United-atom, Mie λ-6 force fields for normal and branched alkanes extrapolate poorly to high pressures when parameterized using vapor-liquid equilibria properties. To be submitted to the Journal of Physical Chemistry, B.

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

Contribution of NIST, an agency of the United States government; not subject to copyright in the United States.

Developing fundamental equations of state requires larges amounts of data over a wide range of temperatures and pressures. Typically, the range of experimental data is limited by cost, safety, and feasibility. Molecular simulation provides an avenue for estimating thermophysical properties at extreme temperatures and pressures. The reliability of this hybrid data set approach (from experiment and molecular simulation) depends on the validity of the force field implemented. Most force fields are optimized to agree with vapor-liquid equilibria data. However, as demonstrated in this study, although state-of-the-art united-atom Mie λ-6 potentials provide accurate estimates for vapor-liquid equilibria properties, they tend to over-predict pressures at high temperatures and densities. The theoretical explanation for this observation is that the repulsive barrier is too steep for the "optimal" united-atom Mie λ-6 potential parameterized with VLE data. 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 for dense supercritical fluids and compressed liquids.

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 predictions 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 near 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

3

data exist covering a wide range of PVT phase space. For these reasons, REFPROP contains 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, i.e. if the VLE optimal parameters are transferable to higher temperatures and pressures. For small compounds, such as noble gases, hydrogen sulfide, and hydrogen chloride, force fields optimized to experimental VLE data adequately predict 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 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 classification of 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 the additional adjustable parameter controls the steepness of the repulsive barrier.

Previous work demonstrated that the UA LJ 12-6 potential cannot adequately estimate both ρ_l^{sat} and P_v^{sat} for n-alkanes.[22,23] For this reason, the TraPPE-UA force field was primarily developed to agree with saturated liquid densities.[12] By contrast, accurate prediction of both ρ_l^{sat} and P_v^{sat} over a wide temperature range is possible by varying the repulsive exponent of the LJ potential (i.e. the Mie λ-6 potential). 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").[24] (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.[24,25])

However, there are some theoretical concerns that increasing the repulsive exponent might have some undesirable consequences, especially at high pressures, where close range interactions will become more prevalent 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* 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 hard, and only slight improvement in the repulsive region is observed for the LJ 9-6 potential. (As a brief aside, 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 (radial distribution functions of liquid and vapor at varying temperatures and densities) support λ values between 6 and 10, while quantum argon dimer energies support λ values between 12 and 14.[26] 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 $\lambda \geq 12$.[25,27] 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.[25,26] 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 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 manuscript 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 from 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), isopentane (2-methylbutane), isohexane (2-methylpentane), isooctane (2,2,4-trimethylpentane), and neopentane (2,2-dimethylpropane).

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.[31] 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. 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 as 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 $1/T^{\mathrm{IT}}$ and $1/T^{\mathrm{sat}}$. Thus, a total of 19 simulations are per-

formed 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 Supporting Information.
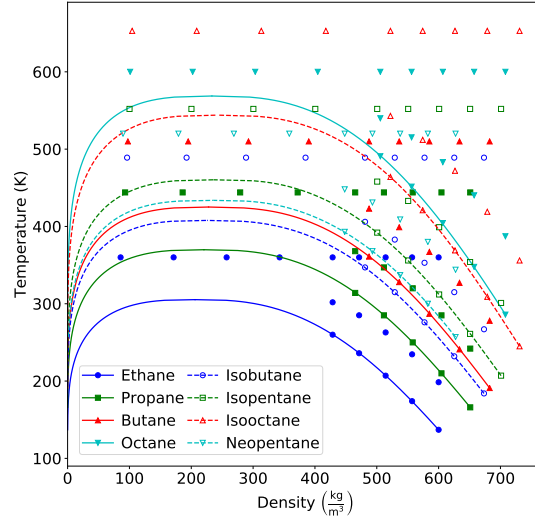

Figure 1: State points simulated for each compound studied. A total of 19 simulations are performed: nine densities along the supercritical isotherm and two temperatures along five liquid density isochores. Filled symbols and solid lines correspond to *n*-alkanes, while empty symbols and dashed lines correspond to branched alkanes. The REFPROP saturation curve for each compound is included as a reference.

Isothermal Isochoric Integration (ITIC) converts the departure internal 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}.[32,33] The equations for ITIC are:

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

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

$$P_v^{\mathrm{sat}} \approx \left(1 + B_2\rho_v^{\mathrm{sat}} + B_3(\rho_v^{\mathrm{sat}})^2\right)\rho_v^{\mathrm{sat}} R_g T^{\mathrm{sat}} \tag{3}$$

$$Z_l^{\text{sat}} = \frac{P_v^{\text{sat}}}{\rho_l^{\text{sat}} R_g T^{\text{sat}}} \tag{4}$$

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.[33] As discussed in our previous work,[33] the B_2 and B_3 values found in Equations 2-3 are calculated using REFPROP correlations.[1] The use of REFPROP correlations introduces a small bias in the resulting ρ_l^{sat} and P_v^{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. 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 containing 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.

Table 1: Effective bond-lengths in units of nm for terminal (CH_3) UA or AUA interaction sites. "Not-applicable" ("N/A") signifies that the force field either does not include these site types (e.g. Exp-6 and TraPPE-2) or that a more complicated notation than a simple effective bond-length is required to adequately represent the force field (i.e. AUA4).

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	N/A	N/A
CH_3-CH	0.154	0.174	N/A	N/A	N/A
CH_3-C	0.154	0.174	N/A	0.1751	N/A

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:

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

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:

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

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 twice 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. The Mie λ-6 potential is:

$$u^{\text{vdw}}(\epsilon, \sigma, \lambda; r) = \left(\frac{\lambda}{\lambda - 6}\right)\left(\frac{\lambda}{6}\right)^{\frac{6}{\lambda-6}} \epsilon \left[\left(\frac{\sigma}{r}\right)^{\lambda} - \left(\frac{\sigma}{r}\right)^6\right] \tag{5}$$

where u^{vdw} is the van der Waals interaction, σ is the distance (r) where $u^{\text{vdw}} = 0$, $-\epsilon$ is the energy of the potential at the minimum $\left(\text{i.e. } u^{\text{vdw}} = -\epsilon \text{ and } \frac{\partial u^{\text{vdw}}}{\partial r} = 0 \text{ for } r = r_{\text{min}}\right)$, 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. 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.[34] For the same reason (i.e. computational efficiency), a common practice to date is to use integer values of λ in Equation 5. The non-bonded force field parameters for TraPPE (and TraPPE-2), Potoff, AUA4, and TAMie are provided in Table 4.

Table 4: Non-bonded (intermolecular) parameters for TraPPE[12,13] (and TraPPE-2[14]), Potoff,[18,19] AUA4,[16,35] 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	N/A	N/A	N/A

The Errington force field utilizes an exponential-6 model:

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

where u^{vdw}, ϵ, and r are the same as in Equation 5, $r_{\min}$ is the distance that corresponds to the minimum in the potential (i.e $u^{\text{vdw}}(r_{\min}) = -\epsilon$), α is a Buckingham exponential-6 parameter, and $r_{\max}$ is the smallest positive value for which $\frac{du^{\text{vdw}}}{dr} = 0$. The Errington 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 6 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[34] for ϵ and σ, an arithmetic mean for the repulsive exponent (λ) (as recommended by Potoff and Bernard-Brunel[18]), and a geometric mean for α (as recommended by Errington and Panagiotopoulos[15]):

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{7}$$

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{8}$$

$$\lambda_{ij} = \frac{\lambda_{ii} + \lambda_{jj}}{2} \tag{9}$$

$$\alpha_{ij} = \sqrt{\alpha_{ii}\alpha_{jj}} \tag{10}$$

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 2 and 3 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 3 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 the Supporting Information.

Figure 2 demonstrates that the existing literature force fields for n-alkanes, while accurate for VLE, do not capture the correct PVT behavior at high pressures, i.e. the higher temperatures and highest isochore densities (ρ_0 and ρ_1). Figure 3 shows that these force fields are typically less reliable at VLE for branched alkanes than for n-alkanes, i.e. notice the large deviations at $Z \approx 0$. More importantly, the same erroneous trend in Z is observed as in Figure 2.

In general, a clear bias is 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 suggests that an exponential repulsive barrier is also too steep, at least for the distances that are sampled at these state points.

The one exception is the TraPPE-2 model for ethane, which reproduces the entire PVT phase space simulated. This result is somewhat surprising considering the TraPPE-2 model 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,
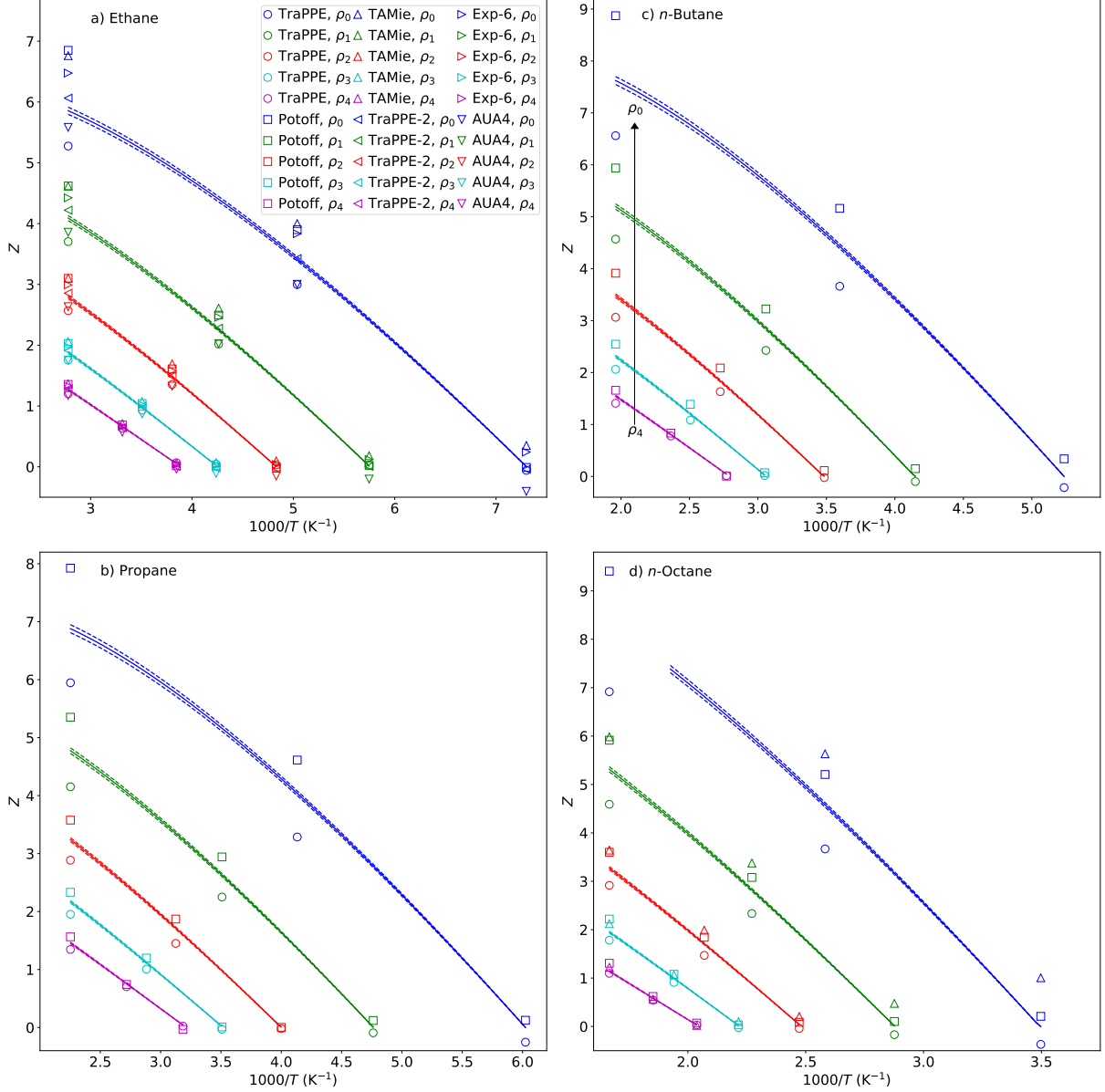
15

Figure 2: 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. TraPPE and Potoff simulation results are depicted using open circles and squares, respectively, with error bars representing two times the standard deviation of the fluctuations from a single simulation. Solid lines represent REFPROP correlations, with dashed lines representing a 1% uncertainty in REFPROP values. Simulation error bars are approximately one symbol size.
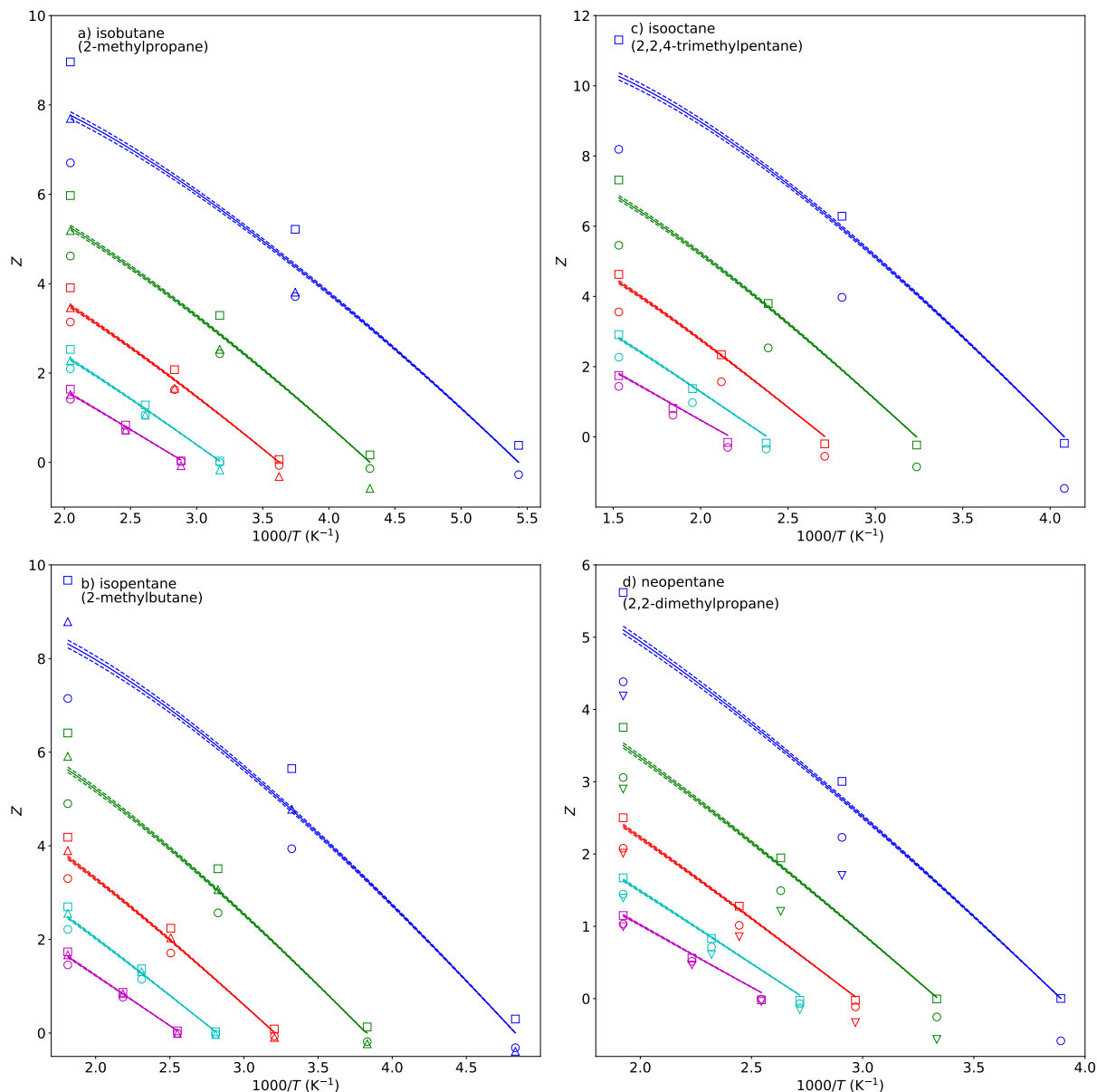
Figure 3: Compressibility factors (Z) along isochores for branched alkanes are not as accurate as normal alkanes at saturation ($Z \approx 0$) and deviate strongly at higher pressures. Panels a)-d) correspond to isobutane, isopentane, isooctane, and neopentane, respectively. Symbols, lines, uncertainties, and formatting are the same as those in Figure 2.

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 over fit 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, a comparison of TraPPE-UA and Potoff is straightforward because they use the same bond-lengths and the same non-bonded Mie λ-6 potential (Equation 5). 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 would 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 ρ_l^{sat}, P_v^{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 ρ_l^{sat} because the UA LJ 12-6 model cannot adequately predict both ρ_l^{sat} and P_v^{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, because there is inherent uncertainty in the non-bonded parameters associated with the VLE data, it is possible that an equally proba-

ble parameter set can adequately predict PVT behavior. Bayesian inference can quantify this parameter uncertainty by determining a distribution of feasible non-bonded parameter sets. If none of the feasible ϵ, σ, and λ values is capable of predicting VLE properties and PVT behavior, 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. We perform this analysis only for n-alkanes, because the CH_3 and CH_2 results suggest that a UA Mie λ-6 potential is not adequate and, therefore, we did not find it necessary to repeat this process for the CH and C interaction sites.

In order to rigorously quantify if the UA Mie λ-6 potential is "adequate", we perform a Bayesian inference analysis. We refer the reader to the literature for a thorough discussion of Bayesian statistics.[36–38] In Section 4.1, we review some basic concepts of Bayes theorem, we define the posterior, likelihood, and prior functions, and we discuss the Markov Chain Monte Carlo (MCMC) sampling approach. As MCMC can be computationally burdensome, especially when coupled with molecular simulations, we use Multistate Bennett Acceptance Ratio (MBAR) as a surrogate model to reduce the computational cost for determining the VLE properties (ρ_l^{sat} and P_v^{sat}) and Z of the supercritical fluids and compressed liquids (see Section 4.2).

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

$$Pr(\theta|D, M) = \frac{Pr(D|\theta, M)Pr(\theta|M)}{Pr(D|M)} \tag{11}$$

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:

$$Pr(M|D) = \frac{Pr(D|M)Pr(M)}{Pr(D)} \tag{12}$$

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.

4.1.2 Application

For clarity, we rewrite Equations 11-12 for the specific case studied

$$Pr(\epsilon, \sigma|\rho_l^{\text{sat}}, P_v^{\text{sat}}, \lambda) = \frac{Pr(\rho_l^{\text{sat}}, P_v^{\text{sat}}|\epsilon, \sigma, \lambda)Pr(\epsilon, \sigma|\lambda)}{Pr(\rho_l^{\text{sat}}, P_v^{\text{sat}}|\lambda)} \tag{13}$$

$$Pr(\lambda|\rho_l^{\text{sat}}, P_v^{\text{sat}}) = \frac{Pr(\rho_l^{\text{sat}}, P_v^{\text{sat}}|\lambda)Pr(\lambda)}{Pr(\rho_l^{\text{sat}}, P_v^{\text{sat}})} \tag{14}$$

where in this context ρ_l^{sat} and P_v^{sat} should 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", $Pr(\rho_l^{\text{sat}}, P_v^{\text{sat}}|\lambda)$ in Equation 14, for different values of λ is determined by integrating the numerator of Equation 13 for all values of ϵ and σ. Note that if we instead sampled from the three-dimensional (ϵ, σ, and λ) distribution, Equation 13 would be viewed as marginalizing with respect to λ.

To compute the Bayes factor between two values of λ, we assume that the "prior" evidence is equal for all values of ϵ, σ, and λ, i.e. we use uniform distributions for $Pr(\epsilon, \sigma | \lambda)$ in Equation 13 and $Pr(\lambda)$ in Equation 14. With this assumption, K depends completely on the "likelihood":

$$K = \frac{\int L(\epsilon, \sigma, |\rho_l^{\text{sat}}, P_v^{\text{sat}}, \lambda_j) d\epsilon d\sigma}{\int L(\epsilon, \sigma, |\rho_l^{\text{sat}}, P_v^{\text{sat}}, \lambda_i) d\epsilon d\sigma} \tag{15}$$

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

4.1.3 Implementation

Markov Chain Monte Carlo (MCMC) is the traditional approach for numerically sampling from the probability distribution $Pr(\epsilon, \sigma | \rho_l^{\text{sat}}, P_v^{\text{sat}} | \lambda)$. A Markov Chain is created 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:

$$\alpha = \min\left(1, \frac{Pr(\epsilon_{i+1}, \sigma_{i+1} | \rho_l^{\text{sat}}, P_v^{\text{sat}}, \lambda)}{Pr(\epsilon_i, \sigma_i | \rho_l^{\text{sat}}, P_v^{\text{sat}}, \lambda)}\right) \tag{16}$$

where α is the acceptance probability, ϵ_i and σ_i are the previous parameter set, and ϵ_{i+1} and σ_{i+1} are the proposed parameter set. The amount to which ϵ or σ is varied ($\Delta\epsilon$ and $\Delta\sigma$) for each MCMC step is 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 $\Delta\epsilon$ and $\Delta\sigma$ do not change. Details for MCMC are provided in Supporting Information (i.e. number of steps for burn-in and production, frequency that step sizes are updated, resulting acceptance percentages, etc.).

Because MCMC moves are accepted based on Equation 16 and the denominator in Equation 13 (i.e. $Pr(\rho_l^{\text{sat}}, P_v^{\text{sat}} | \lambda)$) does not depend on ϵ and σ, the acceptance probability is independent of $Pr(\rho_l^{\text{sat}}, P_v^{\text{sat}} | \lambda)$. Also, we use a "non-informative prior" such that the acceptance probability is independent of $Pr(\epsilon, \sigma | \lambda)$ (although we do include a lower

bound that the parameters are positive, i.e. $Pr(\epsilon, \sigma|\lambda)$ is uniform for all values of ϵ, σ, and λ greater than 0). Therefore, the probability of accepting ϵ_{i+1} and σ_{i+1} is based completely on the likelihood:

$$\alpha = \min\left(1, \frac{L(\epsilon_{i+1}, \sigma_{i+1}|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)}{L(\epsilon_i, \sigma_i|\rho_{\mathrm{l}}^{\mathrm{sat}}, P_{\mathrm{v}}^{\mathrm{sat}}, \lambda)}\right) \tag{17}$$

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 multi-variate normal distribution, $N(\mu, \Sigma)$. The covariance matrix (Σ) accounts for the uncertainties of both the experimental data and the computational analysis (i.e. the methods discussed in Section 4.2 to estimate $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ for a given ϵ, σ, and λ). The uncertainties are assumed to be independent such that the combined variance is the sum of the experimental and computational variances.[37] The experimental data and corresponding uncertainties were obtained through the Thermodynamics Research Center (TRC) and are found in Supporting Information. The computational variances are discussed in Section 4.2.

The parameter sets sampled from MCMC (ϵ_{MCMC} and σ_{MCMC}) provide an estimate of the uncertainty in ϵ and σ (see Figures 4 and 5 in Section 5). This parameter uncertainty propagates when estimating another quantity of interest (QoI), which may or may not be included in D, according to:

$$Pr(QoI|D, M) = \int Pr(QoI|\theta, M)Pr(\theta|D, M)d\theta \approx H\left(QoI(\theta_{\mathrm{MCMC}}|D, M)\right) \equiv H(QoI_{\mathrm{MCMC}}) \tag{18}$$

where H is a normalized histogram, $QoI(\theta_{\mathrm{MCMC}}|D, M)$ are the evaluated values of QoI using the MCMC sampled parameter sets (θ_{MCMC}), and we adopt the more succinct notation QoI_{MCMC} to denote $QoI(\theta_{\mathrm{MCMC}}|D, M)$. From the histogram of QoI_{MCMC} values, standard statistical methods can be implemented to approximate the QoI uncertainty at a desired credible level.

The QoI considered in this study are VLE data, which are included in D, and Z at

high temperatures and pressures. For the specific case of propagating the parameter uncertainties (ϵ and σ for a given λ) from ρ_l^{sat} and P_v^{sat} to Z, Equation 18 becomes:

$$Pr(Z|\rho_l^{\text{sat}}, P_v^{\text{sat}}, \lambda) \approx H\left(Z(\epsilon_{\text{MCMC}}, \sigma_{\text{MCMC}}|\rho_l^{\text{sat}}, P_v^{\text{sat}}, \lambda)\right) \equiv H(Z_{\text{MCMC}}) \qquad (19)$$

Since a large number of MCMC samples are required for adequate representations of $Pr(\theta|D, M)$ and $Pr(QoI|D, M)$, MCMC is computationally infeasible when direct molecular simulation is required for every MCMC step. For this reason, we implement surrogate models to estimate $\rho_{l,\text{MCMC}}^{\text{sat}}$, $P_{v,\text{MCMC}}^{\text{sat}}$ and Z_{MCMC}.

4.2   Surrogate Model

A typical Markov Chain requires $O(10^4\text{-}10^5)$ Monte Carlo steps, where the likelihood function must be evaluated at each step. Since $L(\theta|D)$ depends on the force field parameters (ϵ, σ, and λ), an MCMC approach is computationally infeasible if computing $L(\theta|D)$ requires performing direct molecular simulations for every proposed set of ϵ and σ. For this reason, surrogate models are an essential tool for Bayesian methods such as MCMC. We use a configuration-sampling-based surrogate model, where configurations are sampled using a small group of reference parameter sets ($\epsilon_{\text{ref}}, \sigma_{\text{ref}}$, and λ_{ref}).[33] 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).[39]

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 saturated liquid densities (ρ_l^{sat}) and vapor pressures (P_v^{sat}). This is important since ρ_l^{sat}

and $P_{\mathrm{v}}^{\mathrm{sat}}$ are the data (D) included in $L(\theta|D)$. Details for the implementation of MBAR and ITIC (MBAR-ITIC) is discussed elsewhere.[33]

The ITIC analysis provides VLE properties at only 5 saturation temperature values $(T_{\mathrm{ITIC}}^{\mathrm{sat}})$, while the experimental data set (D) may have hundreds of saturation temperatures $(T_{\mathrm{exp}}^{\mathrm{sat}})$. Although we could fit the force field to empirical correlations of experimental data (i.e. REFPROP, ThermoData Engine (TDE)[40]), raw experimental data are preferred when performing a Bayesian analysis. For this reason, we instead use empirical model fits to interpolate the ITIC VLE properties $(\rho_{\mathrm{l,ITIC}}^{\mathrm{sat}})$ and $(P_{\mathrm{v,ITIC}}^{\mathrm{sat}})$ at any saturation temperature. Specifically, we fit $P_{\mathrm{v,ITIC}}^{\mathrm{sat}}$ and $T_{\mathrm{ITIC}}^{\mathrm{sat}}$ to the Antoine equation:

$$\log_{10}(P_{\mathrm{v}}^{\mathrm{sat}}) = a_0 + \frac{a_1}{T^{\mathrm{sat}} + a_2} \tag{20}$$

where a_i are fitting parameters. We fit $\rho_{\mathrm{l,ITIC}}^{\mathrm{sat}}$ and $T_{\mathrm{ITIC}}^{\mathrm{sat}}$ to a combined rectilinear and density scaling law expression:[23]

$$\rho_{\mathrm{l}}^{\mathrm{sat}} = b_0 + b_1(b_2 - T^{\mathrm{sat}}) + b_3(b_2 - T^{\mathrm{sat}})^{\beta} \tag{21}$$

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 $\rho_{\mathrm{v,ITIC}}^{\mathrm{sat}}$ to a similar expression, but this is unnecessary for our purposes since we are only including $\rho_{\mathrm{l}}^{\mathrm{sat}}$ in D.

In summary, MBAR, ITIC, and Equations 20-21 enable prediction of $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ for any ϵ, σ, and λ by performing direct 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 5), we implement basis functions to efficiently recompute the energies and forces that are required for MBAR and ITIC (for details see Appendix of Messerly et al.[33]). 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 uncertainty, i.e. the standard deviation or variance, due to MBAR, ITIC, and Equations 20-21 is essential for evaluating $L(\theta|D)$. Rather than performing a rigorous statistical assessment, we use an empirical approach for estimating the computational standard deviation (u_c). Specifically, we compute the standard deviation between our estimated ρ_l^{sat} and P_v^{sat} values for TraPPE and Potoff with those reported in the literature obtained using Gibbs Ensemble Monte Carlo (GEMC) or Grand Canonical Monte Carlo (GCMC) histogram reweighting (HR). Although this is a rough approach for estimating uncertainties, this comparison has the added benefit that it incorporates possible deviations associated with the simulation package, and post-simulation analysis, which can be significant.[41]

The error model for ρ_l^{sat} estimates u_c to be 0.15% up to $0.75T_c$ and increasing linearly to 0.75% at the maximum T^{sat}. The error model for P_v^{sat} estimates u_c to be 7% at the minimum Tsat and decreasing linearly to 3% at $0.6T_c$ where it remains constant for higher temperatures. Note that these are conservative estimates of u_c, where other studies suggest ITIC can have significantly smaller uncertainties.[32] In fact, for the compounds investigated in this study, these uncertainties are much larger than the experimental uncertainties (u_{exp}) and, therefore, the size of the parameter space sampled by MCMC depends almost entirely on u_c. The use of a conservative u_c model is intentional in this regard, namely, so that the θ_{MCMC} sampled points represent the only feasible values of ϵ and σ for optimizing ρ_l^{sat} and P_v^{sat}.

5 Results

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. As mentioned in Section 2.2, it is common to limit λ to integer values. Due to the strong correlation between ϵ and λ, advanced sampling methods would be required to achieve good acceptance ratios when sampling from discrete values of λ. For this reason, we perform several independent MCMC runs using fixed values of λ. This approach is computationally more efficient since we are only concerned with a few values of λ (i.e. 12-18).

5.1 Ethane

Figure 4 presents the MCMC results for ethane with $\lambda_{CH_3} = $ 14-18. Panel a) plots the MCMC sampled parameter sets for different values of λ_{CH_3} ($\epsilon_{CH_3,MCMC}$ and $\sigma_{CH_3,MCMC}$). The inset of Panel a) plots the mean absolute percent deviation (MAPD%) of the MCMC predicted VLE properties ($\rho_{l,MCMC}^{sat}$ and $P_{v,MCMC}^{sat}$). Panels b) and d) plot the percent deviation from REFPROP correlations for $\rho_{l,MCMC}^{sat}$ and $P_{v,MCMC}^{sat}$, respectively. Specifically, the upper and lower lines correspond to the uncertainty for each λ, which are determined at the 95% credible level. The insets of Panels b) and d) are histograms of the MAPD% in $\rho_{l,MCMC}^{sat}$ and $P_{v,MCMC}^{sat}$, respectively. Panel c) plots Z with respect to inverse temperature for the two highest isochores (ρ_0 and ρ_1 in Panel a) of Figure 2). The inset of Panel c) plots the distribution of average deviation (AD%) in pressure for the two highest densities along the supercritical isotherm (P^{high}). The AD% is chosen to demonstrate the positive bias in P^{high}, while the MAPD% is used when quantifying the goodness of fit in VLE data.
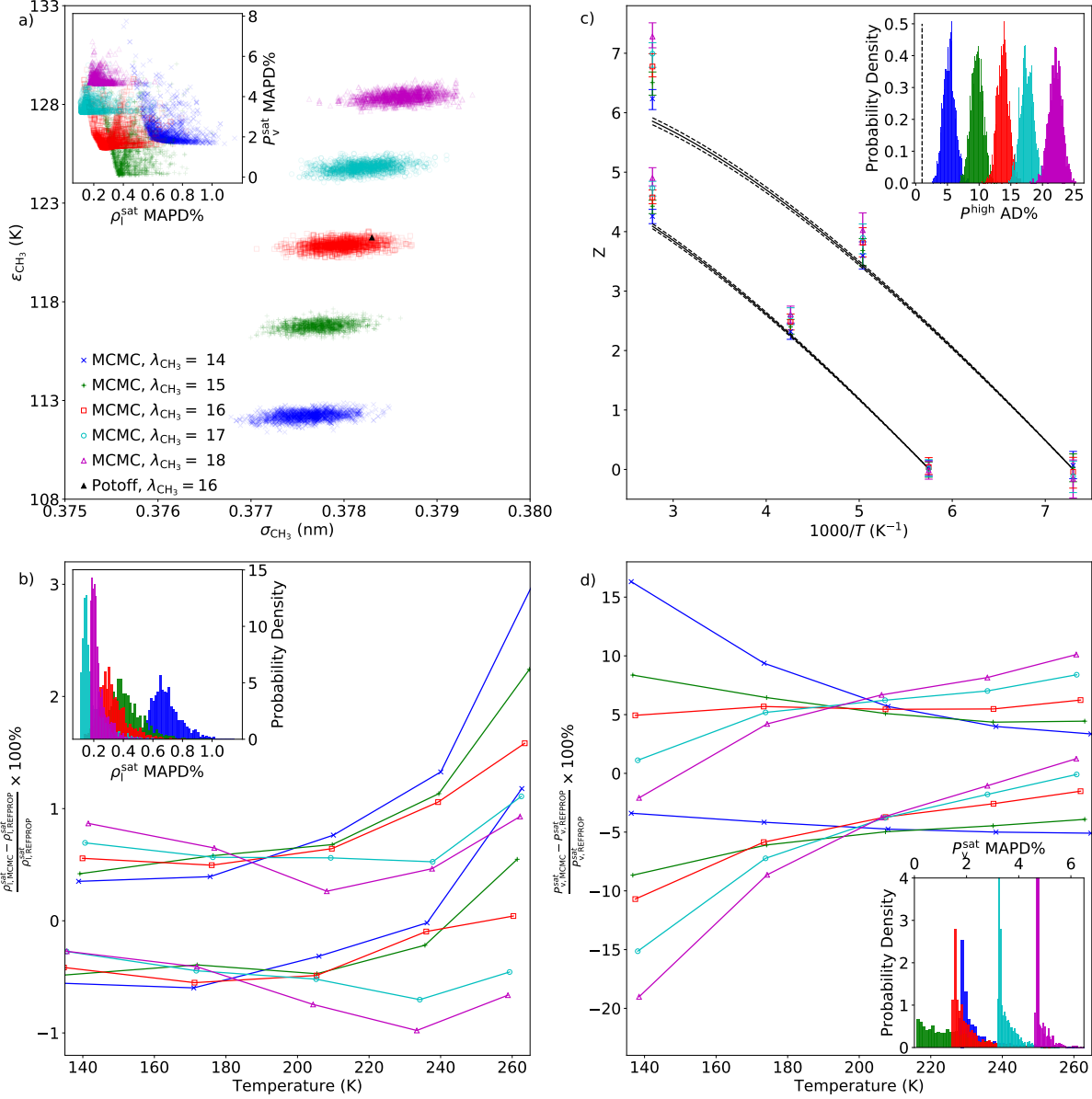
Figure 4: MCMC results confirm that the UA Mie λ-6 potential cannot adequately predict both VLE and high pressures for supercritical fluids and compressed liquids. Potoff parameter set is provided in Panel a) as a reference for $\lambda_{CH_3} = 16$. REFPROP uncertainty in P^{high} is $\pm 1\%$. Experimental data used to compute the likelihood are included in Panels b) and d) as black dots.

Panel a) in Figure 4 demonstrates that ϵ_{CH_3} depends more strongly than σ_{CH_3} on λ_{CH_3}.

Panel b) with the corresponding inset demonstrates that the best prediction of ρ_l^{sat} is ob-

tained for higher values of $\lambda_{\mathrm{CH_3}}$. Panel d) demonstrates that the 14-6 and 18-6 potentials over- and under-predict $P_{\mathrm{v}}^{\mathrm{sat}}$ at low temperatures, respectively, while the 15-6 potential has the least amount of bias and the lowest AD% in $P_{\mathrm{v}}^{\mathrm{sat}}$ (see inset). The inset of Panel a), helps to visualize the overall performance of different values of $\lambda_{\mathrm{CH_3}}$. Notice the Pareto front (i.e. the trade-off) between the two properties included in the objective function, namely, $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$.[22,42,43] Consistent with the insets of Panels b) and d), the 15-6 potential has the lowest MAPD% in $P_{\mathrm{v}}^{\mathrm{sat}}$ while the 16-6, 17-6, and 18-6 have slightly better MAPD% in $\rho_{\mathrm{l}}^{\mathrm{sat}}$.

Finally, and most importantly, Panel c) demonstrates that all of the sampled $\epsilon_{\mathrm{CH_3,MCMC}}$ and $\sigma_{\mathrm{CH_3,MCMC}}$ parameter sets for different values of $\lambda_{\mathrm{CH_3}}$ over-predict Z at high temperatures and densities (P^{high}). As expected, the larger the value of $\lambda_{\mathrm{CH_3}}$, the greater the force field over-predicts P^{high}. For example, the 15-6 and 16-6 potentials, which are the two best potentials based on VLE, over-predict P^{high} by around 10 and 15%, respectively. This supports the fundamental claim of this manuscript, namely, that the UA Mie λ-6 potential cannot adequately predict both VLE and high pressures for supercritical fluids and compressed liquids.

While the 14-6 potential only demonstrates a 5% disagreement in P^{high} (notice that the REFPROP deviation is only 1%), the deprecation in the quality of VLE is significant. This can be quantified by determining the Bayes factor for different values of $\lambda_{\mathrm{CH_3}}$ (normalized by the 14-6 potential). Table 6 shows that for ethane the 15-6 and 16-6 potentials are equally justified while the evidence for 14-6 and 17-6 is much less and the evidence for 18-6 is negligible.

Although a Bayes factor around 4 is typically not considered substantial evidence, these values depend strongly on the VLE data and the error model used to compute $L(\theta|D)$. We have chosen a very conservative error model to demonstrate the inadequacy in predicting P^{high} (see Section 4.2). However, a less conservative error model would pro-

Table 6: Bayes factors for different values of λ (normalized by $\lambda = 14$). CH_3 values depend on ethane while CH_2 values are based on propane and n-butane.

λ	CH_3	CH_2
14	1.0	1.0
15	3.6	-
16	3.5	16.3
17	0.8	-
18	0.1	-

vide more convincing evidence for the λ Bayes factors. Also, recall that ITIC is limited to $T^{\text{sat}} < 0.85 T_c$. Therefore, it is possible that the optimal value of λ_{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 ρ_l^{sat} at higher temperatures (240-260 K) for the 14-6 potential, it appears that higher temperature data would strengthen the evidence that the 14-6 potential is not suitable for VLE. It is unclear whether higher temperature data would support the 15-6 or 16-6 potential, likely the optimal λ_{CH_3} value is some fraction between 15 and 16. In order to include VLE data from $0.85 T_c < T^{\text{sat}} < 0.95 T_c$, we are working on implementing MBAR with GCMC.

5.2 Larger n-alkanes

Figure 5 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. Panel a) contains the MCMC parameter sets for propane, n-butane, and n-octane, while Panel b) contains results for propane and n-butane. Figure 5 also includes contours of the average percent deviations (AD%) in P^{high} ($T = T^{IT}$ and $\rho = rho_0$ or ρ_1) relative to the REFPROP correlations. The "REFPROP uncertainty" region corresponds to AD% of ± 1.

Notice in Figure 5 that the MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets, for a given
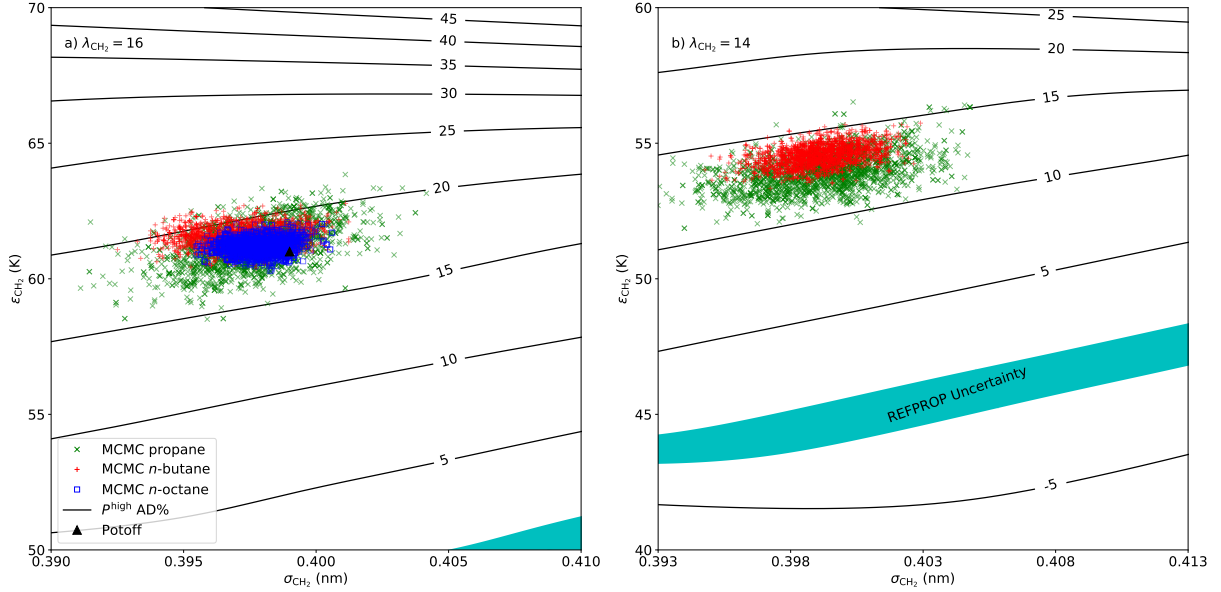
Figure 5: MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets result in large AD% for P^{high}. Panels a) and b) correspond to $\lambda_{CH_2} = 16$ and $\lambda_{CH_2} = 14$, respectively. REFPROP uncertainty in P^{high} is $\pm1\%$. Potoff parameter set is provided as a reference for $\lambda_{CH_2} = 16$.

value of λ_{CH_2}, overlap considerably for the different compounds. This provides statistical evidence to support the common assumption of transferability of CH_2 parameters between different n-alkanes. Also, note that the uncertainty in the parameters is largest for propane and smallest for n-octane. This suggests that, as expected, 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. Notice, in Panel a), that the Potoff parameter set for $\lambda_{CH_2} = 16$ is within the MCMC sample region, suggesting that it is one of the statistically feasible parameter sets.

More importantly, for the purposes of this manuscript, 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 $\approx$ 16-21 and $\approx$ 10-15, much greater than the REFPROP uncertainty of around 1%. Because the "REFPROP uncertainty" contours are roughly parallel to the MCMC region and found at much lower ϵ_{CH_2} (around

45 K for $\sigma_{\mathrm{CH_2}} = 0.399$ nm), in order to accurately predict P^{high}, it is necessary to sacrifice accuracy in ρ_l^{sat} and P_v^{sat}. This suggests that neither the UA Mie 16-6 or 14-6 models are capable of predicting VLE and PVT for supercritical fluids and compressed liquids of n-alkanes.

Finally, although the UA Mie 14-6 AD% in P^{high} is slightly better than the UA Mie 16-6 AD%, the UA Mie 14-6 is significantly less reliable for VLE. Table 6 demonstrates the strong evidence for a 16-6 potential over a 14-6 potential based on VLE data. (The evidence for CH_2 is likely stronger than CH_3 because VLE data from more than one compound were included.) Therefore, considering the deprecation in VLE, the marginal gain in accuracy for P^{high} likely does not merit using a UA Mie 14-6 potential.

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 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.[32,44,45]

Alternatively, 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, it is possible to use the simulation results as an upper constraint when per-

forming a non-linear optimization. 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. Therefore, despite the short-comings of the UA Mie λ-6 potential at high pressures, it can still provide a reasonable trend to help keep the FEOS from diverging.

6.2 Limitations

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]

$$A_{xy}^{\mathrm{dep}} R_g T \equiv (1/T)^x \rho^y \frac{\partial^{x+y} A^{\mathrm{dep}}}{\partial (1/T)^x \partial \rho^y} \tag{22}$$

This approach is advantageous for developing FEOS as it eliminates redundant information found in traditional macroscopic properties.[2–7] 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]

$$Z = 1 + A_{01}^{\mathrm{dep}} \tag{23}$$

$$\left(\frac{-\partial Z}{T \partial (1/T)} \right)_\rho = 1 + A_{01}^{\mathrm{dep}} - A_{11}^{\mathrm{dep}} \tag{24}$$

$$\frac{U^{\mathrm{dep}}}{R_g T} = A_{10}^{\mathrm{dep}} \tag{25}$$

$$\left(\frac{\partial U^{\mathrm{dep}}}{R_g \partial T} \right)_\rho = -A_{20}^{\mathrm{dep}} \tag{26}$$

$$\left(\frac{\rho \partial Z}{\partial \rho} \right)_T = 1 + 2A_{01}^{\mathrm{dep}} + A_{02}^{\mathrm{dep}} \tag{27}$$

Unfortunately, with the exception of *ms2*,[46] we are not aware of any open-source simulation packages that readily provide A_{xy}^{dep}.

It is our opinion that traditional macroscopic properties, such as Z and U^{dep} (with their respective derivatives), are more readily understood and visualized. In other words, 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 23-24, 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 25-27. 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}, and A_{02}^{dep}, because the results are less conclusive they are provided only as Supporting Information. Furthermore, the relationship between Equations 25-27 and the repulsive barrier, λ, are not obvious. By contrast, the observation that larger values of λ overpredict Z with increasing temperature is clear evidence that $\lambda > 12$ is not theoretically justified but, rather, is an empirical observation 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]

7 Conclusions

In principle, a fundamental equation of state can be developed for compounds without any experimental data by using only molecular simulation results. However, due to uncertainties and deficiencies in the force field, experimental data should be favored over molecular simulation values whenever possible.[7] Recently, molecular simulation results at extreme temperatures and pressures have been used to supplement experimental data when developing fundamental equations of state. For this approach to work, it is imperative that the force field be reliable and transferable over different PVT conditions. Unfortunately, parameterizing a united-atom Mie λ-6 potential with VLE data does not ensure reliable extrapolation to extreme temperatures and pressures. The Bayesian statistical analysis performed in this study suggests that the UA Mie λ-6 model type is not adequate for predicting *both* VLE properties and pressures for dense supercritical fluids and compressed liquids. Specifically, no set of ϵ, σ, and λ can adequately predict VLE and PVT behavior. 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.

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

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

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

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

(26) Kulakova, L.; Arampatzis, G.; Angelikopoulos, P.; Hadjidoukas, P.; Papadim-itriou, C.; Koumoutsakos, P. *Scientific Reports* **2017**, *7*, 16576.

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

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

(32) 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.

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

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

(35) 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.

(36) 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.

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

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

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

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

(41) 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.

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

(43) 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.

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

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

(46) 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.