

AUTHOR QUERY FORM

	<p>Journal: J. Chem. Phys.</p> <p>Article Number: 005836JCP</p>	<p>Please provide your responses and any corrections by annotating this PDF and uploading it to AIP's eProof website as detailed in the Welcome email.</p>
---	---	--

Dear Author,

Below are the queries associated with your article. Please answer all of these queries before sending the proof back to AIP.

Article checklist: In order to ensure greater accuracy, please check the following and make all necessary corrections before returning your proof.

1. Is the title of your article accurate and spelled correctly?
2. Please check affiliations including spelling, completeness, and correct linking to authors.
3. Did you remember to include acknowledgment of funding, if required, and is it accurate?

Location in article	Query/Remark: click on the Q link to navigate to the appropriate spot in the proof. There, insert your comments as a PDF annotation.
Q1	Please check that the author names are in the proper order and spelled correctly. Also, please ensure that each author's given and surnames have been correctly identified (given names are highlighted in red and surnames appear in blue).
Q2	Please note that the "Author Contributions" paragraph has been deleted in accordance with standard AIP Publishing guidelines, which do not allow for the listing of the specific contributions of individual authors.
Q3	Please provide a brief description of the supplementary material.
Q4	Please confirm the page number in Refs. 51, as we have inserted the required information.

Thank you for your assistance.

Uncertainty quantification confirms unreliable extrapolation toward high pressures for united-atom Mie λ -6 force field

Richard A. Messerly,^{1,a)} Michael R. Shirts,^{2,b)} and Andrei F. Kazakov^{1,c)}

¹*Thermodynamics Research Center, National Institute of Standards and Technology, Boulder, Colorado 80305, USA*

²*Department of Chemical and Biological Engineering, University of Colorado, Boulder, Colorado 80309, USA*

(Received 8 May 2018; accepted 2 September 2018; published online XX XX XXXX)

Molecular simulation results at extreme temperatures and pressures can supplement experimental data when developing fundamental equations of state. Since most force fields are optimized to agree with vapor-liquid equilibria (VLE) properties, however, the reliability of the molecular simulation results depends on the validity/transferability of the force field at higher temperatures and pressures. As demonstrated in this study, although state-of-the-art united-atom Mie λ -6 potentials for normal and branched alkanes provide accurate estimates for VLE, they tend to over-predict pressures for dense supercritical fluids and compressed liquids. The physical explanation for this observation is that the repulsive barrier is too steep for the “optimal” united-atom Mie λ -6 potential parameterized with VLE properties. Bayesian inference confirms that no feasible combination of non-bonded parameters (ϵ , σ , and λ) is capable of simultaneously predicting saturated vapor pressures, saturated liquid densities, and pressures at high temperatures and densities. This conclusion has both practical and theoretical ramifications, as more realistic non-bonded potentials may be required for accurate extrapolation to high pressures of industrial interest. <https://doi.org/10.1063/1.5039504>

I. INTRODUCTION

An accurate understanding of the relationship between pressure (P), density (ρ), and temperature (T) and caloric properties (such as internal energy U) for a given compound is essential for designing industrial chemical processes. Fundamental equations of state (FEOSs), such as those based on the Helmholtz free energy, are a powerful approach for estimating $P\rho T$ behavior and caloric properties. For example, the National Institute of Standards and Technology (NIST) Reference Fluid Properties (REFPROP) currently provides FEOS for approximately one hundred and fifty 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 FEOSs are semi-empirical and have 50 to 100 fitting parameters, the FEOS predictions can result in large errors at temperatures and pressures that are significantly higher than those used in parameterizing the FEOS, which are typically near or below the critical temperature and pressure. Therefore, improvement in an FEOS at high temperatures and pressures necessitates additional data for those conditions.

The lack of experimental data at high temperatures and pressures, especially, is 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 FEOSs.^{2–6} Although it is possible to fit an FEOS to just molecular simulation results, the recommended approach is to implement hybrid data sets, i.e., from both experiment and molecular simulation.⁷

For example, several recent studies 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, and 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 feed-stock for many petrochemical processes, and, therefore, a large body of experimental data exist covering a wide range of $P\rho T$ phase space for some alkanes. For these reasons, REFPROP provides highly accurate FEOSs 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 FEOSs for industrially relevant hydrocarbons with minimal experimental

^{a)}Electronic mail: richard.messerly@nist.gov

^{b)}Electronic mail: michael.shirts@colorado.edu

^{c)}Electronic mail: andrei.kazakov@nist.gov

82 data, i.e., those with longer chain-lengths or a higher degree
83 of branching.

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

96 The most accurate force fields for estimating hydrocarbon
97 VLE properties, such as ρ_1^{sat} and P_v^{sat} , are the Transferable
98 Potentials for Phase Equilibria (TraPPE)^{12,13} (and, especially, the recent TraPPE-2¹⁴), Errington,¹⁵ fourth generation
99 anisotropic-united-atom (AUA4),^{16,17} Potoff,^{18,19} and Transferable anisotropic Mie potential (TAMie).^{20,21} Note that each
100 is either an united-atom (UA) or an anisotropic-united-atom (AUA) force field. In contrast with the more computationally
101 expensive all-atom (AA) approach, both UA and AUA models group the hydrogen interaction sites with their neighboring
102 carbon atom. Although an AA force field should, in principle, be able to yield more accurate VLE results, it is much easier to
103 locate the “true” optimal parameter set for UA and AUA force fields since fewer (highly correlated) parameters are optimized
104 simultaneously.

105 In addition to the division between UA and AUA force fields,
106 the existing force fields differ in the non-bonded functional form and corresponding parameters. The TraPPE,
107 TraPPE-2, and AUA4 force fields use a Lennard-Jones (LJ)
108 12-6 potential, while the Potoff and TAMie force fields use the Mie λ -6 (or generalized Lennard-Jones) potential, and
109 the Errington force field uses the Buckingham exponential-6 (Exp-6) potential. The Mie λ -6 potential is²²

$$110 u^{\text{vdw}}(\epsilon, \sigma, \lambda; r) = \left(\frac{\lambda}{\lambda - 6} \right) \left(\frac{\lambda}{6} \right)^{\frac{6}{\lambda-6}} \epsilon \left[\left(\frac{\sigma}{r} \right)^\lambda - \left(\frac{\sigma}{r} \right)^6 \right], \quad (1)$$

111 where u^{vdw} is the van der Waals interaction, σ is the distance
112 (r), where $u^{\text{vdw}} = 0$, $-\epsilon$ is the energy of the potential at the
113 minimum (i.e., $u^{\text{vdw}} = -\epsilon$ and $\frac{du^{\text{vdw}}}{dr} = 0$ for $r = r_{\min}$), and λ
114 is the repulsive exponent.

115 The Buckingham exponential-6 model is¹⁵

$$116 u^{\text{vdw}}(\epsilon, r_{\min}, \alpha; r) = \begin{cases} \frac{\epsilon}{1 - \frac{6}{\alpha}} \left[\frac{6}{\alpha} \exp\left(\alpha \left[1 - \frac{r}{r_{\min}}\right]\right) - \left(\frac{r_{\min}}{r}\right)^6 \right] & \text{for } r > r_{\max} \\ \infty & \text{for } r < r_{\max}, \end{cases} \quad (2)$$

117 where u^{vdw} , ϵ , and r are the same as in Eq. (1), r_{\min} is the
118 distance that corresponds to the minimum in the potential (i.e.,
119 $u^{\text{vdw}}(r_{\min}) = -\epsilon$), α is a Buckingham exponential-6 parameter,
120 and r_{\max} is the smallest positive value for which $\frac{du^{\text{vdw}}}{dr} = 0$.

121 The three-parameter Mie λ -6 and Exp-6 potentials are
122 more flexible than the two-parameter LJ 12-6 potential as an
123 additional adjustable parameter controls the steepness of the
124 repulsive barrier. Note that the Mie λ -6 potential reduces to
125 the LJ 12-6 potential for $\lambda = 12$. Therefore, the LJ 12-6 potential
126 can be considered as a special subclass of the Mie λ -6 potential.

127 Previous work demonstrated that the UA LJ 12-6 potential
128 cannot adequately estimate both ρ_1^{sat} and P_v^{sat} for
129 n -alkanes.^{23,24} For this reason, the TraPPE-UA force field was
130 primarily developed to agree with ρ_1^{sat} (and the critical tem-
131 perature, T_c).¹² By contrast, accurate prediction of both ρ_1^{sat}
132 and P_v^{sat} over a wide temperature range is possible by varying
133 the repulsive exponent of the LJ potential (i.e., the Mie λ -6
134 potential). Although an attractive exponent of 6 has a strong
135 theoretical basis, $\lambda = 12$ (LJ 12-6) is a historical artifact that
136 was chosen primarily for computational purposes (see pp. 18
137 and 140–143 of Ref. 25).

138 Typically, when parameterized to VLE data, the optimal
139 value of λ is greater than 12 with a corresponding increase
140 in the well depth (ϵ). Specifically, for most hydrocarbons, the
141 Potoff UA force field^{18,19} uses $\lambda = 16$, while the TAMie force
142 field²⁰ uses $\lambda = 14$. Gordon also demonstrated that reliable
143 viscosities can be obtained from a UA Mie λ -6 model for
144 n -alkanes by using $\lambda = 14$ and $\lambda = 20$ for the CH_3 and CH_2
145 sites, respectively (note that the subtle difference in how Gordon
146 defines the Mie λ -6 potential, also known as “mod-n-6”).²⁶
147 However, it is important to note that Gordon and Galliéro
148 *et al.* reported the λ values of 11 and 10, respectively, for
149 UA methane when optimized with viscosity data.^{26,27}

150 There are some theoretical concerns that increasing
151 the repulsive exponent might have some undesirable conse-
152 quences, especially at high pressures, where particles will
153 spend more time with very short pairwise distances than at
154 VLE conditions. For example, Refs. 28–30 demonstrate that
155 neither an all-atom LJ 12-6 nor an all-atom LJ 9-6 is adequate
156 to reproduce high-level *ab initio* calculations of n -alkanes
157 ranging from methane to n -butane. The studies of Rowley
158 *et al.* suggest that a modified-Morse potential is necessary for
159 accurate representation of *ab initio* dimer energies.^{28,29} Hayes
160 *et al.* confirmed these results while also emphasizing that the
161 short-range repulsive forces, which are most important when
162 computing high pressures in molecular simulation, are poorly
163 represented with an AA LJ 12-6 or an AA LJ 9-6 model.³⁰
164 Specifically, the LJ 12-6 potential is too steep, and only slight
165 improvement in the repulsive region is observed for the LJ
166 9-6 potential. Note that Hayes *et al.* also highlighted deficiencies
167 in the repulsive region for the Buckingham exponential-6
168 potential.

169 Recently, Kulakova *et al.* used the Bayesian inference
170 to conclude that experimental data for argon, specifically the
171 liquid and vapor radial distribution functions at varying tem-
172 peratures and densities, support λ values between 6 and 10,
173 while argon dimer *ab initio* energies support λ values between
174 12 and 14.³¹ They suggest that these larger values of λ should

not be used for liquid phase simulations. By contrast, two other studies of noble gases, including argon, support $\lambda \geq 12$.^{27,32} Specifically, Mick *et al.* reported a 13-6 potential for argon, while Galliéro *et al.* stated that the 12-6 potential is superior for argon than the 10-6, 14-6, 16-6, 18-6, and 20-6 potentials. The likely explanation for this discrepancy is the choice of experimental data. The optimal value of λ from the work of Kulakova *et al.* is based on the radial distribution function, while Galliéro *et al.* used viscosity and pressure, and Mick *et al.* utilized VLE data.

Structural properties, such as the radial distribution function, and *ab initio* calculations provide considerable insight into the true repulsive barrier.^{27,31} However, the “correct” value of λ does not guarantee adequate prediction of VLE and/or $P\rho T$ 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. For example, high pressure properties are sensitive to extremely close-range interactions ($r < 0.8\sigma$), while such distances are rarely sampled with VLE simulations and, thus, do not impact VLE properties. Furthermore, the “optimal” λ is an “effective optimal” as it accounts for numerous model assumptions, such as pair-wise additivity (i.e., excluding three-, four-, etc., body interactions) or the lack of explicit hydrogens. For these reasons, despite theoretical evidence that the repulsive barrier should be softer than $\lambda = 12$, a UA Mie λ -6 potential is simply not capable of predicting VLE properties of ethane for $\lambda < 12$ (see Figs. 1 and 2 of Ref. 18).

The purpose of this study is to determine whether or not the UA Mie λ -6 model is adequate for predicting *both* VLE and $P\rho T$ at high temperatures and pressures for alkanes. Although the theoretical results discussed previously for noble gases and all-atom *n*-alkane models are not necessarily applicable to UA models for normal and branched alkanes, the working hypothesis based on the literature is that a UA Mie λ -6 potential parameterized with VLE data is too repulsive and, thus, performs poorly at high pressures. This assessment is of practical engineering importance for deciding whether or not UA Mie λ -6 force fields should be used when developing fundamental equations of state for alkanes based on hybrid data sets.

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

II. MOLECULAR DYNAMICS

A. Simulation details

Four normal and four branched alkanes of varying chain-length and degree of branching are simulated in this study. Specifically, we simulate ethane, propane, *n*-butane, *n*-octane,

isobutane (2-methylpropane), isohexane (2-methylpentane), isoctane (2,2,4-trimethylpentane), and neopentane (2,2-dimethylpropane). These compounds were chosen to represent a diverse set of normal and branched alkanes available in REFPROP.^{1,33–38}

Molecular dynamics 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 2018.³⁹ Each simulation uses the velocity Verlet integrator with a 2 fs time step,⁴⁰ 1.4 nm cutoff for non-bonded interactions with tail corrections for energy and pressure, Nosé-Hoover thermostat with a thermostat time constant of 1 ps,⁴¹ and fixed bond-lengths constrained using LINear Constraint Solver (LINCS) with a LINCS-order of eight.^{42,43} Note that GROMACS non-bonded tail corrections assume that the long-range contribution from the $r^{-\lambda}$ term is negligible compared to the r^{-6} term. A comparison between the energies and pressures obtained with GROMACS and other (slower) simulation packages verified that the small error introduced with this approximation does not significantly affect our results. For this reason, we do not attempt to modify the GROMACS default tail correction values to include the $r^{-\lambda}$ contribution. Also, Coulombic interactions are not computed as none of the force fields require partial charges for the compounds studied.

The equilibration time is 0.1 ns for ethane and propane, 0.2 ns for *n*-butane, and 0.5 ns for all other compounds. The production time is 1 ns for ethane, 2 ns for propane and *n*-butane, and 4 ns for all other compounds. Replicate simulations are 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 the supplementary material.

The specific state points for each compound studied are depicted in Fig. 1 (for tabulated values, see Sec. SI.I of the supplementary material). These state points correspond to the recommended conditions for the isothermal isochoric integration (ITIC) algorithm discussed in Sec. IV.^{44–46} Simulations are performed along a supercritical isotherm with a reduced temperature ≈ 1.2 ($T_r \equiv \frac{T}{T_c}$, where T_c is the critical temperature) and five saturated liquid density isochores (ρ^{IC}). Nine densities are simulated along the supercritical isotherm (T^{IT}). Simulations along each isochore are performed at three temperatures, namely, T^{IT} , the REFPROP saturation temperature (T^{sat}), and $2/(1/T^{IT} + 1/T^{sat})$. Since five of the isotherm densities correspond to the five different ρ^{IC} values, a total of 19 simulations are performed for each compound and force field.

B. Force field

A united-atom (UA) or anisotropic-united-atom (AUA) representation is used for each compound studied. UA models assume that the UA interaction site is that of the carbon atom, while AUA models assume that the AUA interaction site is displaced from the carbon atom toward the hydrogen atom(s). Note that TraPPE and Potoff are UA force fields, while

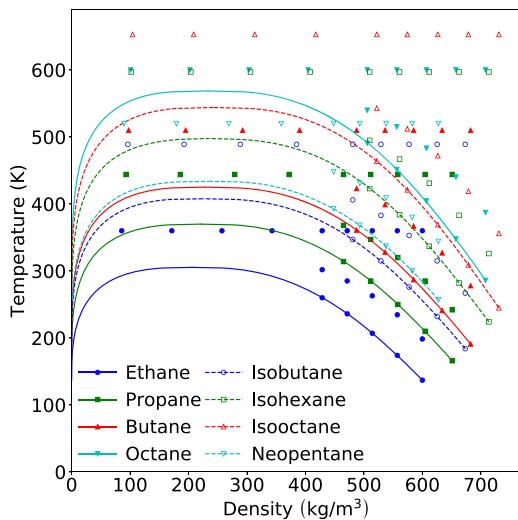


FIG. 1. State points simulated for ethane, propane, *n*-butane, *n*-octane, isobutane, iso hexane, iso octane, and neopentane. A total of 19 simulations are performed: nine densities along the supercritical isotherm and two temperatures along five liquid density isochores. Filled symbols and solid lines 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 [Refs. 1 and 33–38].

the TraPPE-2, Errington, AUA4, and TAMie are AUA force fields.

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. 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 and AUA sites. Although TAMie is an AUA force field, only the terminal CH_3 sites have a displacement in the interaction site. For example, Fig. 2 depicts both the UA and AUA representations of iso octane when only terminal CH_3 interaction sites

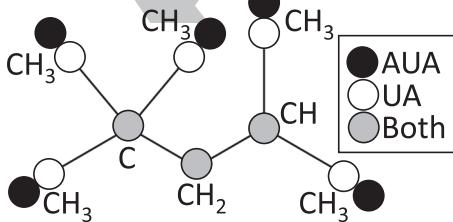


FIG. 2. Comparison between AUA and UA representations of iso octane. AUA force fields have the same complexity as UA force fields if only the terminal (CH_3) sites have an anisotropic displacement, i.e., a longer effective bond-length. Note that the AUA4 approach requires a more complicated displacement of CH_2 and CH sites than that depicted here.¹⁶

TABLE I. Effective bond-lengths in units of nm for terminal (CH_3) UA or AUA interaction sites. Empty table entries for Exp-6¹⁵ and TraPPE-2¹⁴ denote that the force field does not contain the corresponding interaction site type. Empty table entries in AUA4¹⁶ arise because this force field uses a more complicated construction than the simple effective bond-length approach. Specifically, AUA4 requires CH_2 and CH interaction sites that are not along the C-C bond axis.

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

are displaced from the carbon center. This convention is much simpler to implement than other AUA approaches (such as AUA4) where non-terminal (i.e., CH_2 and CH) interaction sites also have a displacement distance. For this reason, we do not attempt to simulate the AUA4 force field for any compounds containing CH_2 and CH interaction sites. For the compounds and force fields simulated, the anisotropic displacement in a terminal interaction site (i.e., CH_3) is treated simply as a longer effective bond-length (see Table I). 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.

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

$$u^{\text{bend}} = \frac{k_\theta}{2}(\theta - \theta_0)^2, \quad (3)$$

where u^{bend} is the bending energy, θ is the instantaneous bond angle, θ_0 is the equilibrium bond angle (see Table II), and k_θ is the harmonic force constant with $k_\theta/k_B = 62\,500 \text{ K}/\text{rad}^2$ for all bonding angles, where k_B is the Boltzmann constant.

Dihedral torsional interactions are determined using a cosine series,

$$u^{\text{tors}} = c_0 + c_1[1 + \cos \phi] + c_2[1 - \cos 2\phi] + c_3[1 + \cos 3\phi], \quad (4)$$

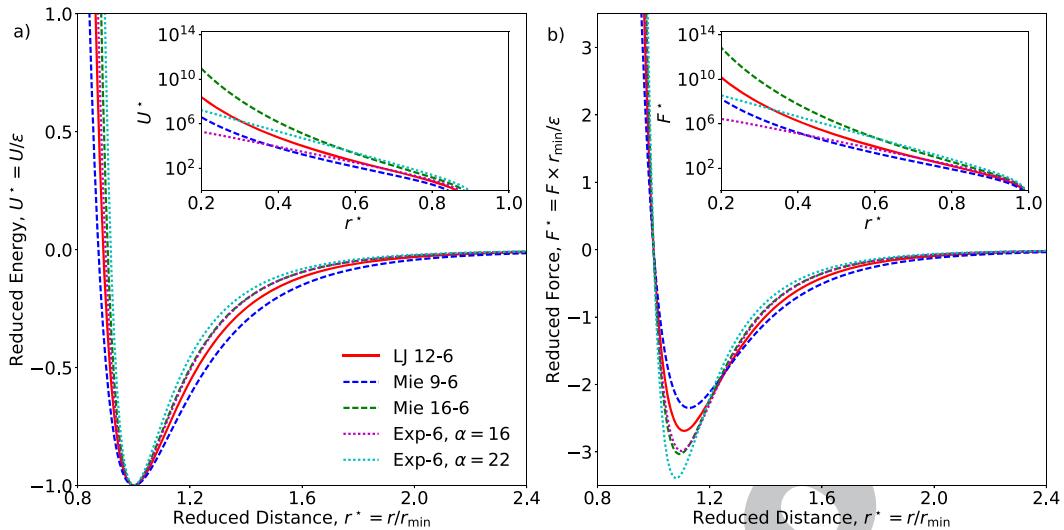
where u^{tors} is the torsional energy, ϕ is the dihedral angle, and c_n are the Fourier constants (see Table III). Note that the

TABLE II. Equilibrium bond angles (θ_0).¹³ CH_i and CH_j represent CH_3 , CH_2 , CH, or C sites.

Bending sites	θ_0 (deg)
$CH_i-CH_2-CH_j$	114.0
$CH_i-CH-CH_j$	112.0
CH_i-C-CH_j	109.5

TABLE III. Fourier constants (c_n/k_B) in units of K.¹³ CH_i and CH_j represent CH_3 , CH_2 , CH, or C sites.

Torsion sites	c_0/k_B	c_1/k_B	c_2/k_B	c_3/k_B
$CH_i-CH_2-CH_2-CH_j$	0.0	355.03	-68.19	791.32
$CH_i-CH_2-CH-CH_j$	-251.06	428.73	-111.85	441.27
$CH_i-CH_2-C-CH_j$	0.0	0.0	0.0	461.29



396 FIG. 3. Comparison between Eqs. (1) and (2). Short distance forces increase and energy wells become more narrow with increasing λ and α . Panels (a) and (b)
397 plot the reduced energy (U^*) and force (F^*) with respect to the reduced distance (r^*), where the energy and distance are scaled by ϵ and r_{\min} , respectively.

398 TABLE IV. Non-bonded (intermolecular) parameters for TraPPE^{12,13} (and TraPPE-2¹⁴), Potoff^{18,19}, AUA4,^{16,47}
399 and TAMie^{20,21} force fields. The “short/long” Potoff CH and C parameters are included in the parentheses. The
400 ethane specific parameters for TAMie are included in the parentheses.

United-atom	TraPPE (TraPPE-2)			Potoff (S/L)		
	ϵ/k_B (K)	σ (nm)	λ	ϵ/k_B (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
<hr/>						
AUA4						
CH ₃	120.15	0.3607	12	136.318 (130.780)	0.360 34 (0.364 63)	14
CH ₂	86.29	0.3461	12	52.9133	0.404 00	14
CH	50.98	0.3363	12	14.5392	0.436 56	14
C	15.04	0.244	12

415 Errington c_n values for CH_i–CH₂–CH₂–CH_j are a factor of
416 two less than those reported in Table III.¹⁵

417 Non-bonded interaction energies and forces between sites
418 located in two different molecules or separated by more than
419 three bonds are calculated using either a Lennard-Jones 12-6,
420 Mie λ -6, or Buckingham Exponential-6 potential [see Eqs. (1)
421 and (2)]. Figure 3 compares the energy and force of the LJ
422 12-6, Mie 9-6, Mie 16-6, and Exp-6 (for $\alpha = 16$ and 22) using
423 the same values of ϵ and r_{\min} .

424 The non-bonded LJ 12-6 or Mie λ -6 force field parameters
425 for TraPPE, TraPPE-2, Potoff, AUA4, and TAMie are provided
426 in Table IV. Note that, for computational purposes, a common
427 practice to date is to use integer values of λ in Eq. (1).

428 The Errington Exp-6 non-bonded parameters are found
429 in Table V. Note that Errington reported values for ϵ , σ , and
430 α . We compute r_{\min} and r_{\max} to facilitate compatibility with
431 Eq. (2) and future validation of our results.

432 Non-bonded interactions between two different site types
433 (i.e., cross-interactions) are determined using the Lorentz-
434 Berthelot combining rules²⁵ for ϵ and σ , an arithmetic mean

435 for the repulsive exponent λ (as recommended in Ref. 18), and
436 a geometric mean for α (as recommended in Ref. 15),

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}, \quad (5) \quad 437$$

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}, \quad (6) \quad 438$$

$$\lambda_{ij} = \frac{\lambda_{ii} + \lambda_{jj}}{2}, \quad (7) \quad 440$$

$$\alpha_{ij} = \sqrt{\alpha_{ii}\alpha_{jj}}, \quad (8) \quad 441$$

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

444 TABLE V. Non-bonded (intermolecular) parameters for Errington Exp-6
445 force field.¹⁵

United-atom	ϵ/k_B (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

452

III. CASE STUDY

453

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 $P\rho T$ behavior at higher temperatures and pressures (with the exception of ethane for the TraPPE-2 potential). Figures 4 and 5 plot the compressibility factor with respect to the inverse temperature for *n*-alkanes and branched alkanes, respectively. Note that saturation corresponds to $Z \approx 0$ for each isochore. The “Potoff” results in Fig. 5 are only for the “short/long” model since the “short/long” model is more accurate than the “generalized” model (available in Sec. SI.II of the [supplementary material](#)).

465

Note that Figs. 4 and 5 include a constant 1% uncertainty in the REFPROP correlations for all compounds at all state points. This is a conservative estimate as the reported REFPROP uncertainty for ρ is typically only 1% at T^{IT} , while it decreases near T^{sat} to a value $\leq 0.2\%$. Furthermore, only ethane,³³ *n*-butane,³⁵ isohexane,³⁷ and neopentane³⁷ have a

reported uncertainty of 1% at T^{IT} . REFPROP uncertainties for propane,³⁴ isobutane,³⁵ and *n*-octane³⁶ are actually 0.1%, 0.4%, and 0.5%, respectively, while isoctane³⁷ does not have a reported uncertainty.

Figure 4 demonstrates that the existing literature force fields for *n*-alkanes, while accurate for VLE ($Z \approx 0$), do not capture the correct $P\rho T$ behavior at high pressures (P^{high}), i.e., Z at the higher temperatures ($T > T^{\text{sat}}$) and highest isochore densities (ρ_3^{IC} and ρ_4^{IC}). Figure 5 shows the same erroneous trend in Z for branched alkanes. Note that the error in Z at high temperatures is less obvious because these force fields are typically not as reliable at predicting VLE for branched alkanes as for *n*-alkanes, i.e., notice the large deviations at $Z \approx 0$. However, it is clear in both Figs. 4 and 5 that none of the force fields adequately reproduces Z over the entire temperature range, or the slope of Z with respect to inverse T .

A surprising trend is that the Errington (AUA Exp-6) model has a positive bias at high pressures. This appears to suggest that the repulsive barrier is too steep, despite the fact

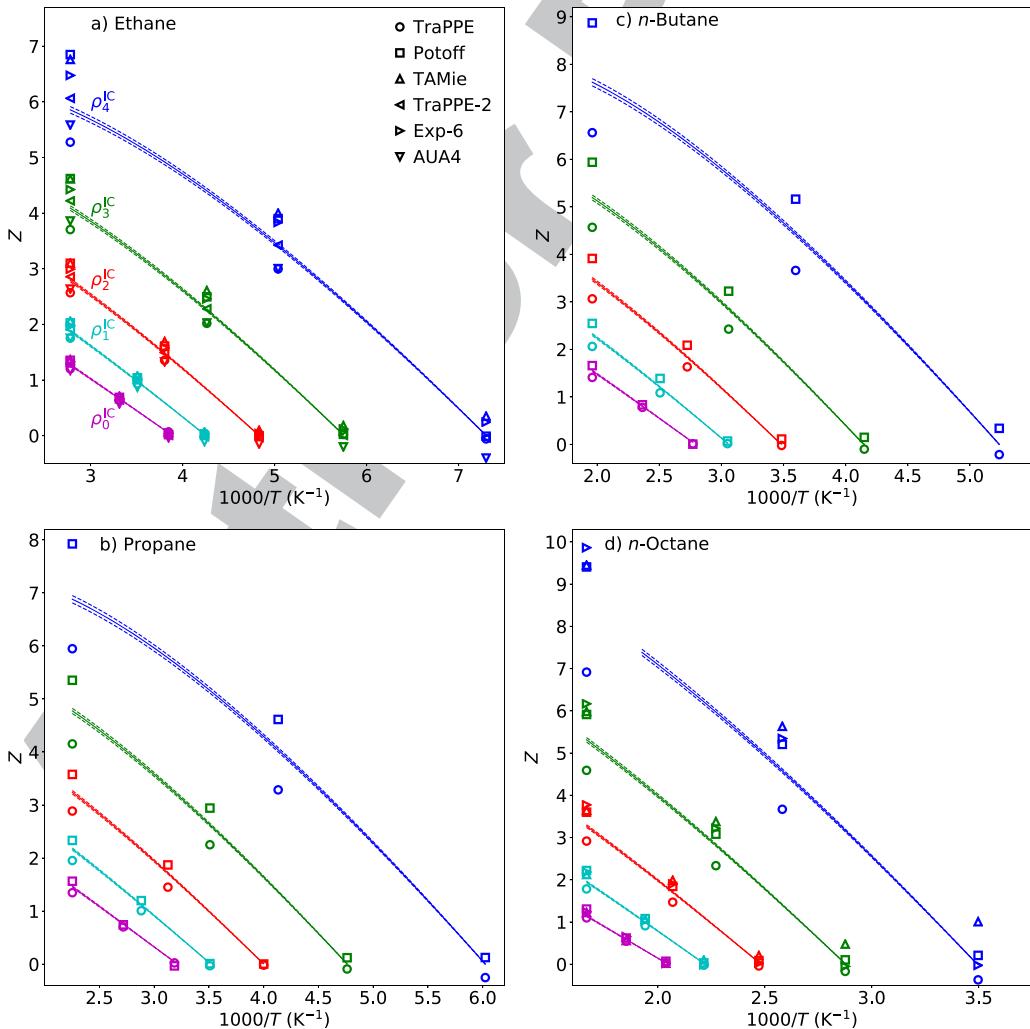


FIG. 4. Compressibility factor (Z) along isochores for *n*-alkanes. Most force fields are accurate at saturation ($Z \approx 0$) but deviate strongly at higher pressures. Different symbol shapes correspond to the various force fields, with the legend in Panel (a). Densities are distinguished by color, increase vertically, and are labeled such that $\rho_0^{\text{IC}} < \rho_1^{\text{IC}} < \rho_2^{\text{IC}} < \rho_3^{\text{IC}} < \rho_4^{\text{IC}}$. Panels (a)–(d) correspond to ethane, propane, *n*-butane, and *n*-octane, respectively. Solid lines represent REFPROP correlations, with dashed lines representing a 1% uncertainty in REFPROP values.^{1,33–36} Simulation error bars computed with block averaging are approximately one symbol size.

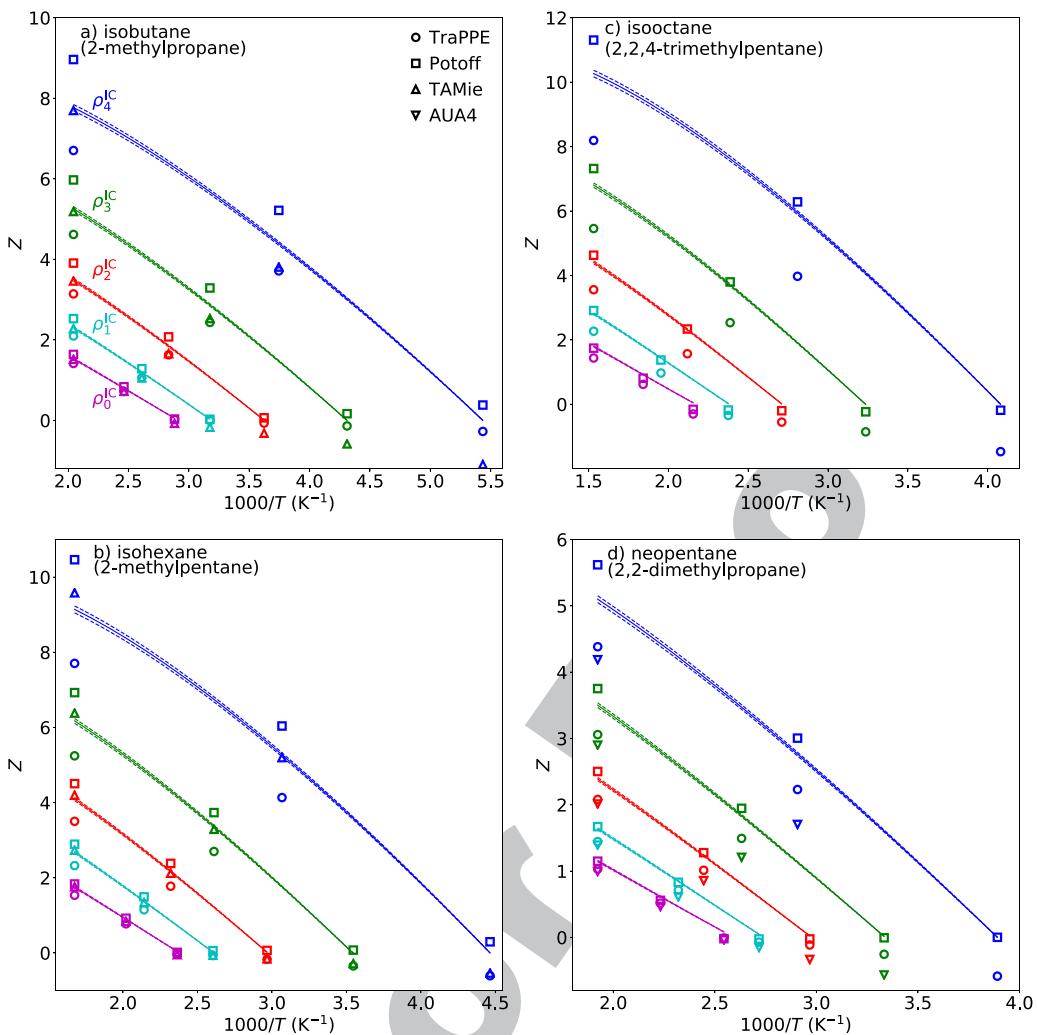


FIG. 5. Compressibility factor (Z) along isochores for branched alkanes. Force fields are not as accurate as normal alkanes at saturation ($Z \approx 0$) and deviate strongly at higher pressures. Panels (a)–(d) correspond to isobutane, isohexane, isoctane, and neopentane, respectively. Symbols, lines, uncertainties, and formatting are the same as those in Fig. 4. The Potoff results for isobutane and neopentane use the “short” parameters, while isohexane and isoctane use the “long” parameters (see Table IV).¹⁹

that the Exp-6 model is typically considered softer than the LJ 12-6. However, the Exp-6 is less repulsive than the LJ 12-6 only at very short distances, e.g., $r < 0.7r_{\min}$ for $\alpha = 16$ and $r < 0.3r_{\min}$ for $\alpha = 22$, while it is actually somewhat more repulsive for the closest-range distances sampled in molecular dynamics at these conditions, i.e., $0.7r_{\min} < r < r_{\min}$ [see the inset of Panel (b) in Fig. 3]. More definitive and straightforward conclusions regarding the shape of the Mie λ -6 repulsive barrier are possible by directly comparing different values of λ .

In general, clear systematic biases are observed for the LJ 12-6 potentials and the Mie λ -6 potentials. Specifically, the LJ 12-6 (TraPPE-UA and AUA4) and Mie λ -6 (Potoff and TAMie) 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.

The one exception to this trend is the TraPPE-2 model for ethane, which has the most accurate prediction of the entire $P\rho T$ phase space simulated. Specifically, TraPPE-2 reproduces the REFPROP Z to within 1% for all state points except

at P^{high} , where the average percent deviation (AD%) relative to the REFPROP correlations is still only 3%.

The performance of TraPPE-2 is somewhat surprising considering that this force field has only three fitting parameters (ϵ , σ , and the effective bond-length), while the TAMie model has these three parameters and an additional fitting parameter (λ). It is possible that a four-parameter optimization, such as that used by TAMie, is overfit to the VLE data and would perform better if high pressure $P\rho T$ data were included in the parameterization. Furthermore, it is important to note that TraPPE-2 uses a much longer effective bond-length of 0.230 nm, while TAMie did not consider bond-lengths larger than 0.194 nm. Therefore, the fact that the TraPPE-2 force field extrapolates to high pressures better than TAMie suggests that, at high pressures, it is important to account for hydrogens with a longer effective bond-length than that typically used for AUA models (see Table I).

Unfortunately, a direct comparison of the non-bonded interactions for AUA force fields is difficult because each model has a different anisotropic displacement, i.e., effective bond-length. By contrast, comparing TraPPE-UA and Potoff

is straightforward because they use the same bond-lengths and the same non-bonded Mie λ -6 potential [Eq. (1)]. For example, since the TraPPE-UA (LJ 12-6) potential under-predicts Z and the Potoff (UA Mie 16-6) potential over-predicts Z, it seems reasonable that a UA Mie 13-6, 14-6, or 15-6 model could demonstrate the proper trend.

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 to be consistent with the TraPPE and Potoff UA models. Specifically, we perform a Bayesian uncertainty quantification analysis to determine if there exists a set of ϵ , σ , and λ that reasonably predicts ρ_1^{sat} , P_v^{sat} , and P^{high} . The results in Sec. V demonstrate that the optimal value of λ for predicting $P\rho T$ of supercritical fluids and compressed liquids is not capable of predicting VLE properties accurately, and vice versa.

IV. UNCERTAINTY QUANTIFICATION

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

Bayesian inference is a rigorous approach to determine all feasible ϵ , σ , and λ parameter sets. We refer the reader to the literature for a thorough discussion of Bayesian statistics.^{31,48–51} In Sec. IV A, we review some basic concepts of Bayes' theorem, define the posterior, likelihood, and prior distributions, and discuss the Markov Chain Monte Carlo (MCMC) approach for sampling from the posterior joint distribution of the parameters. MCMC can be computationally burdensome, especially when molecular simulation is required to compute the likelihood. For this reason, we utilize surrogate models to reduce the computational cost of MCMC by several orders of magnitude. Section IV B demonstrates how these surrogate models estimate ρ_1^{sat} , P_v^{sat} , and Z for a given set of ϵ , σ , and λ . We implement this analysis for n-alkanes to generate joint distributions of $\epsilon_{\text{CH}_3} - \sigma_{\text{CH}_3}$ and $\epsilon_{\text{CH}_2} - \sigma_{\text{CH}_2}$ for different values of λ_{CH_3} and λ_{CH_2} , respectively.

A. Bayesian inference

1. Theory

Bayes' theorem states

$$Pr(\theta|D, M) = \frac{Pr(D|\theta, M)Pr(\theta|M)}{Pr(D|M)}, \quad (9)$$

where Pr denotes a probability distribution function, θ is the parameter set, M is the model, and D are the data. $Pr(\theta|D, M)$ 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,

$$Pr(M|D) = \frac{Pr(D|M)Pr(M)}{Pr(D)}, \quad (10)$$

where $Pr(M)$ is the “model prior,” $Pr(D)$ is a normalization constant, and $Pr(M|D)$ is the “model posterior.” The ratio of $Pr(M|D)$ between two different models (M_i and M_j), known as the Bayes factor (K_{ij}), provides the relative probability of models M_i and M_j , given the data D .

The parameter uncertainty propagates when estimating another quantity of interest (QoI), which may or may not be included in D , according to³¹

$$Pr(QoI|D, M) = \int Pr(QoI|\theta, M)Pr(\theta|D, M)d\theta. \quad (11)$$

This expression is commonly referred to as “robust posterior prediction.” Note that the uncertainty in QoI , obtained from $Pr(QoI|D, M)$, does not account for deficiencies in the model itself, only the uncertainty in the model parameters.

2. Application

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 based on VLE data. For clarity, we rewrite Eqs. (9) and (10) for the specific case studied by substituting ϵ and σ for θ , λ for M , and ρ_1^{sat} and P_v^{sat} for D ,

$$Pr(\epsilon, \sigma|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda) = \frac{L(\epsilon, \sigma|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda)Pr(\epsilon, \sigma|\lambda)}{Pr(\rho_1^{\text{sat}}, P_v^{\text{sat}}|\lambda)}, \quad (12)$$

$$Pr(\lambda|\rho_1^{\text{sat}}, P_v^{\text{sat}}) = \frac{Pr(\rho_1^{\text{sat}}, P_v^{\text{sat}}|\lambda)Pr(\lambda)}{Pr(\rho_1^{\text{sat}}, P_v^{\text{sat}})}, \quad (13)$$

where in this context ρ_1^{sat} and P_v^{sat} are arrays of experimental data values. Note that θ does not include λ since we use λ to distinguish between models. The “model evidence,” $Pr(\rho_1^{\text{sat}}, P_v^{\text{sat}}|\lambda)$ in Eq. (13), for different values of λ is determined by integrating the numerator of Eq. (12) for all values of ϵ and σ .

To compute the Bayes factor between two values of λ (i.e., between different models, M_i and M_j), we assume that the prior evidence is equal for all positive values of ϵ , σ , and λ (within a feasible range). Specifically, we use bounded uniform prior distributions for $Pr(\epsilon, \sigma|\lambda)$ in Eq. (12) and $Pr(\lambda)$ in Eq. (13), where the lower bound is 0 and the upper bound is an order of magnitude greater than the literature values for ϵ , σ , and λ . Due to the large amount of information contained in the data, D , the use of a uniform prior does not impact our results, i.e., the data “overwhelm” the prior. One advantage of using a uniform prior is that the Bayes factor, K_{ij} , depends completely on the likelihood,

$$K_{ij} = \frac{Pr(\lambda_j|\rho_1^{\text{sat}}, P_v^{\text{sat}})}{Pr(\lambda_i|\rho_1^{\text{sat}}, P_v^{\text{sat}})} = \frac{\int L(\epsilon, \sigma|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda_j)d\epsilon d\sigma}{\int L(\epsilon, \sigma|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda_i)d\epsilon d\sigma}, \quad (14)$$

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

We utilize robust posterior prediction [Eq. (11)] to propagate the joint parameter uncertainty in ϵ and σ (for a given λ) to three different *QoI*, specifically, ρ_1^{sat} , P_v^{sat} , and Z . For example, the uncertainty in predicting Z is obtained from

$$\begin{aligned} & Pr(Z|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda) \\ &= \int Pr(Z|\epsilon, \sigma, \lambda) Pr(\epsilon, \sigma|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda) d\epsilon d\sigma. \quad (15) \end{aligned}$$

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

3. Implementation

Markov Chain Monte Carlo (MCMC) is the traditional approach for numerically sampling from the probability distribution $Pr(\epsilon, \sigma|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda)$. We use the Metropolis-Hastings algorithm to create a Markov chain by proposing new ϵ and σ sets and accepting those moves based on the criterion,

$$\alpha = \min\left(1, \frac{Pr(\epsilon_{i+1}, \sigma_{i+1}|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda) Q(\epsilon_{i+1}, \sigma_{i+1}|\epsilon_i, \sigma_i, s_\epsilon^2, s_\sigma^2)}{Pr(\epsilon_i, \sigma_i|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda) Q(\epsilon_i, \sigma_i|\epsilon_{i+1}, \sigma_{i+1}, s_\epsilon^2, s_\sigma^2)}\right), \quad (16)$$

where α is the acceptance probability, ϵ_i and σ_i are the previous parameter sets, ϵ_{i+1} and σ_{i+1} are the proposed parameter sets, and Q is the proposal distribution from which ϵ_{i+1} and σ_{i+1} are sampled. In this study, Q is a bi-variate normal distribution with mean equal to ϵ_i and σ_i , variance of s_ϵ^2 and s_σ^2 , and a covariance of 0.

The amount to which ϵ or σ is varied for each MCMC step (the difference between ϵ_{i+1} and ϵ_i or between σ_{i+1} and σ_i) depends on Q , specifically, on s_ϵ^2 and s_σ^2 . These parameters (s_ϵ^2 and s_σ^2) are tuned such that approximately $\frac{1}{3}$ of the moves are accepted, i.e., s_ϵ^2 and s_σ^2 are decreased if the acceptance rate of MCMC moves is less than $\frac{1}{3}$, and vice versa. This “tuning” period (also referred to as a “burn-in” period) is followed by a production period where s_ϵ^2 and s_σ^2 do not change.

Since ϵ_{i+1} and σ_{i+1} are highly correlated with ϵ_i and σ_i , it is important to “thin” the MCMC ϵ and σ parameter sets (ϵ_{MCMC} and σ_{MCMC}); i.e., every j^{th} parameter set is stored. The parameter sets sampled from MCMC (θ_{MCMC} , or specifically, ϵ_{MCMC} and σ_{MCMC}) provide a joint distribution for the feasible values of ϵ and σ (see Figs. 7 and 12 in Sec. V). Section SI.III of the supplementary material provides an MCMC example with some details (i.e., a number of steps for burn-in and production, frequency that s_ϵ^2 and s_σ^2 are updated, resulting acceptance rates, etc.).

The integration required for robust posterior prediction [see Eqs. (11) and (15)] is approximated from the distribution of *QoI* values evaluated for each MCMC parameter set, i.e., $QoI(\theta_{\text{MCMC}}|D, M)$ or the more succinct notation QoI_{MCMC} . From QoI_{MCMC} , standard statistical methods are used to approximate the *QoI* uncertainty at a desired credible level. For example, the histograms and uncertainties reported in Sec. V are obtained from distributions of ρ_1^{sat} , P_v^{sat} , and Z_{MCMC} .

Because MCMC moves are accepted based on Eq. (16) and the denominator in Eq. (12) (i.e., $Pr(\rho_1^{\text{sat}}, P_v^{\text{sat}}|\lambda)$) does not depend on ϵ and σ , the acceptance probability is independent of $Pr(\rho_1^{\text{sat}}, P_v^{\text{sat}}|\lambda)$. Also, as mentioned previously, we use a weakly informative uniform prior over a larger than feasible range of parameters such that the acceptance probability is independent of $Pr(\epsilon, \sigma|\lambda)$. Furthermore, Q is chosen to be

symmetric such that the Q terms in the numerator and denominator of Eq. (16) cancel. Therefore, the probability of accepting ϵ_{i+1} and σ_{i+1} is based completely on the likelihood,

$$\alpha = \min\left(1, \frac{L(\epsilon_{i+1}, \sigma_{i+1}|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda)}{L(\epsilon_i, \sigma_i|\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda)}\right), \quad (17)$$

where the likelihood, $L(\epsilon, \sigma, |\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda)$, is calculated from a normal distribution,

$$\begin{aligned} L(\epsilon, \sigma, |\rho_1^{\text{sat}}, P_v^{\text{sat}}, \lambda) &= \prod_k \frac{1}{\sqrt{2\pi s_{\text{D,SM}}^2(T_k^{\text{sat}})}} \\ &\times \exp\left(\frac{-\left(\rho_{1,\text{SM}}^{\text{sat}}(\epsilon, \sigma, \lambda; T_k^{\text{sat}}) - \rho_{1,k}^{\text{sat}}\right)^2}{2s_{\text{D,SM}}^2(T_k^{\text{sat}})}\right) \\ &\times \prod_k \frac{1}{\sqrt{2\pi s_{\text{D,SM}}^2(T_k^{\text{sat}})}} \\ &\times \exp\left(\frac{-\left(P_{v,\text{SM}}^{\text{sat}}(\epsilon, \sigma, \lambda; T_k^{\text{sat}}) - P_{v,k}^{\text{sat}}\right)^2}{2s_{\text{D,SM}}^2(T_k^{\text{sat}})}\right), \end{aligned} \quad (18)$$

where the first and second products are over the experimental ρ_1^{sat} and P_v^{sat} data, respectively, T_k^{sat} is the saturation temperature that corresponds to the k^{th} data point, “SM” refers to the surrogate model (see Sec. IVB) used to estimate ρ_1^{sat} or P_v^{sat} for a given ϵ , σ , λ , and T^{sat} , and $s_{\text{D,SM}}^2$ is the combined variance of the experimental data and the surrogate model. The variances are independent, meaning that the combined variance is the sum of the experimental and surrogate model variances, i.e., $s_{\text{D,SM}}^2 = s_{\text{D}}^2 + s_{\text{SM}}^2$.⁴⁹

B. Surrogate model

A typical Markov Chain requires $O(10^4$ to 10^5) Monte Carlo steps, where the likelihood function must be evaluated at each step. Since $L(\theta|D, M)$ depends on the force field parameters (ϵ , σ , and λ), an MCMC approach is computationally infeasible if computing $L(\theta|D, M)$ requires performing

738 direct molecular simulations for every proposed parameter set.
 739 Furthermore, propagation of uncertainty with robust posterior
 740 prediction may require $O(10^2$ to $10^3)$ θ_{MCMC} parameter sets
 741 for adequate representations of $Pr(QoID, M)$ [see Eqs. (11)
 742 and (15)]. For these reasons, surrogate models to estimate
 743 $\rho_{1, \text{MCMC}}^{\text{sat}}$, $P_{v, \text{MCMC}}^{\text{sat}}$, and Z_{MCMC} are essential for this study.

744 1. Multistate Bennett acceptance ratio

745 We use a configuration-sampling-based surrogate model,
 746 where configurations are sampled using a small group of reference
 747 parameter sets (θ_{ref} , consisting of ϵ_{ref} , σ_{ref} , and λ_{ref}).⁴⁵
 748 Specifically, NVT ensemble averages for the MCMC parameter
 749 sets (θ_{MCMC}) are estimated by reweighting the sampled
 750 reference configurations using Multistate Bennett Acceptance
 751 Ratio (MBAR).⁵² MBAR is a nearly exact surrogate model
 752 when a sufficient number of configurations sampled by θ_{ref} are
 753 similar to those that would be sampled with direct simulation
 754 of θ , which can be easily verified by statistical measures (see
 755 the discussion on a number of effective samples in Ref. 45).
 756 For this purpose, we use a single value of ϵ_{ref} with nine evenly
 757 spaced σ_{ref} values for each fixed value of $\lambda_{\text{ref}} = \lambda$.⁴⁵

758 2. Isothermal isochoric integration

759 The properties that are estimated using MBAR are the
 760 departure internal energy ($U^{\text{dep}} \equiv U - U^{\text{ig}}$, where U^{ig} is
 761 the ideal gas internal energy) and the compressibility factor
 762 ($Z \equiv \frac{P}{\rho R_g T}$, where R_g is the universal gas constant). Isotherm-
 763 al isochoric integration (ITIC) converts the MBAR estimated
 764 U^{dep} and Z values at the 19 ITIC state points to saturation
 765 temperatures (T^{sat}), saturated liquid densities (ρ_1^{sat}), saturated
 766 vapor densities (ρ_v^{sat}), and saturated vapor pressures (P_v^{sat}).
 767 This is important since ρ_1^{sat} and P_v^{sat} are the data (D) included
 768 in $L(\theta|D)$. Details for the combined implementation of MBAR
 769 and ITIC (MBAR-ITIC) are discussed elsewhere.⁴⁵

770 The ITIC equations are

$$771 \frac{A^{\text{dep}}}{R_g T^{\text{sat}}} = \int_0^{\rho^{\text{IC}}} \frac{Z-1}{\rho} d\rho|_{T=T^{\text{IT}}} + \int_{T^{\text{IT}}}^{T^{\text{sat}}} U^{\text{dep}} d\left(\frac{1}{R_g T}\right)|_{\rho=\rho^{\text{IC}}}, \quad (19)$$

$$772 \rho_v^{\text{sat}} \approx \rho_1^{\text{sat}} \exp\left(\frac{A^{\text{dep}}}{R_g T^{\text{sat}}} + Z_1^{\text{sat}} - 1 - 2B_2 \rho_v^{\text{sat}} - 1.5B_3 (\rho_v^{\text{sat}})^2\right), \quad (20)$$

$$773 P_v^{\text{sat}} \approx \left(1 + B_2 \rho_v^{\text{sat}} + B_3 (\rho_v^{\text{sat}})^2\right) \rho_v^{\text{sat}} R_g T^{\text{sat}}, \quad (21)$$

$$774 Z_1^{\text{sat}} = \frac{P_v^{\text{sat}}}{\rho_1^{\text{sat}} R_g T^{\text{sat}}}, \quad (22)$$

775 where $A^{\text{dep}} \equiv A - A^{\text{ig}}$ is the Helmholtz free energy departure
 776 from ideal gas for $T = T^{\text{sat}}$ and $\rho = \rho^{\text{IC}} = \rho_1^{\text{sat}}$, Z_1^{sat} is the
 777 saturated liquid compressibility factor, B_2 is the second virial
 778 coefficient, B_3 is the third virial coefficient, T^{IT} is the isothermal
 779 temperature, and ρ^{IC} is the isochoric density. For details
 780 regarding the implementation of ITIC, see Refs. 44–46. As
 781 discussed in our previous work,⁴⁵ the B_2 and B_3 values found
 782 in Eqs. (20) and (21) are calculated using REFPROP correlations.¹ The use of REFPROP correlations introduces a small

783 bias in the resulting ρ_1^{sat} and P_v^{sat} , which is accounted for in the
 784 surrogate model uncertainty.

785 The ITIC analysis provides VLE properties at only 5
 786 saturation temperature values ($T_{\text{ITIC}}^{\text{sat}}$), while the experimental
 787 data set may have hundreds of saturation temperatures (T_D^{sat}).
 788 Although it is possible for D to consist of computed values
 789 from an empirical correlation fit to experimental data [e.g.,
 790 REFPROP, ThermoData Engine (TDE)], it is considered to
 791 be best practice for Bayesian inference that raw experimental
 792 data be used for D . For this reason, we instead use empirical
 793 model fits to interpolate the ITIC VLE properties ($T_{\text{ITIC}}^{\text{sat}}$,
 794 $\rho_{1,\text{ITIC}}^{\text{sat}}$, and $P_{v,\text{ITIC}}^{\text{sat}}$) so that ρ_1^{sat} and P_v^{sat} can be estimated at
 795 any value of T^{sat} . Specifically, we fit $P_{v,\text{ITIC}}^{\text{sat}}$ and $T_{\text{ITIC}}^{\text{sat}}$ to the
 796 Antoine equation
 797

$$798 \log_{10}(P_v^{\text{sat}}) = a_0 + \frac{a_1}{T^{\text{sat}} + a_2}, \quad (23)$$

799 where a_i are fitting parameters. We fit $\rho_{1,\text{ITIC}}^{\text{sat}}$ and $T_{\text{ITIC}}^{\text{sat}}$ to a
 800 combined rectilinear and density scaling law expression²⁴

$$801 \rho_1^{\text{sat}} = b_0 + b_1(b_2 - T^{\text{sat}}) + b_3(b_2 - T^{\text{sat}})^{\beta}, \quad (24)$$

802 where b_i are fitting parameters, and $\beta = 0.326$. b_0 and b_2
 803 only provide rough estimates of the critical density (ρ_c) and
 804 critical temperature (T_c). More reliable estimates of the criti-
 805 cal point require simultaneous fitting of $\rho_{v,\text{ITIC}}^{\text{sat}}$ to a similar
 806 expression, but this is unnecessary for our purposes since D
 807 does not include the critical constants. Note that Eqs. (23)
 808 and (24) are only used to interpolate ITIC values and not to
 809 extrapolate to higher or lower T^{sat} . These equations are reli-
 810 able over the limited temperature range studied ($0.45 < T_r <$
 811 0.85), whereas a wider temperature range would require
 812 more flexible models.^{53,54}

813 In summary, MBAR, ITIC, and Eqs. (23) and (24) enable
 814 prediction of ρ_1^{sat} and P_v^{sat} over a range of T^{sat} for any ϵ , σ ,
 815 and λ by performing a small number of direct NVT sim-
 816 ulations with only a few reference parameter sets. The non-
 817 bonded energies and forces, which are required for MBAR
 818 and ITIC, are rapidly recomputed post-simulation with θ_{MCMC}
 819 for each reference configuration (for details on the so-called
 820 “basis functions,” see Ref. 55 and Sec. SI.IV of the sup-
 821 plementary material in Ref. 45). In total, this methodol-
 822 ogy reduces the computational cost for computing $L(\theta|D)$
 823 by several orders of magnitude compared to direct simu-
 824 lation of VLE, using Gibbs Ensemble Monte Carlo (GEMC)
 825 or Grand Canonical Monte Carlo histogram reweighting
 826 (GCMC-HR).

827 3. Uncertainty model

828 Quantifying the surrogate model variance (s_{SM}^2) is essen-
 829 tial for evaluating $L(\theta|D)$. While only a brief description is
 830 provided here, details are found in Sec. SI.IV of the supple-
 831 mentary material. Rather than performing a rigorous statistical
 832 assessment of MBAR, ITIC, and Eqs. (23) and (24), we use an
 833 empirical approach for estimating s_{SM}^2 . Specifically, we com-
 834 pute the deviation between the surrogate model estimates of
 835 ρ_1^{sat} and P_v^{sat} for TraPPE-UA and Potoff with those reported
 836 in the literature for the respective force fields obtained using
 837 GEMC¹² or GCMC-HR.¹⁸ Although this is a rough approxi-
 838 mation for estimating s_{SM}^2 , the benefit of this inter-laboratory

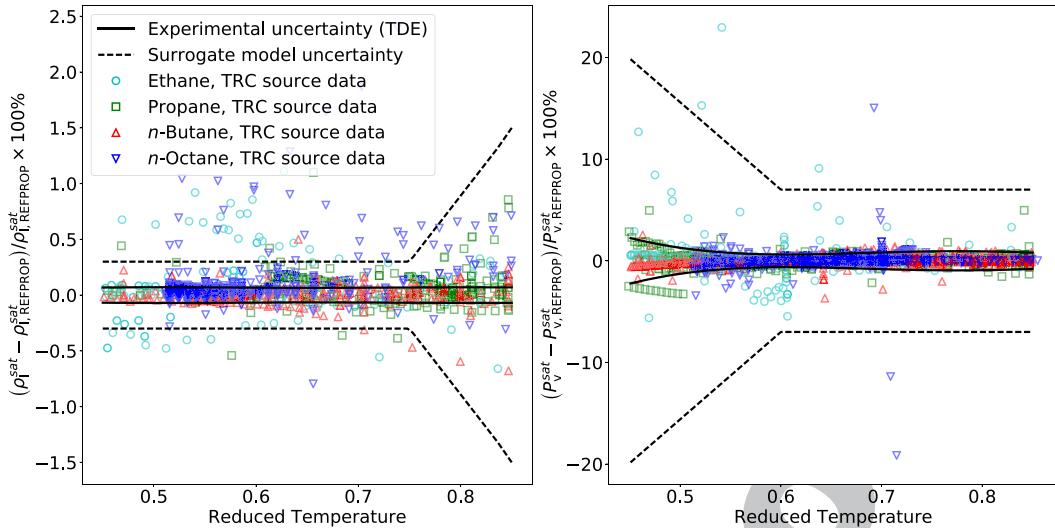


FIG. 6. Uncertainty assessment. Experimental (TDE) uncertainties are negligible compared to surrogate model uncertainties. Panels (a) and (b) plot the uncertainties for ρ_l^{sat} and P_v^{sat} with respect to reduced temperature (absolute temperature divided by the REFPROP T_c). Uncertainties are computed at the 95% confidence level using the respective s_D^2 and s_{SM}^2 values. Also included are percent deviations between the REFPROP values^{1,33–36} and the experimental data used in Eq. (18) to compute the likelihood for ethane, propane, *n*-butane, and *n*-octane (from the Thermodynamics Research Center, TRC, source data).⁵⁸

comparison is that s_{SM}^2 accounts for “dark uncertainty,”⁵⁶ i.e., uncertainties that arise from unknown sources which can lead to unresolvable discrepancies between research groups.⁵⁷ These non-statistical uncertainties are typically associated with different simulation packages, MD instead of MC, finite-size effects, and post-simulation analysis (e.g., ITIC rather than HR).

As shown in Fig. 6, the surrogate model uncertainty (u_{SM} , reported at the 95% confidence level) for ρ_l^{sat} is 0.3% up to $0.75T_c$ and increases linearly to 1.5% at the maximum T^{sat} . The surrogate model uncertainty for P_v^{sat} is 20% at the minimum T_{sat} and decreases linearly to 7% at $0.6T_c$, where it remains constant for higher temperatures. Note that these are conservative estimates of u_{SM} , where other studies suggest smaller uncertainties in MBAR and ITIC.^{44–46} In fact, for the compounds investigated in this study, these uncertainties are much larger than the experimental uncertainties (u_{D} , at the 95% confidence level)⁵⁸ and, therefore, the size of the parameter space sampled by MCMC depends almost entirely on u_{SM} . The use of a conservative u_{SM} model is intentional in this regard, namely, so that the θ_{MCMC} sampled points represent practically all of the feasible ϵ and σ parameter sets (for a given λ) optimized with ρ_l^{sat} and P_v^{sat} .

V. RESULTS

In this section, we use MCMC and the aforementioned surrogate models to determine the parameter uncertainty in CH₃ and CH₂ interaction sites of *n*-alkanes. As the simulation results of branched alkanes are significantly less accurate than those of *n*-alkanes for both VLE and high pressure properties (cf. Figs. 4 and 5), we do not investigate the uncertainties of CH and C interaction sites.

Since the common practice is to limit λ to integer values (see Sec. II B), we perform several independent MCMC runs using a single, fixed, integer value of λ . The Bayesian inference analysis for CH₃ and CH₂ 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 pair of two-dimensional MCMC runs by assuming that the CH₃ parameters from ethane are transferable to propane, *n*-butane, and *n*-octane.

A. Ethane

Figures 7–10 present the MCMC results for ethane with $13 \leq \lambda_{\text{CH}_3} \leq 18$. Figure 7 demonstrates that the feasible region of ϵ_{CH_3} depends strongly on λ_{CH_3} , namely, larger values of λ_{CH_3} require larger values of ϵ_{CH_3} . By contrast, we observe a much smaller shift toward larger values of σ_{CH_3} with increasing λ_{CH_3} . This observation is consistent with Ref. 18.

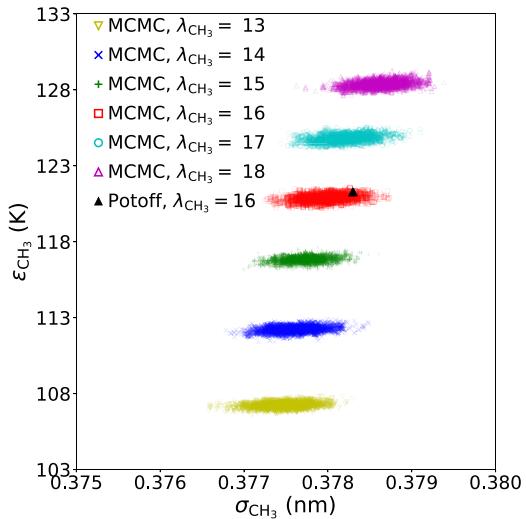


FIG. 7. MCMC sampled parameter sets for different values of λ_{CH_3} , $\epsilon_{\text{CH}_3,\text{MCMC}}$ and $\sigma_{\text{CH}_3,\text{MCMC}}$. ϵ_{CH_3} is strongly correlated with λ_{CH_3} and σ_{CH_3} , while σ_{CH_3} and λ_{CH_3} are only weakly correlated. The Potoff parameter set is included as a reference for $\lambda_{\text{CH}_3} = 16$.¹⁸

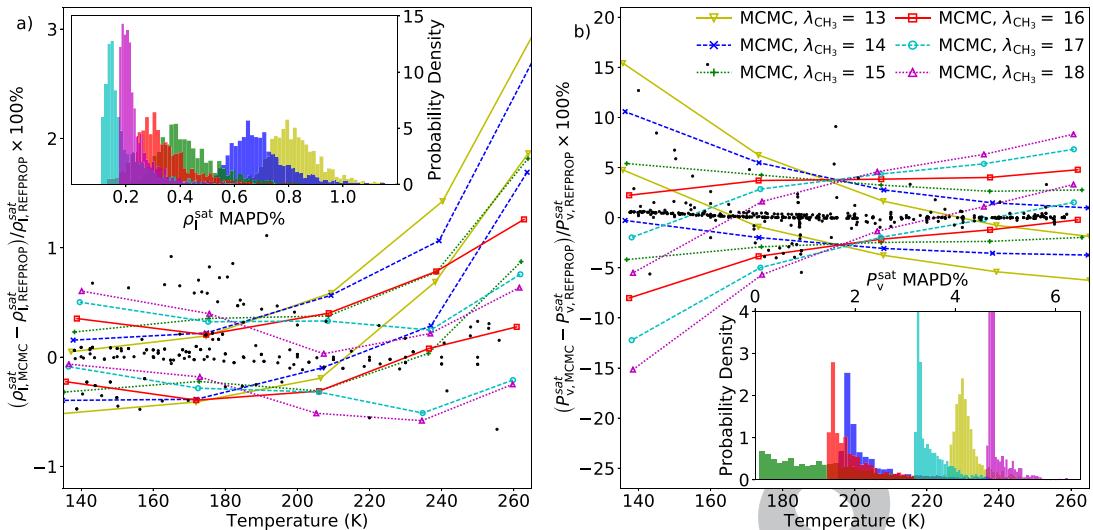


FIG. 8. Ethane VLE properties with respect to temperature for $\lambda_{\text{CH}_3} = 13$ to 18. Different trends are observed depending on the value of λ_{CH_3} . Panels (a) and (b) plot the percent deviation from REFPROP values^{1,33} for $\rho_l^{\text{sat},\text{MCMC}}$ and $P_v^{\text{sat},\text{MCMC}}$, respectively. Robust posterior prediction propagates the joint parameter uncertainties from ϵ_{CH_3} and σ_{CH_3} to ρ_l^{sat} and P_v^{sat} . The upper and lower lines for each λ correspond to the 95% credible interval obtained from QoI_{MCMC} . The insets of Panels (a) and (b) are histograms of the MAPD% in $\rho_l^{\text{sat},\text{MCMC}}$ and $P_v^{\text{sat},\text{MCMC}}$, respectively. Experimental data used to compute the likelihood are included as black dots.⁵⁸

Figures 8–9 compare the performance of different values of λ for ρ_l^{sat} , P_v^{sat} , and Z . Notice that the insets of Fig. 8 plot the mean absolute percent deviation (MAPD%) to quantify the goodness of fit to VLE data, while the inset of Fig. 9 plots the average deviation (AD%) to demonstrate the positive bias in P^{high} . Note also that because MAPD% and AD% are percent deviations they are not directly related to the squared deviations of the normal distribution used to compute the likelihood. We plot MAPD% and AD% as these are easier to conceptualize and quantify.

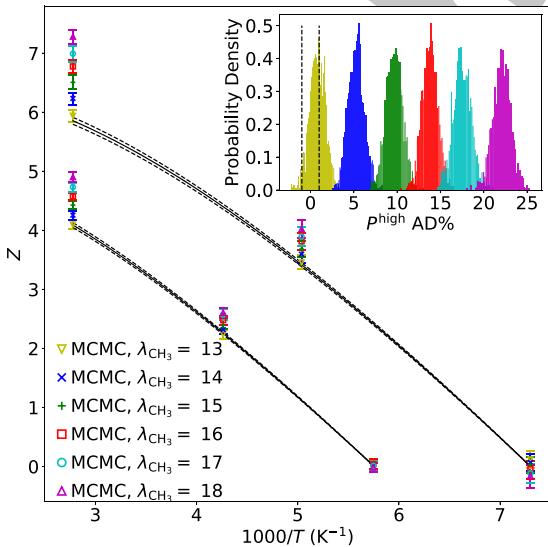


FIG. 9. Ethane isochore results for $\lambda_{\text{CH}_3} = 13$ to 18. Inadequacies of the UA Mie λ -potential are observed in Z for the two highest isochore densities (ρ_3^{IC} and ρ_4^{IC}) by propagating the joint uncertainties in ϵ_{CH_3} and σ_{CH_3} for different values of λ_{CH_3} . The inset plots the distribution of average deviation (AD%) in P^{high} , i.e., P_{MCMC} for $\rho = \rho_3^{\text{IC}}$ and ρ_4^{IC} at $T = T^{\text{IT}}$. REFPROP uncertainty in P^{high} is $\pm 1\%$.^{1,33}

Panel (a) of Fig. 8 with the corresponding inset demonstrates that the best prediction of ρ_l^{sat} is obtained for higher values of λ_{CH_3} . However, while the $\rho_l^{\text{sat}} \text{ MAPD\%}$ for $\lambda_{\text{CH}_3} = 15$ to 18 are similar, $\lambda_{\text{CH}_3} = 13$ and 14 have significantly higher $\rho_l^{\text{sat}} \text{ MAPD\%}$. Panel (b) of Fig. 8 demonstrates that $\lambda_{\text{CH}_3} \leq 14$ and $\lambda_{\text{CH}_3} \geq 17$ over- and under-predict P_v^{sat} at low temperatures, respectively, while $\lambda_{\text{CH}_3} = 15$ and 16 have the best trend for P_v^{sat} . The inset of Panel (b) shows that $\lambda_{\text{CH}_3} = 15$ has the lowest MAPD% in P_v^{sat} .

Finally, Fig. 9 demonstrates that all of the sampled $\epsilon_{\text{CH}_3,\text{MCMC}}$ and $\sigma_{\text{CH}_3,\text{MCMC}}$ parameter sets for $\lambda_{\text{CH}_3} \geq 14$ over-predict Z at high temperatures and densities (P^{high}). As expected, the larger the value of λ_{CH_3} , the more the force field over-predicts P^{high} .

While Figs. 8 and 9 plot the results for ρ_l^{sat} , P_v^{sat} , and Z individually, Fig. 10 helps to visualize the overall performance of different values of λ_{CH_3} for simultaneously predicting all three quantities of interest. In Panel (a), notice that the trade-off between the MAPD% of ρ_l^{sat} and P_v^{sat} . This compromise between two competing properties included in the objective function, namely, ρ_l^{sat} and P_v^{sat} , is known as a Pareto front.^{23,59,60} The optimal location for a Pareto front is the bottom left region of the plot (low MAPD% for both ρ_l^{sat} and P_v^{sat}), while the worst location is the top right region (high MAPD% for both ρ_l^{sat} and P_v^{sat}). Note that the inset of Panel (a) includes an approximate “overall” Pareto front that combines the results for all values of λ_{CH_3} . Although not depicted for visual clarity, the “L” shaped frontier for different colors/symbols demonstrates that each λ_{CH_3} value also has its own Pareto front. Because the overall Pareto front consists of points from the $15 \leq \lambda_{\text{CH}_3} \leq 17$ Pareto fronts, the Pareto optimal λ_{CH_3} value is either 15, 16, or 17, depending on the relative weight assigned to ρ_l^{sat} and P_v^{sat} . By contrast, since the $\lambda_{\text{CH}_3} = 13$, 14, and

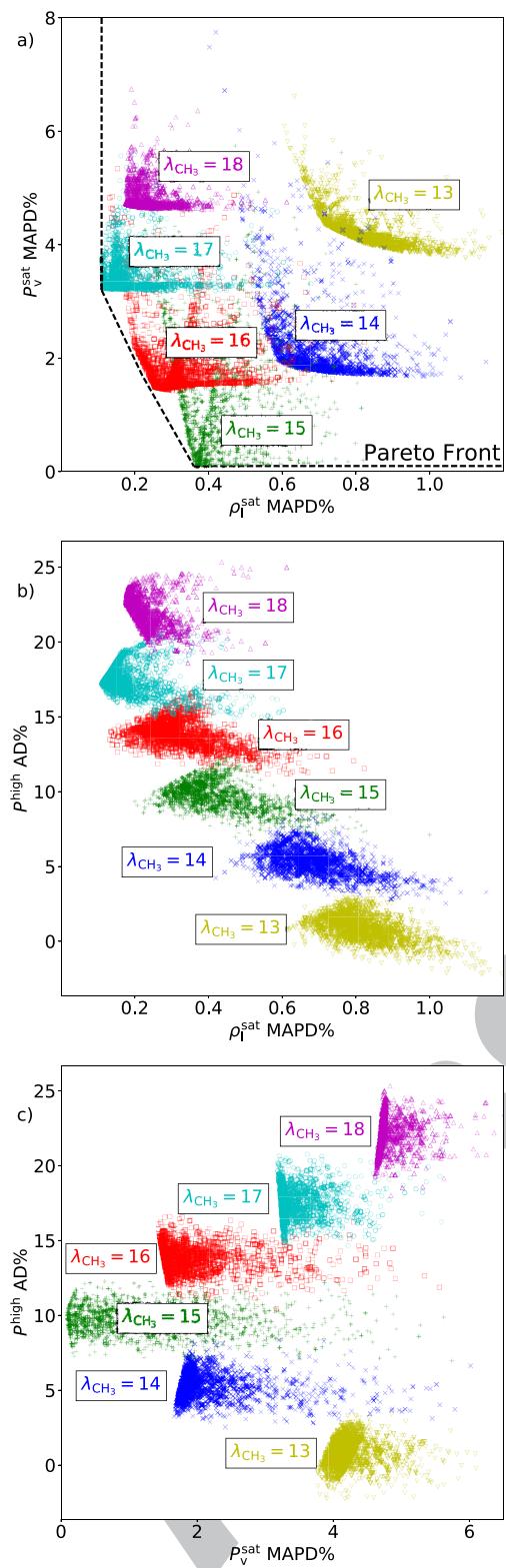


FIG. 10. Overall performance of $\lambda_{\text{CH}_3} = 13$ to 18 for ethane. MCMC confirms that accurate prediction of VLE leads to large deviations in pressures for supercritical fluids and compressed liquids. Panel (a) plots the mean absolute percent deviation (MAPD%) of $\rho_{l,\text{MCMC}}^{\text{sat}}$ and $P_{v,\text{MCMC}}^{\text{sat}}$. Panels (b) and (c) plot the average deviation (AD%) in P^{high} with respect to MAPD% of $\rho_{l,\text{MCMC}}^{\text{sat}}$ and $P_{v,\text{MCMC}}^{\text{sat}}$, respectively.

18 Pareto fronts are completely inside the overall Pareto front, these λ_{CH_3} values are not optimal, regardless of the weighting.

Finally, and most importantly for our purposes, Panels (b) and (c) of Fig. 10 demonstrate the increase in MAPD% for ρ_l^{sat} and P_v^{sat} that accompanies more accurate prediction of P^{high} . For example, although $\lambda_{\text{CH}_3} = 15, 16$, and 17 are the best values based on VLE data, they over-predict P^{high} by around 10%, 14%, and 18%, respectively. By contrast, while $\lambda_{\text{CH}_3} = 13$ is the most accurate for P^{high} , the MAPD% for ρ_l^{sat} and P_v^{sat} are 4 and 40 times larger than the respective minimum MAPD%. These results support the fundamental claim of this work, namely, that the UA Mie λ -6 potential cannot adequately predict both VLE and high pressures for supercritical fluids and compressed liquids.

Figure 11 provides statistical evidence for each value of λ by comparing their respective Bayes factors (normalized with respect to $\lambda = 14$) based solely on ρ_l^{sat} and P_v^{sat} . Bayes' factors from 1 to 3.2, 3.2 to 10, 10 to 32, 32 to 100, and greater than 100 are typically classified as “not substantial,” “substantial,” “strong,” “very strong,” and “decisive” evidence, respectively.⁶¹ Panel (a) shows that, with $\frac{3.6}{0.02} = 180$, there is “decisive” evidence against the use of $\lambda_{\text{CH}_3} = 13$ for predicting ρ_l^{sat} and P_v^{sat} . As $\lambda_{\text{CH}_3} = 13$ is the only value that predicts P^{high} within the REFPROP uncertainty, we conclude that no set of $\epsilon_{\text{CH}_3}, \sigma_{\text{CH}_3}$, and λ_{CH_3} can predict *both* VLE and P^{high} .

In addition, there is “very strong” evidence that the 18-6 potential is not justified by VLE data ($\frac{3.6}{0.1} = 36$). The evidence in favor of the 15-6 or 16-6 potentials over the 14-6 and 17-6 potentials is not as definitive, although it is still considered “substantial” ($\frac{3.6}{1.0}, \frac{3.6}{0.8}, \frac{3.5}{1.0}$, and $\frac{3.5}{0.8} > 3.5$). By contrast, the evidence for $\lambda_{\text{CH}_3} = 15$ instead of $\lambda_{\text{CH}_3} = 16$ is “not substantial” ($\frac{3.6}{3.5} \approx 1.03$).

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

Also, ITIC is limited to $T^{\text{sat}} < 0.85T_c$. 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 to 290 K). Based on the observed bias in ρ_l^{sat} at higher temperatures (240–260 K) for $\lambda_{\text{CH}_3} = 14$, it appears that higher temperature VLE data would strengthen the counter evidence against the 14-6 potential. It is unclear whether higher temperature data would support the 15-6 or 16-6 potential, although the optimal λ_{CH_3} is likely a non-integer value between 15 and 16. Implementing MBAR with GCMC may be necessary to include VLE data from $0.85 < T^{\text{sat}} < 0.95$.

B. Larger *n*-alkanes

The conclusions regarding the UA Mie λ -6 potential for ethane are generally applicable to larger *n*-alkanes. Specifically, we observe that improved accuracy in predicting VLE requires a larger value of λ_{CH_2} . However, this improvement

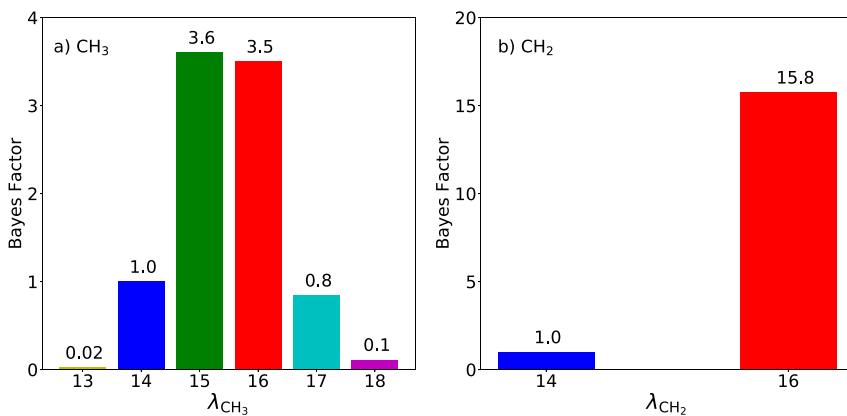


FIG. 11. Bayes' factor (evidence) for different values of λ_{CH_3} [Panel (a)] and λ_{CH_2} [Panel (b)]. Evidence supports $\lambda_{\text{CH}_3} = 15$ or 16 and $\lambda_{\text{CH}_2} = 16$ over $\lambda_{\text{CH}_2} = 14$. CH_3 values depend only on ethane, while CH_2 values are based on propane, *n*-butane, and *n*-octane. Note that all values are normalized with respect to $\lambda = 14$.

comes at the cost of significantly over-predicting P^{high} . Figure 12 presents the MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets with Panels (a) and (b) corresponding to $\lambda_{\text{CH}_2} = 16$ and $\lambda_{\text{CH}_2} = 14$, respectively. Note that these results were obtained using fixed values of ϵ_{CH_3} , σ_{CH_3} , and λ_{CH_3} , where $\lambda_{\text{CH}_3} = \lambda_{\text{CH}_2}$. The values of ϵ_{CH_3} and σ_{CH_3} are the maximum likelihood parameter set from ethane for the corresponding λ_{CH_3} value.

Notice that in Fig. 12 the MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets, for a given value of λ_{CH_2} , overlap considerably for propane, *n*-butane, and *n*-octane. These joint distributions provide statistical evidence in favor of the common assumption that CH_2 parameters are transferable between different *n*-alkanes. To further demonstrate this point, Fig. 12 includes the MCMC results when the posterior is based on the combined likelihoods from all three compounds, referred to as "MCMC transferable."

Panel (a) shows that the Potoff CH_2 parameter set is within the MCMC sample regions for $\lambda_{\text{CH}_2} = 16$. The same result was also observed for ethane (see Fig. 7). This suggests that the Potoff CH_3 and CH_2 parameters are supported by the VLE data used in this study, even though the Potoff force field was parameterized using VLE data in a higher temperature range ($0.6 < T_r^{\text{sat}} < 0.95$).

Also, note that the uncertainty in the parameters is largest for propane and smallest for *n*-octane. Therefore, the sensitivity of ρ_1^{sat} and P_v^{sat} , with respect to the CH_2 parameters, increases with an increasing number of CH_2 interaction sites. Although this result is fairly intuitive, it is a valuable insight when selecting a training set of molecules for force field development. For example, notice that the MCMC transferable region is almost identical to that of *n*-octane, which shows that propane and *n*-butane data provide relatively little additional information that is not contained in the *n*-octane data.

Most importantly, for the purposes of this study, the contours in Fig. 12 demonstrate that the MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets have a strong positive bias (i.e., large AD%) in P^{high} . Specifically, $\lambda_{\text{CH}_2} = 16$ and $\lambda_{\text{CH}_2} = 14$ have AD% of approximately 16%–21% and 10%–15%, respectively, much greater than the REFPROP uncertainty of around 1%. Furthermore, because the 0% contour is roughly parallel to the MCMC region and found at much lower ϵ_{CH_2} values, it is necessary to sacrifice considerable accuracy in ρ_1^{sat} and P_v^{sat} in order to accurately predict P^{high} . It is interesting that, for corresponding values of λ , the AD% for these larger *n*-alkanes is higher than that of ethane. This suggests that longer chain-lengths, with a UA Mie λ -6 force field, exacerbate the erroneous Z trend at high pressures.

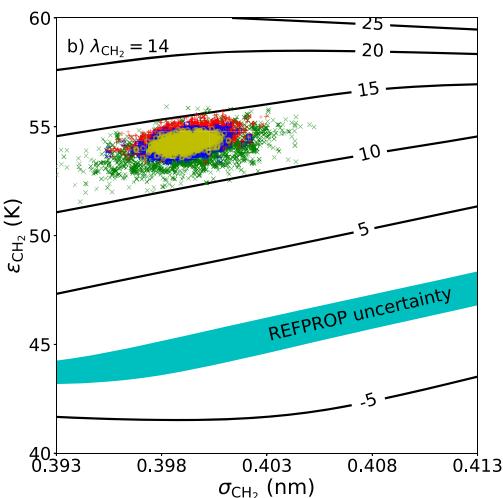
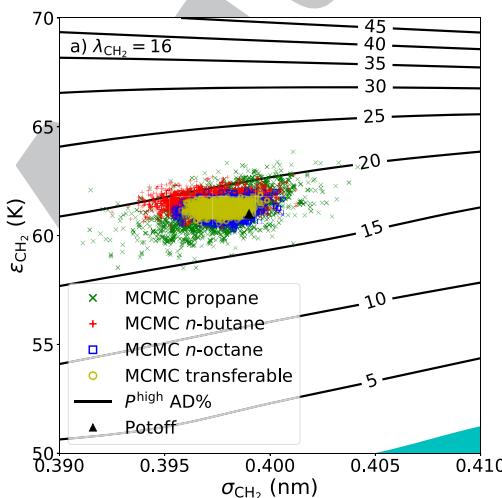


FIG. 12. High pressure performance of $\lambda_{\text{CH}_2} = 16$ [Panel (a)] and $\lambda_{\text{CH}_2} = 14$ [Panel (b)] for propane, *n*-butane, and *n*-octane. MCMC sampled ϵ_{CH_2} and σ_{CH_2} parameter sets result in large AD% for P^{high} . Contours are the AD% in P^{high} relative to the REFPROP values, where the "REFPROP uncertainty" region represents $\pm 1\%$ deviation.^{134–36} Panel (a) includes the Potoff parameter set as a reference for $\lambda_{\text{CH}_2} = 16$.¹⁸

Although the AD% in P^{high} is slightly lower for $\lambda_{\text{CH}_2} = 14$ than for $\lambda_{\text{CH}_2} = 16$, the UA Mie 14-6 potential is significantly less reliable for VLE. Figure 11 demonstrates that there is “strong” evidence for $\lambda_{\text{CH}_2} = 16$ over $\lambda_{\text{CH}_2} = 14$, based on VLE data. Note that the evidence in Fig. 11 for the λ value of CH₂ sites is stronger than that for the CH₃ sites. This suggests that the ethane ρ_1^{sat} and P_v^{sat} results are less sensitive to λ than the larger *n*-alkanes and/or that the ethane VLE data contain less information than the combined data of propane, *n*-butane, and *n*-octane. In conclusion, these results suggest that neither UA Mie 16-6 nor 14-6 force fields are capable of predicting VLE and $P\rho T$ for supercritical fluids and compressed liquids of *n*-alkanes.

VI. DISCUSSION

A. Recommendations

Although the UA Mie λ -6 potential is not quantitatively reliable at high pressures, it may still be of use for FEOS parameterization when considering the insight gained in this study. For example, since the Potoff force field consistently over-predicts high pressures, a non-linear FEOS optimization could utilize the simulation results as an upper constraint for the FEOS pressure.³ Furthermore, the primary purpose to include molecular simulation data for FEOS development is to increase the range of validity by ensuring good behavior of the FEOS at high temperatures and pressures. As FEOSs are based on empirical equations with 50 to 100 fitting parameters, even an inaccurate force field has a more sound theoretical basis. Therefore, the UA Mie λ -6 simulation output for a given property should not demonstrate non-physical oscillations, inflection points, derivative sign-changes, etc., which can plague a poorly-fit FEOS.

Essentially, whether or not a FEOS should be developed using a hybrid data set consisting of UA Mie λ -6 simulation results depends on the quality and quantity of available experimental data. If the data cover a wide range of state points and properties, it is possible that the UA Mie λ -6 potential may still be useful, despite the systematic deviations at high pressures. By contrast, if the experimental data are limited such that the FEOS depends almost entirely on the molecular simulation results, the UA Mie λ -6 force field will lead to large deviations at high pressures. Therefore, in this scenario, we advise against the use of UA Mie λ -6 force fields when developing a FEOS for normal and branched alkanes. For this purpose, we recommend further investigation of alternative potentials with a softer repulsive barrier and a more sound theoretical basis, e.g., Buckingham exponential-6, modified-Morse,^{28–30} or an extended Lennard-Jones.^{44,62,63}

B. Limitations

There are some caveats to the primary conclusion from this study that UA Mie λ -6 force fields parameterized with VLE data should not be used to develop fundamental equations of state for normal and branched alkanes. The main limitation is that the poor extrapolation at high pressures is based solely on the trend of Z with respect to inverse temperature. By contrast, the simulation values that are typically included in hybrid

data sets used to generate FEOSs are derivatives of the departure (or residual) Helmholtz free energy with respect to the inverse temperature and/or density,^{7–11}

$$A_{xy}^{\text{dep}} R_g T \equiv (1/T)^x \rho^y \frac{\partial^{x+y} A^{\text{dep}}}{\partial(1/T)^x \partial \rho^y}, \quad (25)$$

where x and y are the integer values and the derivative is of order $x + y$. There are advantages of using A_{xy}^{dep} for developing FEOSs, as this approach eliminates redundant information found in traditional macroscopic properties.^{2,4–7} For example, the following expressions demonstrate the interdependency of the properties we computed, namely, Z and U^{dep} with their derivatives along isochores and isotherms,⁶

$$Z = 1 + A_{01}^{\text{dep}}, \quad (26)$$

$$\frac{1}{T} \left(\frac{-\partial Z}{\partial(1/T)} \right)_\rho = 1 + A_{01}^{\text{dep}} - A_{11}^{\text{dep}}, \quad (27)$$

$$\frac{U^{\text{dep}}}{R_g T} = A_{10}^{\text{dep}}, \quad (28)$$

$$\frac{1}{R_g} \left(\frac{\partial U^{\text{dep}}}{\partial T} \right)_\rho = -A_{20}^{\text{dep}}, \quad (29)$$

$$\rho \left(\frac{\partial Z}{\partial \rho} \right)_T = 1 + 2A_{01}^{\text{dep}} + A_{02}^{\text{dep}}. \quad (30)$$

However, at least two reasons exist as for why the conclusions of this study are not based on A_{xy}^{dep} . First, with the exception of *ms2*,⁶⁴ we are not aware of any open-source simulation package that readily provides A_{xy}^{dep} . As *ms2* is currently limited to small rigid molecules, it is not amenable to the systems of interest. In addition, macroscopic properties, such as Z and U^{dep} (with their respective derivatives), are more readily understood and visualized than A_{xy}^{dep} . It is also easier to quantify the impact of Z and U^{dep} on process design than A_{xy}^{dep} . For example, as demonstrated in Ref. 8, an inaccurate prediction of some A_{xy}^{dep} does not necessarily result in poor prediction of $P\rho T$ behavior or heat capacities.

Although we do not perform a detailed investigation of A_{xy}^{dep} , we have indirectly investigated each of the A_{xy}^{dep} terms in Eqs. (26)–(30). For example, Secs. III and V present Z and, by inspection, the slope of Z with respect to $1/T$ at constant ρ . Since these properties are equivalent to Eqs. (26) and (27), respectively, Secs. III and V indirectly focus on two of the Helmholtz derivatives, namely, A_{01}^{dep} and A_{11}^{dep} . Section SI.V of the supplementary material also demonstrates 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 Eqs. (28)–(30). Although these additional properties provide insight regarding A_{10}^{dep} , A_{20}^{dep} , A_{01}^{dep} , and A_{02}^{dep} , unfortunately, the results are less conclusive due to the relatively large uncertainties in the REFPROP values, ca. 5% and 10% for U^{dep} and $\left(\frac{\partial U^{\text{dep}}}{R_g \partial T} \right)_\rho$, respectively. Furthermore, the relationship between Eqs. (28)–(30) and the repulsive barrier, λ , is not obvious from these results. 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.

Another potential limitation is that we utilize a single layer Bayes model as opposed to a hierarchical model, where the posterior is proportional to multiple priors that depend on the parameters from different levels of the hierarchy (for a more detailed discussion, see Refs. 31 and 51). Wu *et al.* demonstrated the need for hierarchical models when the data set, D , contains discrepancies, i.e., internal inconsistencies. However, since we use a conservative estimate for the surrogate model uncertainty, i.e., $u_{\text{SM}} \gg u_D$, any discrepancies in the VLE data should not affect the parameter uncertainties. A hierarchical approach is also useful when accounting for model inadequacies, i.e., when the force field is not capable of representing multiple data types. A hierarchical method should thus be favored if determining the parameter uncertainty when simultaneously considering ρ_1^{sat} , P_v^{sat} and P^{high} . Furthermore, a hierarchical model should be used if the parameters are not transferable between molecules, e.g., the Potoff CH and C parameters for “short” and “long” branched alkanes. However, such a hierarchical approach is unnecessary for our purposes since the transferable UA Mie λ -6 force field for *n*-alkanes is capable of reproducing ρ_1^{sat} and P_v^{sat} , which are the only properties included in D .

VII. CONCLUSIONS

Recently, molecular simulation results at extreme temperatures and pressures have supplemented experimental data when developing fundamental equations of state for compounds with limited experimental data. For this hybrid data set approach to be useful, it is imperative that the force field be reliable and transferable over different $P\rho T$ conditions. Unfortunately, literature united-atom force fields that are highly accurate for estimating VLE properties of normal and branched alkanes have systematic deviations in Z at non-VLE conditions. Bayesian inference suggests that the UA Mie λ -6 model type is not adequate for simultaneously predicting ρ_1^{sat} , P_v^{sat} , and P^{high} . In the case of ethane, evidence from VLE data supports $\lambda = 15$ or 16, while Z at high pressures requires $\lambda = 13$. A similar trend is observed for larger *n*-alkanes. Specifically, evidence from VLE data supports $\lambda = 16$, while we observe only slight improvement in Z at high pressures for $\lambda = 14$. Therefore, while considerable improvement in VLE is observed for the Mie λ -6 potential over the traditional Lennard-Jones 12-6, the use of $\lambda > 12$ does not appear to have physical/theoretical justification but, rather, is simply an empirical remedy that performs well for VLE. For these reasons, we recommend that alternative force fields be considered for developing FEOSSs of normal and branched alkanes, which utilize anisotropic-united-atom/all-atom models and/or more physically realistic non-bonded potentials.

SUPPLEMENTARY MATERIAL

See supplementary material for ■.



ACKNOWLEDGMENTS

We are appreciative for the assistance provided by Eric W. Lemmon regarding the reliability and uncertainties of the REFPROP correlations at high pressures. We would also like

to acknowledge the Bayesian expertise provided by Joshua H. Fass of the Chodera lab at the Memorial Sloan Kettering Cancer Center. We are grateful for the internal review provided by Ian H. Bell and Daniel W. Siderius from the National Institute of Standards and Technology (NIST).

This research was performed while Richard A. Messerly held a National Research Council (NRC) Postdoctoral Research Associateship at NIST. Michael R. Shirts acknowledges support from the National Science Foundation (NSF) Grant No. CHE-1738975.

- ¹E. W. Lemmon, M. L. Huber, and M. O. McLinden, *NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP*, Version 9.1 (National Institute of Standards and Technology, 2013).
- ²M. Thol, G. Rutkai, A. Köster, R. Lustig, R. Span, and J. Vrabec, “Equation of state for the Lennard-Jones fluid,” *J. Phys. Chem. Ref. Data* **45**(2), 023101 (2016).
- ³M. Thol, G. Rutkai, R. Span, J. Vrabec, and R. Lustig, “Equation of state for the Lennard-Jones truncated and shifted model fluid,” *Int. J. Thermophys.* **36**(1), 25–43 (2015).
- ⁴G. Rutkai, M. Thol, R. Span, and J. Vrabec, “How well does the Lennard-Jones potential represent the thermodynamic properties of noble gases?,” *Mol. Phys.* **115**(9–12), 1104–1121 (2017).
- ⁵R. Lustig, G. Rutkai, and J. Vrabec, “Thermodynamic correlation of molecular simulation data,” *Mol. Phys.* **113**(9–10), 910–931 (2015).
- ⁶G. Rutkai and J. Vrabec, “Empirical fundamental equation of state for phosgene based on molecular simulation data,” *J. Chem. Eng. Data* **60**(10), 2895–2905 (2015).
- ⁷G. Rutkai, M. Thol, R. Lustig, R. Span, and J. Vrabec, “Communication: Fundamental equation of state correlation with hybrid data sets,” *J. Chem. Phys.* **139**(4), 041102 (2013).
- ⁸M. Thol, F. H. Dubberke, G. Rutkai, T. Windmann, A. Köster, R. Span, and J. Vrabec, “Fundamental equation of state correlation for hexamethyldisiloxane based on experimental and molecular simulation data,” in *Special Issue covering the Nineteenth Symposium on Thermophysical Properties [Fluid Phase Equilib.* **418**, 133–151 (2016).
- ⁹M. Thol, G. Rutkai, A. Köster, F. H. Dubberke, T. Windmann, R. Span, and J. Vrabec, “Thermodynamic properties of octamethylcyclotetrasiloxane,” *J. Chem. Eng. Data* **61**(7), 2580–2595 (2016).
- ¹⁰M. Thol, G. Rutkai, A. Köster, S. Miroshnichenko, W. Wagner, J. Vrabec, and R. Span, “Equation of state for 1,2-dichloroethane based on a hybrid data set,” *Mol. Phys.* **115**(9–12), 1166–1185 (2017).
- ¹¹M. Thol, G. Rutkai, A. Köster, M. Kortmann, R. Span, and J. Vrabec, “Fundamental equation of state for ethylene oxide based on a hybrid dataset,” *Chem. Eng. Sci.* **121**, 87–99 (2015). 2013 Danckwerts Special Issue on Molecular Modelling in Chemical Engineering.
- ¹²M. G. Martin and J. I. Siepmann, “Transferable potentials for phase equilibria. I. United-atom description of n-alkanes,” *J. Phys. Chem. B* **102**(14), 2569–2577 (1998).
- ¹³M. G. Martin and J. Ilja Siepmann, “Novel configurational-bias Monte Carlo method for branched molecules. Transferable potentials for phase equilibria. II. United-atom description of branched alkanes,” *J. Phys. Chem. B* **103**(21), 4508–4517 (1999).
- ¹⁴M. S. Shah, J. Ilja Siepmann, and M. Tsapatsis, “Transferable potentials for phase equilibria. Improved united-atom description of ethane and ethylene,” *AIChE J.* **63**(11), 5098–5110 (2017).
- ¹⁵J. R. Errington and A. Z. Panagiotopoulos, “A new intermolecular potential model for the n-alkane homologous series,” *J. Phys. Chem. B* **103**(30), 6314–6322 (1999).
- ¹⁶P. Ungerer, C. Beauvais, J. Delhommelle, A. Boutin, B. Rousseau, and A. H. Fuchs, “Optimization of the anisotropic united atoms intermolecular potential for n-alkanes,” *J. Chem. Phys.* **112**(12), 5499–5510 (2000).
- ¹⁷E. Bourasseau, P. Ungerer, A. Boutin, and A. H. Fuchs, “Monte Carlo simulation of branched alkanes and long chain n-alkanes with anisotropic united atoms intermolecular potential,” *Mol. Simul.* **28**(4), 317–336 (2002).
- ¹⁸J. J. Potoff and D. A. Bernard-Brunel, “Mie potentials for phase equilibria calculations: Applications to alkanes and perfluoroalkanes,” *J. Phys. Chem. B* **113**(44), 14725–14731 (2009).
- ¹⁹J. R. Mick, M. Soroush Barhaghi, B. Jackman, L. Schwiebert, and J. J. Potoff, “Optimized Mie potentials for phase equilibria: Application to branched alkanes,” *J. Chem. Eng. Data* **62**(6), 1806–1818 (2017).

- 1306 20 A. Hemmen and J. Gross, "Transferable anisotropic united-atom force field
1307 based on the Mie potential for phase equilibrium calculations: n-alkanes
1308 and n-olefins," *J. Phys. Chem. B* **119**(35), 11695–11707 (2015).
- 1309 21 D. Weidler and J. Gross, "Transferable anisotropic united-atom force
1310 field based on the Mie potential for phase equilibria: Aldehydes, ketones,
1311 and small cyclic alkanes," *Ind. Eng. Chem. Res.* **55**(46), 12123–12132
1312 (2016).
- 1313 22 C. Herdes, T. S. Totton, and E. A. Müller, "Coarse grained force field for
1314 the molecular simulation of natural gases and condensates," *Fluid Phase
1315 Equilib.* **406**, 91–100 (2015).
- 1316 23 K. Stöbener, P. Klein, M. Horsch, K. Kufer, and H. Hasse, "Parametrization
1317 of two-center Lennard-Jones plus point-quadrupole force field models
1318 by multicriteria optimization," *Fluid Phase Equilib.* **411**, 33–42
1319 (2016).
- 1320 24 R. A. Messerly, T. A. Knotts IV, and W. Vincent Wilding, "Uncertainty
1321 quantification and propagation of errors of the Lennard-Jones 12–6
1322 parameters for n-alkanes," *J. Chem. Phys.* **146**(19), 194110 (2017).
- 1323 25 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon
1324 Press; Oxford University Press, Oxford, England, New York, 1987).
- 1325 26 P. A. Gordon, "Development of intermolecular potentials for predicting
1326 transport properties of hydrocarbons," *J. Chem. Phys.* **125**(1), 014504
1327 (2006).
- 1328 27 G. Galliéro, C. Boned, A. Baylaucq, and F. Montel, "Molecular dynamics
1329 comparative study of Lennard-Jones α -6 and exponential α -6 potentials:
1330 Application to real simple fluids (viscosity and pressure)," *Phys. Rev. E* **73**,
1331 061201 (2006).
- 1332 28 R. L. Rowley and T. Pakkanen, "Determination of a methane intermolecular
1333 potential model for use in molecular simulations from *ab initio*
1334 calculations," *J. Chem. Phys.* **110**(7), 3368–3377 (1999).
- 1335 29 R. L. Rowley, Y. Yang, and T. A. Pakkanen, "Determination of an ethane
1336 intermolecular potential model for use in molecular simulations from
1337 *ab initio* calculations," *J. Chem. Phys.* **114**(14), 6058–6067 (2001).
- 1338 30 J. M. Hayes, J. C. Greer, and D. A. Morton-Blake, "A force-field description
1339 of short-range repulsions for high density alkane molecular dynamics
1340 simulations," *J. Comput. Chem.* **25**(16), 1953–1966 (2004).
- 1341 31 L. Kulakova, G. Arampatzis, P. Angelikopoulos, P. Hadjoudoukas,
1342 C. Papadimitriou, and P. Koumoutsakos, "Data driven inference for the
1343 repulsive exponent of the Lennard-Jones potential in molecular dynamics
1344 simulations," *Sci. Rep.* **7**(1), 16576 (2017).
- 1345 32 J. R. Mick, M. Soroush Barhaghi, B. Jackman, K. Rushaidat, L. Schwiebert,
1346 and J. J. Potoff, "Optimized Mie potentials for phase equilibria: Application
1347 to noble gases and their mixtures with n-alkanes," *J. Chem. Phys.* **143**(11),
1348 114504 (2015).
- 1349 33 D. Bücker and W. Wagner, "A reference equation of state for the thermody-
1350 namic properties of ethane for temperatures from the melting line to 675 K
1351 and pressures up to 900 MPa," *J. Phys. Chem. Ref. Data* **35**(1), 205–266
1352 (2006).
- 1353 34 E. W. Lemmon, M. O. McLinden, and W. Wagner, "Thermodynamic prop-
1354 erties of propane. III. A reference equation of state for temperatures from
1355 the melting line to 650 K and pressures up to 1000 MPa," *J. Chem. Eng.
1356 Data* **54**(12), 3141–3180 (2009).
- 1357 35 D. Bücker and W. Wagner, "Reference equations of state for the thermo-
1358 dynamic properties of fluid phase n-butane and isobutane," *J. Phys. Chem.
1359 Ref. Data* **35**(2), 929–1019 (2006).
- 1360 36 R. Beckmueller, M. Thol, and R. Span, "Fundamental equation of state for
1361 n-octane," in *International Journal of Thermophysics* (Pending Publication,
1362 2018).
- 1363 37 E. W. Lemmon and R. Span, "Short fundamental equations of state for 20
1364 industrial fluids," *J. Chem. Eng. Data* **51**(3), 785–850 (2006).
- 1365 38 T. M. Blackham, A. K. Lemmon, and E. W. Lemmon, "Fundamental equa-
1366 tion of state for isoocetane," in *International Journal of Thermophysics*
1367 (Pending Publication, 2018).
- 1368 39 M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and GRO-
1369 MACS Development Team, GROMACS User Manual version 2018,
1370 www.gromacs.org, 2018.
- 1371 40 W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A com-
1372 puter simulation method for the calculation of equilibrium constants for
1373 the formation of physical clusters of molecules: Application to small water
1374 clusters," *J. Chem. Phys.* **76**(1), 637–649 (1982).
- 1375 41 W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distribu-
1376 tions," *Phys. Rev. A* **31**, 1695–1697 (1985).
- 1377 42 B. Hess, H. Bekker, J. Herman, C. Berendsen, and J. G. E. M. Fraaije,
1378 "LINCS: A linear constraint solver for molecular simulations," *J. Comput.
1379 Chem.* **18**(12), 1463–1472 (1998).
- 1380 43 B. Hess, "P-LINCS: A parallel linear constraint solver for molecular
1381 simulation," *J. Chem. Theory Comput.* **4**(1), 116–122 (2008).
- 1382 44 S. Mostafa Razavi, "Optimization of a transferable shifted force field for
1383 interfaces and inhomogenous fluids using thermodynamic integration,"
1384 M.S. thesis, The University of Akron, 2016.
- 1385 45 R. A. Messerly, S. Mostafa Razavi, and M. R. Shirts, "Configuration-
1386 sampling-based surrogate models for rapid parameterization of non-bonded
1387 interactions," *J. Chem. Theory Comput.* **14**(6), 3144–3162 (2018).
- 1388 46 S. Mostafa Razavi, R. A. Messerly, and J. Richard Elliott, *Coexistence
1389 Calculation Using the Isothermal-isochoric Integration Method* (Pending
1390 Publication, 2018).
- 1391 47 C. Nieto-Draghi, A. Bocahut, B. Creton, P. Have, A. Ghoufi, A. Wender,
1392 A. Boutin, B. Rousseau, and L. Normand, "Optimisation of the dynam-
1393 ical behaviour of the anisotropic united atom model of branched alkanes:
1394 Application to the molecular simulation of fuel gasoline," *Mol. Simul.* **34**(2),
1395 211–230 (2008).
- 1396 48 F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adal-
1397 steinsson, and O. M. Knio, "Uncertainty quantification in MD simulations.
1398 Part II: Bayesian inference of force-field parameters," *Multiscale Model.
1399 Simul.* **10**(4), 1460–1492 (2012).
- 1400 49 P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, "Bayesian
1401 uncertainty quantification and propagation in molecular dynamics simula-
1402 tions: A high performance computing framework," *J. Chem. Phys.* **137**(14),
1403 144103 (2012).
- 1404 50 F. Cailliez and P. Pascal, "Statistical approaches to forcefield calibration
1405 and prediction uncertainty in molecular simulation," *J. Chem. Phys.* **134**(5),
1406 054124 (2011).
- 1407 51 S. Wu, P. Angelikopoulos, C. Papadimitriou, R. Moser, and P. Koumoutsakos,
1408 "A hierarchical Bayesian framework for force field selection in
1409 molecular dynamics simulations," *Philos. Trans. R. Soc., A* **374**(2060)
1410 20150032 (2016).
- 1411 52 M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of sam-
1412 ples from multiple equilibrium states," *J. Chem. Phys.* **129**, 124105
1413 (2008).
- 1414 53 A. Vetere, "The Riedel equation," *Ind. Eng. Chem. Res.* **30**(11), 2487–2492
1415 (1991).
- 1416 54 M. Funke, R. Kleinrahm, and W. Wagner, "Measurement and correlation
1417 of the (P, ρ , T) relation of ethane II. Saturated-liquid and saturated-vapour
1418 densities and vapour pressures along the entire coexistence curve," *J. Chem.
1419 Thermodyn.* **34**(12), 2017–2039 (2002).
- 1420 55 L. N. Naden and M. R. Shirts, "Rapid computation of thermodynamic
1421 properties over multidimensional nonbonded parameter spaces using adap-
1422 tive multistate reweighting," *J. Chem. Theory Comput.* **12**(4), 1806–1823
1423 (2016).
- 1424 56 P. N. Patrone, A. Dienstfrey, A. R. Browning, S. Tucker, and S. Christ-
1425 tensen, "Uncertainty quantification in molecular dynamics studies of the
1426 glass transition temperature," *Polymer* **87**, 246–259 (2016).
- 1427 57 M. Schappals, A. Mecklenfeld, L. Kröger, V. Botan, A. Köster, S. Stephan,
1428 E. J. García, G. Rutkai, G. Raabe, P. Klein, K. Leonhard, C. W. Glass,
1429 J. Lenhard, J. Vrabec, and H. Hasse, "Round robin study: Molecular sim-
1430 ulation of thermodynamic properties from models with internal degrees of
1431 freedom," *J. Chem. Theory Comput.* **13**(9), 4270–4280 (2017).
- 1432 58 M. Frenkel, R. D. Chirico, V. Diky, X. Yan, Q. Dong, and C. Muzny,
1433 "Thermodata engine (TDE): Software implementation of the dynamic data
1434 evaluation concept," *J. Chem. Inf. Model.* **45**(4), 816–838 (2005).
- 1435 59 K. Stöbener, P. Klein, S. Reiser, M. Horsch, K.-H. Kufer, and H. Hasse,
1436 "Multicriteria optimization of molecular force fields by Pareto approach,"
1437 *Fluid Phase Equilib.* **373**, 100–108 (2014).
- 1438 60 S. Werth, K. Stöbener, P. Klein, K.-H. Kufer, M. Horsch, and H. Hasse,
1439 "Molecular modelling and simulation of the surface tension of real
1440 quadrupolar fluids," *Chem. Eng. Sci.* **121**, 110–117 (2015). 2013 Danckwerts
1441 Special Issue on Molecular Modelling in Chemical Engineering.
- 1442 61 H. Jeffreys, *The Theory of Probability*, 3rd ed. (Clarendon Press, Oxford,
1443 2004).
- 1444 62 P. G. Hajigeorgiou, "The extended Lennard-Jones potential energy function:
1445 A simpler model for direct-potential-fit analysis," *J. Mol. Spectrosc.* **330**,
1446 4–13 (2016). Potentiology and Spectroscopy in Honor of Robert Le Roy.
- 1447 63 F. Kalos and A. E. Grosser, "Intermolecular potentials from differential
1448 cross sections: Ar + Ar," *Can. J. Chem.* **50**(6), 892–896 (1972).
- 1449 64 G. Rutkai, A. Köster, G. Guevara-Carrion, T. Janzen, M. Schappals,
1450 C. W. Glass, M. Bernreuther, A. Wafai, S. Stephan, M. Kohns, S. Reiser,
1451 S. Deublein, M. Horsch, H. Hasse, and J. Vrabec, "ms2: A molecular sim-
1452 ulation tool for thermodynamic properties, release 3.0," *Comput. Phys.
1453 Commun.* **221**, 343–351 (2017).