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

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

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

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

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

SI.I Simulation Set-Up

This section is provided to improve the reproducibility of the results presented in this study.

SI.I.1 State Points

Tables SI.I, SI.II, SI.III, SI.IV, SI.V, SI.VI, SI.VII, and SI.VIII contain the state points that were simulated for ethane, propane, *n*-butane, *n*-octane, isobutane, isohexane, isooctane, and neopentane, respectively. The first ten state points of each table correspond to five isochores while the last nine points are for the supercritical isotherm. The number of state points, the specified reduced temperatures, and the spacing between neighboring densities were recommended by the developers of the ITIC approach (J. Richard Elliott and Seyed Mostafa Razavi). It has been demonstrated that these points are sufficient for accurate calculation of $\rho_{\mathrm{l}}^{\mathrm{sat}}$, $\rho_{\mathrm{v}}^{\mathrm{sat}}$, and $P_{\mathrm{v}}^{\mathrm{sat}}$.[1–3] Note that the temperatures (T_{sim}), box lengths (L_{box}), and number of molecules (N_{M}) are the exact values used in simulation while the density (ρ) is approximate (rounded) since it is calculated from L_{box}, N_{M}, and the molecular weight.

SI.I.2 GROMACS Input Files

We have provided example input files for simulating isooctane at 653.0 K with the TraPPE-UA force field in GROMACS (see attached .gro, .top, and .mdp files).

Table SI.I: State points simulated for ethane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
137.0	3.21680	400	600.01
198.5	3.21680	400	600.01
174.0	3.29730	400	557.13
234.6	3.29730	400	557.13
207.0	3.38640	400	514.30
262.9	3.38640	400	514.30
236.0	3.48610	400	471.42
285.1	3.48610	400	471.42
260.0	3.59860	400	428.58
301.9	3.59860	400	428.58
360.0	6.15360	400	85.712
360.0	4.88410	400	171.43
360.0	4.26660	400	257.15
360.0	3.87650	400	342.85
360.0	3.59860	400	428.58
360.0	3.48610	400	471.42
360.0	3.38640	400	514.30
360.0	3.29730	400	557.13
360.0	3.21680	400	600.01

Table SI.II: State points simulated for propane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
166	3.55643	400	651.13
242	3.55643	400	651.13
210	3.64538	400	604.62
285	3.64538	400	604.62
250	3.74395	400	558.11
320	3.74395	400	558.11
285	3.85413	400	511.60
347	3.85413	400	511.60
314	3.97854	400	465.09
368	3.97854	400	465.09
444	6.80321	400	93.019
444	5.39971	400	186.04
444	4.71708	400	279.06
444	4.28575	400	372.08
444	3.97854	400	465.09
444	3.85413	400	511.60
444	3.74395	400	558.11
444	3.64538	400	604.62
444	3.55643	400	651.13

Table SI.III: State points simulated for *n*-butane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
191	3.83864	400	682.53
278	3.83864	400	682.53
241	3.93465	400	633.78
327	3.93465	400	633.78
287	4.04104	400	585.03
367	4.04104	400	585.03
328	4.15997	400	536.28
399	4.15997	400	536.28
361	4.29425	400	487.52
423	4.29425	400	487.52
510	7.34306	400	97.50
510	5.82819	400	195.01
510	5.09140	400	292.51
510	4.62584	400	390.02
510	4.29425	400	487.52
510	4.15997	400	536.28
510	4.04104	400	585.03
510	3.93465	400	633.78
510	3.83864	400	682.53

Table SI.IV: State points simulated for *n*-octane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
285.92	5.98449	800	708.01
387.29	5.98449	800	708.01
347.68	6.13416	800	657.44
440.25	6.13416	800	657.44
404.46	6.30003	800	606.87
483.20	6.30003	800	606.87
451.48	6.48542	800	556.30
515.25	6.48542	800	556.30
490.78	6.69481	800	505.72
539.92	6.69481	800	505.72
600.00	11.44803	800	101.14
600.00	9.08616	800	202.29
600.00	7.93753	800	303.43
600.00	7.21175	800	404.58
600.00	6.69481	800	505.72
600.00	6.48542	800	556.30
600.00	6.30003	800	606.87
600.00	6.13416	800	657.44
600.00	5.98449	800	708.01

Table SI.V: State points simulated for isobutane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
184	4.85814	800	673.40
267	4.85814	800	673.40
232	4.97964	800	625.30
315	4.97964	800	625.30
276	5.11429	800	577.20
353	5.11429	800	577.20
315	5.26480	800	529.10
383	5.26480	800	529.10
347	5.43475	800	481.00
406	5.43475	800	481.00
489	9.29328	800	96.20
489	7.37608	800	192.40
489	6.44360	800	288.60
489	5.85440	800	384.80
489	5.43475	800	481.00
489	5.26480	800	529.10
489	5.11429	800	577.20
489	4.97964	800	625.30
489	4.85814	800	673.40

Table SI.VI: State points simulated for isohexane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
224	5.43297	800	713.86
326	5.43297	800	713.86
282	5.56885	800	662.87
383	5.56885	800	662.87
337	5.71943	800	611.88
431	5.71943	800	611.88
384	5.88774	800	560.89
467	5.88774	800	560.89
423	6.07780	800	509.90
495	6.07780	800	509.90
597	10.39289	800	101.98
597	8.24884	800	203.96
597	7.20603	800	305.94
597	6.54711	800	407.92
597	6.07780	800	509.90
597	5.88774	800	560.89
597	5.71943	800	611.88
597	5.56885	800	662.87
597	5.43297	800	713.86

Table SI.VII: State points simulated for isooctane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
245	5.92132	800	730.91
356	5.92132	800	730.91
309	6.06941	800	678.71
419	6.06941	800	678.71
369	6.23353	800	626.50
472	6.23353	800	626.50
421	6.41697	800	574.29
512	6.41697	800	574.29
464	6.62411	800	522.08
543	6.62411	800	522.08
653	11.32707	800	104.42
653	8.99031	800	208.83
653	7.85376	800	313.25
653	7.13561	800	417.66
653	6.62411	800	522.08
653	6.41697	800	574.29
653	6.23353	800	626.50
653	6.06941	800	678.71
653	5.92132	800	730.91

Table SI.VIII: State points simulated for neopentane.

T_{sim} (K)	L_{box} (nm)	N_{M} (molecules)	$\rho\left(\frac{\mathrm{kg}}{\mathrm{m}^3}\right)$
257	5.34568	800	627.43
344	5.34568	800	627.43
300	5.47938	800	582.61
380	5.47938	800	582.61
337	5.62754	800	537.79
409	5.62754	800	537.79
368	5.79315	800	492.98
431	5.79315	800	492.98
393	5.98015	800	448.16
448	5.98015	800	448.16
520	10.22592	800	89.63
520	8.11632	800	179.26
520	7.09026	800	268.90
520	6.44193	800	358.53
520	5.98015	800	448.16
520	5.79315	800	492.98
520	5.62754	800	537.79
520	5.47938	800	582.61
520	5.34568	800	627.43

SI.II Potoff Generalized

Figure SI.1 provides the simulation results for the Potoff generalized force field for branched alkanes. Compare to Figure 5 in the main text.

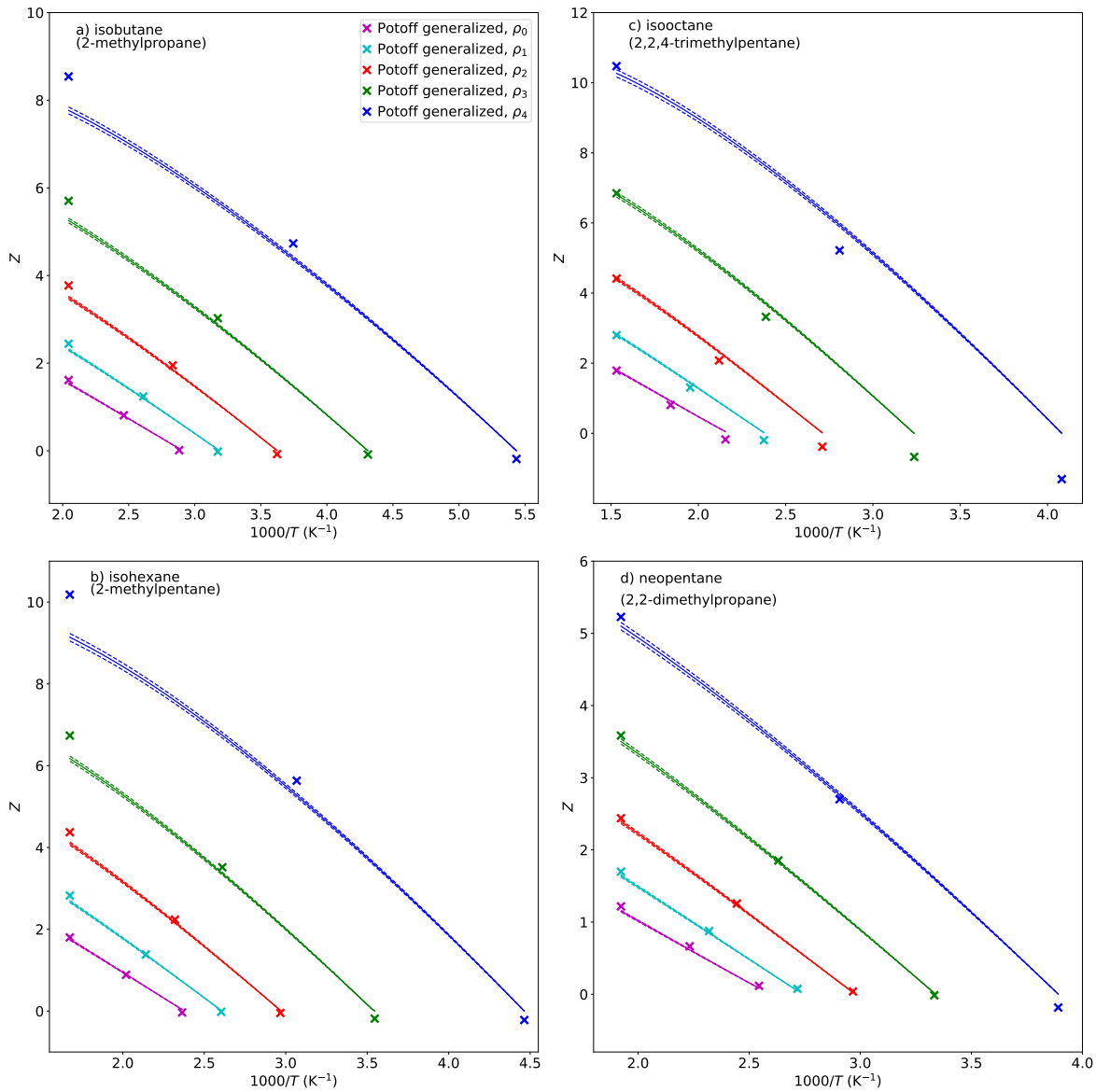

Figure SI.1: Compressibility factors (Z) along isochores for branched alkanes deviate strongly at higher pressures using the Potoff generalized force field. Panels a) to d) correspond to isobutane, isohexane, isooctane, and neopentane, respectively. Symbols, lines, uncertainties, and formatting are the same as those in Figure 4.

SI.III MCMC Example

In this section we present typical results from an MCMC run. All of the MCMC runs performed in this study used 10000 MCMC steps for the burn-in period and an additional 10000 MCMC steps for the production period. The proposal distribution variances (s_ϵ^2 and s_σ^2) were tuned every 100 MCMC steps during the burn-in period. Specifically, if the acceptance ratio was less than 20 %, the variances were decreased by a factor of $0.9^2 = 0.81$. If the acceptance ratio was greater than 50 %, the variances were increased by a factor of 1.1^2. During the production period, every 20 MCMC steps are stored to account for auto-correlation. Thus, the posterior predictive results for $\rho_{\mathrm{l,MCMC}}^{\mathrm{sat}}$, $P_{\mathrm{v,MCMC}}^{\mathrm{sat}}$, and Z_{MCMC} are based on 500 ϵ_{MCMC} and σ_{MCMC} parameter sets.

Figures SI.2-SI.3 provide an example of the MCMC results for ethane with $\lambda_{\mathrm{CH_3}} = 16$. Figure SI.2 Panels a) to c) demonstrate that a burn-in period of 10000 steps is sufficient for an ergodic-like sampling of σ, ϵ and $\log(Pr)$, respectively. Panels d) and e) show that the respective distributions for σ and ϵ appear to be normal. Note that the histograms of Panels d) to f) only include production period samples. The acceptance rate for proposed moves in ϵ and σ was 28.8 %.

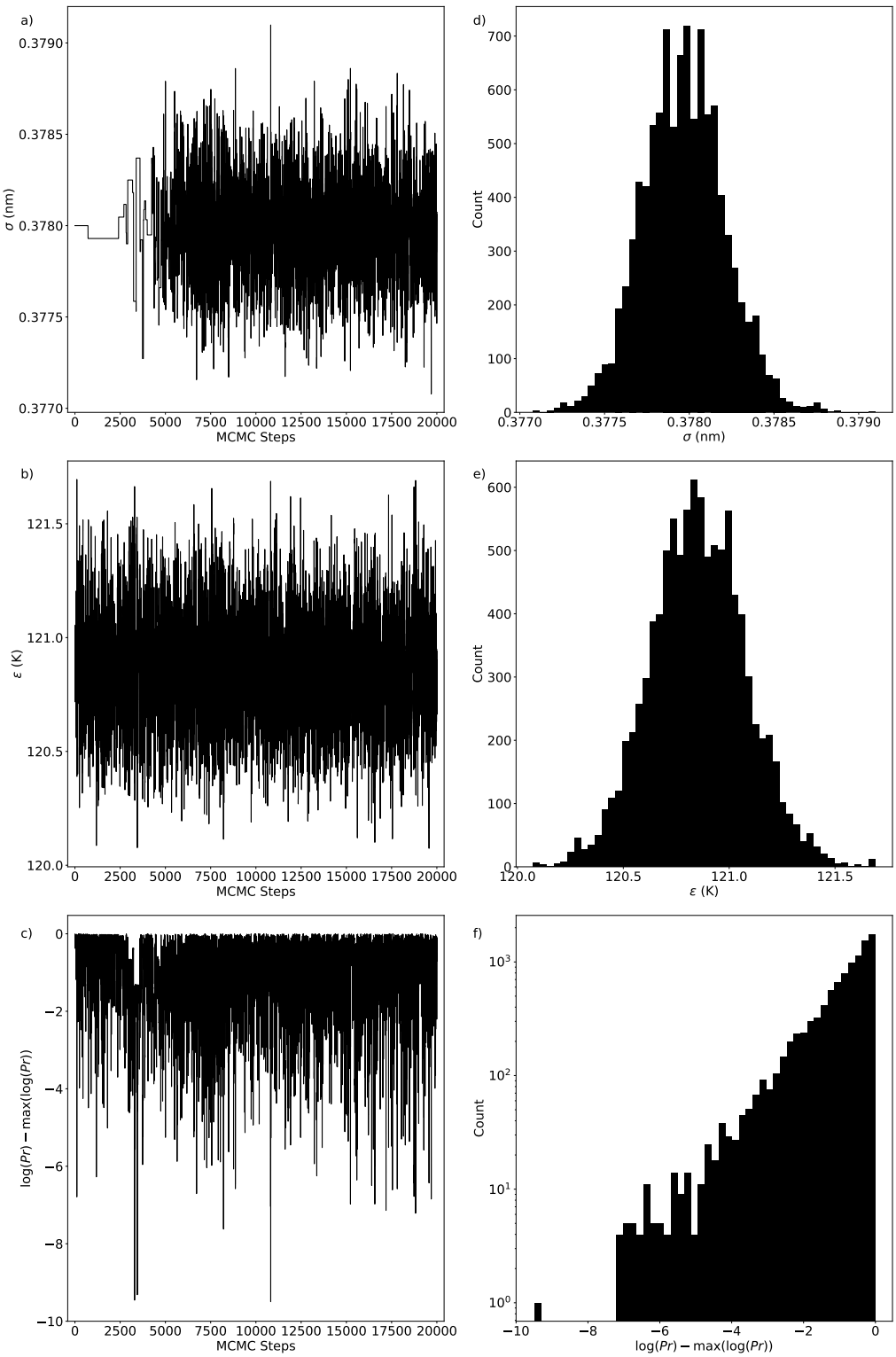

Figure SI.2: Panels a) to c) plot the respective traces of σ, ϵ and $\log(Pr)$, where Pr is the posterior. Panels d) to f) plot histograms of the production period samples for σ, ϵ and $\log(Pr)$, respectively. Note that $\log(Pr)$ is normalized by the maximum $\log(Pr)$.

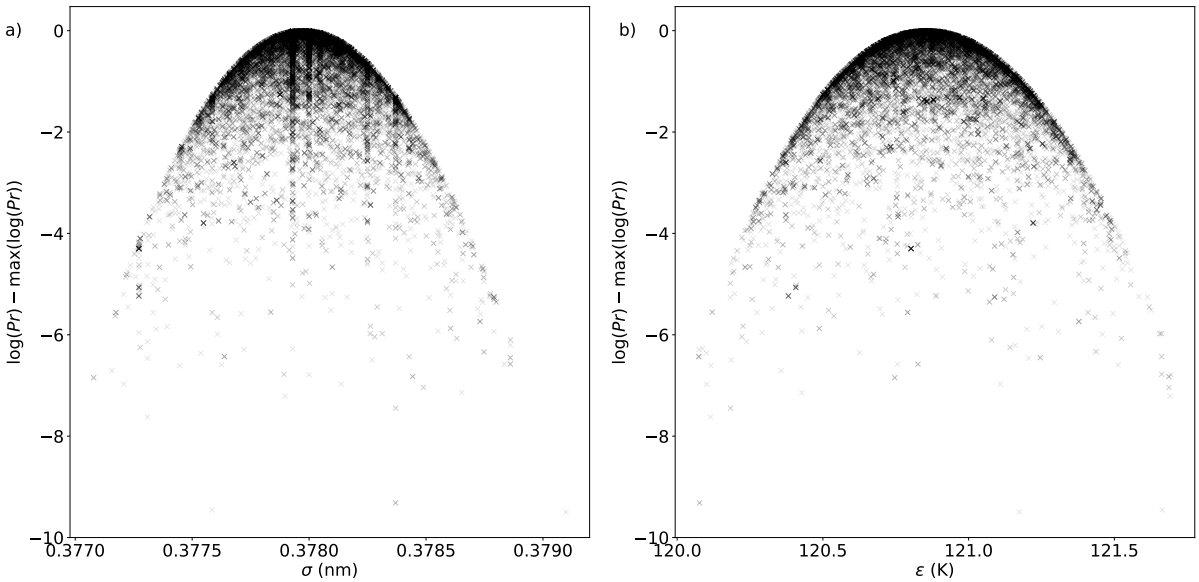

Figure SI.3: Panels a) and b) plot the dependence of $\log(Pr)$ (normalized by the maximum $\log(Pr)$, where Pr is the posterior) with respect to σ and ϵ.

SI.IV Error Model

Figure SI.4 presents the percent deviation between the surrogate model (SM) and literature (lit) VLE values using the TraPPE-UA and Potoff force fields for ethane, propane, *n*-butane, and *n*-octane. From this figure, an approximate surrogate model uncertainty was deduced such that most of the points, with their corresponding error bars, would lie within this uncertainty region. Note that the majority of literature values are for $T_r > 0.6$, therefore, we used our best judgment when extrapolating the surrogate model uncertainties to $T_r = 0.45$.

There are two likely reasons to explain why the percent deviations in $\rho_{\mathrm{l}}^{\mathrm{sat}}$ increase with respect to temperature. First, ITIC neglects the higher order virial terms that become more important at higher temperatures. Second, we use the REFPROP B_2 and B_3 values, rather than those of the force field. By contrast, the larger percent deviations in $P_{\mathrm{v}}^{\mathrm{sat}}$ with decreasing temperature are attributed primarily to the difference in orders of magnitude

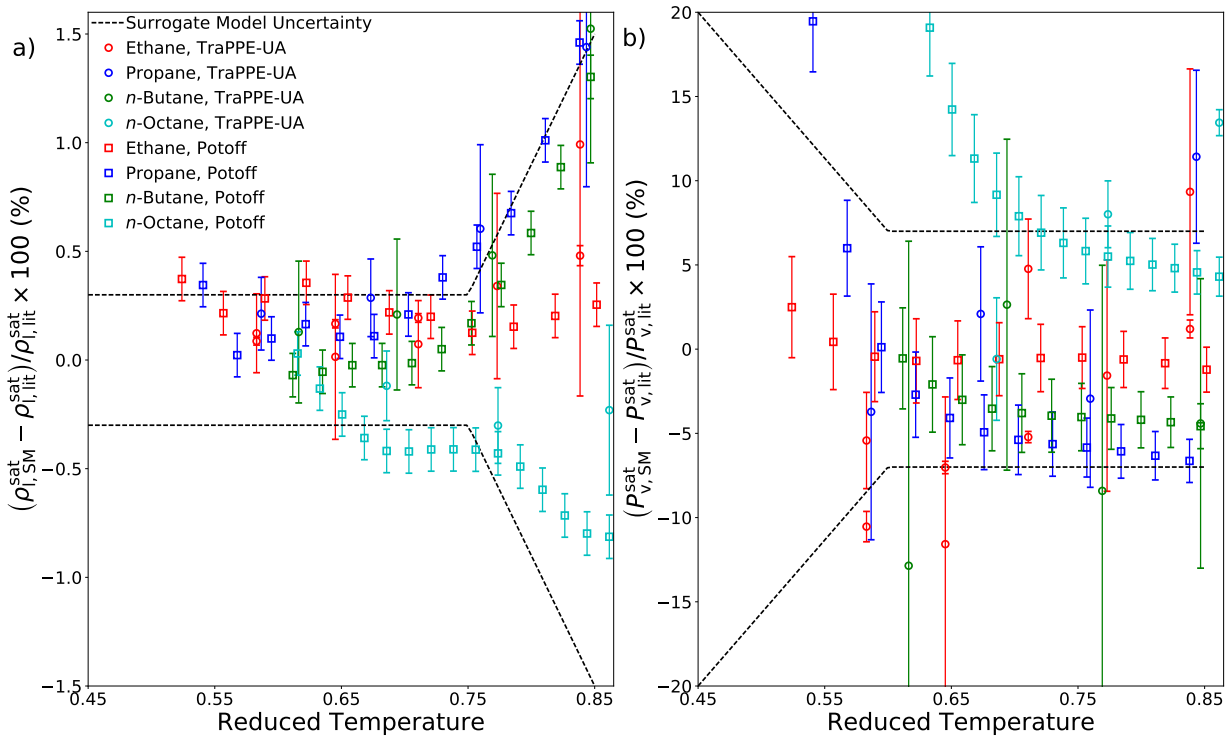

Figure SI.4: Panels a) and b) demonstrate how well the surrogate model uncertainty agrees with the percent deviations between the surrogate model and literature values for $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and $P_{\mathrm{v}}^{\mathrm{sat}}$, respectively. Reduced temperature is relative to the REFPROP T_c. The TraPPE-UA and Potoff literature values were obtained using GEMC[4,5] and GCMC,[6] respectively. Error bars represent statistical uncertainties that were obtained using replicate simulations. The error bars for TraPPE-UA were obtained from References 4,5, or alternatively on the TraPPE website. No attempt was made to determine if these are standard deviations or 95 % confidence intervals. The error bars for the Potoff model are a constant 0.1 % for $\rho_{\mathrm{l}}^{\mathrm{sat}}$ and 0.5 % to 3 % for $P_{\mathrm{v}}^{\mathrm{sat}}$, increasing linearly with decreasing reduced temperature. The values 0.5 % to 3 % were reported in Reference 6. As Reference 6 did not report uncertainties for $\rho_{\mathrm{l}}^{\mathrm{sat}}$, the value of 0.1 % is based on some of Potoff's more recent work.[7]

for $P_{\mathrm{v}}^{\mathrm{sat}}$, since the virial coefficients should not impact $P_{\mathrm{v}}^{\mathrm{sat}}$ at lower temperatures.

Aside from these expected trends, some other unexplained deviation trends are observed for a specific property, compound, and/or force field. However, it is not readily obvious if these deviations are caused by limitations in the surrogate model or the literature values themselves. For example, the TraPPE-UA and Potoff literature values were obtained using GEMC[4,5] and GCMC,[6] respectively. These Monte Carlo methods can suffer from low insertion acceptance at $T_r < 0.7$, which can result in erroneous VLE values, for $P_{\mathrm{v}}^{\mathrm{sat}}$ in particular. It is not our intention to imply that our values are better than either the TraPPE-UA or Potoff literature values, but rather we wish to emphasize that it should be expected that our approach would not yield exactly the same values as other methods. It is for this precise reason that our surrogate model uncertainty is larger than the statistical uncertainties which are typically reported in the literature.

SI.V Additional Properties

Figure SI.5 compares the MCMC results of different λ values for Z along the supercritical isotherms of the *n*-alkanes studied. The purpose of this plot is to demonstrate deviations in the slope of Z with respect to density for a constant temperature, i.e. $\left(\frac{\partial Z}{\partial \rho}\right)_T$, which is related to A_{01}^{dep} and A_{02}^{dep} (see Equation 29). The deviations at high densities along the supercritical isotherm are already explained in the main text. For this reason, the insets focus on the low density region. Inset in Panel a) shows that the 13-6 potential under-predicts Z at low density while it agrees well at high density. Recall that for ethane the 13-6 potential is very accurate at predicting Z and even the slope of Z along the isochores, i.e. $\left(\frac{-\partial Z}{\partial (1/T)}\right)_\rho$. However, the 13-6 potential does not appear to have the correct slope for Z along the isotherm, i.e. $Z\left(\frac{\partial Z}{\partial \rho}\right)_T$. The same trend is observed for the 14-6 potential in all four *n*-alkanes, namely, the improvement for predicting Z at high densities typicaly results in under-prediction of Z at low densities.

Figure SI.6 compare the MCMC results of different λ values for U^{dep} along the isochores of the *n*-alkanes studied. The purpose of this plot is to demonstrate deviations in U^{dep} and the slope of U^{dep} with respect to temperature for a constant density, i.e. $\left(\frac{\partial U^{\text{dep}}}{\partial T}\right)_\rho$, which are related to A_{10}^{dep} and A_{20}^{dep}, respectively (see Equations 27-28). Definitive conclusions are difficult due to the relatively large uncertainty in the REFPROP correlations, approximated to be 5 %. However, Panel a) demonstrates that at T^{sat}, i.e. the lowest temperature for a given isochore, $\lambda = 15$ accurately predicts U^{dep} while $\lambda > 15$ and $\lambda < 15$ consistently under- and over-predict U^{dep} at T^{sat}. By contrast, $\lambda = 13$ agrees more closely with U^{dep} at T^{IT}, i.e. the highest temperature, while $\lambda > 13$ under-predict U^{dep} at T^{IT}. A similar trend is observed for propane, *n*-butane, and *n*-octane, namely, $\lambda = 16$ and $\lambda = 14$ are more accurate at T^{sat} and T^{IT}, respectively. These results suggest that all values of λ are inaccurate for $\left(\frac{\partial U^{\text{dep}}}{\partial T}\right)_\rho$ and are only reliable for U^{dep} at certain conditions.

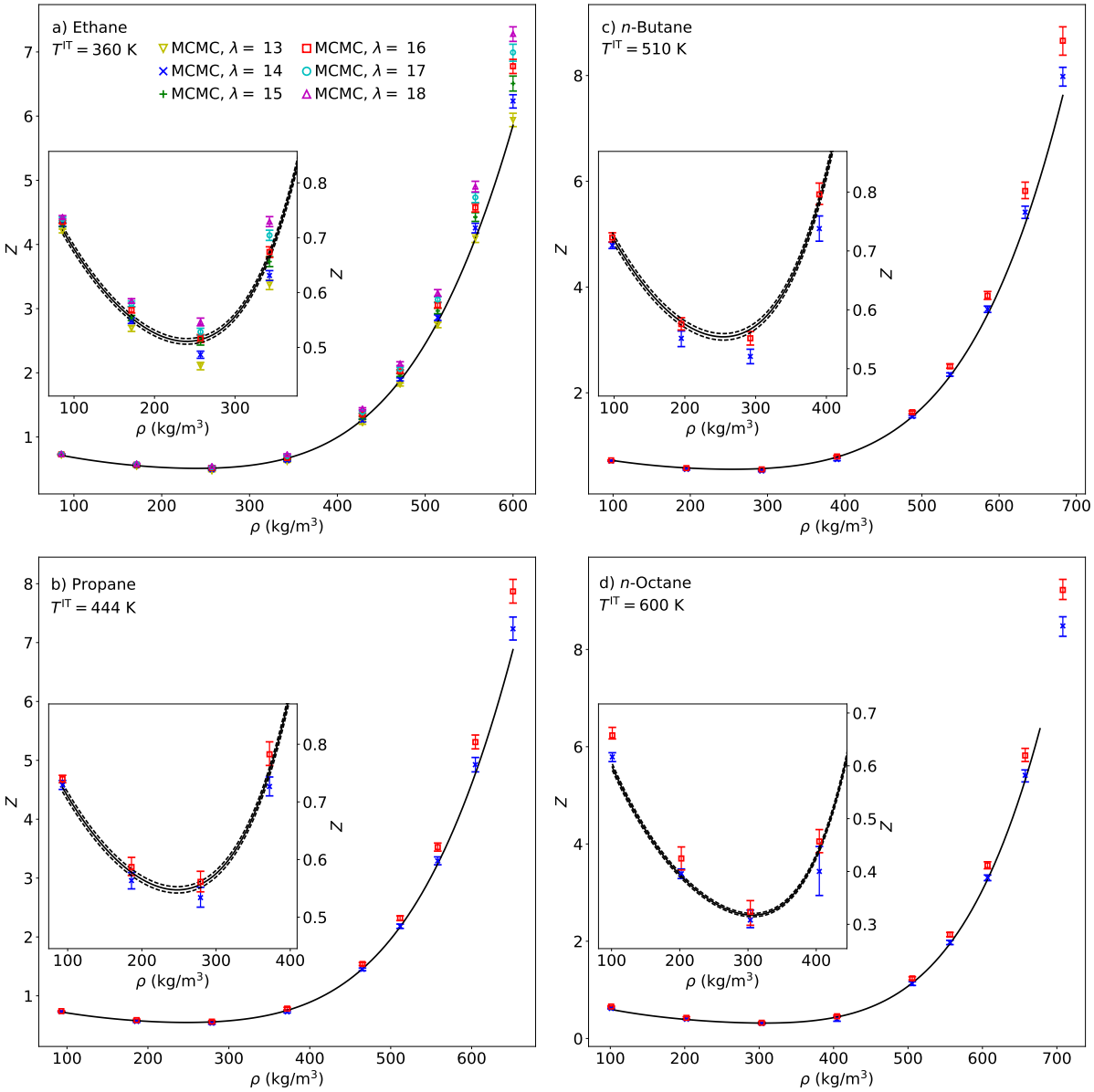

Figure SI.5: Z_{MCMC} along the supercritical isotherms for different λ values. Panels a) to d) correspond to ethane, propane, *n*-butane, and *n*-octane, respectively. Solid lines are the REFPROP correlation with dashed lines (included only in the insets) representing a 1 % uncertainty.

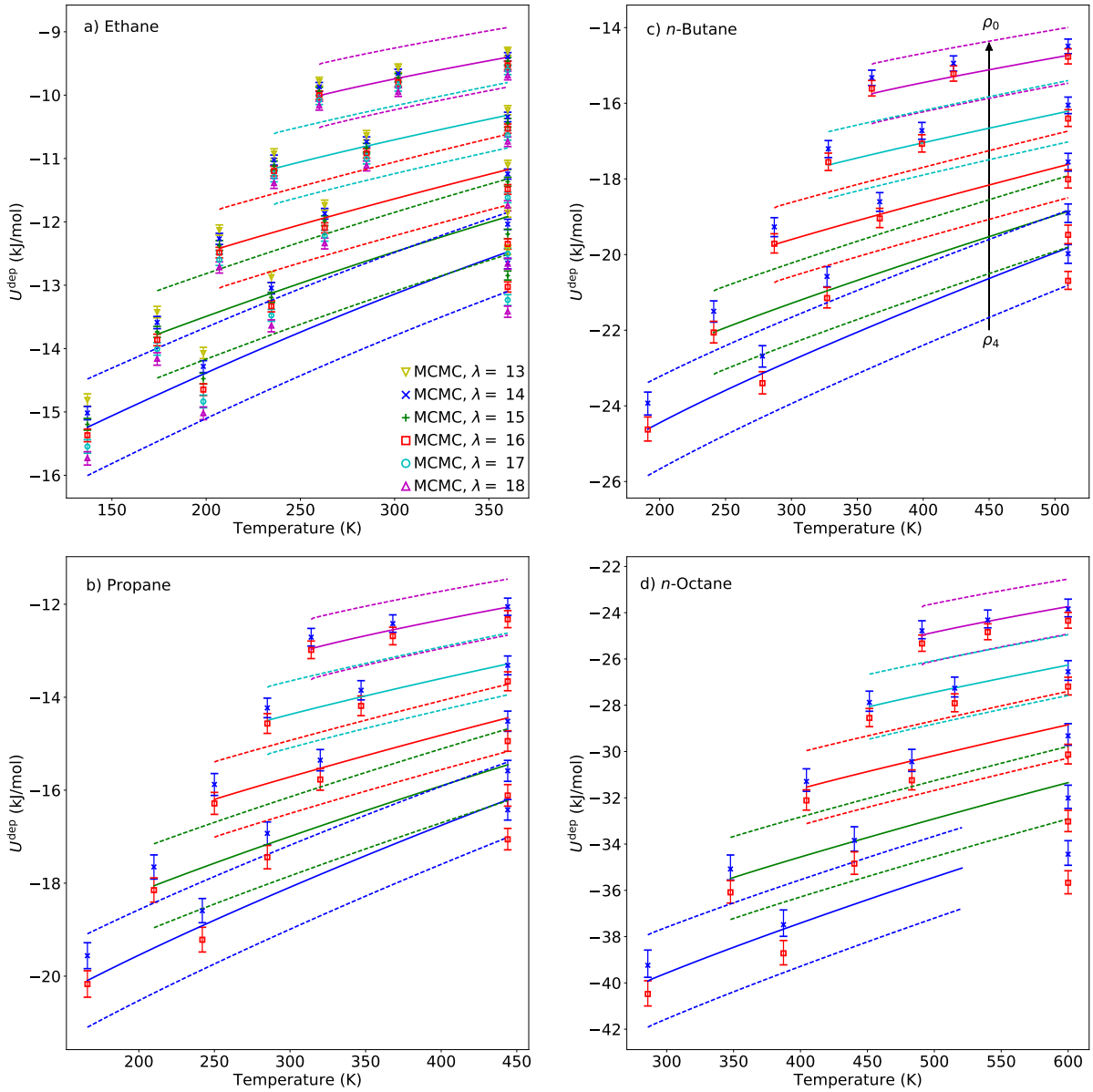

Figure SI.6: $U^{\mathrm{dep}}_{\mathrm{MCMC}}$ along the isochores for different λ values. Panels a) to d) correspond to ethane, propane, *n*-butane, and *n*-octane, respectively. Solid lines are the REFPROP correlation with dashed lines representing a 5 % uncertainty. Line colors of different isochores correspond with Figures 4-5 in main text.

References

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

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

(3) Razavi, S. M.; Messerly, R. A.; Elliott, J. R. **2018**, *Pending publication.*

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

(5) Eggimann, B. et al. T-UA No. 2 ethane. TraPPE Validation Database, University of Minnesota: Minneaoplis, MN. http://www.chem.umn.edu/groups/siepmann/trappe/, http://www.chem.umn.edu/groups/siepmann/trappe/ (accessed 2015 June 11).

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

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