Supplementary Information: Title

Joseph Heindel[1,2], Selim Sami[1], Teresa Head-Gordon[1,2,3]

[1]Kenneth S. Pitzer Theory Center and Department of Chemistry
[2]Chemical Sciences Division, Lawrence Berkeley National Laboratory
[3]Departments of Bioengineering and Chemical and Biomolecular Engineering
University of California, Berkeley, CA, USA
corresponding author: thg@berkeley.edu

Damping function for charge penetration The core-shell interactions are damped according to

$$\boldsymbol{T}_{ij}^{damp} = \begin{bmatrix} 1 & \nabla & \nabla^2 \end{bmatrix} \cdot \left(\frac{1}{r_{ij}} f_{ij}^{damp}(r_{ij}) \right) \tag{1}$$

while the corresponding interaction tensor for shell-shell damping is written as:

$$\boldsymbol{T}_{ij}^{overlap} = \begin{bmatrix} 1 & \nabla & \nabla^2 \\ \nabla & \nabla^2 & \nabla^3 \\ \nabla^2 & \nabla^3 & \nabla^4 \end{bmatrix} \cdot \left(\frac{1}{r_{ij}} f_{ij}^{overlap}(r_{ij}) \right) \tag{2}$$

The damping functions $f_{ij}^{damp}(r_{ij})$ and $f_{ij}^{overlap}(r_{ij})$ themselves take the following forms.

$$f_{ij}^{damp}(r_{ij}) = 1 - \left(1 + \frac{1}{2} b_j r_{ij} \right) e^{-b_j r_{ij}} \tag{3a}$$

$$f_{ij}^{overlap}(r_{ij}) = 1 - \left(1 + \frac{11}{16} b_{ij} r_{ij} + \frac{3}{16} (b_{ij} r_{ij})^2 + \frac{1}{48} (b_{ij} r_{ij})^3 \right) e^{-b_{ij} r_{ij}} \tag{3b}$$

The damping function in Eq. 3a can be derived directly from the form of the Slater density in Eq. **??** by computing its electrostatic potential. The damping function in Eq. 3b can be derived from a symmetrized coulomb integral where each density interacts with the damped potential generated by the other density.[1] Finally, it is important to note that these damping functions are the ones which apply to charge-charge interactions and that as higher-order multipoles are considered, new damping functions are generated alongside the gradients of $1/r_{ij}$. In other words, every interaction tensor in the multipole expansion will be damped differently. We use the same set of damping functions as derived by Rackers and Ponder which are appropriate for the Slater density in Eq. **??**.[1] Very similar models have been applied successfully in MB-UCB[2] based on functional forms proposed by Piquemal[3] and others[4,5].

Environment dependent bonding potential fot the water dimer To summarize the large effect this environment-dependent bonding potential has on the forces within FQCT, let us consider a water dimer where each monomer is fixed at the ωB97X-V/def2-QZVPPD geometry. This simply eliminates the geometric distortion forces which are not relevant here. We now compute the forces due to each term in EDA, a so-called force decomposition analysis[6], and compare them against the same forces predicted by FQCT with and without Eqs. **??** and **??**.

The net effect of the field-dependent Morse potential and charge transfer correction is to shift force from the oxygen atom to the hydrogen atom participating in a hydrogen bond. The improvement in accuracy in electrostatics and polarization forces are particularly notable since this correction involves no free parameters. The improvement in Pauli repulsion is also quite large correcting about half the force error at negligible computational cost. The charge transfer forces also remove about half the force error along the O−H bond. Note that the Pauli repulsion forces can be quite a bit more accurate with the present model, as shown in the parentheses in Table 1. Since we use the Pauli repulsion to correlate errors in the model, as shown in Figure **??**, the Pauli forces compensate for remaining errors in the dispersion and charge transfer forces.

In the course of developing this model, we found that we were unable to reproduce the expected correlation between change in bond length and change in harmonic frequency for O−H stretches.[7] We ultimately determined that the failure of our model to reproduce this correlation was due to a lack of coupling between our bonding potential and the environment. To that end, we extended

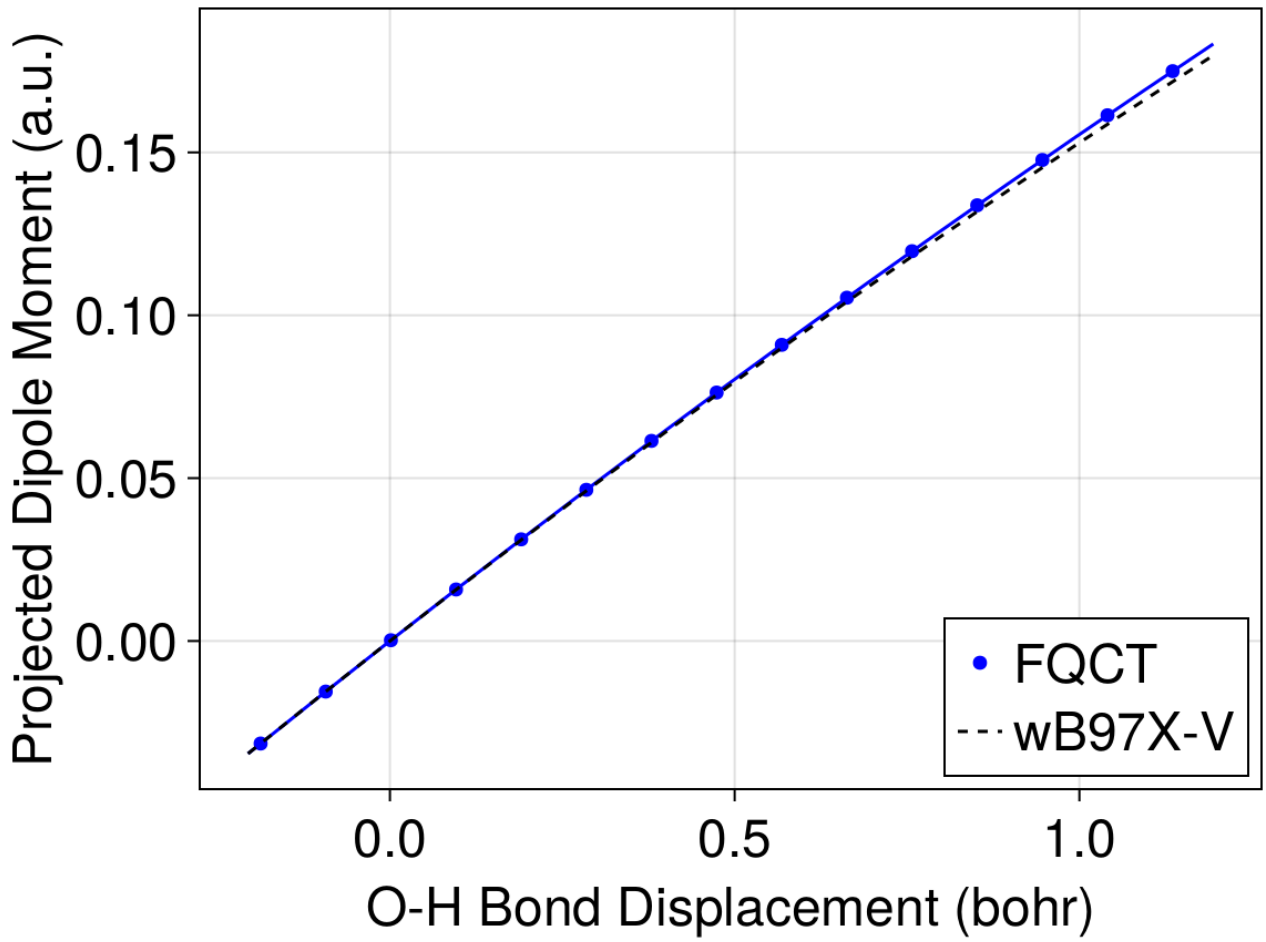
Water Dimer Force Decomposition Analysis Along O−H Bond

Method	Atom	Mod. Pauli	Cls. Elec.	Disp.	Pol.	CT	Total
FQCT	$O_{don.}$	-30.1	1.4	6.0	-3.4	-1.5	-27.5
No F.D. Morse	$H_{don.}$	-73.7	54.5	5.3	15.4	24.9	26.4
	$O_{acc.}$	96.4	-48.9	-10.1	-12.7	-22.3	2.4
FQCT	$O_{don.}$	-15.5 (7.50)	-13.7	6.0	-5.6	-8.7	-37.5
	$H_{don.}$	-88.5 (-111.9)	70.0	5.3	17.7	31.9	36.4
	$O_{acc.}$	96.4 (95.5)	-51.7	-10.1	-13.0	-22.0	-0.4
ωB97X-V	$O_{don.}$	3.92	-15.0	0.75	-8.4	-15.6	-34.3
	$H_{don.}$	-105.7	72.4	10.8	18.9	39.3	35.8
	$O_{acc.}$	92.3	-52.3	-10.3	-10.4	-19.6	-0.50

Table 1: Comparison of the forces projected along the O−H bond of a water dimer due to each component in the EDA. Forces predicted by FQCT are computed with and without the field-dependent morse potential described in Eqs. **??** and **??**. The first block of entries is FQCT with no field-dependent Morse potential, the second block is FQCT, and the third is ωB97X-V/def2-QZVPPD. All forces in KJ/mol/Å. $O_{don.}$ and $H_{don.}$ are the oxygen and hydrogen atoms donating the hydrogen bond and $O_{acc.}$ is the oxygen accepting the hydrogen bond. All forces are projected along the O−H bond vector. The numbers in parentheses are what we forces before error fitting with Pauli repulsion.

our Morse potential to be field-dependent in a manner first described elsewhere.[7] In short, the field-dependence of the potential requires specifying the first and second dipole derivatives along the O−H stretch.

Rather than treating these as fitting parameters, we computed the dipole derivatives from electronic structure as follows. First, we scan along an O−H stretch and compute the total dipole moment, μ, at each point along the scan. This dipole moment is then projected onto the O−H bond vector, $\mu_{\mathrm{OH}} = \frac{R_{\mathrm{OH}} \cdot \mu}{|R_{\mathrm{OH}}|}$. This allows us to isolate how the dipole changes as the O−H vibrates in a particular environment.



Supplementary Figure 1: *Projected dipole moments along various O-H stretches.* The dipole moment of (H_2O) is computed with ωB97X-V/def2-QZVPPD and FQCT as a function of the O−H stretch distance. All other degrees of freedom are fixed. The dipole moment is projected along the O−H stretch unit vector. By fitting a second-order polynomial, we find $\mu^{(1)}_{\mathrm{OH}} = 0.165$ $\mu^{(2)}_{\mathrm{OH}} = -0.012$ for ωB97X-V. The FQCT dipole derivatives are basically identical: $\mu^{(1)} = 0.1658$ and $\mu^{(2)} = -0.0104$.

In principle, the curves in Figure 1 should eventually be equal when sufficient explicit solvent is included in the calculation. Since the dipole derivatives computed from $(H_2O)_{2+6}$ are not sufficient to reproduce the Badger correlation, we increase the dipole derivatives until the appropriate correlation is reproduced. This results in values of $\mu^{(1)}_{\mathrm{OH}} = 1.19$ and $\mu^{(2)}_{\mathrm{OH}} = 3.69$.

We have already shown in Figure 5 that our model produces accurate three-body contributions to polarization and charge transfer for water trimers extracted from ion-water clusters. It is interesting to see, however, the 2-body and many-body contributions to each component of the energy for whole water clusters. To that end, we computed the 2-body and many-body contributions for a subset of the reference structures used in Tables 3 and 4. A comparison of these quantities as computed with ωB97X-V/def2-QZVPPD and FQCT is shown in Table 2.

Comparison of Many-Body Expansion for EDA Components (kcal/mol)

$(H_2O)_n$	Component	Cls. Elec.		Mod. Pauli		Disp.		Pol.		CT	
		FQCT	ωB97X-V	FQCT	ωB97X-V	FQCT	ωB97X-V	FQCT	ωB97X-V	FQCT	ωB97X-V
$(H_2O)_3$	2-Body	-27.02	-26.77	29.51	28.87	-6.31	-6.28	-3.41	-3.48	-6.29	-6.08
	$\geq$ 3-Body	-	-	0.0	-0.24	0.0	0.21	-1.42	-1.63	-0.77	-0.74
	Total	-27.02	-26.77	29.51	28.63	-6.31	-6.07	-4.84	-5.11	-7.07	-6.82
$(H_2O)_6$	2-Body	-79.01	-78.43	88.93	88.40	-19.65	-19.90	-10.67	-10.94	-18.32	-18.41
Prism	$\geq$ 3-Body	-	-	0.0	-0.62	0.0	0.83	-5.82	-6.40	-2.93	-2.80
	Total	-79.01	-78.43	88.93	87.78	-19.65	-19.07	-16.49	-17.34	-21.26	-21.21
$(H_2O)_6$	2-Body	-78.76	-78.67	90.15	90.10	-19.13	-19.36	-11.17	-11.25	-19.41	-19.62
Cage	$\geq$ 3-Body	-	-	0.0	-0.66	0.0	0.79	-5.76	-6.26	-3.15	-3.07
	Total	-78.76	-78.67	90.15	89.44	-19.13	-18.56	-16.94	-17.51	-22.57	-22.69
$(H_2O)_{10}$	2-Body	-161.5	-162.4	190.4	192.2	-37.89	-38.34	-24.69	-24.92	-42.62	-43.27
	$\geq$ 3-Body	-	-	0.0	-0.68	0.0	1.18	-13.50	-14.92	-7.67	-7.42
	Total	-161.5	-162.4	190.4	191.5	-37.89	-37.16	-38.19	-39.84	-50.29	-50.69
$(H_2O)_{16}$	2-Body	-276.8	-277.7	327.2	329.0	-67.23	-67.61	-42.21	-42.44	-72.41	-73.15
Antiboat	$\geq$ 3-Body	-	-	0.0	-1.06	0.0	2.17	-23.93	-27.07	-12.98	-12.37
	Total	-276.8	-277.7	327.2	327.9	-67.23	-65.44	-66.14	-69.51	-85.39	-85.52
$(H_2O)_{20}$	2-Body	-354.7	-355.4	419.4	420.9	-89.17	-89.20	-53.76	-54.26	-92.54	-92.80
ES Prism	$\geq$ 3-Body	-	-	0.0	-1.66	0.0	3.03	-30.53	-34.85	-16.28	-15.21
	Total	-354.7	-355.4	419.4	419.3	-89.17	-86.17	-84.29	-89.11	-108.8	-108.0

Table 2: Comparison of the 2-body, many-body (i.e. $\geq$ 3-Body), and total energies as predicted by FQCT and as computed with ωB97X-V/def2-QZVPPD. Both calculations are done at the *ab initio* optimized geometries. Names of isomers, when applicable, are written below the cluster size.

One of the significant advantages of using data from EDA to parameterize a force field is it allows for very fine-grained analysis of the accuracy of a model. We illustrate this in Table 2 by comparing the 2-body, many-body, and total energies as computed by our force field and ωB97X-V/def2-QZVPPD. First, we can see that the model is generally very accurate at reproducing each term, as should be expected given the excellent performance of the model in reproducing total energies and optimized strcutures. With that being said, there are a couple of apparent shortcomings. First, we currently do not include many-body dispersion which is very small but not entirely negligible, especially considering it is the only term besides Pauli repulsion which is repulsive. It is sometimes argued that many-body dispersion can be ignored since there is also a small many-body exchange effect which is attractive and hence offsets many-body dispersion. It is clear from Table 2 that this is at least approximately true in EDA. Even if the cancellation were perfect, however, that does not guarantee the forces will cancel. For that reason, we may explore adding many-body dispersion in the future, but neglect it for now since many-body charge transfer is generally much more important.

Additionally, it is apparent that our model slightly understimates two-body contributions to both polarization and charge transfer. This likely arises from a limitation of the functional forms and hence suggests improvements that can be explored in the future. Our description of many-body contributions to charge transfer, however, are excellent. Table 2 also provides information about ways in which we have gotten lucky in the parameterization of this model. For instance, in larger clusters the model's dispersion energies are slightly too attractive while our polarization model is not quite attractive enough. This cancellation of errors was not intended but likely contributes, in some small way, to the successes we have seen in Tables 4 and 5.

We are digging into such detail with the hopes that future force fields will do the same. Electronic structure only used to be able to provide coarse-grained information in the form of total energies and forces. This hampered the ability to pinpoint limitations of a force field. Now, however, we can easily identify systematic errors in a force field using the powerful combination of the many-body expansion and energy decomposition analysis.

Supplementary Table S3 presents the mean absolute errors (MAEs) and skewness over each cluster size for each energy category in the EDA. In general, the MAEs obtained by this model are excellent. In addition to the MAEs for each category of the EDA, we also report the skewness of the error distribution.

MAE of Force Field EDA Terms (kcal/mol)

	$(H_2O)_2$	$(H_2O)_3$	$(H_2O)_4$	$(H_2O)_5$
Pauli (no error fit)	0.134 (1.513)	0.209 (0.538)	0.277 (0.491)	0.329 (0.400)
Pauli	0.195 (-0.064)	0.297 (-0.007)	0.387 (0.117)	0.506 (-0.098)
Electrostatics	0.123 (-0.268)	0.206 (0.189)	0.283 (-0.092)	0.348 (-0.061)
Dispersion	0.069 (0.090)	0.092 (-0.124)	0.109 (-0.407)	0.149 (-0.420)
Polarization	0.047 (-0.243)	0.088 (-0.041)	0.122 (0.437)	0.155 (0.506)
Charge Transfer	0.102 (-0.610)	0.159 (-0.207)	0.218 (-0.278)	0.264 (-0.197)
Interaction	0.089 (1.158)	0.166 (0.875)	0.226 (0.399)	0.290 (0.428)

Table 3: Comparison of the mean absolute error (MAE) in kcal/mol of all terms in the EDA against predictions of our water model for hydrogen-bonded water dimers, trimers, tetramers, and pentamers. The numbers in parentheses are the skewness of the error distribution. In total, there are 2400 each of dimers, trimers, tetramers, and pentamers. The first row shows the Pauli repulsion energy without inclusion of error fitting while the second row is the Pauli repulsion used in the final model which is calibrated to maximize error correlation.

Supplementary References

[1] Joshua A Rackers, Roseane R Silva, Zhi Wang, and Jay W Ponder. Polarizable water potential derived from a model electron density. *Journal of chemical theory and computation*, 17(11): 7056–7084, 2021.

[2] Akshaya K Das, Lars Urban, Itai Leven, Matthias Loipersberger, Abdulrahman Aldossary, Martin Head-Gordon, and Teresa Head-Gordon. Development of an advanced force field for water using variational energy decomposition analysis. *Journal of chemical theory and computation*, 15(9):5001–5013, 2019.

[3] Jean-Philip Piquemal, Nohad Gresh, and Claude Giessner-Prettre. Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *The Journal of Physical Chemistry A*, 107 (48):10353–10359, 2003.

[4] Qiantao Wang, Joshua A Rackers, Chenfeng He, Rui Qi, Christophe Narth, Louis Lagardere, Nohad Gresh, Jay W Ponder, Jean-Philip Piquemal, and Pengyu Ren. General model for treating short-range electrostatic penetration in a molecular mechanics force field. *Journal of chemical theory and computation*, 11(6):2609–2618, 2015.

[5] Joshua A Rackers, Qiantao Wang, Chengwen Liu, Jean-Philip Piquemal, Pengyu Ren, and Jay W Ponder. An optimized charge penetration model for use with the amoeba force field. *Physical Chemistry Chemical Physics*, 19(1):276–291, 2017.

[6] Abdulrahman Aldossary, Martí Gimferrer, Yuezhi Mao, Hongxia Hao, Akshaya K Das, Pedro Salvador, Teresa Head-Gordon, and Martin Head-Gordon. Force decomposition analysis: A method to decompose intermolecular forces into physically relevant component contributions. *The Journal of Physical Chemistry A*, 127(7):1760–1774, 2023.

[7] Mark A Boyer, Ondrej Marsalek, Joseph P Heindel, Thomas E Markland, Anne B McCoy, and Sotiris S Xantheas. Beyond badger's rule: The origins and generality of the structure–spectra relationship of aqueous hydrogen bonds. *The Journal of Physical Chemistry Letters*, 10(5): 918–924, 2019.