



Decomposed Mean-Field Simulations of Local Properties in Condensed Phases

Eriksen, Janus Juul

Published in:
Journal of Physical Chemistry Letters

Link to article, DOI:
[10.1021/acs.jpcllett.1c01375](https://doi.org/10.1021/acs.jpcllett.1c01375)

Publication date:
2021

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Eriksen, J. J. (2021). Decomposed Mean-Field Simulations of Local Properties in Condensed Phases. *Journal of Physical Chemistry Letters*, 12, 6048-6055. <https://doi.org/10.1021/acs.jpcllett.1c01375>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Decomposed Mean-Field Simulations of Local Properties in Condensed Phases

Janus J. Eriksen*

DTU Chemistry, Technical University of Denmark

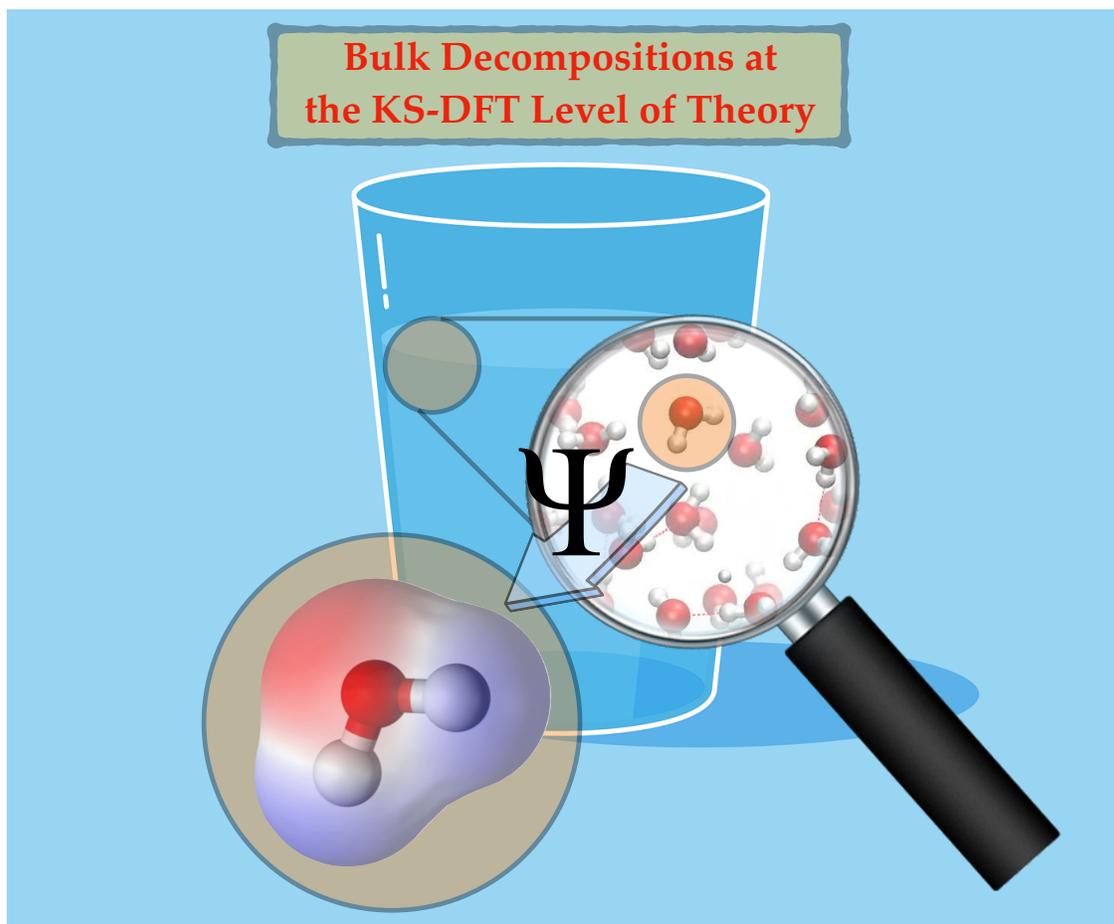
Kemitorvet Bldg. 206, DK-2800 Kgs. Lyngby, Denmark

E-mail: janus@kemi.dtu.dk

Abstract

The present work demonstrates a robust protocol for probing localized electronic structure in condensed-phase systems, operating in terms of a recently proposed theory for decomposing the results of Kohn-Sham density functional theory in a basis of spatially localized molecular orbitals [Eriksen, *J. Chem. Phys.* **153**, 214109 (2020)]. In an initial application to liquid, ambient water and the assessment of the solvation energy and the embedded dipole moment of H₂O in solution, we find that both properties are amplified on average—in accordance with expectation—and that correlations are indeed observed to exist between them. However, the simulated solvent-induced shift to the dipole moment of water is found to be significantly dampened with respect to typical literature values. The local nature of our methodology has further allowed us to evaluate the convergence of bulk properties with respect to the extent of the underlying one-electron basis set, ranging from single- ζ to full (augmented) quadruple- ζ quality. Albeit a pilot example, our work paves the way towards future studies of local effects and defects in more complex phases, e.g., liquid mixtures and even solid-state crystals.

TOC Graphic



Although the concept of local properties in extended bulk systems is easily contested as being ambiguous, its accurate and stringent determination by means of computational methodologies need not necessarily be. To that end, the properties of a single water molecule will be inherently different from those in liquid water, insofar as the electronic structure of an individual monomer in a droplet undoubtedly differs from that of its isolated counterpart due to charge transfer and delocalization effects. Divergences from isolated (vacuum) results will fundamentally be driven by perturbations to the local electronic structure of the embedded monomer, in addition to structural distortions mediated by the environment, and any consistent attempt at distinguishing individual property contributions to that of the bulk will inevitably prove difficult as a result. Although this will hold true for both the experimental and computational sides of things, the latter will benefit from being able to effectively inter-link alterations to the electronic structure in the solvation process that drives the transition from a water molecule in isolation to what constitutes a minimal droplet model.¹

If bulk properties are sought after as resolved into individual monomeric contributions, however, the employed methodology must partition the total quantities into a set of inherent terms that are subject to a number of requirements. First, local contributions must rely on objects that themselves remain spatially local, while simultaneously allowing for the effect of any nearby environment to fold into them. Second, the magnitude of a local (monomeric) property needs to stay satisfactorily invariant upon an enlargement of the underlying one-electron basis set—in addition to augmentations by diffuse functions—to make the involved simulations computationally tractable. Finally, the employed partitioning scheme should ideally stay sensitive enough to vary upon a refinement in the treatment of the electronic structure, if indeed correlation effects happen to be significant in the system at hand (which need not necessarily be the case). While the opposite would be desirable as well, the ability to encode an increased degree of physics into the computations and, in turn, benefit from this by having the added scrutiny reflected in the final results is, at the very least, theoretically

appealing. We will here demonstrate this very point by presenting results for the solvation shifts to both monomer ground-state energies and dipole moments, as obtained at both the uncorrelated Hartree-Fock (HF) level and by using exchange-correlation (xc) functionals belonging to various rungs of the Kohn-Sham density functional theory (KS-DFT) hierarchy.²

In a recent paper,³ we set out to demonstrate how to decompose mean-field (MF) molecular energies and dipole moments by means of a spatially localized orbital basis. Importantly, the decomposition of these properties was mediated by dividing the full 1-electron reduced density matrix (1-RDM) into a set of contributions associated with either the specific bonds or atoms of a given system, without recourse to an explicit partitioning of the electron density. Specifically, Ref. 3 outlined how to decompose first-order electronic properties in a basis of atom-specific 1-RDMs, $\{\boldsymbol{\delta}\}$, which themselves are defined according to the definition:

$$\boldsymbol{\delta}_K = \sum_{\sigma} \boldsymbol{\delta}_{K,\sigma} = \sum_{\sigma} \sum_i^{\mathcal{N}_{\sigma}} \mathbf{d}_{i,\sigma} \mathbf{p}_{i,\sigma}^K . \quad (1)$$

In turn, the objects in Eq. 1 are defined via a set of 1-RDMs, $\mathbf{d}_{i,\sigma} = \mathbf{C}_{i,\sigma} \mathbf{C}_{i,\sigma}^T$, unique to the individual occupied spin- σ molecular orbitals (MOs) of the system, $\mathbf{C}_{i,\sigma}$, and a set of population weights of all \mathcal{N}_{σ} MOs of α -/ β -spin on a given atom K , $\{\mathbf{p}^K\}$. The earlier investigations in Ref. 3 clearly emphasised how the population weights used to assign $\{\mathbf{d}\}$ should ideally not be drawn from regular Mulliken population analyses,⁴ but rather be recast into an alternative basis of reduced dimension, such as, the intrinsic AOs (IAOs) courtesy of Knizia.⁵

Using the atom-specific 1-RDMs in Eq. 1, an MF energy may be reformulated into the following, partitioned form amongst the $\mathcal{M}_{\text{atom}}$ atoms of the system at hand

$$E_{\text{MF}} = \sum_K^{\mathcal{M}_{\text{atom}}} E_{\text{elec},K}(\mathbf{D}, \boldsymbol{\delta}_K) + (E_{xc,K}(\boldsymbol{\rho}, \boldsymbol{\rho}_K)) + E_{\text{nuc},K} , \quad (2)$$

in terms of nuclear and electronic contributions that read as

$$E_{\text{nuc},K} = Z_K \sum_{K>L}^{\mathcal{M}_{\text{atom}}} \frac{Z_L}{|\mathbf{r}_K - \mathbf{r}_L|} \quad (3a)$$

$$E_{\text{elec},K} = \text{Tr}[\mathbf{T}_{\text{kin}}\boldsymbol{\delta}_K] + \frac{1}{2}(\text{Tr}[\mathbf{V}_K\mathbf{D}] + \text{Tr}[\mathbf{V}_{\text{nuc}}\boldsymbol{\delta}_K]) + \frac{1}{2} \sum_{\sigma} \text{Tr}[\mathbf{G}_{\sigma}(\mathbf{D})\boldsymbol{\delta}_{K,\sigma}] \quad (3b)$$

$$E_{xc,K} = \text{Tr}[\epsilon_{xc}(\boldsymbol{\rho})\boldsymbol{\rho}_K] . \quad (3c)$$

In Eq. 3a, Z_K and \mathbf{r}_K denote the nuclear charge and position of atom K , while the kinetic energy and nuclear attraction operators in Eq. 3b are denoted by \mathbf{T}_{kin} and \mathbf{V}_{nuc} , respectively, alongside the attractive potential associated with atom K , \mathbf{V}_K , and an effective Fock potential, \mathbf{G}_{σ} . In Eq. 3b, \mathbf{D} denotes the full, spin-summed 1-RDM, while the xc energy in Eq. 3c is expressed in terms of the computed energy density, ϵ_{xc} , as derived from the total electronic density, $\boldsymbol{\rho}$, and possibly its derivatives, which are quantities that may be trivially defined in an atom-specific manner, $\{\boldsymbol{\rho}_K\}$, by proceeding through $\{\boldsymbol{\delta}_K\}$. Likewise, a molecular dipole moment—irrespective of the employed level of MF theory—may be expressed as

$$\boldsymbol{\mu}_{\text{MF}} = \sum_K^{\mathcal{M}_{\text{atom}}} \boldsymbol{\mu}_{\text{elec},K}(\boldsymbol{\delta}_K) + \boldsymbol{\mu}_{\text{nuc},K} , \quad (4)$$

with nuclear and electronic contributions defined as

$$\boldsymbol{\mu}_{\text{nuc},K}^{\alpha} = Z_K \mathbf{r}_K^{\alpha} \quad (5a)$$

$$\boldsymbol{\mu}_{\text{elec},K}^{\alpha} = -\text{Tr}[\bar{\boldsymbol{\mu}}^{\alpha}\boldsymbol{\delta}_K] , \quad (5b)$$

in terms of AO dipole integrals, $\bar{\boldsymbol{\mu}}^{\alpha}$, for each of the Cartesian components, $\alpha = x, y, z$.

Provided with the theoretical setting above, a number of degrees of freedom in simulating a local bulk property still exist. The electronic contributions in Eqs. 3 and 5 are fundamentally defined in terms of a spatially localized MO basis, so which choice should one

ideally make for constructing this? In the context of the present study, we will use so-called intrinsic bond orbitals (IBOs),⁵ as defined by a standard Pipek-Mezey (PM) optimization scheme,⁶ but using atomic charges derived from IAO- rather than Mulliken-based populations.⁷ In Figs. S1 and S2 of the Supporting Information (SI), results obtained using IBOs are compared to corresponding results obtained using either regular PM or Foster-Boys⁸ (FB) localized MOs, in addition to results based on the equally lossless, but theoretically different energy density analysis (EDA) partitioning by Nakai,^{9,10} in which the total 1-RDM gets partitioned solely on the basis of which atoms the individual AOs are localized on (that is, irrespective of any further population measures). In our earlier study,³ IBOs were found to be generally far superior to both FB and PM localized MOs, as these tended to yield unsystematic results for a selection of different systems. However, all three kinds of localized MOs are found to give comparable results in the present context, cf. Figs. S1 and S2 of the SI, which present results using either of the pc-1 and aug-pc-1 basis sets,¹¹ respectively. In fact, the consistency of the results obtained using any of the three choices of localized MOs increases in augmenting the pc-1 basis set by diffuse functions, whereas the opposite is pronouncedly true in the case of the EDA results, which are devoid of any ground for interpretation in the more realistic of the two basis sets. This discrepancy is due to the fact that the contributions yielded by EDA are intimately tied to the employed AOs rather than actual atoms, which, in turn, renders the decomposition somewhat predefined and insensitive to changes in the electronic structure. Add to that the observation that the EDA partitioning possesses no strict basis set limit, and we are ultimately left with IBOs (alongside IAOs for computing the weights in Eq. 1) as a viable and rigorous option, and we will henceforth make exclusive use of these throughout the remainder of the present work.

In simulating local properties within a liquid bulk phase, one will need to sample dynamical effects both configurationally and radially by saturating results both with respect to fluctuations that come about as a result of temperature and the average local environment

surrounding a central monomer unit. Simulating the actual bulk may be done by means of several treatments, e.g., force fields¹² or *ab initio* molecular dynamics¹³ (AIMD), of which the latter constitutes the arguably most rigorous, albeit most costly option.¹⁴ The radial sampling of a local property is then carried out in a brute-force manner, by which a sphere surrounding a central monomer is incrementally enlarged until it encompasses a sufficient number of neighbours to warrant the intrinsic physics appropriately accounted for. Assuming that AIMD simulations have been properly equilibrated, the use of periodic boundary conditions (PBCs) renders the choice of central monomer to use further on relatively irrelevant. For consistency, however, it is usually wise to not only include different well-separated configurational snapshots in a sampling set, so as to avoid any autocorrelation between these, but also to extend the set of inputs by bulks centred around different monomers, which may be chosen upon at random. For samplings drawn from force-field simulations, it is reasonable just to focus on the monomer in closest proximity to the center of charge of the simulation body being employed.

In the present study, we will simulate the ground-state energy and molecular dipole moment of an embedded water monomer, as examples of prototypical local properties in a condensed phase, and we will base our study on 100 random snapshots from each of three fundamentally different samplings of liquid water: a simplified, yet flexible three-point TIP3P model¹⁸ of a spherical droplet, a more advanced four-point TIP4P/2005 model¹⁹ of a cubic box (with structural rigidity of the individual monomers enforced), and finally an AIMD simulation at the KS-DFT level of theory.²⁰ All bulks are reasonably assumed equilibrated, and the reader is referred to the original Refs. 21–23, respectively, for further details on how the simulations were conducted. The hydrogen bonding a central monomer unit participates in is compared for the three different samplings in Fig. 1, while Figs. S3 through S5 of the SI seek to compare these in terms of their bulk composition, fluctuations to their structural configurations, as well as their local orientational tetrahedral arrangements, respectively. As

is evident from our comparisons of the bulks, quantifiable differences are indeed observed to exist between them, and it will hence prove instructive to study the extent to which these subtle, dynamical effects influence simulations of local properties within the liquid phase.

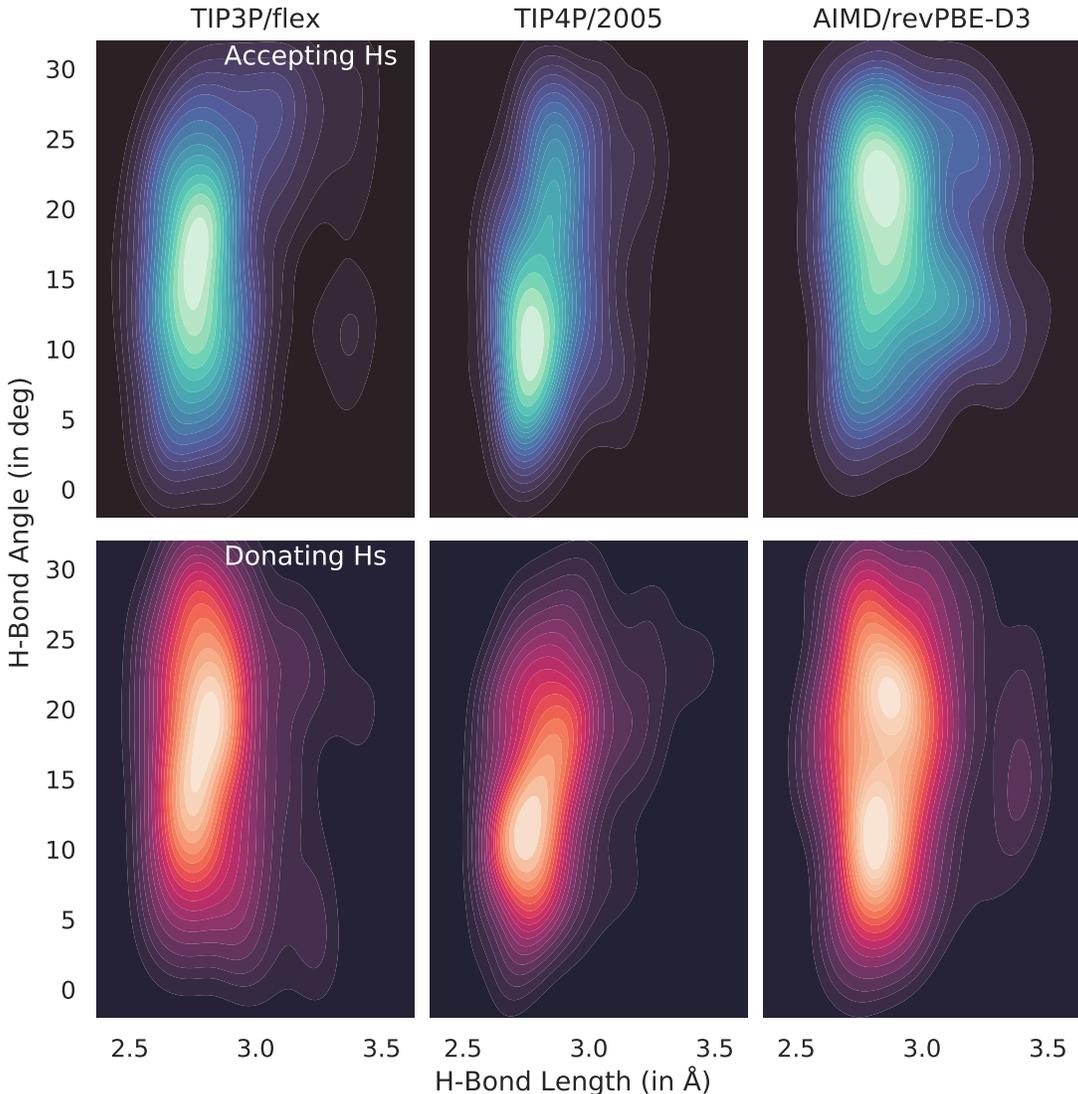


Figure 1: Distributions of hydrogen bonds whereby a central monomer unit either accepts (upper panel) or donates (lower panel) a proton in the three different samplings. As per convention,²⁴ a hydrogen bond is defined in terms of a max. O–O distance of 3.5 Å and a max. angle between the donating O, donating H, and accepting O of 30°.

In terms of the electronic structure treatment, we will assess the quality of standard, restricted HF as well as a total of 4 different *xc* functionals, namely, B3LYP,^{25,26} CAM-B3LYP,²⁷ M06-2X,²⁸ and ω B97M-V.²⁹ This selection of density functional approximations

(DFAs) is motivated, in part, by the favoured candidates of the rigorous evaluation of KS-DFT for the simulation of dipole moments in Ref. 30. Of these, B3LYP is arguably the most popular, not only among its own class of (hybrid) generalized gradient approximations (GGAs), but across all of the available xc functionals to date. The Coulomb-attenuated version of the functional, CAM-B3LYP, differs from the native B3LYP by implementing a mixing of short- and long-range exchange via the standard error function, while ω B97M-V and M06-2X are meta-GGAs of the range-separated and global hybrid kind, respectively.

All KS-DFT calculations have been run with the PySCF program package,^{31–33} while all subsequent decompositions have been performed using the `decodense` code.³⁴ The PySCF default number of (radial, angular Lebedev) grid points was used for the H and O atoms, i.e., (50, 302) and (75, 302), respectively, in combination with the standard pruning scheme, except for the double-grid integration involved in the evaluation of the nonlocal VV10 correlation for describing dispersion (van der Waals) interactions,³⁵ for which reduced SG-1 grids were used.³⁶ The effect of employing denser quadrature grids was assessed for the ω B97M-V and M06-2X functionals, as these involve complex expressions for the exchange inhomogeneity factor.³⁷ For both DFAs, negligible errors were found, i.e., much smaller than those due to the use of a truncated radial extent (see below), and on par with errors due to the density fitting approximation,³⁸ which has similarly been invoked throughout due to its cost reductions.

Before embarking on simulating properties in each of the three bulks discussed above, we will first discuss the design of a suitable computational protocol. In Figs. S6 and S7 of the SI, results of a coarse-grained radial sampling extending outwards from a central monomer under consideration—at the B3LYP/(aug-)pc-1 level of theory—are observed to saturate at a radius well before the largest extension of $r = 5.5 \text{ \AA}$, a distance corresponding safely to the inclusion of the second solvation shell and a sphere encompassing in excess of 30 neighbouring monomers on average in all three samplings. No apparent differences in the

convergence profiles are observed by adding diffuse functions. However, bulks of this extent are bound to render KS-DFT intractable in the enlarged and augmented basis sets that are compulsory to the determination of dipole moments. For this reason, we will here advocate in favour of the use of background point charges, which are observed from Figs. S6 and S7 to accelerate radial convergence—particularly in the case of dipole moments—and thus allow for reduced bulks of $r = 4.0 \text{ \AA}$, compromising the final accuracy of our simulations by only a fraction of a kcal/mol or a Debye in the case of energies and dipole moments, respectively.³⁹

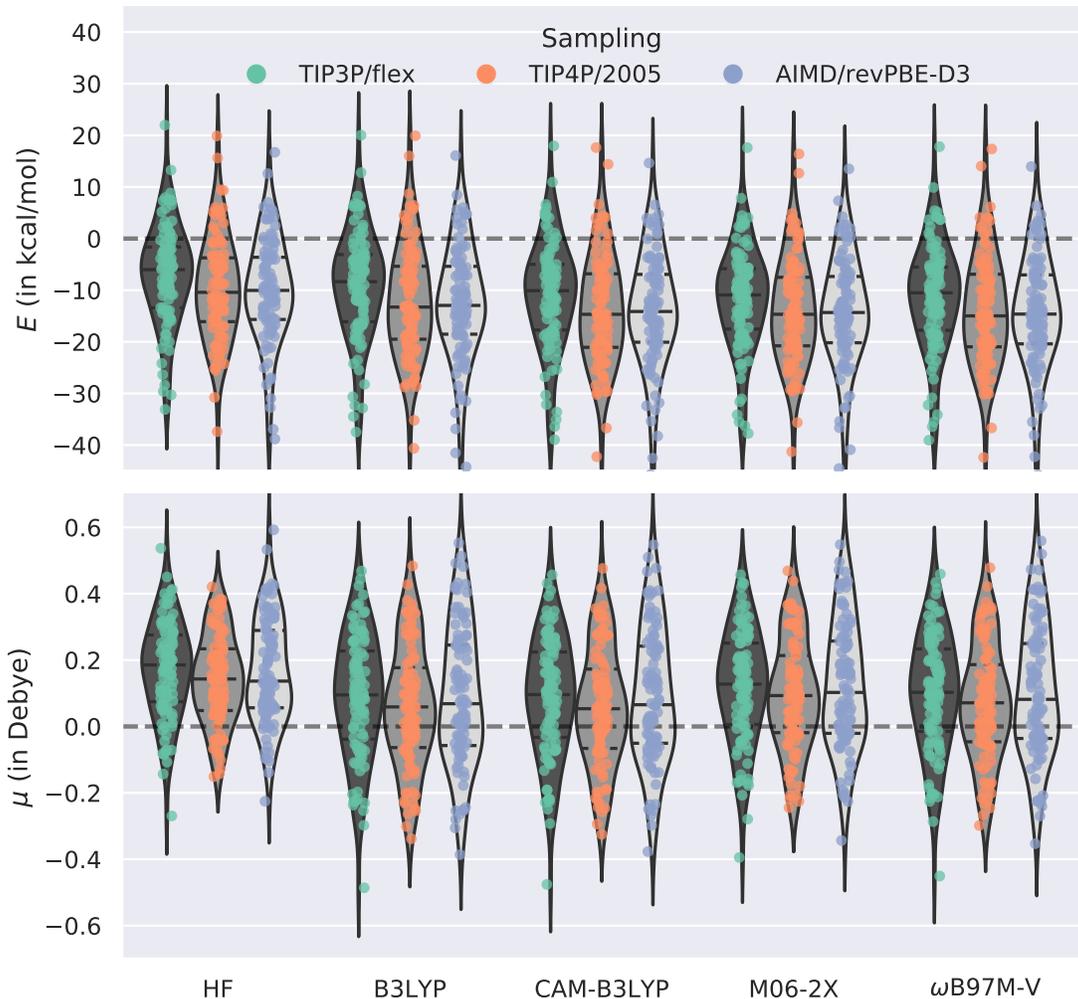


Figure 2: Distributions of solvation energies (upper panel) and dipole moments (lower panel) for each of the three samplings and each of the four xc functionals (using the aug-pc-1 basis set). In each violin, the quartiles of the underlying distribution have been displayed.

Having established an efficient and functional protocol, the main results of the present

study now follow. In Fig. 2, we report energies and molecular dipole moments, both measured as solvation shifts to corresponding vacuum results. In all three bulk samplings, HF and all of the DFAs under consideration have been evaluated in the aug-pc-1 basis set of double- ζ quality. From the results for both properties in Fig. 2, we observe little variation in-between the different bulks. At the same time, we note how the HF results slightly differ from those obtained using KS-DFT. Within any given bulk sampling, the individual distributions are difficult to recognize from one another, bar a minor potential shift, all giving mean solvation energies and dipole moments of approximately -10 kcal/mol and 0.1 D, respectively, with a pronounced positive-valued tail for the latter property. Given these similarities, we may restrict our attention to only a single combination of DFA and bulk sampling in moving towards extended basis sets, e.g., the B3LYP *xc* functional and the AIMD variant. Results in basis sets ranging from augmented single- to quadruple- ζ quality are presented in Fig. 3, which—besides verifying all conclusions based on the aug-pc-1 results in Fig. 2—reiterate the stable and rapid convergence of our decompositions with respect to basis set size first investigated in Ref. 3 (cf. also Fig. S8 of the SI, which presents Fig. 3 on a linear scale).

To inspect the findings of Figs. 2 and 3 in greater detail, we next group individual results on the basis of their local networks of hydrogen bonds, namely, the number of protons which a central monomer accepts and donates in these, cf. Fig. 4. In accordance with expectation, the monomers that accept more protons than they donate are generally stabilized more than those for which the opposite is true. This fact is also exemplified by Fig. S9 of the SI, which presents results for decomposed dipole moments along a constrained scan of the hydrogen bond in a simple water dimer. Further to that, a convincing pattern is observed to emerge for all three samplings in which shifts to ground-state energies and dipole moments are inversely proportional, in the sense that those monomers that energetically favour embedding also have net amplifications of their polar charge distributions. In support of this observation, these two simulated properties are plotted against one another in Fig. S10 of the SI where

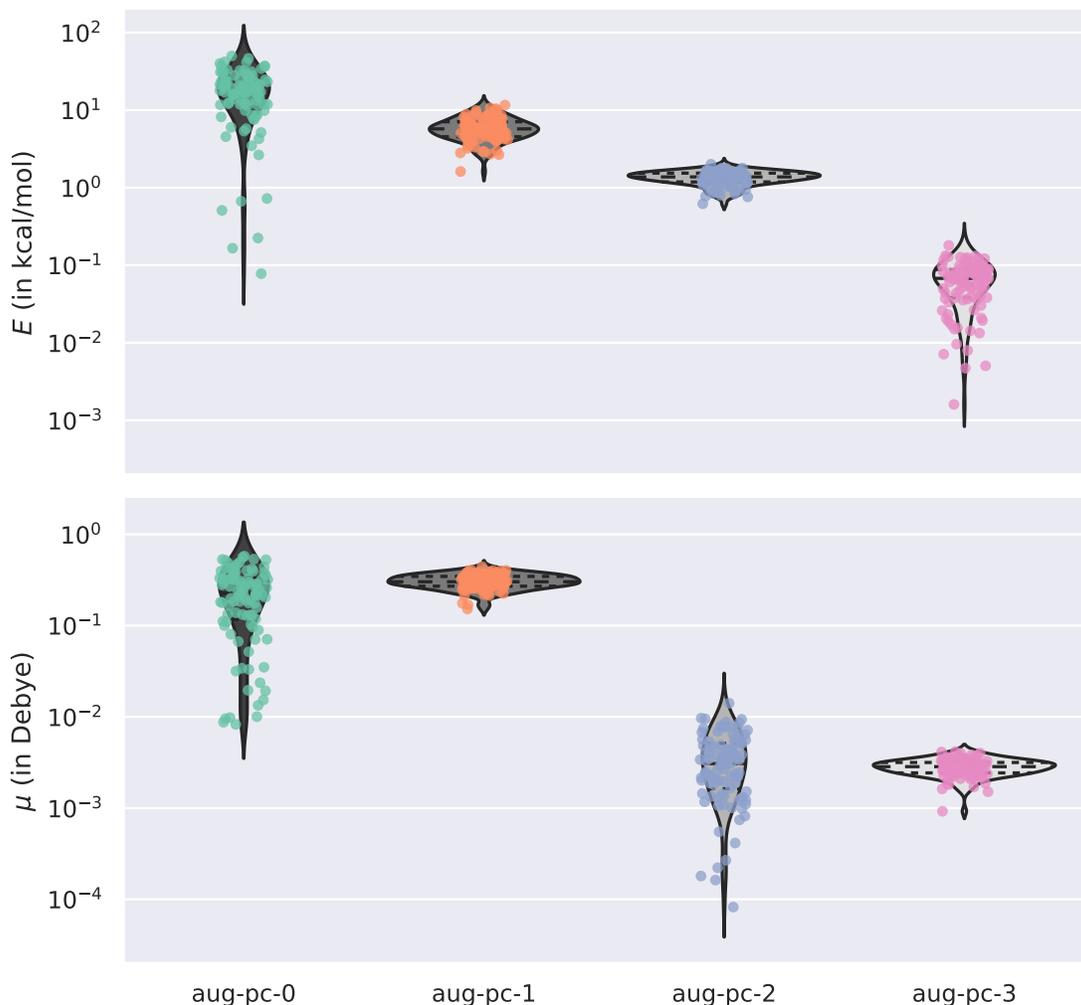


Figure 3: Convergence of (absolute) solvation energies and dipole moments of H_2O in the AIMD/revPBE-D3 sampling, as calculated at the B3LYP/aug-pc- x level of theory ($x = 0-3$) and measured on a log-scale against results in the largest aug-pc-3 basis set.

their signs and—to a somewhat lesser degree—their magnitudes are observed to correlate. However, what neither of Figs. 4, S9, and S10 succeed in quantifying is the extent to which a monomer is actually perturbed in a given configuration; nor do they fully show what the effects of potential asymmetries present in the hydrogen bond network surrounding said monomer amount to, nor how these factors explicitly correlate with the sign and magnitude of changes to the observed properties.⁴⁰ For this reason, all further attempts at unravelling and clarifying such relationships in bulk phases will be postponed to future studies.

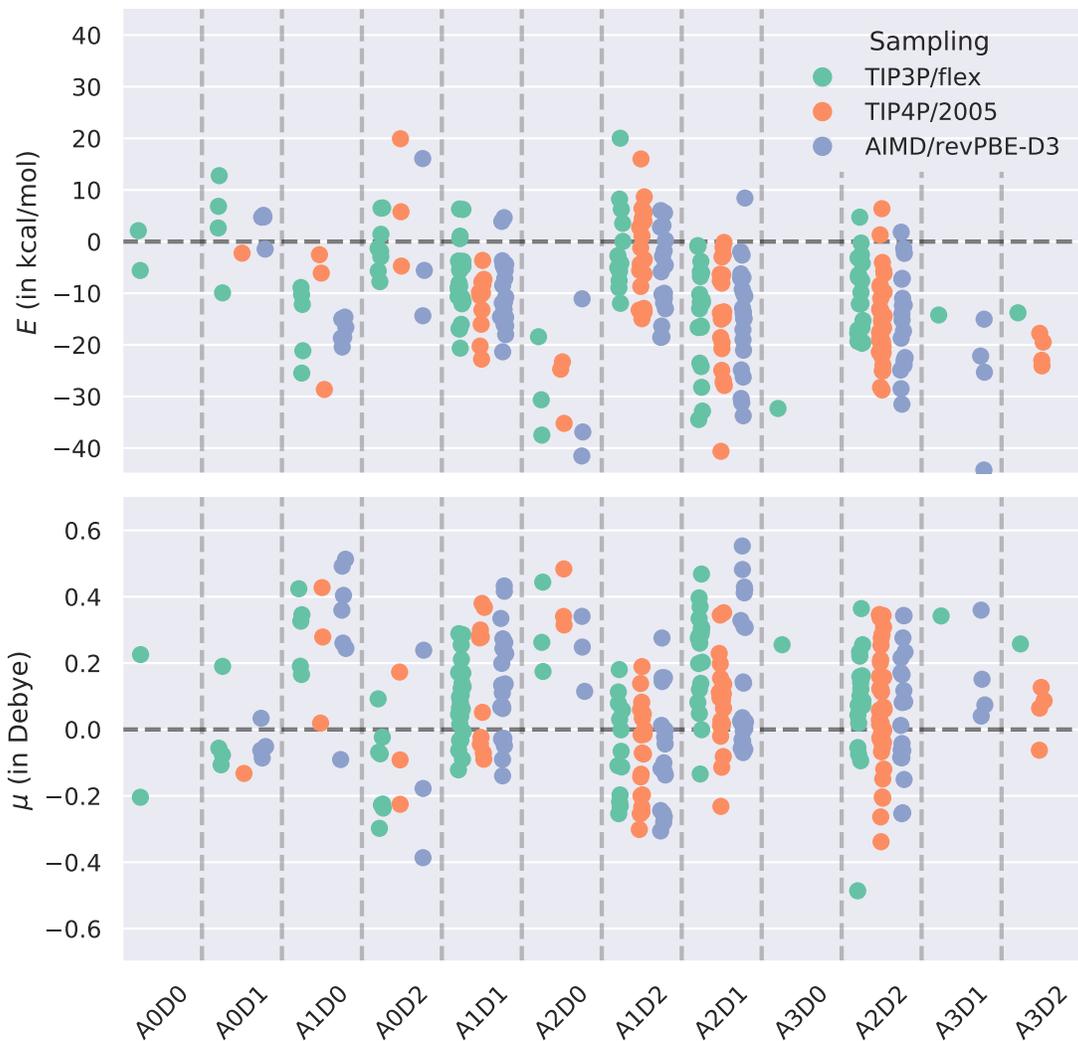


Figure 4: Solvation energies and dipole moments of H_2O in all three samplings, as calculated at the B3LYP/aug-pc-1 level of theory. Results have been grouped on the basis of the number of protons accepted (A) and donated (D) by a central monomer in hydrogen bonds (Fig. 1).

Finally, we note how we predict the overall (averaged) shift to the dipole moment of H_2O to be less than what has been reported by others, e.g., in the two recent studies by Han, Isborn, and Shi (Ref. 22) and by Zhu and Van Voorhis (Ref. 23). In the former of these two, various schemes for assigning partial atomic charges (in response to local embedding environments) were evaluated, while in the latter, the authors employed the self-attractive Hartree (SAH) decomposition of the total electron density—as introduced in Ref. 41—to simulate said dipole moments. Focusing on the latter study, the decomposition used in

Ref. 23 is characterized by a single, adjustable parameter that governs the degree of density localization, whereas the scheme used throughout the present work has a dependency on the local MOs and charge population proxy of choice. However, whereas practically no variance with respect to the employed MO basis is found herein, cf. Figs. S1 and S2 of the SI, the authors behind the SAH study found two distinct pictures, one localized and the other delocalized with less and more polarizable monomer dipoles, respectively, depending on the strength of their control parameter. Agreement between Ref. 23 and the present study is met in the local limit, which yields the weakest solvation shifts, in our case by means of a protocol that remains spatially local in contrast to earlier attempts at employing localized MOs for this very purpose.⁴² In Fig. 5, we show how too crude a truncation of the local environment in a simulation of local properties—particularly in combination with the use of background point charges^{43,44}—will lead to an overestimation of solvation effects, notably, in the case of dipole moments. As such, only upon saturating all possible charge (de)localization and polarization effects, even in a local picture frame, may reliable results be obtained,⁴⁵ and these properties happen to differ only very moderately from results for isolated monomers in the averaged, rather than instantaneous limit. This also serves as a warning of how the common practice of extending on conclusions drawn from smaller, idealized water clusters to the bulk limit may be inherently problematic,⁴⁶ as local properties are anything but insensitive to the heterogeneous local environment of water and its evolution over time. Unlike the situation in ice,^{47,48} networks of hydrogen bonds in ambient water are disordered, exhibiting a dynamical mixture of various identifiable configurations,^{49–51} and the nature of molecular dipoles in solution will necessarily depend on this disorder in complex manners.⁵²

In summary, we have demonstrated how an atom-centric decomposition theory for partitioning mean-field properties in a basis of spatially localized MOs may be leveraged as a robust protocol for probing localized electronic structure in condensed-phase systems,³ with an initial application to solvation properties in liquid water. We find that results for both

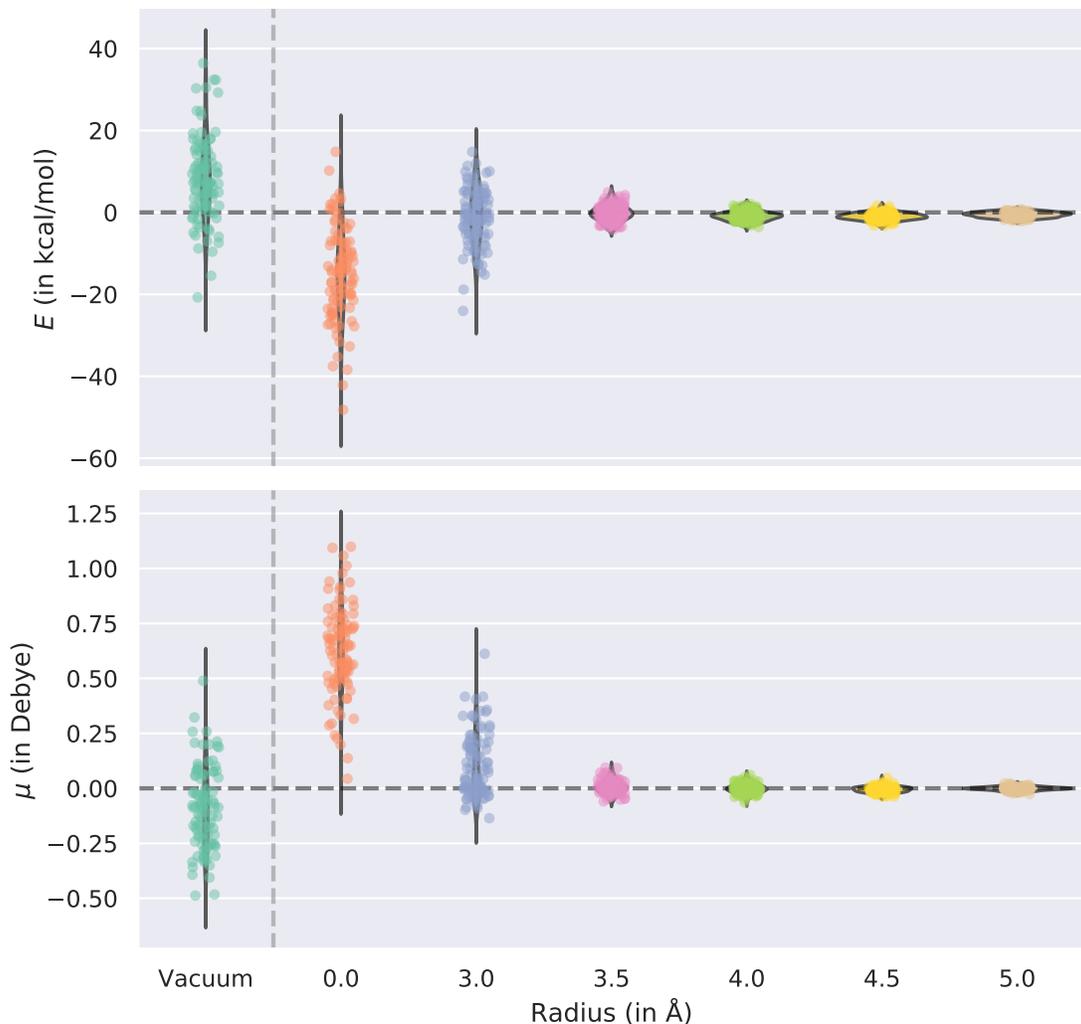


Figure 5: Convergence of solvation energies and dipole moments of H_2O in the TIP3P/flex sampling against results at a radius of $r = 5.5 \text{ \AA}$, as calculated at the B3LYP/aug-pc-1 level of theory and simulated using background point charges, except where noted otherwise.

energies and dipole moments in the bulk phase are pronouncedly invariant, not only with respect to the sophistication of the underlying sampling procedure, but also to that of the employed DFA and the involved level of theory. Importantly, we have been able to move beyond the prevailing use of modest basis sets, restating the stable and rapid convergence of our decompositions through basis sets of augmented quadruple- ζ quality. These results have verified our in-depth analyses in the pragmatic, but in our case sufficient, aug-pc-1 basis set.

It is difficult to ascribe divergences of our simulated dipole moments from literature values, e.g., obtained by means of experimental X-ray diffraction techniques ($\mu = 2.9 \pm 0.6$ D),⁵³ as being caused by missing accounts of residual electron correlation, since all of our tested DFAs agree with one another, despite their functional differences. In addition, even experimental determinations of the monomer dipole in liquid water are not free of ambiguities, as these will all rely on some kind of fitting procedure to a model for the involved charge transfer processes, unlike in vacuum ($\mu = 1.855$ D) where Stark effect measurements can determine equilibrium dipole moments.⁵⁴ Instead, both the assessment and interpretation of this and related properties in solution may need be reevaluated. Measured in close proximity of the isolated monomer alone, the dipole moment of water is clearly enhanced in the condensed phase. At a distance, however, its solvation shift is consistently observed to be significantly dampened due to an effective averaging of effects related to the relocation of electronic density. This is further motivated by the fact that our final solvation energies compare nicely with the enthalpy of vaporization of liquid water (ca. 10.5 kcal/mol at 25 °C).⁵⁵ We will here argue that the present results—given the convincing saturation with respect to both electron correlation and basis set extension—provide a unique view on the dynamical interplay of individual monomers in liquid, ambient water, with potential applications of interest in data-driven models^{56,57} and as a training pool in the design of new force fields.^{58,59}

Acknowledgments

This work was supported by a generous research grant (no. 37411) from VILLUM FONDEN (a part of THE VELUX FOUNDATIONS). Preliminary work was carried out while the author was still based at the School of Chemistry, University of Bristol, and The Independent Research Fund Denmark is gratefully acknowledged for financial support throughout this early phase, during which work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol. Finally, the author wishes to

thank Liang Shi (University of California, Merced) and Tianyu Zhu (Caltech) for sharing the TIP4P/2005 and AIMD/revPBE-D3 samplings of Ref. 22 and 23, respectively.

Supporting Information

The Supporting Information collects a number of additional results in support of the main results reported in the present Letter. The use of IBOs, PM, and FB localized MOs are compared in Figs. S1 and S2 (also in comparison to Nakai’s EDA partitioning), in the pc-1 and aug-pc-1 basis sets, respectively. Fig. S3 depicts the radial distributions of neighbouring monomers within each of the three samplings, while Fig. S4 depicts the distributions of bond lengths and angles for the central monomers within the samplings. Fig. S5 quantifies the distributions of hydrogen bonds in the three sampling through results for the corresponding orientational tetrahedral orders to emphasize any deviations from regular tetrahedral configurations. The convergences of results obtained for the energy and dipole moment of an embedded water monomer with outwards distance are evaluated in Figs. S6 and S7, again for each of the pc-1 and aug-pc-1 basis sets, respectively, and Fig. S8 presents the basis set convergence results behind Fig. 3 plotted on a linear scale, rather than as absolute numbers on a log-scale. Finally, results for decomposed molecular dipole moments along an optimized (constrained) potential energy scan of the water dimer are presented in Fig. S9, while the two properties in question are finally correlated against one another in Fig. S10.

References

- (1) Rognoni, A.; Conte, R.; Ceotto, M. How Many Water Molecules are Needed to Solvate One? Chem. Sci. **2021**, 12, 2060.
- (2) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy. AIP Conf. Proc. **2001**, 577, 1.
- (3) Eriksen, J. J. Mean-Field Density Matrix Decompositions. J. Chem. Phys. **2020**, 153, 214109.
- (4) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. J. Chem. Phys. **1955**, 23, 1833.
- (5) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge Between Quantum Theory and Chemical Concepts. J. Chem. Theory Comput. **2013**, 9, 4834.
- (6) Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for *Ab Initio* and Semiempirical Linear Combination of Atomic Orbital Wave Functions. J. Chem. Phys. **1989**, 90, 4916.
- (7) Lehtola, S.; Jónsson, H. Pipek-Mezey Orbital Localization Using Various Partial Charge Estimates. J. Chem. Theory Comput. **2014**, 10, 642.
- (8) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. Rev. Mod. Phys. **1960**, 32, 300.
- (9) Nakai, H. Energy Density Analysis with Kohn-Sham Orbitals. Chem. Phys. Lett. **2002**, 363, 73.
- (10) Kikuchi, Y.; Imamura, Y.; Nakai, H. One-Body Energy Decomposition Schemes Revisited: Assessment of Mulliken-, Grid-, and Conventional Energy Density Analyses. Int. J. Quantum Chem. **2009**, 109, 2464.

- (11) Jensen, F. Polarization Consistent Basis Sets: Principles. J. Chem. Phys. **2001**, 115, 9113.
- (12) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. Chem. Rev. **2016**, 116, 7501.
- (13) Iftimie, R.; Minary, P.; Tuckerman, M. E. *Ab Initio* Molecular Dynamics: Concepts, Recent Developments, and Future Trends. Proc. Natl. Acad. Sci. **2005**, 102, 6654.
- (14) In terms of accuracy, however, the data-driven many-body models from the Paesani group (described in Refs. 15–17 and collectively reviewed in Ref. 12) constitute the current state-of-the-art for water simulations, providing highly accurate descriptions of a wealth of properties in both the gas and condensed phase.
- (15) Babin, V.; Leforestier, C.; Paesani, F. Development of a “First Principles” Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. J. Chem. Theory Comput. **2013**, 9, 5395.
- (16) Babin, V.; Medders, G. R.; Paesani, F. Development of a “First Principles” Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. J. Chem. Theory Comput. **2014**, 10, 1599.
- (17) Medders, G. R.; Babin, V.; Paesani, F. Development of a “First Principles” Water Potential with Flexible Monomers. III: Liquid Phase Properties. J. Chem. Theory Comput. **2014**, 10, 2906.
- (18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. **1983**, 79, 926.

- (19) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. J. Chem. Phys. **2005**, 123, 234505.
- (20) In all three bulk samplings, the gauge origin of the AO dipole integrals have been fixed to coincide with the position of the central oxygen atom under investigation, which has further been translationally moved to the position, $\mathbf{R}_O = (0, 0, 0)$.
- (21) Faber, F. A.; Christensen, A. S.; Huang, B.; von Lilienfeld, O. A. Alchemical and Structural Distribution Based Representation for Universal Quantum Machine Learning. J. Chem. Phys. **2018**, 148, 241717.
- (22) Han, B.; Isborn, C. M.; Shi, L. Determining Partial Atomic Charges for Liquid Water: Assessing Electronic Structure and Charge Models. J. Chem. Theory Comput. **2021**, 17, 889.
- (23) Zhu, T.; Van Voorhis, T. Understanding the Dipole Moment of Liquid Water from a Self-Attractive Hartree Decomposition. J. Phys. Chem. Lett. **2021**, 12, 6.
- (24) Luzar, A.; Chandler, D. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. Phys. Rev. Lett. **1996**, 76, 928.
- (25) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. J. Chem. Phys. **1993**, 98, 5648.
- (26) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. J. Phys. Chem. **1994**, 98, 11623.
- (27) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). Chem. Phys. Lett. **2004**, 393, 51.

- (28) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. Theor. Chem. Acc. **2008**, 120, 215.
- (29) Mardirossian, N.; Head-Gordon, M. ω B97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation. J. Chem. Phys. **2016**, 144, 214110.
- (30) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. J. Chem. Theory Comput. **2018**, 14, 1969.
- (31) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: The Python-Based Simulations of Chemistry Framework. WIREs Comput. Mol. Sci. **2018**, 8, e1340.
- (32) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H.; Eriksen, J. J.; Gao, Y.; Guo, S.; Hermann, J.; Hermes, M. R.; Koh, K.; Koval, P.; Lehtola, S.; Li, Z.; Liu, J.; Mardirossian, N.; McClain, J. D.; Motta, M.; Mussard, B.; Pham, H. Q.; Pulkin, A.; Purwanto, W.; Robinson, P. J.; Ronca, E.; Sayfutyarova, E. R.; Scheurer, M.; Schurkus, H. F.; Smith, J. E. T.; Sun, C.; Sun, S.-N.; Upadhyay, S.; Wagner, L. K.; Wang, X.; White, A.; Whitfield, J. D.; Williamson, M. J.; Wouters, S.; Yang, J.; Yu, J. M.; Zhu, T.; Berkelbach, T. C.; Sharma, S.; Sokolov, A. Y.; Chan, G. K.-L. Recent Developments in the PySCF Program Package. J. Chem. Phys. **2020**, 153, 024109.
- (33) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent Developments

- in LIBXC — A Comprehensive Library of Functionals for Density Functional Theory. Software X **2018**, 7, 1.
- (34) DECODENSE: A Decomposed Mean-Field Theory Code, See <https://github.com/januseriksen/decodense>.
- (35) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional: The Simpler the Better. J. Chem. Phys. **2010**, 133, 244103.
- (36) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A Standard Grid for Density Functional Calculations. Chem. Phys. Lett. **1993**, 209, 506.
- (37) Dasgupta, S.; Herbert, J. M. Standard Grids for High-Precision Integration of Modern Density Functionals: SG-2 and SG-3. J. Comput. Chem. **2017**, 38, 869.
- (38) Dunlap, B. I. Robust and Variational Fitting. Phys. Chem. Chem. Phys. **2000**, 2, 2113.
- (39) The TIP3P/flex and AIMD/revPBE-D3 samplings both use TIP3P charge distributions, namely, $q(\text{O}) = -0.834$ and $q(\text{H}) = 0.417$, while the TIP4P/2005 sampling makes use of its own three-point model, i.e., $q(\text{M}) = -1.1128$ and $q(\text{H}) = 0.5564$, with the negative charge placed on a dummy atom (M) at a distance of 0.1546 Å away from the oxygen along the $\angle(\text{H}-\text{O}-\text{H})$ bisector.
- (40) Bakó, I.; Daru, J.; Pothoczki, S.; Pusztai, L.; Hermansson, K. Effects of H-bond Asymmetry on the Electronic Properties of Liquid Water — An AIMD Analysis. J. Mol. Liq. **2019**, 293, 111579.
- (41) Zhu, T.; de Silva, P.; Van Voorhis, T. Self-Attractive Hartree Decomposition: Partitioning Electron Density into Smooth Localized Fragments. J. Chem. Theory Comput. **2018**, 14, 92.
- (42) Bakó, I.; Mayer, I. On Dipole Moments and Hydrogen Bond Identification in Water Clusters. J. Phys. Chem. A **2016**, 120, 4408.

- (43) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. Dipole and Quadrupole Moments of Liquid Water Calculated within the Coupled Cluster/Molecular Mechanics Method. Chem. Phys. Lett. **2002**, 364, 379.
- (44) Kemp, D. D.; Gordon, M. S. An Interpretation of the Enhancement of the Water Dipole Moment Due to the Presence of Other Water Molecules. J. Phys. Chem. A **2008**, 112, 4885.
- (45) Bakó, I.; Mayer, I. Hierarchy of the Collective Effects in Water Clusters. J. Phys. Chem. A **2016**, 120, 631.
- (46) Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. The Water Dipole Moment in Water Clusters. Science **1997**, 275, 814.
- (47) Coulson, C. A.; Eisenberg, D. Interactions of H₂O Molecules in Ice I. The Dipole Moment of an H₂O Molecule in Ice. Proc. R. Soc. London. Ser. A **1966**, 291, 445.
- (48) Batista, E. R.; Xantheas, S. S.; Jónsson, H. Molecular Multipole Moments of Water Molecules in Ice Ih. J. Chem. Phys. **1998**, 109, 4546.
- (49) Liu, J.; He, X.; Zhang, J. Z. H. Structure of Liquid Water — A Dynamical Mixture of Tetrahedral and ‘Ring-and-Chain’ Like Structures. Phys. Chem. Chem. Phys. **2017**, 19, 11931.
- (50) Ruiz Pestana, L.; Mardirossian, N.; Head-Gordon, M.; Head-Gordon, T. *Ab Initio* Molecular Dynamics Simulations of Liquid Water Using High Quality Meta-GGA Functionals. Chem. Sci. **2017**, 8, 3554.
- (51) Riera, M.; Lambros, T. T., E. Nguyen; Götz, A.; Paesani, F. Low-Order Many-Body Interactions Determine the Local Structure of Liquid Water. Chem. Sci. **2019**, 10, 8211.
- (52) Torii, H. Extended Nature of the Molecular Dipole of Hydrogen-Bonded Water. J. Phys. Chem. A **2013**, 117, 2044.

- (53) Badyal, Y. S.; Saboungi, M.-L.; Price, D. L.; Shastri, S. D.; Haeffner, D. R.; Soper, A. K. Electron Distribution in Water. J. Chem. Phys. **2000**, 112, 9206.
- (54) Clough, S. A.; Beers, Y.; Klein, G. P.; Rothman, L. S. Dipole Moment of Water from Stark Measurements of H₂O, HDO, and D₂O. J. Chem. Phys. **1973**, 59, 2254.
- (55) Lide, D. R. CRC Handbook of Chemistry and Physics. 85th ed.; CRC Press, 2004.
- (56) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. J. Chem. Theory Comput. **2019**, 15, 3678.
- (57) Veit, M.; Wilkins, D. M.; Yang, Y.; DiStasio Jr., R. A.; Ceriotti, M. Predicting Molecular Dipole Moments by Combining Atomic Partial Charges and Atomic Dipoles. J. Chem. Phys. **2020**, 153, 024113.
- (58) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. J. Phys. Chem. B **2013**, 117, 9956.
- (59) Lambros, E.; Paesani, F. How Good are Polarizable and Flexible Models for Water: Insights from a Many-Body Perspective. J. Chem. Phys. **2020**, 153, 060901.