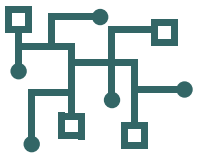


ISE

Industrial and
Systems Engineering

Learning Coarse-Grained Potentials for Binary Fluids

PEIYUAN GAO[1], XIU YANG[2], AND ALEXANDRE M. TARTAKOVSKY[1]

[1]Advanced Computing, Mathematics, and Data Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

[2]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

ISE Technical Report 20T-032

LEHIGH
UNIVERSITY

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

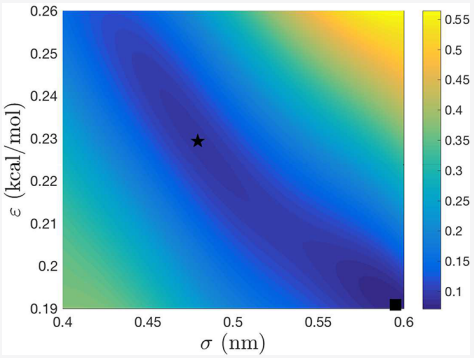
Learning Coarse-Grained Potentials for Binary Fluids

Peiyuan Gao,* Xiu Yang, and Alexandre M. Tartakovsky*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c00337 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: For a multiple-fluid system, CG models capable of accurately predicting the interfacial properties as a function of curvature are still lacking. In this work, we propose a new probabilistic machine learning (ML) model for learning CG potentials for binary fluids. The water−hexane mixture is selected as a typical immiscible binary liquid−liquid system. We develop a new CG force field (FF) using the Shinoda-DeVane-Klein (SDK) FF framework and compute parameters in this CG FF using the proposed probabilistic ML method. It is shown that a standard response-surface approach does not provide a unique set of parameters, as it results in a loss function with multiple shallow minima. To address this challenge, we develop a probabilistic ML approach where we compute the probability density function (PDF) of parameters that minimize the loss function. The PDF has a well-defined peak corresponding to a unique set of parameters in the CG FF that reproduces the desired properties of a liquid−liquid interface. We compare the performance of the new CG FF with several existing FFs for the water−hexane mixture, including two atomistic and three CG FFs with respect to modeling the interface structure and thermodynamic properties. It is demonstrated that the new FF significantly improves the CG model prediction of both the interfacial tension and structure for the water−hexane mixture.



1. INTRODUCTION

We propose a machine learning (ML) method for estimating parameters in coarse-grained (CG) force fields (FFs) and use it for constructing a CG FF for binary fluids. ML methods are often used to construct potential energy response surfaces using quantum chemistry calculations and parametrize atomistic interaction potentials.[1−5] Similar strategies were used to construct coarse-grained models.[6] In this paper, we propose a polynomial-regression-based ML method to construct a response surface that relates parameters in a CG FF to the (curvature−dependent) surface tension of the liquid−liquid interface. Then, the parameters can be identified by minimizing the loss function constructed as a mean square difference between surface tensions given by the response surface and measured in atomistic simulations. We demonstrate that for the considered problem, this approach does not provide a unique set of parameters because it results in a loss function with multiple shallow minima. To address this challenge, we propose a probabilistic approach where we compute the probability density function (PDF) of parameters minimizing the loss function. The PDF has a well-defined peak corresponding to a unique set of parameters in the CG FF that reproduces the desired properties of a liquid−liquid interface.

We focus on liquid−liquid interfaces because of their importance for many physical, chemical, and biological processes, including micelle formation, interfacial polymerization, and protein folding.[7−9] Therefore, understanding interfaces at the molecular level is fundamentally important. Various techniques, including spectroscopy and electrochemical measurement methods, have been proposed to experimentally study interfaces.[10−12] However, due to limited resolution, experimental methods cannot reveal the microscopic structures in full details.

Computational methods, including Molecular Dynamics (MD) techniques, have been used for studying aqueous interfaces since the 1980s,[13] but most of the MD studies were focused on liquid−vapor interfaces.[14,15] In contrast to liquid−vapor interfaces, MD simulations of liquid−liquid interfaces are more challenging because a larger system size is required to stabilize the flexible liquid phases and the interfacial region.[16] The computation of interfacial properties of large systems involving sampling of long time and large length scales remains a challenge for atomistic MD models.[17−21] CG models present an attractive alternative to atomistic MD models because of their ability to simulate much larger time and length scales.[22,23] In this regard, several coarse-graining approaches have been developed for the interface system by averaging over atomistic details and building CG FFs that can reproduce certain essential properties.[24,26] Generally, there are two ways to construct the CG FF, i.e., the bottom-up approach and the top-down approach.[25] In the bottom-up approach, the CG

Received: April 6, 2020
Published: July 15, 2020

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

potentials are extracted and constructed from the atomic system.[26,27] In other words, it focuses on the accurate modeling of the underlying atomistic structural details at a particular state point for a specific system.[28] The top-down CG models that macroscopic properties (e.g., thermodynamic data) are used as the main target of their parametrization are often cheaper and have better representability.[29−32] It is found that several CG FFs developed with the top-down CG approach can accurately reproduce multiple properties of industrial fluids. For example, the Shinoda-DeVane-Klein (SDK) CG FF and its extension SPICA (surface property fitting coarse graining) CG FF were shown to accurately model the surface tension, bulk density, and hydration free energy of water and alkanes.[33−38] The MARTINI FF, originally designed for lipids, surfactants, and biomacromolecules, was used to model the interface system.[23,39−42] The Statistical Associating Fluid Theory (SAFT) CG FF[19,43−48] was developed for many solvents, including water, alkanes, and carbon dioxide, where the effective CG intermolecular interactions between particles are estimated using an accurate description of the macroscopic experimental vapor−liquid equilibria data by means of a molecular-based equation of state. The above-mentioned CG FFs were shown to effectively describe multiple physical properties for some industrial fluids.

Parametrization methods for CG FFs for pure fluids are well established. However, for multiple-fluid systems, parametrization of FFs, especially the potentials acting between beads of different liquids, still remains a challenge. For example, transferable CG models that can reproduce the local structure and free energy in multiple-fluid systems with changing chemical environment are still lacking.[49] For the coarse-grained ML potential, once the coarse-graining map is defined, the definition of the energy function can be seen as a learning problem. In particular, the energy loss function or the force-matching loss function can be used to train the effective energy of the model from the atomistic energies or forces. Such an approach was used to design coarse-grained force fields for different systems with kernel methods[50,51] and deep neural networks.[52] However, few ML methods were focused on learning the CG potential using the top-down CG approach. In this work, we present a novel probabilistic ML method to estimate interaction parameters in the CG FF. This approach is applied to parametrizing a CG FF of a water−alkane system using interfacial tension as the target. We select the water−hexane mixture as it is a typical immiscible binary system and use the proposed CG model to study the interfacial properties of water−hexane as functions of the interface curvature. We demonstrate that the proposed parametrization improves the ability of CG models to predict the interfacial tension and interfacial structure as functions of the interface curvature, even though the interfacial structure is not used as a target in the CG model parametrization. This paper is organized as follows. Section 2 describes the atomic and CG models. Section 3 discusses the atomic and CG simulation results. Section 3.3 introduces the ML method and discusses its application to the water−hexane mixture. Section 4 presents the conclusions and outlook for CG modeling of complex liquid−liquid interfaces.

2. SIMULATION MODELS AND METHODS

2.1. Atomistic Model and Simulation. Several rigid water models have been proposed in the literature, but only the Transferable Intermolecular Potential with 4 Points 2005 (TIP4P2005) model was shown to accurately reproduce the temperature-dependent liquid−vapor surface tension.[53,54] Therefore, we employ the TIP4P2005 water model in our atomistic simulations. The Transferable Potentials for Phase Equilibria (TraPPE) FF[55] was shown to predict surface tension of alkanes in experiments.[55] Also, Neyt et al. demonstrated that the TIP4P2005 water and octane models combined in a TraPPE FF can reproduce the experimentally measured interfacial tension of a water−*n*-octane system.[41] Therefore, in this work, we employ the *n*-hexane model from the TraPPE FF. The interaction potential between the TIP4P2005-modeled water and the TraPPE-modeled alkane is modified following Ashbaugh's protocol.[56] This modification results in the more accurate hydration energy of alkane molecules in water and does not change other properties. In addition, the *n*-hexane model in the TraPPE FF is a united-atom model, where CH_3 and CH_2 groups are represented with a single united atom. Therefore, the interaction between the "TIP4P2005" water and TraPPE *n*-hexane does not include electrostatic interactions, which might affect the local structure of the interface. To study the effect of this potential on interfacial tension, we also test the hexane model in the Optimized Potential for Liquid Simulation All-Atom (OPLS-AA) FF.[57]

We consider the water−hexane mixture with planar and curved interfaces (see Figure 1). In our simulations of planar

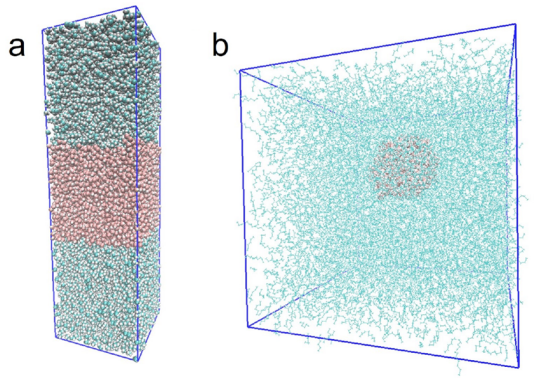


Figure 1. Initial state of the water−hexane planar interface and curved interface in atomic simulation: (a) planar interface and (b) curved interface.

interfaces, we put a pre-equilibrated water slab sandwiched between pre-equilibrated *n*-hexane slabs. The initial simulation box size is $L_x = L_y = 6$ nm and $L_z = 20$ nm. We place 8315 water molecules and 1152 hexane molecules for the TraPPE FF and 1140 hexane molecules for the OPLS-AA FF in the simulation box. Initially, the water and hexane molecules are separated by the plane interface. We also model a spherical water droplet in *n*-hexane with both the TIP4P2005 water−TraPPE *n*-hexane and the TIP4P2005 water−OPLS *n*-hexane models. We simulate droplets with radii of 2 nm (1026 water molecules) and 3 nm (3609 water molecules) in the simulation box with $L_x = L_y = L_z \approx 11$ nm and $L_x = L_y = L_z \approx 15$ nm, respectively. The box size is slightly adjusted during the equilibration process to keep pressure at 1 atm. For both the curved and planar interfaces, the long-range dispersion force correction method is used to obtain the correct density and pressure. These planar and droplet systems are equilibrated for 10 ns using the NP_NAT[58] ensemble (to keep the pressure constant, the box volume is changed by varying L_z) and the

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

NPT ensemble with a V-rescale thermostat and Berendsen barostat, respectively. The temperature and pressure are set to 310 K and 1 atm. Then we run another 10 ns simulation with the canonical ensemble at 310 K to collect data. All bonds between atoms are fixed by the LINCS algorithm.[59] Periodic boundary conditions are used in all three directions. The time step is 2 fs. The cutoffs for vdW interaction are 1.5 and 1.2 nm for the TraPPE and OPLS-AA FFs, respectively. The cutoff for Coulomb interaction is 1.2 nm for OPLS-AA FF. The particle mesh Ewald (PME) method[60] is used for calculating the long-range electrostatic interactions. All the atomistic simulations are performed with GROMACS.

2.2. CG Model and Simulations. We select the MARTINI (including the original and polarized water model) and SAFT CG FFs for modeling the water−hexane interface. The original MARTINI water model freezes under certain conditions.[39] To avoid water freezing, we replaced 12% of CG water beads with antifreeze CG water beads. Our simulation results show that the addition of antifreeze CG water beads does not affect the interfacial tension between water and *n*-hexane as long as the percentage of antifreeze CG water beads does not exceed 50%. For the polarized MARTINI water model, antifreezing CG water beads is not needed. There are several SAFT CG water models. Here, we employ the bio2 CG water model.

We build planar and curved interface systems for all considered CG FFs (see Figure 2). In the planar interface

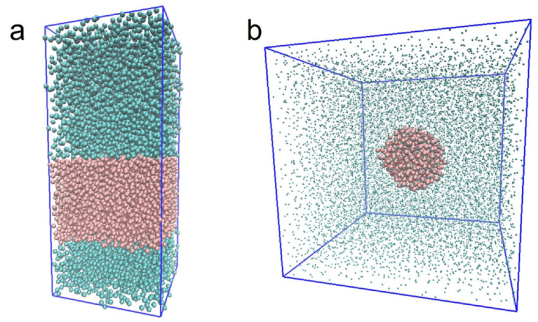


Figure 2. Initial state of the water−hexane planar interface and curved interface in CG simulation: (a) planar interface and (b) curved interface.

simulations, the simulation box size is set to $L_x = L_y > 5$ nm and $L_z > 11$ nm to avoid the boundary effect on the interfacial tension.[61] To study properties of the curved interface, we simulate 2 and 3 nm water droplets in *n*-hexane. To reduce the boundary effect, the initial lengths of the simulation box are set to 11 and 15 nm. The simulation boxes are equilibrated for 20 ns in NP_NAT and NPT ensembles at 310 K and 1 atm, respectively. Then, we perform 30 ns (planar interface) and 10 ns (curved interface) NVT simulations at 310 K for data collection. To get better statistics, we perform five parallel simulations for each curved interface system. The cutoffs for vdW interaction are 1.2 and 1.5 nm for the MARTINI and SAFT CG FFs, respectively. The cutoff for the Coulomb potential is 1.2 nm for the polarized water model in the MARTINI CG FF. The V-rescale thermostat and Berendsen barostat are used to keep constant temperature and pressure during pre-equilibrium. Then, the Nose-Hoover thermostat is employed in the production simulation. The time step is 10 fs. All CG simulations are performed with GROMACS.[16]

3. SIMULATION RESULTS

In this section, we investigate the density profiles, pressure profiles, and interfacial tensions of a water−hexane mixture with planar and curved interfaces using two atomistic and three CG FFs. Our analysis demonstrates that the interfacial structures and interfacial tensions obtained from the two atomistic models are in close agreement with each other. On the other hand, the three considered CG models produce results, which do not agree with each other nor the results of the atomistic models. Finally, we present a novel probabilistic ML approach for learning parameters in the CG FF and demonstrate that this FF significantly improves the prediction of both the interfacial tension and structure for the water−hexane mixture.

3.1. Planar Interface. *3.1.1. Density Profiles.* Here, we describe the interface structure of water−hexane systems using the intrinsic and nonintrinsic density profiles. The nonintrinsic or local mass density $\rho_N(x)$ is defined as the mass of liquid in a cube (centered at point x) divided by the cube's volume. Here, we use the cube size of 0.2 nm. The nonintrinsic density is averaged within each cube over time and over all cubes with the same normal distance to the interface. At the molecular level, the interface is corrugated by thermal capillary waves rather than being flat. To detect the molecular-level interface (so-called intrinsic interface), we use the so-called identification of the truly interfacial molecule (ITIM) method.[62−65] This method identifies interfacial molecules that are exposed to the opposite phase using a probe sphere with radius of 0.2 nm (see Figure 3). The probe sphere is moved along test lines

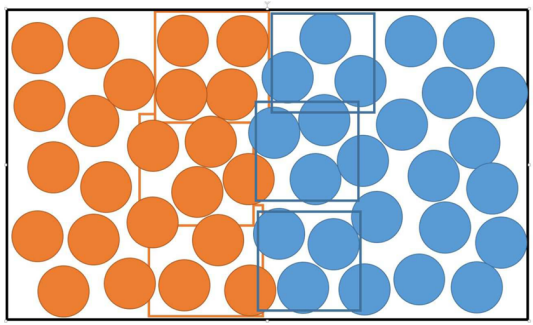


Figure 3. Scheme of the intrinsic density calculation for a water−hexane interface.

perpendicular to the plane of the fluid−fluid interface. Atoms that first encounter the probing ball are identified as the interfacial atoms, and the corresponding molecules are identified as the interfacial molecules. This process is repeated over the entire interfacial area in the simulation.

The intrinsic density provides more information about the interface structure (i.e., the location of the interface and the molecular organization) than the nonintrinsic density.[66] The nonintrinsic density profile is smooth and only contains approximate information about the interface location. The intrinsic density profile has local peaks corresponding to the locations of molecules layers near the interface, with the largest peak corresponding to the location of the interface.[62,67−69]

Figure 4 presents the intrinsic and nonintrinsic density profiles of the water and *n*-hexane of a water−hexane planar interface obtained from atomistic simulations with the TIP4P2005-TraPPE and TIP4P2005-OPLSAA models. Both atomistic models result in the same water density profiles and

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

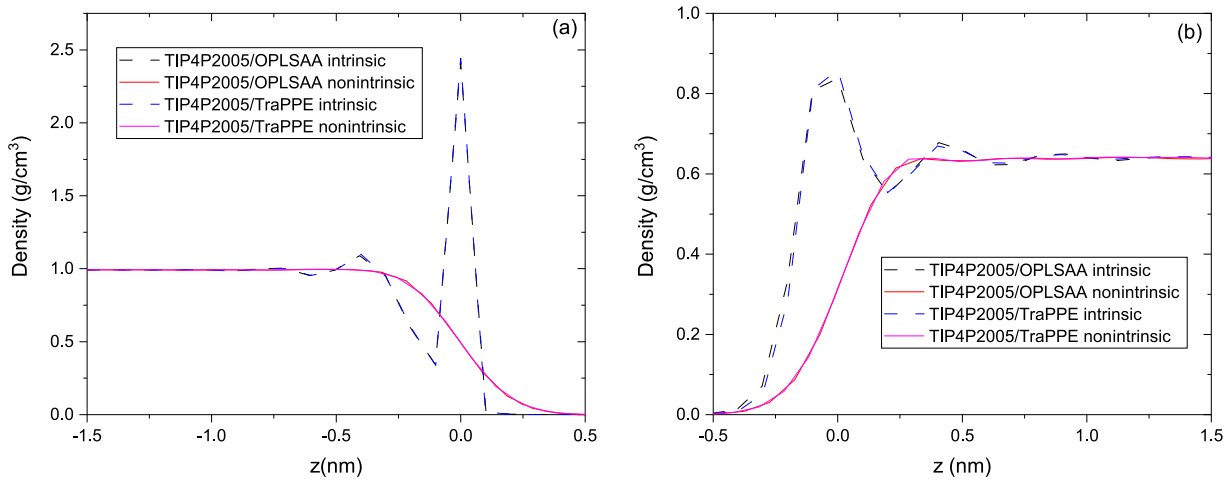


Figure 4. Intrinsic and nonintrinsic density profiles of (a) water and (b) hexane at the water−hexane interface as a function of z obtained from atomistic simulation. The point $z = 0$ corresponds to the position of the outermost water/hexane atoms in the intrinsic density profile and the Gibbs dividing surface of the water−hexane system for the nonintrinsic density profile.

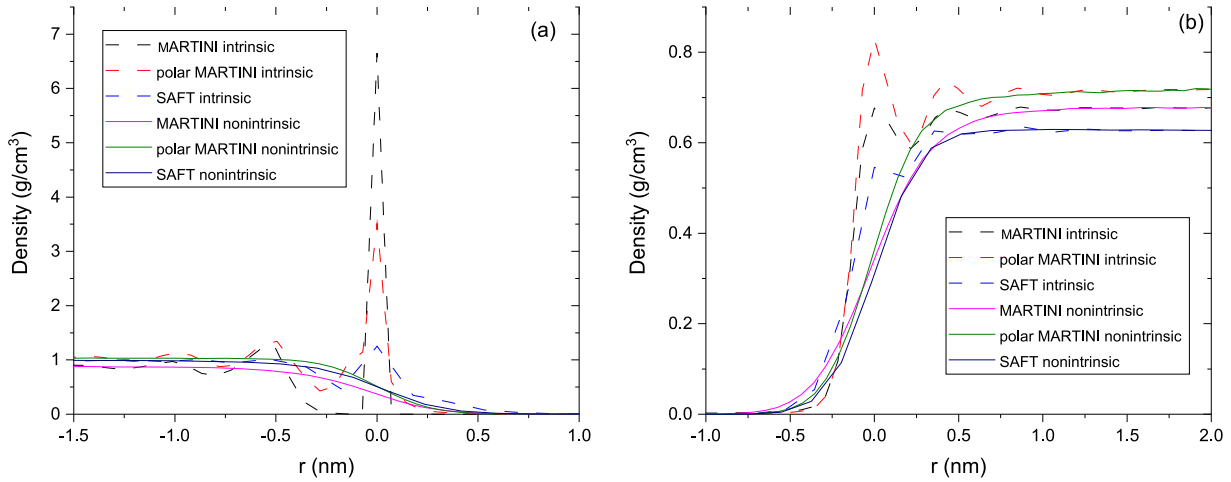


Figure 5. Intrinsic and nonintrinsic density profiles of (a) water and (b) hexane at the water−hexane interface in CG simulations. The zero point of the interface corresponds to the position of the outermost water/hexane atoms in the intrinsic density profile and the Gibbs dividing surface of the water−hexane system for the nonintrinsic density profile.

very similar hexane density profiles. Also, both atomistic models can reproduce the experimental density of water and hexane at 310 K. The intrinsic density profiles show that there are two water layers close to the interface. In addition, the strong directional bonding of water creates a well-defined correlation structure at short distances from the interface, but it does not propagate to longer distances as efficiently as it does for more packed liquid structures such as alkanes. The comparison of parts (a) and (b) of Figure 4 shows longer-range oscillations in alkanes than in water. Similar observations were made for a water−hexane binary system with the SPC/E water model.[70] In the case of hexane, we see that the distribution of the first peak is wider. This is due to the long tail of the alkane molecule. Overall, we find that the intrinsic structure of the water/n-hexane system is insensitive to atomistic FF parameters.

The density profiles in CG simulations are shown in Figure 5. The nonintrinsic and intrinsic densities of water and hexane are different for various CG FFs. The nonintrinsic density profile obtained with the SAFT CG FF is flatter than that obtained with the MARTINI CG FF. It should be noted that the density of water in the n-hexane phase is almost zero in the TIP4P2005 water model that is close to the experimental water density value of 6×10^{-5} g/cm^3.[71] In CG simulations, the density of water in hexane is 3×10^{-4} g/cm^3 for the MARTINI CG FF, which is approximately five times larger than the experimental value. For the SAFT bio2 CG water model in n-hexane, the water density is even greater. In Figure 5(a), the intrinsic water density profile has three peaks (note that two peaks were observed in MD simulations). This indicates that the CG water phase shows a longer-range ordered structure compared to the atomistic simulations. The intrinsic density profiles are similar for the original and polarized MARTINI CG water models, except that the original MARTINI CG water model has a higher interfacial density. The first peak in the SAFT bio2 CG model is lower than in the atomistic models because the CG model produces a wider interface. The positions of the first intrinsic density peaks for the CG n-hexane models are also very close. The hexane intrinsic density profiles, obtained from the MARTINI and SAFT CG n-hexane models, do not have distinct peaks (Figure 5(b)). However, we can observe a peak in the intrinsic density of hexane in the

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

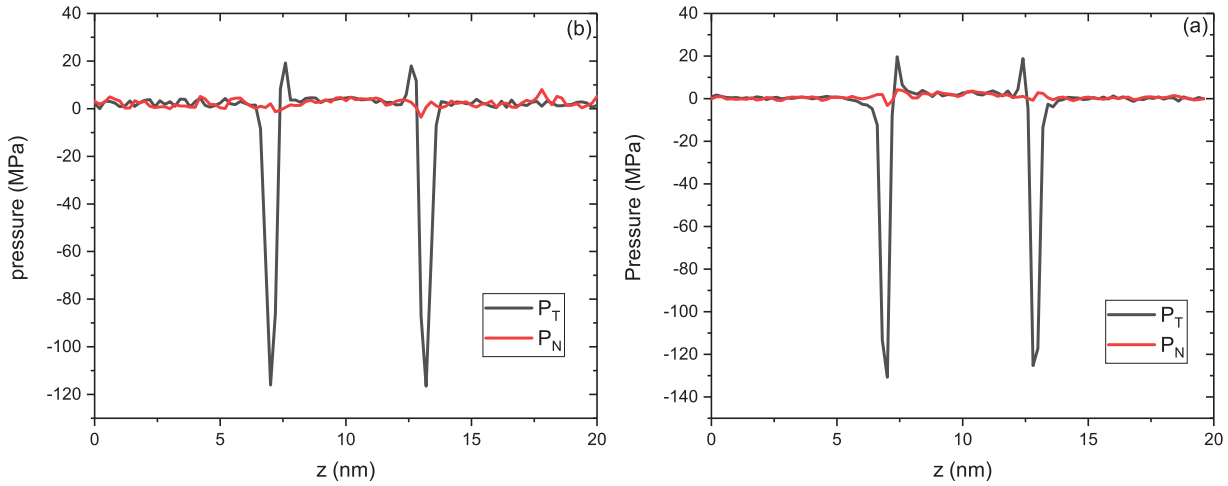


Figure 6. Pressure tensor components of the water−hexane planar interface in the atomistic (a) TIP4P2005 water + hexane in TraPPE FF and (b) TIP4P2005 water + hexane in OPLS-AA FF models.

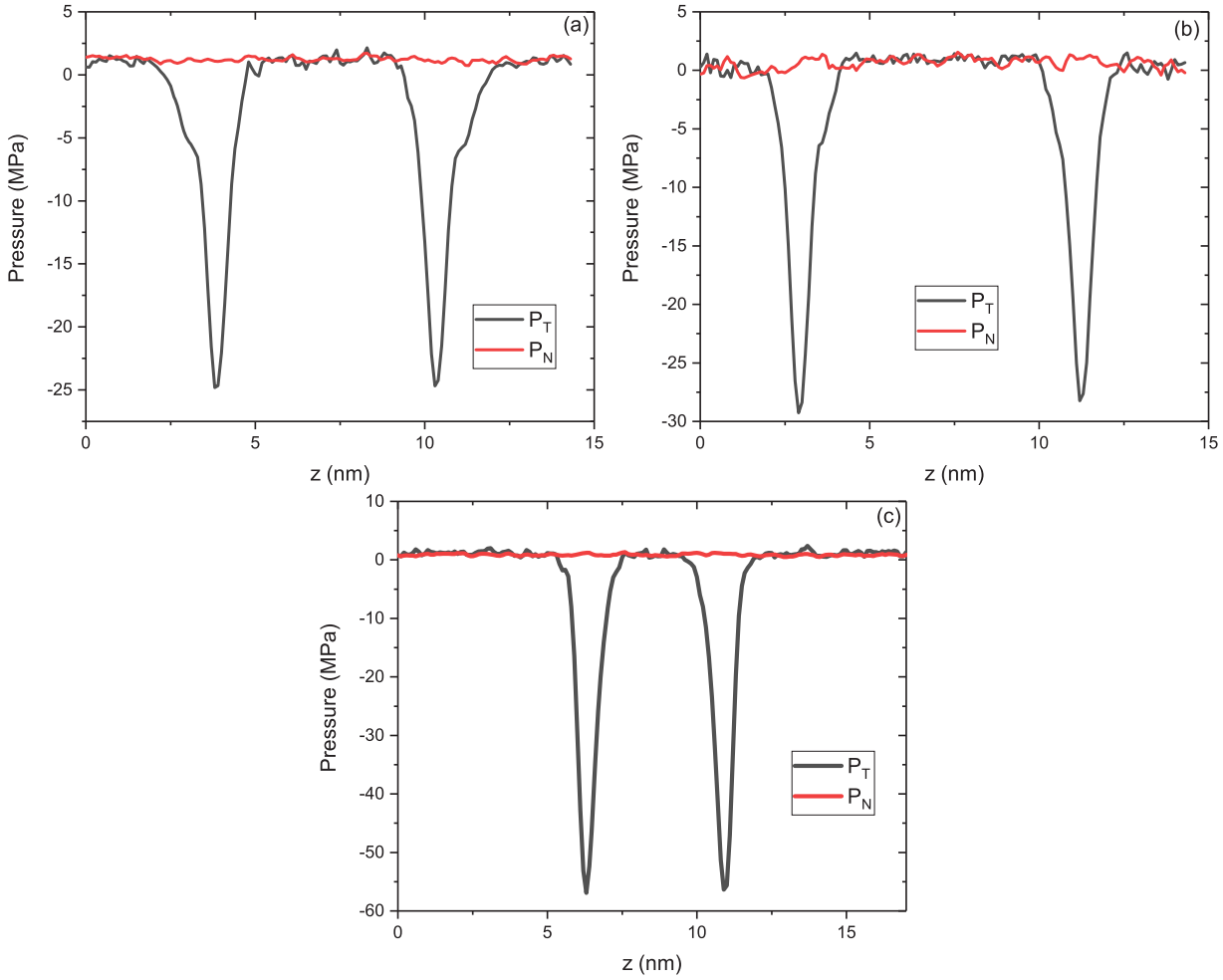


Figure 7. Local pressure components of the water−hexane planar interface in (a) MARTINI FF, (b) MARTINI FF with polarized water, and (c) SAFT FF models.

polarized MARTINI CG model. This is because single CG beads are used for both the MARTINI and SAFT bio2 CG water models. On the other hand, the polarized MARTINI CG water model has a physics-based, three-point structure. Bresme et al. demonstrated that the packing of water molecules will influence the orientation of alkane molecules at the interface.[68] In our CG simulations, we also see that the geometry topology constraints of the CG water model affect the local interface structure of the hexane phase. For the SAFT *n*-hexane model, the CG water beads infiltrate into the hexane phase so deeply

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

that the density of the first peak is lower than that of the bulk phase. In Figure 5(a),(b), the intrinsic density profiles of water or hexane for the three CG models are all different, which illustrates that the intrinsic density profile is sensitive to the choice of water and *n*-hexane CG models.

3.1.2. Pressure Profiles and Interfacial Tension. Previous atomistic simulations demonstrated that the errors in the estimated surface tension and liquid density are closely correlated.[53,72] Therefore, the accurate prediction of density is very important in the calculation of surface tension. Above, we demonstrated that both of the considered atomistic water and hexane models can reproduce the liquid bulk density at 310 K. Here, we calculate the interfacial tension based on the mechanical approach.[73−77] The interfacial tension of a planar interface is computed as[75]

$$\gamma(z) = \int_{-z/2}^{z/2} (P_N(z) - P_T(z))dz \quad (1)$$

where P_N and P_T are the normal and tangential components of the pressure tensor along the normal direction to the surface. For a spheric droplet, the expression for the surface tension takes the form

$$\gamma(r) = \int_{0}^{\infty} (P_N(r) - P_T(r))dr \quad (2)$$

We use the Irving−Kirkwood[73] and Vanegas and Ollila[76,78] approaches for computing pressure components in eqs 1 and 2, respectively. These approaches were originally proposed for MD systems with pairwise interactions. To compute local pressure components due to three-body angular potentials, these potentials are decomposed into pairwise potentials by Vanegas's central force decomposition (CFD) method,[77] which has been implemented in a modified GROMACS code (http://mdstress.org/). The many-body electrostatic interactions are approximated as pairwise interactions. In our pressure calculations, the cutoff of the pairwise interactions is 2.0 nm. There are other methods to calculate surface tension including the thermodynamic methods.[79] We note that the mechanical and thermodynamic methods can yield different surface tension estimates, especially at the interfaces with large curvature. However, both of these methods predict surface tension to decrease with the the decreasing radius of curvature.[79]

The normal and tangent pressure tensor components as functions of z, obtained from the atomistic and CG simulations, are presented in Figures 6 and 7. Atomistic simulations produce two symmetrical positive stress regions in the tangent component of pressure that correspond to the two water−hexane interfaces (Figure 6). They both appear on the water side of the interfaces. A similar pressure profile was also observed in an atomistic simulation with the TIP3P water and CHARMM hexane models.[80] Water molecules cause interface polarization and the positive pressure region on the water side of the interface.

Calculated and experimentally determined interfacial tensions are listed in Table 1. Both atomistic models predict the interfacial tension within 5% of the experimental value. We note that the computational cost of the all-atom model (OPLS-AA FF) is about five times larger than that of the united-atom model (TraPPE FF). The SAFT CG FF can also reproduce the experimental interfacial tension. However, the interfacial tension predicted by the MARTINI CG FF is only half of the experiment value. A previous MARTINI CG FF simulation study of a water−octane system at 298 K also reported an approximately 25% error in the estimated interfacial tension.[41] In addition, we find that using the polarized MARTINI water model instead of the MARTINI water model only slightly improves the interfacial tension prediction.

Table 1. Interfacial Tensions γ_∞ of the Water−Hexane Planar Interface in the Atomistic and CG Simulations and the Experiment[81] at 310 K

model	interfacial tension (mN/m)
experiment	49.4
atomistic TIP4P2005 + TraPPE	52.4 ± 1.1
atomistic TIP4P2005 + OPLS-AA	52.1 ± 1.2
CG MARTINI	25.9 ± 1.0
CG polarized MARTINI	27.8 ± 1.2
CG SAFT	51.6 ± 1.1

3.2. Curved Interfaces. *3.2.1. Density Profiles.* The intrinsic and nonintrinsic density profiles of a 2 nm water droplet in *n*-hexane, obtained in the two atomistic models, are shown in Figure 8. There are two peaks in the intrinsic density profiles in both atomistic models, which is similar to what we observed in the planar interface atomistic simulations. However, the peaks at the curved interface are higher than those at the planar interface. Compared to Figure 4, we also see that the width of the first peak is narrower, implying that the first water layer on the droplet surface is thinner than the one at the planar interface.

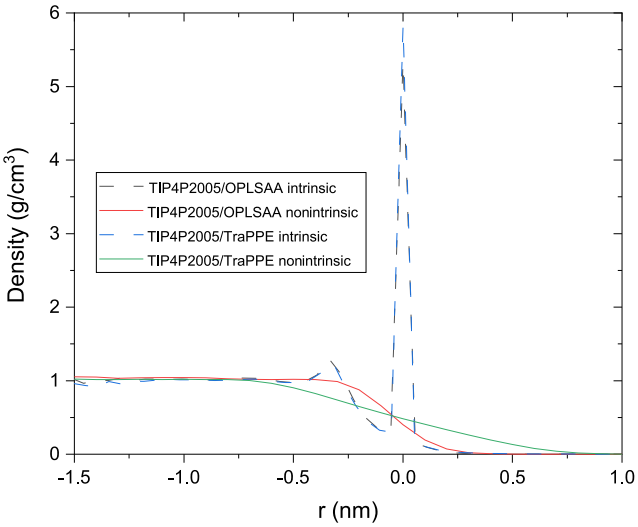


Figure 8. Intrinsic and nonintrinsic density profiles of a 2 nm water droplet in *n*-hexane obtained from atomistic simulations.

The CG water droplets show qualitatively different results. Figure 9 shows the density profiles of a 2 nm water droplet in *n*-hexane with various CG FFs. In the CG simulations of the planar interface, we see three density peaks on the water side. In Figure 9, the intrinsic density profile in the SAFT CG FF simulation has three peaks, while there are only two peaks in the MARTINI FF simulation. This could be caused by a larger cutoff in the SAFT CG FF. Both the CG and atomistic simulations show that the first peak in the water density profile is much higher for a curved interface than a planar interface.

3.2.2. Pressure Profiles and Interfacial Tension. Although it is widely accepted that the Laplace law, which relates the

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

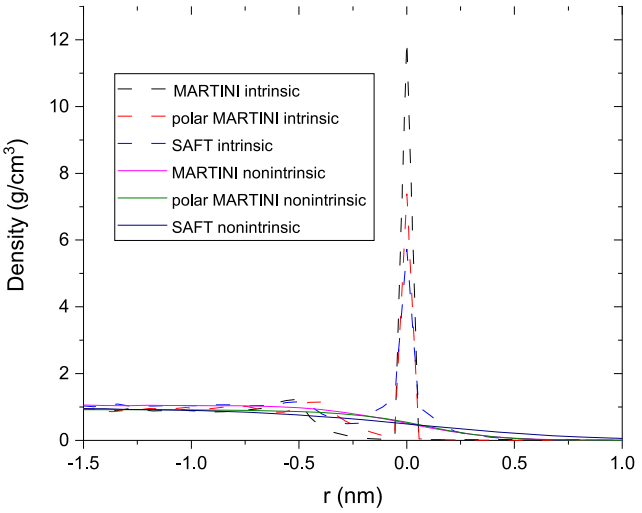


Figure 9. Intrinsic and nonintrinsic density profiles of a 2 nm water droplet in *n*-hexane obtained from CG simulations.

pressure jump across a curved interface to its curvature, fails for nanodroplets, the limit of the Laplace law validity is controversial. Takahashi and Morita concluded that this limit is less than 1 nm.[82] For liquid droplets in vapor environment, this limit was found to be between 5−10 nm.[79,83] Figures 10 and 11 show the normal and tangential components of the pressure tensor for a 2 nm water droplet in *n*-hexane. We see negative peaks in the tangent pressure profile at the interface in all simulations, indicating that the interface is under compression. Similar to the planar interface in atomistic simulations, we find a small peak on the water side of the tangent pressure in the droplet atomistic simulations. The pressure in the water droplet is greater than that in the hexane phase, which is consistent with the Laplace law. Comparing Figures 10 and 11, we find that the inner pressure in the atomistic simulations is higher than that in the CG simulations. In addition, electrostatic interactions in the MARTINI FF slightly increase the inner pressure, as shown in Figure 11(b). Table 2 lists the interfacial tensions of a 2 nm water droplet in *n*-hexane obtained from the atomistic and CG simulations. Both atomistic models result in a similar interfacial tension, which is smaller than the interfacial tension of the planar interface. Similar to the planar interface, the interfacial tension calculated with the MARTINI CG FFs is much smaller than that provided by the corresponding atomistic simulation. The SAFT CG FF, which is able to reproduce the interfacial tension of the planar interface, also results in a nearly 50% smaller interfacial tension than that in the atomistic simulations.

3.3. Stochastic ML Method for Estimating the Interaction Parameters in CG FFs. Our results in the previous section show that the MARTINI CG FF cannot reproduce the interfacial tension and density profile near the interface observed in our atomistic simulations. The SAFT CG FF can predict the interfacial tension of the planar interface but underestimates the interfacial tension of the curved interface by almost 50%. In addition, we find that the bio2 water model in SAFT CG FF overestimates the solubility of water in *n*-hexane. Therefore, we develop a new CG FF for the water−hexane system. We propose using the SDK CG FF[33] because it allows a lower degree of coarse graining, and the current SDK FF does not define the parameters between water and hexane for the low coarse-graining degree water model. We note that there is an SDK FF for the high coarse-graining degree water model, but we find that this water model may lead to crystallization of large water droplets.

In the remainder of this paper, we propose a new approach for learning coarse-grained potentials, apply it to estimating parameters in the water−hexane potential under the SDK CG FF framework, and test the resulting model for the water−hexane system against atomistic simulations. In this work, we use the 1:2 water model (one CG water bead represents two water molecules) and the 1:3 hexane model. The potential between CG water and hexane beads is given as

$$U_{\text{w-h}} = \left(\frac{\lambda_r}{\lambda_r - \lambda_a}\right)\left(\frac{\lambda_r}{\lambda_a}\right)^{\frac{\lambda_a}{\lambda_r - \lambda_a}} \varepsilon\left(\left(\frac{\sigma}{r}\right)^{\lambda_r} - \left(\frac{\sigma}{r}\right)^{\lambda_a}\right) \quad (3)$$

where λ_r and λ_a are repulsive and attractive exponents, respectively, ε is the energy parameter, and σ is the core diameter. The potentials $U_{\text{w-w}}$ and $U_{\text{h-h}}$ between water−water and hexane−hexane beads have the same form, with

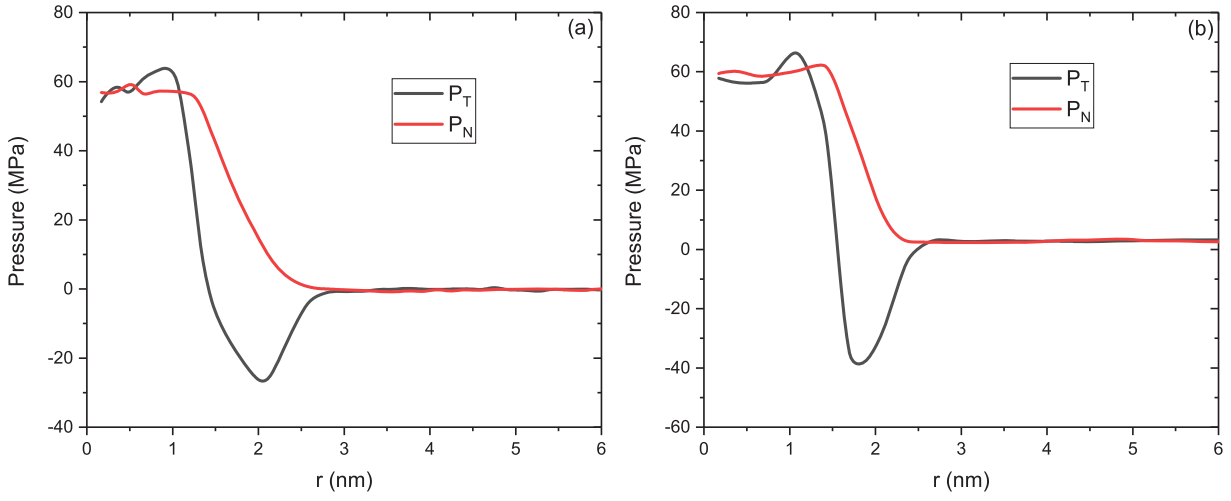


Figure 10. Pressure components as a function of the distance from the center of a 2 nm water droplet in *n*-hexane in atomistic FF simulations, including (a) the TIP4P2005 water model and *n*-hexane in TraPPE FF and (b) the TIP4P2005 water model and *n*-hexane in OPLS-AA FF.

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

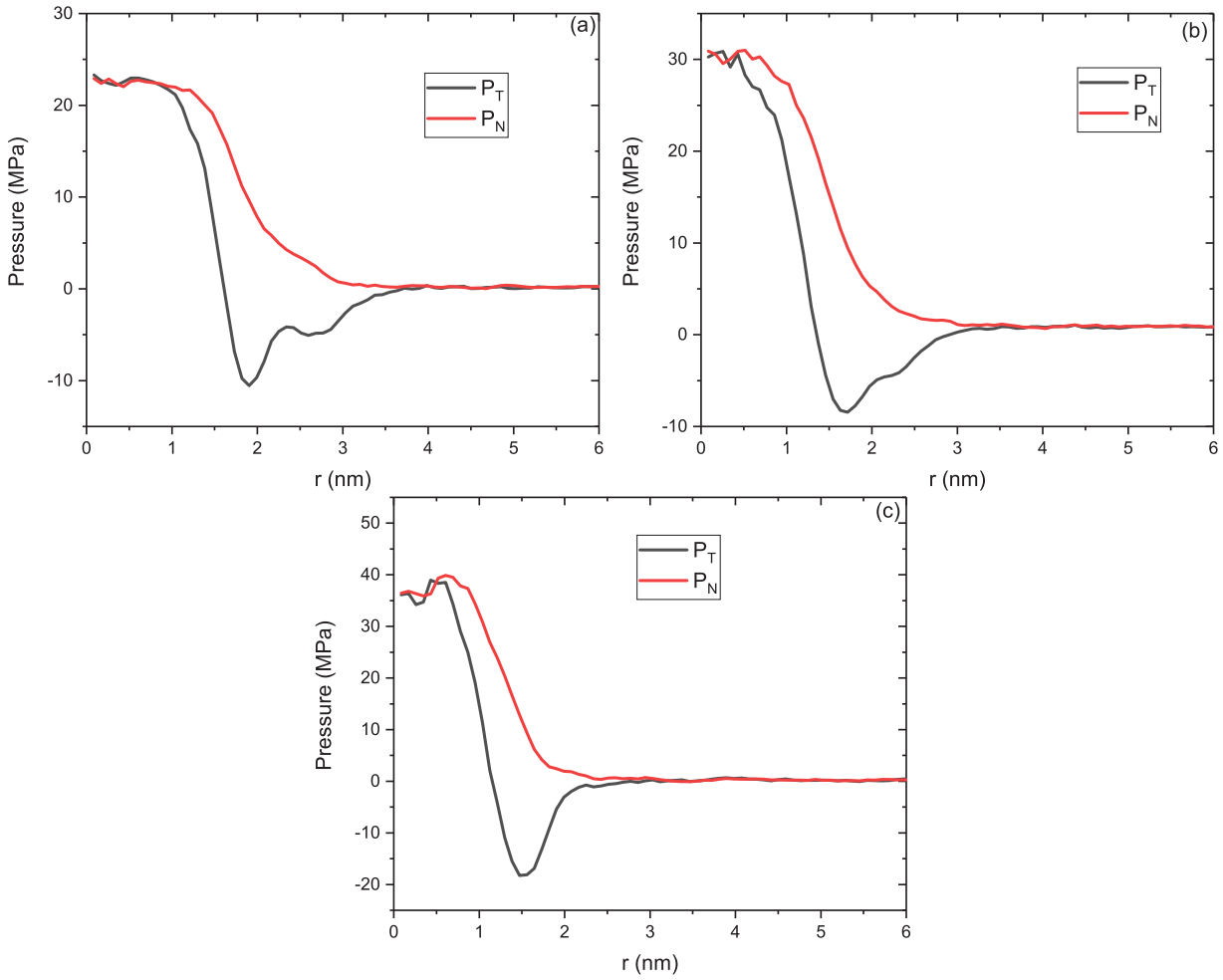


Figure 11. Pressure components as a function of the distance from the center of a 2 nm water droplet in *n*-hexane in CG simulations, including (a) MARTINI FF, (b) MARTINI FF with the polarized water model, and (c) SAFT FF.

Table 2. Interfacial Tensions γ_2 (r = 2 nm) of a Water Droplet in *n*-Hexane for Various Atomistic and CG FFs at 310 K

model	interfacial tension (mN/m)
atomistic TIP4P2005 + TraPPE	47.0 ± 1.1
atomistic TIP4P2005 + OPLS-AA	47.2 ± 1.9
CG MARTINI	21.4 ± 2.0
CG polarized MARTINI	23.9 ± 2.2
CG SAFT	27.7 ± 1.1

parameters λ_r, λ_a, σ, and ε listed in Table 3. In the original SDK framework, there are only two combinations of λ_r and λ_a, (λ_r = 12, λ_a = 4) and (λ_r = 9, λ_a = 6). The former combination results in a sharper interface because of the larger repulsive force corresponding to λ_r = 12. In the atomistic simulations, we observe a relatively sharp water−hexane interface. Therefore, in the $U_{w\text{-}h}$ potential, we set λ_r = 12 and λ_a = 4. Next, we learn the σ and ε parameters in the $U_{w\text{-}h}$ potential using the surface tension of the planar and curved water−hexane interfaces as target properties.

Table 3. CG Interaction Parameters of Water and Hexane[33,90]

CG model	λ_r	λ_a	ε (kcal/mol)	σ (nm)
water	9	6	0.7050	0.2908
hexane	9	6	0.4690	0.4585

We define the parameter vector $\theta = (\sigma,\varepsilon)^T$ and use polynomial regression (PR)[84−86] to construct a surrogate model of the interfacial tension as a function of θ. PR uses a linear combination of a set of orthogonal basis functions of θ to represent the quantity of interest (QoI) f

$$f(\theta) = \sum_{i=1}^{N} c_i \psi_i(\theta) \tag{4}$$

where ψ_i are basis functions (Legendre polynomials), and c_i are constant coefficients. Details of constructing multivariate Legendre polynomials and selection of N can be found in the Supporting Information. Here, f is the interfacial tension obtained from the atomistic simulations.

We search parameters in the space $\sigma \in [\sigma_{min}, \sigma_{max}]$ and $\varepsilon \in [\varepsilon_{min}, \varepsilon_{max}]$ and treat σ and ε as independent uniform random variables given by

$$\begin{pmatrix}\sigma \\ \varepsilon\end{pmatrix} = \begin{pmatrix}\bar{\sigma} \\ \bar{\varepsilon}\end{pmatrix} + (\delta_\sigma, \delta_\varepsilon)\cdot\begin{pmatrix}\xi_1 & 0 \\ 0 & \xi_2\end{pmatrix} \tag{5}$$

where $(\bar{\sigma},\bar{\varepsilon})$ = (0.5, 0.225) are the parameter means and (ξ_1, ξ_2) are independent random variables uniformly distributed on

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

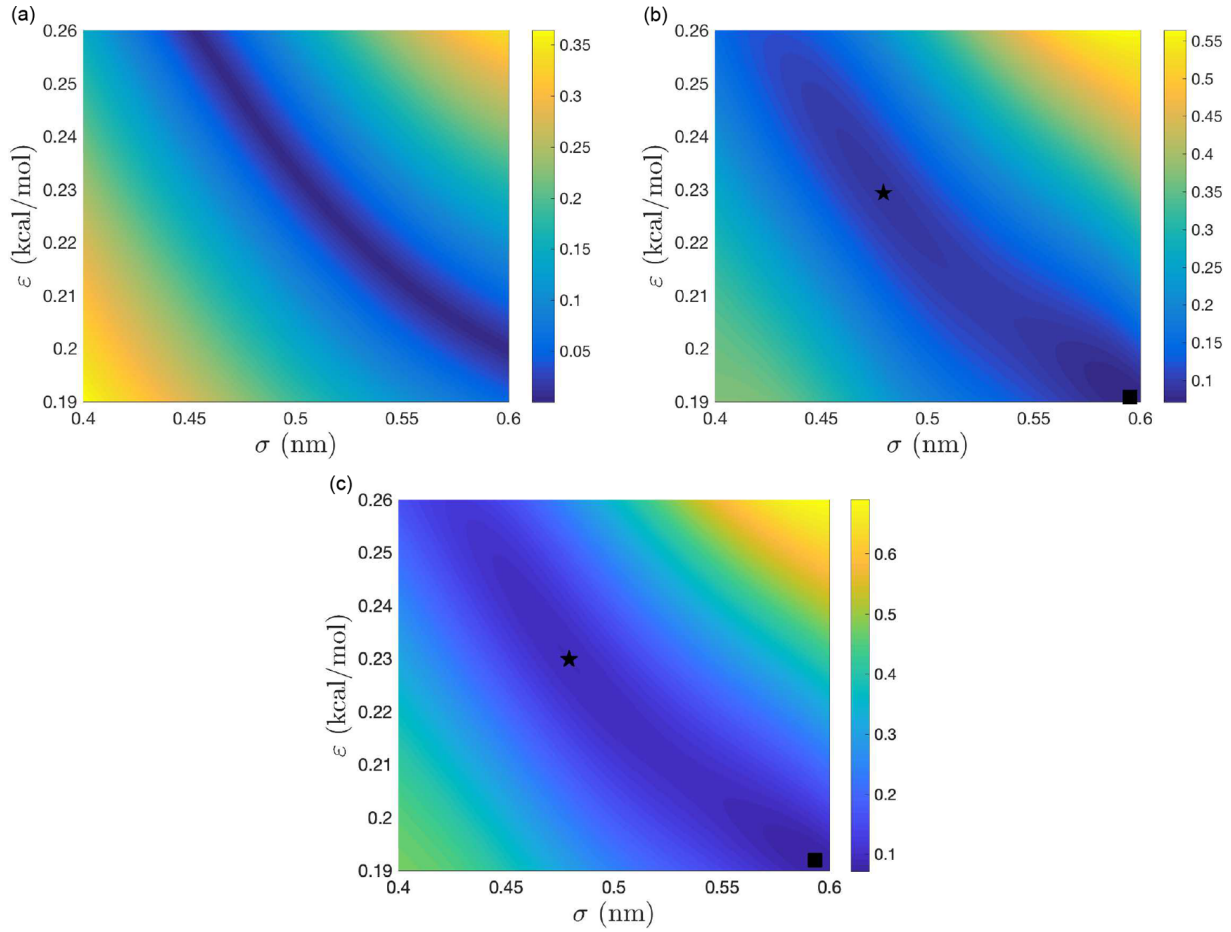


Figure 12. Loss functions L_1 (a), L_2 (b), and L_3 (c) as functions of σ and ε. Star and square symbols denote minima of the loss functions.

$[-1, 1]$. The $\overline{\sigma}$ and $\overline{\varepsilon}$ values are defined as an average of σ and ε in water−water and hexane−hexane potentials, respectively. The parameters $\delta_\sigma = 0.1$ and $\delta_\varepsilon = 0.035$ are found as $\delta_\sigma = (\sigma_{max} - \overline{\sigma})$ and $\delta_\varepsilon = (\varepsilon_{max} - \overline{\varepsilon})$, where $\sigma_{max} = 0.6$ is estimated as the maximum size of the water−hexane molecule cluster, and $\varepsilon_{max} = 0.26$ is estimated as the interaction energy between water and hexane. The parameters σ_{min} and ε_{min} are computed as $\sigma_{min} = 2\overline{\sigma} - \sigma_{max}$ and $\varepsilon_{min} = 2\overline{\varepsilon} - \varepsilon_{max}$, respectively. We generate 49 samples of ξ_1 and ξ_2 using the sparse grids method[87] with one-dimensional Gaussian quadrature points and the tensor product rule (i.e., the number of samples is equal to 7^d, where 7 is the number of one-dimensional quadrature points and d is the number of unknown parameters). Its distribution is shown in Figure S1. We then compute (σ, ε) for each sample (ξ_1, ξ_2) from eq 5, simulate the flat interface using the CG model for these values of (σ, ε), and compute the corresponding interfacial tension. The values of the interfacial tension are used to estimate the coefficients c_i in the PR surrogate model $f_\infty(\sigma, \varepsilon)$ based on the probabilistic collocation method.[88] Here, the subscript ∞ signifies that this is a response surface of a planar interface with the infinite radius of curvature. We find that the interfacial tension changes smoothly in the considered parameter space, and the relative error of the surrogate model, based on 10-fold cross-validation,[89] is less than 1%.

Finally, the surrogate model is used to find parameters σ and ε that correspond to the interfacial tension of the planar water−hexane interface in the atomistic simulation, by solving the minimization problem

$$(\sigma, \varepsilon) = \min_{\sigma,\varepsilon} L_1(\sigma, \varepsilon) \tag{6}$$

where

$$L_1(\sigma, \varepsilon) = \sqrt{\left(\frac{f_0(\sigma, \varepsilon) - \gamma_\infty}{\gamma_\infty}\right)^2} \tag{7}$$

is the "single target" loss function.

Figure 12(a) shows L_1 as a function of σ and ε. There is an infinite number of pairs (σ,ε) that generate γ_∞ lying on the curve $L_1(\sigma, \varepsilon) = 0$. To regularize parametrization, we select the interfacial tension of a 2 nm water droplet in hexane ($\gamma(2) = \gamma_2 = 47$ mN/m) as an additional constraint. We use the same 49 samples of random variables and the corresponding σ and ε to simulate a water droplet in hexane with the CG model. To reduce the statistical error caused by the thermal fluctuations of a small droplet, every sample is averaged over five independent CG simulations. These simulations are used to construct the surrogate model of the surface tension of a 2 nm droplet $f_2(\sigma, \varepsilon)$. Then, the optimal σ and ε are determined by solving the minimization problem

$$(\sigma, \varepsilon) = \min_{\sigma,\varepsilon} L_2(\sigma, \varepsilon) \tag{8}$$

where

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

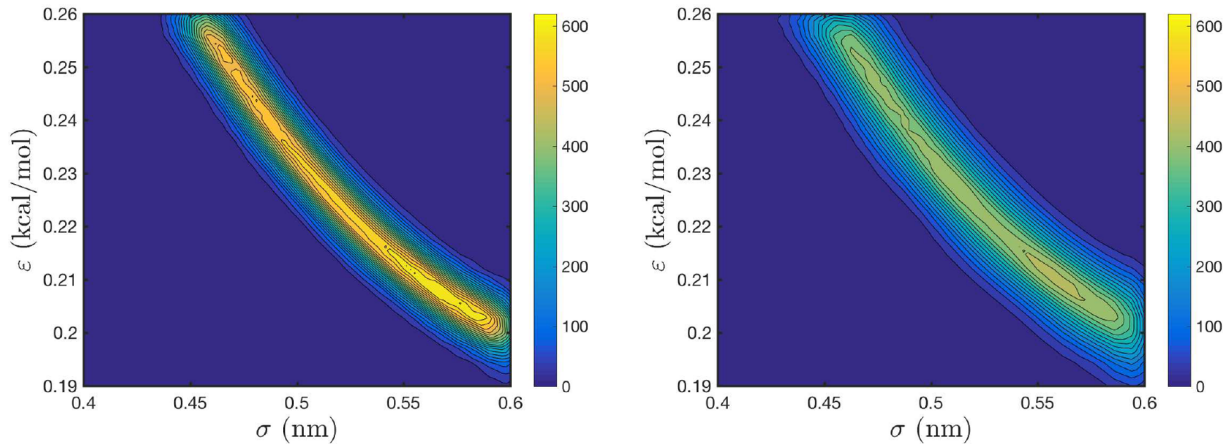


Figure 13. Single target: PDF function of the optimal parameters with (left) 4% Gaussian noise and (right) 8% Gaussian noise.

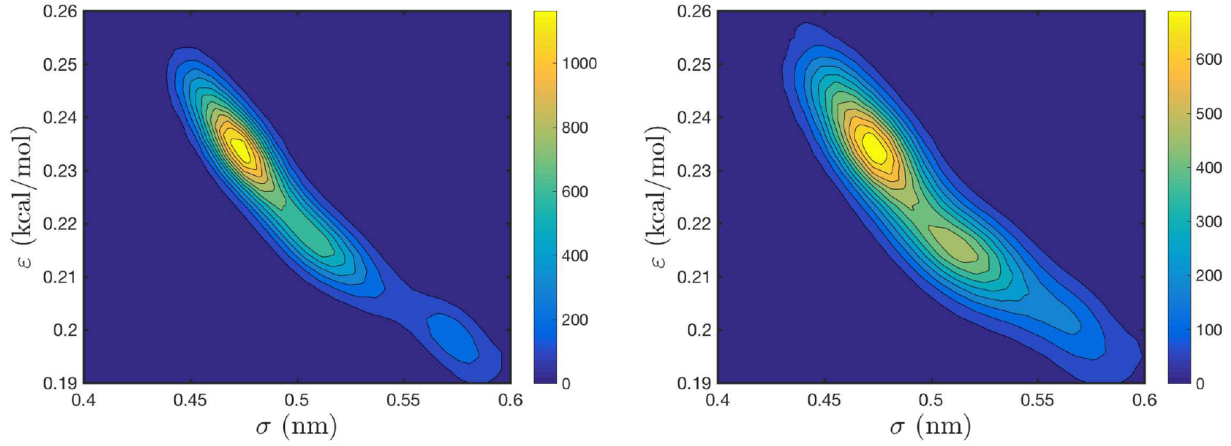


Figure 14. Two targets: PDF function of the optimal parameters with (left) 4% Gaussian noise and (right) 8% Gaussian noise.

$$L_2(\sigma, \varepsilon) = \left[\sqrt{\left(\frac{f_0(\sigma, \varepsilon) - \gamma_\infty}{\gamma_\infty} \right)^2 + \left(\frac{f_2(\sigma, \varepsilon) - \gamma_2}{\gamma_2} \right)^2} \right] \tag{9}$$

is the "two-target" loss function.

Figure 12(b) shows L_2 as a function of σ and ε. It can be seen that there are two sets of optimal parameters: ε = 0.23 kcal/mol and σ = 0.48 nm (the star) and ε = 0.19 kcal/mol and σ = 0.59 nm (the square). The difference between the response surface values at these two points is less than 2%, which is within the range of fluctuations observed in the CG simulations. (Note that if only the f_2 term is included in the loss function L_2, the result is similar to Figure 12(a). See Figure S2 for details.) To further regularize parametrization, we simulate a 3 nm water droplet in *n*-hexane using the atomistic MD model and use the resulting surface tension γ_3 = 49.1 mN/m as the third target to determine the parameters σ and ε as

$$(\sigma, \varepsilon) = \min_{\sigma,\varepsilon} L_3(\sigma, \varepsilon) \tag{10}$$

where

$$L_3(\sigma, \varepsilon) = \sqrt{\left(\frac{f_0(\sigma, \varepsilon) - \gamma_\infty}{\gamma_\infty} \right)^2 + \left(\frac{f_2(\sigma, \varepsilon) - \gamma_2}{\gamma_2} \right)^2 + \left(\frac{f_3(\sigma, \varepsilon) - \gamma_3}{\gamma_2} \right)^2} \tag{11}$$

is the "three-target" loss function, and $f_3(\sigma, \varepsilon)$ is the response surface of the surface tension of a 3 nm droplet as a function of σ and ε. Figure 12(c) shows that $L_3(\sigma, \varepsilon)$ still has two minima. This demonstrates that adding more targets does not make parametrization of this problem unique.

We propose a probabilistic approach to identify a unique set of parameters. We base our approach on the fact that the interfacial tension calculations from CG simulations are noisy due to the particle nature of the CG model. When the interfacial tension is used as a target to estimate parameters, these fluctuations (which can be treated as uncertainty) should be transferred to parameters. In our case, this requires knowledge of the interfacial tension sensitivity with respect to the parameters σ and ε. To perform the sensitivity analysis, we add 4% and 8% Gaussian noise to the values of the interfacial tension obtained from the 49 CG simulations, construct the surrogate model, and determine the optimal parameter set (ε, σ) as described above. We repeat this procedure 100,000 times and compute the probability density function (PDF) of the optimal (ε, σ).

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

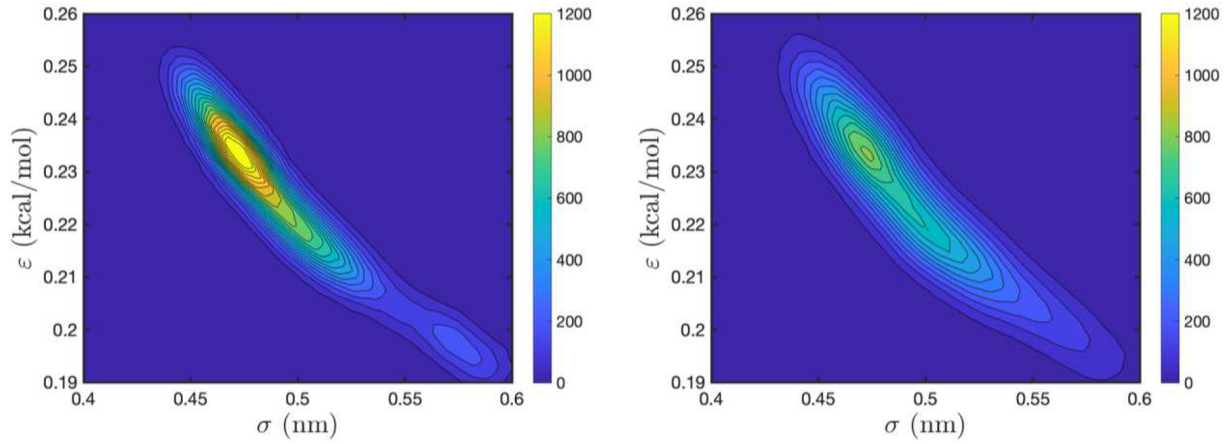


Figure 15. Three targets: PDF function of the optimal parameters with (left) 4% Gaussian noise and (right) 8% Gaussian noise.

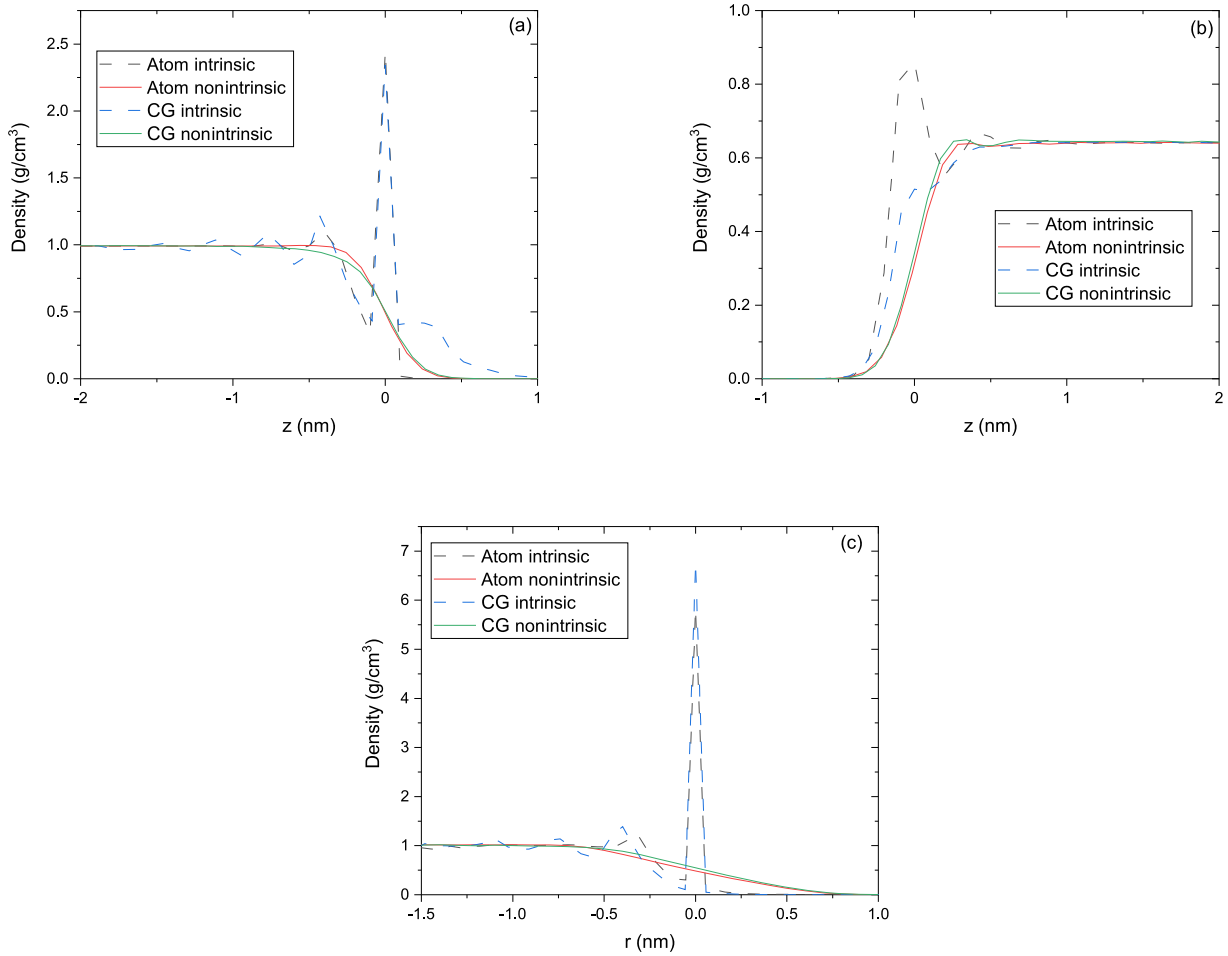


Figure 16. Intrinsic and nonintrinsic density profiles of a water−hexane interface obtained with atomistic TIP4P2005/TraPPE FF the CG FF with learned σ and ε: (a) water at the water−hexane planar interface, (b) hexane at the water−hexane planar interface, and (c) water at the water−hexane curved interface.

We evaluate PDFs of the (ε, σ) parameter sets, which minimize the $L_1(\varepsilon, \sigma)$, $L_2(\varepsilon, \sigma)$, and $L_3(\varepsilon, \sigma)$ loss functions. These PDFs are shown in Figures 13, 14, and 15, respectively. In Figure 13, the PDF has a "ridge" of most-probable parameters minimizing L_1. Therefore, we conclude that one target is not enough to uniquely determine the σ and ε parameters. In Figures 14 and 15, we see that the PDFs have a single sharp maximum at $(\varepsilon,\sigma) \approx (0.23,0.48)$ for both 4% and 8% added noise, with the peak in the 4% case being steeper than that in the 8% case. This demonstrates that the smaller noise (uncertainty) in the surface tension leads to a more certain estimate of the optimal parameters. The difference in the optimal value of parameters in Figures 14 and 15 is less than 1%. Therefore, in the proposed probabilistic approach,

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

two targets are sufficient to uniquely define the unknown parameters, while the standard response surface approach yields a nonunique result even with three targets. The squares in Figure 12(b),(c) correspond to a point with very small probability in Figures 14 and 15.

With (ε,σ) = (0.23,0.48) obtained from the PDF using *two* targets, the CG model produces an interfacial tension of 53.2 mN/m for the planar interface and 42.0 mN/m for the curved interface. These values are within 11% of the values obtained in atomistic simulations. Finally, we test the CG model by simulating a 3 nm water droplet in *n*-hexane and comparing the resulting surface tension with γ_3 found from the atomistic simulation. Note that we do not use γ_3 to obtain the parameters (ε,σ) = (0.23,0.48). We find that the interfacial tension of the 3 nm droplet is 45.1 mN/m, which is within 9% of the 49.1 mN/m interfacial tension value computed from the atomistic simulation of the 3 nm water droplet.

The intrinsic and nonintrinsic density profiles for planar and curved interfaces obtained from the CG model with (ε,σ) = (0.23,0.48) are presented in Figure 16. Figure 16(a),(b) shows that the width of the planar interface is very similar in the atomistic and CG simulations, as well as in the resulting nonintrinsic density profiles for both water and hexane. For the planar interface, there are three peaks in the "CG" intrinsic density profile of water, while only two peaks are observed in the "atomistic" intrinsic water density profile. The CG FF produces a longer-range ordered structure because it uses a larger cutoff than the atomistic FF. On the other hand, the locations and magnitudes of the first two peaks in the CG density profile are close to those in the atomistic simulations. Figure 16(c) also demonstrates good agreement between the intrinsic and nonintrinsic density profiles of a 2 nm water droplet (i.e., curved interface) obtained with our CG model and the atomistic models in terms of the bulk water density, interface width, and structure. There are some disagreements in the intrinsic density profiles of hexane in the CG and atomistic simulations. There is a relatively small peak in the intrinsic atomistic hexane density profile and no apparent peak in the CG intrinsic hexane density profile.

This disagreement is caused by the coarse graining of the one-site CG water and two-site CG hexane models.

4. CONCLUSION

We developed a new probabilistic machine learning framework that combines the polynomial-regression-based response surface with the uncertainty analysis resulting in the probability density function of optimal CG FF parameters. Optimal parameters are defined as parameters that minimize the difference between the CG and atomistic predictions of target properties (here, the interfacial tension of planar and curved interfaces). We applied the probabilistic ML method to parametrize the CG FF for a water−hexane mixture and demonstrated that a such constructed CG FF has better transferability than existing CG FFs with respect to the interface curvature. Specifically, it is more accurate than existing CG FFs for predicting the interfacial tension and coexisting densities as functions of the interface curvature.

First, we tested two existing atomistic FFs (the TIP4P2005 water model and OPLS-AA FF and the TIP4P2005 water model and TraPPE FF) and three existing CG FFs (MARTINI FF, polarized MARTINI FF, and SAFT FF) for a water−hexane mixture. The interface structure and thermodynamic properties were calculated for the planar and curved interfaces (water droplets with a radius of 2 and 3 nm in hexane). We found that the simulation results of both atomistic FFs for the planar interface agree well with published experimental values. Next, the atomistic FFs were used to simulate curved interfaces, and the resulting density profiles and surface tension values were used as reference solutions. The atomistic simulations confirmed that at the considered length scales, the interfacial tension depends on the interface curvature.

Then, we tested three popular CG FFs and found that none of them can accurately reproduce the interfacial structure and interfacial tensions of the planar or curved water−hexane interfaces. Next, we used the proposed probabilistic ML approach to learn new CG interaction parameters σ and ε within the SDK CG FF framework. We chose the interfacial tensions of the planar interface and the curved interfaces (2 and 3 nm droplets) as target properties. It is found that the probabilistic approach produces a unique set of parameters with two targets (the surface tension of the planar interface and the 2 nm droplet) and used the third target (the surface tension of the 3 nm droplet) to validate the FF. The calculated interfacial tension is within 9% of the atomistic prediction. We also demonstrated that the intrinsic and nonintrinsic densities in the learned CG model are in close agreement with those observed in the atomistic simulations. Transferability of the proposed CG FF with respect to the interface curvature also suggests transferability with respect to the chemical composition of the fluid mixture as the interfacial curvature locally defines the ratio of water to hexane.

Finally, we demonstrated that the standard ML response surface approach does not produce a unit set of σ and ε parameters even when three targets are used. Therefore, the proposed probabilistic framework reduces the number of target properties required to uniquely parametrize CG FFs and, thus, significantly reduces the computational cost of CG model parametrization. In addition, our ML method is very efficient. It only took less than 1 s for current model training. The proposed approach is general and can be used for learning interaction parameters in both atomistic and CG models of complex systems with appropriate target properties.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00337.

Method details about constructing multivariate Legendre polynomials and selection of N, distribution of 49 points in training set (Figure S1), and loss function L'_1 as function of σ and ε (Figure S2) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Peiyuan Gao − *Advanced Computing, Mathematics, and Data Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0002-2906-6551; Email: peiyuan.gao@pnnl.gov

Alexandre M. Tartakovsky − *Advanced Computing, Mathematics, and Data Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* Email: alexandre.tartakovsky@pnnl.gov

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Author

Xiu Yang − *Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, Pennsylvania 18015, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c00337

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research as part of the Collaboratory on the Physics-Informed Learning Machines for Multiscale and Multiphysics Problems (PhILMs) project. Pacific Northwest National Laboratory is operated by Battelle for the DOE under Contract DE-AC05-76RL01830.

■ REFERENCES

(1) John, S. T.; Csanyi, G. Many-Body Coarse-Grained Interactions Using Gaussian Approximation Potentials. *J. Phys. Chem. B* **2017**, *121*, 10934−10949.

(2) Sidky, H.; Whitmer, J. K. Learning free energy landscapes using artificial neural networks. *J. Chem. Phys.* **2018**, *148*, 104111.

(3) Chmiela, S.; Sauceda, H. E.; Muller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.

(4) Tang, Y.-H.; Zhang, D.; Karniadakis, G. E. An atomistic fingerprint algorithm for learning ab initio molecular force fields. *J. Chem. Phys.* **2018**, *148*, 034101.

(5) Gong, Z.; Wu, Y.; Wu, L.; Sun, H. Predicting Thermodynamic Properties of Alkanes by High-Throughput Force Field Simulation and Machine Learning. *J. Chem. Inf. Model.* **2018**, *58*, 2502−2516.

(6) Noé, F.; Tkatchenko, A.; Müller, K.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361−390.

(7) Ishiyama, T.; Imamura, T.; Morita, A. Theoretical Studies of Structures and Vibrational Sum Frequency Generation Spectra at Aqueous Interfaces. *Chem. Rev.* **2014**, *114*, 8447−8470.

(8) Piradashvili, K.; Alexandrino, E. M.; Wurm, F. R.; Landfester, K. Reactions and Polymerizations at the Liquid−Liquid Interface. *Chem. Rev.* **2016**, *116*, 2141−2169.

(9) Liu, S. J.; Li, Q.; Shao, Y. H. Electrochemistry at micro- and nanoscopic liquid/liquid interfaces. *Chem. Soc. Rev.* **2011**, *40*, 2236−2253.

(10) Nagata, Y.; Ohto, T.; Backus, E. H. G.; Bonn, M. Molecular Modeling of Water Interfaces: From Molecular Spectroscopy to Thermodynamics. *J. Phys. Chem. B* **2016**, *120*, 3785−3796.

(11) Johnson, C. M.; Baldelli, S. Vibrational Sum Frequency Spectroscopy Studies of the Influence of Solutes and Phospholipids at Vapor/Water Interfaces Relevant to Biological and Environmental Systems. *Chem. Rev.* **2014**, *114*, 8416−8446.

(12) Winter, B. Liquid microjet for photoelectron spectroscopy. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2009**, *601*, 139−150.

(13) Lee, C.; McCammon, J. A.; Rossky, P. J. The structure of liquid water at an extended hydrophobic surface. *J. Chem. Phys.* **1984**, *80*, 4448−4455.

(14) Nair, A. R.; Sathian, S. P. Studies on the effect of curvature on the surface properties of nanodrops. *J. Mol. Liq.* **2014**, *195*, 248−254.

(15) Chiu, S.-W.; Scott, H. L.; Jakobsson, E. A Coarse-Grained Model Based on Morse Potential for Water and n-Alkanes. *J. Chem. Theory Comput.* **2010**, *6*, 851−863.

(16) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS molecule and liquid database. *Bioinformatics* **2012**, *28*, 752−753.

(17) Mayoral, E.; Goicochea, A. G. Modeling the temperature dependent interfacial tension between organic solvents and water using dissipative particle dynamics. *J. Chem. Phys.* **2013**, *138*, 094703.

(18) Yesylevskyy, S. O.; Schafer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, e1000810.

(19) Lobanova, O.; Avendano, C.; Lafitte, T.; Muller, E. A.; Jackson, G. SAFT-gamma force field for the simulation of molecular fluids: 4. A single-site coarse-grained model of water applicable over a wide temperature range. *Mol. Phys.* **2015**, *113*, 1228−1249.

(20) Ghoufi, A.; Malfreyt, P.; Tildesley, D. J. Computer modelling of the surface tension of the gas-liquid and liquid-liquid interface. *Chem. Soc. Rev.* **2016**, *45*, 1387−1409.

(21) Zubillaga, R. A.; Labastida, A.; Cruz, B.; Martinez, J. C.; Sanchez, E.; Alejandre, J. Surface Tension of Organic Liquids Using the OPLS/AA Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 1611−1615.

(22) Peter, C.; Kremer, K. Multiscale simulation of soft matter systems - from the atomistic to the coarse-grained level and back. *Soft Matter* **2009**, *5*, 4357−4366.

(23) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.

(24) Stephan, S.; Thol, M.; Vrabec, J.; Hasse, H. Thermophysical Properties of the Lennard-Jones Fluid: Database and Data Assessment. *J. Chem. Inf. Model.* **2019**, *59*, 4248−4265.

(25) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodríguez-Ropero, F.; van der Vegt, N. F. A. Systematic coarse-graining methods for soft matter simulations − a review. *Soft Matter* **2013**, *9*, 2108−2119.

(26) Garay, P. G.; Barrera, E. E.; Pantano, S. Post-Translational Modifications at the Coarse-Grained Level with the SIRAH Force Field. *J. Chem. Inf. Model.* **2020**, *60*, 964−973.

(27) Sanyal, T.; Shell, M. S. Transferable Coarse-Grained Models of Liquid−Liquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy. *J. Phys. Chem. B* **2018**, *122*, 5678−5693.

(28) Dunn, N. J. H.; Noid, W. G. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *J. Chem. Phys.* **2015**, *143*, 243148.

(29) Bejagam, K. K.; Singh, S.; An, Y.; Berry, C.; Deshmukh, S. A. PSO-Assisted Development of New Transferable Coarse-Grained Water Models. *J. Phys. Chem. B* **2018**, *122*, 1958−1971.

(30) Motevaselian, M. H.; Mashayak, S. Y.; Aluru, N. R. Extended coarse-grained dipole model for polar liquids: Application to bulk and confined water. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2018**, *98*, 052135.

(31) Molinero, V.; Moore, E. B. Water Modeled As an Intermediate Element between Carbon and Silicon. *J. Phys. Chem. B* **2009**, *113*, 4008−4016.

(32) Orsi, M. Comparative assessment of the ELBA coarse-grained model for water. *Mol. Phys.* **2014**, *112*, 1566−1576.

(33) Shinoda, W.; Devane, R.; Klein, M. L. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Mol. Simul.* **2007**, *33*, 27−36.

(34) Seo, S.; Shinoda, W. SPICA Force Field for Lipid Membranes: Domain Formation Induced by Cholesterol. *J. Chem. Theory Comput.* **2019**, *15*, 762−774.

(35) Shinoda, W.; DeVane, R.; Klein, M. L. Coarse-grained molecular modeling of non-ionic surfactant self-assembly. *Soft Matter* **2008**, *4*, 2454−2462.

(36) Shinoda, W.; DeVane, R.; Klein, M. L. Zwitterionic Lipid Assemblies: Molecular Dynamics Studies of Monolayers, Bilayers, and Vesicles Using a New Coarse Grain Force Field. *J. Phys. Chem. B* **2010**, *114*, 6836−6849.

(37) Shinoda, W.; DeVane, R.; Klein, M. L. Coarse-grained force field for ionic surfactants. *Soft Matter* **2011**, *7*, 6178−6186.

(38) Miyazaki, Y.; Okazaki, S.; Shinoda, W. pSPICA: A Coarse-Grained Force Field for Lipid Membranes Based on a Polar Water Model. *J. Chem. Theory Comput.* **2020**, *16*, 782−793.

(39) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(40) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−34.
(41) Neyt, J. C.; Wender, A.; Lachet, V.; Ghoufi, A.; Malfreyt, P. Quantitative Predictions of the Interfacial Tensions of Liquid-Liquid Interfaces through Atomistic and Coarse Grained Models. *J. Chem. Theory Comput.* **2014**, *10*, 1887−99.
(42) Ndao, M.; Devémy, J.; Ghoufi, A.; Malfreyt, P. Coarse-Graining the Liquid−Liquid Interfaces with the MARTINI Force Field: How Is the Interfacial Tension Reproduced? *J. Chem. Theory Comput.* **2015**, *11*, 3818−3828.
(43) Avendaño, C.; Lafitte, T.; Adjiman, C. S.; Galindo, A.; Müller, E. A.; Jackson, G. SAFT-gamma Force Field for the Simulation of Molecular Fluids: 2. Coarse-Grained Models of Greenhouse Gases, Refrigerants, and Long Alkanes. *J. Phys. Chem. B* **2013**, *117*, 2717−2733.
(44) Avendaño, C.; Lafitte, T.; Galindo, A.; Adjiman, C. S.; Jackson, G.; Müller, E. A. SAFT-gamma Force Field for the Simulation of Molecular Fluids. 1. A Single-Site Coarse Grained Model of Carbon Dioxide. *J. Phys. Chem. B* **2011**, *115*, 11154−11169.
(45) Lobanova, O.; Mejía, A.; Jackson, G.; Müller, E. A. SAFT-gamma force field for the simulation of molecular fluids 6: Binary and ternary mixtures comprising water, carbon dioxide, and n-alkanes. *J. Chem. Thermodyn.* **2016**, *93*, 320−336.
(46) Herdes, C.; Ervik, A.; Mejía, A.; Müller, E. A. Prediction of the water/oil interfacial tension from molecular simulations using the coarse-grained SAFT-gamma Mie force field. *Fluid Phase Equilib.* **2018**, *476*, 9−15.
(47) Muller, E. A.; Jackson, G. Force-Field Parameters from the SAFT-γ Equation of State for Use in Coarse-Grained Molecular Simulations. *Annu. Rev. Chem. Biomol. Eng.* **2014**, *5*, 405.
(48) Mejia, A.; Herdes, C.; Muller, E. A. Force Fields for Coarse-Grained Molecular Simulations from a Corresponding States Correlation. *Ind. Eng. Chem. Res.* **2014**, *53*, 4131−4141.
(49) Potter, T. D.; Tasche, J.; Wilson, M. R. Assessing the transferability of common top-down and bottom-up coarse-grained molecular models for molecular mixtures. *Phys. Chem. Chem. Phys.* **2019**, *21*, 1912−1927.
(50) Scherer, C.; Scheid, R.; Andrienko, D.; Bereau, T. Kernel-Based Machine Learning for Efficient Simulations of Molecular Liquids. *J. Chem. Theory Comput.* **2020**, *16*, 3194−3204.
(51) Wang, J.; Chmiela, S.; Müller, K.; Noé, F.; Clementi, C. Ensemble learning of coarse-grained molecular dynamics force fields with a kernel approach. *J. Chem. Phys.* **2020**, *152*, 194106.
(52) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755−767.
(53) Underwood, T. R.; Greenwell, H. C. The Water-Alkane Interface at Various NaCl Salt Concentrations: A Molecular Dynamics Study of the Readily Available Force Fields. *Sci. Rep.* **2018**, *8*, 352.
(54) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
(55) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(56) Ashbaugh, H. S.; Liu, L.; Surampudi, L. N. Optimization of linear and branched alkane interactions with water to simulate hydrophobic hydration. *J. Chem. Phys.* **2011**, *135*, 054510.
(57) Kaminski, G.; Jorgensen, W. L. Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids. *J. Phys. Chem.* **1996**, *100*, 18010−18013.
(58) Zhang, Y.; Feller, S. E.; Brooks, B. R.; Pastor, R. W. Computer simulation of liquid/liquid interfaces. I. Theory and application to octane/water. *J. Chem. Phys.* **1995**, *103*, 10252−10266.
(59) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for Mol. Simulat.s. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(60) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(61) Biscay, F.; Ghoufi, A.; Goujon, F.; Lachet, V.; Malfreyt, P. Calculation of the surface tension from Monte Carlo simulations: Does the model impact on the finite-size effects? *J. Chem. Phys.* **2009**, *130*, 184710.
(62) Pártay, L. B.; Hantal, G.; Jedlovszky, P.; Vincze, A.; Horvai, G. A new method for determining the interfacial molecules and characterizing the surface roughness in computer simulations. Application to the liquid−vapor interface of water. *J. Comput. Chem.* **2008**, *29*, 945−956.
(63) Sega, M.; Fábián, B.; Horvai, G.; Jedlovszky, P. How Is the Surface Tension of Various Liquids Distributed along the Interface Normal? *J. Phys. Chem. C* **2016**, *120*, 27468−27477.
(64) Sega, M.; Kantorovich, S. S.; Jedlovszky, P.; Jorge, M. The generalized identification of truly interfacial molecules (ITIM) algorithm for nonplanar interfaces. *J. Chem. Phys.* **2013**, *138*, 044110.
(65) Sega, M.; Hantal, G.; Fábián, B.; Jedlovszky, P. Pytim: A python package for the interfacial analysis of Mol. Simulat.s. *J. Comput. Chem.* **2018**, *39*, 2118−2125.
(66) Mitrinović, D. M.; Tikhonov, A. M.; Li, M.; Huang, Z.; Schlossman, M. L. Noncapillary-Wave Structure at the Water-Alkane Interface. *Phys. Rev. Lett.* **2000**, *85*, 582−585.
(67) Hantal, G.; Darvas, M.; Partay, L. B.; Horvai, G.; Jedlovszky, P. Molecular level properties of the free water surface and different organic liquid/water interfaces, as seen from ITIM analysis of computer simulation results. *J. Phys.: Condens. Matter* **2010**, *22*, 284112.
(68) Bresme, F.; Chacón, E.; Tarazona, P.; Tay, K. Intrinsic Structure of Hydrophobic Surfaces: The Oil-Water Interface. *Phys. Rev. Lett.* **2008**, *101*, 056102.
(69) Bresme, F.; Chacon, E.; Tarazona, P. Molecular dynamics investigation of the intrinsic structure of water-fluid interfaces via the intrinsic sampling method. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4704−4715.
(70) Bresme, F.; Chacón, E.; Tarazona, P. Force-field dependence on the interfacial structure of oil−water interfaces. *Mol. Phys.* **2010**, *108*, 1887−1898.
(71) Roddy, J. W.; Coleman, C. F. Solubility of water in hydrocarbons as a function of water activity. *Talanta* **1968**, *15*, 1281−1286.
(72) Nicolas, J. P.; Smit, B. Molecular dynamics simulations of the surface tension of n-hexane, n-decane and n-hexadecane. *Mol. Phys.* **2002**, *100*, 2471−2475.
(73) Irving, J. H.; Kirkwood, J. G. The Statistical Mechanical Theory of Transport Processes. IV. The Equations of Hydrodynamics. *J. Chem. Phys.* **1950**, *18*, 817−829.
(74) Ghoufi, A.; Goujon, F.; Lachet, V.; Malfreyt, P. Expressions for local contributions to the surface tension from the virial route. *Phys. Rev. E* **2008**, *77*, 031601.
(75) Lovett, R.; Baus, M. A molecular theory of the Laplace relation and of the local forces in a curved interface. *J. Chem. Phys.* **1997**, *106*, 635−644.
(76) Vanegas, J. M.; Torres-Sánchez, A.; Arroyo, M. Importance of Force Decomposition for Local Stress Calculations in Biomembrane Mol. Simulat.s. *J. Chem. Theory Comput.* **2014**, *10*, 691−702.
(77) Admal, N. C.; Tadmor, E. B. A Unified Interpretation of Stress in Molecular Systems. *J. Elasticity.* **2010**, *100*, 63−143.
(78) Ollila, O. H. S.; Risselada, H. J.; Louhivuori, M.; Lindahl, E.; Vattulainen, I.; Marrink, S. J. 3D Pressure Field in Lipid Membranes and Membrane-Protein Complexes. *Phys. Rev. Lett.* **2009**, *102*, 078101.

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(79) Malijevsky, A.; Jackson, G. A perspective on the interfacial properties of nanoscopic liquid drops. *J. Phys.: Condens. Matter* **2012**, *24*, 464121.
(80) Nicolas, J. P.; de Souza, N. R. Molecular dynamics study of the n-hexane−water interface: Towards a better understanding of the liquid−liquid interfacial broadening. *J. Chem. Phys.* **2004**, *120*, 2464−2469.
(81) Zeppieri, S.; Rodríguez, J.; López de Ramos, A. L. Interfacial Tension of Alkane + Water Systems. *J. Chem. Eng. Data* **2001**, *46*, 1086−1088.
(82) Takahashi, H.; Morita, A. A molecular dynamics study on inner pressure of microbubbles in liquid argon and water. *Chem. Phys. Lett.* **2013**, *573*, 35−40.
(83) Ghoufi, A.; Malfreyt, P. Local pressure components and surface tension of spherical interfaces. Thermodynamic versus mechanical definitions. I. A mesoscale modeling of droplets. *J. Chem. Phys.* **2011**, *135*, 104105.
(84) Xiu, D.; Karniadakis, G. The Wiener−Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM J. Sci. Comput.* **2002**, *24*, 619−644.
(85) Lei, H.; Yang, X.; Zheng, B.; Lin, G.; Baker, N. A. Constructing surrogate models of complex systems with enhanced sparsity: quantifying the influence of conformational uncertainty in biomolecular solvation. *Multiscale Model. Simul.* **2015**, *13*, 1327−1353.
(86) Li, Z.; Bian, X.; Yang, X.; Karniadakis, G. E. A comparative study of coarse-graining methods for polymeric fluids: Mori-Zwanzig vs. iterative Boltzmann inversion vs. stochastic parametric optimization. *J. Chem. Phys.* **2016**, *145*, 044102.
(87) Gerstner, T.; Griebel, M. Numerical integration using sparse grids. *Numer. Algorithms.* **1998**, *18*, 209−232.
(88) Xiu, D.; Hesthaven, J. High-Order Collocation Methods for Differential Equations with Random Inputs. *SIAM J. Sci. Comput.* **2005**, *27*, 1118−1139.
(89) Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Series. B* **1974**, *36*, 111−147.
(90) He, X.; Shinoda, W.; DeVane, R.; Klein, M. L. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Mol. Phys.* **2010**, *108*, 2007−2020.

https://dx.doi.org/10.1021/acs.jcim.0c00337
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX