**Introduction**

In molecular mechanics (MM) simulation, the physical interactions are described using an empirical potential energy function (force field). Simple yet accurate force fields are highly useful but challenging to develop. Force field development procedures usually involve fitting the empirical parameters to reference data from high-level theoretical calculations or experimental measurements. Generally speaking, a rich and diverse data set can lead to a highly accurate model but also a difficult optimization problem.

We sought to create an inexpensive water model using polarizable atomic multipole expansions (Figure 3). The AMOEBA model contains self-consistent polarizable dipoles that incur a large computational cost (up to 80%). Using direct polarization greatly speeds up the calculation (up to 5x) by eliminating the self-consistent component, but the parameters need to be redetermined.

We used the novel approach of optimizing force fields using combined theoretical and experimental data. We hope that the water model presented here will be easy-to-use, inexpensive, and accurate for broad applications in theoretical chemistry.

**Method**

The force field is parameterized by minimizing an objective function in the parameter space. The objective function contains the force field errors with respect to the theoretical and/or experimental reference data.

ForceBalance can compute semi-analytic first and second derivatives of the objective function for optimization. When fitting experimental data, the analytic first derivative of a simulated observable is a two-point correlation function:

\[ V(x) \approx \frac{1}{2} \left[ \langle \psi | \nabla_{\beta} \psi \rangle - \langle \psi | \nabla_{\alpha} \psi \rangle \right] \cdot \nabla_{\alpha} \cdot \nabla_{\beta} = \sum_{A} \langle \psi | \mathbf{r} \cdot \mathbf{E} \rangle_{A} \cdot \langle \psi | \mathbf{r} \cdot \mathbf{E} \rangle_{A} \]

where \(x\) is the Boltzmann average of the observable \(A\), \(x\) represents the force field parameters, and \(Z\) is the canonical partition function. The energy derivative \(\nabla_{\alpha} E\) is evaluated using finite difference.

We apply ideas from Bayesian probability theory to prevent overfitting of parameters. A prior distribution centered at the initial parameter values is imposed on the parameter space. ForceBalance supports Laplacian priors (L1-norm penalty function) and Gaussian priors (L2-norm penalty function).

**Results**

The parameterization was performed using ForceBalance and the OpenMM simulation software which provides an accelerated GPU implementation of AMOEBA. ForceBalance uses OpenMM to perform the simulations, builds the objective function and derivatives from the simulation data, then applies the Levenberg-Marquardt optimization algorithm with a Gaussian prior.

The quality of fit is shown in Figures 4 and 5. The density is accurate to within 0.5% over the temperature range, and the temperature of maximum density is accurate to within 5 °C. The MP2 cluster energies are reproduced to within 6 kJ/mol (20%) but a small systematic error remains.

**Conclusion**

This project demonstrates ForceBalance as a highly general method for optimizing force fields using combined theoretical and experimental data. We hope that the water model presented here will be easy-to-use, inexpensive, and accurate for broad applications in theoretical chemistry.