United-atom, Mie λ-6 force fields cannot simultaneously predict vapor-liquid equilibria and high pressure properties. To be submitted to the Journal of Chemical Physics.

Richard A. Messerly,*,† Michael R. Shirts,*,‡ and Andrei F. Kazakov*,†

†*Thermodynamics Research Center, National Institute of Standards and Technology, Boulder, Colorado, 80305*

‡*Department of Chemical and Biological Engineering, University of Colorado, Boulder, Colorado, 80309*

E-mail: richard.messerly@nist.gov; michael.shirts@colorado.edu; andrei.kazakov@nist.gov

Abstract

Molecular simulation results at extreme temperatures and pressures can be used to supplement experimental data when developing fundamental equations of state for estimating thermodynamic properties over a wide range of state points. Since most

force fields are optimized to agree with vapor-liquid equilibria properties, however, the reliability of the molecular simulation results depends on the validity/transferability of the force field at higher temperatures and pressures. As demonstrated in this study, although state-of-the-art united-atom Mie λ-6 potentials for normal and branched alkanes provide accurate estimates for vapor-liquid equilibria properties, they tend to over-predict pressures for dense supercritical fluids and compressed liquids. The physical explanation for this observation is that the repulsive barrier is too steep for the "optimal" united-atom Mie λ-6 potential parameterized with vapor-liquid properties. Bayesian inference confirms that no feasible combination of ϵ, σ, and λ is capable of simultaneously predicting saturated vapor pressures, saturated liquid densities, and pressures at high temperatures and densities. This conclusion has both practical and theoretical ramifications, as more realistic non-bonded potentials may be required for accurate extrapolation to high pressures.

Purpose

1 Introduction

An accurate understanding of the relationship between pressure, volume (or density, ρ), and temperature (PVT) and caloric properties (such as heat capacity) for a given compound is essential for designing industrial chemical processes. Fundamental equations of state (FEOS), such as those based on the Helmholtz free energy, are a powerful approach for estimating PVT behavior and caloric properties. For example, the National Institute of Standards and Technology (NIST) REFPROP (Reference Fluid Properties) currently provides FEOS for around one hundred chemical species.[1] Unfortunately, most compounds do not have sufficient (reliable) experimental data covering a wide range of pressures, densities, and temperatures to develop a highly-accurate FEOS. Since FEOS

are semi-empirical and have 15-30 fitting parameters, the FEOS predictions can result in large errors at temperatures and pressures that are significantly higher than those used in parameterizing the FEOS. Therefore, improvement in an FEOS at high temperatures and pressures necessitates additional data for those conditions.

The lack of experimental data at high temperatures and pressures, especially, is likely attributed to the inherent safety, cost, and complexity of such experiments. By contrast, molecular simulation (i.e. Monte Carlo, MC, and molecular dynamics, MD) methods at high temperatures and pressures do not suffer from any of these limitations. Therefore, in principle, molecular simulation can aid in developing FEOS.[2–6] Although it is possible to fit an FEOS to just molecular simulation results, most studies implement hybrid data sets (i.e. from both experiment and molecular simulation).[7]

For example, several recent studies by Thol et al. supplement experimental data with molecular simulation results at temperatures and pressures beyond the range of available experimental temperatures and pressures.[8–11] Specifically, experimental data were available for temperatures and pressures up to 580 K and 130 MPa, 590 K and 180 MPa, 450 K and 2 MPa, and 560 K and 100 MPa for hexamethyldisiloxane,[8] octamethylcyclotetrasiloxane,[9] ethylene oxide,[11] and 1,2-dichloroethane,[10] respectively. Molecular simulations were performed for these compounds at temperatures and pressures up to 1200 K and 600 MPa, 1200 K and 520 MPa, 1000 K and 700 MPa, 1000 K and 1200 MPa, respectively. The inclusion of these simulation results improved the performance of the FEOS at extreme temperatures and pressures.

While previous studies have focused on small/hazardous compounds, the present study investigates normal and branched alkanes. Hydrocarbons are a fundamental feedstock for many petrochemical processes and, therefore, large amounts of experimental data exist covering a wide range of PVT phase space. For these reasons, REFPROP provides highly-accurate FEOS for several hydrocarbons, most of which are shorter-chains

(less than 20 carbons) with limited branching (i.e. only methyl branches). The use of hybrid data sets is an appealing approach to develop FEOS for industrially relevant hydrocarbons with less experimental data, i.e. those with longer chain-lengths or a higher degree of branching.

The primary limitation for implementing molecular simulation at extreme temperatures and pressures is whether or not the force field, which is typically parameterized using VLE data, is reliable at those conditions. For example, it was demonstrated that VLE-optimized force fields for small compounds, such as noble gases, hydrogen sulfide, and hydrogen chloride, do adequately represent the homogeneous fluid region.[7] In this study, we investigate how well the traditional force fields for predicting VLE of normal and branched alkanes extrapolate to higher temperatures (supercritical fluid) and pressures (compressed liquid). This analysis is performed for four normal and four branched alkanes by comparing the simulated compressibility factor (Z) with the REFPROP correlations. Note that the simulation conditions do not go beyond the range of validity for the respective compounds, so that we can assume the REFPROP correlations are reliable.

The most accurate force fields for estimating hydrocarbon VLE properties (i.e. ρ_l^{sat} and P_v^{sat}) are the Transferable Potentials for Phase Equilibria (TraPPE)[12,13] (and, especially, the recent TraPPE-2[14]), Errington,[15] anisotropic-united-atom (AUA4),[16,17] Potoff,[18,19] and Transferable anisotropic Mie potential (TAMie).[20,21] The TraPPE and Potoff force fields use a united-atom (UA) model while the TraPPE-2, Errington, AUA4, and TAMie force fields use an anisotropic-united-atom (AUA) model. Both a UA and AUA model group the hydrogen interactions with their neighboring carbon atom. However, the UA model assumes that the UA interaction site is that of the carbon atom, while an AUA model assumes that the AUA interaction site is shifted away from the carbon atom and towards the hydrogen atom(s). Although, in theory, an all-atom (AA) force field should yield more accurate results, from a parameterization standpoint, it is much easier to ensure that a

4

global minimum is obtained when parameterizing UA and AUA force fields since fewer (highly correlated) parameters are optimized simultaneously. Furthermore, the reduced computational cost is an additional benefit of the UA and AUA approach.

In addition to the division between UA and AUA force fields, the existing force fields differ in the non-bonded functional form and corresponding parameters. The TraPPE, TraPPE-2, and AUA4 force fields use a Lennard-Jones (LJ) 12-6 potential, while the Potoff and TAMie force fields use the Mie λ-6 (or generalized Lennard-Jones) potential, and the Errington force field uses the Buckingham exponential-6 (Exp-6) potential.

The three-parameter Mie λ-6 and Exp-6 potentials are more flexible than the two-parameter LJ 12-6 potential as an additional adjustable parameter controls the steepness of the repulsive barrier. Specifically, the Mie λ-6 potential is:[22]

$$u^{\mathrm{vdw}}(\epsilon, \sigma, \lambda; r) = \left(\frac{\lambda}{\lambda - 6}\right) \left(\frac{\lambda}{6}\right)^{\frac{6}{\lambda - 6}} \epsilon \left[\left(\frac{\sigma}{r}\right)^{\lambda} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{1}$$

where u^{vdw} is the van der Waals interaction, σ is the distance (r) where $u^{\mathrm{vdw}} = 0$, $-\epsilon$ is the energy of the potential at the minimum $\left(\text{i.e. } u^{\mathrm{vdw}} = -\epsilon \text{ and } \frac{\partial u^{\mathrm{vdw}}}{\partial r} = 0 \text{ for } r = r_{\min}\right)$, and λ is the repulsive exponent. Note that the Mie λ-6 potential reduces to the LJ 12-6 potential for $\lambda = 12$. Therefore, the LJ 12-6 potential can be considered a special subclass of the Mie λ-6 potential.

The Buckingham exponential-6 model is:[15]

$$u^{\mathrm{vdw}}(\epsilon, r_{\min}, \alpha; r) = \begin{cases} \frac{\epsilon}{1 - \frac{6}{\alpha}} \left[\frac{6}{\alpha} \exp\left(\alpha\left[1 - \frac{r}{r_{\min}}\right]\right) - \left(\frac{r_{\min}}{r}\right)^{6}\right] & \text{for } r > r_{\max} \\ \infty & \text{for } r < r_{\max} \end{cases} \tag{2}$$

where u^{vdw}, ϵ, and r are the same as in Equation 1, $r_{\min}$ is the distance that corresponds to the minimum in the potential (i.e $u^{\mathrm{vdw}}(r_{\min}) = -\epsilon$), α is a Buckingham exponential-6

parameter, and r_{max} is the smallest positive value for which $\frac{du^{\text{vdw}}}{dr} = 0$.

Previous work demonstrated that the UA LJ 12-6 potential cannot adequately estimate both $\rho_{\text{l}}^{\text{sat}}$ and $P_{\text{v}}^{\text{sat}}$ for n-alkanes.[23,24] For this reason, the TraPPE-UA force field was primarily developed to agree with saturated liquid densities.[12] By contrast, accurate prediction of both $\rho_{\text{l}}^{\text{sat}}$ and $P_{\text{v}}^{\text{sat}}$ over a wide temperature range is possible by varying the repulsive exponent of the LJ potential (i.e. the Mie λ-6 potential). It is important to mention that, although an attractive exponent of 6 has a strong theoretical basis, $\lambda = 12$ is a historical artifact that was chosen primarily for computational purposes.[25]

Typically, when parameterized to VLE data, the optimal value of λ is greater than 12 with a corresponding increase in the well depth (ϵ). Specifically, for most hydrocarbons, the Potoff UA force field[18,19] uses $\lambda = 16$ while the TAMie force field[20] uses $\lambda = 14$. Gordon also demonstrated that reliable viscosities can be obtained from a UA Mie λ-6 model for n-alkanes by using $\lambda = 14$ and $\lambda = 20$ for the CH_3 and CH_2 sites, respectively (note the subtle difference in how Gordon defines the Mie λ-6 potential, a.k.a. "mod-n-6").[26] It is interesting to note that Gordon and Galliéro et al. report λ values of 11 and 10, respectively, for UA methane when optimized with viscosity data.[26,27]

However, there are some theoretical concerns that increasing the repulsive exponent might have some undesirable consequences, especially at high pressures, where particles will spend more time with very short pairwise distances than at VLE conditions. For example, several older studies demonstrated that neither an all-atom LJ 12-6 or an all-atom LJ 9-6 is adequate to reproduce high-level *ab initio* calculations of n-alkanes ranging from methane to n-butane.[28–30] The studies of Rowley et al. suggest a Morse potential is necessary for accurate representation of *ab initio* dimer energies. Hayes et al. confirms these results while also emphasizing that the short-range repulsive forces, which are most important when computing high pressures in molecular simulation, are poorly represented with an AA LJ 12-6 or an AA LJ 9-6 model. Specifically, the LJ 12-6 potential is too steep,

and only slight improvement in the repulsive region is observed for the LJ 9-6 potential. It is interesting that Hayes et al. also highlights deficiencies in the repulsive region for the Buckingham exponential-6 potential.

Recently, Kulakova et al. used Bayesian inference to conclude that experimental data for argon, specifically the liquid and vapor radial distribution functions at varying temperatures and densities, support λ values between 6 and 10, while argon dimer *ab initio* energies support λ values between 12 and 14.[31] They suggest that these larger values of λ should not be used for liquid phase simulations. By contrast, two other studies of noble gases, including argon, support $\lambda \geq 12$.[27,32] Specifically, Mick et al. reports a 13-6 potential for argon, while Galliéro et al. states that the 12-6 potential is superior for argon than the 10-6, 14-6, 16-6, 18-6, and 20-6 potentials. The primary explanation for this discrepancy is the choice of experimental data. The conclusions from Kulakova et al. are based on the radial distribution function, while Galliéro et al. used viscosity and pressure, and Mick et al. utilized VLE data.

Structural properties, such as the radial distribution function, and *ab initio* calculations provide considerable insight into the true repulsive barrier.[27,31] However, the "correct" value of λ does not guarantee adequate prediction of VLE and/or PVT behavior. This is primarily because the Mie λ-6 potential is only an approximation to the real potential and, thus, it is not flexible enough to agree with both the repulsive and attractive regions. Instead, only the region that is most sensitive to the target experimental data will be adequately represented, which for VLE data is typically the effective-size, depth of the energy minimum, and the dispersive tail. Furthermore, the "optimal" λ is not necessarily the "true" λ because force fields use effective parameter sets that account for assumptions, such as pair-wise additivity (i.e. excluding three-, four-, etc. body interactions) or the lack of explicit hydrogens. For example, a UA Mie λ-6 potential is simply not capable of predicting VLE properties of ethane for $\lambda < 12$ (see Figures 1-2 of Reference 18), despite

theoretical evidence that the repulsive barrier should be softer than $\lambda = 12$.

The purpose of this study is to determine whether or not the UA Mie λ-6 model is adequate for predicting *both* VLE and PVT at high temperatures and pressures for alkanes. Although it is difficult to know if the theoretical studies for noble gases and all-atom n-alkane models are applicable to UA models for normal and branched alkanes, the working hypothesis based on these previous studies is that a UA Mie λ-6 potential performs poorly at high pressures if parameterized with VLE data. This assessment is of practical engineering importance for deciding whether or not UA Mie λ-6 force fields should be used when developing fundamental equations of state for alkanes based on hybrid data sets.

The outline for this study is the following. Section 2 discusses the simulation and force field details. Section 3 is a case study for normal and branched alkanes using the existing force fields developed based on VLE properties. Section 4 explains how Bayesian inference is employed to investigate the adequacy of the UA Mie λ-6 potential. Section 5 presents the results from the Bayesian analysis with recommendations and limitations in Section 6. Section 7 reports the primary conclusions of this study.

2 Methods I

2.1 Simulation Details

Four normal and four branched alkanes of varying chain-length and degree of branching are simulated in this study. Specifically, we simulate ethane, propane, n-butane, n-octane, isobutane (2-methylpropane), isohexane (2-methylpentane), isooctane (2,2,4-trimethylpentane), and neopentane (2,2-dimethylpropane). These compounds were chosen to represent a diverse set of the normal and branched alkanes available in REFPROP.[1]

Simulations for this study are performed in the NVT ensemble (constant number of molecules, N, constant volume, V, and constant temperature, T) using GROMACS version 2018.[33] Each simulation uses the Velocity Verlet integrator with a 2 fs time-step, 1.4 nm cut-off for non-bonded interactions with tail corrections for energy and pressure, Nosé-Hoover thermostat with a time constant of 1 ps, and fixed bond-lengths are constrained using LINCS with a LINCS-order of eight. Coulombic interactions are not computed as none of the force fields require partial charges for the compounds studied. The equilibration time was 0.1 ns for ethane and propane, 0.2 ns for n-butane, and 0.5 ns for all other compounds. The production time was 1 ns for ethane, 2 ns for propane and n-butane, and 4 ns for all other compounds. Replicate simulations were performed for n-octane to validate that a single MD run of this length agrees with the average of several replicates, to within the combined uncertainty. A system size of 400 molecules is used for ethane, propane, and n-butane, while all other compounds use 800 molecules. Example input files are provided in Section SI.I of Supporting Information.

Simulations are performed along a supercritical isotherm (with a reduced temperature, $T_r \approx 1.2$) and five saturated liquid density isochores. Nine densities are simulated along the supercritical isotherm (T^{IT}) with five densities being the same as the isochore densities. Two additional temperatures are simulated along each isochore, with one being the REFPROP saturation temperature (T^{sat}) while the inverse of the second isochore temperature is the average of $1/T^{IT}$ and $1/T^{sat}$. Thus, a total of 19 simulations are performed for each compound and force field. The specific state points for each compound studied are depicted in Figure 1, with the REFPROP saturation curve included as a reference. Tabulated values for the state points of each compound are provided in Section SI.I of Supporting Information.

The state points depicted in Figure 1 correspond to the recommended conditions for the isothermal isochoric integration (ITIC) algorithm.[34,35] ITIC converts the departure in-
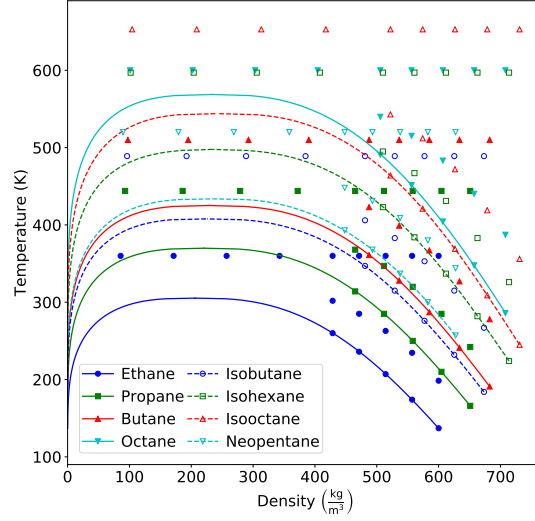
Figure 1: State points simulated for each compound studied. A total of 19 simulations are performed: nine densities along the supercritical isotherm and two temperatures along five liquid density isochores. Filled symbols and solid lines correspond to *n*-alkanes, while empty symbols and dashed lines correspond to branched alkanes. The REFPROP saturation curve for each compound is included as a reference.

ternal energies (U^{dep}) and compressibility factors (Z) obtained at the 19 state points to saturated VLE properties, namely, ρ_l^{sat} and P_v^{sat}. The equations for ITIC are:

$$\frac{A^{\text{dep}}}{R_g T^{\text{sat}}} = \int_0^{\rho_l^{\text{sat}}} \frac{Z-1}{\rho} \partial\rho|_{T=T^{\text{IT}}} + \int_{T^{\text{IT}}}^{T^{\text{sat}}} U^{\text{dep}} \partial\left(\frac{1}{R_g T}\right)|_{\rho=\rho_l^{\text{sat}}} \tag{3}$$

$$\rho_v^{\text{sat}} \approx \rho_l^{\text{sat}} \exp\left(\frac{A^{\text{dep}}}{R_g T^{\text{sat}}} + Z_l^{\text{sat}} - 1 - 2B_2\rho_v^{\text{sat}} - 1.5B_3(\rho_v^{\text{sat}})^2\right) \tag{4}$$

$$P_v^{\text{sat}} \approx \left(1 + B_2\rho_v^{\text{sat}} + B_3(\rho_v^{\text{sat}})^2\right)\rho_v^{\text{sat}} R_g T^{\text{sat}} \tag{5}$$

$$Z_l^{\text{sat}} = \frac{P_v^{\text{sat}}}{\rho_l^{\text{sat}} R_g T^{\text{sat}}} \tag{6}$$

where $A^{\text{dep}} \equiv A - A^{\text{ig}}$ is the Helmholtz free energy departure from ideal gas for temperature (T) equal to the saturation temperature (T^{sat}) and density (ρ) equal to the saturated liquid density (ρ_l^{sat}), $U^{\text{dep}} \equiv U - U^{\text{ig}}$ is the internal energy departure, Z_l^{sat} is the saturated

liquid compressibility factor (Z), B_2 is the second virial coefficient, B_3 is the third virial coefficient, T^{IT} is the isothermal temperature, and R_{g} is the universal gas constant. For details regarding the implementation of ITIC, see Reference 35. As discussed in our previous work,[35] the B_2 and B_3 values found in Equations 4-5 are calculated using REFPROP correlations 1. The use of REFPROP correlations introduces a small bias in the resulting $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$, which is accounted for in Section 4.1.

2.2 Force field

A united-atom (UA) or anisotropic-united-atom (AUA) representation is used for each compound studied. The UA and AUA groups required for normal and branched alkanes are sp^3 hybridized CH_3, CH_2, CH, and C sites. For most literature models, a single (transferable) parameter set is assigned for each interaction site. However, two exceptions exist for the force fields studied. First, TAMie implements a different set of CH_3 parameters for ethane and other alkanes. Second, Potoff reports a "generalized" and "short/long" (S/L) CH and C parameter set. The Potoff "generalized" CH and C parameter set is an attempt at a completely transferable set. However, since the "generalized" parameters performed poorly for some compounds, the S/L parameter set was proposed, where the "short" and "long" parameters are implemented when the number of carbons in the backbone is ≤ 4 and > 4, respectively.

A fixed bond-length is used for each bond between UA or AUA sites. Although TAMie is an AUA force field, only the terminal CH_3 sites have a displacement in the interaction site. For example, Figure 2 depicts both the UA and AUA representations of isooctane when only terminal CH_3 interaction sites are shifted from the carbon center. This convention is much simpler to implement than other AUA approaches (such as AUA4) where non-terminal (i.e. CH_2 and CH) interaction sites also have a displacement distance. For this reason, we do not attempt to simulate the AUA4 force field for any compounds con-

taining CH_2 and CH interaction sites. For the compounds and force fields simulated, the anisotropic shift in a terminal interaction site (i.e. CH_3) is treated simply as a longer effective bond-length (see Table 1). The bond-length for all non-terminal sites is 0.154 nm, except for the Errington Exp-6 force field which uses 0.1535 nm for CH_2-CH_2 bonds.
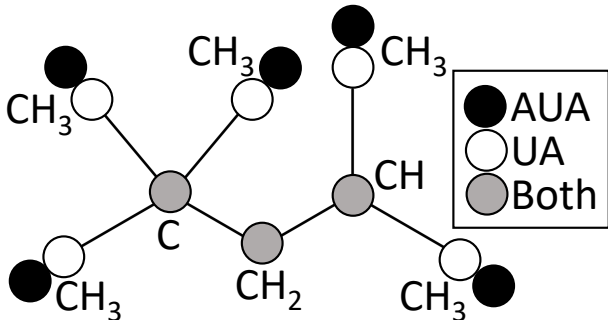
Figure 2: If only the terminal sites have an anisotropic displacement (e.g. TAMie), AUA force fields have the same complexity as UA force fields since the only difference is a longer effective bond-length for the CH_3 sites. Note that the AUA4 representation of isooctane requires a more complicated shifting of CH_2 and CH sites than that depicted here.

Table 1: Effective bond-lengths in units of nm for terminal (CH_3) UA or AUA interaction sites. Empty table entries for Exp-6 and TraPPE-2 denote that the force field does not contain the corresponding interaction site type. Empty table entries in AUA4 arise because this force field uses a more complicated construct than the simple effective bond-length approach. Specifically, AUA4 requires CH_2 and CH interaction sites that are not along the C-C bond axis.

Bond	TraPPE, Potoff	TAMie	Exp-6	AUA4	TraPPE-2
CH_3-CH_3	0.154	0.194	0.1839	0.1967	0.230
CH_3-CH_2	0.154	0.174	0.1687	–	–
CH_3-CH	0.154	0.174	–	–	–
CH_3-C	0.154	0.174	–	0.1751	–

The angle and dihedral energies are computed using the same functional forms for each force field. Angular bending interactions are evaluated using a harmonic potential:

$$u^{\text{bend}} = \frac{k_\theta}{2} (\theta - \theta_0)^2 \tag{7}$$

12

where u^{bend} is the bending energy, θ is the instantaneous bond angle, θ_0 is the equilibrium bond angle, and k_θ is the harmonic force constant which is equal to 62500 K/rad^2 for all bonding angles. Dihedral torsional interactions are determined using a cosine series:

$$u^{\text{tors}} = c_1[1 + \cos \phi] + c_2[1 - \cos 2\phi] + c_3[1 + \cos 3\phi] \tag{8}$$

where u^{tors} is the torsional energy, ϕ is the dihedral angle and c_i are the Fourier constants. The equilibrium bond angles and torsional parameters are found in Tables 2-3, respectively. Note that the Errington c_i values for CH_x-CH_2-CH_2-CH_y are a factor of two less than those reported in Table 3.

Table 2: Equilibrium bond angles (θ_0). x and y are values between 0-3.

Bending sites	θ_0 (degrees)
CH_x-CH_2-CH_y	114.0
CH_x-CH-CH_y	112.0
CH_x-C-CH_y	109.5

Table 3: Fourier constants (c_i) in units of K. x and y are values between 0-3.

Torsion sites	c_0	c_1	c_2	c_3
CH_x-CH_2-CH_2-CH_y	0.0	355.03	-68.19	791.32
CH_x-CH_2-CH-CH_y	-251.06	428.73	-111.85	441.27
CH_x-CH_2-C-CH_y	0.0	0.0	0.0	461.29
CH_x-CH-CH-CH_y	-251.06	428.73	-111.85	441.27

Non-bonded interaction energies and forces between sites located in two different molecules or separated by more than three bonds are calculated using either a Lennard-Jones 12-6, Mie λ-6, or Buckingham Exponential-6 potential (see Equations 1-2). The non-bonded LJ 12-6 or Mie λ-6 force field parameters for TraPPE, TraPPE-2, Potoff, AUA4,

13

and TAMie are provided in Table 4. Note that, for computational purposes, a common practice to date is to use integer values of λ in Equation 1.

Table 4: Non-bonded (intermolecular) parameters for TraPPE[12,13] (and TraPPE-2[14]), Potoff,[18,19] AUA4,[16,36] and TAMie[20,21] force fields. The "short/long" Potoff CH and C parameters are included in parenthesis. The ethane specific parameters for TAMie are included in parenthesis.

United-atom	TraPPE (TraPPE-2)			Potoff (S/L)		
	ϵ (K)	σ (nm)	λ	ϵ (K)	σ (nm)	λ
CH_3	98 (134.5)	0.375 (0.352)	12	121.25	0.3783	16
CH_2	46	0.395	12	61	0.399	16
CH	10	0.468	12	15 (15/14)	0.46 (0.47/0.47)	16
C	0.5	0.640	12	1.2 (1.45/1.2)	0.61 (0.61/0.62)	16
	AUA4			TAMie		
CH_3	120.15	0.3607	12	136.318 (130.780)	0.36034 (0.36463)	14
CH_2	86.29	0.3461	12	52.9133	0.40400	14
CH	50.98	0.3363	12	14.5392	0.43656	14
C	15.04	0.244	12	–	–	–

The Errington Exp-6 non-bonded parameters are found in Table 5. Note that Errington reported values for ϵ, σ, and α. We computed $r_{\min}$ and $r_{\max}$ to facilitate compatibility with Equation 2 and future validation of our results.

Table 5: Non-bonded (intermolecular) parameters for Errington Exp-6 force field.

United-atom	ϵ (K)	σ (nm)	α	$r_{\min}$ (nm)	$r_{\max}$ (nm)
CH_3	129.6	0.3679	16	0.4094	0.0574
CH_2	73.5	0.400	22	0.436	0.0221

Non-bonded interactions between two different site types (i.e. cross-interactions) are determined using Lorentz-Berthelot combining rules[25] for ϵ and σ, an arithmetic mean for the repulsive exponent (λ) (as recommended in Reference 18), and a geometric mean

for α (as recommended in Reference 15):

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{9}$$

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{10}$$

$$\lambda_{ij} = \frac{\lambda_{ii} + \lambda_{jj}}{2} \tag{11}$$

$$\alpha_{ij} = \sqrt{\alpha_{ii}\alpha_{jj}} \tag{12}$$

where the ij subscript refers to cross-interactions and the subscripts ii and jj refer to same-site interactions.

3 Case study for alkanes

The purpose of this case study is to demonstrate that the existing UA and AUA force fields for normal and branched alkanes that were parameterized with VLE properties do not predict the proper PVT behavior at higher temperatures and pressures (with the exception of ethane for the TraPPE-2 potential). Figures 3 and 4 plot the compressibility factor with respect to inverse temperature for n-alkanes and branched alkanes, respectively. Note that saturation corresponds to $Z \approx 0$ for each isochore. The "Potoff" results in Figure 4 are only for the the "short/long" model, since the "short/long" model is more accurate than the "generalized" model. The results for the "generalized" model do not provide any additional insight but are found in Section SI.II of Supporting Information.

Figure 3 demonstrates that the existing literature force fields for n-alkanes, while accurate for VLE, do not capture the correct PVT behavior at high pressures (P^{high}), i.e. the higher temperatures and highest isochore densities (ρ_3 and ρ_4). Figure 4 shows the same erroneous trend in Z for branched alkanes. Note that the error in Z at high temperatures
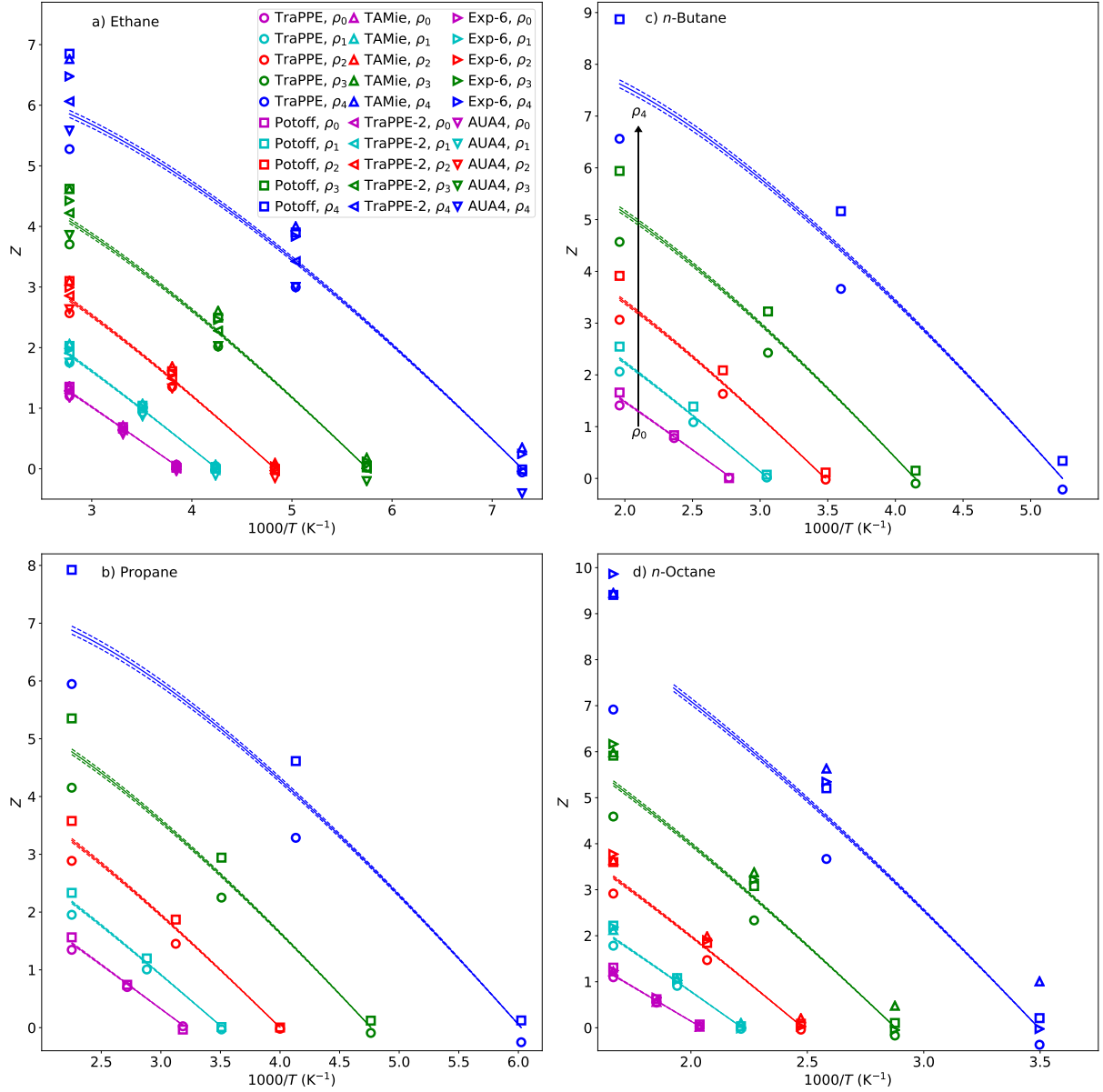
Figure 3: Compressibility factors (Z) along isochores agree at saturation ($Z \approx 0$) but deviate strongly at higher pressures. Densities are distinguished by color, increase vertically, and are labeled such that $\rho_0 < \rho_1 < \rho_2 < \rho_3 < \rho_4$. Panels a)-d) correspond to ethane, propane, *n*-butane, and *n*-octane, respectively. Solid lines represent REFPROP correlations, with dashed lines representing a 1% uncertainty in REFPROP values. Simulation error bars are approximately one symbol size.
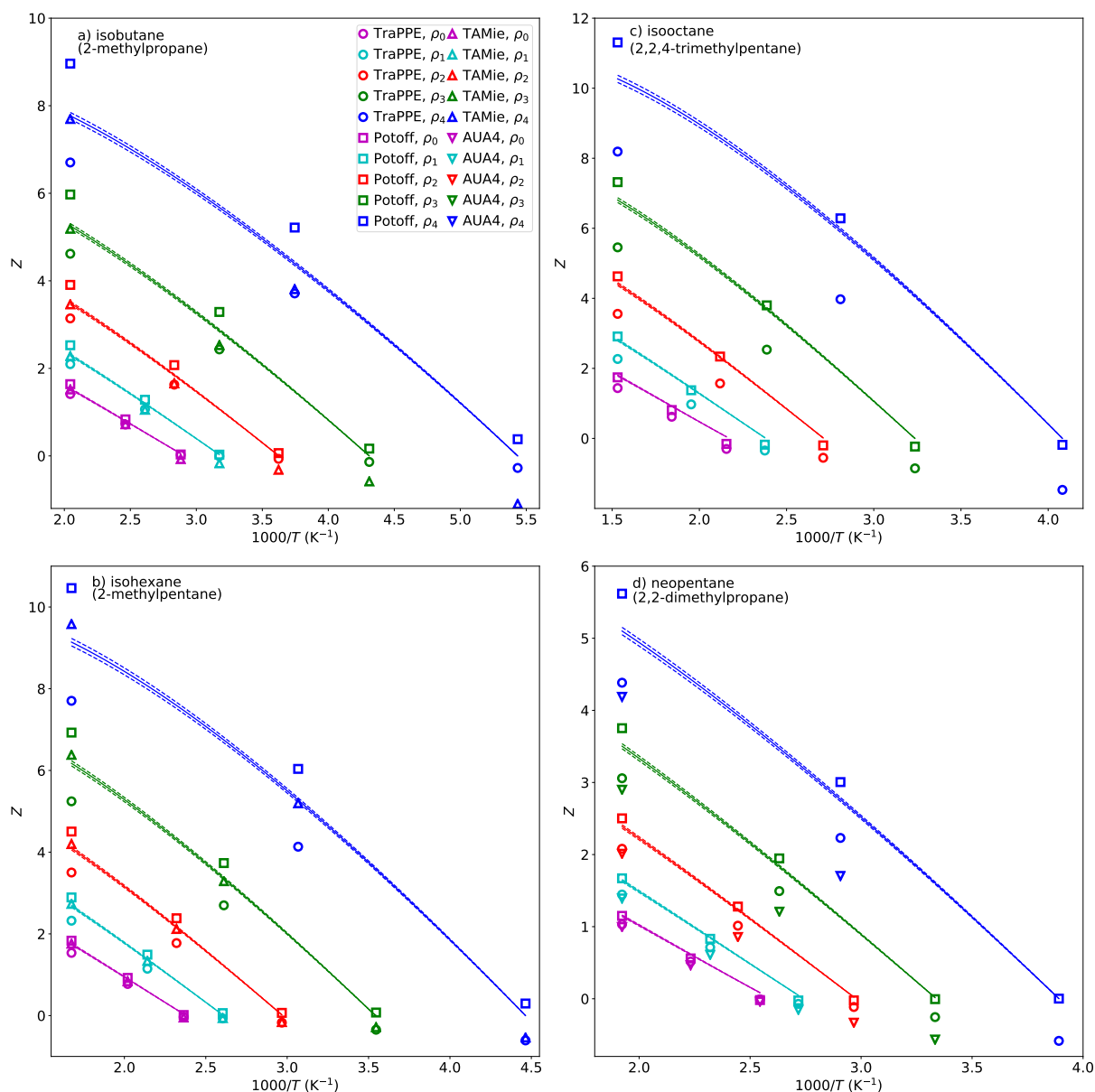
Figure 4: Compressibility factors (Z) along isochores for branched alkanes are not as accurate as normal alkanes at saturation ($Z \approx 0$) and deviate strongly at higher pressures. Panels a)-d) correspond to isobutane, isohexane, isooctane, and neopentane, respectively. Symbols, lines, uncertainties, and formatting are the same as those in Figure 3. The Potoff results for isobutane and neopentane use the "short" parameters, while isohexane and isooctane use the "long" parameters (see Table 4).

is less obvious because these force fields are typically not as reliable at predicting VLE for branched alkanes as for n-alkanes, i.e. notice the large deviations at $Z \approx 0$. That being said, it is clear in both Figures 3-4 that none of the force fields adequately reproduces Z over the entire temperature range, or the slope of Z with respect to inverse T.

In general, clear systematic biases are observed for the LJ 12-6 potentials (TraPPE-UA and AUA4) and the Mie λ-6 potentials (Potoff and TAMie). Specifically, the LJ 12-6 and Mie λ-6 potentials under- and over-predict Z at high pressures, respectively. These results are intuitive as the repulsive barriers are steeper for the respective Mie 16-6 and 14-6 potentials of the Potoff and TAMie force fields. A more surprising trend is that the Errington (AUA Exp-6) model also has a positive bias at high pressures. This appears to suggest that the repulsive barrier is too steep, however, the Exp-6 model is softer than the LJ 12-6 at very short distances. Therefore, it is more likely that the Exp-6 repulsive forces are over-predicted for the close-range distances that are sampled in molecular dynamics, i.e. $0.8\sigma < r < r_{\min}$. Alternatively, since the exponential contribution is non-negligible for $r_{\min} < r < 1.5\sigma$, it is possible that the Exp-6 under-predicts the attractive forces in this region. More definitive conclusions regarding the repulsive barrier are possible for the Mie λ-6 potential by directly comparing different values of λ.

The one exception to this trend is the TraPPE-2 model for ethane, which has the most accurate prediction of the entire PVT phase space simulated. Specifically, TraPPE-2 reproduces Z to within the REFPROP uncertainties for all state points except at the two highest densities along the supercritical isotherm, i.e. $T = T^{IT}$ and $\rho = \rho_3$ or ρ_4. The average percent deviation (AD%) relative to the REFPROP correlations is only 3% for pressure at these state points (P^{high}).

The performance of TraPPE-2 is somewhat surprising considering this force field has only three fitting parameters (ϵ, σ, and the effective bond-length) while the TAMie model has these three parameters and an additional fitting parameter (λ). It is important to note

that TraPPE-2 uses a much longer effective bond-length of 0.230 nm while TAMie did not consider bond-lengths larger than 0.194 nm. Therefore, the fact that the TraPPE-2 force field extrapolates to high pressures better than TAMie suggests that, at high pressures, it is important to account for hydrogens either explicitly (AA model) or with a longer effective bond-length than that typically used for AUA models (see Table 1). It is also possible that a four parameter optimization, such as that used by TAMie, is overfit to the VLE data and would perform better if high pressure PVT data were included in the parameterization.

Unfortunately, a direct comparison of the non-bonded interactions for AUA force fields is difficult because each model has a different anisotropic displacement, i.e. effective bond-length. By contrast, comparing TraPPE-UA and Potoff is straightforward because they use the same bond-lengths and the same non-bonded Mie λ-6 potential (Equation 1). For example, since the TraPPE-UA (LJ 12-6) potential under-predicts Z and the Potoff (UA Mie 16-6) potential over-predicts Z, it seems reasonable that a UA Mie 13-6, 14-6, or 15-6 model could demonstrate the proper trend, if parameterized appropriately. To investigate this hypothesis, the remainder of this document focuses on the UA Mie λ-6 potential, where all bond-lengths are 0.154 nm.

Specifically, we perform a Bayesian uncertainty quantification analysis to determine if there exists a set of ϵ, σ, and λ that reasonably predicts $\rho_{\mathrm{l}}^{\mathrm{sat}}$, $P_{\mathrm{v}}^{\mathrm{sat}}$, and PVT of supercritical fluids and compressed liquids. The results in Section 5 suggest that the optimal value of λ (13 or 14) for predicting PVT of supercritical fluids and compressed liquids is not capable of predicting VLE properties accurately. To understand this point, it is important to remember that the TraPPE-UA LJ 12-6 force field was optimized using only $\rho_{\mathrm{l}}^{\mathrm{sat}}$ because the UA LJ 12-6 model cannot adequately predict both $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$.

4 Methods II

The results presented in Section 3 demonstrate that none of the literature UA or AUA force fields, parameterized with VLE data, can reproduce the PVT behavior for supercritical fluids and compressed liquids. However, there is uncertainty in the non-bonded parameters inherited from the VLE data. Therefore, by considering the inherent uncertainty, it is possible that a feasible parameter set exists that adequately predicts VLE and P^{high}. By contrast, if none of the feasible ϵ, σ, and λ values is capable of simultaneously predicting VLE properties and PVT behavior at high pressures, we can conclude that the UA Mie λ-6 potential (and Lennard-Jones 12-6 as a special case) is inadequate for this purpose and, therefore, should not be used when developing FEOS with molecular simulation results.

Bayesian inference is a rigorous approach to quantify if the UA Mie λ-6 potential is "adequate". We refer the reader to the literature for a thorough discussion of Bayesian statistics.[31,37–40] In Section 4.1, we review some basic concepts of Bayes theorem, we define the posterior, likelihood, and prior distributions, and we discuss the Markov Chain Monte Carlo (MCMC) approach for sampling from the posterior joint distribution of the parameters. MCMC can be computationally burdensome, especially when molecular simulation is required to compute the likelihood. Section 4.2 demonstrates how we utilize surrogate models to estimate $\rho_{\text{l}}^{\text{sat}}$, $P_{\text{v}}^{\text{sat}}$, and Z for a given set of ϵ, σ, and λ. The use of surrogate models reduces the computational cost of MCMC by several orders of magnitude. We implement this analysis for n-alkanes to generate joint distributions of ϵ_{CH_3}-σ_{CH_3} and ϵ_{CH_2}-σ_{CH_2} for different values of λ_{CH_3} and λ_{CH_2}, respectively.

4.1 Bayesian Analysis

4.1.1 Theory

Bayesian inference is used to quantify the uncertainty in the non-bonded parameters (ϵ and σ) and to determine the evidence for different values of λ. Bayes theorem states

$$Pr(\theta|D, M) = \frac{Pr(D|\theta, M)Pr(\theta|M)}{Pr(D|M)} \tag{13}$$

where Pr denotes a probability distribution function, θ is the parameter set (ϵ and σ), M is the model (Mie λ-6 for a given value of λ) and D are the data (ρ_l^sat and P_v^sat). $Pr(\theta|D, M)$ is commonly referred to as the "posterior", $Pr(D|\theta, M)$ is the "likelihood" (alternatively expressed as $L(\theta|D, M)$), $Pr(\theta|M)$ is the "prior", and $Pr(D|M)$ is a normalization constant which is also the "model evidence".

The "model evidence" is used in model selection, by computing the probability of different models given the data:

$$Pr(M|D) = \frac{Pr(D|M)Pr(M)}{Pr(D)} \tag{14}$$

where $Pr(M)$ is the "model prior", $Pr(D)$ is a normalization constant, and $Pr(M|D)$ is the "model posterior". The ratio of $Pr(M|D)$, known as the Bayes factor (K), provides the relative probability of two models.

The parameter uncertainty propagates when estimating another quantity of interest (QoI), which may or may not be included in D, according to:[31]

$$Pr(QoI|D, M) = \int Pr(QoI|\theta, M)Pr(\theta|D, M)d\theta \tag{15}$$

This expression is commonly referred to as "robust posterior prediction." Note that the

uncertainty in QoI, obtained from $Pr(QoI|D, M)$, does not account for deficiencies in the model itself, only the uncertainty in the model parameters.

4.1.2   Application

For clarity, we rewrite Equations 13-14 for the specific case studied

$$Pr(\epsilon, \sigma|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda) = \frac{L(\epsilon, \sigma|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)Pr(\epsilon, \sigma|\lambda)}{Pr(\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}|\lambda)} \tag{16}$$

$$Pr(\lambda|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}) = \frac{Pr(\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}|\lambda)Pr(\lambda)}{Pr(\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}})} \tag{17}$$

where in this context $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ should each be interpreted as an array of experimental data values. Notice that θ does not include λ, since we use λ to distinguish between models. The "model evidence", $Pr(\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}|\lambda)$ in Equation 17, for different values of λ is determined by integrating the numerator of Equation 16 for all values of ϵ and σ.

To compute the Bayes factor between two values of λ (i.e. two different models, M), we assume that the prior evidence is equal for all positive values of ϵ, σ, and λ (within a feasible range), i.e. we use uniform distributions for $Pr(\epsilon, \sigma|\lambda)$ in Equation 16 and $Pr(\lambda)$ in Equation 17. Due to the large amount of information contained in the data, D, the use of a uniform prior does not impact our results, i.e. the data "overwhelms" the prior. The advantage of this assumption is that the Bayes factor, K, depends completely on the likelihood:

$$K = \frac{Pr(\lambda_j|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}})}{Pr(\lambda_i|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}})} = \frac{\int L(\epsilon, \sigma|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda_j)d\epsilon d\sigma}{\int L(\epsilon, \sigma|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda_i)d\epsilon d\sigma} \tag{18}$$

where λ_i and λ_j are the different (fixed) values of λ being compared.

We utilize robust posterior prediction (Equation 15) to propagate the parameter uncertainty for three different QoI, specifically, $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$, which are included in D, as well as Z. For example, the uncertainty in predicting Z due to the ϵ–σ parameter uncertainties

(for a given λ, based on VLE data) is obtained from:

$$Pr(Z|\rho_1^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda) = \int Pr(Z|\epsilon, \sigma, \lambda)Pr(\epsilon, \sigma|\rho_1^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)\partial\epsilon\partial\sigma \qquad (19)$$

Similar expressions exist for $Pr(\rho_1^{\mathrm{sat}}|\rho_1^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)$ and $Pr(P_{\mathrm{v}}^{\mathrm{sat}}|\rho_1^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)$. Note that posterior prediction allows for uncertainty estimates in ρ_1^{sat} and $P_{\mathrm{v}}^{\mathrm{sat}}$ at any temperature, not just those included in D.

4.1.3 Implementation

Markov Chain Monte Carlo (MCMC) is the traditional approach for numerically sampling from the probability distribution $Pr(\epsilon, \sigma|\rho_1^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)$. We use the Metropolis-Hastings algorithm to create a Markov Chain by proposing new ϵ or σ values and accepting those moves based on the ratio of the probability between the previous parameter set and the proposed parameter set:

$$\alpha = \min\left(1, \frac{Pr(\epsilon_{i+1}, \sigma_{i+1}|\rho_1^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)Q(\epsilon_{i+1}, \sigma_{i+1}|\epsilon_i, \sigma_i, s_\epsilon^2, s_\sigma^2)}{Pr(\epsilon_i, \sigma_i|\rho_1^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)Q(\epsilon_i, \sigma_i|\epsilon_{i+1}, \sigma_{i+1}, s_\epsilon^2, s_\sigma^2)}\right) \qquad (20)$$

where α is the acceptance probability, ϵ_i and σ_i are the previous parameter set, ϵ_{i+1} and σ_{i+1} are the proposed parameter set, and Q is the proposal distribution from which ϵ and σ are sampled. In this study, Q is a bi-variate normal distribution with mean equal to ϵ_i and σ_i, variance of s_ϵ^2 and s_σ^2, and a covariance of 0. The amount to which ϵ or σ is varied ($\Delta\epsilon$ and $\Delta\sigma$) for each MCMC step depends on s_ϵ^2 and s_σ^2. These parameters are tuned such that approximately $\frac{1}{3}$ of the moves are accepted. This "tuning" period (also referred to as a "burn-in" period) is followed by a production period where s_ϵ^2 and s_σ^2 do not change. Since ϵ_{i+1} and σ_{i+1} are highly correlated with ϵ_i and σ_i, it is important to "thin" the MCMC ϵ and σ parameter sets (ϵ_{MCMC} and σ_{MCMC}), i.e. every j^{th} parameter

set is stored. Section SI.III of Supporting Information provides an MCMC example with some details (i.e. number of steps for burn-in and production, frequency that s_ϵ^2 and s_σ^2 are updated, resulting acceptance rates, etc.).

The joint distribution of the parameter sets sampled from MCMC (θ_{MCMC}, or specifically, ϵ_{MCMC} and σ_{MCMC}) provide an estimate of the uncertainty in the ϵ and σ (see Figure 6 and 11 in Section 5). The integration required for robust posterior prediction (see Equations 15 and 19) is approximated from the distribution of QoI values evaluated for each MCMC parameter set, i.e. $QoI(\theta_{\mathrm{MCMC}}|D, M)$ or the more succinct notation QoI_{MCMC}. From QoI_{MCMC}, standard statistical methods are used to approximate the QoI uncertainty at a desired credible level. For example, the histograms and uncertainties reported in Section 5 are obtained from distributions of $\rho_{\mathrm{l,MCMC}}^{\mathrm{sat}}$, $P_{\mathrm{v,MCMC}}^{\mathrm{sat}}$, and Z_{MCMC}.

Because MCMC moves are accepted based on Equation 20 and the denominator in Equation 16 (i.e. $Pr(\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}|\lambda)$) does not depend on ϵ and σ, the acceptance probability is independent of $Pr(\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}|\lambda)$. Also, we use a "weakly informative prior" such that the acceptance probability is independent of $Pr(\epsilon, \sigma|\lambda)$. Specifically, $Pr(\epsilon, \sigma|\lambda)$ is uniform for all feasible values of ϵ and σ, i.e. the lower bound for the distribution is 0 (K or nm) and the upper bound is an order of magnitude larger than literature values for ϵ and σ. Furthermore, Q is chosen to be symmetric such that the Q terms in the numerator and denominator of Equation 20 cancel. Therefore, the probability of accepting ϵ_{i+1} and σ_{i+1} is based completely on the likelihood:

$$\alpha = \min\left(1, \frac{L(\epsilon_{i+1}, \sigma_{i+1}|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)}{L(\epsilon_i, \sigma_i|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)}\right) \tag{21}$$

For this reason, we discuss in some detail how we calculate $L(\epsilon, \sigma, |\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)$.

The likelihood is calculated using a normal distribution, $N(D, s_{\mathrm{D,SM}}^2)$, where $s_{\mathrm{D,SM}}^2$ is the combined variance of the experimental data and the surrogate model used to estimate

ρ_l^{sat} and P_v^{sat} for a given ϵ, σ, and λ (see Section 4.2). The variances are independent meaning that the combined variance is the sum of the experimental and surrogate model variances, i.e. s_D^2 and s_{SM}^2.[38] The experimental and surrogate model uncertainties (u_D and u_{SM}, respectively), are computed at the 95% confidence level using the respective s_D^2 and s_{SM}^2 values. Figure 5 compares the experimental data for ρ_l^{sat} and P_v^{sat} with the corresponding values of u_D and u_{SM}. The experimental data and uncertainties are obtained through the Thermodynamics Research Center (TRC) ThermoData Engine (TDE).[41] Section 4.2 discusses the key observation from this figure, namely, that the experimental uncertainties are negligible compared to the surrogate model uncertainties.
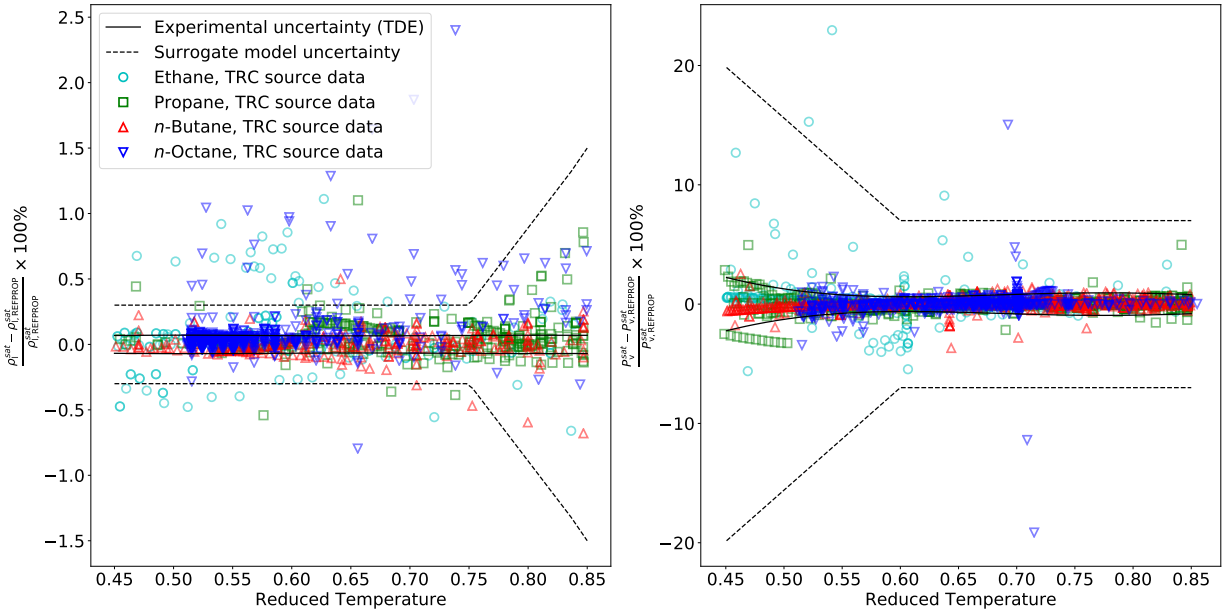
Figure 5: Surrogate model uncertainties are larger than experimental (TDE) uncertainties by at least a factor of five. Panel a) and b) plot the uncertainties for ρ_l^{sat} and P_v^{sat} with respect to reduced temperature (absolute temperature divided by the REFPROP T_c). Also included are percent deviations between the experimental data for ethane, propane, n-butane, and n-octane relative to the REFPROP correlations.

4.2 Surrogate Model

A typical Markov Chain requires $O(10^4–10^5)$ Monte Carlo steps, where the likelihood function must be evaluated at each step. Since $L(\theta|D, M)$ depends on the force field parameters (ϵ, σ, and λ), an MCMC approach is computationally infeasible if computing $L(\theta|D, M)$ requires performing direct molecular simulations for every proposed set of ϵ and σ. Furthermore, posterior predictive uncertainty propagation may require $O(10^2–10^3)$ θ_{MCMC} parameter sets for adequate representations of $Pr(QoI|D, M)$ (see Equations 15-19). For these reasons, surrogate models to estimate $\rho^{\mathrm{sat}}_{\mathrm{l,MCMC}}$, $P^{\mathrm{sat}}_{\mathrm{v,MCMC}}$ and Z_{MCMC} are essential for this study.

We use a configuration-sampling-based surrogate model, where configurations are sampled using a small group of reference parameter sets ($\epsilon_{\mathrm{ref}}, \sigma_{\mathrm{ref}}$, and λ_{ref}).[35] As discussed in our previous study, we use a single value of ϵ_{ref} with nine evenly spaced σ_{ref} values for a fixed value of λ_{ref}. Ensemble averages for the MCMC parameter sets (θ_{MCMC}) are estimated by reweighting the sampled reference configurations using Multistate Bennett Acceptance Ratio (MBAR).[42] MBAR is a nearly exact surrogate model when there are a sufficient number of effective samples.[35]

The properties that are estimated using MBAR are the departure internal energy (U^{dep}) and the compressibility factor (Z). As discussed in Section 2, Isothermal Isochoric Integration (ITIC) converts the MBAR estimated U^{dep} and Z values at the 19 ITIC state points to saturation temperatures (T^{sat}), saturated liquid densities ($\rho^{\mathrm{sat}}_{\mathrm{l}}$) and saturated vapor pressures ($P^{\mathrm{sat}}_{\mathrm{v}}$). This is important since $\rho^{\mathrm{sat}}_{\mathrm{l}}$ and $P^{\mathrm{sat}}_{\mathrm{v}}$ are the data (D) included in $L(\theta|D)$. Details for the implementation of MBAR and ITIC (MBAR-ITIC) is discussed elsewhere.[35]

The ITIC analysis provides VLE properties at only 5 saturation temperature values ($T^{\mathrm{sat}}_{\mathrm{ITIC}}$), while the experimental data set may have hundreds of saturation temperatures ($T^{\mathrm{sat}}_{\mathrm{D}}$). Although it is possible to use computed values from an empirical correlation fit to experimental data (i.e. REFPROP, ThermoData Engine (TDE)) as the data set, it is

considered best practice for Bayesian inference that D consist of raw experimental data. For this reason, we instead use empirical model fits to interpolate the ITIC VLE properties $(\rho_{\text{l,ITIC}}^{\text{sat}})$ and $(P_{\text{v,ITIC}}^{\text{sat}})$ so that $\rho_{\text{l}}^{\text{sat}}$ and $P_{\text{v}}^{\text{sat}}$ can be estimated at any saturation temperature. Specifically, we fit $P_{\text{v,ITIC}}^{\text{sat}}$ and $T_{\text{ITIC}}^{\text{sat}}$ to the Antoine equation:

$$\log_{10}(P_{\text{v}}^{\text{sat}}) = a_0 + \frac{a_1}{T^{\text{sat}} + a_2} \tag{22}$$

where a_i are fitting parameters. We fit $\rho_{\text{l,ITIC}}^{\text{sat}}$ and $T_{\text{ITIC}}^{\text{sat}}$ to a combined rectilinear and density scaling law expression:[24]

$$\rho_{\text{l}}^{\text{sat}} = b_0 + b_1(b_2 - T^{\text{sat}}) + b_3(b_2 - T^{\text{sat}})^\beta \tag{23}$$

where b_i are fitting parameters, and $\beta = 0.326$. b_0 and b_2 only provide rough estimates of the critical density (ρ_c) and critical temperature (T_c). More reliable estimates of the critical point require simultaneous fitting of $\rho_{\text{v,ITIC}}^{\text{sat}}$ to a similar expression, but this is unnecessary for our purposes since D does not include the critical constants. Note that Equations 22-23 are only used to interpolate ITIC values, and not to extrapolate to higher or lower T^{sat}. These equations are reliable over the limited temperature range studied ($0.45 < T_{\text{r}} < 0.85$), whereas a wider temperature range would require more flexible models.[43,44]

In summary, MBAR, ITIC, and Equations 22-23 enable prediction of $\rho_{\text{l}}^{\text{sat}}$ and $P_{\text{v}}^{\text{sat}}$ at any T^{sat} for any ϵ, σ, and λ by performing direct NVT simulations with only a few reference parameter sets. In addition, since the Mie λ-6 potential is linear with respect to r^{-6} and $r^{-\lambda}$ (see Equation 1), we implement basis functions to efficiently recompute the energies and forces that are required for MBAR and ITIC (for details see Section SI.IV of Supporting Information from Reference 35 and Reference 45). In total, this methodology reduces the computational cost for computing $L(\theta|D)$ by several orders of magnitude compared to direct simulation of VLE, using Gibbs Ensemble Monte Carlo (GEMC) or

Grand Canonical Monte Carlo (GCMC) histogram reweighting (HR).

Quantifying the surrogate model variance (s_{SM}^2) is essential for evaluating $L(\theta|D)$. Rather than performing a rigorous statistical assessment of MBAR, ITIC, and Equations 22-23, we use an empirical approach for estimating s_{SM}^2. Specifically, we compute the deviation between the surrogate model estimates for $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ values for TraPPE and Potoff with those reported in the literature for the respective force fields obtained using Gibbs Ensemble Monte Carlo (GEMC) or Grand Canonical Monte Carlo (GCMC) histogram reweighting (HR). Although this is a rough approximation for estimating s_{SM}^2, this comparison has the added benefit that it incorporates possible deviations associated with the simulation package, use of MD instead of MC, finite-size effects, and post-simulation analysis (e.g. ITIC rather than HR), which can be significant.[46] Details are provided in Section SI.IV of Supporting Information.

As shown in Figure 5, the surrogate model uncertainty (u_{SM}, reported at the 95% confidence level) for $\rho_{\mathrm{l}}^{\mathrm{sat}}$ is 0.3% up to $0.75T_c$ and increases linearly to 1.5% at the maximum T^{sat}. The surrogate model uncertainty for $P_{\mathrm{v}}^{\mathrm{sat}}$ is 20% at the minimum T_{sat} and decreases linearly to 7% at $0.6T_c$, where it remains constant for higher temperatures. Note that these are conservative estimates of u_{SM}, where other studies suggest ITIC can have significantly smaller uncertainties.[34] In fact, for the compounds investigated in this study, these uncertainties are much larger than the experimental uncertainties (u_{D}) and, therefore, the size of the parameter space sampled by MCMC depends almost entirely on u_{SM}. The use of a conservative u_{SM} model is intentional in this regard, namely, so that the θ_{MCMC} sampled points represent practically all of the feasible values of ϵ and σ for optimizing $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$.

5   Results

In this section, we use MCMC and the aforementioned surrogate models to determine the parameter uncertainty in CH_3 and CH_2 interaction sites of n-alkanes. Since the common practice is to limit λ to integer values (see Section 2.2), we perform several independent MCMC runs using a single, fixed, integer value of λ. The Bayesian inference analysis for CH_3 and CH_2 sites is performed sequentially. Specifically, rather than sampling from a four-dimensional parameter space (i.e. ϵ_{CH_3}, ϵ_{CH_2}, σ_{CH_3}, and σ_{CH_2} for a given value of λ_{CH_3} and λ_{CH_2}), we implement a sequential two-dimensional approach by assuming the CH_3 parameters from ethane are transferable to propane, n-butane, and n-octane.

5.1   Ethane

Figures 6-9 present the MCMC results for ethane with λ_{CH_3} = 13-18. Figure 6 demonstrates that the feasible region of ϵ_{CH_3} depends strongly on λ_{CH_3}, namely, larger values of λ_{CH_3} require larger values of ϵ_{CH_3}. By contrast, we observe a much smaller shift towards larger values of σ_{CH_3} with increasing λ_{CH_3}. This observation is consistent with the literature.[18]

The results depicted in Figure 7 Panels a)-b) and Figure 8 are useful for comparing the performance of different values of λ for ρ_l^{sat}, P_v^{sat}, and Z, respectively. Notice that the insets in Figure 7 plot the mean absolute percent deviation (MAPD%) to quantify the goodness of fit to VLE data, while the inset in Figure 8 plots the average deviation (AD%) to demonstrate the positive bias in P^{high}. Note that because MAPD% and AD% are percent deviations they are not directly related to the squared deviations of the normal distribution used to compute the likelihood. However, we plot MAPD% and AD% as these are easier to conceptualize and quantify.

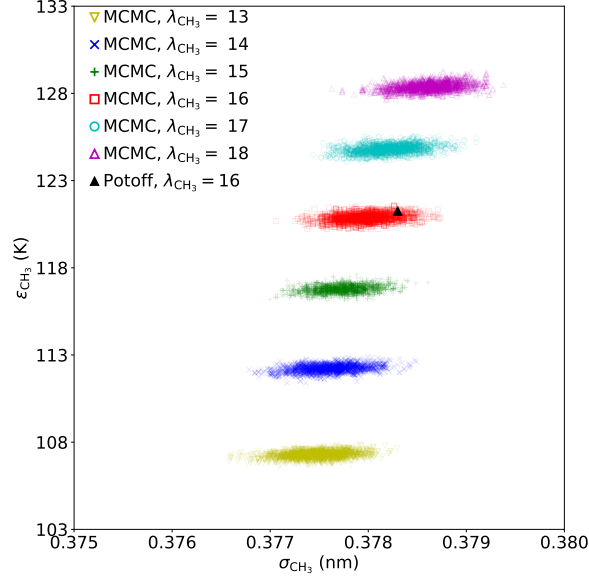Figure 7 Panel a) with the corresponding inset demonstrates that the best prediction of

Figure 6: MCMC sampled parameter sets for different values of λ_{CH_3} ($\epsilon_{CH_3,MCMC}$ and $\sigma_{CH_3,MCMC}$) quantify the uncertainty in ϵ_{CH_3}–σ_{CH_3}. The Potoff parameter set is included as a reference for $\lambda_{CH_3} = 16$.

ρ_l^{sat} is obtained for higher values of λ_{CH_3}. However, while the ρ_l^{sat} MAPD% for $\lambda_{CH_3} = 15$–18 are similar, $\lambda_{CH_3} = 13$–14 have significantly higher ρ_l^{sat} MAPD%. Figure 7 Panel b) demonstrates that $\lambda_{CH_3} = 13$–14 and $\lambda_{CH_3} = 17$–18 over- and under-predict P_v^{sat} at low temperatures, respectively, while $\lambda_{CH_3} = 15$–16 are the most reliable. $\lambda_{CH_3} = 15$ has the least amount of bias and the lowest MAPD% in P_v^{sat} (see inset).

Finally, Figure 8 demonstrates that all of the sampled $\epsilon_{CH_3,MCMC}$ and $\sigma_{CH_3,MCMC}$ parameter sets for $\lambda_{CH_3} \geq 14$ over-predict Z at high temperatures and densities (P^{high}). As expected, the larger the value of λ_{CH_3}, the more the force field over-predicts P^{high}.

While Figures 7-8 plot the results for each property individually, Figure 9 helps to visualize the overall performance of different values of λ_{CH_3} for simultaneously predicting all three properties of interest. In Panel a), notice the trade-off between the MAPD% of ρ_l^{sat} and P_v^{sat}. This compromise between two competing properties included in the objective function, namely, ρ_l^{sat} and P_v^{sat}, is known as a Pareto front.[23,47,48] The optimal location for a

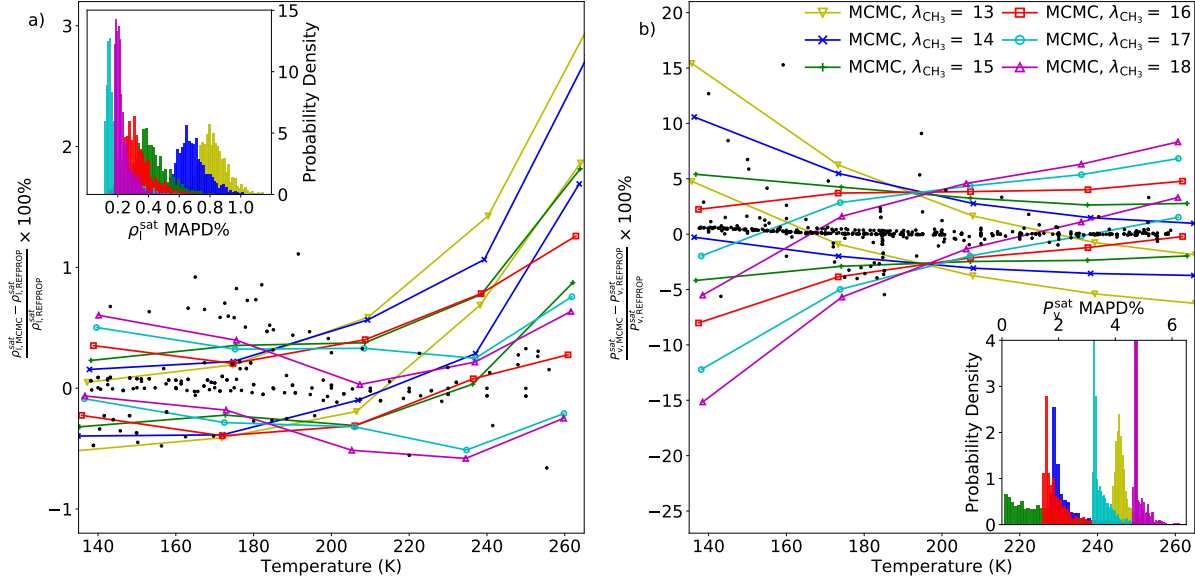Figure 7: Uncertainties in ϵ_{CH_3}–σ_{CH_3} propagate to ρ_l^{sat} and P_v^{sat}. Panels a)–b) plot the percent deviation from REFPROP correlations for $\rho_{l,MCMC}^{sat}$ and $P_{v,MCMC}^{sat}$, respectively. The upper and lower lines for each λ correspond to the 95% credible interval obtained from QoI_{MCMC}. Insets of Panels a)–b) are histograms of the MAPD% in $\rho_{l,MCMC}^{sat}$ and $P_{v,MCMC}^{sat}$, respectively. Experimental data used to compute the likelihood are included in Panels a)–b) as black dots.

Pareto front is the bottom left region of the plot (low MAPD% for both ρ_l^{sat} and P_v^{sat}) while the worst location is the top right region (high MAPD% for both ρ_l^{sat} and P_v^{sat}). Note that the inset of Panel a) includes an approximate "overall" Pareto front that combines the results for all values of λ_{CH_3}. Although not depicted for visual clarity, the "L" shaped frontier for different colors/symbols demonstrates that each λ_{CH_3} value also has its own Pareto front. Because the overall Pareto front consists of points from the $\lambda_{CH_3} = 15$–17 Pareto fronts, the Pareto optimal λ_{CH_3} value is either 15, 16, or 17, depending on the relative weight assigned to ρ_l^{sat} and P_v^{sat}. By contrast, since the $\lambda_{CH_3} = 13$, 14, and 18 Pareto fronts are completely inside the overall Pareto front, these λ_{CH_3} values are not optimal, regardless of the weighting.

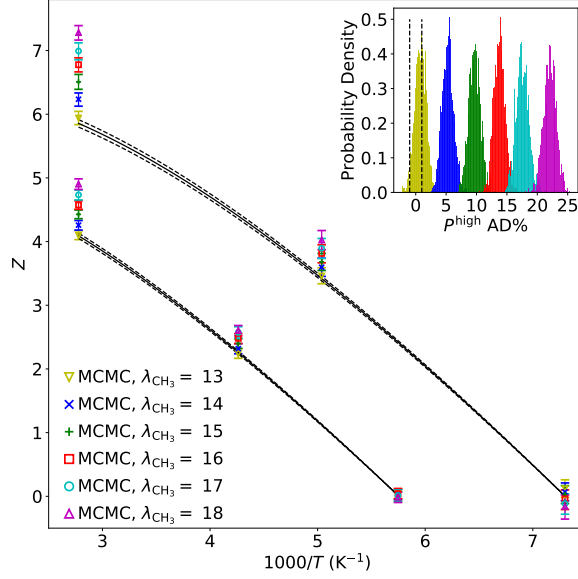Finally, and most importantly for our purposes, Panels b)-c) of Figure 9 demonstrate

Figure 8: Inadequacies of the UA Mie λ-6 potential are observed in Z for the two highest isochore densities (ρ_3 and ρ_4) by propagating the uncertainties in ϵ_{CH_3}–σ_{CH_3} for different values of λ_{CH_3}. The inset plots the distribution of average deviation (AD%) in P^{high}, i.e. P_{MCMC} for $\rho = \rho_3$–ρ_4 and $T = T^{IT}$. REFPROP uncertainty in P^{high} is $\pm 1\%$.

the depreciation in ρ_l^{sat} and P_v^{sat} that is required to achieve more accurate prediction of P^{high}. For example, although $\lambda_{CH_3} = 15$ and 16 are the best values based on VLE data, they over-predict P^{high} by around 10 and 15%, respectively. By contrast, while $\lambda_{CH_3} = 13$ is the most accurate for P^{high}, the MAPD% for ρ_l^{sat} and P_v^{sat} are, respectively, 4 and 40 factors larger than the optimal MAPD%. This supports the fundamental claim of this work, namely, that the UA Mie λ-6 potential cannot adequately predict both VLE and high pressures for supercritical fluids and compressed liquids.

To statistically verify this statement, Figure 10 provides the Bayes factor for each value of λ_{CH_3} (normalized with respect to $\lambda_{CH_3} = 14$) based solely on ρ_l^{sat} and P_v^{sat}. Bayes Factors between 1–3.2, 3.2–10, 10–32, 32–100, and > 100 are typically classified as "not substantial", "substantial", "strong", "very strong", and "decisive" evidence, respectively.[49] Therefore, there is "decisive" evidence against the use of $\lambda_{CH_3} = 13$ for predicting ρ_l^{sat} and P_v^{sat}. Since $\lambda_{CH_3} = 13$ is the only value that predicts P^{high} within the REFPROP uncer-
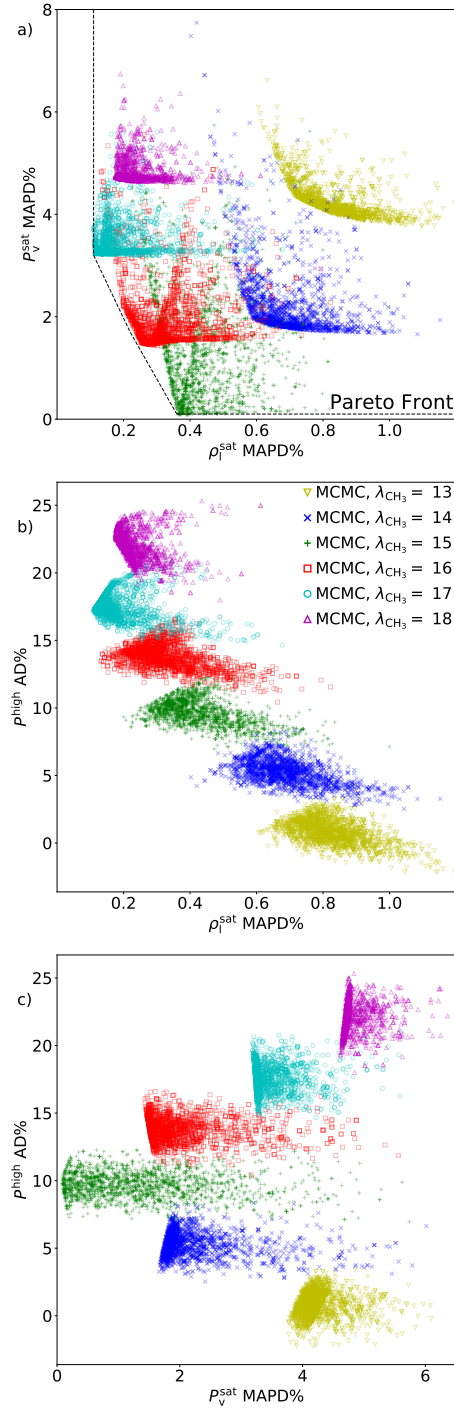
Figure 9: MCMC results confirm that the UA Mie λ-6 potential cannot adequately predict both VLE and high pressures for supercritical fluids and compressed liquids. Panel a) plots the mean absolute percent deviation (MAPD%) of $\rho^{\text{sat}}_{\text{l,MCMC}}$ and $P^{\text{sat}}_{\text{v,MCMC}}$. Panels b)-c) plot the average deviation (AD%) in P^{high} with respect to MAPD% of $\rho^{\text{sat}}_{\text{l,MCMC}}$ and $P^{\text{sat}}_{\text{v,MCMC}}$, respectively.

33

tainty, we conclude that no set of $\epsilon_{\mathrm{CH_3}}$, $\sigma_{\mathrm{CH_3}}$, and $\lambda_{\mathrm{CH_3}}$ can predict *both* VLE and P^{high}.
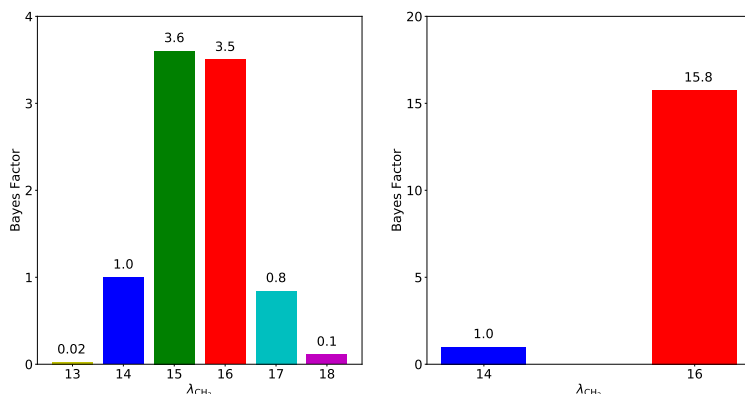

Figure 10: Evidence (Bayes factor) in Panel a) supports $\lambda_{\mathrm{CH_3}} = 15$ or 16, while evidence in Panel b) strongly supports $\lambda_{\mathrm{CH_2}} = 16$ over $\lambda_{\mathrm{CH_2}} = 14$. $\mathrm{CH_3}$ values depend only on ethane while $\mathrm{CH_2}$ values are based on propane, *n*-butane, and *n*-octane. Note that all values are normalized with respect to $\lambda = 14$.

In addition, there is "very strong" evidence that the 18-6 potential is not justified by VLE data. The evidence in favor of the 15-6 and 16-6 potentials over the 14-6 and 17-6 potentials is not as definitive, although it is still considered "substantial." By contrast, the evidence for $\lambda_{\mathrm{CH_3}} = 15$ instead of $\lambda_{\mathrm{CH_3}} = 16$ is "not substantial."

It is important to mention that these Bayes factors depend primarily on the VLE data and the $s_{\mathrm{D,SM}}^2$ model used to compute $L(\theta|D)$. We use a very conservative error model for $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ (see Figure 5) so that our MCMC samples cover a large region of parameter space. This is done primarily to demonstrate that the UA Mie λ-6 is inadequate for predicting VLE and P^{high}. However, a less conservative error model would provide more convincing evidence regarding the optimal λ value based solely on VLE data.

Also, recall that ITIC is limited to $T^{\mathrm{sat}} < 0.85 T_c$. Therefore, it is possible that the optimal value of $\lambda_{\mathrm{CH_3}}$ could be deduced (i.e. larger Bayes factors) if higher temperature VLE data were included (say from 260–290 K). Based on the observed bias in $\rho_{\mathrm{l}}^{\mathrm{sat}}$ at higher temperatures (240–260 K) for $\lambda_{\mathrm{CH_3}} = 14$, it appears that higher temperature VLE data

would strengthen the counter evidence against the 14-6 potential. It is unclear whether higher temperature data would support the 15-6 or 16-6 potential, likely the optimal λ_{CH_3} is a non-integer value between 15 and 16. Implementing MBAR with GCMC may be necessary to include VLE data from $0.85 < T_r^{sat} < 0.95$.

5.2 Larger *n*-alkanes

This section demonstrates that the conclusions regarding the UA Mie λ-6 potential for ethane are general for larger *n*-alkanes. Specifically, we observe that improved accuracy in predicting VLE requires a larger value of λ_{CH_2}. However, this improvement comes at the cost of significantly over-predicting P^{high}. Figure 11 presents the MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets with Panels a) and b) corresponding to $\lambda_{CH_2} = 16$ and $\lambda_{CH_2} = 14$, respectively. Note that these results were obtained using fixed values of ϵ_{CH_3}, σ_{CH_3}, and λ_{CH_3}, where $\lambda_{CH_3} = \lambda_{CH_2}$. The values of ϵ_{CH_3} and σ_{CH_3} are the maximum likelihood parameter set from ethane for the corresponding λ_{CH_3} value.

Notice in Figure 11 that the MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets, for a given value of λ_{CH_2}, overlap considerably for propane, *n*-butane, and *n*-octane. This provides statistical evidence to support the common assumption of transferability of CH_2 parameters between different *n*-alkanes. To further demonstrate this point, Figure 11 includes the MCMC results when we use the combined likelihoods from all three compounds, referred to as "MCMC transferable."

Panel a) shows that the Potoff CH_2 parameter set is within the MCMC sample regions for $\lambda_{CH_2} = 16$. The same result was also observed for ethane (see Figure 6). This suggests that the Potoff CH_3 and CH_2 parameters are supported by the VLE data used in this study, even though the Potoff force field was parameterized using VLE data in a higher temperature range $(0.6 < T_r^{sat} < 0.95)$.

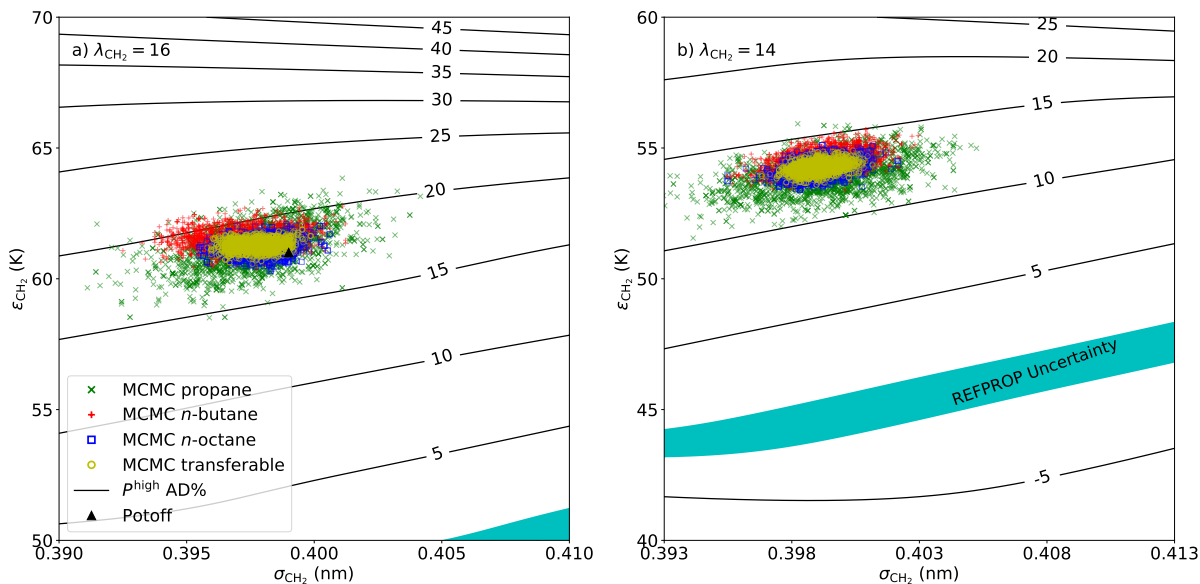Also, note that the uncertainty in the parameters is largest for propane and smallest

Figure 11: MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets result in large AD% for P^{high}. Panels a) and b) correspond to $\lambda_{CH_2} = 16$ and $\lambda_{CH_2} = 14$, respectively. Contours are the AD% in P^{high} relative to the REFPROP correlations, where the "REFPROP uncertainty" region represents $\pm 1\%$ deviation. Panel a) includes the Potoff parameter set as a reference for $\lambda_{CH_2} = 16$.

for n-octane. This suggests that the sensitivity of ρ_l^{sat} and P_v^{sat}, with respect to the CH_2 parameters, increases with increasing number of CH_2 interaction sites. Although this result is fairly intuitive, it is a valuable insight when selecting a training set of molecules for force field development. For example, notice that the MCMC transferable region is almost identical to that of n-octane, this suggests that propane and n-butane provide relatively little additional information.

For the purposes of this study, it is important to focus on the contours in Figure 11, which correspond to AD% in P^{high}. Notice that the MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets have a strong positive bias (i.e. large AD%) in P^{high}. Specifically, $\lambda_{CH_2} = 16$ and $\lambda_{CH_2} = 14$ have, respectively, AD% of $\approx 16\text{-}21$ and $\approx 10\text{-}15$, much greater than the REFPROP uncertainty of around 1%. It is interesting that, for corresponding values of λ, the AD% for these larger n-alkanes is higher than that of ethane. This suggests that longer

chain-lengths, with a UA Mie λ-6 force field, exacerbate the erroneous PVT trend at high pressures.

Although the AD% in P^{high} is slightly lower for $\lambda_{\text{CH}_2} = 14$ than for $\lambda_{\text{CH}_2} = 16$, the UA Mie 14-6 potential is significantly less reliable for VLE. Figure 10 demonstrates the "strong" evidence, based on VLE data, for $\lambda_{\text{CH}_2} = 16$ over $\lambda_{\text{CH}_2} = 14$. Note that the evidence in Figure 10 for CH_2 is likely stronger than CH_3 because VLE data from more than one compound were included. Furthermore, because the "REFPROP uncertainty" contours are roughly parallel to the MCMC region and found at much lower ϵ_{CH_2} values, it is necessary to sacrifice considerable accuracy in ρ_l^{sat} and P_v^{sat} in order to accurately predict P^{high}. In conclusion, these results suggest that neither UA Mie 16-6 or 14-6 force fields are capable of predicting VLE and PVT for supercritical fluids and compressed liquids of n-alkanes.

6 Discussion

6.1 Recommendations

The Mie λ-6 potential is capable of reproducing ρ_l^{sat} and P_v^{sat} for $\lambda = 15$ or 16. However, this improved accuracy does not appear to have physical/theoretical justification. For this reason, λ values greater than 13 over-predict pressures for supercritical fluids and compressed liquids. Therefore, for developing FEOS of normal and branched alkanes, we recommend further investigation of alternative potentials with a softer repulsive barrier and a more sound theoretical basis, e.g. Buckingham exponential-6, Morse,[28–30] or an extended Lennard-Jones.[34,50,51]

Although the UA Mie λ-6 potential is not quantitatively reliable at high pressures, it may still be of use for FEOS parameterization when considering the insight gained in

this study. For example, since the Potoff force field consistently over-predicts high pressures, a non-linear FEOS optimization could utilize the simulation results as an upper constraint for the FEOS pressure.[3] Furthermore, the primary purpose to include molecular simulation data for FEOS development is to increase the range of validity by ensuring good behavior of the FEOS at high temperatures and pressures. While FEOS are based on empirical equations with 10–30 fitting parameters, even an inaccurate force field has a more sound theoretical basis and, therefore, the UA Mie λ-6 simulation output for a given property should not demonstrate non-physical oscillations, inflection points, derivative sign-changes, etc., which can plague a poorly-fit FEOS.

Essentially, whether or not a FEOS should be developed using a hybrid data set consisting of UA Mie λ-6 simulation results depends on the quality and quantity of available experimental data. If the data cover a wide range of state points and properties, it is possible that the UA Mie λ-6 potential may still be useful, despite the systematic deviations at high pressures. By contrast, if the experimental data are limited such that the FEOS depends almost entirely on the molecular simulation results, the UA Mie λ-6 force field will lead to large deviations at high pressures.

6.2 Limitations

This section discusses limitations to the primary conclusion from this study, namely, that UA Mie λ-6 force fields parameterized with VLE data should not be used to develop fundamental equations of state for normal and branched alkanes. The main limitation is that the poor extrapolation at high pressures is based solely on the trend of Z with respect to inverse temperature. By contrast, the simulation values that are typically included in hybrid data sets are derivatives of the residual (or departure) Helmholtz free energy with

respect to inverse temperature and/or density:[7–11]

$$A_{xy}^{\text{dep}} R_g T \equiv (1/T)^x \rho^y \frac{\partial^{x+y} A^{\text{dep}}}{\partial (1/T)^x \partial \rho^y} \qquad (24)$$

We would like to emphasize the advantage of using A_{xy}^{dep} for developing FEOS, as this approach eliminates redundant information found in traditional macroscopic properties.[2,4–7] For example, the following expressions demonstrate the interdependency of the properties we computed, namely, Z and U^{dep} with their derivatives along isochores and isotherms:[6]

$$Z = 1 + A_{01}^{\text{dep}} \qquad (25)$$

$$\frac{1}{T} \left(\frac{-\partial Z}{\partial (1/T)} \right)_\rho = 1 + A_{01}^{\text{dep}} - A_{11}^{\text{dep}} \qquad (26)$$

$$\frac{U^{\text{dep}}}{R_g T} = A_{10}^{\text{dep}} \qquad (27)$$

$$\frac{1}{R_g} \left(\frac{\partial U^{\text{dep}}}{\partial T} \right)_\rho = -A_{20}^{\text{dep}} \qquad (28)$$

$$\rho \left(\frac{\partial Z}{\partial \rho} \right)_T = 1 + 2A_{01}^{\text{dep}} + A_{02}^{\text{dep}} \qquad (29)$$

Unfortunately, with the exception of *ms2*,[52] we are not aware of any open-source simulation packages that readily provide A_{xy}^{dep}.

In addition, macroscopic properties, such as Z and U^{dep} (with their respective derivatives), are more readily understood and visualized than A_{xy}^{dep}. Also, it is easier to quantify their impact on process design than derivatives in the residual Helmholtz free energy. For example, as demonstrated by Thol et al., an inaccurate prediction of some A_{xy}^{dep} does not necessarily result in poor prediction of PVT behavior or heat capacities.[8] For these reasons, we did not base our conclusions on the adequacy of the UA Mie λ-6 potential to predict A_{xy}^{dep}.

Instead, we have presented Z and, by inspection, the slope of Z with respect to $1/T$ at constant ρ. Since these properties are equivalent to Equations 25-26, respectively, we have indirectly focused on only two of the Helmholtz derivatives, namely, A_{01}^{dep} and A_{11}^{dep}. However, we also observed some deviations in U^{dep}, the slope of U^{dep} with respect to T at constant ρ, and the slope of Z with respect to ρ at constant T, which are equivalent to Equations 27-29. Unfortunately, deficiencies are not as obvious for these properties due to the relatively large uncertainties in the REFPROP correlations, ca. 5% and 10% for U^{dep} and $\left(\frac{\partial U^{\mathrm{dep}}}{R_g \partial T}\right)_\rho$. Although these additional properties provide insight regarding A_{10}^{dep}, A_{20}^{dep}, A_{01}^{dep}, and A_{02}^{dep}, because the results are less conclusive they are provided only in Section SI.V of Supporting Information. Furthermore, the relationship between Equations 27-29 and the repulsive barrier, λ, is not obvious. By contrast, the observation that larger values of λ over-predict Z with increasing temperature is clear evidence that $\lambda > 12$ is not theoretically justified but, rather, is an empirical remedy that performs well for VLE.

Future work should investigate more thoroughly the adequacy of UA Mie λ-6 (or other) force fields to predict U^{dep} and isochoric/isobaric heat capacities at high temperatures and pressures. For example, rigorous studies of this type have been performed on the single-site Lennard-Jones fluid.[2–4]

Another limitation is that we utilize a single layer Bayes model as opposed to a hierarchical model.[31,40] Wu et al. demonstrated that it is important to use hierarchical models when there are discrepancies in the data set, D. Fortunately, since we use a conservative estimate for the surrogate model uncertainty, i.e. $u_{\mathrm{SM}} \gg u_{\mathrm{D}}$, any discrepancies in the VLE data should not affect the parameter uncertainties.

The hierarchical approach is also useful when accounting for model inadequacies, i.e. when the force field is not capable of representing multiple data types. A hierarchical method should thus be favored if determining the parameter uncertainty when simultaneously considering ρ_l^{sat}, P_v^{sat} and P^{high}. Furthermore, a hierarchical model should be

used if the parameters are not transferable between molecules, i.e. if the CH_2 parameters were different for propane, n-butane, and n-octane. However, the hierarchical approach is unnecessary for our purposes, since the transferable UA Mie λ-6 force field for n-alkanes is capable of reproducing ρ_l^{sat} and P_v^{sat}, which are the only properties included in D.

7 Conclusions

Recently, molecular simulation results at extreme temperatures and pressures have been used to supplement experimental data when developing fundamental equations of state for compounds with limited experimental data. For this hybrid data set approach to work, it is imperative that the force field be reliable and transferable over different PVT conditions. Unfortunately, literature united-atom force fields that are highly accurate for estimating VLE properties of normal and branched alkanes have systematic deviations in Z at non-VLE conditions. Bayesian inference suggests that the UA Mie λ-6 model type is not adequate for simultaneously predicting saturated liquid density, vapor pressure, and Z at high pressures. In the case of ethane, evidence from VLE data supports $14 \leq \lambda \leq 17$, while Z at high pressures requires $\lambda = 13$. A similar trend is observed for larger n-alkanes. Specifically, evidence from VLE data supports $\lambda = 16$, while we observe slight improvement in Z at high pressures for $\lambda = 14$. Therefore, while considerable improvement in VLE is observed for $\lambda > 12$, the use of UA Mie λ-6 potentials over the traditional UA Lennard-Jones 12-6 potential appears to be more empirical than theoretical. For these reasons, we recommend that alternative models be considered for developing FEOS of normal and branched alkanes, such as force fields that use anisotropic-united-atom, all-atom, and/or alternative non-bonded potentials.

Acknowledgments

References

(1) Lemmon, E. W.; Huber, M. L.; McLinden, M. O. NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 9.1, National Institute of Standards and Technology. 2013; https://www.nist.gov/srd/refprop.

(2) Thol, M.; Rutkai, G.; Köster, A.; Lustig, R.; Span, R.; Vrabec, J. *Journal of Physical and Chemical Reference Data* **2016**, *45*, 023101.

(3) Thol, M.; Rutkai, G.; Span, R.; Vrabec, J.; Lustig, R. *International Journal of Thermophysics* **2015**, *36*, 25–43.

(4) Rutkai, G.; Thol, M.; Span, R.; Vrabec, J. *Molecular Physics* **2017**, *115*, 1104–1121.

(5) Lustig, R.; Rutkai, G.; Vrabec, J. *Molecular Physics* **2015**, *113*, 910–931.

(6) Rutkai, G.; Vrabec, J. *Journal of Chemical & Engineering Data* **2015**, *60*, 2895–2905.

(7) Rutkai, G.; Thol, M.; Lustig, R.; Span, R.; Vrabec, J. *The Journal of Chemical Physics* **2013**, *139*, 041102.

(8) Thol, M.; Dubberke, F.; Rutkai, G.; Windmann, T.; Köster, A.; Span, R.; Vrabec, J. *Fluid*

Phase Equilibria **2016**, *418*, 133 – 151, Special Issue covering the Nineteenth Symposium on Thermophysical Properties.

(9) Thol, M.; Rutkai, G.; Köster, A.; Dubberke, F. H.; Windmann, T.; Span, R.; Vrabec, J. *Journal of Chemical & Engineering Data* **2016**, *61*, 2580–2595.

(10) Thol, M.; Rutkai, G.; Köster, A.; Miroshnichenko, S.; Wagner, W.; Vrabec, J.; Span, R. *Molecular Physics* **2017**, *115*, 1166–1185.

(11) Thol, M.; Rutkai, G.; KÃűster, A.; Kortmann, M.; Span, R.; Vrabec, J. *Chemical Engineering Science* **2015**, *121*, 87 – 99, 2013 Danckwerts Special Issue on Molecular Modelling in Chemical Engineering.

(12) Martin, M. G.; Siepmann, J. I. *The Journal of Physical Chemistry B* **1998**, *102*, 2569–2577.

(13) Martin, M. G.; Siepmann, J. I. *The Journal of Physical Chemistry B* **1999**, *103*, 4508–4517.

(14) Shah, M. S.; Siepmann, J. I.; Tsapatsis, M. *AIChE Journal* **2017**, *63*, 5098–5110.

(15) Errington, J. R.; Panagiotopoulos, A. Z. *The Journal of Physical Chemistry B* **1999**, *103*, 6314–6322.

(16) Ungerer, P.; Beauvais, C.; Delhommelle, J.; Boutin, A.; Rousseau, B.; Fuchs, A. H. *The Journal of Chemical Physics* **2000**, *112*, 5499–5510.

(17) Bourasseau, E.; Ungerer, P.; Boutin, A.; Fuchs, A. H. *Molecular Simulation* **2002**, *28*, 317–336.

(18) Potoff, J. J.; Bernard-Brunel, D. A. *The Journal of Physical Chemistry B* **2009**, *113*, 14725–14731.

(19) Mick, J. R.; Soroush Barhaghi, M.; Jackman, B.; Schwiebert, L.; Potoff, J. J. *Journal of Chemical & Engineering Data* **2017**, *62*, 1806–1818.

(20) Hemmen, A.; Gross, J. *The Journal of Physical Chemistry B* **2015**, *119*, 11695–11707.

(21) Weidler, D.; Gross, J. *Industrial & Engineering Chemistry Research* **2016**, *55*, 12123–12132.

(22) Herdes, C.; Totton, T. S.; Müller, E. A. *Fluid Phase Equilibria* **2015**, *406*, 91 – 100.

(23) Stöbener, K.; Klein, P.; Horsch, M.; Kufer, K.; Hasse, H. *Fluid Phase Equilibria* **2016**, *411*, 33 – 42.

(24) Messerly, R. A.; KnottsIV, T. A.; Wilding, W. V. *The Journal of Chemical Physics* **2017**, *146*, 194110.

(25) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press ; Oxford University Press: Oxford England New York, 1987; pp xix, 385 p.

(26) Gordon, P. A. *The Journal of Chemical Physics* **2006**, *125*, 014504.

(27) Galliéro, G.; Boned, C.; Baylaucq, A.; Montel, F. m. c. *Phys. Rev. E* **2006**, *73*, 061201.

(28) Rowley, R. L.; Pakkanen, T. *The Journal of Chemical Physics* **1999**, *110*, 3368–3377.

(29) Rowley, R. L.; Yang, Y.; Pakkanen, T. A. *The Journal of Chemical Physics* **2001**, *114*, 6058–6067.

(30) Hayes, J. M.; Greer, J. C.; MortonâĂŞBlake, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1953–1966.

(31) Kulakova, L.; Arampatzis, G.; Angelikopoulos, P.; Hadjidoukas, P.; Papadimitriou, C.; Koumoutsakos, P. *Scientific Reports* **2017**, *7*, 16576.

(32) Mick, J. R.; Soroush Barhaghi, M.; Jackman, B.; Rushaidat, K.; Schwiebert, L.; Potoff, J. J. *The Journal of Chemical Physics* **2015**, *143*, 114504.

(33) Abraham, M.; van der Spoel, D.; Lindahl, E.; B.Hess,; the GROMACS development team, GROMACS User Manual version 2018, www.gromacs.org (2018). 2018.

(34) Razavi, S. M. Optimization of a Transferable Shifted Force Field for Interfaces and In-homogenous Fluids using Thermodynamic Integration. M.Sc. thesis, The University of Akron, 2016.

(35) Messerly, R. A.; Shirts, M. R. *Journal of Chemical Theory and Computation* **2018**, *Pending publication.*

(36) Nieto-Draghi, C.; Bocahut, A.; Creton, B.; Have, P.; Ghoufi, A.; Wender, A.; ; Boutin, A.; Rousseau, B.; Normand, L. *Molecular Simulation* **2008**, *34*, 211–230.

(37) Rizzi, F.; Najm, H. N.; Debusschere, B. J.; Sargsyan, K.; Salloum, M.; Adalsteins-son, H.; Knio, O. M. *Multiscale Modeling & Simulation* **2012**, *10*, 1460–1492.

(38) Angelikopoulos, P.; Papadimitriou, C.; Koumoutsakos, P. *The Journal of Chemical Physics* **2012**, *137*, 144103.

(39) Cailliez, F.; Pernot, P. *The Journal of Chemical Physics* **2011**, *134*, 054124.

(40) Wu, S.; Angelikopoulos, P.; Papadimitriou, C.; Moser, R.; Koumoutsakos, P. *Philo-sophical Transactions of the Royal Society of London A: Mathematical, Physical and Engi-neering Sciences* **2016**, *374*.

(41) Frenkel, M.; Chirico, R. D.; Diky, V.; Yan, X.; Dong, Q.; Muzny, C. *Journal of Chemical Information and Modeling* **2005**, *45*, 816–838.

(42) Shirts, M. R.; Chodera, J. D. *The Journal of Chemical Physics* **2008**, *129*, 124105.

(43) Vetere, A. *Industrial & Engineering Chemistry Research* **1991**, *30*, 2487–2492.

(44) Funke, M.; Kleinrahm, R.; Wagner, W. *The Journal of Chemical Thermodynamics* **2002**, *34*, 2017 – 2039.

(45) Naden, L. N.; Shirts, M. R. *Journal of Chemical Theory and Computation* **2016**, *12*, 1806–1823.

(46) Schappals, M.; Mecklenfeld, A.; Kröger, L.; Botan, V.; Köster, A.; Stephan, S.; García, E. J.; Rutkai, G.; Raabe, G.; Klein, P.; Leonhard, K.; Glass, C. W.; Lenhard, J.; Vrabec, J.; Hasse, H. *Journal of Chemical Theory and Computation* **2017**, *13*, 4270–4280.

(47) Stöbener, K.; Klein, P.; Reiser, S.; Horsch, M.; Kufer, K.-H.; Hasse, H. *Fluid Phase Equilibria* **2014**, *373*, 100 – 108.

(48) Werth, S.; Stöbener, K.; Klein, P.; Kufer, K.-H.; Horsch, M.; Hasse, H. *Chemical Engineering Science* **2015**, *121*, 110 – 117, 2013 Danckwerts Special Issue on Molecular Modelling in Chemical Engineering.

(49) Jeffreys, H. *The theory of probability. Third edition*; Clarendon Press: Oxford, 2004.

(50) Hajigeorgiou, P. G. *Journal of Molecular Spectroscopy* **2016**, *330*, 4 – 13, Potentiology and Spectroscopy in Honor of Robert Le Roy.

(51) Kalos, F.; Grosser, A. E. *Canadian Journal of Chemistry* **1972**, *50*, 892–896.

(52) Rutkai, G.; K oster, A.; Guevara-Carrion, G.; Janzen, T.; Schappals, M.; Glass, C. W.; Bernreuther, M.; Wafai, A.; Stephan, S.; Kohns, M.; Reiser, S.; Deublein, S.; Horsch, M.; Hasse, H.; Vrabec, J. *Computer Physics Communications* **2017**, *221*, 343 – 351.