

Supporting Information

Partial molar properties from molecular simulation using multiple linear regression

Tyler R. Josephson^a, Ramanish Singh^{a,b}, Mona S. Minkara^a, Evgenii Fetisov^{a,c}, and J. Ilja Siepmann^{a,b}

^a Department of Chemistry and Chemical Theory Center, ^b Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, USA ^c Chemical Physics and Analysis, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

ARTICLE HISTORY

Compiled August 20, 2019

1. Extended derivation

1.1. Transfer properties

Consider a closed system with two phases in equilibrium. The total internal energy is given by:

$$U = U^{\text{II}} + U^{\text{I}} \quad (1)$$

For a multicomponent system, these total internal energies can be expressed in terms of their partial molar contributions:

$$U = \sum_i N_i^{\text{II}} \bar{U}_i^{\text{II}} + \sum_i N_i^{\text{I}} \bar{U}_i^{\text{I}} \quad (2)$$

Now, the system will change from an initial state 0 to a final state f , during which some amount of species j is transferred between the two phases.

$$\Delta U_{\text{tr},j} = U^f - U^0 = U^{\text{II},f} - U^{\text{II},0} + U^{\text{I},f} - U^{\text{I},0} \quad (3)$$

$$\Delta U_{\text{tr},j} = \sum_i N_i^{\text{II},f} \bar{U}_i^{\text{II},f} + \sum_i N_i^{\text{I},f} \bar{U}_i^{\text{I},f} - \sum_i N_i^{\text{II},0} \bar{U}_i^{\text{II},0} - \sum_i N_i^{\text{I},0} \bar{U}_i^{\text{I},0} \quad (4)$$

Separating out the contribution from species j gives:

$$\Delta U_{\text{tr},j} = N_j^{\text{II},f} \bar{U}_j^{\text{II},f} + N_j^{\text{I},f} \bar{U}_j^{\text{I},f} - N_j^{\text{II},0} \bar{U}_j^{\text{II},0} - N_j^{\text{I},0} \bar{U}_j^{\text{I},0} + \sum_{i \neq j} N_i^{\text{II},f} \bar{U}_i^{\text{II},f} + \sum_{i \neq j} N_i^{\text{I},f} \bar{U}_i^{\text{I},f} - \sum_{i \neq j} N_i^{\text{II},0} \bar{U}_i^{\text{II},0} - \sum_{i \neq j} N_i^{\text{I},0} \bar{U}_i^{\text{I},0} \quad (5)$$

Because the number of molecules $N_{i \neq j}$ in each phase is not changing, this simplifies to:

$$\Delta U_{\text{tr},j} = N_j^{\text{II},f} \bar{U}_j^{\text{II},f} + N_j^{\text{I},f} \bar{U}_j^{\text{I},f} - N_j^{\text{II},0} \bar{U}_j^{\text{II},0} - N_j^{\text{I},0} \bar{U}_j^{\text{I},0} + \sum_{i \neq j} N_i^{\text{II}} (\bar{U}_i^{\text{II},f} - \bar{U}_i^{\text{II},0}) + \sum_{i \neq j} N_i^{\text{I}} (\bar{U}_i^{\text{I},f} - \bar{U}_i^{\text{I},0}) \quad (6)$$

Assuming a sufficiently small perturbation, the partial molar internal energies in each phase are invariant to the perturbation in the system, and this simplifies to:

$$\Delta U_{\text{tr},j} = (N_j^{\text{II},f} - N_j^{\text{II},0}) \bar{U}_j^{\text{II}} + (N_j^{\text{I},f} - N_j^{\text{I},0}) \bar{U}_j^{\text{I}} \quad (7)$$

because the number of molecules $i \neq j$ is not changing. Note - if the perturbation in j causes the partial molar internal energy of any species to change significantly, then the total change in internal energy may be influenced not only by changes in j but also by changes in \bar{U}_i .

The mass balance gives us:

$$\Delta N_j^{I \rightarrow II} = N_j^{\text{II},f} - N_j^{\text{II},0} = N_j^{\text{I},0} - N_j^{\text{I},f} \quad (8)$$

which allows us to simplify $\Delta U_{\text{tr},j}$ to

$$\Delta U_{\text{tr},j} = \Delta N_j^{I \rightarrow II} \bar{U}_j^{\text{II}} - \Delta N_j^{I \rightarrow II} \bar{U}_j^{\text{I}} \quad (9)$$

$$\frac{\Delta U_{\text{tr},j}}{\Delta N_j^{I \rightarrow II}} = \bar{U}_j^{\text{II}} - \bar{U}_j^{\text{I}} \quad (10)$$

which we abbreviate to:

$$\Delta \bar{U}_{\text{tr},j} = \bar{U}_j^{\text{II}} - \bar{U}_j^{\text{I}} \quad (11)$$

Likewise, the enthalpy of transfer is given by:

$$\Delta H_{\text{tr},j} = \Delta U_{\text{tr},j} + p \Delta V_{\text{tr},j} \quad (12)$$

$$\Delta H_{\text{tr},j} = N_j^{\text{II},f} \bar{H}_j^{\text{II},f} + N_j^{\text{I},f} \bar{H}_j^{\text{I},f} - N_j^{\text{II},0} \bar{H}_j^{\text{II},0} - N_j^{\text{I},0} \bar{H}_j^{\text{I},0} + \sum_{i \neq j} N_i^{\text{II}} (\bar{H}_i^{\text{II},f} - \bar{H}_i^{\text{II},0}) + \sum_{i \neq j} N_i^{\text{I}} (\bar{H}_i^{\text{I},f} - \bar{H}_i^{\text{I},0}) \quad (13)$$

which can be decomposed into its separate contributions from \bar{U} and $p\bar{V}$ (p being

equal for both phases).

$$\begin{aligned}\Delta H_{\text{tr},j} = & N_j^{\text{II},f} \bar{U}_j^{\text{II},f} + N_j^{\text{I},f} \bar{U}_j^{\text{I},f} - N_j^{\text{II},0} \bar{U}_j^{\text{II},0} - N_j^{\text{I},0} \bar{U}_j^{\text{I},0} + \\ & \sum_{i \neq j} N_i^{\text{II}} (\bar{U}_i^{\text{II},f} - \bar{U}_i^{\text{II},0}) + \sum_{i \neq j} N_i^{\text{I}} (\bar{U}_i^{\text{I},f} - \bar{U}_i^{\text{I},0}) + \\ & p(N_j^{\text{II},f} \bar{V}_j^{\text{II},f} + N_j^{\text{I},f} \bar{V}_j^{\text{I},f} - N_j^{\text{II},0} \bar{V}_j^{\text{II},0} - N_j^{\text{I},0} \bar{V}_j^{\text{I},0}) + \\ & \sum_{i \neq j} N_i^{\text{II}} (\bar{V}_i^{\text{II},f} - \bar{V}_i^{\text{II},0}) + \sum_{i \neq j} N_i^{\text{I}} (\bar{V}_i^{\text{I},f} - \bar{V}_i^{\text{I},0})\end{aligned}\quad (14)$$

Equation 14 holds in general. For a sufficiently small perturbation such that partial molar internal energies and partial molar volumes in each phase are invariant, it may be simplified to:

$$\begin{aligned}\Delta H_{\text{tr},j} = & (N_j^{\text{II},f} - N_j^{\text{II},0}) \bar{U}_j^{\text{II}} + (N_j^{\text{I},f} - N_j^{\text{I},0}) \bar{U}_j^{\text{I}} + \\ & p((N_j^{\text{II},f} - N_j^{\text{II},0}) \bar{V}_j^{\text{II}} + (N_j^{\text{I},f} - N_j^{\text{I},0}) \bar{V}_j^{\text{I}})\end{aligned}\quad (15)$$

which, after applying the mass balance, becomes:

$$\frac{\Delta H_{\text{tr},j}}{\Delta N_j^{\text{I} \rightarrow \text{II}}} = \bar{U}_j^{\text{II}} - \bar{U}_j^{\text{I}} + p(\bar{V}_j^{\text{II}} - \bar{V}_j^{\text{I}}) = \bar{H}_j^{\text{II}} - \bar{H}_j^{\text{I}} \quad (16)$$

We use this rather extended derivation in order to highlight a potential pitfall. Equation 12 is generally valid, but

$$\Delta \bar{H}_{\text{tr},j} = \bar{H}_j^{\text{II}} - \bar{H}_j^{\text{I}} \quad (17)$$

is ONLY valid when a sufficiently small perturbation does not result in changes to \bar{H}_i within each phase.

1.2. Enthalpies of adsorption

We also use this derivation to explore the validity of this approach in constant-volume systems. Consider adsorption into a porous material with a constant volume V_{ad} . Phase II is the adsorbent and phase I is a fluid mixture. The expression for enthalpy of transfer of species j in an NpT -Gibbs ensemble is given by:

$$\Delta H_{\text{tr},j} = \Delta U_{\text{tr},j} + p \Delta V_{\text{tr},j} \quad (18)$$

Here, $\Delta V_{\text{tr},j}$ needs clarification. We can express it as

$$\Delta V_{\text{tr},j} = V^f - V^0 = V^{\text{II},f} - V^{\text{II},0} + V^{\text{I},f} - V^{\text{I},0}, \quad (19)$$

noticing that $V^{\text{II},f} - V^{\text{II},0} = 0$ for an incompressible adsorbent. Consequently, $\Delta V_{\text{ads},j}$ only depends on changes in the fluid,

$$\Delta V_{\text{ads},j} = V^{\text{I},f} - V^{\text{I},0} \quad (20)$$

which, for a sufficiently small perturbation, may be evaluated by changes in the partial molar volume.

2. SI for natural gas condensate

2.1. Benchmarking with numerical differentiation

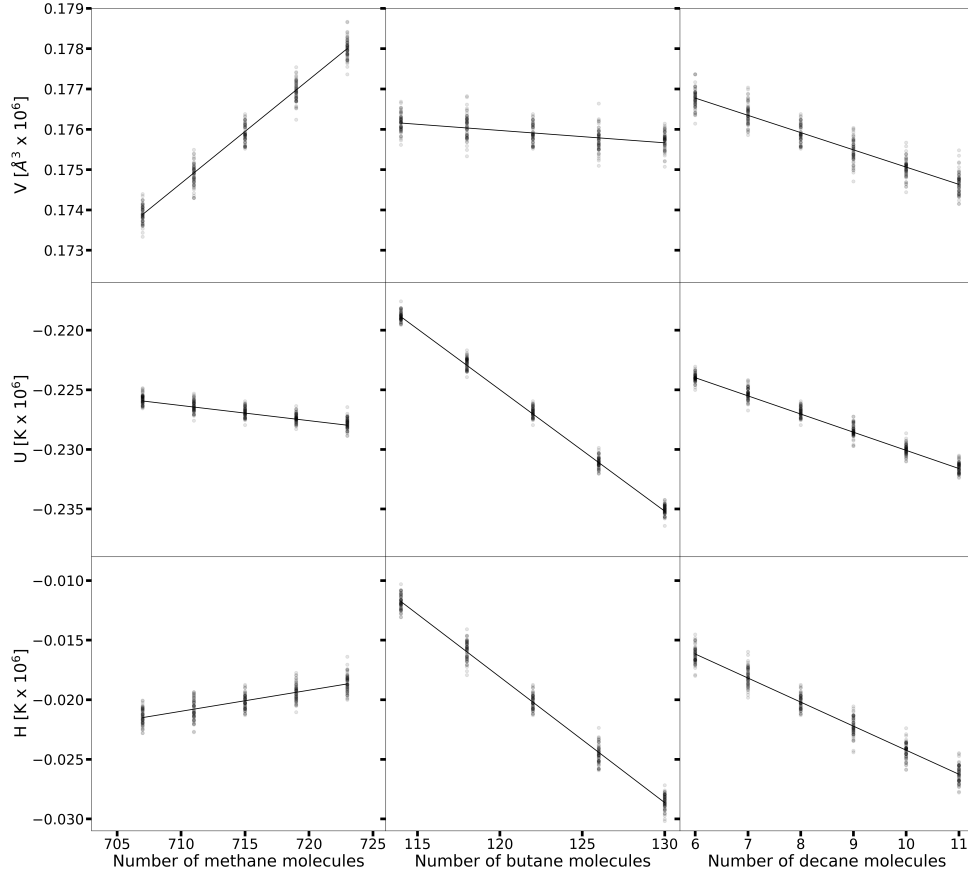


Figure S1.: V , U , and H are calculated in the NpT ensemble at different compositions by varying the number of molecules of methane, butane and decane separately. Partial molar properties are calculated by fitting a straight line to the data. Data shown here for the vapor phase.

2.2. Correlations in V , U , and p

Figure S2 is provided to visualise correlations between V , U , and p in the liquid and the vapor of the model natural gas condensate. V is strongly correlated with U in both the liquid and the vapor phase - this is to be expected as a larger simulation box can accommodate more molecules and a larger energy. The instantaneous pressure is uncorrelated with both V and U .

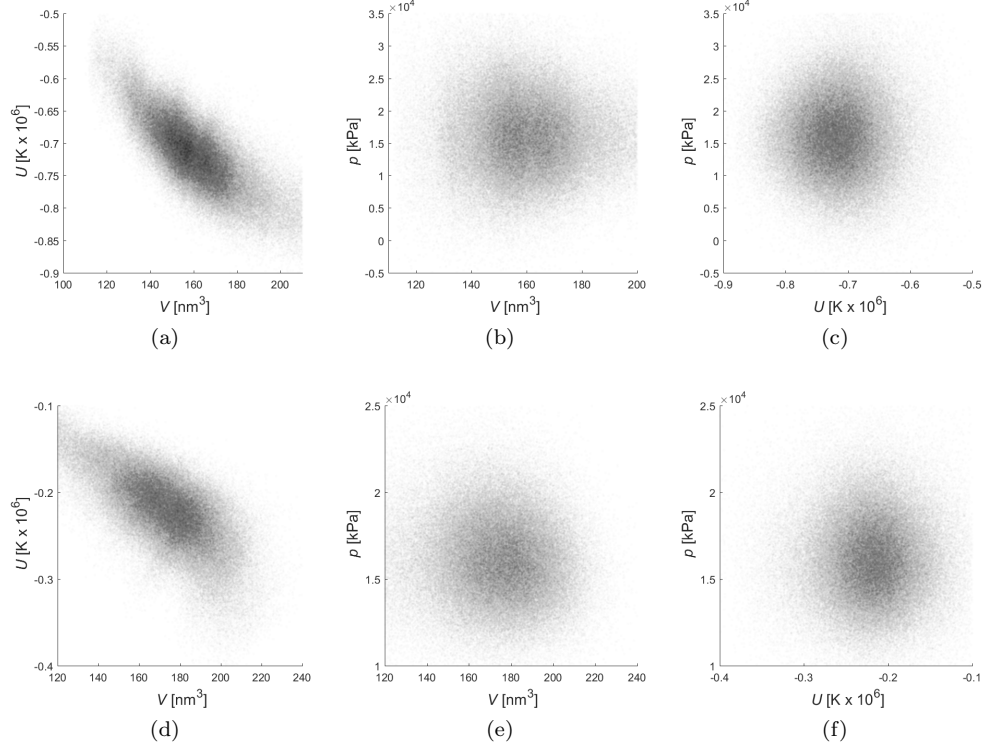


Figure S2.: Correlations between instantaneous observables V , U , and p in the liquid (a-c) and vapor (d-f) phase of the model natural gas condensate with $N = 1825$ in the NpT -Gibbs ensemble.

2.3. Uncertainty analysis

Table S1.: Volume, internal energy, and enthalpy balances on model natural gas condensate at 333 K, in the NVT -Gibbs and NpT -Gibbs ensembles. Uncertainties reported are 95% confidence intervals from 64 independent measurements of \bar{V}_i . $\bar{H}_{\text{inst},i}$ were calculated using $\mathbf{N} \cdot \bar{\mathbf{H}} = \mathbf{H}$ and instantaneous measurements of the enthalpy of each fluid box at each frame. $\bar{H}_{\text{ave},i}$ were calculated using $\bar{H}_i = \bar{U}_i + p\bar{V}_i$, and the ensemble average pressure of the vapor box for each independent simulation.

		NVT -Gibbs		NpT -Gibbs	
		Liquid	Vapor	Liquid	Vapor
$\bar{V}_i N_i$ [nm ³]	C1	89 ₃	168 ₄	86 ₃	178 ₄
	C4	39.7 ₁₃	6.8 ₁₇	47.6 ₁₂	-3.4 ₆
	C10	28.4 ₇	0.30 ₂₅	27.2 ₉	-3.3 ₃
	$\sum \bar{V}_i N_i$	157 ₃	175 ₄	161 ₃	171 ₄
	V_{box}	157 ₃	175 ₃	161 ₃	171 ₄
$\bar{U}_i N_i$ [K x 10 ³]	C1	5 ₅	-116 ₃	-45 ₃	-91 ₂
	C4	-466 ₇	-105 ₄	-421 ₆	-115 ₂
	C10	-245 ₇	-6.0 ₆	-251 ₇	-12.5 ₆
	$\sum \bar{U}_i N_i$	-706 ₁₁	-228 ₅	-716 ₁₀	-219 ₃
	U_{box}	-706 ₈	-228 ₈	-716 ₉	-220 ₈
$\bar{H}_{i,\text{inst}} N_i$ [K x 10 ³]	C1	58 ₅	100 ₄	57 ₄	118 ₃
	C4	-368 ₉	-111 ₅	-365 ₉	-119 ₅
	C10	-212 ₅	-10.7 ₉	-219 ₆	-16.4 ₁₂
	$\sum \bar{H}_i N_i$	-522 ₁₁	-22 ₇	-527 ₁₂	-18 ₆
	H_{box}	-521 ₅	-22 ₄	-526 ₅	-18 ₃
$\bar{H}_{i,\text{ave}} N_i$ [K x 10 ³]	C1	110 ₇	82 ₆	59 ₄	117 ₃
	C4	-419 ₁₀	-97 ₆	-367 ₉	-119 ₅
	C10	-212 ₆	-6.0 ₈	-219 ₆	-16.4 ₁₂
	$\sum \bar{H}_i N_i$	-522 ₁₄	-22 ₉	-527 ₁₂	-18 ₆
	H_{box}	-521 ₅	-22 ₄	-527 ₅	-18 ₃

2.4. System size effects

Table S2 provides the thermodynamic properties of the natural gas condensate in the NpT -Gibbs ensemble with $N = 873$ and $N = 3761$ in order to compare with the $N = 1825$ system described in Table 1. n -decane, which is only a trace compound in the vapor phase, exhibits a highly skewed distribution in loading that appears to be truncated in the $N = 873$ system (Fig. S3). The factor of 4.3 increase in system size from $N = 873$ to $N = 3761$ does not change any measured property, whether bulk or partial molar, within the 95% confidence interval in either ensemble. A small deviation may be present in the minor component, n -decane, which had a large variance in the vapor phase in the small NpT system ($x_{\text{C10}} = 0.0137 \pm 38$ as compared to around 0.0097 for the others). This is associated with a departure in \bar{V} in the vapor phase, but with the exception of this property, all properties are consistent, within error, across a factor of 4.3 change in system size. As the system gets larger and the fluctuations decrease, we find partial molar properties are also calculated with more precision. As the system size approaches the thermodynamic limit, we may, in theory, be able to compute partial molar properties using MLR, but reaching sufficient precision will become more challenging because the fluctuations are infinitesimally

small and numerical issues associated with ill-conditioned matrices may become more pronounced.

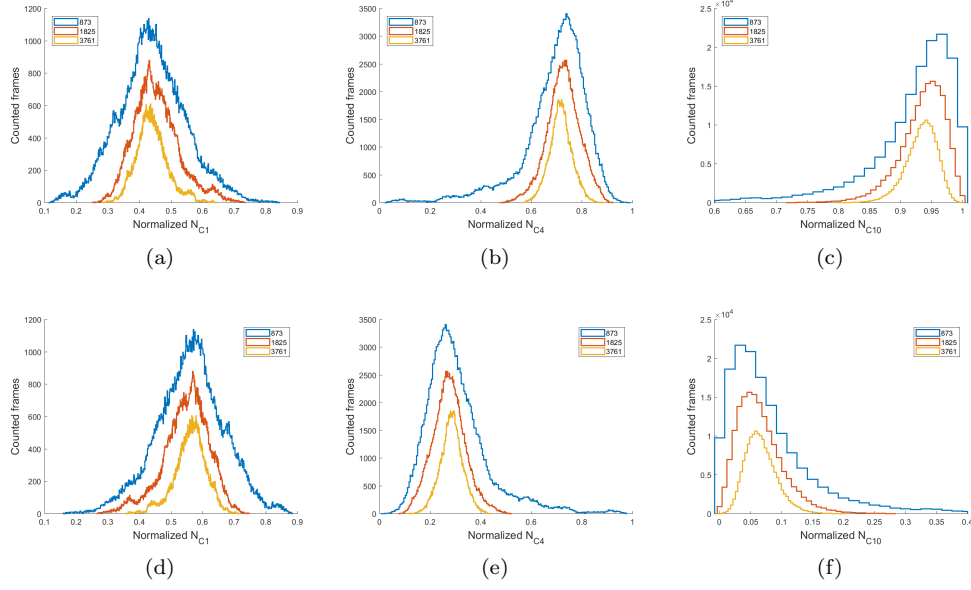


Figure S3.: Distributions of molecules in the liquid (a-c) and vapor (d-f) phases of the model natural gas condensate in the NpT -Gibbs ensemble as a function of system size. Molecule populations are normalised by the total number of molecules of species i in the two-box system, N_i/N_i^{total} . Bin widths for each system size are set to $1/N_i^{\text{total}}$, so that each step corresponds to a discrete change in N_i .

Table S2.: System size effects are explored by comparing the model natural gas condensate with $N = 873$ to that with $N = 3761$ in the NpT -Gibbs ensemble with $p = 16220$ MPa. Each system has 50,000 MC cycles of production, though some simulations of the small system finished with fewer than 50,000 MC cycles when one box shrank to less than 28 Å (twice the force field cutoff of 14 Å). Uncertainties reported are 95% confidence intervals from 64 independent measurements of each quantity.

	Liquid $N = 873$	Liquid $N = 3761$	Vapor $N = 873$	Vapor $N = 3761$
N_{C1}	275 ₁₇	1157 ₂₇	335 ₁₇	1473 ₂₇
N_{C4}	142 ₆	626 ₇	61 ₆	248 ₇
N_{C10}	53.8 ₂₀	239.9 ₈	6.2 ₂₀	17.1 ₈
x_{C1}	0.580 ₁₀	0.571 ₃	0.840 ₁₁	0.848 ₂
x_{C4}	0.303 ₆	0.310 ₂	0.146 ₇	0.142 ₂
x_{C10}	0.117 ₅	0.119 ₂	0.014 ₄	0.009 ₃
p [kPa]	16260 ₇₀	16220 ₂₀	16240 ₆₀	16230 ₁₀
V [nm ³]	76 ₄	325 ₅	82 ₄	362 ₆
U [K]	-330 ₁₀	-1458 ₁₀	-117 ₁₃	-468 ₁₄
$E_{pV, \text{ inst}}$ [K x 10 ³]	89 ₄	381 ₆	96 ₅	426 ₈
$E_{pV, \text{ ave}}$ [K x 10 ³]	89 ₄	381 ₆	97 ₅	426 ₈
H_{inst} [K x 10 ³]	-241 ₁₀	-1077 ₈	-21 ₁₀	-42 ₅
H_{ave} [K x 10 ³]	-241 ₁₀	-1076 ₈	-21 ₁₀	-42 ₅
\bar{V}_{C1} [L/mol]	0.091 ₄	0.0885 ₁₀	0.152 ₂	0.1531 ₈
\bar{V}_{C4} [L/mol]	0.089 ₆	0.092 ₂	-0.012 ₁₂	-0.015 ₅
\bar{V}_{C10} [L/mol]	0.14 ₂	0.145 ₅	-0.20 ₃	-0.250 ₁₈
\bar{U}_{C1} [kJ/mol]	-0.69 ₁₉	-0.66 ₇	-1.01 ₇	-1.08 ₃
\bar{U}_{C4} [kJ/mol]	-11.70 ₄₃	-11.42 ₁₉	-8.6 ₄	-8.33 ₁₅
\bar{U}_{C10} [kJ/mol]	-16.9 ₁₃	-17.6 ₅	-13.2 ₁₁	-13.0 ₅
$\bar{H}_{C1, \text{ inst}}$ [kJ/mol]	0.51 ₈₂	0.79 ₉	1.46 ₆	1.41 ₃
$\bar{H}_{C4, \text{ inst}}$ [kJ/mol]	-9.90 ₅₆	-9.9 ₃	-8.8 ₃	-8.60 ₂₀
$\bar{H}_{C10, \text{ inst}}$ [kJ/mol]	-13 ₅	-15.4 ₆	-16.5 ₁₇	-17.0 ₇
$\bar{H}_{C1, \text{ ave}}$ [kJ/mol]	0.8 ₂	0.77 ₈	1.46 ₈	1.41 ₃
$\bar{H}_{C4, \text{ ave}}$ [kJ/mol]	-10.3 ₄	-9.9 ₂	-8.8 ₅	-8.58 ₂₁
$\bar{H}_{C10, \text{ ave}}$ [kJ/mol]	-14.7 ₁₆	-15.2 ₅	-16.4 ₁₄	-17.1 ₇

3. SI for liquid-liquid equilibrium

$$\begin{aligned}
N_1^a \bar{V}_1 + N_2^a \bar{V}_2 + N_3^a \bar{V}_3 &= V^a \\
N_1^b \bar{V}_1 + N_2^b \bar{V}_2 + N_3^b \bar{V}_3 &= V^b \\
N_1^c \bar{V}_1 + N_2^c \bar{V}_2 + N_3^c \bar{V}_3 &= V^c
\end{aligned} \tag{21}$$

It is possible to use Eq. 21 to predict these partial molar properties using data at multiple system compositions. This approach would be akin to experimentally characterizing the bulk properties at each of these state points, and solving the system of equations to compute the partial molar properties. Anderson et al. used this method to calculate partial molar volumes in binary mixtures of water and silica using least squares regression, with multiple simulations at constant NpT [1]. By counting each n -hexane molecule as 1/2 n -dodecane, and using the average values for N_i , V , and U for each of the three EtOH-rich phases, \bar{V}_i , \bar{U}_i , and \bar{H} may be calculated, and once again for the three n -dodecane-rich phases (Table S4). Eq. 21 should in principle, give estimates of the partial molar properties that are *averaged* across the composition space of the three state points at hand, but in several cases, gives values outside this range. For example, \bar{U}_{C12} in the EtOH-rich phase ranges from -6 to -11.1 kJ/mol by MLR, but \bar{U}_{C12} by Eq. 21 is -16.2 kJ/mol. Likewise, \bar{U}_{H2O} in the n -dodecane-rich phase is consistently -32 kJ/mol according to MLR, but only -18.2 kJ/mol by Eq. 21. These issues arise because of relatively large fluctuations in the loading of trace species (e.g. H_2O); perturbing the mean values of these trace species within the uncertainty of the simulation ensemble averages and repeating the MLR can cause large variations in the partial molar properties from Eq. 21. Consequently, we do not recommend using the system of equations approach for calculating partial molar properties.

Table S3.: Compositions of LLE systems at $T = 333$ K and $p = 0.1$ MPa, from Harwood et al. [2].

			EtOH	n -Hexane	n -Dodecane	Water
$x_{H_2O}^{total}=0.1$	EtOH-rich	N_i	479 ₃₄	1.6 ₃	29 ₁₁	89 ₄
		x_i	0.801 ₁₃	0.0026 ₄	0.048 ₁₆	0.149 ₆
	C12-rich	N_i	113 ₃₄	3.1 ₃	269 ₁₁	8 ₄
		x_i	0.279 ₅₅	0.0078 ₆	0.69 ₆	0.020 ₇
$x_{H_2O}^{total}=0.25$	EtOH-rich	N_i	552 ₁₁	1.3 ₁	7.6 ₈	292 ₃
		x_i	0.647 ₃	0.00153 ₇	0.0089 ₁₀	0.342 ₂
	C12-rich	N_i	41 ₁₁	3.43 ₅	289.9 ₈	4.1 ₂₇
		x_i	0.12 ₂	0.0102 ₃	0.86 ₃	0.012 ₇
$x_{H_2O}^{total}=0.36$	EtOH-rich	N_i	566 ₅	1.68 ₈	3.7 ₄	492.5 ₁₁
		x_i	0.5319 ₁₇	0.00157 ₇	0.0035 ₃	0.4631 ₁₆
	C12-rich	N_i	28 ₅	3.08 ₇	293.8 ₄	2.6 ₁₁
		x_i	0.083 ₁₂	0.0094 ₂	0.899 ₁₅	0.008 ₃

Table S4.: Partial molar properties calculated using Eq. 21, and the ensemble average N_i , V , and U for three EtOH-rich and three C12-rich mixtures.

i	\bar{V}_i [L/mol]		\bar{U}_i [kJ/mol]		\bar{H}_i [kJ/mol]	
	EtOH-rich	C12-rich	EtOH-rich	C12-rich	EtOH-rich	C12-rich
EtOH	0.0612	0.0618	−33.3	−32.9	−33.3	−32.9
C12	0.242	0.234	−16.2	−16.4	−16.2	−16.4
H ₂ O	0.0179	0.0492	−38.6	−18.2	−38.6	−18.2

4. Code and example files

Monte Carlo for Complex Chemical Systems (MCCCS-MN) is the Monte Carlo simulation program developed and used by the Siepmann research group. The source code for this software, as well as sample input files (fort.4, fort.77, and topmon.inp) and output files (run1a.dat and fort.12) for a production run of the ternary methane/*n*-butane/*n*-decane mixture at 333 K and 16220 kPa in the NpT -Gibbs ensemble, are available at <https://github.com/SiepmannGroup/PartialMolarProperties>.

References

- [1] K.E. Anderson, L.C. Grauvilardell, M.M. Hirschmann and J.I. Siepmann, J. of Phys. Chem. B **112** (41), 13015–13021 (2008).
- [2] D.B. Harwood, C.J. Peters and J.I. Siepmann, Fluid Phase Equilib. **407**, 269–279 (2016).