Humboldt Research Fellowship

Richard Messerly[1]

[1]*Thermodynamics Research Center (TRC), National Institute of Standards and Technology (NIST), Boulder, Colorado, 80305, USA*

ABSTRACT

1 Description on website

The current state of research should first be briefly described and supported by approximately five revelant publications from the research area (one page max).

The outline should focus on a clear description of the questions you intend to address in your research, their originality and significance for the advancement of the research field (approx. two pages).

Furthermore, the academic methods to be used to achieve these goals should be clearly described and referenced, if appropriate (approx. two pages).

The research outline should comprise approximately five pages in total (including references). Should you significantly exceed this length, you may be asked to cut it down to approximately five pages.

For the purposes of evaluation it must be clearly demonstrated that you yourself have drawn up the main contents independently and agreed them beforehand with your host. Any contents contributed by the host institute must be attributed accordingly.

2 Rough draft: Self-consistent fundamental equations of state and force fields of "technical accuracy"

The design of efficient and reliable chemical processes requires accurate prediction of thermophysical properties over a wide range of temperatures and pressures. Fundamental equations of state (FEOS), such as those based on the Helmholtz free energy (e.g. REFPROP), are a powerful approach for estimating pressure, density, temperature ($P\rho T$) behavior and caloric properties, such as internal energy (U) and isochoric/isobaric heat capacities (c_v and c_p, respectively). For example, the state-of-the-art FEOS in terms of the Helmholtz free energy is:

$$\frac{A}{R_g T} \equiv \alpha = \alpha^{\mathrm{ig}} + \alpha^{\mathrm{r}} \tag{1}$$

where A is the Helmholtz free energy, R_g is the gas constant, T is temperature, α is reduced Helmholtz free energy, α^{ig} and α^{r} are the ideal gas and residual contributions to the reduced Helmholtz free energy. As α^{ig} is typically obtained using first principles (*ab initio*) calculations, the primary focus of FEOS development is modeling α^{r}. A semi-empirical model for α^{r} with a large number of non-linear fitting parameters (between 50 and 100) is optimized to reproduce available experimental data.

An FEOS correlation is considered to be of "technical accuracy" when it is reliable over the entire fluid region of technological interest. Unfortunately, most compounds do not have sufficient *reliable* experimental data for a diverse set of thermodynamic properties covering a wide range of $P\rho T$ conditions to develop a technical accuracy FEOS. Mixture experimental data are even more scarce and of questionable quality. Therefore, FEOS predictions can result in large errors when extrapolated to higher temperatures, pressures, and compositions than the conditions for the training data. Improvement in an FEOS at high temperatures and pressures necessitates additional data for those conditions. The lack of experimental data at high temperatures and pressures, especially, is attributed to the inherent safety, cost, and complexity of such experiments.

By contrast, molecular simulation (i.e. Monte Carlo, MC, and molecular dynamics, MD) methods at high temperatures and pressures do not suffer from any of these limitations. For this reason, molecular simulation results at extreme temperatures and pressures can supplement experimental data.[?,?,?,?] Recently, the host institute successfully applied this "hybrid data set" approach to dramatically extend the FEOS' range of applicability for several compounds.[?,?,?,?,?]

The simulation values that are typically included in hybrid data sets are derivatives of the residual Helmholtz free energy with respect to inverse temperature and/or density:[?,?,?,?]

$$A^{\mathrm{r}}_{xy} R_g T \equiv (1/T)^x \rho^y \frac{\partial^{x+y} A^{\mathrm{r}}}{\partial (1/T)^x \partial \rho^y} \tag{2}$$

The inclusion of A^{r}_{xy} simulation results greatly improved the extrapolation of the FEOS to extreme temperatures and pressures. Note that *ms2*,[?] the open-source simulation package developed by Jadran Vrabec's group, already computes A^{r}_{xy}. Therefore, the infrastructure is already in place to implement the hybrid data set approach at the Technische Universitat Berlin.

To improve the performance of this so-called "hybrid data set" approach, it is imperative that the force field be transferable over different $P\rho T$ conditions. In a previous study, we demonstrated the poor transferability to high pressures of the popular united-atom Mie *n*-6 potential (of which the traditional Lennard-Jones 12-6 is a subclass).

Therefore, we propose that more theoretical and flexible potential models be considered for developing FEOS. Specifically, we propose the development of extended Lennard-Jones (ex-LJ) force fields to improve the fundamental equations of state at high pressures.

The primary limitation of this approach is that most force fields are not "transferable" over a wide range of $P\rho T$ conditions. For example, in a previous study, we demonstrated the poor transferability to high pressures of the popular united-atom Mie *n*-6 potential (of which the traditional Lennard-Jones 12-6 is a subclass). The primary reason for this is that the optimal value of *n* to reproduce saturated vapor and liquid properties is typically greater than 12 (e.g. the Potoff 16-6 potential). Unfortunately, $n \gg 12$ is not theoretically justified and leads to an overly repulsive potential at high densities.

The most general expression for the extended Lennard-Jones non-bonded potential is:

$$u_{\mathrm{nb,exLJ}} = \sum_m C_m r^{-m} \tag{3}$$

where *m* are integer values and C_m are the coefficients for the corresponding r^{-m} terms. Note that the traditional Lennard-Jones 12-6 potential is obtained if C_6 and C_{12} are negative and positive, respectively, while all other C_m values are zero. Traditionally, there are only two to four non-zero C_m values, e.g., 16-6, 12-8-6, 14-12-8-6. The primary reasons for this are to simplify the parameterization and to avoid over-fitting if the data are insufficient to characterize more parameters. In principle, however, any number of C_m terms could be non-zero.

Two body potentials developed from *ab initio* values typically utilize several fitting parameters because of the high information content in *ab initio* calculations at different intermolecular distances.

The r^{-6}, r^{-8}, and r^{-10} terms can be derived from dispersive interactions and, therefore, it seems most reasonable for C_6, C_8, and C_{10} to be negative. However, it is unclear how well a 12-10-8-6 potential would perform if only the C_{12} coefficient is positive. Therefore, from an empirical standpoint, it is possible to allow all coefficients to be either positive or negative.

The ex-LJ 12-10-8-6 potential is more flexible than the two-parameter (ε and σ) LJ 12-6, and more theoretically justified than the three-parameter Mie n-6, particularly when $n \gg 12$. However, it has not been tested as extensively as the LJ 12-6, Mie n-6, and exponential-6 potentials. Although the ex-LJ was proposed over two decades previously, the main reason for the lack of popularity is the additional complexity in parameterizing the ex-LJ potential. Due to the additional model parameters of the ex-LJ, the traditional approach of performing molecular simulation with each proposed parameter set (θ) is computationally infeasible when several C_m terms are non-zero.

To facilitate parameterization of ex-LJ potentials, we propose the use of Multistate Bennett Acceptance Ratio (MBAR) combined with basis functions (Φ). Previous publications demonstrate that MBAR-Φ reduces the computational cost by several orders of magnitude compared to direct molecular simulation. The combined MBAR-Φ approach estimates ensemble averages for MBAR has been shown to be a reliable approach for reweighting conf has been shown to greatly accelerate force field parameterization

MBAR is a statistical method that reweights configurations sampled with a reference force field(s) to predict ensemble averages for a non-simulated force field. MBAR requires "recalculating" the non-bonded energies (and sometimes the forces, depending on the property of interest) for each configuration sampled. Basis functions greatly accelerate the cost of this recalculation step. The total non-bonded internal energy ($U_{\text{nb,total}}$) is computed from $\sum_{i=1}^{N_{\text{sites}}-1} \sum_{j=i+1}^{N_{\text{sites}}} \sum_m C_{m,ij} r_{ij}^{-m}$, where N_{sites} is the number of interacting sites in the system, $C_{m,ij}$ is the C_m term for the ij interaction and r_{ij} is the intermolecular distance between sites ij. Since there is a linear relationship between $U_{\text{nb,total}}$ and Equation BLANK,

the ex-LJ potential is amenable to basis functions. Rather than storing the configurations of all N molecules, basis functions store for different θ ($C_{m,ij}$) by storing the $\sum_{i=1}^{N_{\text{sites}}-1} \sum_{j=i+1}^{N_{\text{sites}}} r_{ij}^{-m}$ contributions for different values of m. Basis functions are amenable to the extended Lennard-Jones potential. For example, the

Basis functions store the contributions from the r^{-m} terms for the nonbonded energies and forces

The primary question of this study is whether or not it is possible to develop a pairwise additive force field of "technical accuracy." Specifically, if enough terms are included in Equation BLANK (ex-LJ), can the force field fit higher order derivatives of the Helmholtz free energy over the entire fluid region of technological interest. Some intermediary questions that will be addressed are, which terms of Equation BLANK provide the greatest improvement in the force field? Without over-fitting, how many terms are required?

As the ex-LJ has received relatively little attention in the literature, some additional questions are: which coefficients should be negative? To facilitate fitting and physical interpretation, should traditional parameters such as ε, σ, etc. be used to describe the ex-LJ? What type of combining rules should be used when more than two terms are included in the expansion?

In previous hybrid data set studies, the LJ 12-6 parameters were optimized with VLE properties. The two reasons for this practice are the availability of VLE data and the difficulty of optimizing force field parameters. If several model parameters are included in Equation BLANK, the traditional approach of force field parameterization, namely, trial-and-error optimization is not computationally feasible. Furthermore, it is unlikely that VLE data alone can provide a unique set of parameters. By contrast, derivatives of Helmholtz free energy provide high information content regarding the non-bonded potential. Unfortunately, reliable A_{xy}^{r} values require an accurate FEOS. For this reason, we propose an iterative hybrid data set approach where the number of non-zero C_m terms increases at each iteration.

$$A_{xy}^{\mathrm{r}}(\boldsymbol{\theta}) = F_i[\langle U(q_{\mathrm{ref}}, \boldsymbol{\theta})^i \rangle_{\mathrm{MBAR}}] + F_j[\langle U(q_{\mathrm{ref}}, \boldsymbol{\theta})^j \rangle_{\mathrm{MBAR}}] \tag{4}$$

where q_{ref} are configurations sampled using the reference force field(s), $\langle \rangle_{\mathrm{MBAR}}$ are ensemble averages estimated using MBAR, and F_i and F_j are functionals that depend on different powers of the internal energy ($U^{i,j}$).