Lessons Learned from Theory and Simulation of Step Potentials

J. Richard Elliott

Chemical Engineering Department, The University of Akron, Akron, OH 44325-3906, USA

Abstract

This article provides a brief review of the model referred to as Step Potentials for Equilibria and Discontinuous Molecular Dynamics (SPEADMD) with observations that may be instructive in formulating the next generation of SAFT models. Effects of branches, fused spheres, rings, and molecular flexibility are demonstrated. The evolution of contributions of the equation of state from low to high molecular weight is shown to be sensitive to details of the molecular structure. Statistical mechanical details are shown to affect the formulation of fundamental theory and trends to be expected. Recent results for contributions beyond second order in temperature are shown to substantially improve predictions in the critical region. For mixtures, it is shown how molecular simulation of the off-lattice excess entropy and energy can be achieved, with implications for mixing rules. A heightened emphasis is placed on the confounded nature of molecular scale site-site interactions relative to macroscopic pure component and mixture data. This leads to general conclusions about how SAFT models should be articulated going forward, and how progress of the thermodynamic modeling community in general might be improved.

Keywords: Physical property prediction; Vapor pressure; Density; Molecular simulation; transferable force fields; molecular interactions.

Introduction.....	2
Fundamentals of DMD/TPT for Pure Fluids	3
Definition of Step Potentials.....	3
Proof of Principle.....	5
Definitions of TPT Contributions and Behavior for n-Alkanes.....	6
Trends in TPT Contributions with Molecular Structure	7
Subtle Influences on A2.....	8
Association and Wertheim's Theory.....	9
Stiffness and Branching	10
Beyond Second Order TPT.....	12
Softness and Compressed Fluid Density.....	13
Mixtures	14
Inference of the Excess Entropy from Off-Lattice DMD Simulations	14
Inference of the Excess Energy.....	16
Confounded Site-Site Interaction Energies.....	17
Molecular Interactions and Databases	19
Experimental data for phase equilibria	19
Molecular Simulation Data.....	20
Transferable vs. Customized Potentials.....	21
Conclusions.....	21

Introduction

The SAFT framework has been a tremendous step forward in the formulation of practical equations of state that have a verifiable basis in statistical mechanics and well-defined potential functions. Inevitably, however, discrepancies arise between the original framework (based on flexible, tangent-sphere chains) and more detailed molecular structures and their models. Effects of branches, fused spheres, rings, and flexibility can be demonstrated. To a certain extent, many of these effects have been anticipated through the model referred to as Step Potentials for Equilibria and Discontinuous Molecular Dynamics (SPEADMD). As progress is contemplated to address detailed effects beyond the original SAFT framework, it may be useful to consider the lessons learned from the SPEADMD model. To that end, this article provides a brief review of the SPEADMD model with observations that may be instructive in formulating the next generation of SAFT models.

As engineers apply molecular perspective to the topic of "Fluid Phase Equilibria," our primary interest is in models of molecular interactions that can be applied to predictions of phase equilibrium and transport properties for engineering applications. This includes phase equilibrium of pure fluids, as characterized by vapor pressure, as well as vapor-liquid equilibria (VLE), liquid-liquid (LLE), and solid-liquid equilibria (SLE) of mixtures. These kinds of properties have been modeled in the past with equations of state and activity models that are often specific to a particular application. For example, independent UNIFAC models exist for predicting vapor pressure, critical temperature, critical pressure, boiling temperature, vapor-liquid equilibria, and liquid-liquid equilibria. No two of these models share the same UNIFAC parameters. In the context of transferable potential models, all of these properties should be predicted by a single set of parameters. This is an entirely different scale of demand for an engineering model. At first glance, it may seem a "fool's errand" to attempt such a feat. Nevertheless, we show in the course of this review that predictions of vapor pressure with transferable step potentials are now at a level of outperforming group contribution predictions, naturally subsuming predictions of critical properties and boiling points. Furthermore, predictions of mixed phase equilibria are competitive with existing engineering models. In addition to the practical benefits of having a single general model to predict many properties, there is the pedagogical benefit of relating the macroscopic properties directly to the model of molecular interactions: addressing deficiencies in the macroscopic predictions leads to increasingly sophisticated insights at the nanoscopic level. One might surmise (not over glibly) that knowing the force laws between all atoms would enable chemical engineers to design nanostructures in the same way that civil engineers design infrastructure.

In work occurring over the past 10-15 years, we have considered the problem of transferable potential models within the context of discontinuous molecular dynamics (DMD) and thermodynamic perturbation theory (TPT). DMD was necessary because we formulated our potentials as a sequence of steps, as described below. DMD was also convenient because it was much more computationally efficient than MD of continuous potentials (CMD) using the serial processors prevalent at the time. TPT was beneficial because the depths of the steps comprise the perturbations, which are applied post-simulation to infer the entire equation of state $A(T, \rho)$. Through TPT, a specific equation of state for each molecular structure and instance of assumed well-depths, all of the thermodynamic properties could be computed and compared to experimental data. By minimizing the deviations between computations and experiment, the optimal well-depths could be inferred. We have now characterized the potential functions of roughly 60 site types and 500 molecules.¹

Major components of this transferable potential model were the blister potentials and related theory of associating molecules developed by Wertheim.² Note that blister potentials are inherently discontinuous, requiring DMD for simulation of their dynamics from the outset.³ Noting that the TPT contributions are computed rigorously through the statistical mechanical averages, and the association contribution is computed from Wertheim's theory, our implementation of DMD/TPT is entirely consistent with the spirit of the "statistical associating fluid theory" (SAFT) originated by Chapman et al.⁴ The primary refinements are in the detailed treatment of molecular characteristics such as flexibility, branching, and rings. These details preclude some of the generalized trends that are often convenient within implementations of SAFT models to date. For example, the local compressibility approximation is not found to be accurate for the second order term, and the perturbation contributions do not scale linearly with a single shape factor applied to the hard sphere reference term. Nevertheless, we consider our model to conform to the SAFT perspective in its most fundamental sense of rigorously connecting assumed potential models with their macroscopic properties. We hope that this review may inform future developers of the SAFT perspective who may be interested in detailed molecular interactions. b

The manuscript is organized under three primary headings: pure fluids, mixtures, and databases. The discussion of pure fluids includes mathematical definitions of the manner in which the molecular interactions are characterized through step potentials and their molecular simulations. The quantitative nature of TPT in characterizing the thermodynamics of the full potential is demonstrated and trends in the TPT contributions are analyzed, emphasizing the ways in which molecular details like step depths, softness, flexibility, and branching alter the thermodynamics. The discussion of mixtures shows how molecular simulation leads to functional forms for interpolating simulations with varying composition. These interpolation formulas lead to unambiguous mixing rules at the equation of state level. The consideration of mixtures also leads to emphasis on the site-site nature of molecular interactions and the need for data to characterize these interactions. The discussion of databases addresses the current status and ways that data management can be improved.

Fundamentals of DMD/TPT for Pure Fluids

A number of preliminary considerations must be established before proceeding to review the findings. Among these are basic definitions of the fundamental terms and statistical mechanical averages. The connection between thermodynamic perturbation theory (TPT) and results of the simulated potential must also be established. These considerations lead to a few observations worthy of special mention in clarifying how step potentials provide a theory that is intermediate between simple square well models and continuous force fields.

Definition of Step Potentials

Step potentials follow a fairly intuitive definition. The potential function is broken down into a sequence of steps that could, in principle, reproduce the square well potential with a single step or the Lennard-Jones potential with roughly a 10-20 steps.⁵ The engineering question is how many steps suffice to obtain a potential that represents what can be achieved with a continuous potential? Chapela⁵ addressed that question in the context of the single-site Lennard-Jones model. A more general perspective pertains to transferable potential models like n-alkanes. The assumption of transferability involves its own level of approximation. Assumptions inherent in interaction site models for multi-site molecules also involve approximations. Therefore, the

answer to this engineering question is not as simple in the context of transferable interaction site models for multi-site molecules. Cui and Elliott addressed this question for n-alkanes and found that four wells were sufficient, with well transitions at $r/\sigma = 1.2, 1.5, 1.8, 2.0$, where r is the site-site center to center distance and σ is the site diameter.⁶ Unlu et al. further determined that reasonable accuracy could be retained while reducing the number of parameters by interpolating the middle wells.⁷ Since the work of Unlu et al., we have referred to this model of step potentials for equilibria and discontinuous molecular dynamics as the SPEADMD model ("*speed-MD*").

Table 1 lists samples of the site parameters from the latest characterization,¹ and Figure 1a illustrates a few key site types.⁷ One trend is clear at the outset; as a site is located near to other sites such that the ranges of attraction overlap, the optimized potential becomes less steep as the extent of overlap increases. Keep in mind that the potentials are optimized relative to experimental data, the fundamental source of this inference. When we apply potential models with a fixed steepness for all site types (e.g. r^{-6}), we assume a uniform steepness. The r^{-6} dependency was derived for spherically symmetric molecules like argon. It is not clear that it should be exactly the same for overlapping interaction sites, or united atom models. This observation lends added perspective to recent efforts with Mie potentials.⁸

We can gain additional insight into the nature of step potentials by considering the attractive overlaps for interaction site models with a square well potential, as illustrated in Figure 1b.⁶ The site diameters, bond lengths, and bond angles are drawn to scale for n-alkanes in the trans conformation. A sample trajectory shows how a probe particle would be deflected as it crosses first one attractive well then another of an ethane molecule. Each of these wells operates independently under the assumptions of interaction site models. The potential energy is double in positions where a single attractive overlap occurs, and triple where double overlaps occur, as in n-butane. Molecules like benzene or neopentane may exhibit triple overlaps. Therefore, a square well potential is actually a step potential when applied as an interaction site model.

Figure 1. Schematic diagrams illustrating key features of step potential models. (a) Uniform bond lengths and bond angles are applied for n-alkanes through the intramolecular potential wells centered at 0.154 and 0.265 nm, while variable intermolecular interactions apply at distances of 0.27 nm or greater. (b) A square well potential model for each individual interaction site leads to multiple steps in the interaction site model of the overall molecule, illustrated here for united atom models of ethane and n-butane.

Table 1. Samples of site types and potential parameters.

Type	ε_1/k_B (K)	ε_A/k_B	σ (nm)	Example
CH3a-	65.5	32.8	0.3630	1' terminus e.g. nOctane
CH3b-	54.6	32.9	0.3630	2' branch e.g. 2,3 dimethylbutane
CH3c-	42.6	42.6	0.3630	3' branch e.g. 2,2dimethylHexane
>CH2a	35	16.9	0.3570	nAlkyl, ether, ester. e.g. nOctane
>CH-a	7.2	7.2	0.3900	Alkyl branch, e.g. 2MethylHexane
>C<a	0.2	0.2	0.3425	3' alkyl branch, e.g. 2,2DiMethylHexane,
=ACH-a	27.8	22.5	0.3425	Aromatic carbon, e.g. benzene
>AC=	15.5	11.7	0.3300	Aromatic branch, e.g. Toluene
CH2=a	60.4	27.7	0.3500	Alpha olefin bonded to =CH-, e.g. 1Hexene
=CH-a	23.4	20.8	0.3500	Alkene, e.g. 1Hexene or 3Pentene

Proof of Principle

A key step in establishing the connection between theory and simulation is demonstrating that the theory provides an accurate characterization of the simulation results. This fundamental requirement is illustrated in Figure 2, where U is the internal energy, Z is the compressibility factor, and η is the packing fraction. There are slight deviations in internal energy near the critical packing fraction of roughly 0.15, but this is to be expected from the scaling in the critical region. The compressibility factor shows remarkable accuracy at packing fractions greater than roughly 0.32. For small molecules, packing fractions greater than 0.32 correspond to saturated or compressed liquids at reduced temperatures ($T_r \equiv T/T_c$, where T_c is the critical temperature) less than 0.9. The observation that the compressibility factor is so linear along isochores is a reflection of the small slope of the second order term with respect to density, especially in the liquid regime. In this way we begin to develop an appreciation for interpreting the shape of the second order term.

Figure 2. Comparison of DMD/TPT to DMD of the full potential for n-butane. (a) Internal energy. (b) Compressibility factor. Points represent simulations with the full potential. Lines represent applications of TPT where only the reference potential is simulated.

Definitions of TPT Contributions and Behavior for n-Alkanes

Additional insight requires articulation of the TPT contributions in terms of their statistical mechanical averages. These are given to second order for step potentials by,

$$\frac{A - A^{id}}{RT} = A0(\eta) + \frac{A1(\eta)}{T} + \frac{A2(\eta)}{T^2} + \dots \quad (1)$$

$$A0(\eta) = \int_0^\eta (Z0 - 1) \frac{d\rho}{\rho} \quad (2)$$

$$A1 = -\frac{1}{N} \sum_i^{\text{sites}} \sum_j^{\text{sites}} \sum_m^{\text{wells}} u_{ijm} \langle N_{ijm} \rangle \quad (3)$$

$$A2 = -\frac{1}{2N} \sum_i^{\text{sites}} \sum_j^{\text{sites}} \sum_k^{\text{sites}} \sum_l^{\text{sites}} \sum_m^{\text{wells}} \sum_n^{\text{wells}} (\langle N_{ijm} N_{ikn} \rangle - \langle N_{ijm} \rangle \langle N_{ikn} \rangle) u_{ijm} u_{ikn} \quad (4)$$

$$\frac{U - U^{ig}}{RT} = \frac{A1}{T} + \frac{2A2}{T^2} + \dots \quad (5)$$

$$Z - Z_0 = \frac{Z1}{T} + \frac{Z2}{T^2} + \dots \quad (6)$$

where N_i gives the independent pairs in a given well. These are counted by summing only site numbers greater than the given site.^{9,6} N_m 's are tabulated at intervals between $r/\sigma_{ij} = 1, 1.05, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0$, then averaged post-simulation to expedite simulations. $Z0$ is computed directly from the virial summed over all collisions. In this way, the potential definition can be instantly altered to achieve various characterizations of the steps. For example, a step potential with wells at 1.3, 1.5, 1.7, 2.0 could be formulated as easily as one with wells at 1.2, 1.5, 1.8, 2.0. Note that the order of the TPT term is indicated deliberately with a suffix rather than a subscript. Subscripts are reserved for characterizing mixing rules in later discussion. Examples of the reference ($A0$), first order ($A1$) and second order terms are illustrated for a broad range of molecular weights of n-alkanes in Figure 3.¹⁰

Figure 3. Reference and TPT contributions for n-alkanes. The reference fluid is renormalized to emphasize the detailed behavior of the various chain lengths. The solid lines represent a generalized correlation that is analogous to PC-SAFT. The dashed line in (a) represents infinitely long, flexible tangent sphere chains (TSinf).

In viewing Eq. (4) it is apparent that the second order term ($A2$) represents the fluctuations in the free energy. True to the TPT perspective, $A2$ should approach zero at high

density where the thermodynamics are dominated by packing effects inherent in $\langle N_{ijm} \rangle_0$, and fluctuations become smaller as higher densities constrain the molecular movements. Furthermore, one should expect A_2 to be largest in magnitude near the critical density, where fluctuations are largest. Figure 3 shows that this is indeed true for n-alkanes. We have also simulated many other molecular structures and find that it is true for all cases of step potentials.¹¹

Trends in TPT Contributions with Molecular Structure

Figure 3 also shows the manner in which each contribution approaches its long-chain limit. For each contribution, the distinction between molecular weights decreases with increasing molecular weight until an asymptote is approached at roughly 1200 molecular weight. Coincidentally, this molecular weight also corresponds roughly to the entanglement threshold.¹² Beyond the entanglement threshold, the equation of state scales linearly with molecular weight as expected for a true polymer. In this way, the progression from oligomers to polymers can be easily appreciated. Once again, this pattern extends to other molecular structures.¹¹

It is interesting to contemplate a generalized equation of state based on n-alkanes in a similar manner to the way that the perturbed chain SAFT (PC-SAFT) model is based on using freely jointed tangent sphere chains as its reference system. One might refer to this formulation as the perturbed alkane SAFT model (PASAFT). Ghobadi and Elliott provided generalized relations for estimating the TPT contributions of n-alkanes for such a model, as illustrated by the solid curves in Figure 3. Unfortunately, the limitations of such a model become apparent when one considers its adaptation for any other molecular structure extrapolated to the polymer limit. For example, Figure 4 illustrates the TPT contributions of polyisoprene with the PASAFT contributions superposed as solid curves. Comparing the A_0 and A_1 terms, it is apparent that one cannot superpose the trends in the TPT terms without adjusting the effective well-depths and site diameters. But the well-depths and site diameters have been inferred specifically for the sites in each molecular structure. Adjusting them in an ad hoc manner would undermine any predictive capability that might be desired. Hence there is little point in developing a generalized alternative to the freely jointed chain, because it would be just as limited as tangent sphere chains in recognizing that each molecular structure possesses its own "fingerprint." Those "fingerprints" can be characterized as in Figure 3 and Figure 4 but they cannot be superseded by a universal curve for all molecular structures.

Figure 4. Reference and TPT contributions for polyisoprene. Symbols correspond to the oligomer number where, for example, 8-mer refers to $H-(C_5H_{10})_8-H$. Curves represent the equivalent n-alkane with same molecular weight as polyisoprene oligomer.

Subtle Influences on A_2

The previous discussion has especially emphasized the nature of A_2 because its behavior is not always as simple as its seemingly intuitive rationale might suggest. Furthermore, it must be recognized that accurate characterization of A_2 is essential to an accurate implementation of TPT. Approximations like the local compressibility approximation or macroscopic compressibility approximation are not sufficient for quantitative agreement between theory and simulation at $T_r < 0.9$.¹³ To clarify, we would like to explain two instances in which A_2 deviates from the expected behavior and how those deviations arise: the square well potential and the Barker-Henderson formulations of TPT for Lennard-Jones spheres.

Figure 5a illustrates the trends in A_2 for three potential models that are closely related. They all apply to two-step spheres and the variations are easily computed using different step depths. Beginning with the square well model with $\lambda = 2.0$, we can imagine that this model is composed of two steps ($r/\sigma = 1-1.5, 1.5-2$) where the second well is equal to the first (triangles). When the depth of the second well is equal to zero, we have the square well fluid with $\lambda=1.5$ (diamonds). When the depth of the second well is equal to 25% of the first well, we have a very simple step potential (squares). The square well potentials both indicate a shoulder in A_2 at high density. Observing the behavior of A_2 while varying the second well depth shows that A_2 is first intermediate between $\lambda=2$ and $\lambda=1.5$, then diminishes below $\lambda=1.5$, then rises back to the $\lambda=1.5$ result as the second well approaches zero. The smallest values for A_2 are obtained for the illustrated 25%. Thus it becomes apparent that the shape of A_2 is related to diminishing attractive effects in the potential with increasing range. The reason the shoulder appears relates to the sharp change in energy at the end of the square well range. Owing to this sharp change, fluctuations in the vicinity lead to large swings in energy. The impacts of these fluctuations are then reflected in A_2 , of course. Briefly, the high density shoulder in the A_2 of square well sphere fluids is artificial, and should be discounted when contemplating general implementations of TPT.

Another instance in which A_2 deviates from the expected behavior is more subtle. It relates to Lennard-Jones (LJ) spheres and the manner of separating the reference potential from the attractive part of the potential. Understanding this separation would be important to incorporating influences of softness in a step potential, but also in its own right. The softness of the potential function is an important detail that is discussed more at a later point in this review. The Barker-Henderson (BH) separation is given by Eq. (7).¹³ The separation of Weeks, Chandler, and Andersen (WCA), on the other hand, is given by Eq. (8).¹⁴

$$u^{ref} = \begin{cases} u^{LJ} & r < \sigma \\ 0 & r > \sigma \end{cases}; \quad u^{att} = \begin{cases} 0 & r < \sigma \\ u^{LJ} & r > \sigma \end{cases}; \quad (7)$$

$$u^{ref} = \begin{cases} u^{LJ} + \epsilon & r < r_{min} \\ 0 & r > r_{min} \end{cases}; \quad u^{att} = \begin{cases} -\epsilon & r < r_{min} \\ u^{LJ} & r > r_{min} \end{cases}; \quad (8)$$

As shown in Figure 5b,¹⁵ the BH separation leads to monotonically increasing values of A_2 with increasing density, while the WCA separation leads to the intuitive trend. The explanation for this behavior is similar to that for square well spheres, but the sharp change in the potential occurs close to σ in this case. For the BH separation, the attractive part of the potential changes sharply between σ and r_{min} . With increasing density, the fluctuations occur more narrowly close to σ , but that is the exact range where the potential is rapidly varying. Therefore, the second order term increases with increasing density over the density range of interest. On the other hand, the WCA attractive potential is flat, like the step potential. Therefore the tighter fluctuations simply give smaller influences and A_2 approaches zero in the usual way.

Figure 5. Peculiar behaviors of A_2 for (a) square well vs. two-step spheres and (b) the Barker-Henderson vs. WCA separation of the LJ potential.

These peculiar fundamental behaviors may seem like more than most people would like to know about TPT, but they make an important point. Details of the potential function and its connection with the macroscopic equation of state are readily apparent in the TPT contributions. Each detail can be interpreted. Understanding these connections lends greater insight into our nanoscopic understanding. This is why the TPT contributions really function as valuable "molecular fingerprints."

Association and Wertheim's Theory.

SPEADMD implements the same conception of solvation and association contributions as all other SAFT implementations: Wertheim's theory. The essence of Wertheim's theory is that complexation can be represented by small blister potentials as depicted in Figure 6a. In accordance with specified rules of steric hindrance and polarity, an analytical solution is obtained for the reaction equilibrium network.

At first glance, the blister potentials may seem unrealistic compared to the point charges that are applied in continuous potential models. Nevertheless, Figure 6b illustrates how Wertheim's theory can be used to interpolate state points simulated with a point charge model. This means that there is an equivalent blister potential that effectively characterizes the point charge model through the adjustment of the blister's position, diameter, and strength. To clarify, the point charge result was obtained by simulating a WCA reference model with point charges assigned according to the TraPPE model.¹⁶ The WCA reference without charges was also simulated and subtracted to obtain the isolated point charge effect. Similar results have been obtained for models of water and propionitrile.¹⁷ To be fair, the characterization of internal energy is not quite as accurate as that for compressibility factor, but the point is that Wertheim's theory has the capability to characterize more sophisticated models of molecular interactions, even if it appears to be excessively simple. One must simply interpret the parameters in Wertheim's theory as one would adjustable parameters in any model.

Figure 6. Schematic diagrams of association models for alcohols and water. (a) Wertheim's theory is characterized by discontinuous "blister" potentials. (b) Adapting Wertheim's theory to characterize the point charge contributions of continuous potential models shows that a convenient correspondence is feasible.

The SPEADMD model makes one possible exception to the way that Wertheim's theory is implemented relative to most implementations. We apply the square-root combining rule ("ECR")^{18,19} as an initial approximation before checking the influence of asymmetric association. Symmetric association describes molecules with equal numbers of donors and acceptors, like hydrocarbons, alcohols, and water. Asymmetric association includes ethers, acids, ketones, and esters. For ternary systems, the simplifications inherent in the ECR result in roughly 2x faster execution than the full implementation of Michelsen-Hendricks (MH).²⁰ For multicomponent systems, we can expect the enhanced speed of execution for ECR to be much less dependent on the number of components than the MH algorithm. ECR involves the solution of a single nonlinear equation regardless of the number of components, while MH requires solution of a non-linear solvation/association matrix. In applications to chemical processes or petroleum reservoir simulations that may involve 50 components or more, enhanced computational efficiency may justify some simplification of the molecular models. Hence, we reserve this approximation to initialize the general solution. When asymmetry is detected, we call the MH algorithm for the general solution. This is easily implemented through the combining rule,

$$\varepsilon_{ij}^{AD} = \left(\varepsilon_{ii}^A \varepsilon_{jj}^D \right)^{1/2} + k_{ij}^{AD} \left(\varepsilon_{ii}^A + \varepsilon_{jj}^D \right) \quad (9)$$

In summary, the role of Wertheim's theory in the SPEADMD model is effectively the same as it is in other SAFT implementations. Our experience with the method is that it provides a great improvement over molecular models based solely on physical interactions and that it can encompass more sophisticated molecular models of complexation to a large degree.

Stiffness and Branching

A careful review of Figure 3a shows a curve for tangent sphere chains in the long chain limit (TSinf). This curve is inferred from the SAFT reference contribution^{21,22} and the approximation of seven carbons per segment as estimated by Ghobadi and Elliott.¹⁰ Since the ordinate of the infinite alkane is slightly higher than the infinite tangent sphere chain, it means that the alkanes are slightly more repulsive. This can be explained by the added stiffness inherent in n-alkanes owing to the exclusion of dihedral angles less than roughly $\pm 45^\circ$. Gerek has studied

the dihedral angle distribution in n-alkanes according to the SPEADMD model, as shown in Figure 7.²³ Angles greater than roughly 100° are equally probable. There is no inhibition between the gauche and trans position as would be obtained with a conventional continuous potential. This behavior is attributable solely to the excluded volume effects of the intramolecular site diameters and bond lengths and bond angles. The effect is athermal in the same sense that all the interactions in the reference fluid are athermal. The reference potential is either infinite or zero at all temperatures. This stiffness is expected to enhance repulsion by making it more difficult to squeeze the chains into their most compact configuration,²⁴ and our observations appear to be qualitatively consistent with this expectation for long chains.

Figure 7. Dihedral angle distribution of n-alkanes in the SPEADMD model.

The effect of branching is illustrated cursorily by comparing the n-alkanes to polyisoprene in Figure 4a. Several additional examples of branched molecular structure have been presented by Vahid and Elliott.¹¹ The effect is to make the branched chain appear to be less repulsive, like the flexible chain. This is qualitatively consistent with recent findings of Marshall and Chapman.²⁵ On the other hand, it is not entirely clear why the polyisoprene appears to be less repulsive. Adding a CH₃ branch exposes a larger volume than adding a CH₂ link. For example, the SPEADMD volume of 2-methyl-butane is 0.0872 nm³ compared to 0.0855 nm³ for n-pentane. Therefore, the packing fraction shifts to the right even if the pressure is the same. This would not be a consideration for tangent sphere chain models. It should also be noted that branching affects the chain conformation. For example, polystyrene would show a different dihedral angle distribution function than n-alkanes, even with the simple athermal specification inherent in SPEADMD. The mingling of these effects might appear to be a disadvantage of SPEADMD to those who would like to identify specific effects of branching vs. stiffness. On the other hand, real molecules are not like tangent sphere chains; the real effects are commingled. Once again the "fingerprint" nature of molecular structure arises. The question then is whether the characterization of these effects inherent in SPEADMD is sufficiently accurate compared to real molecules. We cannot answer that question in great detail at this time. All we can say is that the qualitative behavior is reasonable.

Beyond Second Order TPT

Second order TPT is impressive for its accuracy at $T_r < 0.9$, but it would be desirable to characterize behavior approaching the critical point. Recent results for the third order^{8,15} and fourth order^{26,27} TPT contributions provide encouragement in this regard. For example, the statistical formulas for the square well fluid can be expressed in a relatively simple form:²⁷

$$A_3 = -\frac{\langle (M - \langle M \rangle_0)^3 \rangle_0}{6N} \quad (10)$$

$$A_4 = -\frac{\left[\langle (M - \langle M \rangle_0)^4 \rangle_0 - 3 \langle (M - \langle M \rangle_0)^2 \rangle_0^2 \right]}{24N} \quad (11)$$

These formulas show how the third and fourth order terms relate to higher order variance measures. Ghobadi and Elliott have shown how these higher order contributions exhibit trends in the ratio of A_i/A_{i-1} .²⁸ Briefly, plotting this ratio vs. density shows a roughly Gaussian function with its peak near the critical density. Extrapolating this trend from second order to infinite order leads to the following form for the Helmholtz energy,

$$\frac{(A - A^{ig})}{Nk_B T} = A_0 + \frac{A_1}{T} + \frac{A_2}{T^2} \left(1 + \left(\frac{2}{\mathcal{G}^2} \right) \left(\exp(\mathcal{G}) - 1 - \mathcal{G} - \frac{\mathcal{G}^2}{2} \right) \right) \quad (12)$$

Where $\mathcal{G}$ is the Gaussian function, $\mathcal{G} \equiv A_i/A_{i-1}T = \mathcal{G}_0 \exp\{-\mathcal{G}_1((\eta - \eta_0)^2 - \eta_0^2)\}/T$.

We can adapt this result to provide a semi-empirical extension to the SPEADMD model, treating $\mathcal{G}_0$, $\mathcal{G}_1$, and η_0 as adjustable parameters. Figure 8 illustrates the improvement that "Gaussian extrapolation" in this manner can have on binodal phase diagrams. Even better results are obtained when third and fourth order TPT contributions are known from simulation.²⁸

Figure 8. Binodal phase diagrams for n-alkanes. Dashed lines show the result of second order TPT and solid lines are obtained by "Gaussian extrapolation" of estimated TPT contributions beyond second order.

Softness and Compressed Fluid Density

One last detail of pure fluid behavior should be addressed before moving on to mixtures. Plots of compressibility factors along isochores can provide convenient insights about the nature of the molecular interactions and the manner of their characterization. Figure 9 illustrates several key points in this regard.²⁹ The intercept along the x-axis characterizes the saturation temperature, where typically $Z_L^{sat} < 0.01$. With a few simple trials, a well width of 1.7 is found to be the most satisfactory. A similar analysis yields 1.5 as the optimal well width for methane and 1.8 for ethane, where ethane is modeled with a two-site model.

It should also be noted that the y-intercepts on this plot characterize the reference fluid equation of state when an athermal reference is applied, as for step potentials. For spherical molecules, it is a simple matter to vary the site diameter until the hard sphere fluid matches extrapolations of the isochores.³⁰

Finally, it is clear from the high temperature behavior that an athermal reference fluid is not appropriate at temperatures above roughly 500K. At these high temperatures, the softness of the molecular interaction becomes apparent. It should also be noted that the pressures are extremely high under these conditions. If interest is primarily limited to conditions near ambient, the athermal reference should be satisfactory, but recent applications in tribology and oil and gas recovery are approaching conditions where these considerations may be of practical significance.

Characterizing the high temperature behavior with step potentials is a challenge. In principle, it should be feasible to add an additional step with a repulsive shoulder at a height of 300-500K and a diameter below the ambient temperature diameter. Preliminary efforts with this approach have not been promising, however.²⁹ More promising results have been obtained using the WCA reference fluid and continuous potentials.¹⁵

Long term, it seems likely that continuous potentials will show more promise for correlating dynamic properties like viscosity and inhomogeneities like vapor-liquid interfaces. Understanding these properties requires simulating the full potential instead of just the reference potential, at least at our current level of understanding. DMD loses its computational advantage in that case because the number of collision events rapidly increases with increasing sophistication of the potential model. Furthermore, DMD is not convenient for massive parallelization. Incorporation of these kinds of properties into the SPEADMD model would seem to require an ad hoc approach, in which the fundamentals are established using continuous potentials and correlating forms are applied to enable computations with SPEADMD.

Figure 9. Isochores of compressed fluid density for xenon. Solid points represent data from the NIST Webbook. Dashed lines represent correlation of the isochore at 2.92 g/cm³ and predictions of the isochore at 2.65 g/cm³ with square well spheres and three values of well width.

Mixtures

When the behaviors of energy and entropy become functions of composition, their characterization requires care and attention to subtle details. The results are rewarding, however, in that fundamental and challenging questions can be unambiguously resolved within the SPEADMD methodology. For example, the excess entropy of mixing has been an outstanding challenge for many years. Most treatments have relied on lattice theory, such as the Guggenheim-Staverman treatment that has been adopted in UNIQUAC and UNIFAC models or the Flory-Huggins treatment that is commonly applied in polymer solution theories. Owing to its athermal reference state, the "combinatorial" contribution to the excess entropy can be clearly isolated as a direct result of simulation within the SPEADMD model. The residual entropy contribution is a natural result of the attractive TPT contributions. A similar approach yields the excess internal energy. These results lead to characterization of mixing rules that can be traced directly to the results of molecular simulations.

With simulation-based mixing rules for each of the TPT contributions, applications to mixtures ought to be entirely predictive. While some predictions are quite accurate, others lead to the realization that a subtle assumption has been applied while characterizing pure fluids. The implications of relaxing this assumption and explicitly characterizing the required interactions leads to a fresh perspective on the broad scope of the general problem of characterizing molecular interactions. Large databases are required, and efficient protocols for translating the data into molecular insights.

Inference of the Excess Entropy from Off-Lattice DMD Simulations

One benefit of the athermal reference fluid in the SPEADMD model is that it is well-defined in the infinite temperature limit, while influences from the attractive part of the potential are exactly zero. This is not true for continuous potentials. The closest one might come would be to simulate something like the WCA reference fluid. Unfortunately, the internal energy contribution is not exactly zero in that case and it introduces variability such that precision is lost. As discussed below, precision is essential, and precision is exactly what DMD delivers.

DMD is an event driven algorithm that applies the exact solution for hard sphere collisions at every event. Therefore, the only imprecision that can arise is machine precision, which is in roughly the 14th significant figure. This imprecision is random such that the energy is conserved to roughly 13 significant figures regardless of duration, even for square well interactions. For reference interactions, energy is maintained at exactly zero throughout the simulation. Typical energy conservation for continuous potentials is roughly five significant figures, or less if an unshifted cutoff potential is applied. The precision of the cumulant that appears in the sum of the DMD virial is also exact, so estimates of the pressure are similarly precise.

During DMD reference initialization, overlaps are eliminated such that each site begins in a location of zero energy. Since all collisions are with hard interactions, overlaps are prevented, such that the energy is exactly zero at all times. The pressure is non-zero, of course, and varies substantially with packing fraction. A convenient characterization of pressure is in the form of the compressibility factor for the reference fluid, Z_0 . The Helmholtz energy is then,

$$\frac{(A_0 - A^{ig})}{RT} = \int_0^\eta (Z_0 - 1) \frac{d\rho}{\rho} \quad (13)$$

Noting that $U_0 = 0$, and $A_0 \equiv U_0 - T^*S_0$ gives, for reference potentials,

$$\frac{(S0 - S^{ig})}{R} = \int_0^{\eta} (Z0 - 1) \frac{d\rho}{\rho} \quad (14)$$

Therefore, the reference fluid entropy can be accurately determined by integrating the compressibility factor over a range of densities at constant composition. Figure 10 illustrates the procedure for a mixture of reference "n-octanol" and "water."³¹ Keep in mind that the reference fluid for this "water" is simply a hard sphere with a diameter of 0.27 nm. The excess entropy is the difference between the computed value by integration and the molar average. Vahid et al. found that the excess was largest when a small molecule was mixed with a large molecule, as in the case of octanol+water. This finding was not strictly a function of volume ratio, however. For example, if toluene is mixed with polystyrene, the excess would be small, even if the volume ratio was the same as that for octanol+water. In the case of toluene+polystyrene, precision is especially important because one encounters a small difference between two large numbers.

Apparently, small molecules like water, methane, etc. are rich in entropy, as reflected by their entropy departure function per unit volume. This leads to a valuable guideline for estimating the excess entropy. Vahid et al. defined an entropic solubility parameter in terms of the entropy density at a packing fraction of 0.4.³¹ The framework of Scatchard-Hildebrand theory could then be used to characterize the excess entropy for most mixtures to a fair degree of accuracy. Owing to the relationship between excess entropy and the reference contribution to the Helmholtz energy, this framework also serves as a mixing rule for the reference contribution. Note that Scatchard-Hildebrand theory is consistent with a van der Waals 1-fluid (vdW1) mixing rule. While many mixing rules have been proposed over the years, and vdW1 is one of the oldest and simplest, this rule has been inferred as a straightforward interpolation of simulation results. The rule has been tested for mixtures of high and low molecular weight, branched and ring compounds, and every combination.³¹ In every case, it is clearly the appropriate interpolating formula. If applying this rule does not fit experimental phase equilibrium data for some reason, then explanations other than the mixing rule for athermal excess entropy should be sought.

It should be noted that these observations of excess entropy have been formulated at constant packing fraction, not constant pressure. Transforming to a constant pressure basis would be best achieved with a consistent attractive contribution. We discuss analysis of mixing rules for the attractive contributions next. Nevertheless, we can gain some insight about the nature of these observations by comparing to the mixing rule for $A0$ in the SAFT model as originally formulated by Chapman et al.²¹ Chapman's mixing rule can also be expressed on a basis of constant packing fraction. It may be gratifying to admirers of the SAFT model to know that Chapman's mixing rule has proved to be qualitatively correct in all the cases where we tested it. Quantitatively, it tends to underestimate the magnitude of the excess entropy by about 30%. Chapman's mixing rule is largely driven by differences in the segment diameter through the hard sphere mixing rule of Mansoori et al.³² Considering the complex manner in which segment diameters and segment numbers are correlated in the SAFT framework, it is noteworthy that Chapman's mixing rule works as well as it does.

Figure 10. Thermodynamics of reference fluid mixtures. (a) Compressibility factor, where points indicate state points simulated over 21 packing fractions at each composition. (b) Excess entropy at several packing fractions. The dashed line indicates the best fit obtained from a mixing rule similar to the Flory-Huggins model. The solid line indicates regression assuming the van der Waals 1-fluid mixing rule (aka. Scatchard-Hildebrand).

Inference of the Excess Energy

The same simulations that lead to inference of A_0 also lead to A_1 and A_2 . In fact, the attractive contributions are somewhat more direct because they require no integration for their interpretation. We have also analyzed the A_1 and A_2 contributions for mixtures over a broad range of molecular types and combinations.^{33,34} Typical results for A_1 are illustrated in Figure 11. The A_2 contributions are difficult to interpret because they exhibit a lack of precision owing to their small magnitude in the liquid range and their statistical definition. The A_1 contributions, however, show clear adherence to the vdW1 mixing rule. Perhaps this is not surprising given that it is the first order attractive contribution and the vdW analysis is effectively first order. Even so, it is reassuring to have this validation.

When taken in conjunction with the A_0 term, we are led to a couple of interesting observations. Since both terms follow vdW1 mixing, it is feasible to compensate for deviations in one contribution by an equal and opposite deviation in the other term, by correlating the molecular energetic binary parameter k_{ij} for example. This cancellation of errors would be satisfactory over some finite range of packing fractions. On a related note, we should observe that k_{ij} has no basis for being temperature dependent, as is often assumed. On the other hand, Figure 11 does appear to show some dependence on density. Since higher temperatures generally lead to reductions in density, it may be possible to treat k_{ij} as temperature dependent over a very narrow range, but there is no theoretical basis for suggesting whether the temperature dependence should be linear with temperature, reciprocal temperature, or exponential temperature.

Figure 11. Excess $A1$ contribution, reflecting excess energy in the SPEADMD model. Points correspond to various packing fractions. Curves are regressed according to the vdW1 mixing rule assuming a constant value for the molecular energetic binary parameter, k_{ij} .

Confounded Site-Site Interaction Energies

The discussion of the molecular energetic binary parameter (k_{ij}) in the previous section touches on a subtle thread that unravels into a rather large mess. Nevertheless, the nature of the problem before us and the essential need for TPT (or some equally efficient theory) cannot be comprehended without full illumination of this issue. Completely resolving this issue could involve an army of engineers working closely together for years. The crux of the issue is this: molecules do not interact, sites do.

To clarify, the "molecular" energetic binary parameter (k_{ij}) is not fundamental. The proper parameter to characterize is the site-site interaction parameter (k_{ij}^{ss}). We can envision site-site interaction parameters as analogous to UNIFAC parameters. Through a limited number of site-site parameters, we can infer an infinite number of molecule-molecule interactions. But there are two key distinctions between the site-site k_{ij}^{ss} and UNIFAC parameters. First, there are fewer of them because the k_{ij}^{ss} matrix is symmetric whereas $a_{ij} \neq a_{ji}$ in UNIFAC. Second, each k_{ij}^{ss} is coupled to many, many more properties than any UNIFAC parameter. These seeming contradictions derive from the fundamental nature of the interaction potentials and the large number of properties that are implied by the potential functions and Newton's laws of motion. It is a blessing from the perspective of theory based predictions of engineering properties, and a curse from the perspective of the engineers who must infer the optimal estimates of these miraculous parameters.

To understand k_{ij}^{ss} better, consider two molecules: n-pentanol and n-pentane. We can optimize the site-site parameters as discussed in the pure components section above. Next, suppose we would like to predict the vapor-liquid equilibrium (VLE) of a mixture of n-pentanol and n-pentane. We would not be surprised to see a discrepancy, which we would normally correlate with k_{ij} . But now we recognize that k_{ij} is not fundamental, and we should more properly

correlate the k_{ij}^{ss} for some pair(s) of sites. The OH-CH₂ interaction would be a logical place to start. So we correlate k_{ij}^{ss} for OH-CH₂. Unfortunately, the OH-CH₂ interaction also exists in the n-pentanol molecule itself. While characterizing the site-site parameters for pure fluids, we implicitly assumed $k_{ij}^{ss} = 0$. The mixture data prove that assumption to be wrong. Therefore, the pure component properties need to be revisited, which alters the mixture predictions, which alters the $\{k_{ij}^{ss}\}$, which returns us to the pure component properties in an iterative cycle until self-consistency is achieved. Then we turn to liquid-liquid equilibria. Then we turn to viscosity. In other words, the molecular scale k_{ij}^{ss} is confounded with the macroscopic data in such a way that its inference becomes quite a challenge.

Now imagine trying to perform this optimization without a theory like TPT. Molecular simulation of a well-equilibrated single state point requires hours on a multicore+GPU processor. State points over a range of densities and temperatures are required for pure fluids, and mixtures add the dimension of composition. And that covers just 1-2 compounds. A single iteration for a single instance of $\{k_{ij}^{ss}\}$ for 500 pure compounds and a small number of mixtures could easily require CPU years, even with thousands of dedicated nodes.

Through TPT, no simulation is required in order to perform this optimization, simply a resummation of the weighted averages. Therefore, the same computation requires minutes in the SPEADMD framework. Optimization of 500 compounds with roughly hundreds of trial step depths and a single instance of $\{k_{ij}^{ss}\}$ takes a few hours. Considering the specific nature of molecular "fingerprints," the implications of stiffness, branching, and ring formation, and the large scope of the problem, engineers should be urged to abandon more and more empirical correlations based on less and less verifiable physical hypotheses. TPT makes it feasible to perform a similar quality of regression in a similar timeframe while maintaining verifiable physical paradigms. This approach would involve a large collaborative project with shared databases and programs.

There is some good news when it comes to engaging the problem of confounded k_{ij}^{ss} . First, Jackson and coworkers have made a start within the SAFT-VR framework. So far, their scope has been limited to using $\{k_{ij}^{ss}\}$ to improve vapor pressure correlations. Given the broad scope of the complete problem, it is reasonable to start with a limited scope. Second, the validity of the vdW1 mixing rule leads to an identity between the k_{ij} and k_{ij}^{ss} that may prove useful.³⁴

$$A1 = \frac{1}{Nk_B} \sum_i \sum_j \sum_m \langle N_{ijm} \rangle \sqrt{u_{im} u_{jm}} (1 - k_{ij}^{ss}) = \frac{\sum_i \sum_j x_i x_j (b_i b_j)^{1/2} \sqrt{A1_{ii} A1_{jj}} (1 - k_{ij})}{\sum_i x_i b_i} \quad (15)$$

The significance of this relation is that $\{k_{ij}\}$ can be compiled by evaluations at the molecular level, as has been done traditionally, and used as a surrogate for the experimental data. In this way, experimental data that might involve hundreds of state points are reduced to a single value of k_{ij} . Once again, TPT provides tremendous leverage in efficiency. One caveat with this approach is that, technically, the pure component $A1$ values change as $\{k_{ij}^{ss}\}$ evolves. Therefore, this relation is only valid if changes to $A1$ are small. Nevertheless, it should be feasible to enhance efficiency by performing a few iterations of $\{k_{ij}^{ss}\}$ before re-evaluating $\{k_{ij}\}$. Once a reasonable set of $\{k_{ij}^{ss}\}$ has been obtained, maintenance as more compounds or data are added should be feasible without frequent re-evaluation of $\{k_{ij}\}$.

Molecular Interactions and Databases

The role of experimental data in characterizing molecular interactions has arisen throughout this review. For pure fluids, vapor pressure, saturated liquid density, and compressed fluid density are important properties that should not be overlooked. Heats of vaporization could also be included, but are somewhat redundant with vapor pressure data. Recently, surface tension capability has been developed,¹⁷ and should be included for most future efforts. For mixtures, VLE and LLE are obviously important, but solid-liquid equilibria (SLE) could also be valuable. The inclusion of surface tension for pure fluids coincides with the capability to predict interfacial tensions in mixtures. Interfacial properties can be very sensitive to the molecular interactions, and they may provide great insight into the validity of molecular scale assumptions. These are early days for the treatment of interfacial properties, but ultimately they should be included in conjunction with phase equilibrium data.

When it comes to molecular simulation models, however, the simulation data can also be valuable. These data can be time consuming to generate and occasionally subject to varying interpretations. Furthermore, one researcher might have an additional perspective on interpreting the simulation results that was not anticipated by the original developers. All of these scenarios could be facilitated by a more progressive approach to data in the internet age.

Experimental data for phase equilibria

Unfortunately, various developers of molecular models use various databases, and this undermines progress in the field. It is difficult to know whether a particular model development provides a general step forward when it is evaluated with respect to a very specific database. One explanation of this situation is that database compilation can be a time consuming task. Projects like DETHERM, TRC, and DIPPR are understandably protective of the time and effort they have put into compiling data and cannot be expected to share their entire efforts openly. Nevertheless, the recently available THERMOLIT facility through the TRC does make it much easier to access the data in the open literature. With a small change in attitude toward the publication of model developments, a new standard could be set that substantially improves our rate of progress.

Admittedly, it requires an altruistic approach to resolve this problem. As one example, Danner and Gess have made an initial effort with a database for binary VLE that spans a range of compound polarity and combinations.³⁵ Unfortunately, this database has rarely formed the basis for comprehensive model comparisons.³³ The inclusion of carboxylic acids in the database is somewhat problematic because these have been challenging to characterize from a molecular perspective.³⁶ Nevertheless, we cannot conclude that standardizing databases is an impossible task.

Another example of such an effort was put forth in the context of the SPEADMD model for pure fluid properties.¹ The supplemental data of that article include vapor pressures and saturated liquid densities for over 500 compounds ranging from 0.0001 MPa to the critical point, where experimental data are available. One way of implementing this database might be for particular developers to limit the scope to families of compounds during a particular development effort, primary amines, or perfluorocarbons, for example. Then all of the compounds and data for that family should be included in the effort. It might be reasonable to extend the database and include additional compounds, but then the additional data should also be published as supplemental data. In this way, systematic progress can be achieved. Absent such an approach, one must question whether the scientific method is being faithfully applied.

A similar effort might solve the problem for mixture phase equilibria. Instead of ignoring the Danner-Gess database, it should be expanded such that families of interactions are well-represented, instead of lightly sampled. For example, alcohols and amines exhibit acid-base interactions, often leading to negative deviations from ideal solution behavior.³⁴ These interactions vary with composition and steric hindrance depending on the locations of the acid-base sites and the sites around them. It would be reasonable for model developers to limit their interest to a subset of a general database that focuses on such interactions. To advance the process, we are including a database of experimental data for these types of systems. Going forward, it should be the expectation that model developers submit their databases as supplemental data whenever the data are not openly available elsewhere.

Molecular Simulation Data

The *Journal of Chemical and Engineering Data* has recently called for inclusion of molecular simulation data in the chemical engineering literature. This is a positive step forward and should pave the way for progress, but there are issues to be resolved. In principle, scientific results should be reproducible, but subtle distinctions may undermine that reproducibility in the context of molecular simulations. As an example, consider the treatment of truncation of continuous potential functions for dispersion forces, and the various ways in which it might be approached. Truncation is advisable because continuous potentials decay continuously to infinity and the simulation time increases as the cube (roughly) as the range of the potential. One recent approach is to apply Ewald sums to the dispersion forces, which extrapolates a periodic image of the dispersions to infinity. Another approach is to apply tail corrections, based on the assumption that pair distribution functions have approached a uniform value outside the range of truncation, typically 1.2-1.4 nm. These approaches become problematic in the simulation of interfaces because the spherical asymmetry on which they are based is no longer valid. This fact leads to multiple approaches to correct for asymmetry, and consequent inconsistency. Another approach is to apply an interpolating function from the truncation value to zero, in order to avoid the fundamental failure to conserve energy when tail corrections are applied. Another approach is to truncate the potential and shift it such that its value goes exactly to zero at the truncation point, but this approach ignores the discontinuity in force, which can be expected to strongly alter the pressure. Finally, one may shift both the force and the potential to zero to provide continuous functions for each, but this adds a term proportional to distance, r , to the original potential function. All of these implementations may be referred to as "the Lennard-Jones potential," for example, but all give different results. By the way, step potentials suffer no such ambiguity.

To ensure reproducibility, sufficient simulation results must be publicly available to ensure that competent simulation personnel can implement the procedures described and obtain the same results. Ultimately, this may require that the programs used to generate the simulation results must be publicly accessible. As a demonstration of this approach, supplemental data of the article describing SPEADMD optimization¹ also includes simulation results and a protocol that facilitates access to the data through simple machine commands. These data include the site diameters and bond lengths applied during the simulation, the integrated values of molecular volume averaged over all densities, the step depths applied, and the resultant TPT contributions for every molecular structure, for example.

Transferable vs. Customized Potentials

Transferable potential models have the advantage of being predictive, but suffer in accuracy when compared to potential models developed for a specific compound. Transferability is basically the assumption that the potential surrounding a CH₂ site in the middle of a n-pentane molecule is identical to that in the middle of a n-pentadecane molecule. For n-alkanes, this assumption suffices to achieve roughly 10% error in predicted vapor pressures. Nevertheless, a customized potential can generally achieve less than 3% error in correlated vapor pressures. Of course, one would like to have the best of both worlds.

Fortunately, TPT facilitates the development of both transferable and customized potentials. The statistical averages are compiled in such a way that new instances of the well depths can be imposed and all the thermodynamic properties recomputed. By imposing potential functions selected from transferable site types to multiple families of compounds simultaneously, the optimal transferable potentials are obtained. With an alternate protocol, the molecular size and step depths can be optimized individually for a customized thermodynamic model. Once again, it is difficult to imagine how such efficiency could be achieved without a level of theory comparable to TPT.

Table 2 shows a summary of the deviations in vapor pressure and saturated liquid density achieved for a training set of 393 compounds. One particular advantage of the customized models is that the vapor pressure deviations are sufficiently small that VLE correlations are not overly biased by deviations in vapor pressure.

Table 2. Summary of vapor pressure and saturated liquid density deviations for 393 compounds sampling 27 different families of compounds.

Type	%PAAD	%PBias	%pAAD	%pBias
Transferable:	10.2	0.7	3.5	1.3
Customized:	1.9	-0.1	1.2	-0.1

Conclusions.

The recurrent theme throughout our experience with DMD/TPT has been that universal equations are quite uncommon. It seems that every molecular structure possesses a distinct degree of stiffness, branching, and cyclic character. Furthermore, the TPT contributions reflect these distinctions like "molecular fingerprints." In this way, it does not seem feasible to superpose the TPT contributions for a single molecular structure on all other molecular structures and achieve optimal predictions, whether the base molecular structure is flexible tangent sphere chains, or slightly stiffer fused sphere alkanes.

On the other hand, certain trends are recurrent, and these make it feasible within the TPT framework to characterize intermolecular potentials while simultaneously achieving timely engineering predictions. Each molecular structure is distinct, but trends in molecular weight lead to asymptotic behavior in the long chain limit that can be easily interpolated for any molecular weight. The TPT framework dictates that only the reference fluid needs to be rigorously simulated, while the thermodynamics of the full potential can be computed quantitatively through the theory. The range of the reference potential is roughly one third that of the full potential, implying faster simulations by a factor near 30. Also, since the attractive part of the potential is the perturbation, its influences can be computed without further simulation leading to optimization of this part of the potential with engineering efficiency. The result is transferable potential models that permit predictions of vapor pressure with accuracy near 10%, if the molecular structures are simple enough to conform to transferability. When customized

potential models are developed for each molecule, the accuracy approaches 2% for vapor pressure and 1% for saturated liquid density correlation. Incorporating the Gaussian extrapolation technique leads to characterization of the critical point that approaches the experimental values, while providing more accurate characterization of the liquid density over the entire temperature range than cubic equations of state. Although the approach has its foundation in specific descriptions of the site-site molecular interactions, the summative equations are similar in form to other engineering equations and can be easily incorporated into existing thermodynamics tools and chemical process simulators.

In this context, the findings of DMD/TPT hint at just one step in a natural evolution. Much effort has been expended over the past 25 years to construct universal predictive frameworks based on a reference fluid of flexible tangent sphere chains. Those efforts are reaching maturity and useful capabilities are being achieved. Nevertheless, findings reviewed here, along with other findings related to flexibility and branching, suggest that the limits of this reference fluid are being approached. The DMD/TPT approach is available in the form of the SPEADMD model to incorporate a higher degree of resolution in characterizing these effects, but molecular models will continue to evolve. For example, united atom models cannot address the differences between isotactic and syndiotactic polyvinylchloride. Therefore, explicit atom models will be required to address the entire scope of possible problems. The challenge is to assess the level of detail that is required to solve a given problem without sacrificing computational efficiency. An advantage of the DMD/TPT approach is that it substantially broadens the scope of molecular detail while sacrificing very little in computational efficiency.

Altogether, these findings suggest the need for certain protocols in order to achieve systematic progress in thermodynamic modeling. The experimental data that have been used to characterize models based on the flexible tangent sphere chain reference are the same data required for characterizing more sophisticated molecular models. Unfortunately, the manners in which these databases have been managed so far do not promote progress. A related comment applies to publication of molecular simulation data. It is not entirely unfair to question the reproducibility of certain scientific findings under the current protocols. We have made suggestions for ways that the community may collaborate more effectively and initiated efforts as a show of our good faith.

Molecular modeling of engineering properties continues to be a fascinating subject. At the molecular level, disciplinary lines are blurred. The same molecular interactions that control VLE and interfacial tension also control plasticizer compatibility, composite adhesion, protein aggregation, and membrane selectivity. Engineering properties are much more distinctly measurable and fundamental in many cases, however. The fluid phase engineering community has an opportunity to articulate potential models that are continuously improving and demonstrably better than other alternatives. Experience with DMD/TPT shows how this development can leverage engineering expertise in optimization and data management to achieve this goal, while serving conventional engineering applications with timely and seamless integration.

Acknowledgements

Initial funding for the SPEADMD.org project was received from the US National Science Foundation. Additional support by Chemstations, Inc. is also gratefully acknowledged.

Reference:

- ¹ A.D. Sans, A.V. Vahid, and J.R. Elliott, *J. Chem. Eng. Data* **59**, 3069 (2014).
- ² M. S. Wertheim, *J. Stat. Phys.* **42**, 459 (1986).
- ³ J-X. Liu and J.R. Elliott, *Ind. Eng. Chem. Res.* **35**, 2369 (1996).
- ⁴ W. G. Chapman, K. E. Gubbins, G. Jackson et al., *Fluid Phase Equilib.* **52**, 31 (1989).
- ⁵ G. A. Chapela, L. E. Scriven, and H. T. Davis, *J. Chem. Phys.* **91** (7), 4307 (1989).
- ⁶ J. Cui and J.R. Elliott Jr., *J. Chem. Phys.* **116**, 8625 (2002).
- ⁷ O. Unlu, N.H. Gray, Z.N. Gerek et al., *Ind. Eng. Chem. Res* **43** (7), 1788 (2004).

8 Thomas Lafitte, Anastasia Apostolakou, Carlos Avendaño et al., The Journal of Chemical
 Physics **139** (15), 154504 (2013).
 9 J.A. Barker and D. Henderson, J. Chem. Phys **47**, 2856 (1967).
 10 Ahmadreza F Ghobadi and J Richard Elliott, Fluid Phase Equilibria **306** (1), 57 (2011).
 11 A. Vahid and J.R. Elliott, Fluid Phase Equilibria **351**, 61 (2013).
 12 Z.N. Gerek and J.R. Elliott, Ind. Eng. Chem. Res. **49**, 3411 (2010).
 13 J. A. Barker and D. Henderson, J. Chem. Phys. **47**, 4714 (1967).
 14 J.D. Weeks, D. Chandler, and H.C. Andersen, J. Chem. Phys **54** (12), 5237 (1971).
 15 Ahmadreza F Ghobadi and J Richard Elliott, The Journal of Chemical Physics **139** (23),
 234104 (2013).
 16 Jeffrey J Potoff and J Ilja Siepmann, AIChE Journal **47** (7), 1676 (2001).
 17 Ahmadreza F Ghobadi and J Richard Elliott, The Journal of Chemical Physics **141**,
 094708 (2014).
 18 J. R. Elliott, Ind. Eng. Chem. Res. **35**, 1624 (1996).
 19 Georgis M. Kontogeorgis, Michael Michelsen, M., Georgis K. Folas et al., Ind. Eng.
 Chem. Res. **45** (14), 4855 (2006).
 20 M.L. Michelsen and E.M. Hendriks, Fluid Phase Equilibria **180**, 165 (2001).
 21 W.G. Chapman, K.E. Gubbins, G. Jackson et al., Industrial & engineering chemistry
 research **29** (8), 1709 (1990).
 22 J. Gross and G. Sadowski, Ind. Eng. Chem. Res. **40**, 1244 (2001).
 23 Z.N. Gerek, Ph.D. Dissertation, The University of Akron, 2006.
 24 T. van Westen, B. Oyarzúna, T.J.H. Vlugt et al., Molecular Physics **112** (7), 919 (2014).
 25 B.D. Marshall and W.G. Chapman, J. Chem. Phys. **138**, 174109 (2013).
 26 Yu T Pavlyukhin, Journal of Structural Chemistry **53** (3), 476 (2012).
 27 S. Zhou and J.R. Solana, J. Chem. Phys. **138**, 244115 (2013).
 28 A.F. Ghobadi and J.R. Elliott, J. Chem. Phys. **Submitted** (2015).
 29 E. Moharrerri and J Richard Elliott, in *Midwest Thermodynamics and Statistical
 Mechanics Symposium* (Minneapolis, MN, 2012).
 30 J.R. Elliott and C.T. Lira, *Introductory Chemical Engineering Thermodynamics, 2ed.*
 (Prentice-Hall, Upper Saddle River, NJ, 2012).
 31 A.V. Vahid, N.H. Gray, and J Richard Elliott, Macromolecules **47**, 1514 (2014).
 32 G.A. Mansoori, N.F. Carnahan, K.E. Starling et al., J. Chem. Phys **54**, 1523 (1971).
 33 A. Vahid, A.D. Sans, and J. R. Elliott, Ind. Eng. Chem. Res. **47**, 7955 (2008).
 34 J.R. Elliott, A. Vahid, and A.D. Sans, Fluid Phase Equilibria **256** (1-2), 4 (2007).
 35 R.P. Danner and M.A. Gess, Fluid Phase Eq. **56**, 285 (1990).
 36 A. Vahid and J.R. Elliott, AIChE Journal **56** (2), 485 (2010).