

Supplementary Information: Title

Joseph Heindel^{1,2}, Selim Sami¹, Teresa Head-Gordon^{1,2,3}

¹Kenneth S. Pitzer Theory Center and Department of Chemistry

²Chemical Sciences Division, Lawrence Berkeley National Laboratory

³Departments of Bioengineering and Chemical and Biomolecular Engineering

University of California, Berkeley, CA, USA

corresponding author: thg@berkeley.edu

9 **Damping function for charge penetration** The core-shell interactions are damped accord-
 10 ing to

$$\mathbf{T}_{ij}^{damp} = [1 \quad \nabla \quad \nabla^2] \cdot \left(\frac{1}{r_{ij}} f_{ij}^{damp}(r_{ij}) \right) \quad (1)$$

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

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

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

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

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

14 The damping function in Eq. 3a can be derived directly from the form of the Slater density in
 15 Eq. ?? by computing its electrostatic potential. The damping function in Eq. 3b can be derived
 16 from a symmetrized coulomb integral where each density interacts with the damped potential
 17 generated by the other density.¹ Finally, it is important to note that these damping functions
 18 are the ones which apply to charge-charge interactions and that as higher-order multipoles are
 19 considered, new damping functions are generated alongside the gradients of $1/r_{ij}$. In other words,
 20 every interaction tensor in the multipole expansion will be damped differently. We use the same
 21 set of damping functions as derived by Rackers and Ponder which are appropriate for the Slater
 22 density in Eq. ??.¹ Very similar models have been applied successfully in MB-UCB² based on
 23 functional forms proposed by Piquemal³ and others^{4,5}.

24 **Environment dependent bonding potential for the water dimer** To summarize the
 25 large effect this environment-dependent bonding potential has on the forces within FQCT, let us
 26 consider a water dimer where each monomer is fixed at the ω B97X-V/def2-QZVPPD geometry.
 27 This simply eliminates the geometric distortion forces which are not relevant here. We now compute
 28 the forces due to each term in EDA, a so-called force decomposition analysis⁶, and compare them
 29 against the same forces predicted by FQCT with and without Eqs. ?? and ??.

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

39 In the course of developing this model, we found that we were unable to reproduce the expected
 40 correlation between change in bond length and change in harmonic frequency for O-H stretches.⁷
 41 We ultimately determined that the failure of our model to reproduce this correlation was due to
 42 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
ω B97X-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 ω B97X-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.

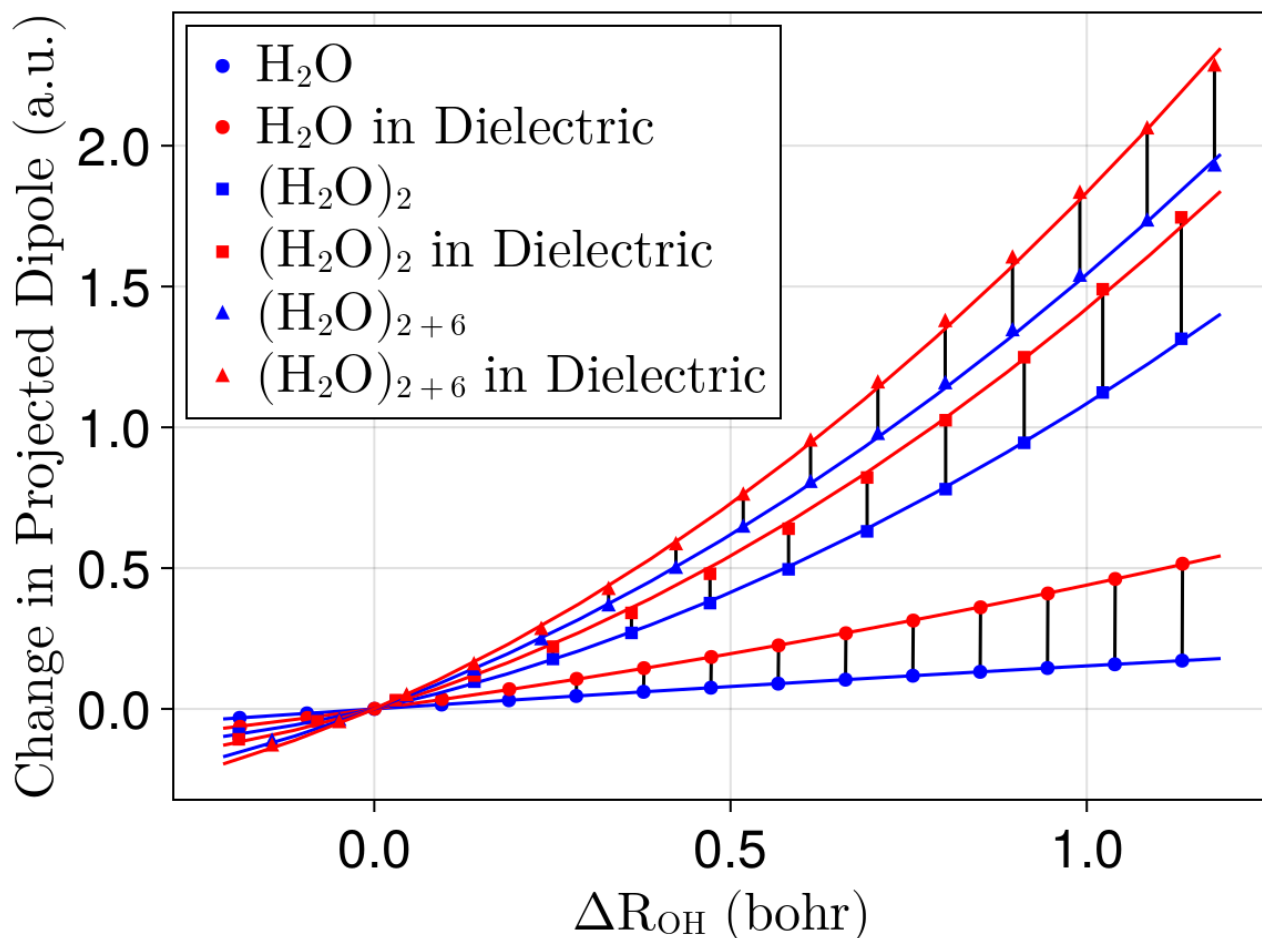
Rather than treating these as fitting parameters, we computed the dipole derivatives from electronic structure as follows. First, we scan along an O–H stretch and compute the total dipole moment, μ , at each point along the scan. This dipole moment is then projected onto the O–H bond vector, $\mu_{\text{OH}}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

We considered three different environments to see how the dipole derivatives depend on the local environment and ultimately chose a set of parameters we deemed appropriate for polarization and one to represent the full system. Remember that in EDA, the polarization energy represents the response of a molecule to the rest of the environment plus mutual polarization. To that end, we computed an O–H stretch of an isolated H₂O in a polarizable continuum described by the COSMO model.⁸ This resulted in the dipole surface shown with circles in Fig. 1. The lines are fits with a second order polynomial, $\Delta\mu_{\text{OH}} = \mu_{\text{OH}}^{(1)}\Delta R_{\text{OH}} + \mu_{\text{OH}}^{(2)}\Delta R_{\text{OH}}^2$. For H₂O in a dielectric, we find $\mu_{\text{OH}}^{(1)} = 0.345$ and $\mu_{\text{OH}}^{(2)} = 0.095$. This is to be compared to the same values for an isolated H₂O of $\mu_{\text{OH}}^{(1)} = 0.165$ $\mu_{\text{OH}}^{(2)} = -0.012$. The effect of an environment is, therefore, to promote dissociation to ionic products rather than radicals which is what the change in sign of $\mu_{\text{OH}}^{(2)}$ implies and to make it easier to elongate the O–H bond which is what the magnitude of $\mu_{\text{OH}}^{(1)}$ means.

Interestingly, when the environment is represented by an actual molecule, as with (H₂O)₂, the second dipole derivative increases by a lot and increases even further when a dielectric continuum is used. This is seen in the points marked with squares in Fig. 1. We attribute this large non-linearity in the dipole surface to the role of charge delocalization, which makes these

parameters suitable for the full potential which includes charge transfer.

A water dimer, however, is not a representative environment for liquid water, or even large water clusters. To remedy this, we constructed a tetrahedrally coordinated water dimer, $(\text{H}_2\text{O})_{2+6}$. The notation $(\text{H}_2\text{O})_{2+6}$ indicates that this is a water dimer solvated by six other water molecules. The cartesian coordinates are available with the manuscript. When this larger solvation environment is used, we see a further increase in the dipole derivatives beyond just a water dimer. The dipole derivatives for the central O–H stretch in $(\text{H}_2\text{O})_{2+6}$ are $\mu_{\text{OH}}^{(1)} = 1.082$ and $\mu_{\text{OH}}^{(2)} = 1.54$. When these values are used in the model, the $\Delta\omega$ vs ΔR_e slope is $\approx -17.1 \text{ cm}^{-1}/.001$ over the same clusters.



Supplementary Figure 1: *Projected dipole moments along various O–H stretches.* The dipole moment of (H_2O) , $(\text{H}_2\text{O})_2$, and $(\text{H}_2\text{O})_{2+6}$ are computed with $\omega\text{B97X-V/def2-QZVPPD}$ 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. $(\text{H}_2\text{O})_{2+6}$ is a water dimer with six additional water molecules placed so that the water dimer is tetrahedrally coordinated. The blue points and lines are gas-phase systems while the red points and lines include a polarizable continuum. Each point is connected vertically to make it clear they are the same structure but in the presence of a polarizable continuum. Circles are used for a single water molecule, squares for a water dimer, and triangles for the tetrahedrally solvated water dimer. See text for more details and values of computed dipole derivatives.

74 In principle, the curves in Figure 1 should eventually be equal when sufficient explicit solvent
 75 is included in the calculation. Since the dipole derivatives computed from $(\text{H}_2\text{O})_{2+6}$ are not suffi-
 76 cient to reproduce the Badger correlation, we increase the dipole derivatives until the appropriate
 77 correlation is reproduced. This results in values of $\mu_{\text{OH}}^{(1)} = 1.19$ and $\mu_{\text{OH}}^{(2)} = 3.69$.

78 We have already shown in Figure 5 that our model produces accurate three-body contributions
 79 to polarization and charge transfer for water trimers extracted from ion-water clusters. It is inter-
 80 esting to see, however, the 2-body and many-body contributions to each component of the energy
 81 for whole water clusters. To that end, we computed the 2-body and many-body contributions for
 82 a subset of the reference structures used in Tables 3 and 4. A comparison of these quantities as
 83 computed with $\omega\text{B97X-V/def2-QZVPPD}$ and FQCT is shown in Table 2.

Comparison of Many-Body Expansion for EDA Components (kcal/mol)											
$(\text{H}_2\text{O})_n$	Component	Cls. Elec.		Mod. Pauli		Disp.		Pol.		CT	
		FQCT	$\omega\text{B97X-V}$	FQCT	$\omega\text{B97X-V}$	FQCT	$\omega\text{B97X-V}$	FQCT	$\omega\text{B97X-V}$	FQCT	$\omega\text{B97X-V}$
$(\text{H}_2\text{O})_3$	2-Body	-27.02	-26.77	29.51	28.87	-6.31	-6.28	-3.41	-3.48	-6.29	-6.08
	≥ 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
$(\text{H}_2\text{O})_6$ Prism	2-Body	-79.01	-78.43	88.93	88.40	-19.65	-19.90	-10.67	-10.94	-18.32	-18.41
	≥ 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
$(\text{H}_2\text{O})_6$ Cage	2-Body	-78.76	-78.67	90.15	90.10	-19.13	-19.36	-11.17	-11.25	-19.41	-19.62
	≥ 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
$(\text{H}_2\text{O})_{10}$	2-Body	-161.5	-162.4	190.4	192.2	-37.89	-38.34	-24.69	-24.92	-42.62	-43.27
	≥ 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
$(\text{H}_2\text{O})_{16}$ Antiboat	2-Body	-276.8	-277.7	327.2	329.0	-67.23	-67.61	-42.21	-42.44	-72.41	-73.15
	≥ 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
$(\text{H}_2\text{O})_{20}$ ES Prism	2-Body	-354.7	-355.4	419.4	420.9	-89.17	-89.20	-53.76	-54.26	-92.54	-92.80
	≥ 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. ≥ 3 -Body), and total energies as predicted by FQCT and as computed with $\omega\text{B97X-V}/\text{def2-QZVPPD}$. 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 ω B97X-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 structures. 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 underestimates 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 ₂ O) ₂	(H ₂ O) ₃	(H ₂ O) ₄	(H ₂ O) 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.

- 138 [7] Mark A Boyer, Ondrej Marsalek, Joseph P Heindel, Thomas E Markland, Anne B McCoy, and
139 Sotiris S Xantheas. Beyond badger’s rule: The origins and generality of the structure–spectra
140 relationship of aqueous hydrogen bonds. *The Journal of Physical Chemistry Letters*, 10(5):
141 918–924, 2019.
- 142 [8] Andreas Klamt. The cosmo and cosmo-rs solvation models. *Wiley Interdisciplinary Reviews:*
143 *Computational Molecular Science*, 1(5):699–709, 2011.