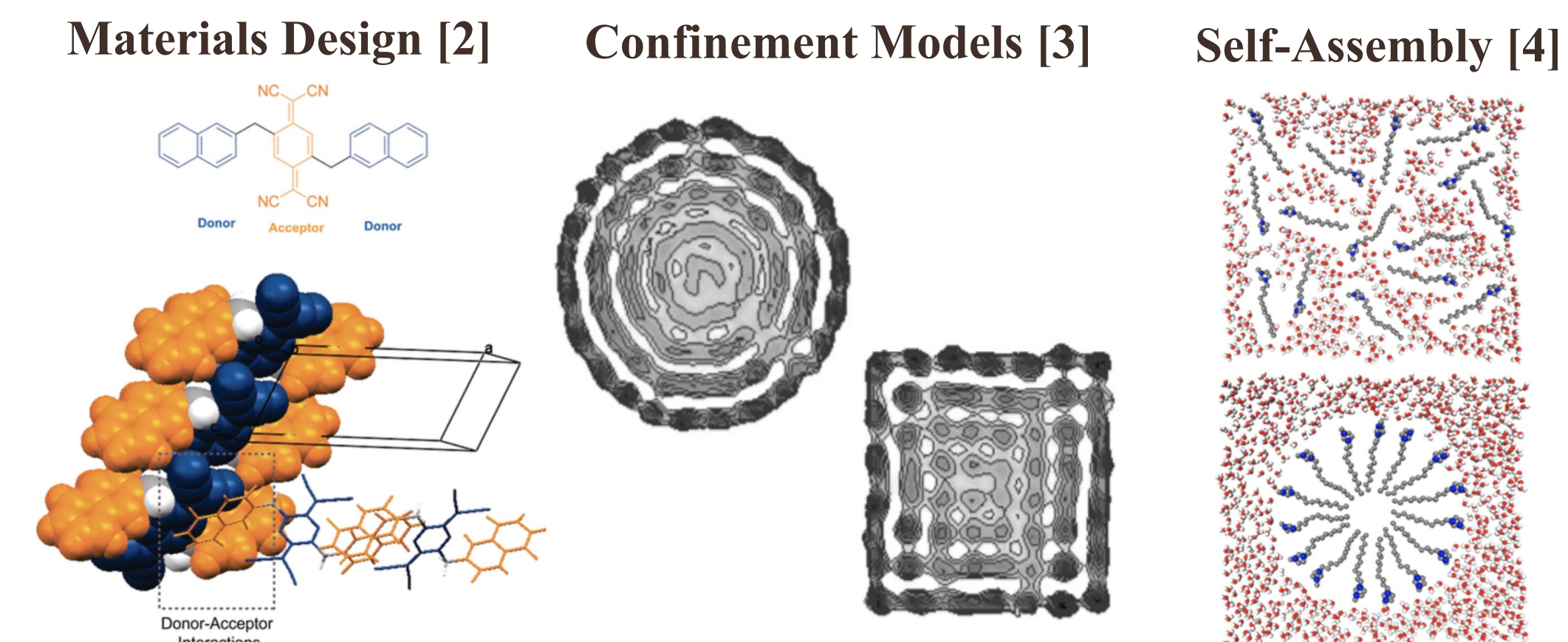
We find that force field parameters for Mie-type fluids can (in practice) be accurately recovered from structure data from current neutron instruments.

Local maxima in PDFs challenges the general view that the structure factor is insensitive to dispersion energy and repulsive exponent.

Bayesian optimization is an accurate technique for structure inversion.

Broader Perspectives and Applications

1. Building Novel Force Fields for Structure Based Applications



2. Coarse-Graining

3. Neutron Diffraction Analysis

4. Quantifying Many-Body Effects in Fluid Ensembles

General Conclusions

1. Local GP surrogate models enable Bayesian optimization to high-dimensional neutron diffraction measurements.
2. Low uncertainty structure factors contain sufficient information to extract local force parameters in Mie-type fluids.
3. Existing neutron instruments (NOMAD/NIMROD) are accurate enough for force field applications, **threshold accuracy is approximately 0.01 at 30 Å⁻¹**.
4. Only collision diameter is predicted accurately for high-uncertainty instruments (Omega West/D4B) within our model.
5. Possibly explains why early studies concluded that only particle size impacted structure factor measurements and structure factor was insensitive to the interatomic interactions.

Remaining Scientific Questions

1. Can Bayesian optimization be extended to molecular liquids / mixtures?
2. How do posterior distributions depend on thermodynamic state?
3. How accurate are structure-based force fields at reproducing other thermodynamic properties?

References

1. Shanks, B. L., Potoff, J. J. & Hoepfner, M. P. Transferable Force Fields from Experimental Scattering Data with Machine Learning Assisted Structure Refinement. *J. Phys. Chem. Lett.* 11512–11520 (2022) doi:10.1021/acs.jpclett.2c03163.
2. Venkataraman, D., Yurt, S., Venkataraman, B. H. & Gavvalapalli, N. Role of Molecular Architecture in Organic Photovoltaic Cells. *J. Phys. Chem. Lett.* 1, 947–958 (2010).
3. Elmejlawy, Mahmoud A. Thermodynamic Modeling of Hydrocarbons Confined in Mesopores: Insights on Entropy, Density, and Phase Behavior. University of Utah Thesis, 2022.
4. Vicent-Luna, J. M., Romero-Enrique, J. M., Calero, S. & Anta, J. A. Micelle Formation in Aqueous Solutions of Room Temperature Ionic Liquids: A Molecular Dynamics Study. *J. Phys. Chem. B* 121, 8348–8358 (2017).

Acknowledgements

Funding: National Science Foundation Award No. CBET-1847340

Contributors:

1. Brennon L. Shanks: Conceptualization, coding, writing, poster preparation and presentation
2. Harry W. Sullivan: Conceptualization, coding
3. Michael P. Hoepfner: Funding acquisition, conceptualization, poster preparation and editing

