

# Making Many-Body Interactions Nearly Pairwise Additive: The Polarized Many-Body Expansion Approach

Srimukh Prasad Veccham, Joonho Lee,\* and Martin Head-Gordon\*

*Department of Chemistry, University of California, Berkeley, California 94720, USA*

*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

E-mail: linusjoonho@gmail.com; mhg@cchem.berkeley.edu

## Abstract

The Many-Body Expansion (MBE) is a useful tool to simulate condensed phase chemical systems, often avoiding the steep computational cost of usual electronic structure methods. However, it often requires higher than 2-body terms to achieve quantitative accuracy. In this work, we propose the Polarized MBE (PolBE) method where each MBE energy contribution is treated as an embedding problem. In each energy term, a smaller fragment is embedded into a larger, polarized environment and only a small region is treated at the high-level of theory using embedded mean-field theory. The role of polarized environment was found to be crucial in providing quantitative accuracy at the 2-body level. PolBE accurately predicts non-covalent interaction energies for a number of systems, including  $\text{CO}_2$ , water, and hydrated ion clusters, with a variety of interaction mechanisms, from weak dispersion to strong electrostatics considered in this work. We further demonstrate that the PolBE interaction energy is predominantly pairwise unlike the usual vacuum MBE which requires higher-order terms to achieve similar accuracy. We numerically show that PolBE often performs better than other widely used embedded MBE methods such as the electrostatically embedded MBE. Owing to the lack of expensive diagonalization of Fock matrices and its embarrassingly parallel nature, PolBE is a promising way to access condensed phase systems with hybrid density functionals that are difficult to treat with currently available methods.

# Introduction

Simulating condensed phase systems is considered the holy grail of quantum chemistry. However historically, the majority of quantum chemistry research has focused on molecular properties rather than bulk properties of liquids. The major hurdle towards calculating and predicting properties of bulk materials is the steep computational scaling in modeling them. For decades, a major research goal of the quantum chemistry community has been to formulate more accurate and inexpensive computational methodologies to predict properties of bulk materials.

For systems comprising of non-covalently bonded fragments, the many-body expansion method has been used as way to circumvent the steep hyper-linear scaling of quantum chemistry methods. The Many-Body Expansion (MBE) is a technique for calculating the properties of a large system by (1) partitioning the system into  $N$  fragments, (2) calculating fragment properties, and (3) reassembling them using the many-body expansion equation. This approach reduces a large supersystem calculation into a large number of much smaller fragment calculations, and each of them is completely independent of others. Thereby, MBE calculations are “embarrassingly parallel.” While the MBE is in principle exact, it provides computational advantages only when it is truncated up to a certain order. Truncated many-body expansions are not exact, and its accuracy can heavily depend on the underlying chemical interactions in the system. Some applications of the many-body expansion method are determination of lattice energies<sup>1</sup> and studying polymorphism in molecular crystals.<sup>2</sup> Many-body based cluster corrections to calculations with periodic boundary conditions have also served as powerful tools to study crystal structures using highly accurate wavefunction techniques.<sup>3</sup> MBEs have also been used to understand and incorporate the effects of mutual polarization in classical molecular dynamics force fields.<sup>4</sup>

The convergence of MBE can be accelerated using embedding, in which each  $k$ -body term is calculated in the presence of the remaining  $(N - k)$  fragments (hereafter referred to as the ‘environment’). An active research question is then how one models this environment

and its interaction with the  $k$ -body term at a given electronic structure level. Quantum embedding can be performed in the context of Green's functions,<sup>5–8</sup> densities or potentials,<sup>9–15</sup> or density matrices.<sup>16,17</sup> In many chemical phenomena, the interesting chemical activity like binding or chemical reaction occurs only in a small part of the chemical system. Embedding techniques, taking advantage of this “nearsightedness”<sup>18</sup> in order to divide the system into two parts: A - the region of interest where an actual chemical phenomenon occurs and B - the environment which surrounds the system but does not directly take part in the actual chemical process. Even though the environment does not directly participate in the chemical phenomenon, it sufficiently modulates the chemical process occurring in the system that it cannot be neglected. Embedding can also be done by modelling the environment using molecular mechanics which constitute widely used QM:MM methods like ONIOM.<sup>19</sup> ONIOM and its extensions<sup>20</sup> have been extensively used to study a wide range of systems and phenomena from heterogeneous catalysis to biological macromolecules.<sup>21</sup> However, ONIOM does not contain any explicit system-environment coupling terms in the hamiltonian, and cannot accurately capture the effects of environment. Embedded MBEs (eMBE) make use of quantum embedding techniques to describe the  $k$ -body term using an accurate and often more expensive computational method while using a much cheaper method to represent the environment. An eMBE truncated at the 2-body level aims to reduce the relative importance of the higher order terms (3-body and higher) by folding in these many-body effects into the 1-body and the 2-body terms. Different models of the environment, which can capture distinct physical effects, are capable of encoding these many-body effects with varying degrees of accuracy and computational expense.

A simple way to go beyond vacuum MBE, where we do not include any environment, is to model the environment using atom-centered point charges, termed the Electrostatically-Embedded Many-Body Expansion (EE-MBE).<sup>22–24</sup> These atom-centered point charges incorporate some effect of electrostatic interaction while the finite-extent of electron density, exchange interaction, and polarization of the environment are neglected. EE-MBE is fairly

simple to implement, however its accuracy is very dependent on the actual charges used for embedding.<sup>25</sup>

One way to go beyond a simple charge embedding model is to use the Embedded Many-Body Expansion of Manby and co-workers which uses atom-centered gaussians to represent electrostatic interactions of the environment and a simple empirical model for Pauli repulsion.<sup>26</sup> A slightly more sophisticated embedding method is the Variational Many-Body Expansion (VMBE),<sup>27</sup> which builds upon the 1-body X-Pol wavefunction,<sup>28–33</sup> introduced by Jiali Gao and co-workers. In the X-Pol wavefunction, the supersystem wavefunction is written as a Hartree-product of the monomer wavefunctions. Inter-fragment coulombic interactions are simplified by using point charges in order to avoid computing two-electron integrals for the whole system. The X-Pol wavefunction can either be optimized variationally or by using a double self-consistent field procedure.<sup>28</sup> Dispersion, exchange repulsion, and charge transfer effects are approximated using empirical Lennard-Jones terms. A variation of the X-Pol wavefunction, X-Pol-X, can be used to incorporate the effect of exchange repulsion non-empirically.<sup>34</sup> While the X-Pol wavefunction treats these interactions empirically, VMBE makes an attempt to treat them quantum mechanically at the 2-body level by redefining the fragment in the X-Pol wavefunction to contain two monomers. It is important to note that the environment density is updated for each of the dimer calculations in order to include complete mutual polarization between the dimer and its environment. The Fragment Molecular Orbital (FMO) method is another idea based on the MBE, which uses densities obtained from monomer wavefunctions to represent the fragments in the environment.<sup>35–38</sup> It has been identified that the right choice of a zeroth-order wavefunction, a wavefunction that can be used to form an inexpensive model for the environment, can lead to significantly accelerated convergence of the many-body expansion.<sup>13,27</sup> A good zeroth-order wavefunction is one that is able to incorporate the most important interactions present in the system in a computationally efficient manner.

Our choice for the zeroth order wavefunction is the Self-Consistent Field for Molecular

Interaction (SCFMI) wavefunction, which was initially introduced as an efficient representation for weakly interacting systems.<sup>39–43</sup> SCFMI utilizes localized molecular orbitals and enforces fragment sparsity in the molecular orbitals. In SCFMI, molecular orbitals are expanded in local subsets of atomic orbitals and are called Absolutely Localized Molecular Orbitals (ALMOs).<sup>39</sup> Like other localized molecular orbitals, ALMOs can be used in chemical interpretation<sup>44–47</sup> and formulation of reduced scaling methods using local correlation techniques.

Partitioning of AOs into subsets naturally arises for systems consisting of non-covalently bound entities like molecular clusters. ALMOs are orthogonal on fragment but non-orthogonal between fragments. ALMOs have been routinely used to perform interaction energy calculations which are basis set superposition error free,<sup>40</sup> employed to carry out computationally cheaper SCF alternative for weakly interacting systems,<sup>41</sup> and used to decompose interaction energy in energy decomposition analysis (i.e. ALMO-EDA) schemes.<sup>44–47</sup> Unlike SCF, SCFMI avoids diagonalization of the full system Fock matrix which reduces the prefactor for linear algebra in comparision to an SCF calculation.

Moving forward using eMBE with SCFMI, another computational bottleneck will appear for Density Functional Theory (DFT) calculations. The most accurate hybrid Exchange Correlation (XC) functionals involve some amount of exact exchange, which in turn becomes a bottleneck for large-scale calculations.<sup>48</sup> One way to circumvent this is to further approximate the environment beyond SCFMI. We have noticed that it is sufficient to capture only the important interactions between each of the computed terms and its surrounding in order to ensure expedited convergence of the many-body expansion. In this paper, we represent the 1-body and 2-body terms using hybrid density functionals with appropriate dispersion corrections because of their ability to accurately represent non-covalent interactions. The environment is represented using a semi-local functional due to its cheaper computational cost. The Embedded Mean-Field Theory (EMFT) developed by Manby and Miller is well-suited for embedding a hybrid density functional in a semi-local density functional.<sup>12</sup> EMFT is not

only simple to implement, but it has also been shown to reduce the cost of exact exchange computation while maintaining accuracy.<sup>12</sup> EMFT treats the system and its environment as open quantum systems incorporating quantum entanglement at a mean-field level. The combination of eMBE, SCFMI, and EMFT will be referred to as PolBE.

This paper is organized as follows: the theoretical background behind the MBE will be introduced. ALMOs are defined and the procedure to obtain them self-consistently, namely SCFMI, will be described. The EMFT method and models for computing exact exchange will be presented. The performance of PolBE for several chemical systems will be assessed in comparison with other existing eMBE methods. Finally, we will discuss the computational cost of a preliminary implementation of PolBE.

# Theory

## Embedded Many-body Expansion (eMBE)

The Many-Body Expansion (MBE) for a system of  $F$  fragments is written as a sum of 1-body ( $V_1$ ), 2-body ( $V_2$ ), ...,  $F$ -body ( $V_F$ ) terms as shown in Eq. (1).

$$E_{\text{total}} = \sum_{i=1}^F V_i \quad (1)$$

where

$$V_1 = \sum_{i=1}^F E_i(B), \quad (2)$$

$$V_2 = \sum_{i < j}^F (E_{ij}(B) - E_i(B) - E_j(B)), \quad (3)$$

$$\begin{aligned} V_3 = & \sum_{i < j < k}^F [(E_{ijk}(B) - E_i(B) - E_j(B) - E_k(B)) \\ & - (E_{ij}(B) - E_i(B) - E_j(B)) - (E_{ik}(B) - E_i(B) - E_k(B)) - (E_{jk}(B) - E_j(B) - E_k(B))] \end{aligned} \quad (4)$$

and so on,  $E_i(B)$  is the energy of the  $i$ -th monomer computed in the presence of the environment denoted as ‘B’,  $E_{ij}(B)$  is the energy of dimers ( $ij$ ) computed in the presence of the environment, and  $E_{ijk}(B)$  is the energy of the trimers ( $ijk$ ) computed in the presence of the environment, and so on.

Different eMBE methods differ only in their treatment of the environment. For example, in the Electrostatically-Embedded Many-Body Expansion, the 2-body term is  $E_{ij}(B)$  is calculated in the presence of atom-centered point charges for all monomers  $k \neq i, j$ . This embedding accelerates the convergence of the MBE, in comparison to vacuum MBE in which there is no environment. PolBE represents the environment using the polarized SCFMI

density. In this paper, unless otherwise stated, all the MBEs and the eMBEs have been truncated at the second order, applying the pairwise approximation.

## SCFMI model for the environment

Absolutely Localized Molecular Orbitals (ALMOs) on a fragment X ( $|\psi_{X_i}\rangle$ ) are linear combinations of atomic orbitals belonging to the same fragment ( $|\phi_{X_\mu}\rangle$ ) as shown in Eq. (5).

$$|\psi_{X_i}\rangle = \sum_{\mu} |\phi_{X_\mu}\rangle C_{\bullet X_i}^{X_\mu \bullet} \quad (5)$$

where  $\mathbf{C}$  is the MO coefficient matrix, in which all the MO coefficients  $C_{\bullet X_i}^{Y_\mu \bullet}$  when  $X \neq Y$  are constrained to be zero (i.e. fragment sparsity), giving the  $\mathbf{C}$  matrix a block diagonal structure.

Solving for the MOs with the ALMO constraint gives the SCFMI equations for each fragment X as shown in Eq.(6).

$$[\mathbf{F}]_{XX} [\mathbf{C}]_{XX} = [\mathbf{S}]_{XX} [\mathbf{C}]_{XX} [\epsilon]_{XX} \quad (6)$$

$$[\mathbf{F}]_{XX} = (\mathbf{I} - \mathbf{SP} + \mathbf{SP}^X) \mathbf{F} (\mathbf{I} - \mathbf{PS} + \mathbf{P}^X \mathbf{S}) \quad (7)$$

$$(\mathbf{P}^X)^{Z\nu, X\mu} = (\mathbf{T})_{\bullet Zl}^{Z\nu \bullet} (\boldsymbol{\sigma}^{-1})^{Zl, Xj} (\mathbf{T}^T)_{Xj \bullet}^{\bullet X\mu} \quad (8)$$

where  $[\mathbf{F}]_{XX}$  is the projected Fock matrix constructed as shown in (7),  $[\mathbf{C}]_{XX}$  is the ALMO coefficient matrix,  $[\mathbf{S}]_{XX}$  is the fragment AO overlap matrix, and  $[\epsilon]_{XX}$  contains the fragment orbital energies. The projected Fock matrix ( $[\mathbf{F}]_{XX}$ ) for each fragment is constructed using the density matrix ( $\mathbf{P}$ ), full system AO overlap ( $\mathbf{S}$ ) and, the Stoll projector  $\mathbf{P}^X$  as shown in Eq. (6).  $\mathbf{T}$  is the occupied orbital coefficient matrix and  $\boldsymbol{\sigma}$  is the occupied MO overlap matrix. We follow the projection operator form proposed by Stoll et al. shown in Eq. (8).<sup>39</sup> Diagonalization of the fragment Fock matrix,  $[\mathbf{F}]_{XX}$  of fragment AO dimension, yields MOs which are absolutely localized on that fragment.

The SCFMI wavefunction allows ALMOs on each fragment to polarize in the presence of ALMOs on all other fragments. In terms of physical interactions, the SCFMI wavefunction can capture electrostatics, exchange repulsion, dispersion (as permitted by the XC functional), and polarization between fragments as described by the mean-field theory used. One important interaction *not captured* by the SCFMI wavefunction is the effect of charge transfer among fragments. One way to capture this charge delocalization effect is through perturbation theory, which works decently well for weakly interacting systems.<sup>41</sup> Another more expensive but systematic method to capture the effects of charge delocalization is the Generalized X-Pol method which treats dimer states in the spirit of non-orthogonal configuration interaction theory.<sup>49</sup> In terms of computational cost, SCFMI removes the diagonalization bottleneck involved in the SCF iterations by requiring only fragment Fock matrix diagonalizations. However, there is no cost saving in constructing the Fock matrix for each SCF iteration.

## EMFT

The Embedded Mean-Field Theory (EMFT) is a simple embedding scheme for mean-field electronic structure theories, which partitions the system and the environment at the single-particle basis set level.<sup>12</sup> EMFT uses a more accurate and usually more expensive mean-field theory (referred to as ‘high level’) for the region of interest, while using an approximate and inexpensive mean-field theory (referred to as ‘low level’) for the rest of the system. The region of interest is usually one in which a chemical transformation takes place (denoted as A), and the surrounding is usually not directly involved in the chemical transformation (denoted as B). The EMFT energy is obtained by minimizing the EMFT energy function given by Eq. (9).

$$E_{\text{EMFT}} = E_{\text{low}}[\mathbf{P}] - E_{\text{low}}[\mathbf{P}^{AA}] + E_{\text{high}}[\mathbf{P}^{AA}] \quad (9)$$

where  $E_{\text{low}}[\mathbf{P}]$  is the low level energy computed using the full density matrix  $\mathbf{P}$ ,  $E_{\text{low}}[\mathbf{P}^{\text{AA}}]$  is the low level energy computed using the AA block of the density matrix, and  $E_{\text{high}}[\mathbf{P}^{\text{AA}}]$  is the high level energy computed using the AA block of the density matrix. The core idea is to compute interactions in subsystem A, represented by the AA block of the density matrix  $\mathbf{P}^{\text{AA}}$ , using the high-level mean-field theory. In the context of mean-field theories, hybrid density functionals may serve as the best choice for the high-level theory, while semi-local functionals would be suitable choices for low level. For this combination of hybrid density functional in semi-local density functional, the EMFT energy expression in Eq. (9) can be written as Eq. (10) where ‘high’ and ‘low’ denote the the hybrid density functional and semi-local density functional respectively, and  $\alpha$  is the fraction of exact exchange present in the hybrid density functional.

$$E_{\text{EMFT}} = E_{\text{xc,low}}[\mathbf{P}] + (E_{\text{xc,high}}[\mathbf{P}^{\text{AA}}] - E_{\text{xc,low}}[\mathbf{P}^{\text{AA}}]) + \alpha E_{\mathbf{K}}[\mathbf{P}] \quad (10)$$

$$\mathbf{F}_{\text{EMFT}} = \frac{\partial E_{\text{EMFT}}}{\partial \mathbf{P}} \quad (11)$$

$$= \mathbf{F}_{\text{xc,low}}[\mathbf{P}] + (\mathbf{F}_{\text{xc,high}}[\mathbf{P}^{\text{AA}}] - \mathbf{F}_{\text{xc,low}}[\mathbf{P}^{\text{AA}}]) + \alpha \mathbf{K}^{\text{AA}}[\mathbf{P}] \quad (12)$$

The EMFT Fock matrix expression follows directly from the energy expression (shown in Eq. (12)). The EMFT paper reports two models for computing exact exchange, EX0 and EX1 (shown in Eq. 13 and Eq. 14 respectively).

$$E_{\mathbf{K},\text{EX0}}[\mathbf{P}^{\text{AA}}] = -\frac{1}{4} \sum_{\mu,\nu,\sigma,\lambda \in A} \mathbf{P}_{\lambda,\mu}^{\text{AA}}(\mu\nu|\sigma\lambda) \mathbf{P}_{\nu\sigma}^{\text{AA}} \quad (13)$$

$$E_{\mathbf{K},\text{EX1}}[\mathbf{P}^{\text{AA}}] = -\frac{1}{4} \sum_{\substack{\nu,\sigma \\ \mu,\lambda \in A}} \mathbf{P}_{\lambda,\mu}^{\text{AA}}(\mu\nu|\sigma\lambda) \mathbf{P}_{\nu\sigma} \quad (14)$$

The EX0 requires the construction of only the AA block of the exact exchange matrix ( $\mathbf{K}$ ) from the AA block of the density matrix ( $\mathbf{P}^{\text{AA}}$ ), providing substantial savings in cost. However, this model of computing the exact exchange suffers from severe numerical instability due to unnatural density shift to the part of the system containing exact exchange (A), as

shown in Fig. S1. This is caused, in part, due to the lack of balancing counterpart terms in the semi-local functionals to the negative-definite exact exchange contribution of  $\alpha E_{\text{EX0}}[\mathbf{P}^{\text{AA}}]$  of hybrid DFT to the EMFT energy expression. In the original EMFT paper,<sup>12</sup> this is overcome by using the more expensive EX1 model for exact exchange which computes the exact exchange matrix using the full density matrix, instead of the just the AA block. Another technique to alleviate this problem, while still using the the EX0 model, comprises of doing the computations in AO basis orthogonal between A and B.<sup>50</sup> This approach in turn breaks the fragment sparsity in  $\mathbf{C}$  and is therefore not suitable for our purposes. In this work, we employ the EX0 model for computing exact exchange but constrain the particle number on each fragment by solving for the SCFMI equations instead of the full SCF equations. This fixes the numerical instability problem in EMFT. While the problem of unphysical collapse in the wavefunction is solved, particle number fluctuation (i.e. charge delocalization) is allowed at a two-body level.

## PolBE theory

As mentioned before, we have identified that the SCFMI wavefunction is well-suited to achieve the delicate balance between computational feasibility and incorporating many-body effects into lower order terms. We use EMFT to treat the system using an accurate, expensive mean-field theory while treating the environment using a more approximate and relatively inexpensive mean-field theory. Further computational advantages can be achieved by using a smaller basis set for the environment. For the rest of the paper, we will use XC1 to denote the high level functional and XC2 to denote the low level functional. Similarly, BS1 will refer to the larger basis set and BS2 to the smaller basis set.

1. The first step in PolBE is to calculate the polarized density of the full system at XC2/BS2. The MO coefficient matrix, illustrated in Fig. 1(a), is completely represented in BS2. This approximate polarized density is used for embedding all  $k$ -body terms treated using XC1/BS1.

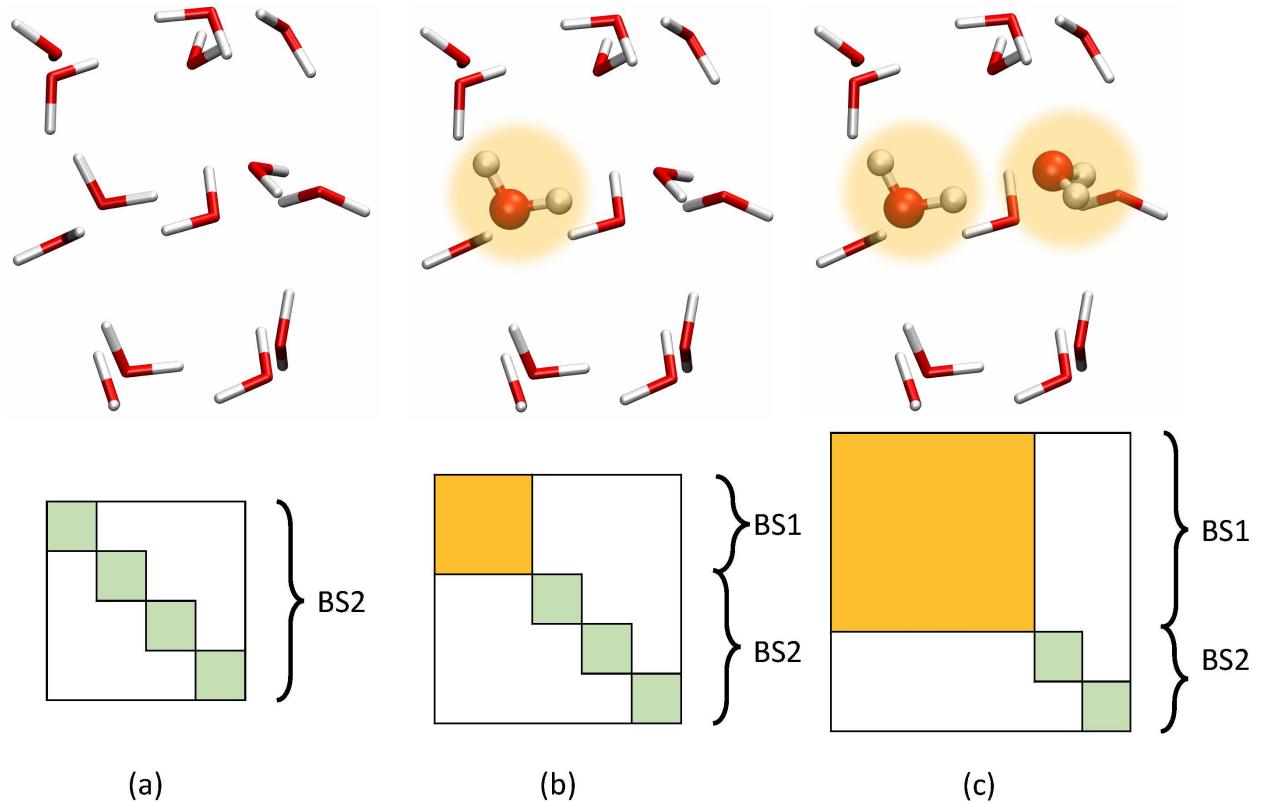


Figure 1: A schematic of the PolBE algorithm. The top panel schematically shows the density functional and basis set used for each of the monomers. Highlighted monomers are treated at XC1/BS1, and the others are treated at XC2/BS2. The bottom panel shows the structure of the MO coefficient matrix and the size of each of the blocks (a)  $V_0$ : All fragments are treated at XC2/BS2 (b)  $V_1$ : One of the fragments is treated at XC1/BS1 while the others are treated at XC2/BS2 (c)  $V_2$ : Two of the fragments are treated at XC1/BS1 while the others are treated at XC2/BS2

2. For a system consisting of  $F$  fragments, each 1-body term is calculated at XC1/BS1 in the presence of the approximate polarized density of the remaining  $(F - 1)$  fragments. Fig. 1(b) shows one of the  $F$  1-body terms in which the monomer is represented in BS1 while the rest of the environment is represented in BS2. In each monomer calculation, EMFT with EX0 is used to build Fock matrix treating the monomer at XC1 while the rest of environment is treated at XC2. Instead of diagonalizing the full Fock matrix, SCFMI is used to construct a projected Fock matrix only for the monomer fragment which is diagonalized, and its density is updated until convergence is reached. It should be emphasized that the density of the environment is not updated.
3. At a 2-body level, dimers are formed by combining all possible monomers into a larger fragment and treated at XC1/BS1 as described earlier. One of the  ${}^N C_2$  dimers is shown as a large fragment consisting of two monomers and is represented in BS1 as shown in Fig. 1(c).

## Computational Details

$\omega$ B97M-V, a combinatorially optimized range-separated hybrid meta-GGA functional with VV10 non-local correlation,<sup>51</sup> has been known to provide very good performance for a range of properties, including binding energies for non-covalent dimers and clusters.<sup>48</sup> We used this functional to represent the system denoted earlier as XC1. We used PBE, a more inexpensive and less accurate GGA functional, to represent the environment (XC2).<sup>52</sup> Def2-TZVPPD, a triple-zeta basis set with polarization and diffuse functions was used for BS1.<sup>53</sup> Def2-SV(P), a double-zeta split-valence basis set with polarization functions only on heavy atoms was used for BS2.<sup>54</sup> A quadrature grid consisting of 99 Euler-MacLaurin radial points and 590 Lebedev angular points was used for integrating both the XC functionals. SG-1 was used for the non-local VV10 part.<sup>55</sup> Binding energies reported below are vertical binding energies, computed as the difference between the energy of the full system and the sum of fragment energies at the super system geometry. The reference SCF binding energy was computed using XC1/BS1 ( $\omega$ B97M-V/def2-TZVPPD); the goal of PolBE is to accurately reproduce the full SCF energy of the supersystem. PolBE was implemented in the new SCF manager of Q-Chem<sup>56</sup> which was developed over Refs. 57–60.

## Results and Discussion

### Water Trimer PES

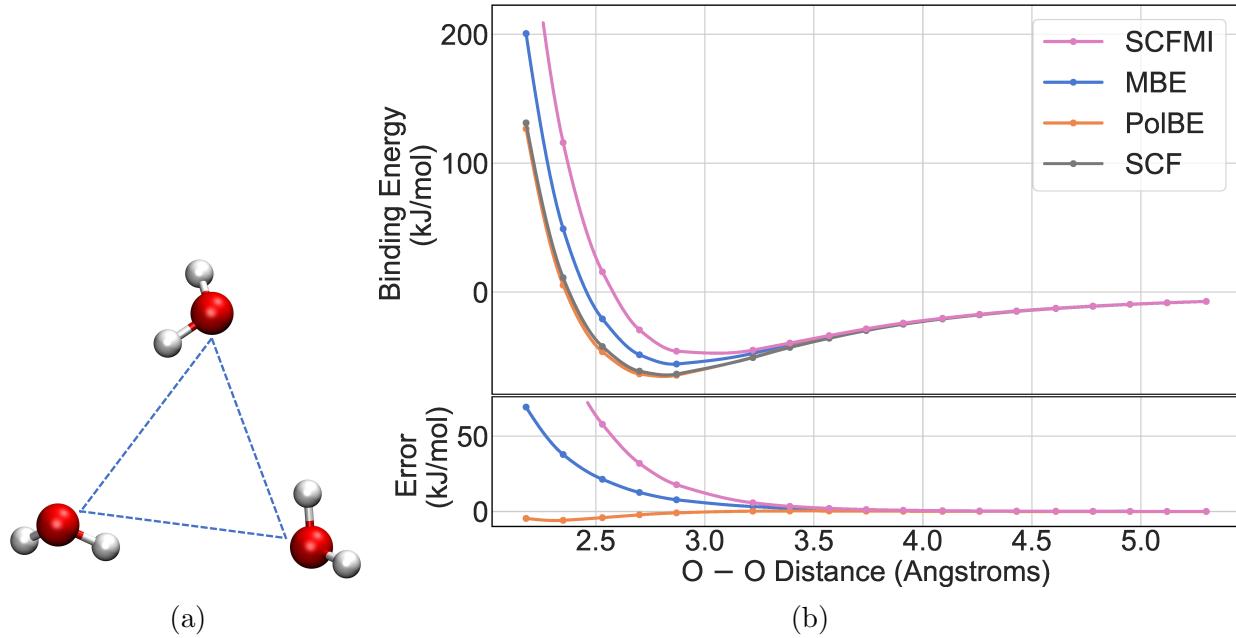


Figure 2: (a) Water trimer formed by stretching molecules away from each other (b) Water PES as computed by full SCF, vacuum MBE, and PolBE. Associated errors are shown in the bottom panel.

A desirable property of an embedded Many-Body expansion is that it should provide a smooth Potential Energy Surface (PES) and consequently a continuous nuclear gradient so that it can be used for performing geometry optimizations, transition state searches, and molecular dynamics simulations.<sup>61,62</sup> In order to evaluate the performance of PolBE as a function of nuclear position, we have investigated the PES of water trimer. PolBE relies on natural fragmentation arising from the lack of covalent bonding between different molecules in the cluster. There are no cut-off parameters which depend on real-space distances, and thus PolBE provides smooth potentials as shown in Fig. 2. PolBE is also trivially size-consistent as it recovers the non-interacting limit at large distances. In the interacting regime, PolBE binding energies are in good agreement with the SCF reference binding energies. PolBE also successfully recovers the repulsive part of the PES, while the EX0 model of

EMFT cannot as shown in Fig. S1.

This numerical instability does not appear in PolBE despite using the inexpensive EX0 model for exact exchange as the ALMO constraint prevents any charge transfer between fragments. However, it should be emphasized that the lack of charge transfer between fragments simply yields inadequate binding energies (as shown in the SCFMI curve in Fig. 2(b)). Therefore, allowing for pairwise charge transfer as is done in PolBE is very effective in recovering the missing charge transfer interaction energy. Non-additive dispersion interactions are also not captured by PolBE. These higher-order interactions are typically very small. In the case of water trimer at equilibrium geometry, higher order dispersion interaction would contribute an additional 0.06 kJ/mol to the PolBE binding energy.

## Water pentamer conformers

Table 1: Binding energy errors for the cage and cyclic conformers. Relative binding energy of the cage conformer with respect to the cyclic conformer as predicted by vacuum MBE truncated at 2-body (MBE-2), vacuum MBE truncated at 3-body (MBE-3), and PolBE truncated at 2-body (PolBE-2). The SCF binding energies for the cage and cyclic conformers are  $-153.7$  and  $-160.7$  kJ/mol. All values are in kJ/mol.

	E <sub>bind</sub> error		E <sub>bind</sub> (cage) – E <sub>bind</sub> (cyclic)
	cage	cyclic	
MBE-2	28.4	45.4	-9.9
MBE-3	1.1	5.9	2.3
PolBE-2	-7.2	-4.5	4.3

When a many-body expansion is truncated at the 2-body level, higher order binding cooperativity is not captured. The change in the binding affinity of molecule upon being bound to another water molecule is termed cooperativity. The phenomenon of cooperativity is found ubiquitously in chemical systems including water,<sup>63</sup> methanol,<sup>64</sup> and hydrogen fluoride.<sup>65</sup> The importance of understanding cooperativity cannot be overemphasized as it is crucial for understanding the bulk phase properties of these molecules. For instance, it enhances the strength of hydrogen bonding in water at a molecular level leading to anomalous adsorption properties in bulk water.<sup>66</sup> Cooperativity or non-pairwise additivity can be quantified as the

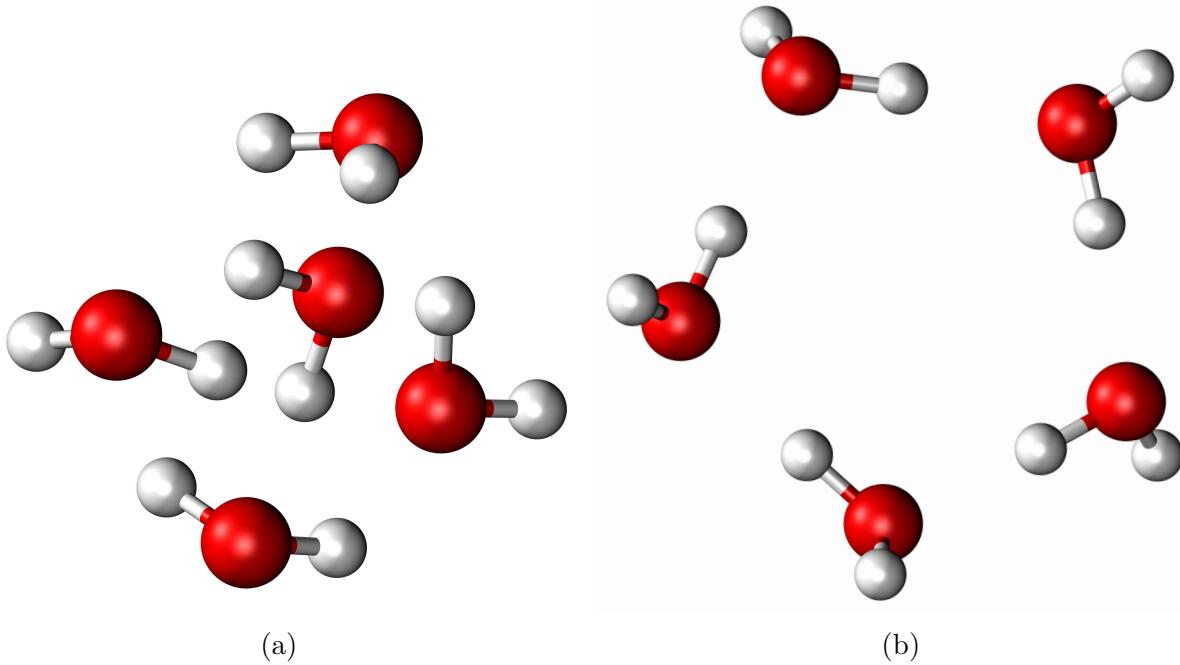


Figure 3: (a) Cage conformer of water pentamer (b) Cyclic conformer of water pentamer.

amount of interaction energy that is not captured by the many-body expansion truncated at the 2-body level. In this work, we show how the higher order effect of cooperativity may be captured at a 2-body level using the PolBE.

There has been considerable interest in predicting the relative energies of various molecular clusters, both qualitatively and quantitatively.<sup>67–69</sup> More often than not, hydrogen bonded molecules in clusters interact in a non-pairwise fashion,<sup>64,65</sup> making the total binding energy not a simple sum of pairwise interaction energies. It has been shown that homodromic hydrogen bonding ring-like networks are associated with the highest cooperativity.<sup>63</sup> This defining characteristic of these structures make them the most stable conformer in small water clusters, from trimer through pentamer, even though they contain only one hydrogen bond per water molecule while other conformers have more than one hydrogen bond per water molecule.

In this section, we investigate the case of two conformers of water pentamer: cyclic and cage using geometries taken from Ref. 70. The cyclic conformer, with the homodromic ring-like structure, exhibits more cooperativity in binding than the cage conformer. Vacuum MBE

truncated at the 2-body level makes errors of 28.4 and 45.4  $\text{kJmol}^{-1}$  for the cage and cyclic conformer respectively showing that that the cyclic conformer has more cooperativity than the cage variant. Vacuum MBE not only predicts quantitatively wrong binding energies, but also incorrectly predicts that the cage conformer is the more stable than the cyclic one. Vacuum MBE needs the 3-body term in order to capture this many-body effect, as shown in Table. 1. A full DFT calculation ( $\omega\text{B97M-V}/\text{def2-TZVPPD}$ ) on these conformers shows that the cyclic conformer is more stable than the cage by 7.0  $\text{kJ/mol}$ . PolBE which correctly embeds the 1-body and 2-body terms in an approximate representation of the environment makes much smaller binding energy errors of 7.2 and 4.5  $\text{kJ/mol}$ . PolBE is able to correctly pre-polarize the monomers, thus decomposing the interactions in a pairwise additive fashion. This demonstrates the ability of our 0<sup>th</sup> order wavefunction (SCFMI wavefunction) to precisely capture the mutual polarizations between the different water molecules which is clearly a many-body effect. PolBE also predicts the relative energetics correctly, predicting that the cyclic conformer is more stable with a relative energy error of 2.7  $\text{kJ/mol}$ , smaller than chemical accuracy of 1 kcal/mol. In the context of molecular simulation, it is more important to get the relative binding energy ordering and separations correctly rather than absolute binding energies. While vacuum MBE needs to be carried out till the third order to even get the relative ordering of conformers correct qualitatively, PolBE can achieve both the qualitative and quantitative relative energy ordering correct at the two-body level.

The water simulation community has recognized the importance of these higher-body effects and has tried to incorporate them either implicitly through inclusion of classical polarization<sup>71,72</sup> or explicitly by including 3-body and other higher order terms.<sup>73-77</sup> This parametrization is further complicated by the fact that different phases of water (clusters, liquid water, and ice) show different magnitudes of non-additivities.<sup>78,79</sup> PolBE does not contain any parametrization, and can approximate the ‘high level’ density functional with reasonable accuracy. PolBE also suggests an alternative way to account for these many-body effects, through the use of an inexpensive, albeit crude, zeroth order wavefunction.

## CO<sub>2</sub> clusters

There has been widespread interest in studying CO<sub>2</sub> molecular clusters because of their presence in earth's atmosphere and that of other planets.<sup>80</sup> There is a significant interest in the green chemistry community as supercritical CO<sub>2</sub> can be used as a benign solvent on an industrial scale.<sup>81</sup> Theoretical studies on CO<sub>2</sub> clusters have served as a bridge to understand its condensed phase properties.

CO<sub>2</sub> are classic examples of weakly-bound clusters. A study on the intermolecular potential of CO<sub>2</sub> has shown that dispersion is one of the leading terms governing the interaction.<sup>82</sup> We performed an ALMO-based energy decomposition analysis<sup>44,83,84</sup> of the binding energy of a single CO<sub>2</sub> molecule completely surrounded by other CO<sub>2</sub> molecules for all the clusters considered in this study. This decomposition, as summarized in Fig. 4(a), shows that polarization, which is the energy associated with the relaxation of molecular orbitals on each fragment in response to molecular orbitals on other fragments, contributes 38% of the interaction energy with charge transfer causing an additional stabilization of 27%. Nevertheless, as shown in Fig. 4(a), dispersion is the main mechanism which binds these CO<sub>2</sub> clusters, and the ability to describe it accurately is critical to understanding the molecular processes governing the chemistry of carbon dioxide. In fact, the semi-local functional PBE used for representing the environment, which does not contain any dispersion corrections, predicts that the CO<sub>2</sub> clusters are unbound. The VV10 non-local correlation functional is a sophisticated method to accurately capture the effect of dispersion. The high level functional used in this study,  $\omega$ B97M-V, contains the VV10 non-local correlation functional and correctly predicts the interaction energy of CO<sub>2</sub> trimer to be -14.98 kJ/mol (CCSD(T)<sup>85</sup> interaction energy is -15.96 kJ/mol). This shows that the correct treatment of dispersion is very much essential in predicting binding energies of such systems accurately.

We have investigated the ability of PolBE to predict interaction binding energies for a series of CO<sub>2</sub> clusters with geometries taken from Ref. 86. In Ref. 86, the geometries for the CO<sub>2</sub> clusters were optimized at M06/aug-cc-pVDZ. The binding energy per monomer

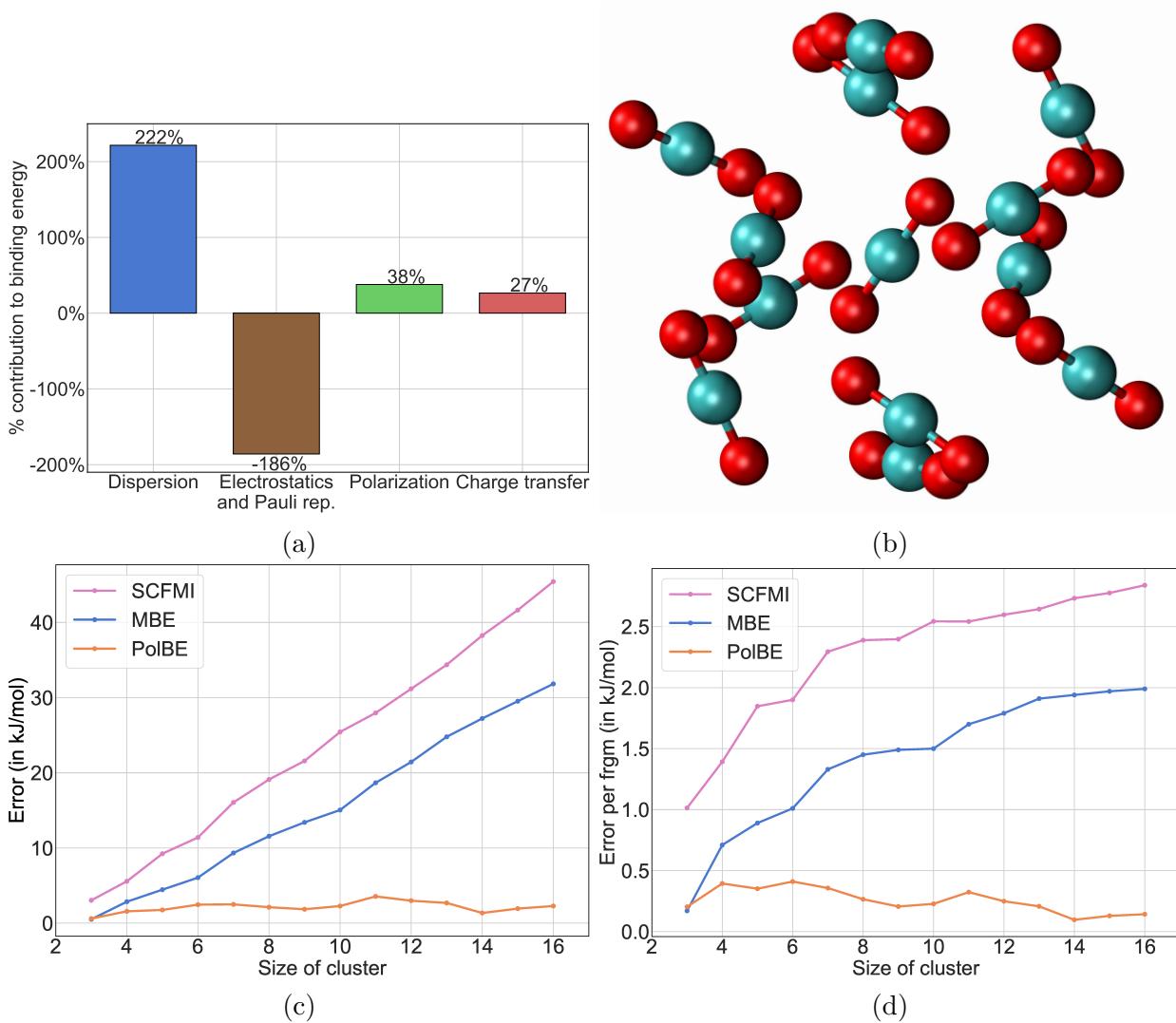


Figure 4: (a) Energy decomposition analysis for all CO<sub>2</sub> clusters considered in this study. Dispersion is the interaction between induced dipoles of different fragments. Pauli repulsion is the steric energetic penalty that occurs when two non-bonded atoms are forced to occupy the same space. Electrostatics is the energy associated with the coulombic interactions between fragments. Polarization is the energy associated with relaxation of molecular orbitals. Charge transfer is the energy associated with the flow of electrons between the fragments. (b) A CO<sub>2</sub> cluster containing 13 molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of CO<sub>2</sub> cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of CO<sub>2</sub> cluster size. For comparison, the reference binding energy for 16-mer is 225.3 kJ/mol.

increases with system size from  $-5.1$  kJ/mol for a trimer to  $-14.1$  kJ/mol for the 16-mer. The binding energy per monomer shows signs of saturation starting at the 13-mer, as it only increases by  $0.5$  kJ/mol from 13-mer to 16-mer. The saturation of binding energy per monomer at such a small size of cluster is a consequence of two effects: 1. the loose packing of  $\text{CO}_2$  molecules in the cluster and 2. Dispersion interactions which dominate the binding energy of  $\text{CO}_2$  are short-range effects. The errors in the predicted binding energies for SCFMI (XC1/BS1), vacuum MBE and PolBE have been plotted in Figure 4(c) as a function of cluster size. The binding energy error predicted by SCFMI and vacuum MBE truncated at the 2-body level increases continuously with system size. At all fragment sizes, SCFMI, vacuum MBE, and PolBE underestimate the binding energy but the magnitude of underestimation varies significantly. The binding energy error prediction of the SCFMI wavefunction is the largest, followed by that of vacuum MBE and then PolBE. For the 16-mer, the largest  $\text{CO}_2$  cluster considered in this study, the vacuum MBE error is  $31.83$  kJ/mol while the PolBE error is only  $2.28$  kJ/mol, smaller by more than an order of magnitude.

The error per monomer saturates for all the three methods as shown in Figure 4(d). For the largest clusters investigated in this study, the binding energy error per monomer saturates at around  $0.15$  kJ/mol for PolBE whereas it saturates at  $2$  kJ/mol for vacuum MBE and  $2.8$  kJ/mol for SCFMI. This saturation of error per monomer is very encouraging and its magnitude is expected to be transferable to large condensed phase systems of  $\text{CO}_2$ .

## Water clusters

Water clusters are of fundamental importance in biological systems and processes and in atmospheric chemistry.<sup>87</sup> Unlike bulk water, water clusters contain molecules at the surface with unsaturated hydrogen-bonds serving as models for liquid-vapor interfaces. Almost every possible experimental tool has been used to characterize the structure of water clusters in an attempt to understand its anomalous properties.<sup>88</sup> It is also possible to use highly accurate ab initio quantum chemical methods to study small water clusters, enabling direct comparison

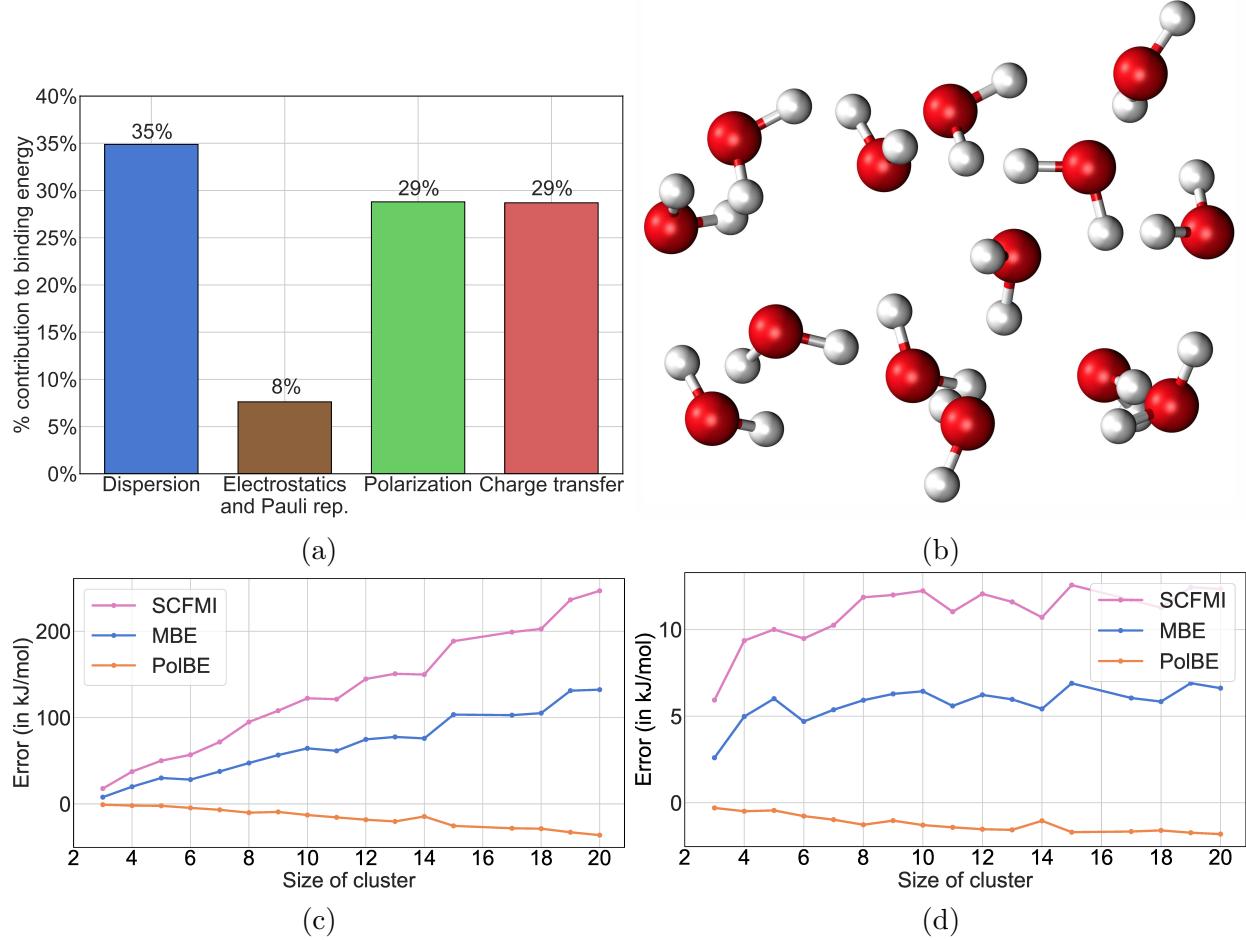


Figure 5: (a) Energy decomposition analysis for all water clusters considered in this study (b) A water cluster containing 14 water molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of water cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of water cluster size. For comparison, the total reference binding energy for the water 20-mer is 845.2 kJ/mol.

between experiment and theory.

Water molecules interact with each other through hydrogen bonding, participating both as donor and acceptor. Water molecules in both water clusters and bulk water interact with each other in a highly non-pairwise additive fashion.<sup>63</sup> Dispersion is an important mechanism contributing 35% to the total interaction energy of water clusters as shown in Fig. 5(a). The gas phase dipole moment of 1.855D of water is enhanced in the condensed phase by about 40%.<sup>89-92</sup> This enhancement in dipole moment coupled with the increasing polarizability of water clusters with size of the cluster<sup>93</sup> leads to stronger binding through magnified polarization contributing about 29% to the interaction energy. While different energy decomposition schemes differ on the actual magnitude of charge transfer in a hydrogen bond in water, all methods agree on that fact that charge transfer plays a significant role in hydrogen bonding placing its relative importance somewhere between 20% to 30%.<sup>44,94</sup> Performing an ALMO-based energy decomposition of the interaction energy for each term of the many-body expansion has shown that polarization and charge transfer play an important role in the 2-body terms, while polarization dominates the 3-body terms.<sup>95</sup> Charge transfer is mostly a pairwise additive effect and higher order terms are negligible.

In this section, we investigate the binding interactions of water clusters using geometries taken directly from a paper by Gadre and co-workers.<sup>96</sup> Binding in water clusters is much stronger than that in CO<sub>2</sub> clusters as can be seen from the larger interaction energies at every cluster size. The binding energy of the water trimer considered in this study has an interaction energy of 63.6 kJ/mol while the CO<sub>2</sub> trimer has a much smaller interaction energy of 15.4 kJ/mol. The presence of long-range interactions like dipole-dipole interactions slows down the saturation of binding energy per monomer as well. Even for the 20 water molecule cluster, the binding energy per water molecule is 42.3 kJ/mol and does not seem to be saturated. As a consequence, SCFMI, vacuum MBE, and PolBE errors per fragment also show slower signs of saturation. As in the case of CO<sub>2</sub> clusters, SCFMI and vacuum MBE systematically underbinds water clusters. The magnitude of the vacuum MBE errors, i.e. the

errors arising from truncating the many-body expansion at the 2-body level, is an indicator of the importance of 3-body and higher body effects in the binding of water clusters. For a water 20-mer, vacuum MBE underestimates the binding energy by 132 kJ/mol which is about 16 % of the total interaction energy. PolBE errors are much smaller, increasing from less than 1 kJ/mol for water trimer to  $-36.2$  kJ/mol for the water 20-mer. It is interesting to note that PolBE consistently overbinds water clusters while it underbinds  $\text{CO}_2$  clusters.

The error per fragment of vacuum MBE increases, though not monotonically, from the trimer value of 2.6 kJ/mol to 6.6 kJ/mol for the 20-mer. SCFMI errors per fragment are much larger, consistently upwards of 11 kJ/mol for large clusters. PolBE makes much smaller errors per fragment starting at 0.3 kJ/mol for trimer and increasing to 1.8 kJ/mol for the 20-mer. The underlying many-body effect in water clusters, the polarization term, is treated in a complete many-body sense in PolBE. Furthermore, charge transfer is predominantly pairwise additive in nature and is treated upto 2-body in PolBE. The success of PolBE can be attributed to treating these underlying physical interactions accurately and consistently. Both vacuum MBE and PolBE error curves show kinks as a function of cluster size (see Fig. 5(d)) due to the lack of a systematic structure in water clusters with increasing size of the cluster. Certain geometries of water clusters are less compact than others, leading to fewer hydrogen bonds and consequently a smaller interaction energy in comparison to its neighboring clusters. The kinks closely follow the discontinuities in the binding energy per fragment plot as a function of water cluster size. Unlike the case of  $\text{CO}_2$  where both the binding energy per fragment and the errors per fragment are saturated at the largest cluster size considered, these quantities continue to increase even at the 20-mer for water. Long-range electrostatics are, in part, responsible for this behavior and pose a considerable challenge to simulation of large systems of water. In order to investigate the saturation of SCF reference binding energy per fragment and PolBE binding energy error per fragment, we have investigated much larger water clusters  $n = 32$  and  $n = 64$ . The SCF binding energy for  $n = 32$  and  $n = 64$  is  $-26.36$  and  $-30.15$  kJ/mol, and are not saturated. The

PolBE binding energy error per fragment is  $-3.4$  and  $-3.9$  kJ/mol for  $n = 32$  and  $n = 64$  respectively. We see that the the PolBE binding energy error per fragment is saturating with cluster size, and hence showing that PolBE is size-extensive.

## Hydrated ion clusters

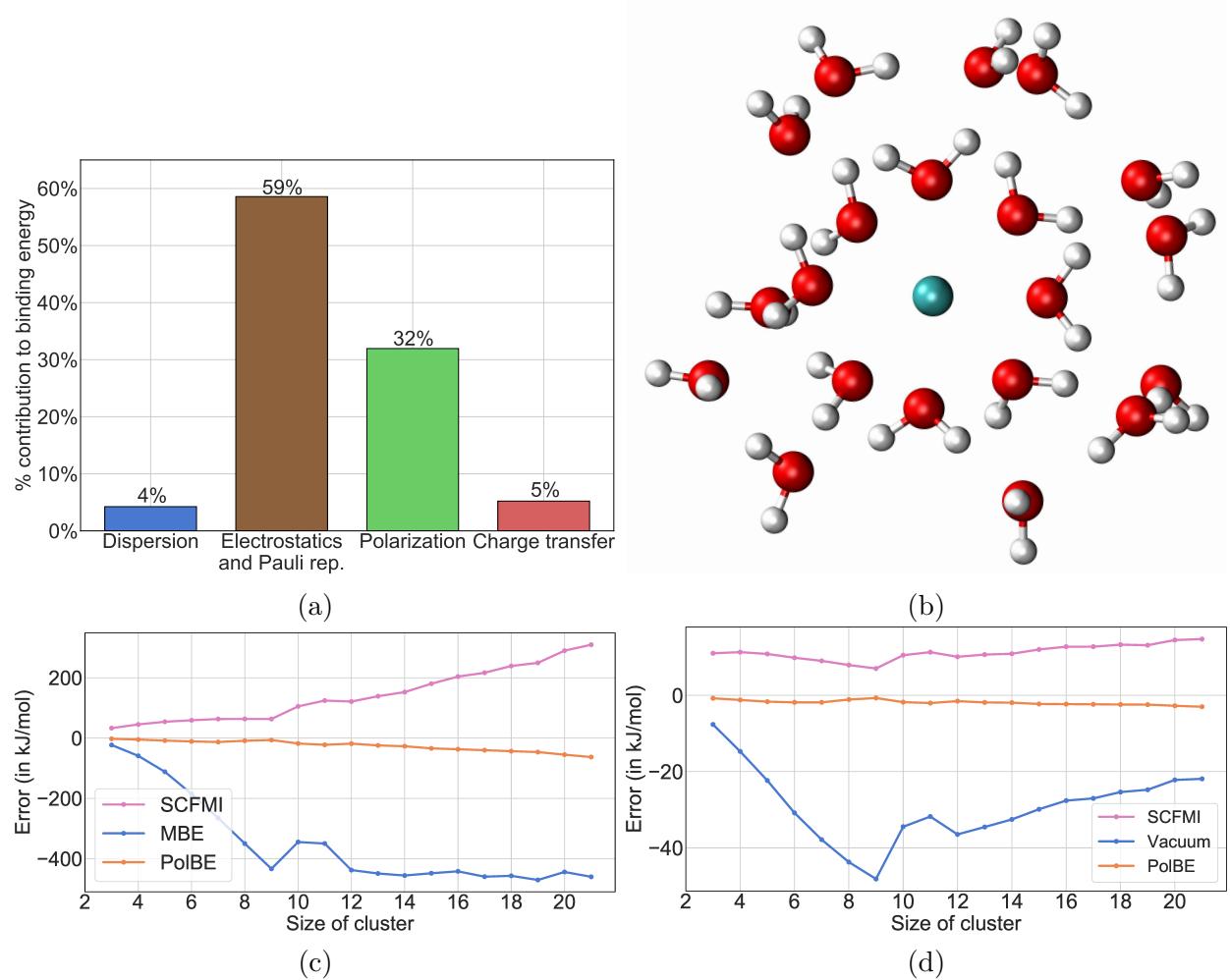


Figure 6: (a) Energy decomposition analysis for hydrated  $\text{Ca}^{2+}$  water clusters (b) Hydrated  $\text{Ca}^{2+}$  containing 20 water molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of cluster size. For comparison, the binding energy of  $\text{Ca}^{2+}(\text{H}_2\text{O})_{20}$  is 1935.6 kJ/mol.

Aqueous ions play a crucial role in many biological processes, and help sustain life on earth. While studying bulk aqueous ions has its own set of challenges, studying ion(s) in water clusters serve as a surrogate to understanding their properties. The small size of

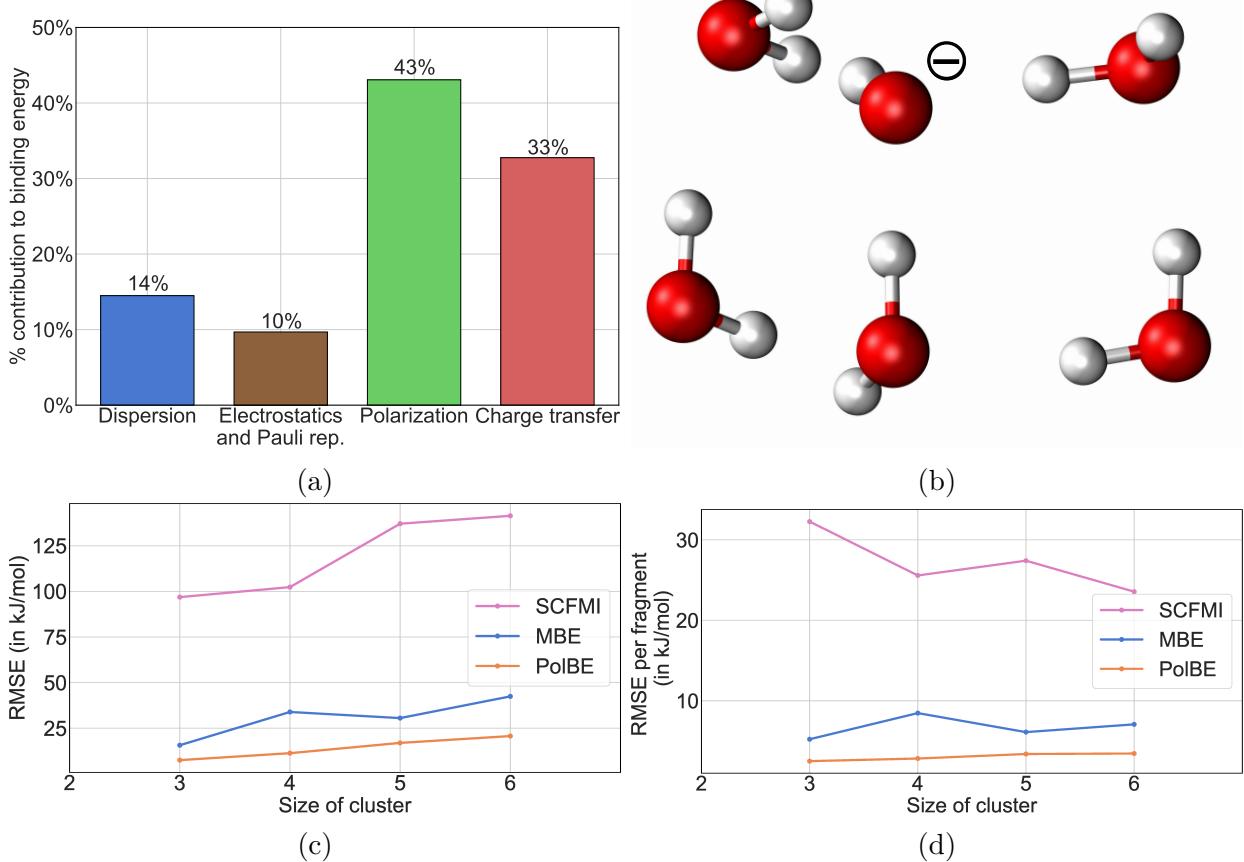


Figure 7: (a) Energy decomposition analysis for hydrated  $\text{OH}^-$  water clusters (b) Hydrated  $\text{OH}^-$  containing 5 water molecules (c) SCFMI, vacuum MBE, and PolBE errors as function of cluster size (d) SCFMI, vacuum MBE, and PolBE errors per fragment as function of cluster size. As multiple conformers of each cluster size were investigated, the error presented here are the root mean square errors of all conformers of a given size. For comparison, the reference RMS binding energy for  $\text{OH}^-(\text{H}_2\text{O})_5$  is 435.6 kJ/mol.

these clusters also enable the usage of accurate and expensive ab initio computational tools and direct comparison with experiment. These clusters also serve as models for studying hydration of ions and their coordination shell structure. In this section, we present the performance of PolBE and vacuum MBE for cationic water clusters ( $\text{Ca}^{2+}(\text{H}_2\text{O})_n$ ,  $n = 2-20$ ) and anionic water clusters ( $\text{OH}^-(\text{H}_2\text{O})_n$ ,  $n = 2-6$ ).  $\text{Ca}^{2+}(\text{H}_2\text{O})_n$  cluster geometries were taken from Ref. 97 and  $\text{OH}^-(\text{H}_2\text{O})_n$  geometries are taken from Ref. 98.

The interaction of ions with water is very strong as can be seen from the large binding energies of the hydrated ion clusters. The binding energy of  $\text{Ca}^{2+}(\text{H}_2\text{O})_{20}$  per fragment is about 92 kJ/mol and that of  $\text{OH}^-(\text{H}_2\text{O})_5$  is 72 kJ/mol. Decomposition the interaction energy between the  $\text{Ca}^{2+}$  ion and its surrounding water molecules shows that the interaction is mainly dominated by electrostatics and polarization. The bare ion carries a large charge density giving rise to large frozen and polarization terms. At a frozen density level, this monopole interacts with the surrounding water dipoles which makes permanent electrostatics strong. These effects are also more long-ranged than in pure water due to the slower decay of monopole-dipole interaction. The length scale of these effects makes this class of systems poorly suited for fragmentation methods, and hence an extreme challenge for PolBE. The large charge density also creates a strong field polarizing the surrounding water molecules very significantly, thus explaining the large polarization term.

In the case of hydrated  $\text{Ca}^{2+}$  ions, the SCFMI wavefunction underestimates the interaction energy as it is unable to capture all the effects of charge transfer. The magnitude of underestimation increases from 33 kJ/mol for  $\text{Ca}^{2+}(\text{H}_2\text{O})_2$  to 310 kJ/mol for  $\text{Ca}^{2+}(\text{H}_2\text{O})_{20}$ . On the other hand, vacuum MBE error increases from  $-22.9$  kJ/mol to  $-460.0$  kJ/mol for  $\text{Ca}^{2+}(\text{H}_2\text{O})_{20}$ . For the  $\text{Ca}^{2+}(\text{H}_2\text{O})_{20}$  cluster, this error is about 23% of the total interaction energy. Vacuum MBE truncated at the two-body level misses certain important higher order effects. Vacuum MBE also overestimates the amount of charge transfer at the 2-body level when the ion comprises one of the two fragments. Overestimating the interaction energies at a 2-body level causes net overbinding, while neglecting higher order terms in the MBE

causes underestimation of interaction energies. These two effects compete with each other to cause a temporary saturation in the total vacuum MBE error at around 450 kJ/mol in clusters of sizes 14 to 21. The saturation of total error with increasing cluster size causes the error per fragment to decrease after the first coordination shell of  $\text{Ca}^{2+}$  has been saturated. PolBE performs significantly better by making much smaller total errors, starting from 2.3 kJ/mol for  $\text{Ca}^{2+}(\text{H}_2\text{O})_2$  and increasing steadily to 63.0 kJ/mol for  $\text{Ca}^{2+}(\text{H}_2\text{O})_{20}$ . PolBE error, even in these strongly interacting systems, ranges from 0.5 % to around 3 % of the total interaction energies, roughly an order of magnitude smaller than vacuum MBE errors. The PolBE error per fragment is only 3.0 kJ/mol for  $\text{Ca}^{2+}(\text{H}_2\text{O})_{20}$  compared to 1.8 kJ/mol for the water 20-mer.

The first important difference between hydrated  $\text{Ca}^{2+}$  clusters and hydrated  $\text{OH}^-$  clusters is the charge density. The single negative charge of the  $\text{OH}^-$  anion has a much smaller charge density causing smaller total interaction energies. Comparing the clusters of the same size, hydrated hydroxide ion interaction energies are roughly half of the hydrated calcium ion interaction energies due to the reduced strength of the monopole. The second difference is the contribution of charge transfer component to the interaction energies. In hydrated calcium clusters, charge transfer contributes only 5% to the total interaction energy, while in hydrated hydroxide anion clusters it contributes 33%. This is expected because  $\text{Ca}^{2+}$  is not a good Lewis acid, but  $\text{OH}^-$  is a strong Lewis base and donates electrons to water readily.

As the contribution of charge transfer to interaction energy is significant, the SCFMI wavefunction severely underestimates the interaction energy. Both vacuum MBE and PolBE overestimate the binding energies for hydrated hydroxide anions, but the errors are shown as positive values as error metric used is root mean square error. Vacuum MBE binding energy error increases from 15.7 kJ/mol to 42.4 kJ/mol for  $\text{OH}^-(\text{H}_2\text{O})_5$ . PolBE errors are only about half the magnitude, overestimating the binding energy of  $\text{OH}^-(\text{H}_2\text{O})_5$  by 20.7 kJ/mol. Similar trends are seen in the binding energy errors per fragment.

## Error analysis of different environment models

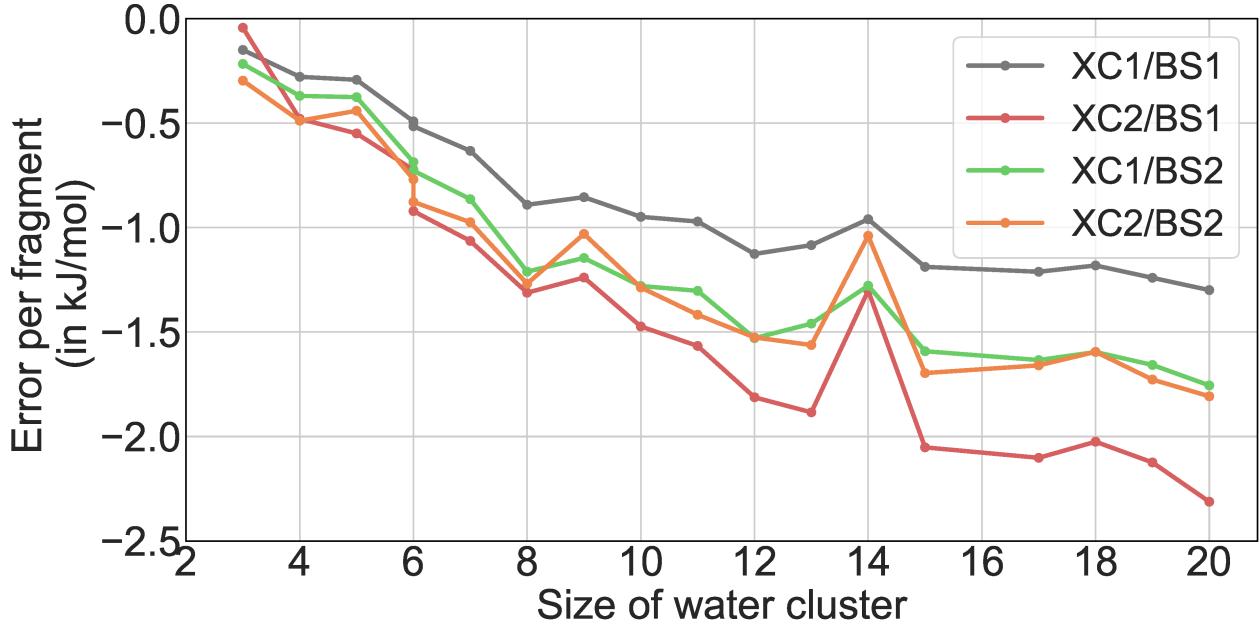


Figure 8: PolBE error per monomer for water clusters as function of size of the cluster with different representations of the environment. In this figure, XC1 is  $\omega$ B97M-V, XC2 is PBE, BS1 is def2-TZVPPD, and BS2 is def2-SV(P). The system is always treated at XC1/BS1.

The PolBE method uses EMFT to build the Fock matrix and solves the SCFMI equations to obtain ALMOs. EMFT relies on the fact that an approximate description of the environment (B) is sufficient to accurately capture the essential chemistry occurring in the system (A). PolBE relies on this assumption, providing an approximate embedding density for computing 1-body and 2-body terms accurately. In this section, we discuss the validity of this assumption. As a mean-field theory, PolBE attempts to represent the environment approximately by using a semi-local functional (XC2) in a much smaller basis set (BS2). The PBE functional, while much cheaper to compute than  $\omega$ B97M-V (XC1), is also grossly inadequate for computing the binding energies of  $\text{CO}_2$ , as it does not contain any dispersion and predicts that the  $\text{CO}_2$  clusters are unbound. The def2-SV(P) basis set used for the environment, denoted as BS2, is much smaller than the def2-TZVPPD used for the system. A single water molecule has 74 basis functions in BS1 and has only 18 basis functions in BS2. This makes the number of contributions of an environment water molecule to the Fock

matrix smaller by a factor of 4 in comparison to the number for a system water molecule.

In this section, we analyze the sources of error coming from the functional and basis set approximations used to represent the environment. In Fig. 8, different approximations of the environments are compared in order to isolate the source of error that comes from each of them. The best representation of the environment uses XC1/BS1 for both the system and the embedding environment. This treatment of the environment density on the same footing as the system does not use EMFT to build the Fock matrix, and is equivalent to SCFMI for each of the 1-body and 2-body terms. Therefore, it is the baseline error to compare against (shown as the gray curve in Fig. 8). When the basis set of the environment is changed to BS2 and the functional is fixed at XC1, an additional error of 0.5 kJ/mol per monomer is added at the water 20-mer. Similarly, changing the environment functional to XC2 while keeping the basis set fixed at BS1 causes the largest error, adding an error of 1 kJ/mol per water molecule to the baseline error. The method proposed in this paper, which represents the environment using XC2 in a small basis set BS2, adds saturating error of only 0.5 kJ/mol to the baseline error. This shows that an approximate representation of the environment is sufficient to accurately and efficiently capture the interaction at a 1-body and 2-body level, decomposing the binding energy in an almost pairwise additive fashion. A systematic analysis of error for different representations of the environment (different local/semi-local functionals and basis sets) is beyond the scope of this paper.

## Comparison with existing eMBE methods

Developing different eMBEs has been an active research topic of many research groups for the last couple of decades. These embedded MBE methods differ only in their description of the environment, ranging from using fixed atom-centered point charges,<sup>22</sup> using variationally-optimized atom-centered point charges,<sup>27</sup> to using actual monomer densities.<sup>36</sup>

Of these methods, the Electrostatically Embedded Many-Body Expansion (EE-MBE)<sup>22</sup> developed by Truhlar and co-workers, is the possibly the simplest and most widely used. In

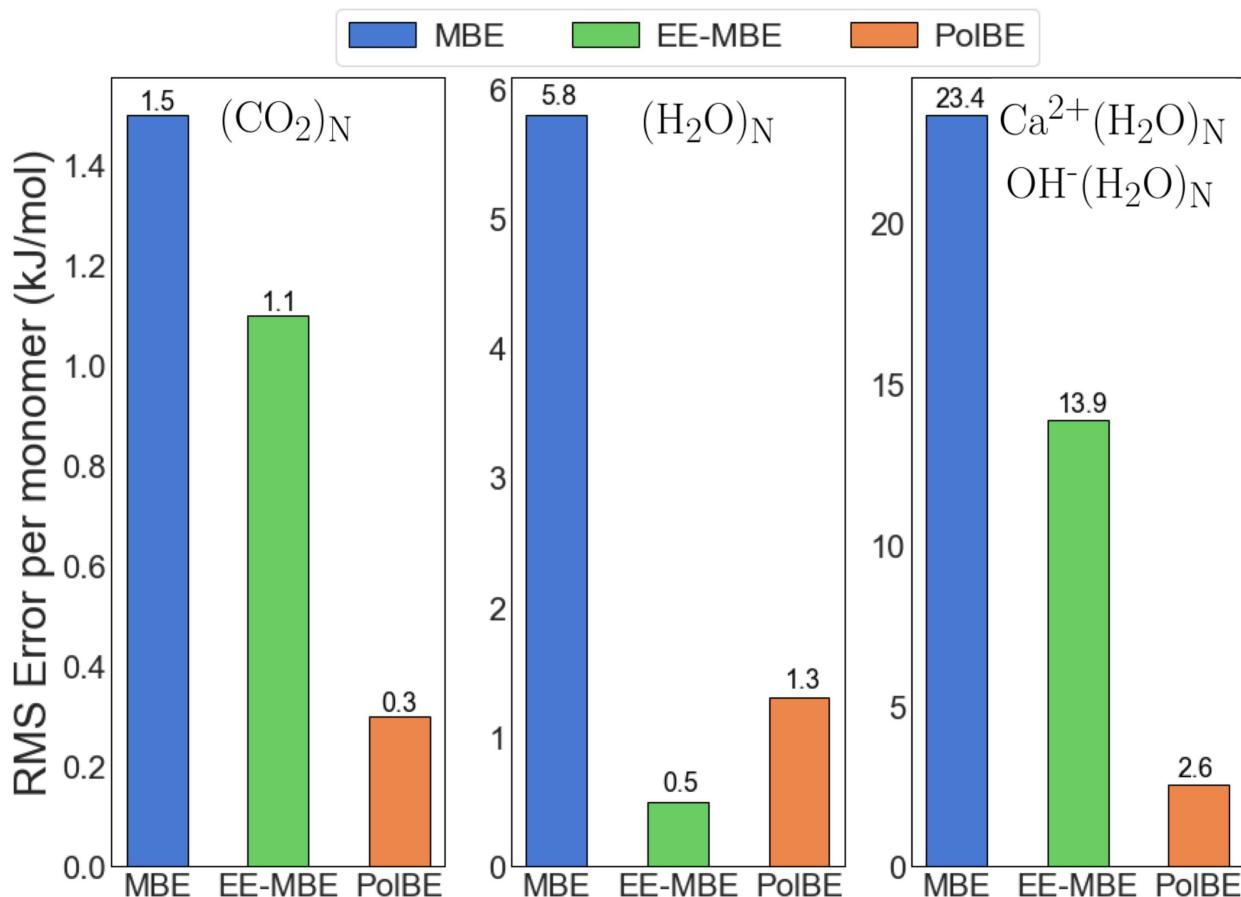


Figure 9: Root Mean Square Error per fragment for all carbon dioxide, water, and hydrated ion clusters. The errors for vacuum MBE, EE-MBE, and PolBE are shown. For  $(CO_2)_N$ , N ranges from 3 to 16. For  $(H_2O)_N$ , N ranges from 3 to 20. For  $Ca^{2+}(H_2O)_N$ , N ranges from 2 to 20 and for  $OH^-(H_2O)_N$ , N ranges from 2 to 5.

this method, each  $k$ -body term is computed in the presence of atom-centered point charges for the rest of system. While an approximate representation of the environment can be sufficient to attain useful accuracy, EE-MBE oversimplifies it, which causes important interactions like exchange repulsion, coulombic interaction of electron clouds, and polarization to be omitted, consequently leading to poorer performance. EE-MBE is also not parameter-free as the binding energies depend on the charges as shown by Herbert and co-workers previously.<sup>25</sup> Including higher-order terms in EE-MBE can lead to significant precision problems which would require tightening of SCF convergence and integral screening thresholds making them more expensive.<sup>99</sup> Fig. 9 shows the errors of EE-MBE performed using CHELPG charges, which are atom-centered charges fitted to reproduce the molecular electrostatic potential at a number of points around the molecule.<sup>100</sup> The CHELPG charges were determined for a water monomer at its equilibrium geometry, and the same set of charges were used for representing all environment water molecules. In comparison to EE-MBE, PolBE errors are much smaller in the case of  $\text{CO}_2$  clusters and hydrated ion clusters. For all the conformers of hydrated ion clusters studied in this work, the RMS error per fragment of PolBE is only 2.6 kJ/mol whereas that of EE-MBE is 13.9 kJ/mol as seen in Fig. 9. For the water clusters investigated, EE-MBE errors are smaller than PolBE errors for the water clusters investigated. However, the performance of EE-MBE is heavily dependent on the embedding charges used as shown in Table S8. For instance, EE-MBE performed using Mulliken charges increases the error per fragment for water molecules to 2.8 kJ/mol. The errors of EE-MBE also depend on the functional and basis set used to run these calculations. The performance of PolBE is robust across the range of systems tested in this work and is also parameter-free. PolBE consistently matches the SCF binding energies computed using XC1/BS1 with modest errors irrespective of the choice of XC1 and BS1. Different choices for representation of the environment, viz. different XC2 and BS2, does not affect the predicted binding energies too significantly.

Table 2: A table showing the computational cost scaling with system size for construction of Fock matrix and diagonalization steps in SCF and PolBE

	Fock build		Diagonalization	
	SCF	PolBE	SCF	PolBE
2-body screening	$\mathcal{O}(F^2)$	$\mathcal{O}(F^2) + \mathcal{O}(F^3) + \mathcal{O}(F^3)$	$\mathcal{O}(F^3)$	$\mathcal{O}(F) + \mathcal{O}(F) + \mathcal{O}(F)$
No 2-body screening	$\mathcal{O}(F^2)$	$\mathcal{O}(F^2) + \mathcal{O}(F^3) + \mathcal{O}(F^4)$	$\mathcal{O}(F^3)$	$\mathcal{O}(F) + \mathcal{O}(F) + \mathcal{O}(F^2)$

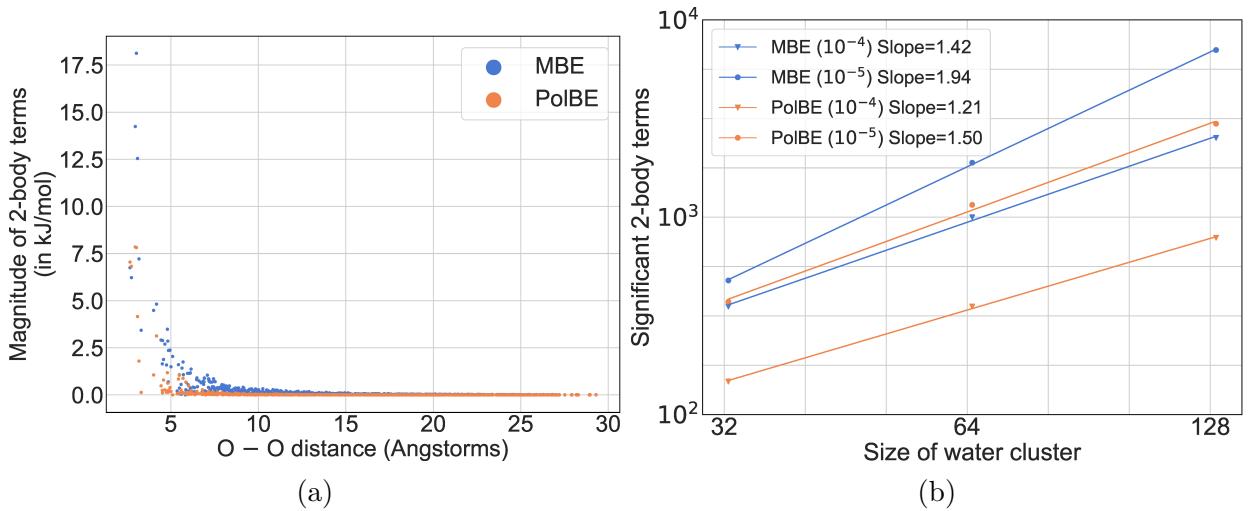


Figure 10: (a) A plot of the magnitude of 2-body terms in MBE and PolBE as a function of the distance between the fragments in a water cluster containing 512 monomers (b) A log-log plot of the number of significant 2-body terms for MBE and PolBE as a function of size of the water cluster. The number of significant terms are indicated determined as the number of terms exceeding the threshold as indicated in parenthesis

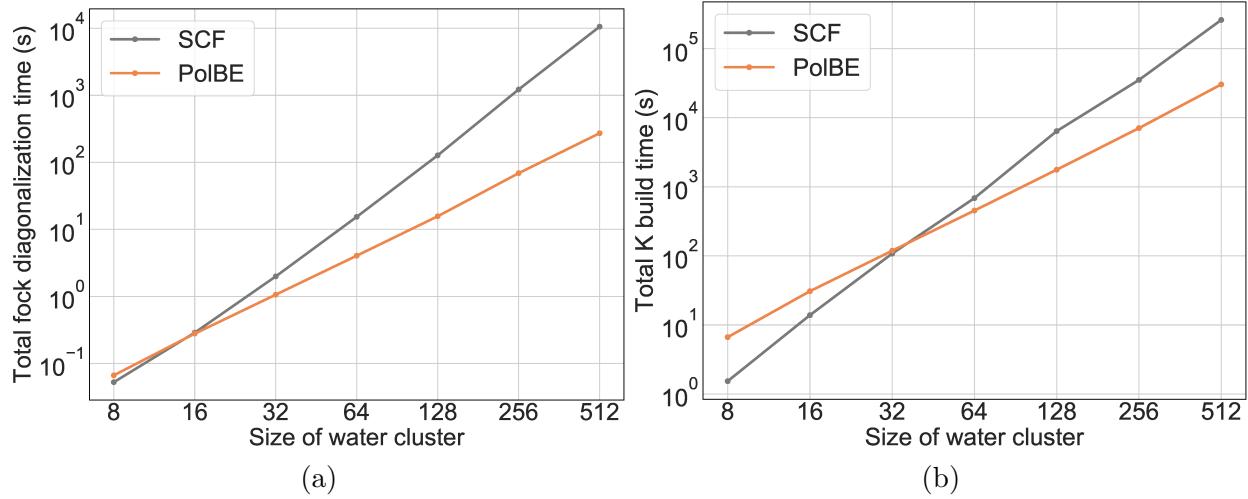


Figure 11: (a) Wall clock time taken for Fock matrix diagonalization(s) in SCF and PolBE as a function of cluster size (b)Wall clock time taken for construction of exact exchange matrix(ces) for SCF and PolBE as a function of cluster size.  $F$  is the number of monomers in the system which indicates system size. Timings reported are the wall clock timings for one iteration of SCF procedure and the wall clock total time for one iteration for all terms in PolBE.

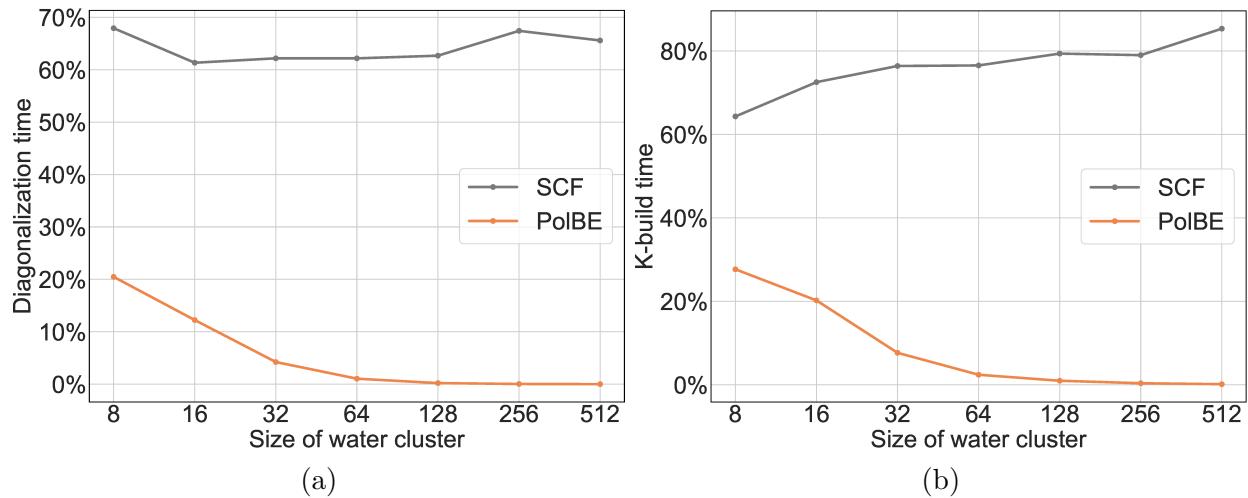


Figure 12: (a) Wall clock time taken for diagonalization of Fock matrix(ces) as a percentage of total time taken for all linear algebra manipulations for SCF and PolBE as a function of cluster size (b) Wall clock time taken for construction of the exact exchange bit as a percentage of total time taken for construction of the Fock matrix(ces) for SCF and PolBE as a function of cluster size. Timings reported are the wall clock timings for one iteration of SCF procedure and the wall clock total time for one iteration for all terms in PolBE.

## Computational timing analysis

The computational expense in the SCF and PolBE procedures can be classified into two categories: diagonalization and Fock build. In this section, we discuss the computational cost of SCF and PolBE in terms of these two categories. All timings reported in this section were run on a 2.3 GHz 16-core AMD Opteron<sup>TM</sup> processor. Timings reported are the wall clock timings for one iteration of SCF and the wall clock total time for one iteration for all terms in PolBE. B3LYP<sup>101,102</sup> and PBE<sup>52</sup> were used for XC1 and XC2, and def2-SVPD<sup>53</sup> and STO-3G<sup>103</sup> were used for BS1 and BS2 in PolBE. For the SCF calculation, the reported timings are for running B3LYP/def2-SVPD. Different density functional and smaller basis sets were used for this part of the timings analysis in order to access larger system sizes. Spherical water clusters used in this study were obtained from Ref. 104.

The number of terms at each order increases combinatorially with increasing order of the MBE. While the number of terms in theory increases combinatorially in PolBE as well, not all of them are numerically significant. This effect can be attributed to embedding, which incorporates physical effects like electrostatics and many-body polarization in the 0-body term as captured at XC2/BS2. These effects are further corrected using a the high-level functional XC1 in a large basis set BS1 at a 1-body level. After incorporating 1-body effects, the only physical interactions that are not yet captured are the corrections to the many-body polarization at XC1/BS1 and the effect of charge transfer. The 2-body terms capture only the effect of 2-body charge transfer and 2-body correction to the polarization, both of which decay rapidly with increasing distance between the pairs. Fig. 10(a) shows rapid decrease in the magnitude of 2-body terms with distance between the pair. It also shows that the magnitude of 2-body terms in PolBE are much smaller and decay much faster than 2-body vacuum MBE terms. This results in the linear increase in the number of 2-body terms in contrast to the quadratic number of all 2-body terms. Employing a conservative distance-based cutoff will limit the number of 2-body computations to scale linearly with system size. Considering cluster sizes of 32, 64, and 128 water molecules with a threshold

of  $10^{-5}$  Hartrees, Fig. 10(b) shows that the number of significant 2-body terms scales sub-quadratically with system size. Using looser thresholds and/or larger system sizes will show a scaling that is much closer to unity.

Consider a system with  $F$  fragments. Let  $N_1$  and  $N_2$  represent the total number of basis functions in BS1 and BS2 respectively. Let  $n_1$  and  $n_2$  represent the numbers of basis functions per monomer in BS1 and BS2 respectively. The cost of Fock matrix construction for SCF scales as  $\mathcal{O}(N_1^2)$  leading to  $\mathcal{O}(F^2)$  scaling with system size.<sup>105</sup> PolBE involves construction of three types of Fock matrices: (a) One Fock matrix at XC2/BS2 for  $V_0$  (b) A linear number of Fock matrices for the  $V_1$  term (c) A linear or quadratic number of Fock matrices for the  $V_2$  term depending on whether screening is employed or not employed. For each 1-body term, the size of the XC2 Fock matrix is roughly  $N_2$ .  $V_1$  Fock matrices would also involve the computation of two additional Fock contributions from XC2 and XC1, but their size is limited to  $n_1$  and does not scale with system size. The total cost of computation of the  $V_1$  terms scales as  $F \times \mathcal{O}(N_2^2)$ . Each Fock matrix for the  $V_2$  term also exhibits similar computational costs. Summing up, the cost of constructing all the Fock matrices scales as  $\mathcal{O}(F^2) + \mathcal{O}(F^3) + \mathcal{O}(F^3)$  after employing 2-body screening.

The density update step, which involves diagonalization of the effective hamiltonian matrix, scales as  $\mathcal{O}(M^3)$ , where  $M$  is the size of the system. For large system sizes, diagonalization of the Fock matrix becomes the most expensive step computationally. Diagonalization accounts for the majority of the computational cost of all linear algebra manipulations, constituting more than 60% of the cost across all fragment sizes as shown in Fig. 12(a). PolBE, which is based on the SCFMI wavefunction, removes this diagonalization bottleneck for large systems as shown in Fig. 12(a). As the cost of diagonalization scales as the third power of the dimension of the matrix, density update step in SCF scales as  $\mathcal{O}(N_1^3)$ . In PolBE at the  $V_0$  level, the total cost of diagonalization scales as  $F \times \mathcal{O}(n_2^2)$ . As  $n_2$  is constant, the cost of diagonalization for  $V_0$  term scales as  $\mathcal{O}(F)$ . For each term in  $V_1$  and  $V_2$ , only the MOs of the monomer and dimers are updated while the MOs of the environment are fixed. This

requires diagonalization of projected Fock matrices with dimension  $n_1$  and  $2n_1$  for the  $V_1$  and  $V_2$  terms respectively. In conclusion, the diagonalization step scales as  $\mathcal{O}(F) + \mathcal{O}(F) + \mathcal{O}(F)$  with system size. The total cost of diagonalization in SCF quickly exceeds that of PolBE for cluster sizes of 16 water molecules (624 basis functions) as shown in Fig. 11.

Fock matrix construction, comprises coulomb, XC and exact exchange contributions, and scales as the 2<sup>nd</sup> power of its size. Exact exchange dominates the computational cost as shown in Fig. 12(b). Hybrid density functionals, known to alleviate the self-interaction energy problem, provide a more reliable and accurate prediction of system properties, including non-covalent interaction energies.<sup>48</sup> However, addition of exact exchange relative to pure density functionals makes them more expensive to compute.<sup>106–108</sup> Different methods, ranging from screening based on locality to density-fitting have been proposed for reducing the computational cost of forming the Fock matrix.<sup>108</sup> However, computation of exact exchange has remained the bottleneck step. As shown in Fig. 12(b), construction of the exact exchange part of the Fock matrix takes up upwards of 60% of the total Fock construction time. PolBE, which embeds hybrid DFT in semi-local DFT, requires construction of only the AA block of the exact exchange matrix which is much smaller than the entire Fock matrix. The size of the AA block is only  $n_1$  for each 1-body term and  $2n_1$  for each 2-body term. The total SCF K-build time exceed the total K-build time for all the  ${}^F C_1 + {}^F C_2$  terms in PolBE at 32 water molecules (1248 basis functions) as shown in Fig. 11(b). PolBE thus accelerates the computational bottleneck step in both linear algebra manipulations and Fock construction. In terms of the total computational cost, PolBE costs more than a full SCF calculation owing to the large number of Fock constructions at a 2-body level. However, all of each of these computations are trivially parallel as described below.

A practical challenge in most DFT software packages is to be able to take full advantage of the multiple cores available in modern day high performance computing architectures.<sup>109</sup> PolBE provides a very simple way to take advantage of this. Once the embedding density has been formed, PolBE involves computation of  ${}^F C_1 + {}^F C_2$  terms, all of them completely

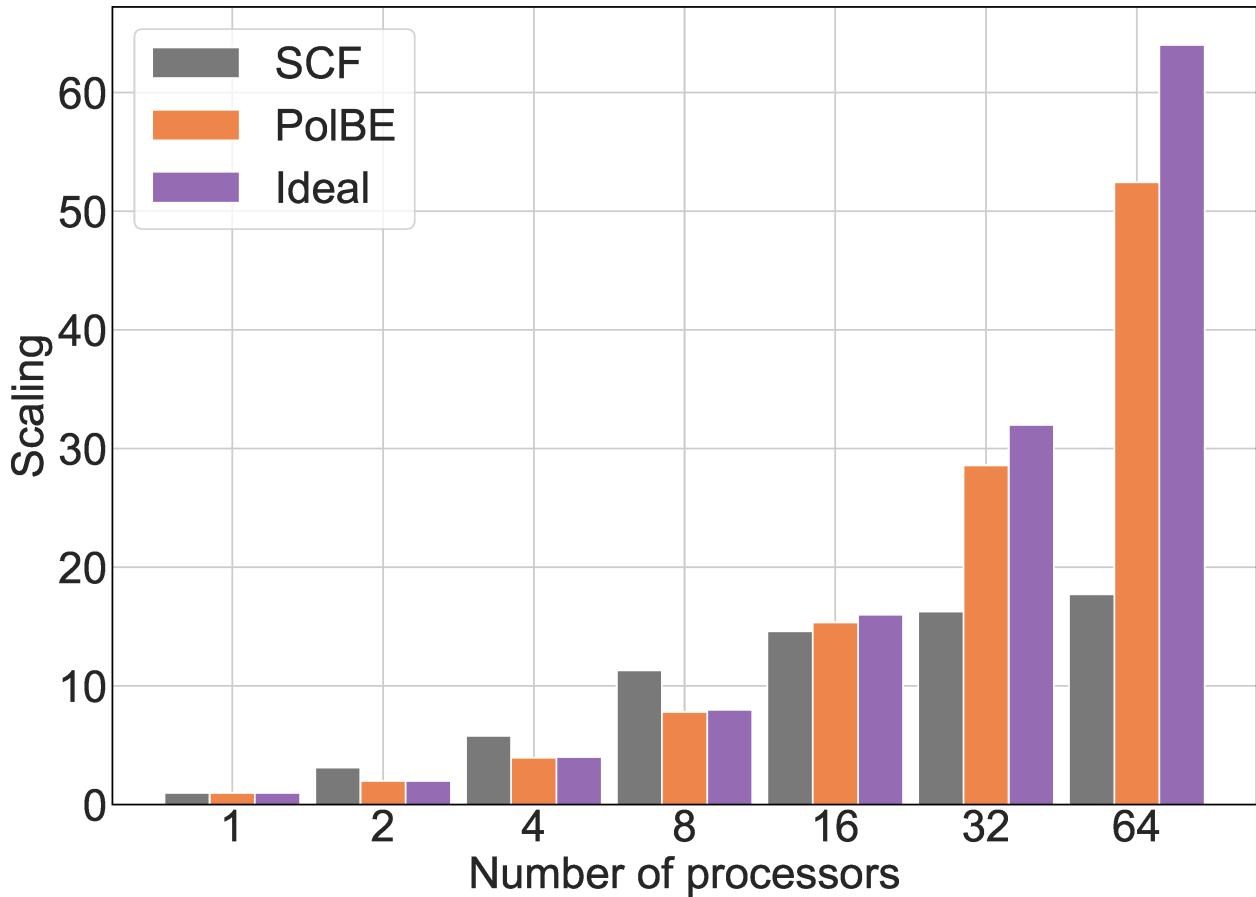


Figure 13: Scaling of SCF and PolBE with increasing number of processors for water 20-mer. Computations were performed using  $\omega$ B97M-V/def2-TZVPPD in PBE/def2-SV(P) for PolBE and  $\omega$ B97M-V/def2-TZVPPD for SCF.

independent of each other. This property of PolBE, inherited from the many-body expansion, makes it trivial to distribute these jobs among any number of given processors making PolBE “embarrassingly parallel.” As shown in Fig. 13, PolBE for water clusters exhibits nearly ideal parallel scaling. In principle, one can achieve even better scaling with PolBE by conceiving smarter master-slave process which are aware of the approximate time it would take to run each of the jobs *a priori*.

## Conclusions and Outlook

In this work, we investigated the importance of accurately representing the environment in order to accelerate the convergence of the many-body expansion. In particular, we found that capturing the effects of the environment using EMFT while partitioning the fragments using SCFMI accurately captures the many-body cooperativity effects commonly found in condensed phase systems. We have shown that the PolBE method proposed in this paper reproduces binding energies of a wide range of molecular clusters of various sizes and consisting of different interaction mechanisms. PolBE consistently performs better than other popular embedded many-body expansion methods EE-MBE. The computational cost of PolBE and its ability to remove the computational bottleneck steps was analyzed and demonstrated.

Comparison of the performance of PolBE to other embedded many-body expansion methods<sup>26</sup> would be of significant interest. The implementation of PolBE reported here can potentially be improved using various strategies. Larger systems can be made computationally more tractable by using a real-space cutoff for computing two-body terms. These cut-offs must depend on the dominant long-range interactions present in the system and the accuracy desired. Such cut-offs have not been investigated in this paper and are an interesting topic for future study. As PolBE is an MBE based method, its accuracy can be systematically improved by including higher order terms in the expansion. While the number of higher order terms increases with each order, some screening of important terms, either based on real-space distances or on the magnitude of lower terms, can be employed. The description of the environment can be made coarser to achieve additional speedups. PolBE, as suggested in this work, uses different basis sets and functionals for the system and environment. The extension of PolBE to use different DFT quadrature grids for the system and the environment is straightforward. Similar ideas can be extended to use different density fitting basis sets for the coulombic interaction term. The ALMO-based representation of the system used in PolBE yields fragment-specific occupied and virtual orbitals. These fragment-specific virtuals, which are both fragment-sparse and smaller in number in comparison to the virtual

space of the whole system, provides a compact representation of the fragment virtual space which can be exploited in the extension of PolBE to double hybrids density functionals<sup>110</sup> and to correlated wavefunction methods like MP2, orbital-optimized MP2,<sup>57–60,111,112</sup> and coupled cluster theories.<sup>113</sup> Extension of these ideas to correlated wavefunction methods is currently being investigated.

## Supplementary Material

See supplementary material for EDA analysis, EMFT EX0 numerical instability, numerical values of binding energy, binding energy errors, charges used for EE-MBE, and details of the computational timings analysis.

## Acknowledgements

This work was supported by the National Science Foundation (NSF) under award CHE-1665315. The authors declare the following competing financial interest(s): M. H.-G. is a part owner of Q-Chem, Inc.

## References

- (1) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schütz, M.; Chan, G. K.-L. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* **2014**, *345*, 640–643.
- (2) Beran, G. J. Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567–5613.
- (3) Boese, A. D.; Sauer, J. Embedded and DFT Calculations on the Crystal Structures of Small Alkanes, Notably Propane. *Cryst. Growth Des.* **2017**, *17*, 1636–1646.
- (4) Demerdash, O.; Yap, E.-H.; Head-Gordon, T. Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu. Rev. Phys. Chem.* **2014**, *65*, 149–174.
- (5) Zgid, D.; Chan, G. K.-L. Dynamical mean-field theory from a quantum chemical perspective. *J. Chem. Phys.* **2011**, *134*, 094115.
- (6) Lan, T. N.; Zgid, D. Generalized Self-Energy Embedding Theory. *J. Phys. Chem. Lett.* **2017**, *8*, 2200–2205.
- (7) Ishida, H. Surface-embedded Green-function method: A formulation using a linearized-augmented-plane-wave basis set. *Phys. Rev. B* **2001**, *63*, 165409.
- (8) Chibani, W.; Ren, X.; Scheffler, M.; Rinke, P. Self-consistent Green's function embedding for advanced electronic structure methods based on a dynamical mean-field concept. *Phys. Rev. B* **2016**, *93*, 165106.
- (9) Elliott, P.; Burke, K.; Cohen, M. H.; Wasserman, A. Partition density-functional theory. *Phys. Rev. A* **2010**, *82*, 24501.
- (10) Huang, C. Extending the density functional embedding theory to finite temperature and an efficient iterative method for solving for embedding potentials. *J. Chem. Phys.* **2016**, *144*, 124106.

- (11) Huang, C.; Carter, E. A. Potential-functional embedding theory for molecules and materials. *J. Chem. Phys.* **2011**, *135*, 194104.
- (12) Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller, T. F. Embedded mean-field theory. *J. Chem. Theory Comput.* **2015**, *11*, 568–580.
- (13) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.
- (14) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.
- (15) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891–5928.
- (16) Knizia, G.; Chan, G. K.-L. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.* **2012**, *109*, 186404.
- (17) Wouters, S.; Jiménez-Hoyos, C. A.; Sun, Q.; Chan, G. K.-L. A Practical Guide to Density Matrix Embedding Theory in Quantum Chemistry. *J. Chem. Theory Comput.* **2016**, *12*, 2706–2719.
- (18) Prodan, E.; Kohn, W. Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11635–11638.
- (19) Vreven, T.; Morokuma, K. In *Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM*; Spellmeyer, D. C., Ed.; Annu. Rep. Comput. Chem.; Elsevier, 2006; Vol. 2; pp 35 – 51.
- (20) Guo, W.; Wu, A.; Xu, X. XO: An extended ONIOM method for accurate and efficient geometry optimization of large molecules. *Chem. Phys. Lett.* **2010**, *498*, 203–208.

- (21) Chung, L. W.; Sameera, W. M.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H. B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.
- (22) Dahlke, E. E.; Truhlar, D. Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.
- (23) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1–6.
- (24) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller-Plesset perturbation theory energies for large water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.
- (25) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for Benchmark Accuracy with the Many-Body Expansion. *Acc. Chem. Res.* **2014**, *47*, 2828–2836.
- (26) Bygrave, P. J.; Allan, N. L.; Manby, F. R. The embedded many-body expansion for energetics of molecular crystals. *J. Chem. Phys.* **2012**, *137*, 164102.
- (27) Gao, J.; Wang, Y. Communication: Variational many-body expansion: Accounting for exchange repulsion, charge delocalization, and dispersion in the fragment-based explicit polarization method. *J. Chem. Phys.* **2012**, *136*, 71101.
- (28) Song, L.; Han, J.; Lin, Y. L.; Xie, W.; Gao, J. Explicit polarization (X-poi) potential using ab initio molecular orbital theory and density functional theory. *J. Phys. Chem. A* **2009**, *113*, 11656–11664.
- (29) Gao, J. Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulations. *J. Phys. Chem. B* **1997**, *101*, 657–663.

- (30) Gao, J. A molecular-orbital derived polarization potential for liquid water. *J. Chem. Phys.* **1998**, *109*, 2346–2354.
- (31) Xie, W.; Gao, J. Design of a next generation force field: The X-POL potential. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.
- (32) Liu, J.; Rana, B.; Liu, K.-Y.; Herbert, J. M. Variational Formulation of the Generalized Many-Body Expansion with Self-Consistent Charge Embedding: Simple and Correct Analytic Energy Gradient for Fragment-Based ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3877–3886.
- (33) Lao, K. U.; Herbert, J. M. Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion. *J. Phys. Chem. Lett.* **2012**, *3*, 3241–3248.
- (34) Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. On the Interfragment Exchange in the X-Pol Method. *J. Chem. Theory Comput.* **2010**, *6*, 2469–2476.
- (35) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
- (36) Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (37) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632–672.
- (38) Fedorov, D. G.; Kitaura, K. In *Modern Methods for Theoretical Physical Chemistry*

of *Biopolymers*; Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier Science: Amsterdam, 2006; Chapter 1, pp 3 – 38.

- (39) Stoll, H.; Wagenblast, G.; Preuss, H. On the Use of Local Basis-Sets for Localized Molecular-Orbitals. *Theor. Chim. Acta* **1980**, *57*, 169–178.
- (40) Gianinetti, E.; Raimondi, M.; Tornaghi, E. Modification of the Roothaan equations to exclude BSSE from molecular interaction calculations. *Int. J. Quantum Chem.* **1996**, *60*, 157–166.
- (41) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124*, 204105.
- (42) Nagata, T.; Takahashi, O.; Saito, K.; Iwata, S. Basis set superposition error free self-consistent field method for molecular interaction in multi-component systems: Projection operator formalism. *J. Chem. Phys.* **2001**, *115*, 3553.
- (43) Cullen, J. M. An examination of the effects of basis set and charge transfer in hydrogen-bonded dimers with a constrained Hartree–Fock method. *Int. J. Quantum Chem.* **1991**, *40*, 193–207.
- (44) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J. Phys. Chem. A* **2007**, *111*, 8753–8765.
- (45) Horn, P. R.; Sundstrom, E. J.; Baker, T. A.; Head-Gordon, M. Unrestricted absolutely localized molecular orbitals for energy decomposition analysis: Theory and applications to intermolecular interactions involving radicals. *J. Chem. Phys.* **2013**, *138*, 30–32.
- (46) Thirman, J.; Head-Gordon, M. An energy decomposition analysis for second-order Møller-Plesset perturbation theory based on absolutely localized molecular orbitals

ARTICLES YOU MAY BE INTERESTED IN Energy decomposition analysis of covalent bonds and intermolecular interactions An energy decomposition analysis for second-order Møller-Plesset perturbation theory based on absolutely localized molecular orbitals. *J. Chem. Phys.* **2015**, *143*, 84124.

- (47) Mo, Y.; Gao, J.; Peyerimhoff, S. D. Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* **2000**, *112*, 5530–5538.
- (48) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *89*, 76, 1–58.
- (49) Gao, J.; Cembran, A.; Mo, Y. Generalized X-Pol Theory and Charge Delocalization States. *J. Chem. Theory Comput.* **2010**, *6*, 2402–2410, PMID: 26613495.
- (50) Ding, F.; Manby, F. R.; Miller, T. F. Embedded Mean-Field Theory with Block-Orthogonalized Partitioning. *J. Chem. Theory Comput.* **2017**, *13*, 1605–1615.
- (51) Mardirossian, N.; Head-Gordon, M. B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
- (52) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (53) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.
- (54) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

- (55) Gill, P. M.; Johnson, B. G.; Pople, J. A. A standard grid for density functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506–512.
- (56) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kus, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C. M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; Distasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T. C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S. P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; Oneill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y. C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z. Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J. D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C. P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.;

- Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.
- (57) Lee, J.; Head-Gordon, M. Regularized Orbital-Optimized Second-Order Møller–Plesset Perturbation Theory: A Reliable Fifth-Order-Scaling Electron Correlation Model with Orbital Energy Dependent Regularizers. *J. Chem. Theory Comput.* **2018**, *14*, 5203–5219.
- (58) Lee, J.; Head-Gordon, M. Distinguishing artificial and essential symmetry breaking in a single determinant: Approach and application to the C<sub>60</sub>, C<sub>36</sub>, and C<sub>20</sub> fullerenes. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4763–4778.
- (59) Lee, J.; Head-Gordon, M. Two single-reference approaches to singlet biradicaloid problems: Complex, restricted orbitals and approximate spin-projection combined with regularized orbital-optimized Møller-Plesset perturbation theory. *J. Chem. Phys.* **2019**, *150*, 244106.
- (60) Bertels, L. W.; Lee, J.; Head-Gordon, M. Third-Order Møller–Plesset Perturbation Theory Made Useful? Choice of Orbitals and Scaling Greatly Improves Accuracy for Thermochemistry, Kinetics, and Intermolecular Interactions. *J. Phys. Chem. Lett.* **2019**, *10*, 4170–4176.
- (61) Mata, R. A.; Werner, H. J. Calculation of smooth potential energy surfaces using local electron correlation methods. *J. Chem. Phys.* **2006**, *125*, 184110.
- (62) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. The limits of local correlation theory: Electronic delocalization and chemically smooth potential energy surfaces. *J. Chem. Phys.* **2008**, *128*, 34103.
- (63) Xantheas, S. S. Cooperativity and hydrogen bonding network in water clusters. *Chem. Phys.* **2000**, *258*, 225–231.

- (64) Sum, A.; Sandler, S. Ab Initio Calculations of Cooperativity Effects on Clusters of Methanol, Ethanol, 1-Propanol, and Methanethiol. *J. Phys. Chem. A* **2000**, *104*, 1121–1129.
- (65) Rincón, L.; Almeida, R.; García-Aldea, D.; Diez Y Riega, H. Hydrogen bond cooperativity and electron delocalization in hydrogen fluoride clusters. *J. Chem. Phys.* **2001**, *114*, 5552–5561.
- (66) Luck, W. A. The importance of cooperativity for the properties of liquid water. *J. Mol. Struct.* **1998**, *448*, 131–142.
- (67) Yuan, D.; Li, Y.; Ni, Z.; Pulay, P.; Li, W.; Li, S. Benchmark Relative Energies for Large Water Clusters with the Generalized Energy-Based Fragmentation Method. *J. Chem. Theory Comput.* **2017**, *13*, 2696–2704.
- (68) Temelso, B.; Renner, C. R.; Shields, G. C. Importance and reliability of small basis set CCSD(T) corrections to MP2 binding and relative energies of water clusters. *J. Chem. Theory Comput.* **2015**, *11*, 1439–1448.
- (69) Friedrich, J.; Fiedler, B. Accurate calculation of binding energies for molecular clusters - Assessment of different models. *Chem. Phys.* **2016**, *472*, 72–80.
- (70) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark structures and binding energies of small water clusters with anharmonicity corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.
- (71) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; Head-Gordon, M.; Clark, G. N.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

- (72) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- (73) Tainter, C. J.; Pieniazek, P. A.; Lin, Y. S.; Skinner, J. L. Robust three-body water simulation model. *J. Chem. Phys.* **2011**, *134*, 234505.
- (74) 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*, 5395–5403.
- (75) 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*, 7501–7528.
- (76) Medders, G. R.; Paesani, F. Infrared and Raman Spectroscopy of Liquid Water through "First-Principles" Many-Body Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1145–1154.
- (77) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **2016**, *145*, 194504.
- (78) Dang, L. X.; Chang, T.-M. Molecular dynamics study of water clusters, liquid, and liquid-vapor interface of water with many-body potentials. *J. Chem. Phys.* **1997**, *106*, 8149–8159.
- (79) Burnham, C. J.; Li, J.; Xantheas, S. S.; Leslie, M. The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the

study of water clusters (n=2-21) and the phonon spectrum of ice Ih. *J. Chem. Phys.* **1999**, *110*, 4566–4581.

- (80) Wang, C. C.; Zielke, P.; Sigurbjörnsson, O. F.; Viteri, C. R.; Signorell, R. Infrared Spectra of C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, and CO<sub>2</sub> Aerosols Potentially Formed in Titan's Atmosphere. *J. Phys. Chem. A* **2009**, *113*, 11129–11137.
- (81) DeSimone, J. M. Practical Approaches to Green Solvents. *Science* **2002**, *297*, 799–803.
- (82) Bukowski, R.; Sadlej, J.; Jeziorski, B.; Jankowski, P.; Szalewicz, K.; Kucharski, S. A.; Williams, H. L.; Rice, B. M. Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory. *J. Chem. Phys.* **1999**, *110*, 3785–3803.
- (83) Horn, P. R.; Head-Gordon, M. Polarization contributions to intermolecular interactions revisited with fragment electric-field response functions. *J. Chem. Phys.* **2015**, *143*, 114111.
- (84) Horn, P. R.; Mao, Y.; Head-Gordon, M. Defining the contributions of permanent electrostatics, Pauli repulsion, and dispersion in density functional theory calculations of intermolecular interaction energies. *J. Chem. Phys.* **2016**, *144*, 114107.
- (85) The CCSD(T) interaction energy reported in this work have been extrapolated to the Complete Basis Set (CBS) limit using the focal point approximation analysis.  $E_{CCSD(T)/CBS} \approx E_{HF/aug-cc-pV5Z}^{SCF} + E_{MP2/aug-cc-pVQZ \rightarrow aug-cc-pV5Z}^{corr} + (E_{CCSD(T)/aug-cc-pVTZ}^{corr} - E_{MP2/aug-cc-pVTZ}^{corr})$ . The MP2 correlation energy was extrapolated to uses the 2-point extrapolation formula<sup>7</sup> using the the aug-cc-pVQZ and the aug-cc-pV5Z basis sets.
- (86) Lemke, K. H.; Seward, T. M. Thermodynamic properties of carbon dioxide clusters by M06-2X and dispersion-corrected B2PLYP-D theory. *Chem. Phys. Lett.* **2013**, *573*, 19–23.

- (87) Jordan, K. D.; Sen, K. *Chemical Modelling: Volume 13*; The Royal Society of Chemistry, 2017; Vol. 13; pp 105–131.
- (88) Ludwig, R. Water: From Clusters to the Bulk. *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 1808 – 1827.
- (89) Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. The water dipole moment in water clusters. *Science* **1997**, *275*, 814–817.
- (90) 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.
- (91) Batista, E. R.; Xantheas, S. S.; Jonsson, H. Molecular multipole moments of water molecules in ice Ih. *J. Chem. Phys.* **1998**, *109*, 4546–4551.
- (92) Silvestrelli, P. L.; Parrinello, M. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.* **1999**, *111*, 3572–3580.
- (93) Ghanty, T. K.; Ghosh, S. K. Polarizability of water clusters: An ab initio investigation. *J. Chem. Phys.* **2003**, *118*, 8547–8550.
- (94) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Electron donation in the water-water hydrogen bond. *Chem. Eur. J.* **2009**, *15*, 851–855.
- (95) Cobar, E. A.; Horn, P. R.; Bergman, R. G.; Head-Gordon, M. Examination of the hydrogen-bonding networks in small water clusters ( $n = 2\text{-}5, 13, 17$ ) using absolutely localized molecular orbital energy decomposition analysis. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15328–15339.
- (96) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. Structure and stability of water clusters ( $\text{H}_2\text{O}$ ) $n$ ,  $n = 8\text{-}20$ : An ab initio investigation. *J. Phys. Chem. A* **2001**, *105*, 10525–10537.

- (97) González, B. S.; Hernández-Rojas, J.; Wales, D. J. Global minima and energetics of Li+(H<sub>2</sub>O)<sub>n</sub> and Ca<sup>2+</sup>(H<sub>2</sub>O)<sub>n</sub> clusters for n > 20. *Chem. Phys. Lett.* **2005**, *412*, 23–28.
- (98) Egan, C. K.; Paesani, F. Assessing Many-Body Effects of Water Self-Ions. I: OH – (H<sub>2</sub>O)<sub>n</sub> Clusters. *J. Chem. Theory Comput.* **2018**, *14*, 1982–1997.
- (99) Richard, R. M.; Lao, K. U.; Herbert, J. M. Understanding the many-body expansion for large systems. I. Precision considerations. *J. Chem. Phys.* **2014**, *141*, 014108.
- (100) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.
- (101) Becke, A. Density Functional Thermochemistry III The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (102) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (103) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- (104) ErgoSCF. 2019; <http://www.ergoscf.org/> [Accessed: 13 May 2019].
- (105) Almlöf, J.; Faegri, K.; Korsell, K. Principles for a direct SCF approach to LCAO-MO ab-initio calculations. *J. Comput. Chem.* **1982**, *3*, 385–399.
- (106) Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.* **2015**, *143*, 24113.
- (107) Manzer, S. F.; Epifanovsky, E.; Head-Gordon, M. Efficient implementation of the pair atomic resolution of the identity approximation for exact exchange for hybrid and range-separated density functionals. *J. Chem. Theory Comput.* **2015**, *11*, 518–527.

- (108) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. *Reviews in Computational Chemistry*; John Wiley Sons, Ltd, 2007; Chapter 1, pp 1–82.
- (109) Hu, W.; Lin, L.; Yang, C. DGDFT: A massively parallel method for large scale density functional theory calculations. *J. Chem. Phys.* **2015**, *143*, 124110.
- (110) Mardirossian, N.; Head-Gordon, M. Survival of the most transferable at the top of Jacob's ladder: Defining and testing the  $\omega$ B97M(2) double hybrid density functional. *J. Chem. Phys.* **2018**, *148*, 241736.
- (111) Lochan, R. C.; Head-Gordon, M. Orbital-optimized opposite-spin scaled second-order correlation: An economical method to improve the description of open-shell molecules. *J. Chem. Phys.* **2007**, *126*, 164101.
- (112) Stück, D.; Head-Gordon, M. Regularized orbital-optimized second-order perturbation theory. *J. Chem. Phys.* **2013**, *139*, 244109.
- (113) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479 – 483.