



# Simple and systematic optimization of a polarizable water model

Lee-Ping Wang, Teresa Head-Gordon<sup>†</sup>, Jay Ponder<sup>°</sup>, Todd J. Martinez and Vijay S. Pande

Department of Chemistry, Stanford University, Stanford CA 94305

<sup>†</sup>Department of Chemistry, U.C. Berkeley, Berkeley CA 94720

<sup>°</sup>Department of Chemistry, Washington University, St. Louis, MO 63130



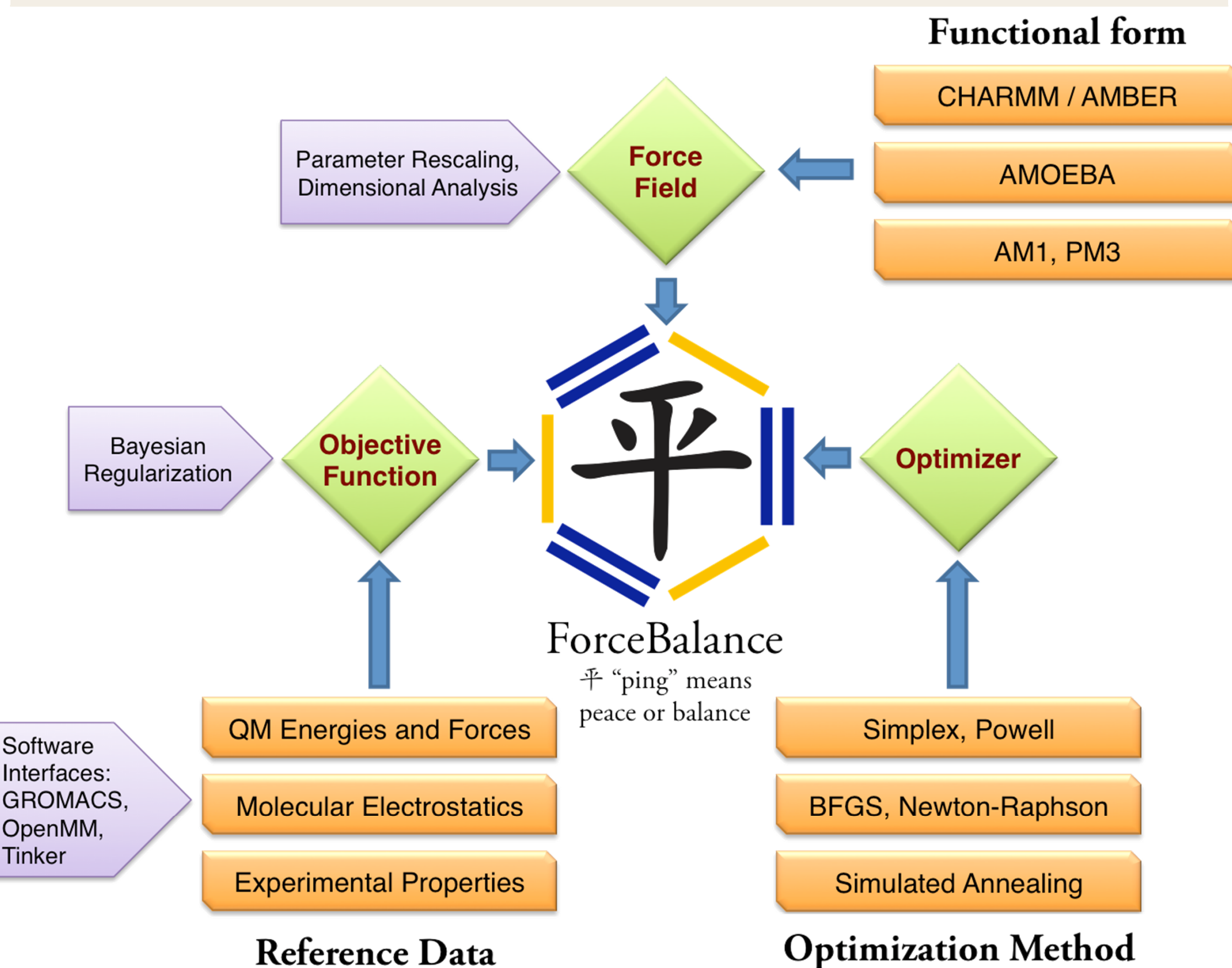
## Introduction

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 developed an open-source optimization program called **ForceBalance** to meet challenges in force field development. ForceBalance constructs force fields from accurate *theoretical data combined with experimental data* using systematic optimization methods and strict regularization schemes.

Here we apply ForceBalance to optimize an inexpensive polarizable water model largely based on the AMOEBA model but using *direct polarization* (AMOEBA uses self-consistent or mutual polarization). Our new model surpasses AMOEBA in accuracy for several properties of water.



**Figure 1:** The class structure of ForceBalance (above) affords a high degree of flexibility with respect to the force field's functional form, the experimental and theoretical reference data, and the optimization method.

## 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.

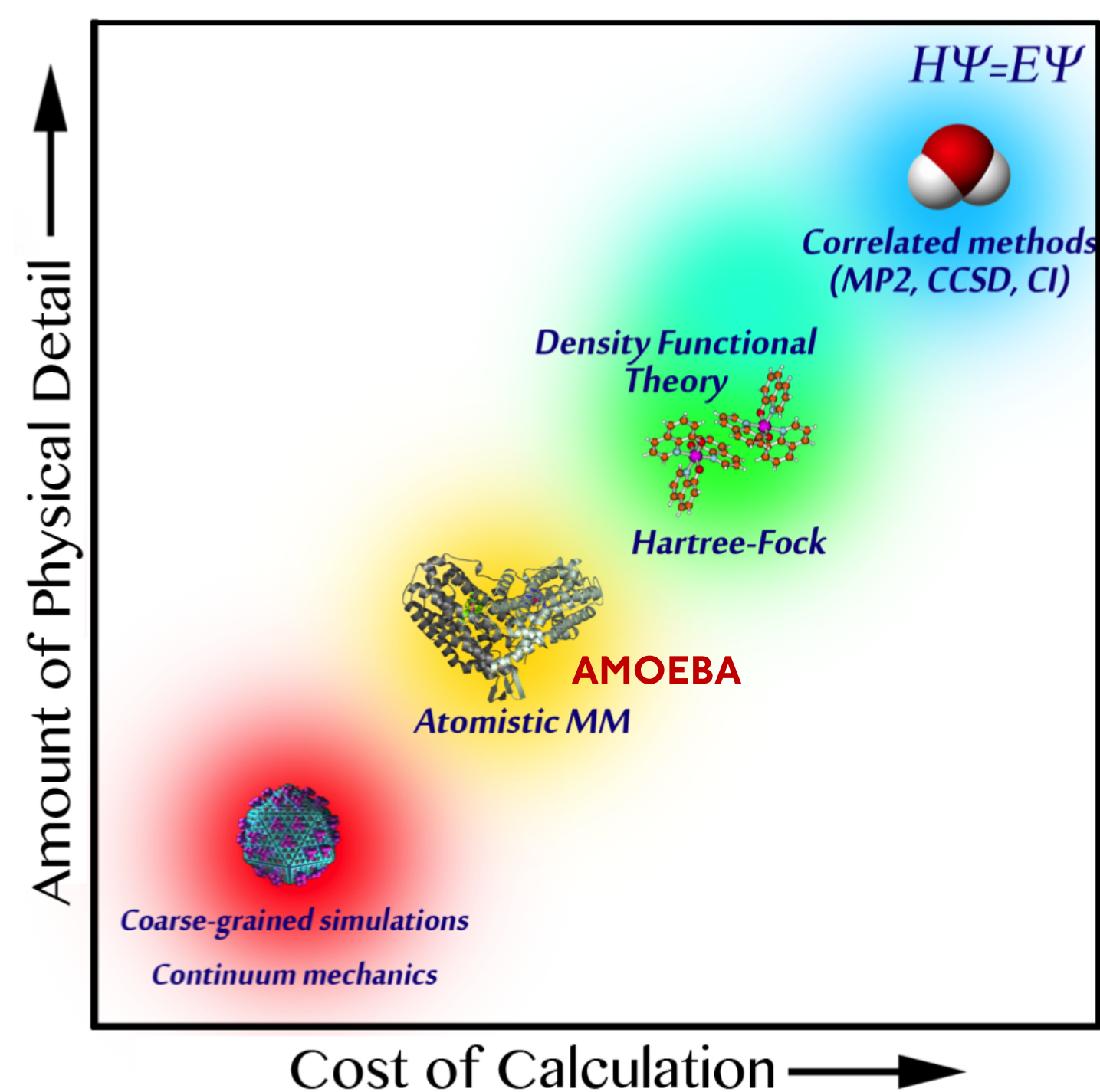
ForceBalance can compute semi-analytic first and second derivatives of the objective function for optimization. When fitting experimental data, the analytic first derivative of a simulated observable is a two-point correlation function:

$$\nabla_{\kappa} \langle A \rangle \equiv \nabla_{\kappa} \left( \frac{\int d\mathbf{r} e^{-\beta E(\mathbf{r}, \kappa)} A(\mathbf{r})}{Z} \right) = -\beta \left( \langle \nabla_{\kappa} E \cdot A \rangle - \langle \nabla_{\kappa} E \rangle \langle A \rangle \right),$$

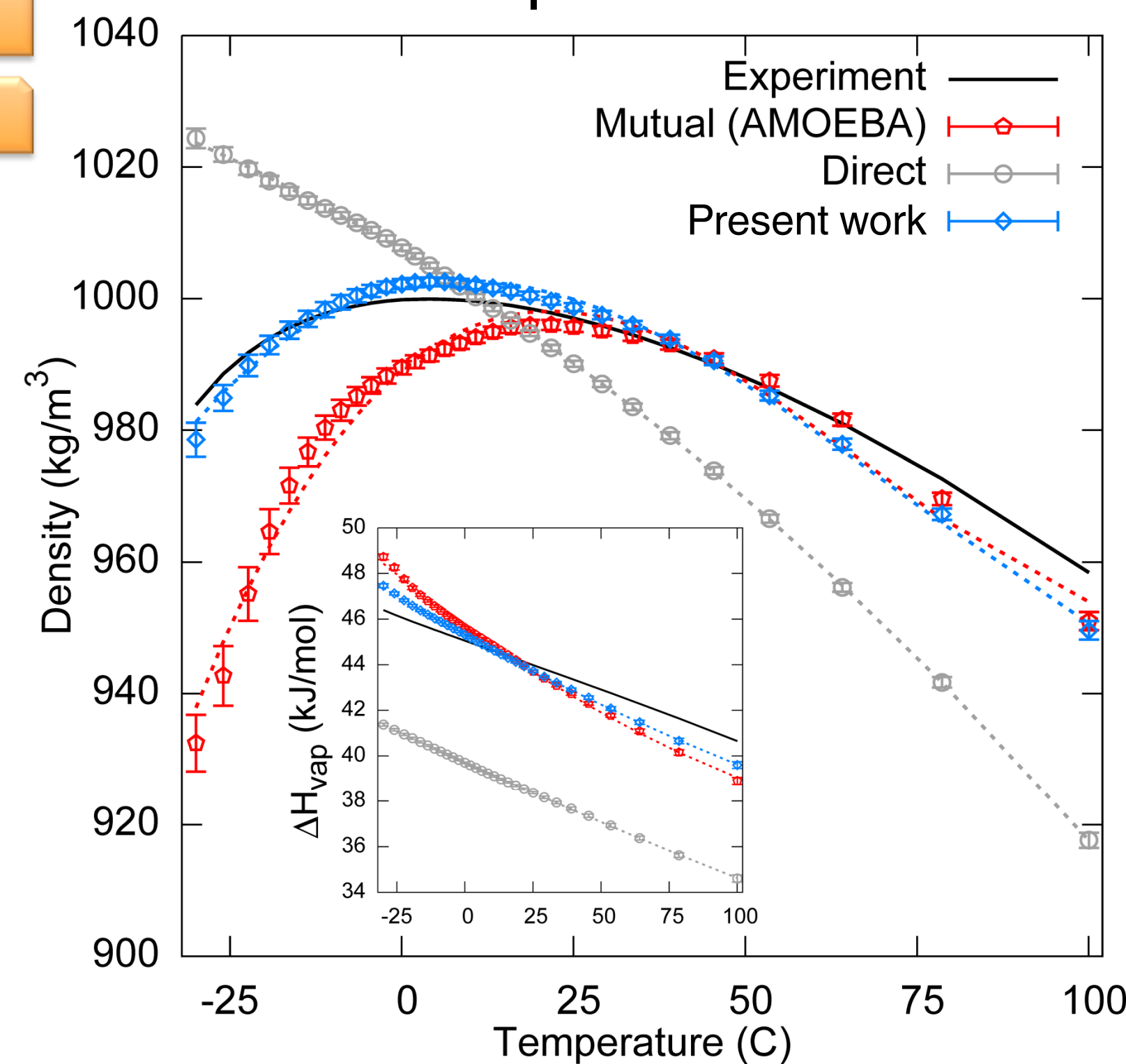
where  $\langle A \rangle$  is the Boltzmann average of the observable  $A$ ,  $\kappa$  represents the force field parameters, and  $Z$  is the canonical partition function. The energy derivative  $\nabla_{\kappa} E$  is evaluated using finite difference.

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).

## Overview of Theoretical Methods



## Fit to Experimental Data



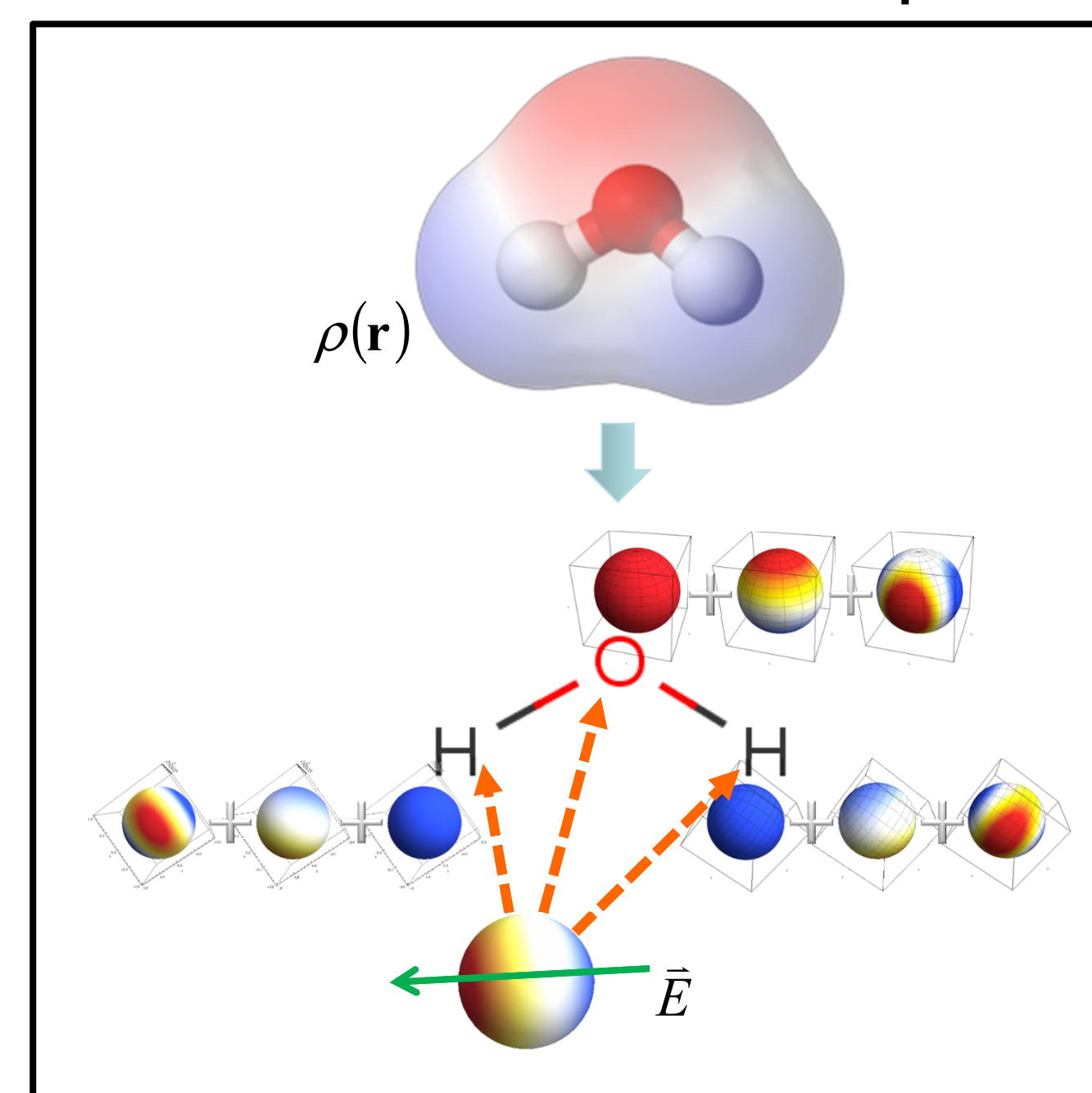
**Figure 2** (Upper left): In the big picture, this work belongs in the category of highly detailed atomistic MM simulation.

**Figure 3** (Upper right): The water model in this work uses fixed multipoles through second order and direct polarization.

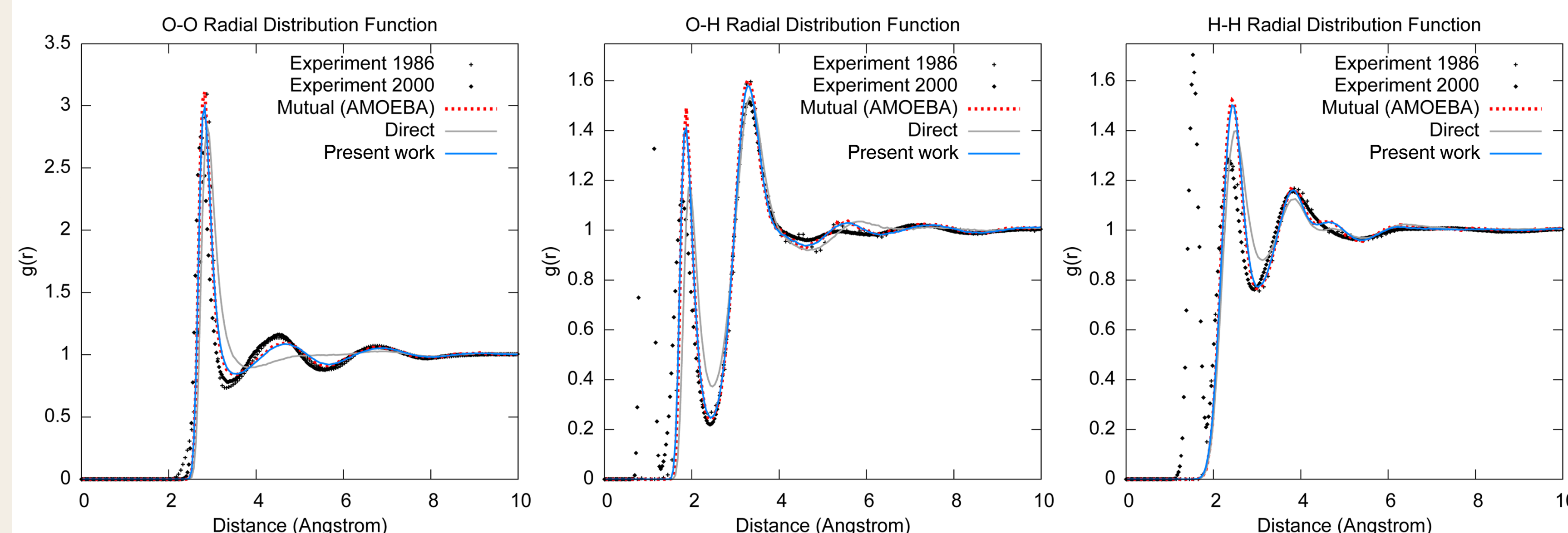
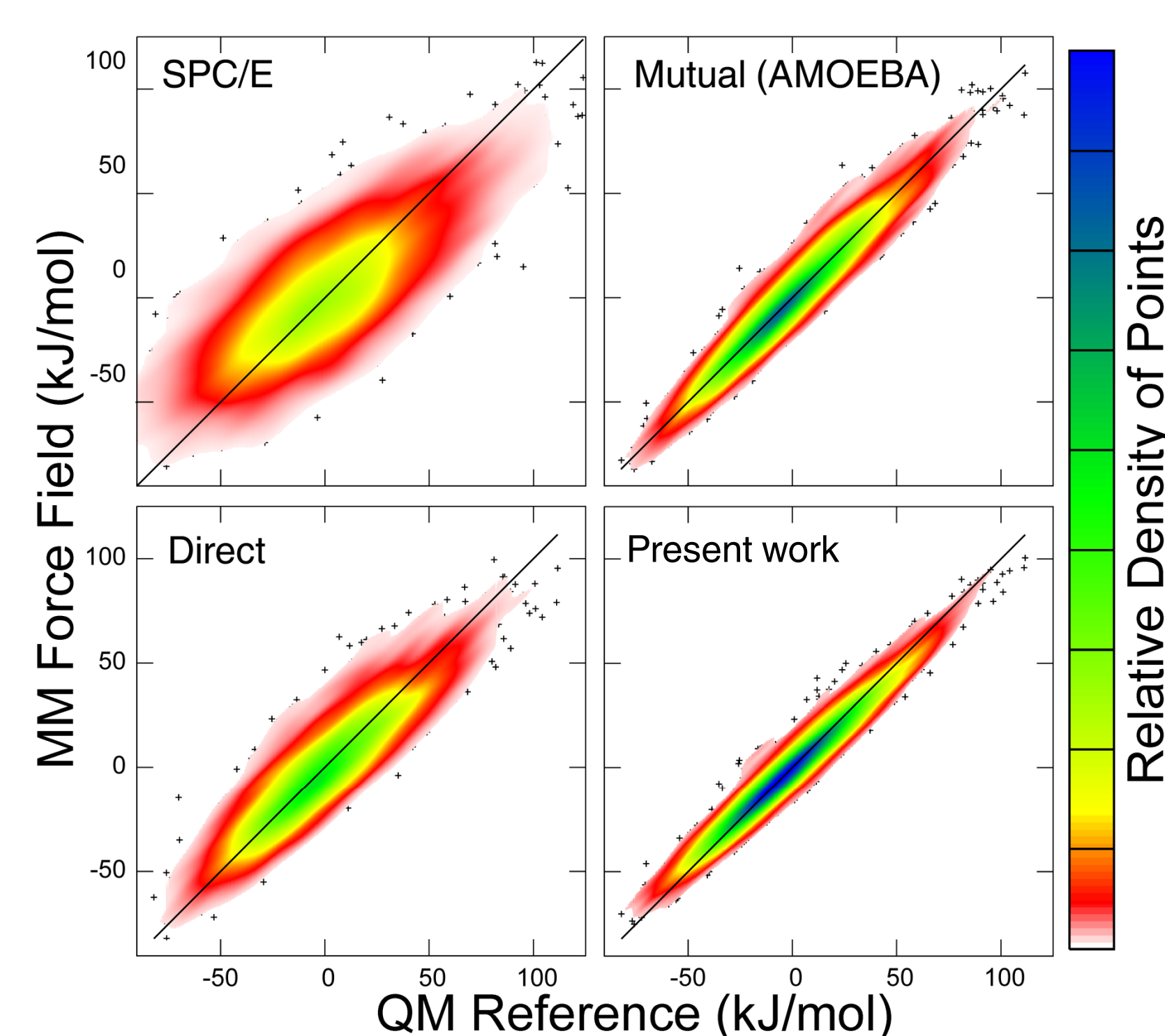
**Figure 4** (Lower left): Temperature profile of water density and  $\Delta H_{\text{vap}}$  (inset). The blue curve is the optimized direct model.

**Figure 5** (Lower right): MP2 vs. MM water cluster energies for 4 models. The optimized direct model is on the bottom right.

## Polarizable Atomic Multipoles



## Fit to Theoretical Calculations



**Figure 6:** Radial distribution functions of water at 298.15 K, 1 atm computed using the optimized direct model (blue) compared to AMOEBA (red), before optimization (gray), and two experimental measurements (black). The optimized direct model matches the AMOEBA results nearly exactly, and agreement with experiment is generally good.

## Application: Direct polarization water model

We sought to create an inexpensive water model using polarizable atomic multipoles (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.

Our reference data set includes elements from theory and experiment. We calculated single-point energies and forces using MP2/aug-cc-pVTZ for 5,100 water cluster geometries. We also considered the density and enthalpy of vaporization ( $\Delta H_{\text{vap}}$ ) ranging from -30 °C to 100 °C at atmospheric pressure.

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.  $\Delta H_{\text{vap}}$  is reproduced to within 1 kJ/mol, but the derivative (heat capacity) is too large. The MP2 cluster energies are reproduced to within 6 kJ/mol (20%) but a small systematic error remains.

Radial distribution functions and other experimental validation results are given in Figure 6 and Table 1. The radial distribution functions were not fitted, but the force field prediction is highly accurate. Kinetic properties such as the diffusion constant and dielectric constant are also accurately reproduced.

Property	AMOEBA	This work	Experiment
Density (kg m <sup>-3</sup> )	1000 ± 1	999 ± 1	997
Enthalpy of vaporization (kJ mol <sup>-1</sup> )	43.8 ± 0.1	43.8 ± 0.1	44.0
Dielectric constant	81 ± 10	81 ± 5	78.4
Diffusion constant (10 <sup>-5</sup> cm <sup>2</sup> s <sup>-1</sup> )	2.0 ± 0.1	2.3 ± 0.1	2.3
Density maximum (°C)	15 - 25	0 - 10	4

**Table 1:** Bulk properties of water at 298.15 K, 1 atm.

Parameter Name	Units	AMOEBA	This work	Δ	Δ (%)
O-H Equilibrium Bond Length	Å	0.9572	0.9641	0.0069	0.7
O-H Bond Force Constant	kcal/mol/Å <sup>2</sup>	529.60	547.64	18.04	3.4
H-O-H Equilibrium Angle	Degree	108.50	107.08	-1.42	-1.3
H-O-H Angle Force Constant	kcal/mol/Å <sup>2</sup>	34.05	29.48	-4.57	-13.4
H-H Urey-Bradley Equilibrium Length	Å	1.5537	1.5382	-0.0155	-1.0
H-H Urey-Bradley Force Constant	kcal/mol/Å <sup>2</sup>	38.25	24.60	-13.65	-35.7
Oxygen Vdw Sigma	Å	3.4050	3.4275	0.0225	0.7
Oxygen Vdw Epsilon	kcal/mol	0.1100	0.1084	-0.0016	-1.5
Hydrogen Vdw Sigma	Å	2.6550	2.5868	-0.0682	-2.6
Hydrogen Vdw Epsilon	kcal/mol	0.0135	0.0123	-0.0012	-8.9
Hydrogen Vdw Reduction Factor	None	0.910	0.890	-0.020	-2.2
Oxygen Charge	e	-0.51966	-0.52418	-0.0045	0.9
Oxygen Dipole Z-component	e bohr	+0.14279	+0.13597	-0.00682	-4.8
Oxygen Quadrupole XX-component	e bohr <sup>2</sup>	+0.37928	+0.38508	0.0058	1.5
Oxygen Quadrupole YY-component	e bohr <sup>2</sup>	-0.41869	-0.33563	0.0825	-19.7
Oxygen Quadrupole ZZ-component	e bohr <sup>2</sup>	+0.03881	-0.04945	-0.0883	-227
Hydrogen Dipole X-component	e bohr	-0.03859	-0.02930	0.0093	-24.1
Hydrogen Dipole Z-component	e bohr	-0.05818	-0.06185	-0.0037	6.3
Hydrogen Quadrupole XX-component	e bohr <sup>2</sup>	-0.03673	-0.02949	0.0072	-19.7
Hydrogen Quadrupole YY-component	e bohr <sup>2</sup>	-0.10739	-0.12096	-0.0136	12.6
Hydrogen Quadrupole XZ-component	e bohr <sup>2</sup>	-0.00203	+0.02103	0.0231	-1136
Hydrogen Quadrupole ZZ-component	e bohr <sup>2</sup>	+0.14412	+0.15045	0.0063	4.4
Oxygen Polarizability	Å <sup>3</sup>	0.837	0.660	-0.178	-21.2
Hydrogen Polarizability	Å <sup>3</sup>	0.496	0.444	-0.052	-10.5

**Table 2:** Force field parameters of the optimized direct model compared to the original AMOEBA model. The parameter modifications are small. Note that the atomic polarizabilities are significantly reduced in the direct model.

## 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.

ForceBalance and OpenMM are freely available at <https://simtk.org/>.