

# Humboldt Research Fellowship

Richard Messerly<sup>1</sup>

<sup>1</sup>*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 relevant 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:

$$\frac{A}{R_g T} \equiv \alpha = \alpha^{\text{ig}} + \alpha^{\text{r}} \quad (1)$$

where  $A$  is the Helmholtz free energy,  $R_g$  is the gas constant,  $T$  is temperature,  $\alpha$  is reduced Helmholtz free energy,  $\alpha^{\text{ig}}$  and  $\alpha^{\text{r}}$  are the ideal gas and residual contributions to the reduced Helmholtz free energy. As  $\alpha^{\text{ig}}$  is typically obtained using first principles (*ab initio*) calculations, the primary focus of FEOS development is modeling  $\alpha^{\text{r}}$ . A semi-empirical model for  $\alpha^{\text{r}}$  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 FEOS often require additional fitting parameters although experimental data are even more scarce and of questionable quality. Therefore, FEOS predictions can result in large errors when extrapolated to higher temperatures and pressures than those of 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.<sup>?,?,?,?,?</sup> Recently, the host institute, in collaboration with BLANK, successfully applied this “hybrid data set” approach to dramatically extend the FEOS’ range of applicability for several compounds.<sup>?,?,?,?,?</sup>

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:<sup>?,?,?,?,?</sup>

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

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

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). This deficiency in the force field causes the FEOS to be inconsistent with the force field.

To improve this method, we propose that more theoretical and flexible non-bonded potential functions be considered for developing FEOS. For example, two body potentials developed from *ab initio* values typically utilize five to ten fitting parameters because of the high information content in *ab initio* calculations at different intermolecular distances.

Specifically, we propose the development of extended Lennard-Jones (ex-LJ) force fields to improve the fundamental equations of state at high pressures.

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

$$u_{\text{nb,exLJ}} = \sum_m C_m r^{-m} \quad (3)$$

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.

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 primary question of this study is whether or not it is possible to develop a pairwise additive force field of “technical accuracy.” As the non-bonded potential plays a key role, we focus primarily on developing extremely flexible intermolecular potential functions. 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.

As the ex-LJ has received relatively little attention in the literature, some additional questions naturally arise. For example, which coefficients should be negative? Which terms of Equation BLANK provide the greatest improvement in the force field? How many terms should be non-zero such that the model is sufficiently flexible but not over-fit. What type of combining rules should be used for cross interactions?

For nearly half a century the Lennard-Jones 12-6 potential has inundated the molecular simulation literature. Only in the past decade has the Mie  $n$ -6 potential received considerable attention. The ex-LJ 12-10-8-6 potential is more flexible than the two-parameter ( $\epsilon$  and  $\sigma$ ) 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. By demonstrating significant improvement at high pressures, the aim of this research is to initiate a paradigm shift in non-bonded potentials, namely, to commence the dawn of the extended Lennard-Jones era.

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 ( $\theta$ ) 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 ( $\Phi$ ). Previous publications demonstrate that MBAR- $\Phi$  reduces the computational cost by several orders of magnitude compared to direct molecular simulation. The combined MBAR- $\Phi$  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  $N_{\text{sites}}$  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  $\theta$  ( $C_{m,ij}$ ) 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  $m$ .

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.

Despite this advantage for fitting  $C_m$  to  $A_{xy}^r$ , it remains to be seen how many non-zero  $C_m$  terms can be utilized.

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

where  $q_{\text{ref}}$  are configurations sampled using the reference force field(s),  $\langle \rangle_{\text{MBAR}}$  are ensemble

averages estimated using MBAR, and  $F_i$  and  $F_j$  are functionals that depend on different powers of the internal energy ( $U^{i,j}$ ).

The methodology proposed in this study is to:

1. Develop a FEOS over the temperatures and pressures where reliable experimental data exist
2. Fit the ex-LJ parameters to FEOS derivative Helmholtz free energy properties in range where FEOS is reliable
3. Perform molecular simulations at state points where experimental data are not available
4. Iterate:
  - (a) Refit the FEOS to the hybrid data set
  - (b) Re-optimize the force field parameters to match FEOS (using MBAR so that no additional molecular simulations are required)

With this iterative approach, it is possible to ensure that the Helmholtz free energy derivative properties are internally consistent between the FEOS and the ex-LJ force field. The purpose for this iterative approach is to improve the extrapolation of FEOS and the reliability of the force field. Also, as the FEOS and force field become more internally consistent with each successive iteration, it is possible to increase the number of fitting parameters in both the FEOS and in the ex-LJ.

We identify four deliverables from this project. First, an improved FEOS for the molecules studies. Second, a force field of "technical accuracy." Third, an increased theoretical understanding of non-bonded potentials. Fourth, an infrastructure for rapid force field parameterization using Helmholtz free energy derivatives.