Bayesian inference analysis of transferable, united-atom, Mie λ-6 force fields for normal and branched alkanes. To be submitted to the Journal of Physical Chemistry, B.

Richard A. Messerly*

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

E-mail: richard.messerly@nist.gov

Abstract

Purpose

The aim of this study is to demonstrate, using Bayesian inference, that a UA Mie force field cannot adequately predict VLE and PVT of compressed liquids and supercritical

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

fluids for normal and branched alkanes. For adequate prediction of VLE and compressed liquid pressures, we recommend using AUA or AA models.

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. Using an FEOS to extrapolate to temperatures and pressures that are significantly higher than those used in parameterizing the FEOS can result in large errors. 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 could aid in developing FEOS.[2–7] 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.[8–10] Specifically, experimental data were available for temperatures and pressures up to 580 K and 130 MPa, 590 K and 180 MPa, and 560 K and 100 MPa

for hexamethyldisiloxane, octamethylcyclotetrasiloxane, and 1,2-dichloroethane, 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 1200 MPa, respectively. The inclusion of these simulation results improved the performance of the FEOS at extreme temperatures and pressures.

Hydrocarbons are a fundamental feed-stock for many petrochemical processes and, therefore, large amounts of experimental 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). To develop FEOS for a wide assortment of hydrocarbons, molecular simulation results at extreme temperatures and pressures could be useful.

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. In this study, we investigate how well the traditional force fields for predicting VLE 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, which are assumed to be reliable at these conditions.

The most accurate force fields for estimating hydrocarbon VLE properties (i.e. $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$) are Transferable Potentials for Phase Equilibria (TraPPE)[11,12] (and, especially, the recent TraPPE-2[13]), Errington,[14] anisotropic-united-atom (AUA4),[15,16] Potoff,[17,18] and Transferable anisotropic Mie potential (TAMie).[19,20] 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 neighobring 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 parameters are optimized simultaneously. 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 $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ for *n*-alkanes.[21,22] For this reason, the TraPPE-UA force field was primarily developed to agree with saturated liquid densities.[11] By contrast, accurate prediction of both $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ over a wide temperature range is possible by varying the repulsive exponent of the LJ potential (i.e. the Mie λ-6 potential). Typically, the optimal value of λ is greater than 12 with a corresponding increase in the well depth (ϵ). Specifically for hydrocarbons, the Potoff UA force field uses $\lambda = 16$ while the TAMie force field uses $\lambda = 14$. However, there is some concern that increasing the repulsive exponent might have some undesirable consequences, especially at high pressures, where close range interactions will become more prevalent than at vapor-liquid equilibria. The purpose of this study is to determine whether or not the UA Mie potential is adequate for predicting both VLE

and PVT at higher temperatures and pressures.

Note that the simulation values used by Thol et al. were derivatives of the residual Helmholtz free energy ($\partial^n a^r$) with respect to inverse temperature and/or density,[8–10] while in this study we simply compare the PVT behavior. Aside from the advantage of simplicity (most simulation packages do not provide $\partial^n a^r$), this choice is based on the fact that PVT is more readily understood and easier to visualize. In other words, it is easier to quantify the impact on process design caused by deviations in PVT behavior than derivatives in the residual Helmholtz free energy. Furthermore, as demonstrated by Thol et al., an inaccurate prediction of some $\partial^n a^r$ does not necessarily result in poor prediction of PVT behavior.[8] Although we did not use $\partial^n a^r$ for our analysis, including higher order derivatives of the residual Helmholtz free energy from molecular simulation is a significant advantage of the approach implemented by Thol et al. for FEOS development.

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 potential. Section 5 presents the results from the Bayesian analysis. Section 6 reports the primary conclusions of this study.

2 Methods I

2.1 Simulation Details

As normal and branched alkanes are the focus of this study, 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.[23] 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. 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, $T_{\mathrm{r}}, \approx 1.2$) and five saturated liquid density isochores. Nine densities are simulated along the supercritical isotherm (with five being those of the isochore densities), and two sub-critical temperatures are simulated along each isochore. 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 Supporting Information.

We use isothermal isochoric integration (ITIC) to convert the departure internal energies (U^{dep}) and compressibility factors (Z) obtained at the 19 state points to saturated VLE properties, namely, $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$.[24,25] The equations for ITIC are:

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

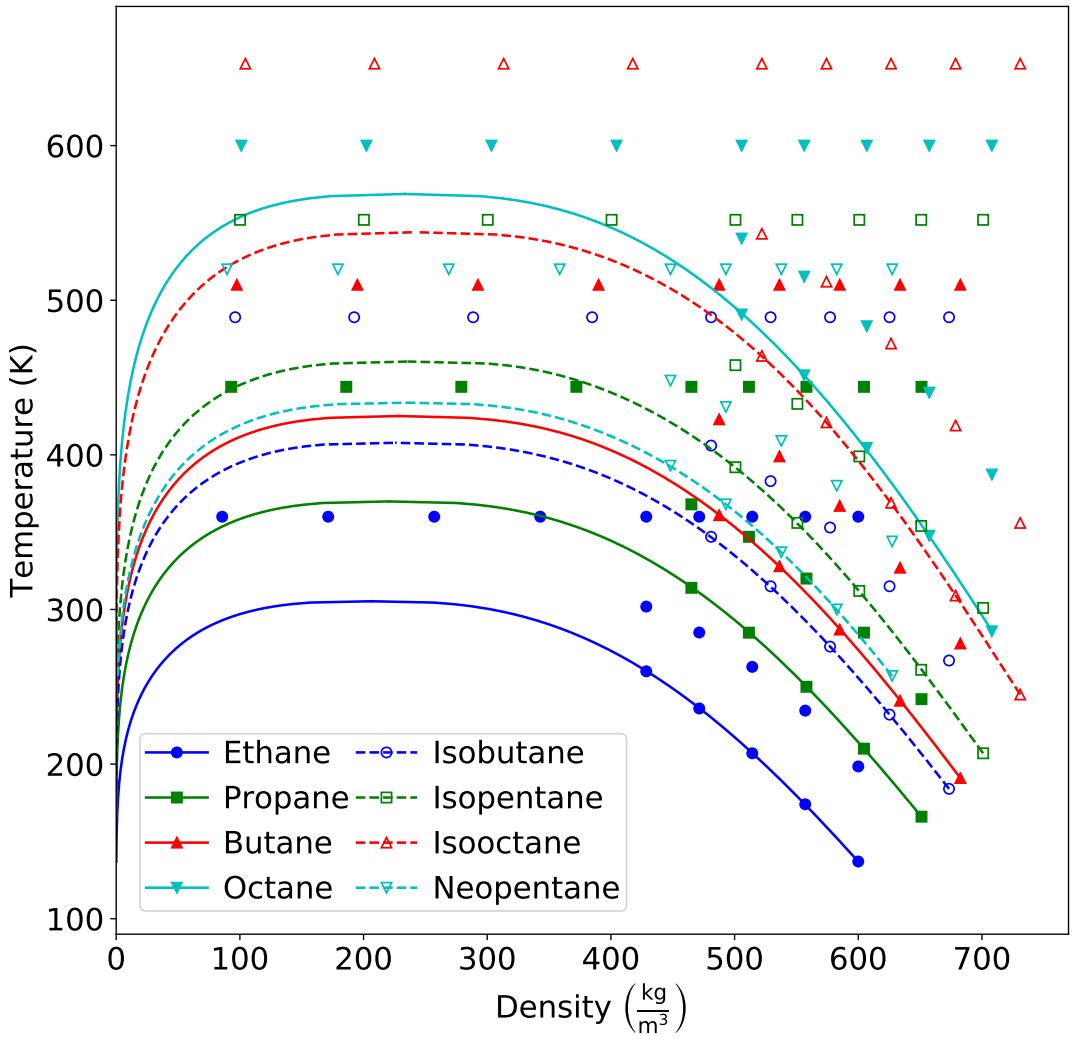

Figure 1: State points simulated for each compound studied. A total of 19 simulations are performed: nine densities along the supercritical isotherm and two temperatures along liquid density isochores. Filled symbols and solid lines correspond to *n*-alkanes, while empty symbols and dashed lines correspond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

$$\rho_{\text{v}}^{\text{sat}} \approx \rho_{\text{l}}^{\text{sat}} \exp\left(\frac{A^{\text{dep}}}{R_{\text{g}} T^{\text{sat}}} + Z_{\text{l}}^{\text{sat}} - 1 - 2B_2\rho_{\text{v}}^{\text{sat}} - 1.5B_3\rho_{\text{v}}^{\text{sat}^2}\right) \tag{2}$$

$$P_{\text{v}}^{\text{sat}} \approx (1 + B_2\rho_{\text{v}}^{\text{sat}} + B_3\rho_{\text{v}}^{\text{sat}^2})\rho_{\text{v}}^{\text{sat}} R_{\text{g}} T^{\text{sat}} \tag{3}$$

$$Z_{\mathrm{l}}^{\mathrm{sat}} = \frac{P_{\mathrm{v}}^{\mathrm{sat}}}{\rho_{\mathrm{l}}^{\mathrm{sat}} R_{\mathrm{g}} T^{\mathrm{sat}}} \tag{4}$$

where $A^{\mathrm{dep}} \equiv A - A^{\mathrm{ig}}$ 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 $(\rho_{\mathrm{l}}^{\mathrm{sat}})$, $U^{\mathrm{dep}} \equiv U - U^{\mathrm{ig}}$ is the internal energy departure, $Z_{\mathrm{l}}^{\mathrm{sat}}$ 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. As discussed and validated in a previous work,[25] the B_2 and B_3 values found in Equations 2-3 are calculated using REFPROP correlations.[1]

2.2 Force field

A united-atom (UA) representation is used for each compound studied. The UA groups are defined following the Transferable Potentials for Phase Equilibria (TraPPE-UA) force field convention, namely, a single parameter set is used for CH_3, CH_2, CH, and C sites. The bonded interactions use the same functional form and parameters as the TraPPE-UA (and Potoff) potential. Specifically, a fixed bond-length of 0.154 nm is used for each bond between united-atom sites. Angular bending interactions are evaluated using a harmonic potential:

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

where θ is the instantaneous bond angle, θ_0 is the equilibrium bond angle, and k_θ is the harmonic force constant. For the molecules studied in this work, $\theta_0 = 114.0°$ and $k_\theta = 62500$ K/rad^2 for each angle. Dihedral torsional interactions are determined using a cosine series:

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

where ϕ is the dihedral angle and the Fourier constants are $c_1/k_{\mathrm{B}} = 355.03$ K, $c_2/k_{\mathrm{B}} = -68.19$ K, and $c_3/k_{\mathrm{B}} = 791.32$ K.

Non-bonded interactions between two different molecules and united-atom sites separated by more than three bonds are calculated using a Mie λ-6 potential:

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

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 $\left(\text{i.e. } u^{\mathrm{vdw}} = -\epsilon \text{ and } \frac{\partial u^{\mathrm{vdw}}}{\partial r} = 0 \text{ for } r = r_{\mathrm{min}}\right)$, and λ is the repulsive exponent. Note that the Mie λ-6 potential reduces to the LJ 12-6 potential for $\lambda = 12$. Therefore, Equation 5 can be considered a generalized Lennard-Jones where the repulsive exponent is a parameter. Although an attractive exponent of 6 has a strong theoretical basis, $\lambda = 12$ is a historical artifact that was chosen primarily for computational purposes.[26] For the same reason (i.e. computational efficiency), a common practice is to use integer values of λ in Equation 5. The non-bonded force field parameters for TraPPE-UA and Potoff are provided in Table 1.

Table 1: Non-bonded (intermolecular) parameters for TraPPE-UA and Potoff force fields. The "short/long" Potoff CH and C parameters are included in parenthesis.

	TraPPE-UA			Potoff		
United-atom	ϵ (K)	σ (nm)	λ	ϵ (K)	σ (nm)	λ
CH_3	98	0.375	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

Non-bonded interactions between two different site types (i.e. cross-interactions) are determined using Lorentz-Berthelot combining rules[26] for ϵ and σ with an arithmetic

mean for the repulsive exponent (λ) (as recommended by Potoff and Bernard-Brunel[17]):

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{6}$$

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{7}$$

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

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. 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 use the "short/long" model, i.e. the CH and C groups depend on the length of the carbon backbone, since the "short/long" model is more accurate than the "generalized" model for these compounds. The "generalized" Potoff results are not included in Figure 3, as they do not aid in the discussion, 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 PρT behavior at high pressures (i.e. high densities and temperatures). Figure 3 shows that for branched alkanes these force fields are

typically less reliable at VLE, but, more importantly, the same erroneous trend in PρT is observed.

Since the TraPPE (LJ 12-6) potential under predicts Z and the Potoff (Mie 16-6) potential over predicts Z, it seems reasonable that perhaps a Mie 13-6, 14-6, or 15-6 potential would demonstrate the proper trend. However, as demonstrated in Section 5, there does not exist 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 for the UA model. To understand this point, it is important to remember that the UA LJ 12-6 (TraPPE-UA) force field cannot adequately predict both $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$. In other words, determining the optimal value of λ for predicting PVT of supercritical fluids and compressed liquids does not guarantee accurate prediction of $P_{\mathrm{v}}^{\mathrm{sat}}$.

Although the TAMie potential uses a softer repulsive exponent than Potoff (14 instead of 16), it consistently over predicts the pressure for supercritical fluids and compressed liquids. The same is true for the Errington Exp-6 model. By contrast, the TraPPE-2 potential provides extremely reliable estimates of $\rho_{\mathrm{l}}^{\mathrm{sat}}$, $P_{\mathrm{v}}^{\mathrm{sat}}$, and the PVT behavior at high pressures and temperatures. Although the TraPPE-2 model is only available for ethane and ethylene, these results suggest that at extreme pressures either an AUA or AA model should be used to better account for the hydrogens. That being said, alternative functional forms, such as the extended Lennard-Jones, should also be considered.

1. Several force fields in the literature have been optimized to agree with VLE properties (TraPPE, Potoff, TraPPE-2, TAMie, Errington, AUA4)

2. TraPPE, Potoff, and AUA4 have parameters for each compound studied, while TraPPE-2 only has parameters for ethane, TAMie has parameters for all except isooctane and neopentane (containing a C group), and ErrExp-6 only has parameters for the *n*-alkanes

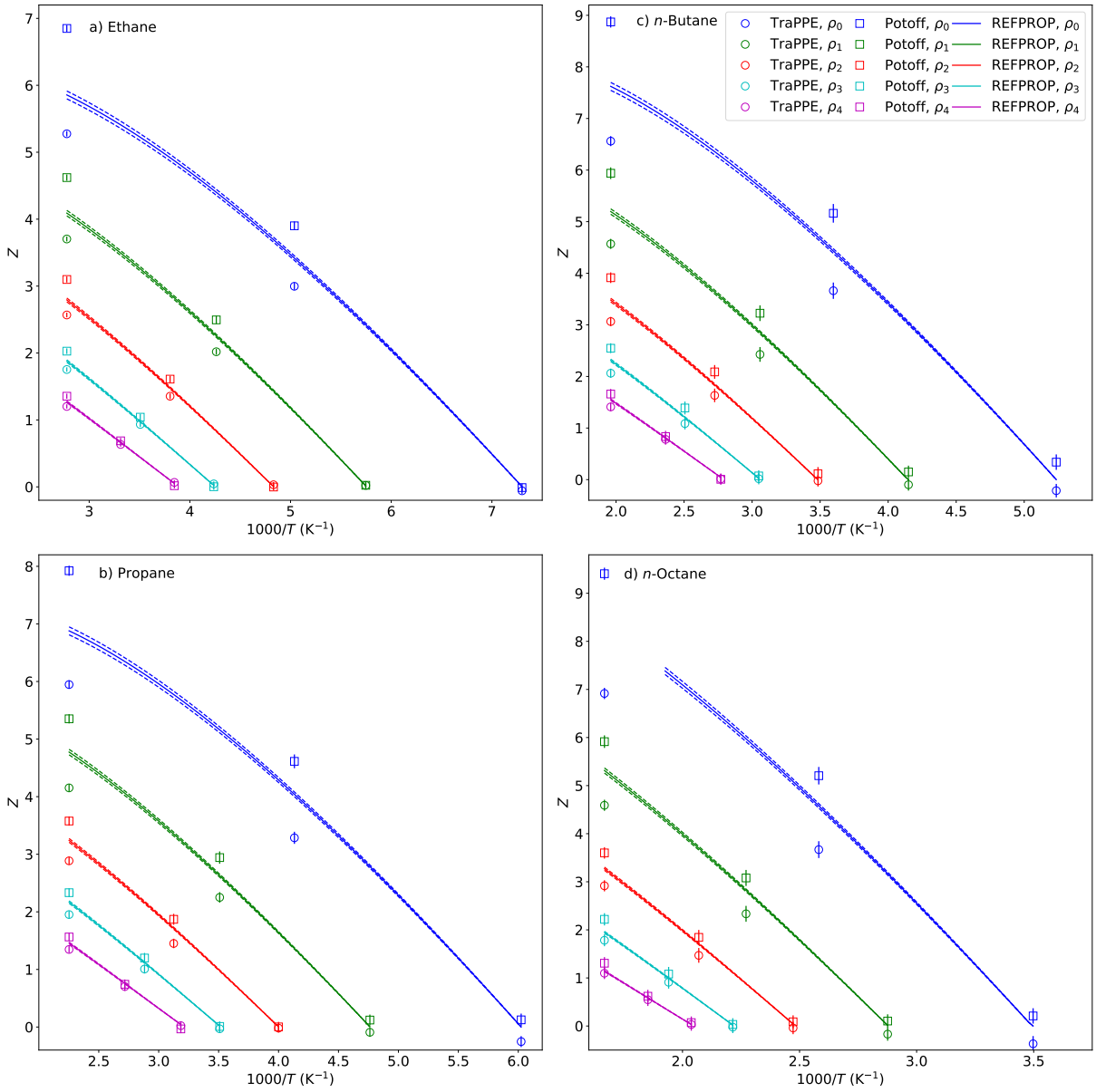

Figure 2: Compressibility factors (Z) along isochores agree at saturation $(Z \approx 0)$ but deviate strongly at higher pressures. Densities are such that $\rho_0 > \rho_1 > \rho_2 > \rho_3 > \rho_4$. Panels a)-d) correspond to ethane, propane, *n*-butane, and *n*-octane, respectively. TraPPE and Potoff 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.

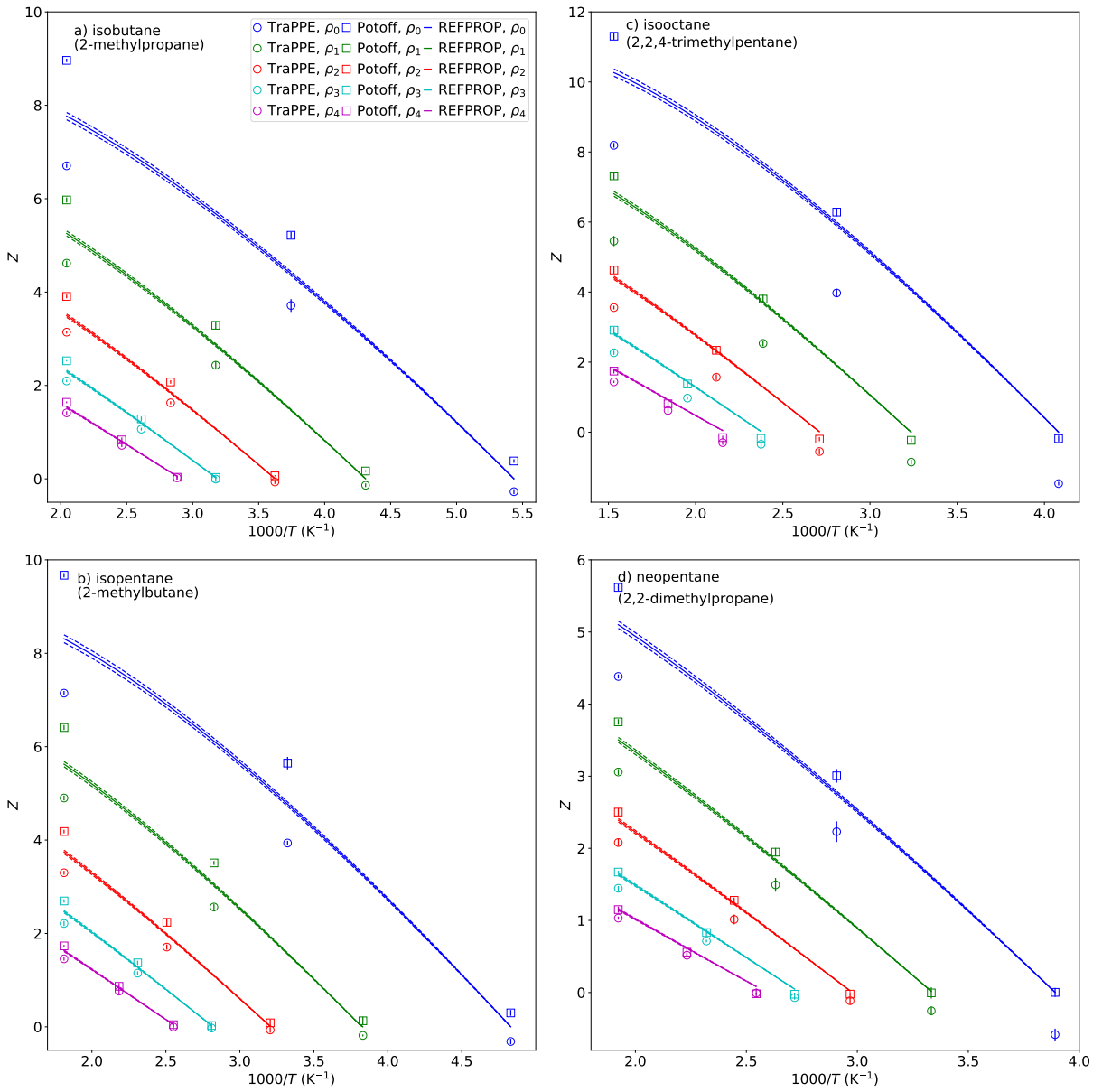

Figure 3: Compressibility factors (Z) along isochores agree at saturation $(Z \approx 0)$ but deviate strongly at higher pressures. Densities are such that $\rho_0 > \rho_1 > \rho_2 > \rho_3 > \rho_4$. Panels a)-d) correspond to isobutane, isopentane, isooctane, and neopentane, respectively. Symbols and lines are the same as those in Figure 2.

3. Figure: Z vs 1000/T for ethane, propane, n-butane, and n-octane

4. Figure: Z vs 1000/T for isobutane, isopropane, isohexane, isooctane and neopentane

5. PVT (Z) trends are inaccurate for both TraPPE and Potoff at high pressures, i.e. non-VLE conditions. Specifically, the TraPPE 12-6 under predicts while 16-6 over predicts

6. Although these results might suggest that a 14-6 potential would work best, recall that the TraPPE 12-6 does not accurately predict Pvsat.

7. Since TraPPE and Potoff use slightly different objective functions we want to perform an equivalent analysis for different values of lambda

8. Hypothesis that we want to test is that there does not exist a set of epsilon, sigma, and lambda that provides reasonable VLE, supercritical fluid, and compressed liquid (for ethane I already know this is not feasible. For n-alkanes I know that the 16-6 cannot accomplish this, but I am not sure about 14-6 or 15-6.)

9. AUA LJ 12-6 is much more accurate for ethane

10. AUA Mie 14-6 potential is not much better than UA Mie 16-6 for n-alkanes

11. AUA Exp-6 force field is not much better than UA Mie 16-6 for n-alkanes

12. No results for AUA4

4 Methods II

The results presented in Section 3 demonstrate that none of the existing force fields studied reproduce the PVT behavior for supercritical fluids and compressed liquids. However, recall that each of these force fields was parameterized using only VLE properties. Therefore, it is possible that including both VLE and PVT properties in the parameterization objective function will improve the results. However, if no combination of

ϵ, σ, and λ is capable of predicting VLE properties and PVT behavior, we can conclude that the UA Mie λ-6 potential is inadequate for this purpose and, therefore, should not be used when developing FEOS with molecular simulation results.

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. 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, Section 4.2 explains the surrogate models used to reduce the computational cost for determining the VLE properties $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ and Z of the supercritical fluids and compressed liquids.

1. We use Bayesian inference with MCMC to quantify the uncertainty in the force field parameters

2. We perform direct molecular simulation for several reference force field parameter sets

3. We use MBAR-ITIC to predict rholsat and Pvsat with non-simulated force field parameters

4. We use MBAR to propagate the parameter uncertainties from VLE to Z of compressed liquids/supercritical

5. We perform this analysis for CH3 and CH2 sequentially and independently. In other words, the Markov Chain only samples a 2-dimensional space.

4.1 Bayesian Analysis

Bayesian inference is used to quantify the uncertainty in ϵ and σ for a given λ. Bayes theorem is

$$Pr(\theta|D) = \frac{Pr(D|\theta)Pr(\theta)}{Pr(D)} \tag{9}$$

where Pr denotes a probability distribution function, θ is the parameter set (i.e. ϵ, σ, and λ for the Mie λ-6 potential), and D are the data. $Pr(\theta|D)$ is commonly referred to as the "posterior, $Pr(D|\theta)$ is the "likelihood" (alternatively expressed as $L(\theta|D)$), $Pr(\theta)$ is the "prior", and $Pr(D)$ is a normalization constant.

Markov Chain Monte Carlo (MCMC) is the traditional approach for sampling from $Pr(\theta|D)$. We use a "non-informative prior" with a lower bound that the parameters are positive, i.e. $Pr(\theta)$ is uniform for all values of ϵ, σ, and λ greater than 0. Because MCMC moves in parameter space are accepted based on the ratio of $Pr(\theta_i|D)$ and $Pr(\theta_{i+1}|D)$ and the denominator in Equation 9 (i.e $(Pr(D))$ does not depend on θ, $Pr(D)$ is not needed. Therefore, the probability of accepting θ_{i+1} is based completely

1. By quantifying the uncertainty in epsilon and sigma for a given lambda, we can determine if the Mie potential is adequate for reproducing VLE and compressed liquid/supercritical pressures

2. Posterior includes saturated liquid density and vapor pressure

3. Markov Chain Monte Carlo is used to sample

4. 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.)

5. The parameter uncertainty is propagated when predicting high pressures

4.2 Surrogate Model

1. A Bayesian analysis is computationally too expensive if direct molecular simulations are performed for every MCMC step

2. As demonstrated in a previous publication, MBAR can reweight the configurations that are sampled from different force fields without direct simulation

3. In other words, a set of reference force fields are simulated for each molecule and MBAR is used instead of direct simulation for each MCMC step

4. As demonstrated in a previous publication, MBAR is used to predict Udep and Z while ITIC is used to convert Udep and Z to rholsat and Pvsat

5. ITIC state points are fit to rectilinear and Antoine equation to interpolate rholsat and Pvsat, this allows for comparison with experimental data at hundreds of temperatures

6. The likelihood includes the experimental uncertainties but, more importantly, the numerical uncertainties. In other words, the numerical uncertainties account for the uncertainties that arise from the simulations themselves, the MBAR reweighting, the ITIC algorithm, and fitting to rectilinear and Antoine.

7. To leave no room for doubt in our conclusions, we use very conservative (and empirical) estimates of numerical uncertainty for rholsat and Pvsat (see Supporting Information)

4.3 Propagation of Uncertainty

1. From the MCMC parameter sets, we randomly sample a subset of 100-1000 parameter sets

2. We use MBAR to predict Z for each of those parameter sets

3. We plot these results as a histogram to determine the 95% credible interval for Z at each state point

5 Results

5.1 Parameter uncertainties

CH3 Results:

Figure: Ethane CH3 uncertainties. Panel a) 14-6, 15-6, 16-6. Panel b) Z plots for each exponent Figure: Evidence for 14-6, 15-6, 16-6 based on VLE

CH2 results:

Combine these into one figure Figure: CH2 uncertainties. Panel a) 16-6 Panel b) 14-6

Figure 4 presents the MCMC sampled $\epsilon_{\mathrm{CH_2}}$ and $\sigma_{\mathrm{CH_2}}$ parameter sets with $\lambda_{\mathrm{CH_2}} = 16$. Notice that the MCMC sampled $\epsilon_{\mathrm{CH_2}}$ and $\sigma_{\mathrm{CH_2}}$ parameter sets overlap considerably for propane, *n*-butane, and *n*-octane. This supports the common assumption of transferability of $\mathrm{CH_2}$ parameters between different *n*-alkanes. Note that the uncertainty in the parameters is largest for propane and smallest for *n*-octane. This suggests that, as expected, the sensitivity of $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$ with respect to the $\mathrm{CH_2}$ parameters increases with increasing number of $\mathrm{CH_2}$ interaction sites. Also, notice that the Potoff parameter set is within the MCMC sample region. More importantly, for the purposes of this manuscript, the MCMC sampled $\epsilon_{\mathrm{CH_2}}$ and $\sigma_{\mathrm{CH_2}}$ parameter sets have P^{high} AAD% $\approx$ 16-23, while the REFPROP uncertainty in P^{high} is estimated to be around 1%.

Figure 4 suggests that the UA Mie 16-6 potential is not capable of predicting VLE and PVT for supercritical fluids and compressed liquids of propane, *n*-butane, and *n*-octane. Figure 5 demonstrates that the UA Mie 14-6 potential is also not capable of predicting

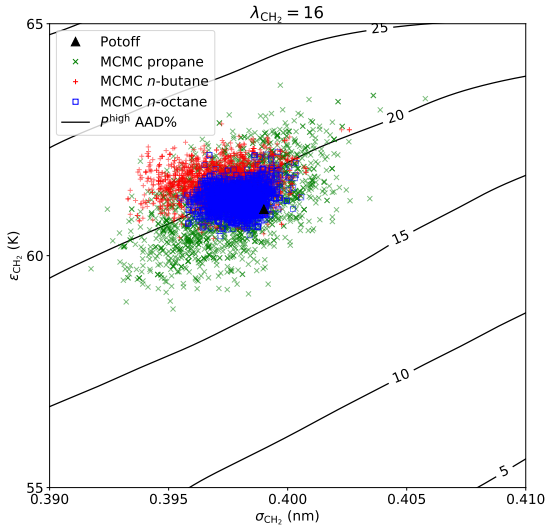

Figure 4: MCMC sampled $\epsilon_{\mathrm{CH_2}}$ and $\sigma_{\mathrm{CH_2}}$ parameter sets with $\lambda_{\mathrm{CH_2}} = 16$ result in AAD% $\approx$ 16-23 for P^{high}. REFPROP uncertainty in P^{high} is $\pm 1\%$. Potoff parameter set is provided as a reference.

both properties for propane and *n*-butane. Specifically, all of the MCMC sampled $\epsilon_{\mathrm{CH_2}}$ and $\sigma_{\mathrm{CH_2}}$ parameter sets in Figure 5 have P^{high} AAD% $\approx$ 10-15. Although this is improvement relative to the UA Mie 16-6 AAD%, recall that the UA Mie 14-6 is less reliable for VLE. Therefore, considering the significant deprecation in VLE, the marginal gain in accuracy for P^{high} likely does not merit using a UA Mie 14-6 potential.

Figure 5 also includes the "REFPROP uncertainty" region which corresponds to AAD% of ± 1. Because the "REFPROP uncertainty" contours are parallel to the MCMC region and found at much lower $\epsilon_{\mathrm{CH_2}}$ (around 45 K for the same $\sigma_{\mathrm{CH_2}}$), in order to accurately predict P^{high}, it is necessary to sacrifice accuracy in ρ_{l}^{sat} and $P_{\mathrm{v}}^{\mathrm{sat}}$. Figures 4-5 provide convincing evidence that, regardless of the value of λ, the UA Mie λ-6 model is not capable of predicting both VLE and PVT at high pressures.

1. Figure: The uncertainty regions for CH3, CH2, CH, and C. I can include 14-6, 15-6, and 16-6. Perhaps I will only do this rigorous analysis for CH3 or for CH3 and

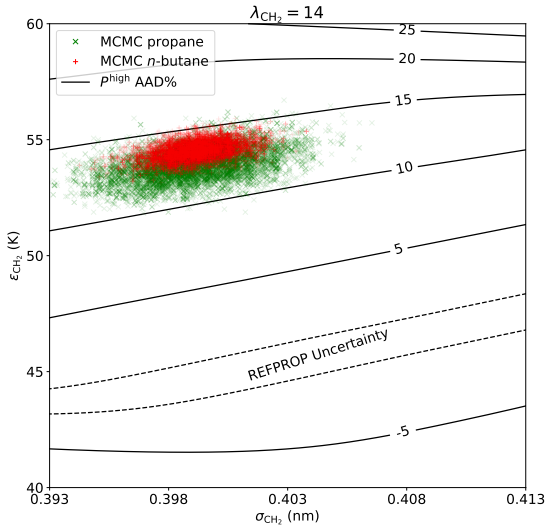

Figure 5: MCMC sampled $\epsilon_{\mathrm{CH_2}}$ and $\sigma_{\mathrm{CH_2}}$ parameter sets with $\lambda_{\mathrm{CH_2}} = 14$ result in AAD% $\approx$ 10-15 for P^{high}. REFPROP uncertainty in P^{high} is included at $\pm 1\%$.

CH2. Probably not for all. I could include the results from the alternative posterior (excluding Pvsat and including high pressures) but then it might be out of place in this section.

2. Bayes factors demonstrate that, for VLE, a 15-6 or 16-6 potential are favored significantly more than a 14-6 (could include 17-6 or 18-6 as well)

3. CH2 credible regions overlap considerably between propane, n-butane, and n-octane

4. By comparing the Bayes factor of a transferable CH2 site and three independent CH2 sites we observe that the CH2 sites are indistinguishable

5. Statement about CH credible regions for isobutane, isopentane, and isohexane

6. Statement about C credible region for isooctane and neopentane

5.2 Propagation of uncertainties

1. Figure: Uncertainties in rholsat and Pvsat for n-alkanes. Include 14-6, 15-6, 16-6.
2. Figure: Uncertainties in rholsat and Pvsat for branched alkanes. Include only 16-6.
3. Clearly the uncertainties are fairly conservative due to the relatively large numerical uncertainties we assigned in the posterior
4. Figure: Z vs 1000/T for n-alkanes where the error bars represent the Bayesian uncertainties from VLE. Include 14-6, 15-6, and 16-6 results.
5. The 16-6 potential is not able to predict both VLE and compressed liquid/supercritical pressures
6. VLE is much worse for 14-6, about the same for 15-6
7. Condensed liquid pressures are slightly better for 14-6 and 15-6 but still over predict
8. Figure: Z vs 1000/T for branched alkanes. Results are only included for the 16-6 potential.
9. Same results as for normal alkanes.

1. We modify the posterior by excluding the Pvsat data and including the REFPROP correlations at high pressures
2. Figure: I can either include the parameter uncertainties here or back in the Parameter Uncertainties section. I could even move this to supporting information
3. Bayes ratios show the evidence for different values of λ
4. We recommend that lower values of λ be favored

6 Conclusions

Recently, molecular simulation results at extreme temperatures and pressures have been used to supplement experimental data when developing a fundamental equation of state. As discussed by Thol et al., due to uncertainties and deficiencies in the force field, experimental data should be favored over molecular simulation values whenever possible. However, in principle, a FEOS could be developed for compounds without any experimental data by using only molecular simulation results, if the force field were reliable and transferable over different PVT conditions. In part, one of our aims was to determine whether the united-atom Mie λ-6 potential for normal and branched alkanes was reliable enough that a FEOS could be developed strictly from molecular simulation results. Unfortunately, the Bayesian statistical analysis performed in this study suggests that this model type (UA Mie λ-6) is not adequate for predicting both VLE properties and high pressures for supercritical fluids and compressed liquids. Specifically, no set of ϵ, σ, and λ can adequately predict VLE and PVT behavior. Therefore, we recommend that alternative models be considered for developing FEOS, such as force fields using anisotropic-united-atom, all-atom, and/or alternative non-bonded potentials, e.g. Buckingham exponential-6, extended Lennard-Jones, etc.

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.; Thol, M.; Lustig, R.; Span, R.; Vrabec, J. *The Journal of Chemical Physics* **2013**, *139*, 041102.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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