An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion

Leif D. Jacobson and John M. Herbert[a)]
Department of Chemistry, The Ohio State University, Columbus, Ohio 43210, USA

(Received 31 December 2010; accepted 8 February 2011; published online 7 March 2011)

We report a fragment-based electronic structure method, intended for the study of clusters and molecular liquids, that incorporates electronic polarization (induction) in a self-consistent fashion but treats intermolecular exchange and dispersion interactions perturbatively, as post-self-consistent field corrections, using a form of pairwise symmetry-adapted perturbation theory. The computational cost of the method scales quadratically as a function of the number of fragments (monomers), but could be made to scale linearly by exploiting distance-dependent thresholds. Extensive benchmark calculations are reported using the S22 database of high-level *ab initio* binding energies for dimers, and we find that average errors can be reduced to <1 kcal/mol with a suitable choice of basis set. Comparison to *ab initio* benchmarks for water clusters as large as $(H_2O)_{20}$ demonstrates that the method recovers $\gtrsim$90% of the binding energy in these systems, at a tiny fraction of the computational cost. As such, this approach represents a promising path toward accurate, systematically improvable, and parameter-free simulation of molecular liquids. © *2011 American Institute of Physics*. [doi:10.1063/1.3560026]

I. INTRODUCTION

It is well-established that wavefunction-based quantum chemistry using Gaussian basis sets can provide accurate ground-state properties for molecular systems. Application of these same robust methods to condensed-phase systems—be they periodic (crystalline solids) or nonperiodic (liquids or amorphous solids)—represents a tremendous challenge, given that the cost of wavefunction methods scales incredibly poorly with system size. In fact, such methods scale *unphysically* with system size,[1] as a result of the use of delocalized molecular orbitals, and the distance dependence of different intermolecular interactions *must* be exploited if we are to apply these methods to condensed phases.[2] From this point of view, Gaussian basis sets possess an inherent advantage over plane-wave basis sets, in that the basis functions are highly localized and the length scale between different basis functions can be assigned in a straightforward way.[3] At present, however, most condensed-phase electronic structure calculations are performed using delocalized plane-wave basis sets, which is appealing for periodic systems because the basis functions are periodic, but may be less advantageous for nonperiodic systems. Moreover, despite many advances in the development of "linear scaling" electronic structure methods,[4,5] routine application of wavefunction-based quantum chemistry remains limited to systems not larger than $\sim$100 atoms.[6]

In the past decade, a variety of fragment-based methods have been introduced, whose goal is to reduce the cost of quantum chemistry calculations in large systems. Such techniques include methods based upon a many-body expansion of the supersystem energy,[7–13] the most sophisticated of which is the "fragment molecular orbital" (FMO) method;[14–16] divide-and-conquer approaches;[17] the "self-consistent field for molecular interactions" (SCF-MI) technique;[18,19] the "natural linear scaling" coupled-cluster approach;[20,21] the effective fragment potential (EFP) method;[22–24] and the "explicit polarization" (XPol) method.[25–28] A complete discussion of the strengths and weaknesses of each of these methods is beyond the scope of the present article. Here, it suffices to note that each one—save for EFP, which is an automated way to parameterize polarizable force fields using *ab initio* calculations—requires iterative construction of Fock matrices for dimers of fragments, or sometimes the supersystem itself, or else requires additional, empirical parameters. The present article reports a method that avoids these requirements.

For maximum versatility, a fragment-based quantum chemistry method should allow for the possibility of fragmenting the system across covalent bonds, and several of the aforementioned methods do facilitate this possibility.[28–31] In the present work, however, we exclude fragmentation across covalent bonds, with the aim of developing an accurate and efficient method for molecular clusters, liquids, and solids that are composed of relatively small monomers, each of which constitutes one fragment.

To construct a low-scaling quantum chemistry method, one can imagine at least two (rather disparate) strategies. One strategy is to make well-defined approximations to an existing *ab initio* method, then examine the extent to which the approximate method is faithful to the original one. Examples of this approach include local correlation methods,[4,32–39] dual basis procedures,[40,41] and density fitting or resolution-of-identity techniques.[42–46] Alternatively, one might construct a method that is promising in its efficiency, and is based upon

a)Electronic mail: herbert@chemistry.ohio-state.edu.

either well-defined approximations or else some observations about the physical nature of interactions. The latter approach, which is the one pursued here, does not seek to reproduce or approximate the energetics of any existing quantum chemistry method.

The method that we propose herein was motivated in part by the electrostatically-embedded, many-body expansion method introduced recently by Dahlke and Truhlar.[7–10] In a simple many-body expansion, the energy of N interacting molecules is decomposed into a sum of one-body terms, two-body terms, etc., $V = V_1 + V_2 + V_3 + \cdots$. For example, V_2 is the sum of the energies of all $N(N-1)/2$ dimers, minus the sum of the N monomer energies to avoid double-counting. Application to water clusters at the level of second-order Møller–Plesset perturbation theory (MP2) demonstrates that four-body terms must be retained in order to obtain accurate results.[47,48] However, Dahlke and Truhlar[7–10] demonstrated that convergence of the many-body expansion is greatly accelerated if low-order n-body calculations are computed in the presence of a set of point charges that approximate the electrostatic potential due to the other monomers. The best results, as compared to supersystem MP2, were obtained using a supersystem Hartree–Fock (HF) calculation followed by a two-body expansion of the MP2 correlation energy.[7,8] Alternatively, accurate results for hydrogen-bonded clusters have been obtained by computing V_1 and V_2 at the MP2 level while using a polarizable force field to evaluate V_3 and V_4, without the need to perform a supersystem HF calculation.[11–13]

We glean two important conclusions from these observations. First, it is crucial to include the electrostatic effects of the environment in the one-body (monomer) calculations. Second, although polarization (induction) is an inherently many-body phenomenon, intermolecular electron correlation is largely a two-body effect. These observations suggest that if one can efficiently incorporate many-body induction in a fragment-based scheme, then it may be possible to approximate intermolecular electron correlation in a pairwise fashion, without resorting to supersystem calculations or high-order terms in the many-body expansion.

To accomplish this, we will use the XPol method of Xie *et al.*[26] to perform electrostatically embedded one-body calculations in a variational, self-consistent manner. The XPol method, which is detailed in Sec. II, involves solving single-fragment SCF equations in the presence of point charges that represent the electrostatic potential due to the other fragments. These point charges are derived from the fragment wavefunctions, but unlike both the FMO method[49,50] and the electrostatic embedding method of Dahlke and Truhlar,[7–10] variation of the point charges is included (exactly) within the single-fragment Fock matrices. Operationally, the method is a "dual SCF" procedure, with an outer loop over fragments and an inner loop to solve the single-fragment Roothaan equations. As such, the XPol method incorporates many-body induction (albeit in an approximate way), but ignores electron exchange between fragments. The original XPol method developed by Xie *et al.*[25–28] accounts for dispersion and exchange-repulsion interactions using empirical Lennard-Jones potentials. Our goal is to replace these empirical corrections with *ab initio*, post-XPol corrections based on a two-body form of symmetry-adapted perturbation theory (SAPT).[51,52]

Traditionally, SAPT has been used as a benchmark method for computing dimer binding energies and for decomposing intermolecular interactions into induction, exchange-repulsion, dispersion, and other components.[51,52] SAPT calculations can be quite expensive, although reduced-cost variants based on density functional theory (DFT) have recently shown great promise.[53–58] Here, we use the variant known as SAPT(0), which employs HF orbitals for the monomers and does not include monomer electron correlation. We also explore SAPT(KS),[53] which is analogous to SAPT(0) but substitutes Kohn–Sham (KS) orbitals in place of HF orbitals.

Our proposed method incorporates self-consistent many-body induction, but assumes that other interactions, such as exchange-repulsion and dispersion, are pairwise additive. Calculation of nonadditive three-body interactions within the SAPT formalism requires computationally expensive triple excitations,[59–61] but SAPT results for $OH^-(H_2O)_2$ indicate that the two-body terms are about an order of magnitude larger than the three-body terms. Furthermore, the three-body terms are dominated by the induction correction,[62] that is, by the change in electrostatics due to the presence of other molecules. In addition, energy decomposition analysis of $(H_2O)_6$ isomers reveals that electrostatic, exchange-repulsion, and dispersion interactions *are* pairwise additive (or nearly so), whereas polarization exhibits many-body effects of $\sim$10 kcal/mol.[63] In our proposed scheme, many-body polarization is mostly captured within the zeroth-order wavefunctions generated by the XPol procedure, and electrostatic embedding serves to reduce the magnitude of the intermolecular perturbation. The fact that the MP2 correlation energy is approximated to high accuracy with only a two-body expansion[8] suggests that we need only extend SAPT to pairwise fragment interactions, which will make the method highly efficient for applications to molecular liquids.

II. THEORY

The method introduced here is essentially an amalgam of two existing methods, XPol and SAPT, so we call the new method XPol/SAPT or XPS. In what follows, we will use indices A and B to label fragments; i and j to label electrons; $a \in A$ and $b \in B$ to label occupied MOs belonging to fragments A and B, respectively; $r \in A$ and $s \in B$ to label virtual orbitals; Greek letters (μ, ν, λ, σ) to label atomic orbital (AO) basis functions; and I, J, K, ... to label nuclei. We restrict our attention to closed-shell spin-restricted calculations. Atomic units are used throughout this section.

A. XPol

XPol is an approximate, fragment-based molecular orbital method that was developed to be a "next-generation" force field.[25,28] This method starts from an *ansatz* in which the supersystem wavefunction is written as a direct product of fragment wavefunctions,[26,64]

$$|\Psi\rangle = \prod_{A}^{N_{\text{frag}}} |\Psi_A\rangle, \tag{1}$$

where N_{frag} is the number of fragments. We assume here that the fragments are molecules and that covalent bonds remain intact, although XPol has been extended to fragmentation of the system across covalent bonds.[26,28] The fragment wavefunctions are antisymmetric with respect to exchange of electrons within a fragment, but not to exchange between fragments. As such, exchange interactions between fragments are neglected. In the original version of XPol, Lennard-Jones potentials are used to model exchange-repulsion and dispersion interactions,[26] but our aim is to replace these empirical interaction terms with perturbation theory. Very recently, Gao and co-workers extended XPol to include interfragment exchange within the SCF iterations,[65] in a manner highly analogous to the SCF-MI approach.[19] In contrast, our method treats exchange as a post-SCF correction.

For closed-shell fragments, the XPol energy is[26]

$$E_{\text{XPol}} = \sum_{A} \left[2 \sum_{a} \mathbf{c}_a^{\dagger} \left(\mathbf{h}^A + \mathbf{J}^A - \tfrac{1}{2}\mathbf{K}^A \right) \mathbf{c}_a + E_{\text{nuc}}^A \right] + E_{\text{embed}}. \tag{2}$$

The term in square brackets is the ordinary HF energy expression,[66] for fragment A. Thus, $\mathbf{c}_a$ is a vector of occupied MO expansion coefficients (in the AO basis) for the occupied MO $a \in A$; $\mathbf{h}^A$ consists of the one-electron integrals; and $\mathbf{J}^A$ and $\mathbf{K}^A$ are the Coulomb and exchange matrices, respectively, constructed from the density matrix for fragment A. The additional terms in Eq. (2),

$$E_{\text{embed}} = \frac{1}{2} \sum_{A} \sum_{B \neq A} \sum_{J \in B} \left(-2 \sum_{a} \mathbf{c}_a^{\dagger} \mathbf{I}_J \mathbf{c}_a + \sum_{I \in A} L_{IJ} \right) q_J, \tag{3}$$

arise from the electrostatic embedding. The matrix $\mathbf{I}_J$ is defined by its AO matrix elements,

$$(\mathbf{I}_J)_{\mu\nu} = \langle \mu | \frac{1}{|\vec{r} - \vec{R}_J|} | \nu \rangle, \tag{4}$$

and L_{IJ} is given by

$$L_{IJ} = \frac{Z_I}{|\vec{R}_I - \vec{R}_J|}. \tag{5}$$

According to Eqs. (2) and (3), each fragment is embedded in the electrostatic potential arising from a set of point charges, $\{q_J\}$, on all of the other fragments; the factor of 1/2 in Eq. (3) avoids double-counting. Exchange interactions between fragments are ignored, and the electrostatic interactions between fragments are approximated by interactions between the charge density of one fragment and point charges on the other fragments.

Crucially, the vectors $\mathbf{c}_a$ are constructed within the absolutely localized MO (ALMO) *ansatz*,[19] meaning that the MOs for each fragment are represented in terms of only those AOs that are centered on atoms in the same fragment. This partition of the AO basis leads to significant computational savings, and affords a method whose cost grows linearly with respect to N_{frag}. The ALMO *ansatz* also excludes basis set superposition error (BSSE) by construction, and in compact basis sets it excludes interfragment charge transfer as well.[19]

The original XPol method of Xie *et al.*[25–28] uses Mulliken charges for the q_J in Eq. (3), though other charge schemes could be envisaged. In addition to Mulliken charges, we will examine Löwdin charges[66] and charges derived from the electrostatic potential (CHELPG).[67] This aspect of the method is discussed in Sec. II B, with additional details given in the supplementary material.[68]

Derivation of the XPol working equations follows closely that of the HF equations. We require the energy expression in Eq. (2) to be stationary with respect to variation of the MO coefficients, subject to the constraint that MOs *within* each fragment are orthonormal. This leads to the XPol SCF equations,[26]

$$\mathbf{F}^A \mathbf{C}^A = \mathbf{S}^A \mathbf{C}^A \boldsymbol{\epsilon}^A. \tag{6}$$

Here, $\mathbf{F}^A$, $\mathbf{C}^A$, $\mathbf{S}^A$, and $\boldsymbol{\epsilon}^A$ are the fragment Fock matrix, MO coefficient matrix, AO overlap matrix, and Lagrange multiplier matrix, respectively. The dimension of these matrices equals the number of AOs centered on fragment A. Upon diagonalizing $\boldsymbol{\epsilon}^A = \left(\mathbf{C}^A\right)^{\dagger} \mathbf{F}^A \mathbf{C}^A$, the diagonal elements of $\boldsymbol{\epsilon}^A$ are the eigenvalues of the fragment Fock matrix, $\mathbf{F}^A$. In the AO basis, $\mathbf{F}^A$ has matrix elements

$$F_{\mu\nu}^A = f_{\mu\nu}^A - \frac{1}{2} \sum_{J \notin A} (I_J)_{\mu\nu} q_J + \sum_{J \in A} M_J (\Lambda_J)_{\mu\nu}, \tag{7}$$

where $\mathbf{f}^A = \mathbf{h}^A + 2\mathbf{J}^A - \mathbf{K}^A$ is the Fock matrix for fragment A in isolation. The additional XPol terms consist of an "M-vector" defined by

$$M_J = \frac{\partial E_{\text{embed}}}{\partial q_J} = \frac{1}{2} \sum_{\substack{B \\ (J \notin B)}} \left(-2 \sum_{b} \mathbf{c}_b^{\dagger} \mathbf{I}_J \mathbf{c}_b + \sum_{I \in B} L_{IJ} \right), \tag{8}$$

and also

$$(\Lambda_J)_{\mu\nu} = \frac{\partial q_J}{\partial P_{\mu\nu}}. \tag{9}$$

Here, $\mathbf{P}$ represents the one-electron density matrix, which is block diagonal in the fragment index.

Our notation differs slightly from that used by Xie *et al.*,[26] though we believe that ours is more transparent. We have written Eq. (7) in a general form that is valid for any charge scheme, and explicit formulas for the charge derivatives, $(\Lambda_J)_{\mu\nu}$, will be presented below. As pointed out by Xie *et al.*,[26] Eq. (7) indicates that each fragment is polarized by the rest of the system, with half of this polarization stemming from point charges on the other fragments and half from the true charge density of the rest of the system, which is contained in the M-vector. In principle, the electrostatic embedding could be systematically improved, by using higher order multipoles or the fragment densities themselves, but we will not explore this possibility here.

Solution of the XPol equations requires a dual SCF procedure, since each fragment Fock matrix depends upon the

electron density of the other fragments. Implementation of the dual SCF is straightforward and will be discussed only briefly, in Sec. III.

B. Charge schemes

We investigate three different charge schemes: Mulliken, Löwdin, and CHELPG. In this section we will derive formulas necessary to incorporate Mulliken charges into the XPS method. Details pertaining to the use of Löwdin and CHELPG charges are given in the supplementary material.[68]

The Mulliken charges stem from a simple partitioning of the electron density,[66]

$$q_J^{\text{Mull}} = Z_J - \sum_{\substack{\mu\in J \\ \nu}} S_{\mu\nu} P_{\nu\mu}. \tag{10}$$

The derivative with respect to a density matrix element is quite simple and can be written as

$$\left(\Lambda_J^{\text{Mull}}\right)_{\mu\nu} = -\frac{1}{2}(S_{\mu\nu}\delta_{\mu\in J} + S_{\nu\mu}\delta_{\nu\in J}), \tag{11}$$

where $\delta_{\mu\in J} = 1$ if the basis function μ is centered on atom J, and $\delta_{\mu\in J} = 0$ otherwise.

In what follows, we will require an operator, $\hat{\Lambda}_J$, whose matrix elements are equal to $(\Lambda_J)_{\mu\nu}$ as given by Eq. (11). It can be verified that a reasonable choice is

$$\hat{\Lambda}_J^{\text{Mull}} = -\frac{1}{2}\sum_{\mu,\nu\in J}\left(|\mu\rangle\left(\mathbf{S}_J^{-1}\right)_{\mu\nu}\langle\nu| + |\nu\rangle\left(\mathbf{S}_J^{-1}\right)_{\nu\mu}\langle\mu|\right). \tag{12}$$

The quantity $\mathbf{S}_J^{-1}$ in this equation refers to the inverse of the *fragment* overlap matrix, for the fragment that contains atom J. At no point it is necessary to invert the supersystem's overlap matrix.

C. Symmetry-adapted perturbation theory

Two issues prevent us from applying perturbation theory to the XPol wavefunction, Eq. (1), in a straightforward manner. First, the fragment wavefunctions (and ALMOs) are not mutually orthogonal amongst the fragments. Second, while the fragment wavefunctions are properly antisymmetric with respect to exchange of electrons *within* a fragment, the direct product *ansatz* in Eq. (1) is not antisymmetric with respect to exchanges *between* fragments. Symmetry-adapted perturbation theory was developed to overcome precisely these two problems. Here, we review only the most relevant details of SAPT; see Refs. 51 and 52 for a complete introduction to the method.

In SAPT, the Hamiltonian for the $A\cdots B$ dimer is written as[51,69]

$$\hat{H} = \hat{F}^A + \hat{F}^B + \xi\hat{W}^A + \eta\hat{W}^B + \zeta\hat{V}, \tag{13}$$

where $\hat{W}^A$ and $\hat{W}^B$ are Møller–Plesset fluctuation operators for fragments A and B, and the intermolecular perturbation, $\hat{V}$, is conveniently written as

$$\hat{V} = \sum_{i\in A}\sum_{j\in B}\hat{v}(ij) \tag{14}$$

with

$$\hat{v}(ij) = \frac{1}{|\vec{r}_i - \vec{r}_j|} + \frac{\hat{v}_A(j)}{N_A} + \frac{\hat{v}_B(i)}{N_B} + \frac{V_0}{N_A N_B}. \tag{15}$$

The quantity V_0 is the nuclear interaction energy between the two fragments, and

$$\hat{v}_A(j) = -\sum_{I\in A}\frac{Z_I}{|\vec{r}_j - \vec{R}_I|} \tag{16}$$

describes the interaction of electron $j \in B$ with nuclei $I \in A$. In what one might call "traditional" SAPT, the interaction energy is expanded in a triple perturbation series in the parameters ξ, η, and ζ.[51,70]

Here, we expand only with respect to ζ. When the zeroth-order monomer wavefunctions come from HF theory, this is usually termed SAPT(0), where the "(0)" means zeroth-order in the monomer fluctuation potentials.[51] Within this formalism, the interaction energy is given by a symmetrized Rayleigh–Schrödinger perturbation expansion,[52,70–72]

$$E_{\text{int}}(\zeta) = \frac{\langle\Psi_0|\zeta\hat{V}\hat{\mathcal{A}}_{AB}|\Psi(\zeta)\rangle}{\langle\Psi_0|\hat{\mathcal{A}}_{AB}|\Psi(\zeta)\rangle}, \tag{17}$$

where $\hat{\mathcal{A}}_{AB}$ is an antisymmetrizer for the $A\cdots B$ supersystem that projects out the Pauli-forbidden components of the supersystem wavefunction, $|\Psi(\zeta)\rangle$. The zeroth-order wavefunction, $|\Psi_0\rangle$, is taken to be a direct product of the monomer wavefunctions, and the interaction energy is expanded with respect to ζ. At every order in the perturbative expansion of Eq. (17), there is a polarization term analogous to what would be obtained in ordinary Rayleigh–Schrödinger perturbation theory, along with an exchange term.[52] Here, we consider the expansion through second order, in which case the interaction energy can be decomposed as[53]

$$E_{\text{int}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{pol}}^{(2)} + E_{\text{exch}}^{(2)}. \tag{18}$$

The various terms in this expression are discussed below.

The antisymmetrizer in Eq. (17) can be written as[51,70]

$$\hat{\mathcal{A}}_{AB} = \frac{N_A!N_B!}{(N_A+N_B)!}\,\hat{\mathcal{A}}_A\hat{\mathcal{A}}_B(\hat{1} + \hat{\mathcal{P}}^{AB} + \hat{\mathcal{P}}'), \tag{19}$$

where $\hat{\mathcal{A}}_A$ and $\hat{\mathcal{A}}_B$ are antisymmetrizers for the two monomers, and $\hat{\mathcal{P}}^{AB}$ is a sum of all one-electron exchange operators between the two monomers. The operator $\hat{\mathcal{P}}'$ in Eq. (19) denotes all of the three-electron and higher order exchanges. This operator is neglected in what is known as the "single-exchange" approximation.[51,52,60] This approximation is expected to be quite accurate at typical van der Waals distances,[52] and we invoke it here.

The electrostatic part of the first-order energy correction is denoted $E_{\text{elst}}^{(1)}$ in Eq. (18), and represents the Coulomb interaction between the two monomer electron densities.[52] The quantity $E_{\text{exch}}^{(1)}$ in Eq. (18) is the corresponding first-order (Hartree–Fock) exchange correction. Explicit formulas for these corrections can be found in the literature.[51,52,70,73]

The second-order polarization correction in Eq. (18) can be further decomposed into induction and dispersion contributions, $E_{\text{pol}}^{(2)} = E_{\text{ind}}^2 + E_{\text{disp}}^2$. The induction correction can be written as

$$E_{\rm ind}^{(2)} = E_{\rm ind}^{(2)}(A \leftarrow B) + E_{\rm ind}^{(2)}(B \leftarrow A), \tag{20}$$

where the notation $(A \leftarrow B)$ indicates that the frozen charge density of B polarizes the density of A. In detail,[51]

$$E_{\rm ind}^{(2)}(A \leftarrow B) = 2\sum_{ar} t_{ar}(w_B)_{ar}, \tag{21}$$

where

$$(w_B)_{ar} = (\hat{v}_B)_{ar} + \sum_b (ar|bb) \tag{22}$$

and $t_{ar} = (w_B)_{ar}/(\epsilon_a - \epsilon_r)$. The second term in Eq. (20), in which A polarizes B, is obtained from Eqs. (21) and (22) by interchanging the following labels: $A \leftrightarrow B$, $a \leftrightarrow b$, and $r \leftrightarrow s$.[51] Finally, the dispersion correction is

$$E_{\rm disp}^{(2)} = 4\sum_{abrs} \frac{(ar|bs)(ra|sb)}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s}. \tag{23}$$

The induction and dispersion corrections both have accompanying exchange corrections (termed exchange-induction and exchange-dispersion, $E_{\rm exch\text{-}ind}^{2}$ and $E_{\rm exch\text{-}disp}^{2}$). Expressions for these corrections are quite lengthy and can be found in the literature.[51,58]

Although an exact first-order exchange correction that avoids the single-exchange approximation has been derived,[74] to the best of our knowledge the exchange-induction and exchange-dispersion formulas have only been derived within the single-exchange approximation. We have tested this exact formula for $E_{\rm exch}^{(1)}$, and we find that the single-exchange approximation is quite robust. Moreover, the single-exchange approximation does not require inversion of the dimer overlap matrix, which *is* required in the exact formula for $E_{\rm exch}^{(1)}$. For this reason, and in the interest of consistency, we invoke the single-exchange approximation for all of the exchange interaction terms.

It is quite common to replace $E_{\rm ind}^{(2)}$ and $E_{\rm exch\text{-}ind}^{(2)}$ with their "response" analogues, $E_{\rm ind,resp}^{(2)}$ and $E_{\rm exch\text{-}ind,resp}^{(2)}$, which afford the infinite-order correction for polarization arising from a frozen partner density.[51] Operationally, this substitution involves replacing the second-order induction amplitudes, t_{ar} in Eq. (21), with amplitudes obtained from solution of the coupled-perturbed Hartree–Fock equations.[75] (The perturbation is simply the electrostatic potential of the other monomer.) In addition, it is common to correct the SAPT(0) binding energy for "higher order" polarization effects, by adding a correction term

$$\delta E_{\rm int}^{\rm HF} = E_{\rm int}^{\rm HF} - \left(E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm ind,resp}^{(2)} + E_{\rm exch\text{-}ind,resp}^{(2)}\right) \tag{24}$$

to the interaction energy.[51] Here, $E_{\rm int}^{\rm HF}$ is the counterpoise-corrected HF binding energy for $A \cdots B$.

If one simply replaces the MOs and eigenvalues of the SAPT(0) corrections with KS MOs and eigenvalues, then the energy expressions above define the method known as SAPT(KS).[53] [This approach was originally termed SAPT(DFT),[53] but this terminology is now reserved for an alternative DFT-based SAPT method.[55–58]] Initially, the SAPT(KS) approach was not deemed very successful, as the electrostatic and induction energies failed to reproduce (traditional) benchmark SAPT values.[54] However, these discrepancies were ultimately determined to result from the incorrect asymptotic behavior of common exchange-correlation (XC) functionals used in DFT. Applying an asymptotic correction to the XC potential improved the agreement with benchmark values,[54] though the dispersion correction was still poor. This led to the development of the method that is nowadays called SAPT(DFT),[55–58] in which the sum-over-states formula for the dispersion interaction [Eq. (23)] is replaced with a generalized Casimir–Polder-type expression based on frequency-dependent density susceptibilities for the monomers, which are calculated by solving time-dependent coupled HF or KS equations.[55–58] We have not yet implemented SAPT(DFT), but the SAPT(KS) approach will be considered here, in addition to SAPT(0).

D. XPol/SAPT

There are two difficulties in combining XPol with SAPT. First, the perturbation in SAPT is not appropriate when XPol is used to obtain zeroth-order wavefunctions and energies, due to the fact that some intermolecular interactions have already been included, albeit approximately, by means of electrostatic embedding. In addition, SAPT(0) must be extended to incorporate an arbitrary number of monomers.

Taking the zeroth-order Hamiltonian to be the sum of the fragment Fock operators defined in Eq. (7), we can write

$$\hat{H}_0 = \sum_A \sum_{i\in A} \hat{f}^A(i) - \sum_A \sum_{i\in A}\left[\frac{1}{2}\sum_{J\notin A} q_J \hat{I}_J(i) - \sum_{J\in A} M_J \hat{\Lambda}_J(i)\right], \tag{25}$$

where $\hat{f}^A(i)$ is the gas-phase (isolated fragment) Fock operator for electron i in fragment A. Equation (25) can be rewritten as

$$\begin{aligned}\hat{H}_0 = &\sum_A \sum_{i\in A} \hat{f}^A(i) \\ &- \sum_A \sum_{B>A}\left[\sum_{i\in A}\left(\frac{1}{2}\sum_{J\in B} q_J \hat{I}_J(i) - \sum_{I\in A} M_I^B \hat{\Lambda}_I(i)\right)\right. \\ &\left. + \sum_{j\in B}\left(\frac{1}{2}\sum_{I\in A} \hat{I}_I(j) q_I - \sum_{J\in B} M_J^A \hat{\Lambda}_J(j)\right)\right],\end{aligned} \tag{26}$$

where M_J^A is the Jth element of an M-vector that contains only contributions from fragment A. This partitioning of $\hat{H}_0$ suggests that we replace $\hat{v}_A$ and $\hat{v}_B$ in Eq. (15) with

$$\hat{v}_A(j) = -\sum_{I\in A}\left(Z_I - \frac{1}{2}q_I\right)\hat{I}_I(j) - \sum_{J\in B} M_J^A \hat{\Lambda}_J(j), \tag{27a}$$

$$\hat{v}_B(i) = -\sum_{J\in B}\left(Z_J - \frac{1}{2}q_J\right)\hat{I}_J(i) - \sum_{I\in A} M_I^B \hat{\Lambda}_I(i). \tag{27b}$$

Once these substitutions have been made, we use the standard SAPT(0) corrections, in a pairwise manner, to calculate all dimer interaction energies, $E_{\rm int}^{AB}$, using Eq. (18) in conjunction

with the perturbation defined in Eq. (27). The total energy within the XPS *ansatz* is then given by

$$E_{\text{XPS}} = \sum_A \left(\sum_a \left[2\,\epsilon_a^A - \mathbf{c}_a^\dagger (\mathbf{J}^A - \tfrac{1}{2}\mathbf{K}^A)\mathbf{c}_a \right] + E_{\text{nuc}}^A + \sum_{B>A} E_{\text{int}}^{AB} \right). \tag{28}$$

In this expression, we have removed the over-counting of two-electron interactions present in HF theory, effectively taking the *intrafragment* perturbation to first order. The generalization from a HF to a KS description of the monomers is straightforward.

To better understand what we have just done, let us first generalize the SAPT Hamiltonian, Eq. (13), to an arbitrary number of fragments by writing

$$\hat{H} = \sum_A (\hat{F}^A + \xi_A \hat{W}^A) + \sum_A \sum_{B>A} \zeta_{AB} \hat{V}_{AB}. \tag{29}$$

In this expression, there are N_{frag} intramolecular perturbations ($\xi_A \hat{W}^A$) and $N_{\text{dimer}} = N_{\text{frag}}(N_{\text{frag}} - 1)/2$ intermolecular perturbations ($\zeta_{AB} \hat{V}_{AB}$). In principle, one should therefore employ a ($N_{\text{frag}} \times N_{\text{dimer}}$)-tuple perturbative expansion for this Hamiltonian, analogous to the triple perturbation expansion that is applied to Eq. (13). However, we neglect intrafragment electron correlation, analogous to SAPT(0) for dimers. In addition, we neglect second-order terms arising from coupling between first-order perturbations on different dimers.

We have also made a third (and somewhat more subtle) approximation. Neglecting monomer electron correlation ($\hat{W}^A$) in the Hamiltonian of Eq. (29), the first-order (intermolecular) energy correction is [cf. Eq. (17)]

$$E^{(1)} = \frac{\langle \Psi_0 \left| \left(\sum_A \sum_{B>A} \hat{V}_{AB} \right) \hat{\mathcal{A}} \right| \Psi_0 \rangle}{\langle \Psi_0 | \hat{\mathcal{A}} | \Psi_0 \rangle}. \tag{30}$$

Here, $|\Psi_0\rangle$ is the zeroth-order, direct-product wavefunction for the supersystem. Note that $\hat{\mathcal{A}}$ is a supersystem antisymmetrizer and is not pairwise additive. This operator can be expressed as

$$\hat{\mathcal{A}} = \frac{\left(\prod_A N_A! \hat{\mathcal{A}}_A \right)}{\left(\sum_A N_A \right)!} \left(1 + \sum_A \sum_{B>A} \hat{\mathcal{P}}^{AB} + \hat{\mathcal{P}}' \right), \tag{31}$$

where the operator $\hat{\mathcal{P}}^{AB}$ generates all pairwise electron exchanges between fragments A and B. Higher order exchange terms contained in $\hat{\mathcal{P}}'$ are neglected within the single-exchange SAPT approximation that was introduced in Sec. II C. In XPS, however, there are multiple dimers of fragments, and in developing this method we have tacitly introduced a further approximation in which only "diagonal" terms such as $\langle \Psi_0 | \hat{V}_{AB} (\hat{1} + \hat{\mathcal{P}}^{AB}) | \Psi_0 \rangle$ are retained in the numerator of Eq. (30). This approximation neglects some single-exchange terms involving three or more fragments. For example, in trimers $A \cdots B \cdots C$, fragments A and B can be coupled by $\hat{\mathcal{P}}^{AB}$ while B and C are coupled by $\hat{V}_{BC}$, but such terms are neglected in the present formulation of XPS. The impact of neglecting these terms is unclear, although such terms are only likely to be important when all three fragments are in very close proximity. In future work, we plan to reformulate our XPS procedure in a more rigorous and systematic manner based on Eqs. (29)–(31), which includes coupling between first-order perturbations and also all single-exchange terms.

At present, we have provided an *ad hoc* physical motivation for the XPS method, rather than a rigorous derivation starting from Eq. (29). For this reason, we refer to XPS as a *parameter-free* quantum-chemical model, rather than an *ab initio* model. Our aim for the time being is to replace the Lennard-Jones terms used in the XPol calculations of Xie *et al.*[25–28] with pairwise SAPT corrections. To this end, we have assumed that the leading many-body effect is induction, and we incorporate this effect within the zeroth-order Hamiltonian in an efficient (albeit approximate) manner. We assume that the remaining induction corrections, as well as all intermolecular exchange and dispersion interactions, can be described in a pairwise fashion. To the extent that the method is successful, it succeeds by reducing the size of the many-body perturbation, to the point where low-order, pairwise perturbation theory provides sufficient accuracy.

The inclusion of many-body induction within the zeroth-order Hamiltonian makes the subsequent SAPT corrections less meaningful in terms of energy decomposition analysis. For instance, the first-order electrostatic correction in XPS is *not* the total electrostatic energy, since the former corrects for errors in the approximate electrostatic treatment at zeroth order (i.e., the electrostatic embedding). The dispersion correction may be less contaminated, since all of the XPS modifications to the traditional SAPT perturbation are one-electron operators [see Eq. (27)], and therefore the pairwise dispersion correction differs from its traditional SAPT analogue only insofar as the MOs are perturbed by the electrostatic embedding. As such, we will continue to interpret this as a true dispersion correction.

Finally, some discussion of basis sets is warranted. Typically, SAPT calculations are performed in the so-called dimer-centered basis set (DCBS),[76] which means that the combined $A + B$ basis set is used to calculate the zeroth-order wavefunctions for both A and B. This leads to the unusual situation that there are more MOs than basis functions: one set of occupied and virtual MOs for each monomer, both expanded in the same (dimer) AO basis. As an alternative to the DCBS, one might calculate $|\Psi_A\rangle$ using only A's basis functions (similarly for B), in which case the SAPT calculation is said to employ the monomer-centered basis set (MCBS).[76] Because XPol derives its efficiency by restricting MOs on fragment A to be built from AOs on fragment A, we consider only the MCBS for the purpose of converging the fragment wavefunctions. (In Sec. IV B, however, we will introduce a post-XPol pseudocanonicalization in the DCBS, in order to recover intermolecular charge transfer.) Use of the MCBS means that our SAPT(0) corrections are most likely *not* converged with respect to basis-set expansion,[76] and thus rely on a cancellation of errors to provide meaningful results. As stated above, we intend XPS not as a benchmark method but rather as an efficient, parameter-free method to study molecular clusters and liquids. As such, we see no problem with relying on error cancellation, provided that the accuracy and robustness

of the cancellation are established by thorough comparisons to benchmark calculations. Such comparisons are reported in Secs. IV and V.

III. COMPUTATIONAL DETAILS

We have implemented the XPol, SAPT(0), and XPS methods within a locally modified version of the Q-CHEM software package.[77] The XPol SCF equations are solved by means of a dual SCF procedure similar to that described by Xie *et al.*[26] In brief, we iterate Eq. (6) to convergence for each fragment A, using a fixed set of point charges derived from the previous SCF solutions for the other fragments. Once all N_{frag} sets of SCF equations have been converged in this manner, we evaluate the supersystem SCF error as the average of fragment errors, each of which is defined as the root mean square of the off-diagonal Fock matrix elements in the ALMO basis. The dual SCF is considered to be converged when this error is $<10^{-8}$ a.u. The integral threshold is set to 10^{-14} a.u. for all calculations reported here, and all SAPT calculations employ Cartesian Gaussian basis functions.

To generate the CHELPG charges, we evaluate the electrostatic potential on a cubic grid, with a grid spacing of 1.0 Å. We discard grid points that lie within the van der Waals radius of any nucleus, using the van der Waals radii suggested by Bondi.[78,79] The grid edges extend 3.0 Å from the nearest atomic surface, as defined by these radii.

As we have not yet implemented the analytic gradient for the XPS method, geometry optimizations are performed using a three-point finite difference of the total energy, with atomic displacements of 10^{-3} bohr. For calculations using CHELPG charges,[67] the number of grid points used to evaluate the electrostatic potential may change as the nuclei are displaced. To avoid discontinuities in the potential energy surface, we therefore use a modified CHELPG procedure based upon a weighted least-squares fit to the electrostatic potential. Details of this procedure can be found in the supplementary material.[68]

In this report, we use the standard density functionals B3LYP,[80,81] BOP,[82] and PBE0.[83–85] We also employ several "long-range corrected" (LRC) functionals including LRC-ωPBEh[86,87] and LRC-μBOP.[88,89] The LRC-μBOP functional uses the Coulomb attenuation parameter $\mu = 0.47$ bohr^{-1} recommended in Ref. 89, while the LRC-ωPBEh functional uses the parameters recommended in Ref. 87 ($\omega = 0.2$ bohr^{-1} and 20% short-range HF exchange).

IV. DIMER BENCHMARKS

In this section, we evaluate dimer binding energies predicted by the XPS method, in comparison to benchmark values. Cembran *et al.*[65] note that the XPol method is not intended to reproduce HF or DFT energies, but is instead intended as an efficient way to obtain energies and forces for simulations of macromolecules and liquids. Similarly, our XPS method is not intended to reproduce any particular model chemistry but rather to allow efficient, accurate, and parameter-free simulations of molecular systems. As such, we compare to dimer SAPT(0) results not with the expectation of reproducing SAPT(0) binding energies exactly, but simply to demonstrate that our procedure does not significantly degrade the results of a method that is known to perform reasonably well for dimer binding energies. We also compare to complete basis set (CBS) extrapolations of binding energies computed at the MP2 level, and at the coupled-cluster level with single, double, and perturbative triple excitations [CCSD(T)].

The use of large basis sets is not consistent with our goal of fast quantum chemistry, so for XPS we consider only double-ζ and a few Pople-type triple-ζ basis sets. In a sense, one may think of the basis set and point-charge embedding scheme as parameters of the method.

A. S22 database

The S22 database of dimer binding energies was assembled in Ref. 90, although we use the revised binding energies from Ref. 91 in this work. The dataset contains 22 biologically relevant molecular dimers, including seven hydrogen-bonded dimers, eight dispersion-dominated complexes, and seven complexes where both dispersion and hydrogen-bonding contribute significantly to the binding energy. The benchmark binding energies for these complexes range from ~0.5–20.0 kcal/mol, and are estimates of CCSD(T)/CBS binding energies. (See Ref. 91 for details of the CBS extrapolation procedure.)

We will compare SAPT(0) and SAPT(KS) binding energies with and without inclusion of the XPol procedure. When XPol is used to obtain the zeroth-order wavefunction, and the perturbation is therefore modified according to Eq. (27), we will refer to these methods as XPS(0) and XPS(KS). We have neglected the $\delta E_{\text{int}}^{\text{HF}}$ correction, as the need for a counterpoise-corrected supersystem binding energy makes this an unattractive option for large systems. We will use the notation SAPT(PBE0)/6-31G*, for example, to indicate a SAPT(KS) calculation using the PBE0 density functional and the 6-31G* basis set, and the notation XPS(PBE0)/6-31G* to denote the corresponding XPS(KS) method. For XPS calculations, the additional symbol "*resp*" will be used to indicate that we have replaced the second-order induction amplitudes with amplitudes obtained from the solution of coupled-perturbed HF or KS equations.[51,75]

Table I shows errors for the SAPT(0), SAPT(KS), XPS(0), and XPS(KS) methods, using Mulliken charges for the XPS calculations. (Additional results, for a wider variety of basis sets and density functionals, can be found in the supplementary data.[68]) We note that XPS(0) generally outperforms XPS(KS), and among variants of the latter, the LRC functionals generally outperform their uncorrected counterparts. The XPS(0) procedure typically results in slightly smaller errors than SAPT(0), but this is not always the case. The inclusion of infinite-order induction (*resp*) universally improves the results when using XPS(0) and SAPT(0), but sometimes degrades the results of SAPT(KS) calculations, especially for the non-LRC functionals.

The fact that traditional generalized gradient approximations (GGAs), such as BOP, and global hybrid functionals, such as B3LYP and PBE0, fare poorly in SAPT(KS)

TABLE I. Mean absolute errors and (in parentheses) maximum absolute errors for the S22 database, in kcal/mol. A variety of SAPT(X) and XPS(X) variants are considered; note that SAPT(HF) is equivalent to the method that is traditionally called SAPT(0). All XPS methods use Mulliken charges for the electrostatic embedding.

	SCF method, X=				
Method	HF	BOP	PBE0	LRC-μBOP	LRC-ωPBEh
SAPT(X)/6-31G*	0.79 (3.27)	2.02 (9.55)	1.52 (6.47)	1.56 (7.31)	1.45 (6.75)
SAPT(X)-*resp*/6-31G*	0.65 (2.48)	2.50 (9.21)	1.73 (6.50)	1.52 (7.33)	1.51 (7.36)
XPS(X)/6-31G*	0.56 (1.46)	2.00 (9.66)	1.33 (6.24)	1.10 (3.79)	0.99 (3.57)
XPS(X)-*resp*/6-31G*	0.90 (3.16)	1.87 (9.01)	1.28 (5.99)	0.80 (3.14)	0.82 (3.11)
SAPT(X)/cc-pVDZ	1.06 (4.45)	2.45 (9.58)	1.83 (6.27)	1.89 (8.30)	1.81 (7.76)
SAPT(X)-*resp*/cc-pVDZ	0.91 (3.72)	2.92 (9.41)	2.06 (7.46)	1.87 (8.34)	1.88 (8.37)
XPS(X)/cc-pVDZ	0.71 (1.94)	2.58 (10.11)	1.73 (6.81)	1.62 (5.93)	1.59 (5.83)
XPS(X)-*resp*/cc-pVDZ	0.50 (1.81)	2.52 (9.00)	1.66 (5.90)	1.38 (5.43)	1.42 (5.45)

calculations is well-documented.[53,54,56–58] In particular, SAPT(KS) has been shown to overestimate dispersion energies, owing to highest occupied/lowest unoccupied MO (HOMO/LUMO) gaps that are too small.[53] Figure 1 shows the S22 database errors for SAPT(KS) and XPS(KS) methods in more detail, and the same trend is evident among traditional (i.e., non-LRC) density functionals. LRC functionals, however, widen the HOMO/LUMO gap,[88] leading to larger

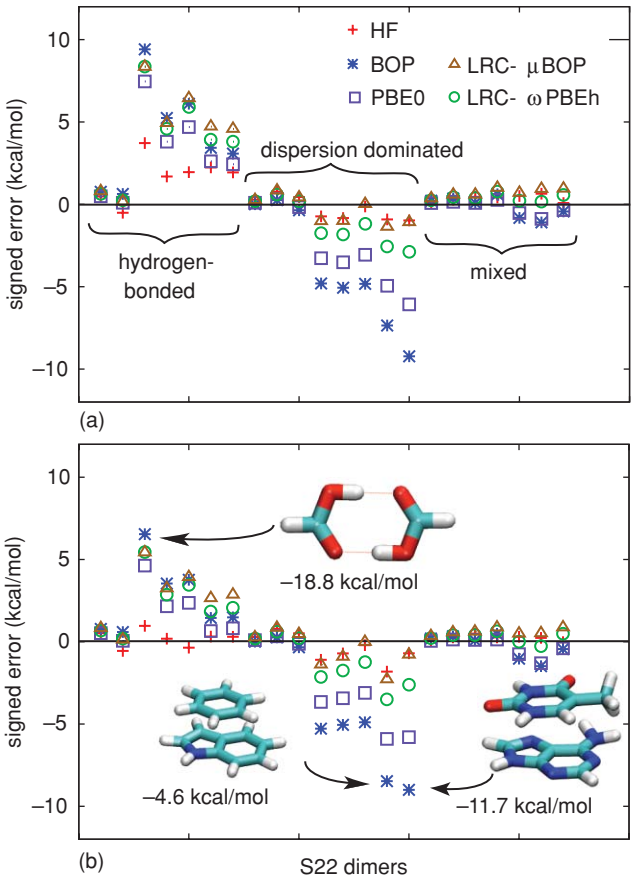

FIG. 1. Binding energy errors (in kcal/mol) across the S22 database, as computed at (a) the SAPT(X)/cc-pVDZ level and (b) the XPS(X)/cc-pVDZ level (with Mulliken embedding charges), for a variety of different SCF methods, X. A few difficult cases are highlighted in panel (b). Starting at the top and moving clockwise around panel (b), these are formic acid dimer, an indole–benzene π stack, and an adenine–thymine π stack.

energy denominators in Eq. (23), and therefore smaller dispersion energies. Although the LRC procedure appears to eliminate the strong overestimation of the dispersion energy, *all* of the density functionals tend to underbind the hydrogen-bonded complexes, often to a considerable extent.

The LRC correction scheme that is employed here differs from the asymptotic correction[92,93] (AC) that is typically used in SAPT(DFT) and SAPT(KS) calculations.[54,58] Although the traditional AC guarantees that the KS potential has the correct asymptotic behavior,[92,93] it has two major drawbacks that, in our view, make it unappealing for XPS calculations. First, construction of the AC requires an accurate ionization potential as input, and second, the AC-KS potential does not correspond to the functional derivative of any known energy functional.[92,93] (In practice, the AC is applied to correct the asymptotic behavior of the potential, but the energy functional is not modified.) These may not be serious problems for SAPT(KS) calculations, where the binding energy is obtained directly and the total system energy is not needed. However, in XPS some of the interaction energy is wrapped up in the zeroth-order XPol energy. We aim to use this method as an efficient way to perform *ab initio* molecular dynamics simulations in clusters and condensed-phase systems, where the total supersystem energy is obviously an important and meaningful quantity. That said, in the context of SAPT(KS) or SAPT(DFT) calculations, it is unclear to us whether the use of LRC functionals is superior to the AC procedure.

Looking carefully at the errors listed in Table I, it may seem strange that the SAPT(0) and XPS(0) approaches are not equivalent for hydrogen-bonded complexes when the infinite-order induction correction is applied. In fact, XPS(0) outperforms SAPT(0) in many of these cases. This is because the response correction is an infinite-order (nonperturbative) correction for induction *in the presence of a frozen partner density*, whereas XPS treats polarization self-consistently, if not exactly. This means that the XPS procedure attempts to include higher order induction effects into the zeroth-order energy. In SAPT(0) calculations, such terms are typically incorporated by means of the $\delta E_{\rm int}^{\rm HF}$ correction, which is omitted here.

Although these initial tests of XPS appear promising, we are generally unable to converge the XPol procedure when

TABLE II. Mean absolute errors and (in parentheses) maximum absolute errors for the S22 database, in kcal/mol. A variety of XPS(X) variants are considered, using either Löwdin or CHELPG embedding charges. For several of the basis sets, the corresponding response (*resp*) result is also listed. The primed and projected ("proj") basis sets are defined in the text (Sec. IV B). MP2 results, with and without counterpoise correction, are also listed for comparison.

	XPS-Löwdin			XPS-CHELPG				
Basis	HF	LRC-μBOP	LRC-ωPBEh	HF	LRC-μBOP	LRC-ωPBEh	MP2	MP2 (counterpoise)
6-31G*	0.73 (2.31)	0.91 (3.91)	0.93 (3.86)	0.87 (2.47)	0.87 (3.56)	0.90 (3.53)	1.44 (3.34)	2.09 (4.47)
6-31G*, *resp*	0.75 (2.40)	0.90 (3.88)	0.93 (3.88)	0.87 (2.55)	0.87 (3.54)	0.91 (3.55)	...	...
6-311G*	0.57 (2.76)	1.20 (6.52)	1.38 (6.92)	0.56 (2.32)	1.21 (6.20)	1.40 (6.67)	1.54 (4.86)	1.82 (5.81)
6-311G*, *resp*	0.54 (2.55)	1.18 (6.41)	1.37 (6.88)	0.54 (2.18)	1.20 (6.15)	1.40 (6.65)	...	...
cc-pVDZ	0.55 (2.04)	1.46 (6.39)	1.48 (6.30)	0.39 (1.12)	1.35 (5.72)	1.39 (5.68)	1.68 (4.75)	1.98 (4.73)
cc-pVDZ, *resp*	0.51 (1.75)	1.44 (6.26)	1.47 (6.27)	0.38 (1.02)	1.35 (5.70)	1.39 (5.69)	...	...
aug-cc-pVDZ	1.52 (4.48)	2.46 (10.24)	2.71 (9.63)	1.26 (3.38)	2.21 (8.43)	2.49 (7.91)	3.15 (11.15)	1.00 (2.80)
aug-cc-pVDZ, *resp*	1.40 (3.80)	2.39 (9.91)	2.67 (9.52)	1.25 (3.39)	2.23 (8.40)	2.49 (7.91)	...	...
aug-cc-pVDZ′	...	...	...	1.31 (3.86)	...	...	1.76 (2.64)	1.02 (2.96)
aug-cc-pVDZ-proj	...	...	...	1.31 (4.42)	1.66 (4.36)	2.05 (6.51)	...	...
aug-cc-pVDZ′-proj	...	...	...	0.75 (3.38)	...	...	...	...

Mulliken embedding charges are used in conjunction with diffuse basis functions. This is perhaps not entirely surprising, given the well-known instability of Mulliken charges with respect to basis-set expansion. (In their work on XPol, Xie *et al.*[25–28] use Mulliken charges exclusively, but have only reported calculations in small basis sets.) For small, compact basis sets, we find that Mulliken, Löwdin, and CHELPG embedding charges all perform similarly, and henceforth we discontinue the use of Mulliken charges in favor of these other two charge schemes.

Having demonstrated that the LRC functionals are superior to traditional GGAs and global hybrid functionals for use with XPS, all remaining calculations focus on HF, LRC-μBOP, and LRC-ωPBEh. Table II displays XPS statistical errors, evaluated over the S22 database, for a larger range of basis sets and charge schemes. (Results for some additional basis sets can be found in the supplementary material.[68]) Also listed in Table II are statistical errors for binding energies computed using standard supersystem MP2 calculations, both with and without counterpoise correction.

The XPS methods are generally more accurate than MP2 in the smaller basis sets, whereas counterpoise-corrected MP2/aug-cc-pVDZ and MP2/aug-cc-pVDZ′ results both exhibit mean errors of 1.0 kcal/mol, versus a mean error of 1.3 kcal/mol for XPS(0)-CHELPG calculations in the same basis sets. For the MP2 calculations, however, counterpoise correction is essential in order to obtain errors this low, whereas this correction is unnecessary in XPS calculations. This represents a significant advantage in the context of larger clusters.

As in the case of Mulliken embedding charges, XPS(0) outperforms XPS(KS) for the S22 database. Oddly, inclusion of diffuse basis functions slightly degrades the performance of XPS(0) results but greatly degrades XPS(KS) results. The compact basis sets perform quite well, and the smallest errors are obtained using the cc-pVDZ basis set, CHELPG charges, and HF orbitals. This combination affords a mean unsigned error of only 0.4 kcal/mol and a maximum error of 1.0 kcal/mol. It is interesting to note that the infinite-order induction correction has very little effect on the errors when CHELPG embedding charges are used. We take this as an indication that these charges better reproduce the electrostatic potential outside of the molecular core, which is precisely what CHELPG charges are designed to do.

The large errors observed at the XPS(KS) level, particularly where diffuse basis sets are employed, result from an underestimation of the binding energies in strongly H-bonded complexes. (The S22 dataset contains five H-bonded complexes whose binding energies exceed 15 kcal/mol in magnitude.) In particular, the XPS binding energy of formic acid dimer is almost always underestimated, and often this species affords the largest error. This was also observed by Hohenstein and Sherrill,[94] in both SAPT(0) calculations as well as SAPT calculations that include intramonomer electron correlation. These authors suggest that it is "imperative" to include the $\delta E_{\rm int}^{\rm HF}$ correction for H-bonded complexes. However, it appears that the XPol procedure recovers some of the higher order induction effects that the $\delta E_{\rm int}^{\rm HF}$ correction is intended to incorporate. In fact, the errors reported here are competitive with those reported in Ref. 94, where intramonomer electron correlation was included and the aug-cc-pVDZ DCBS was used.

Hohenstein and Sherrill[94] report that intramonomer correlation is especially important in the formic acid dimer. It is therefore curious that the XPS(KS) errors are much larger for this species than are the XPS(0) errors, given that XPS(KS) includes some intramolecular electron correlation whereas XPS(0) does not. Where does XPS(KS) go wrong? Rephrasing this question: if we assume that XPS(0) is doing something *right*, then what is so different about XPS(KS)?

To answer this question, we define the total Coulomb and exchange energies for the $A\cdots B$ dimer according to

$$E_{\rm Coul} = E_0^{AB} + E_{\rm elst}^{(1)} + E_{\rm ind}^{(2)} - E_0^A - E_0^B \tag{32}$$

and

$$E_{\rm exch} = E_{\rm exch}^{(1)} + E_{\rm exch\text{-}ind}^{(2)}, \tag{33}$$

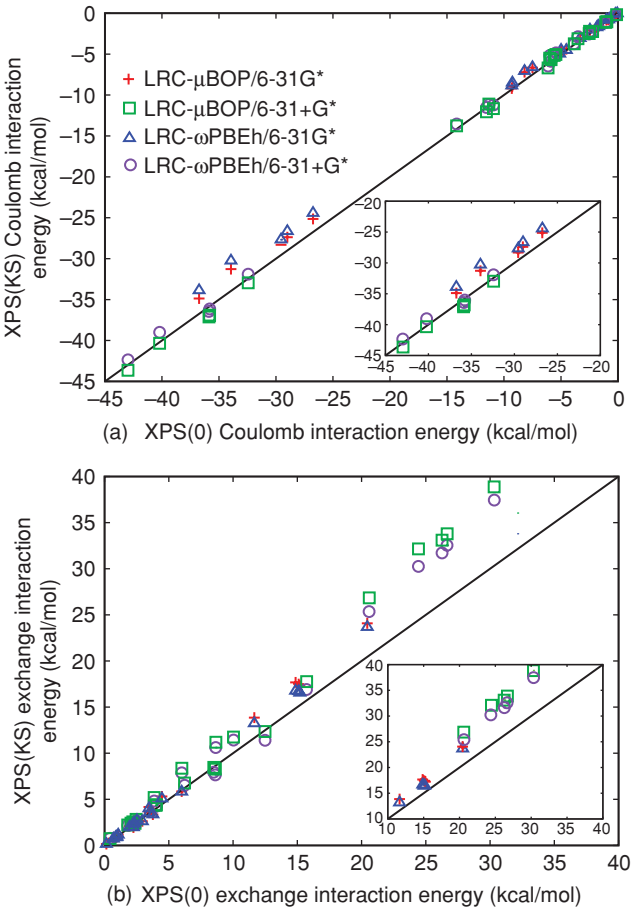

FIG. 2. Comparison of (a) the total Coulomb and (b) the total exchange interaction energies computed at the XPS(0) and XPS(KS) levels, for the S22 database. CHELPG charges are used in each case. The insets present the data for the five strongly H-bonded complexes whose binding energies exceed 15 kcal/mol.

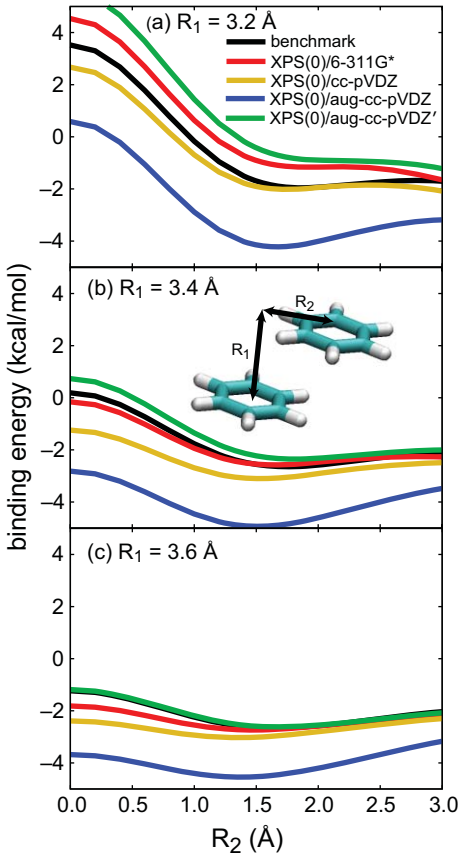

FIG. 3. Binding energy curves for the parallel-displaced benzene dimer, computed at the XPS(0) level using CHELPG charges. CCSD(T)/CBS benchmarks are taken from Ref. 96.

respectively, where E_0^{AB}, E_0^A, and E_0^B are the zeroth-order energies of the dimer and the two monomers. Figure 2 shows that the XPS(0) and XPS(KS) methods predict nearly identical Coulomb energies, but that the XPS(KS) methods predict much larger exchange energies, especially for dimers that exhibit strong hydrogen bonding. The net result is a less favorable cancellation of errors, and therefore an underestimation of the binding energies, when XPS(KS) methods are applied to these complexes. As the quality of the basis set increases, we expect the total electrostatic energy to become more negative while the exchange energy will become more positive (since an increasingly diffuse basis set will allow the monomer wavefunctions to overlap to a larger extent). When HF orbitals are used, errors in the Coulomb and exchange energies due to basis incompleteness must cancel, approximately, as relatively accurate binding energies are obtained. This is not the case when KS orbitals are used. It is tempting to attribute this to the well-known "delocalization error" in DFT,[95] which might exaggerate the degree of overlap and therefore the exchange energy. This artifact would tend to cancel out in SAPT(DFT) calculations, if the $\delta E_{\rm int}^{\rm HF}$ correction were used. For this reason, it may come as no surprise that the HF variant of XPS outperforms KS variants.

B. Potential energy curves

The results of Sec. IV A demonstrate that XPS(0) calculations, with a suitable choice of basis set and point-charge embedding scheme, can approach the accuracy of complete-basis CCSD(T) benchmarks for the S22 database of dimers. The best results are obtained using the cc-pVDZ basis set, which is far from complete, hence our method must benefit from some cancellation of errors. It is important to understand whether that cancellation is robust across potential energy surfaces. In this section, we examine some potential energy curves for $(H_2O)_2$ and $(C_6H_6)_2$. The benzene dimer was selected because it is a stringent test of the accuracy of dispersion interactions, and because CCSD(T)/CBS potential energy curves are available.[96] The water dimer was chosen because in Sec. V we will examine the performance of XPS(0) for binding energies in larger water clusters.

Figure 3 shows binding energy curves for the "parallel-displaced" benzene dimer at three different values of R_1, the distance between the two C_6H_6 planes. For nonpolar molecules, Löwdin and CHELPG charges produce nearly identical results, so only the latter are used here. We observe that 6-311G*, cc-pVDZ, and aug-cc-pVDZ qualitatively capture the profile of the binding energy curves (Fig. 3), but the 3-21G* and 6-31G* basis sets do not (see the supplementary material[68]). The latter basis sets exhibit too small of a barrier

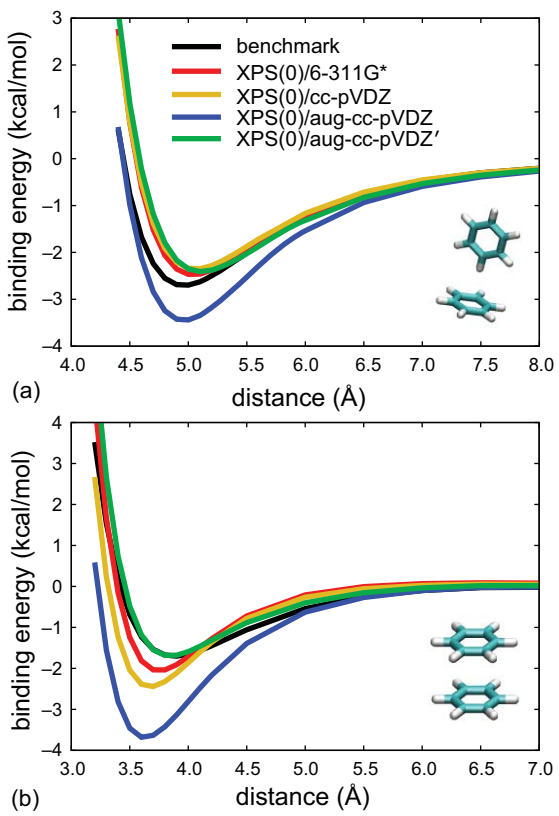

FIG. 4. Binding energy curves for (a) the T-shaped and (b) the sandwich isomer of the benzene dimer. Benchmark CCSD(T)/CBS values are taken from Ref. 96. The distance coordinate in both panels is the center-to-center distance between the benzene rings.

at $R_2 = 0$. In our view, the cc-pVDZ and 6-311G* basis sets exhibit acceptable errors (∼1 kcal/mol).

The largest basis set that we examine, aug-cc-pVDZ, overbinds the benzene dimer, which is not surprising given that XPS employs an MP2-like dispersion formula, and the MP2 method is known to overestimate the interaction energy of dispersion-bound complexes.[96] The dispersion energy in SAPT(0) generally increases as the size of the basis increases,[76] leading to a fortuitous cancellation of errors in small- to medium-sized basis sets, particularly ones that lack diffuse basis functions. For MP2-like methods, Hohenstein and Sherrill[94,97] recommend a modified form of aug-cc-pVDZ that they call aug-cc-pVDZ′, wherein the diffuse functions on hydrogen atoms are removed along with the diffuse *d* functions on the carbon atoms. For the parallel-displaced benzene dimer, we find that this basis leads to a remarkably good cancellation of errors, such that for $R_1 = 3.6$ Å the XPS(0) curve is indistinguishable from the benchmark.

A comparison of the potential energy curves for the three different values of R_1 shown in Fig. 3 suggests that the XPS exchange repulsion energy decays too rapidly with respect to monomer separation. This is more obvious in the case of the "T-shaped" and "sandwich" isomers of $(C_6H_6)_2$, potential energy curves for which are shown in Fig. 4. The aug-cc-pVDZ basis substantially overbinds these isomers at their minimum-energy geometries, although the cc-pVDZ and 6-311G* basis sets perform fairly well, affording errors of the order of ∼1 kcal/mol at the minimum. The aug-cc-pVDZ′ basis set performs extremely well for the sandwich dimer but slightly worse for the T-shaped dimer. For the sandwich configuration all of the basis sets afford a minimum at shorter separations than the benchmark result, except for aug-cc-pVDZ′, where the entire potential curve is quite accurate. With the exception of XPS(0)/aug-cc-pVDZ′, all of the XPS(0) methods perform better for the T-shaped isomer than they do for the sandwich conformation, which probably results from underestimating the induction interactions while simultaneously overestimating the dispersion energy.

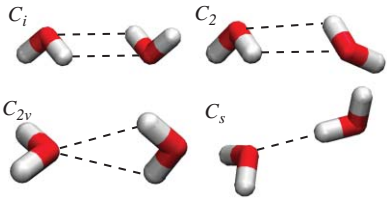

FIG. 5. $(H_2O)_2$ structures consider in this work.

To explore the $(H_2O)_2$ potential surface, we follow Burnham and Xantheas[98] in examining four different $(H_2O)_2$ isomers with distinct point-group symmetries. These are pictured in Fig. 5. We investigate minimum energy paths (MEPs) along the oxygen–oxygen distance coordinate, relaxing the other degrees of freedom subject to the constraint that point-group symmetry is maintained. To obtain benchmark MEPs, we optimized the geometries at the MP2/aug-cc-pVTZ level and then used counterpoise-corrected MP2/aug-cc-pVXZ calculations (X = D, T, Q) to estimate the MP2/CBS binding energy. The HF energy was extrapolated using the three-point *ansatz*

$$E(X) = E(\infty) + ae^{-bX}, \tag{34}$$

where a and b are fitting parameters.[99] The correlation energy was extrapolated using a two-point formula (X = T, Q),

$$E(\infty) = E(X) + cX^{-3}, \tag{35}$$

where c is a fitting parameter.[100]

Figure 6 compares these MP2/CBS benchmark MEPs to XPS(0)-CHELPG results, using a variety of basis sets. The salient features of the MP2 benchmarks that we would like to capture with XPS are

(i) the global minimum along the C_s curve at $R_{O-O} \approx 2.9$ Å;
(ii) a C_{2v} curve with a minimum at the same value of R_{O-O} but higher in energy by ≈2 kcal/mol;
(iii) C_2 and C_i curves having minima at $R_{O-O} \approx 2.75$ Å located about 1 kcal/mol above the C_s minimum; and
(iv) C_2 and C_i curves that coalesce at $R_{O-O} \approx 2.5$ Å, which results from a collapse to C_{2h} symmetry.[98]

Panels (b)–(e) of Fig. 6 show that it is relatively easy to obtain features (i), (iii), and (iv), even if the binding energies are not in agreement with the benchmark values. However, feature (ii) is reproduced only if we use an augmented basis set. With the aug-cc-pVDZ basis, the XPS(0) method

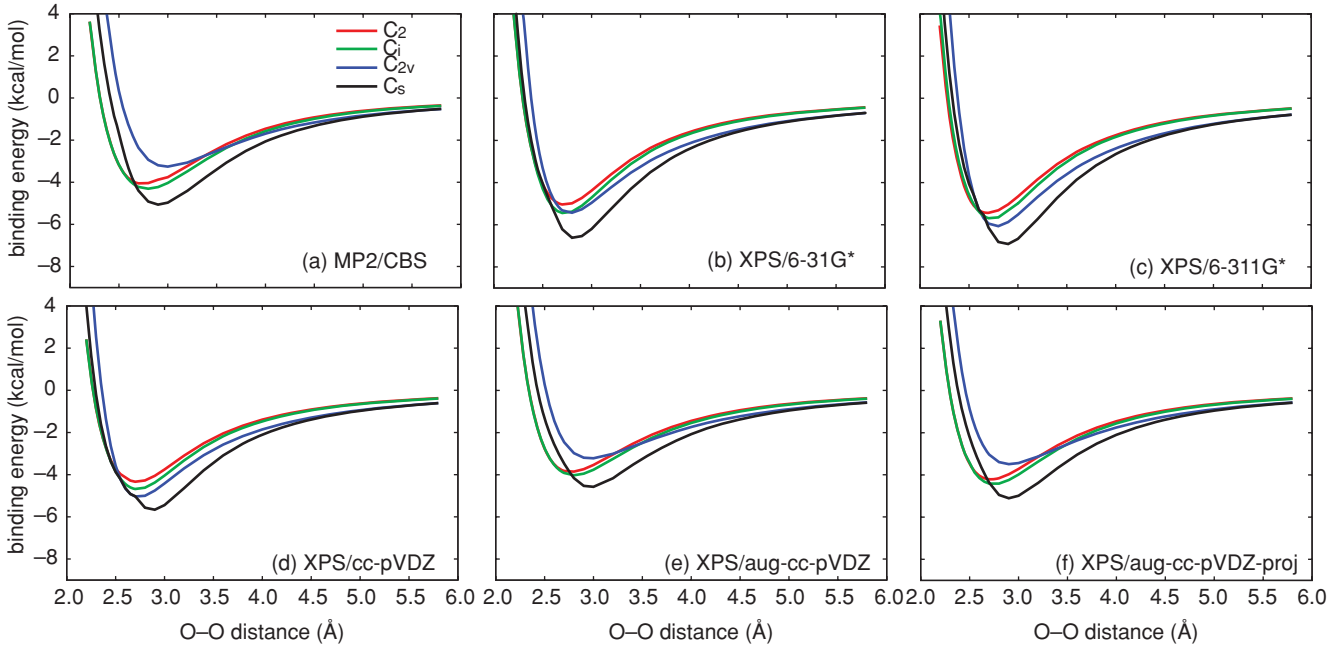

FIG. 6. Binding energy curves for the symmetry-distinct $(H_2O)_2$ isomers shown in Fig. 5: (a) MP2/CBS benchmarks and (b)–(f) XPS(0) results using a variety of basis sets. Löwdin and CHELPG embedding charges afford essentially identical potential curves, so only the latter are shown.

reproduces the relative energetics of the four MEPs quite well but the curves are slightly (≈0.5 kcal/mol) underbound at the C_s minimum, which is pushed out to about 3.0 Å. While the 0.5 kcal/mol underbinding represents only about a 10% error at the minimum, this will add up to fairly significant errors when applied to large water clusters.

Our XPS calculations use the MCBS, in which the MOs of fragment A are expanded in terms of only those AOs that are centered on atoms in fragment A. For this reason, one could argue that we have neglected charge-transfer interactions. (However, some charge-transfer-like interaction must certainly be present, since the basis functions on fragment A do extend over fragment B.) It would be useful to have a basis that mimics the DCBS that is often used in SAPT calculations for dimers,[76] but it is not clear how to generalize this idea to the case of more than two fragments.

As an alternative, we have utilized what we call the projected (proj) basis set, borrowing an idea from dual-basis MP2 calculations.[41,101] We first solve the XPol SCF equations and then, for a particular pairwise SAPT(0) correction, we construct the XPol Fock matrices for fragments A and B in the dimer $(A + B)$ basis set. We then separately diagonalize the occupied–occupied and virtual–virtual blocks of these matrices, which is sometimes called "pseudocanonicalization". This procedure leaves the fragment densities and zeroth-order fragment energies unchanged, but provides a larger set of virtual orbitals that extend over the partner fragment. We use this larger virtual space to perform the perturbative correction. Because the occupied–virtual block of the Fock matrix is nonzero, the pseudocanonical MOs are not rigorous eigenfunctions of the fragment Fock matrices. In principle, we could include a perturbative correction to the zeroth-order energies, of the form $\sum_{ar} F^A_{ar}/(\epsilon_a - \epsilon_r)$ for fragment A. (In the context of MP2-like methods, this is sometimes called the "non-Brillouin singles" term.[19]) We decline to do so, however, as this would have the effect of reintroducing BSSE. Instead, our aim is to enlarge the virtual space in a manner that can account for interfragment charge transfer.

As compared to SAPT(0) calculations performed with the DCBS, we find that the use of this "projected" basis set (aug-cc-pVDZ-proj) results in about a 10% error in the $(H_2O)_2$ binding energy. However, most of this error is contained in the $E^{(1)}_{\rm exch}$ correction, and the components involving virtual orbitals carry an error of <1%. Therefore, if the aim is to increase the virtual space without changing the zeroth-order density, then this is a successful strategy. If the first-order corrections are converged in the MCBS, then this procedure should incur very little additional error.

XPS(0) binding energy curves for $(H_2O)_2$, using the projected basis, are shown in Fig. 6(f). They are qualitatively similar to those in the aug-cc-pVDZ basis but now the binding energies are in a good agreement with MP2/CBS results. Interestingly, if we take the difference in the binding energies computed in the aug-cc-pVDZ MCBS and in the corresponding projected basis set as an estimate of the charge-transfer interaction energy, then charge transfer accounts for only about 10% of the $(H_2O)_2$ interaction energy, a figure that is substantially smaller than that estimated by energy decomposition analysis in the ALMO basis.[102] In Ref. 102, it was also reported that the charge-transfer component of the interaction energy is significantly larger when DFT is used to compute the ALMOs. In the context of the present work, this observation may indicate that the strong underbinding of hydrogen-bound complexes by XPS(KS) may be an artifact of the use of the MCBS. Use of the projected basis significantly decreases the XPS(KS) errors for the S22 database (see Table II). We plan to explore this issue further in future work.

V. WATER CLUSTERS

We have demonstrated that our method does not degrade the results of SAPT(0) for the S22 database, and furthermore that we can describe binding energy curves of benzene dimer with reasonable accuracy, and those of water dimer with high accuracy. Our ultimate goal, however, is application to larger clusters and molecular liquids. In this section, we evaluate the binding energies predicted by our method for a set of water clusters, in order to determine whether many-body effects are accurately reproduced by XPS.

We have assembled a database of 19 $(H_2O)_n$ isomers ranging from $n = 2$ to $n = 20$. Structures and benchmark binding energies for these clusters are taken from the work of Xantheas and co-workers.[103–108] The dataset includes MP2/CBS binding energies for the dimer; the cyclic trimer, tetramer, and pentamer; the ring, book, cage, and prism isomers of the hexamer; and the S_4 and D_{2d} isomers of the octamer. In addition, it includes binding energies for five different $(H_2O)_{11}$ isomers, computed at the MP2/aug-cc-pVQZ//MP2/aug-cc-pVTZ level. (Following Ref. 107, these isomers are labeled 43′4, 44′3′, 515, 551, and 44′12.) Finally, we include MP2/CBS binding energies for four $(H_2O)_{20}$ isomers,[106] one from each of the four families of low-lying minima (dodecahedron, fused cubes, face-sharing pentagonal prisms, and edge-sharing pentagonal prisms) exhibited by the 20-mer. Binding energies for all of the benchmarks are computed relative to relaxed monomers.

In larger clusters, a meaningful comparison of binding energies between different levels of theory should employ geometries that are optimized, separately, at either level of theory. For XPS calculations, geometries were optimized using a three-point finite-difference algorithm in Cartesian coordinates, and were considered to be converged when the change in energy dropped below 10^{-6} hartree. This procedure is quite demanding, computationally, and for the $(H_2O)_{20}$ clusters with the larger basis sets (aug-cc-pVDZ and aug-cc-pVDZ-proj), it was necessary to reduce the convergence threshold to 10^{-4} hartree. Tighter optimization would necessarily increase the binding energies, which (as will become clear in what follows) would improve the agreement between XPS results and benchmark binding energies. However, we expect that the binding energies would increase by not more than a few kcal/mol, for clusters whose binding energies are ∼200 kcal/mol. As such, we believe that these $(H_2O)_{20}$ tests are still meaningful.

Figure 7(a) shows the correlation between the XPS(0) and the benchmark binding energies. In general, the basis sets that were overbinding for $(H_2O)_2$ are also overbinding in larger clusters. In addition, Löwdin embedding tends to afford lower binding energies than CHELPG embedding, indicating that the Löwdin charge scheme underestimates the dipole moments of the H_2O monomers. As was the case for the S22 benchmarks, the cc-pVDZ basis set affords a superb cancellation of errors and yields results in a good agreement with the benchmark values.

Figure 7(b) plots the binding energy errors per hydrogen bond as a function of the number of hydrogen bonds. In all cases, this error grows rapidly from one to five hydrogen bonds, but beyond this it is nearly a constant with respect to

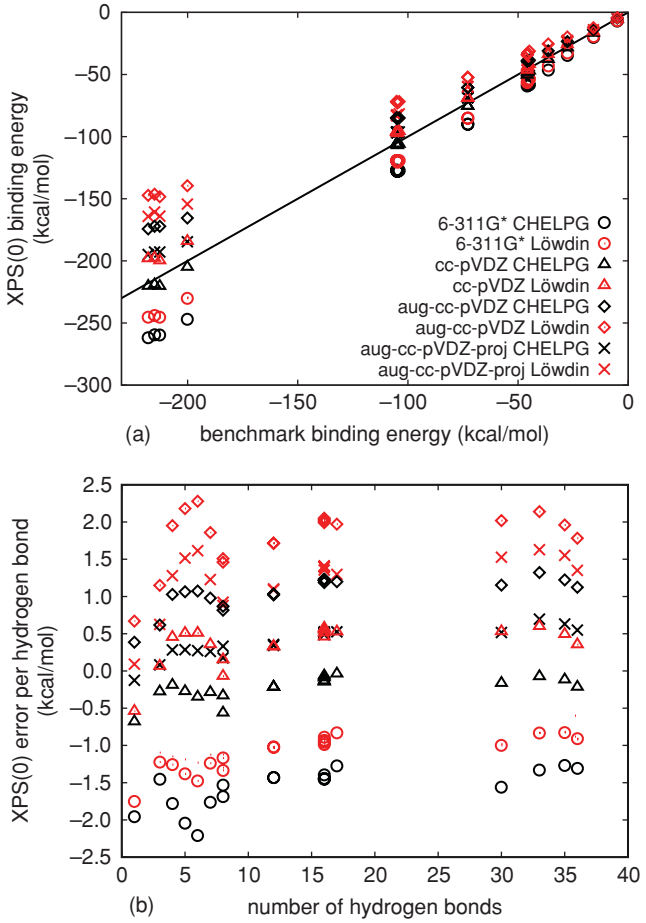

FIG. 7. (a) Correlation between XPS(0) binding energies and MP2 benchmarks. (b) XPS(0) error per hydrogen bond, as a function of the number of hydrogen bonds.

the number of hydrogen bonds. We interpret this as evidence that XPS(0) recovers a constant fraction of the many-body interaction energy in large water clusters.

Binding energies computed at the XPS(0)-CHELPG level, as well as percentage errors relative to MP2 benchmarks, are listed in Table III. The best results are obtained using the cc-pVDZ and aug-cc-pVDZ-proj basis sets. The cc-pVDZ, aug-cc-pVDZ, and aug-cc-pVDZ-proj basis sets all reproduce the correct energetic ordering of the $(H_2O)_6$ isomers, even though the error in the binding energy is greater than the energetic difference between these isomers. The same is not true for the octamers and endecamers, although the energy differences among these isomers amount to only about 1% of the total binding energies. It is difficult to compare the relative energies of the $(H_2O)_{20}$ isomers since our geometries are not fully relaxed, but in all cases the dodecahedron is correctly identified as the highest energy isomer.

As can be seen in Table III, the water dimer is overbound by 14% using the cc-pVDZ basis while the larger water clusters are overbound by ∼2%. In contrast, the aug-cc-pVDZ-proj basis accurately reproduces the dimer binding energy but underestimates the $(H_2O)_{20}$ binding energies by about 10%. Using the aug-cc-pVDZ basis set, the error grows from 8% at $n = 2$ to 20% at $n = 20$. We interpret these findings as an indication that, for a fixed pairwise error, our method recovers ∼90% of the interaction energy when used with

TABLE III. Negative binding energies for $(H_2O)_n$ cluster isomers, in kcal/mol. Percent errors in the XPS binding energies, relative to the benchmarks, are listed in parentheses.

		XPS(0)-CHELPG				
n	Isomer	6-311G*	cc-pVDZ	aug-cc-pVDZ	aug-cc-pVDZ-proj	Benchmark
2		6.9 (39.4)	5.6 (13.7)	4.6 (7.8)	5.1 (2.5)	4.97
3		20.2 (27.6)	16.6 (5.3)	14.0 (11.7)	15.6 (1.6)	15.82
4		34.8 (25.8)	28.4 (2.7)	23.5 (14.9)	26.5 (4.1)	27.63
5		46.5 (28.1)	37.7 (3.8)	31.0 (14.7)	34.9 (3.9)	36.31
6	Book	57.9 (27.0)	47.6 (4.4)	38.8 (15.0)	43.8 (4.0)	45.61
6	Cage	58.0 (26.7)	48.4 (5.8)	38.8 (15.2)	43.8 (4.2)	45.79
6	Cyclic	58.1 (29.6)	47.0 (4.7)	38.4 (14.3)	43.2 (3.7)	44.86
6	Prism	59.4 (29.4)	50.4 (9.8)	39.3 (14.2)	44.5 (3.1)	45.86
8	D_{2d}	90.1 (23.6)	75.5 (3.5)	60.5 (17.0)	68.6 (5.9)	72.88
8	S_4	90.0 (23.6)	75.4 (3.5)	60.6 (16.8)	68.7 (5.6)	72.83
11	$43'4$	126.8 (20.6)	105.8 (0.6)	84.8 (19.4)	96.2 (8.5)	105.16
11	$44'3'$	128.0 (22.2)	107.1 (2.2)	85.3 (18.6)	96.6 (7.8)	104.76
11	515	127.4 (21.2)	106.1 (1.0)	85.4 (18.8)	96.7 (8.0)	105.09
11	551	128.0 (22.0)	106.4 (1.4)	85.5 (18.5)	96.8 (7.8)	104.95
11	$44'12$	127.2 (22.4)	106.0 (2.0)	85.0 (18.3)	96.2 (7.5)	103.97
20	Dodecahedron	247.0 (23.4)	205.0 (2.5)	165.5 (17.3)	184.6 (7.7)	200.10
20	Edge-sharing	261.8 (20.1)	220.3 (1.1)	174.3 (20.0)	194.9 (10.6)	217.90
20	Face-sharing	259.4 (20.6)	219.1 (1.9)	172.2 (19.9)	192.9 (10.3)	215.00
20	Fused cubes	259.7 (22.1)	220.4 (3.6)	172.2 (19.0)	192.8 (9.3)	212.60

CHELPG embedding charges. A smaller fraction of the interaction energy is recovered using Löwdin charges, as shown in Table IV, and by a similar argument we conclude that the XPS(0)-Löwdin method recovers ∼80% of the interaction energy. By performing single-point energy calculations on the $(H_2O)_{20}$ isomers at the optimized XPS(0)-CHELPG geometries, we find that traditional pairwise SAPT(0) in the aug-cc-pVDZ-proj basis recovers ∼70% of the binding energy. This indicates that roughly 30% of the binding energy in these clusters comes from many-body effects, and we recover about 2/3 of this using XPS(0) with CHELPG embedding charges.

VI. COMPUTATIONAL EXPENSE

We intend XPS as a method for large systems, so let us comment on its computational scaling. The first step in an XPS calculation, solving the XPol SCF equations, scales linearly with N_{frag}, assuming that construction and diagonalization of the fragment Fock matrices is much more demanding than formation of the one-electron integrals needed to compute the electrostatic interactions between the fragment densities and the embedding charges. (Even the latter step can ultimately be made to scale linearly by exploiting fast-multipole techniques.[3]) Increasing the size of the basis set formally scales as $\mathcal{O}(N_{\text{basis}}^4)$ for Fock matrix construction and $\mathcal{O}(N_{\text{basis}}^3)$ for diagonalization.

In the second step of XPS, we perform $N_{\text{frag}}(N_{\text{frag}}-1)/2$ independent, pairwise SAPT(0) corrections, so this step scales as $\mathcal{O}(N_{\text{frag}}^2)$ and dominates the total cost in our present, serial implementation. Figure 8 shows actual timings for water clusters, as compared to timings for supersystem HF and MP2 calculations. Already in its present implementation, XPS can be scaled up to quite large water clusters.

Each pairwise correction in SAPT(0) and XPS(0) requires integrals of the form $(aX|bY)$, where $a \in A$ and $b \in B$ are occupied MOs, whereas X and Y range over all occupied and virtual MOs on both fragments. We compute these integrals by first computing all N_{basis}^4 AO integrals $(\mu\nu|\lambda\sigma)$, which represents some unnecessary overhead in the MCBS. The AO integrals are transformed in four steps that scale as $\mathcal{O}(N_o^A N_{\text{basis}}^4)$, $\mathcal{O}(N_o^A N_o^B N_{\text{basis}}^3)$, $\mathcal{O}(N_o^A N_o^B N_{MO}^{AB} N_{\text{basis}}^2)$, and $\mathcal{O}[N_o^A N_o^B (N_{MO}^{AB})^2 N_{\text{basis}}]$, where N_o^A is the number of occupied MOs on fragment A and N_{MO}^{AB} is the total number of MOs (occupied + virtual) on fragments A and B. The most time-consuming contraction step in the SAPT(0) correction is the accumulation of $E_{\text{exch-disp}}^{(2)}$, the bottleneck of which scales as $\mathcal{O}[(N_o^A)^2 N_o^B N_v^A N_v^B]$, where N_v^A is the number of virtual MOs

TABLE IV. Negative binding energies for $(H_2O)_{20}$ clusters, in kcal/mol. Percent errors in the XPS binding energies, relative to the benchmarks, are listed in parentheses.

	XPS(0)-Löwdin		XPS(0)-CHELPG		
Isomer	cc-pVDZ	aug-cc-pVDZ-proj	cc-pVDZ	aug-cc-pVDZ-proj	Benchmark
Dodecahedron	184.2 (7.9)	154.3 (22.9)	220.4 (3.6)	184.6 (7.7)	200.10
Edge-sharing	198.1 (9.1)	164.2 (24.7)	220.3 (1.1)	192.8 (9.3)	217.90
Face-sharing	197.8 (8.0)	160.7 (25.3)	219.1 (1.9)	192.9 (10.3)	215.00
Fused cubes	199.8 (6.0)	164.0 (22.9)	220.4 (3.6)	192.8 (9.3)	212.60

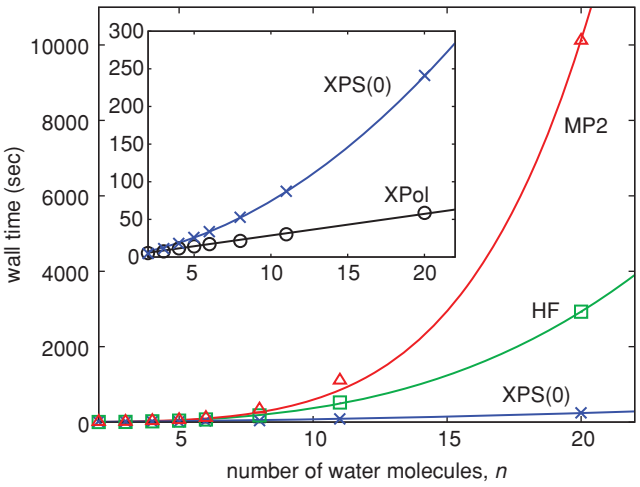

FIG. 8. Wall time required for XPS(0), HF, and MP2 calculations in $(H_2O)_n$ clusters. The inset shows the XPS(0) and XPol timings in more detail. The solid curves are fits to polynomials representing theoretical scalings: n^5 for MP2, n^3 for HF, n^2 for XPS(0), and n for XPol.

associated with fragment A. [As such, the scaling of this step is usually reported as $\mathcal{O}(o^3v^2)$.[56,58,97]] We find that the integral transformation is at least one order of magnitude more expensive than the contractions.

In practice, the SAPT(0) corrections that we use in XPS exhibit a scaling similar to that of dimer MP2 calculations, with respect to either the size of the fragments or the size of the basis set. Stand-alone dimer SAPT(0) calculations are in general less expensive than supersystem MP2 calculations, since the occupied and virtual spaces in SAPT(0) are partitioned into components from fragments A and B, leading to an overall computational expense of $\mathcal{O}(N_o^A N_{\text{basis}}^4)$ for SAPT(0) versus $\mathcal{O}[(N_o^A + N_o^B)N_{\text{basis}}^4]$ for MP2.[97]

The exchange interactions in SAPT(0) decay rapidly as a function of interfragment distance, which could be exploited to reduce the cost of large XPS calculations by introducing cutoff schemes, such that the exchange corrections are evaluated only for nearby fragments. Thresholds could also be used to avoid accumulating dispersion and induction corrections for distant pairs. At very long range, the electrostatic interactions included at the XPol level may be sufficiently accurate to avoid computing $E_{\text{elst}}^{(1)}$ altogether. Parallelization of the $N_{\text{frag}}(N_{\text{frag}} - 1)/2$ independent SAPT(0) corrections is another obvious way to reduce the cost. We plan to explore such cost-reduction techniques in the future.

VII. SUMMARY AND OUTLOOK

We have introduced a new quantum chemistry method for studying intermolecular interactions, which we call XPol/SAPT, or XPS. This method incorporates electronic induction, intermolecular electrostatic interactions, and intramolecular interactions at the SCF level, using a charge embedding scheme whose computational cost grows linearly with the number of monomers. Dispersion, exchange-repulsion, and intermolecular charge-transfer effects, along with corrections to the electrostatic charge embedding, are introduced by means of a pairwise, perturbative post-SCF correction. The monomers are allowed to be fully flexible.

In developing this method, our intention was to replace the need for Lennard-Jones parameters in the XPol procedure,[26] while preserving the favorable scaling of that method with respect to system size. We have demonstrated that the XPS method does not degrade, and in many cases improves upon, the results of traditional SAPT(0) calculations for molecular dimers. Given an appropriate choice of basis set and electrostatic embedding, XPS recovers ~90% of the binding energy of large water clusters, as compared to MP2/CBS benchmarks, whereas traditional pairwise SAPT(0) recovers ~70%. In our present implementation, the cost of the post-XPol corrections scales quadratically with the number of monomers. The computational cost is already quite low for large clusters, if the monomers are small, and can ultimately be made to scale linearly with the number of monomers, by introducing appropriate distance-dependent cutoffs. Work along these lines, including an implementation using periodic boundary conditions, is currently in progress.

While the XPS method is promising with respect to both accuracy and efficiency, many future improvements must be explored. The poor scaling [$\mathcal{O}(N^5)$] of the SAPT corrections with respect to fragment size can be improved by using density-fitting techniques that have previously been introduced in the context of traditional SAPT calculations.[56,58,94,97] A method for fragmenting the system across covalent bonds, such as that used in the original XPol method[25,28] or in the fragment MO method,[29–31] will be needed in order to handle large monomers. These developments are currently being explored in our group.

The XPS method is systematically improvable, which may help to further improve the accuracy. In particular, a more rigorous formulation of the method—which goes beyond the pairwise approximation—is possible, as outlined in Sec. II D, and work along these lines is in progress. It may also be possible to incorporate Casimir–Polder-type dispersion formulas, as currently used in SAPT(DFT),[56,58] in order to obtain better results when DFT is used to describe the monomers. At present, XPS results using KS orbitals are notably inferior to those obtained using HF orbitals.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation CAREER award (CHE-0748448). Calculations were performed at the Ohio Supercomputer Center under project no. PAS-0291. J.M.H. gratefully acknowledges a fellowship from the Alfred P. Sloan Foundation. L.D.J. acknowledges a Presidential Fellowship from The Ohio State University. We thank Sotiris Xantheas for providing coordinates for the water clusters studied herein.

[1] P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[2] S. Hirata, Phys. Chem. Chem. Phys. **11**, 8397 (2009).
[3] C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon, Chem. Phys. Lett. **230**, 8 (1994).
[4] S. Saebo, in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 2002), Vol. 7, pp. 63–87.

[5]C. Ochsenfeld, J. Kussmann, and D. S. Lambrecht, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari (Wiley-VCH, New York 2007), Vol. 23, Chap. 1, pp. 1–82.
[6]R. A. Friesner, Proc. Natl. Acad. Sci. U. S. A. **102**, 6648 (2005).
[7]E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).
[8]E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).
[9]E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **4**, 1 (2008).
[10]E. E. Dahlke, H. R. Leverentz, and D. G. Truhlar, J. Chem. Theory Comput. **4**, 33 (2008).
[11]G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).
[12]A. Sebetci and G. J. O. Beran, J. Chem. Theory Comput. **6**, 155 (2010).
[13]G. J. O. Beran and K. Nanda, J. Phys. Chem. Lett. **1**, 3480 (2010).
[14]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).
[15]D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).
[16]D. G. Fedorov and K. Kitaura, in *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, edited by D. G. Fedorov and K. Kitaura (CRC, Boca Rotan, FL, 2009), Chap. 2, pp. 5–36.
[17]W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[18]H. Stoll, G. Wagenblast, and H. Preuß, Theor. Chem. Acc. **57**, 169 (1980).
[19]R. Khaliullin, M. Head-Gordon, and A. T. Bell, J. Chem. Phys. **124**, 204105 (2006).
[20]N. Flocke and R. J. Bartlett, J. Chem. Phys. **121**, 10935 (2004).
[21]T. F. Hughes, N. Flocke, and R. J. Bartlett, J. Phys. Chem. A **112**, 5994 (2008).
[22]P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, J. Chem. Phys. **105**, 1968 (1996).
[23]M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, J. Phys. Chem. A **105**, 293 (2000).
[24]D. Ghosh, D. Kosenkov, V. Vanovschi, C. F. Williams, J. M. Herbert, M. S. Gordon, M. W. Schmidt, L. V. Slipchenko, and A. I. Krylov, J. Phys. Chem. A **114**, 12739 (2010).
[25]W. Xie and J. Gao, J. Chem. Theory Comput. **3**, 1890 (2007).
[26]W. Xie, L. Song, D. G. Truhlar, and J. Gao, J. Chem. Phys. **128**, 234108 (2008).
[27]L. Song, J. Han, Y.-L. Lin, W. Xie, and J. Gao, J. Phys. Chem. A **113**, 11656 (2009).
[28]W. Xie, M. Orozco, D. G. Truhlar, and J. Gao, J. Chem. Theory Comput. **5**, 459 (2009).
[29]T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **318**, 614 (2000).
[30]D. G. Fedorov, J. H. Jensen, R. C. Deka, and K. Kitaura, J. Phys. Chem. A **112**, 11808 (2008).
[31]Y. Komeiji, Y. Mochizuki, and T. Nakano, Chem. Phys. Lett. **484**, 380 (2010).
[32]S. Saebo and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[33]M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).
[34]M. Schütz and F. R. Manby, Phys. Chem. Chem. Phys. **5**, 3349 (2003).
[35]H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
[36]H.-J. Werner and K. Pflüger, Annu. Rep. Comp. Chem. **2**, 53 (2006).
[37]J. E. Subotnik and M. Head-Gordon, J. Phys.: Condens. Matter **20**, 294211 (2008).
[38]T. S. Chwee, A. B. Szilva, R. Lindh, and E. A. Carter, J. Chem. Phys. **128**, 224106 (2008).
[39]W. Li, P. Piecuch, J. R. Gour, and S. Li, J. Chem. Phys. **131**, 114109 (2009).
[40]W. Liang and M. Head-Gordon, J. Phys. Chem. A **108**, 3206 (2004).
[41]R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong, and M. Head-Gordon, J. Chem. Phys. **125**, 74108 (2006).
[42]F. Weigend, M. Häser, J. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[43]F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).
[44]A. Sodt, J. E. Subotnik, and M. Head-Gordon, J. Chem. Phys. **125**, 194109 (2006).
[45]T. B. Pedersen, F. Aquilante, and R. Lindh, Theor. Chem. Acc. **124**, 1 (2009).
[46]C. D. Sherrill, J. Chem. Phys. **132**, 110902 (2010).
[47]R. A. Christie and K. D. Jordan, in *Intermolecular Forces and Clusters II, Structure and Bonding*, Vol. 116, edited by D. Wales and R. A. Christie (Springer, New York, 2005), pp. 27–41.
[48]J. Cui, H. Liu, and K. D. Jordan, J. Phys. Chem. B **110**, 18872 (2006).
[49]D. G. Fedorov, P. V. Avramov, J. H. Jensen, and K. Kitaura, Chem. Phys. Lett. **477**, 169 (2009).
[50]T. Nagata, D. G. Fedorov, and K. Kitaura, Chem. Phys. Lett. **492**, 302 (2010).
[51]B. Jeziorski, R. Moszynski, A. Ratkiewicz, S. Rybak, K. Szalewicz, and H. L. Williams, in *Methods and Techniques in Computational Chemistry: METECC-94*, edited by E. Clementi (STEF, Cagliari, 1993), Vol. B, Chap. 3, pp. 79–129.
[52]B. Jeziorski, R. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994).
[53]H. L. Williams and C. F. Chabalowski, J. Phys. Chem. A **105**, 646 (2001).
[54]A. J. Misquitta and K. Szalewicz, Chem. Phys. Lett. **357**, 301 (2002).
[55]A. J. Misquitta, B. Jeziorski, and K. Szalewicz, Phys. Rev. Lett. **91**, 033201 (2003).
[56]A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **123**, 214103 (2005).
[57]A. Heßelmann and G. Jansen, Chem. Phys. Lett. **367**, 778 (2003).
[58]A. Heßelman, G. Jansen, and M. Schutz, J. Chem. Phys. **122**, 014103 (2005).
[59]V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **106**, 9668 (1997).
[60]V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **112**, 112 (2000).
[61]E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **133**, 104107 (2010).
[62]N. Turki, A. Milet, A. Rahmouni, O. Ouamerali, R. Moszynski, E. Kochanski, and P. E. S. Wormer, J. Chem. Phys. **109**, 7157 (1998).
[63]Y. Chen and H. Li, J. Phys. Chem. A **114**, 11719 (2010).
[64]J. Gao, J. Chem. Phys. **109**, 2346 (1998).
[65]A. Cembran, P. Bao, Y. Wang, L. Song, D. G. Truhlar, and J. Gao, J. Chem. Theory Comput. **6**, 2469 (2010).
[66]A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (Macmillan, New York, 1982).
[67]C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 361 (1990).
[68]See supplementary material at http://dx.doi.org/10.1063/1.3560026 for additional details of the methodology and additional benchmark results.
[69]K. Patkowski, K. Szalewicz, and B. Jeziorski, J. Chem. Phys. **125**, 154107 (2006).
[70]R. Moszynski, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **100**, 1312 (1994).
[71]B. Jeziorski, K. Szalewicz, and G. Chalasinski, Int. J. Quantum Chem. **14**, 271 (1978).
[72]T. Ćwiok, B. Jeziorski, W. Kołos, R. Moszynski, and K. Szalewicz, J. Chem. Phys. **97**, 7555 (1992).
[73]R. Moszynski, B. Jeziorski, S. Rybak, K. Szalewicz, and H. L. Williams, J. Chem. Phys. **100**, 5080 (1994).
[74]B. Jeziorski, B. Bulski, L. Piela, Int. J. Quantum Chem. **10**, 281 (1976).
[75]J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem. Symp. **13**, 225 (1979).
[76]H. L. Williams, E. M. Mas, and K. Szalewicz, J. Chem. Phys. **103**, 7374 (1995).
[77]Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **8**, 3172 (2006).
[78]A. Bondi, J. Phys. Chem. **68**, 441 (1964).
[79]R. S. Rowland and R. Taylor, J. Phys. Chem. **100**, 7384 (1996).
[80]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[81]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[82]T. Tsuneda, T. Suzumura, and K. Hirao, J. Chem. Phys. **110**, 10664 (1999).

[83]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[84]C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).
[85]M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999).
[86]T. M. Henderson, B. G. Janesko, and G. E. Scuseria, J. Chem. Phys. **128**, 194105 (2008).
[87]M. A. Rohrdanz, K. M. Martins, and J. M. Herbert, J. Chem. Phys. **130**, 54112 (2009).
[88]H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).
[89]J. Song, T. Hirosawa, T. Tsuneda, and K. Hirao, J. Chem. Phys. **126**, 154105 (2007).
[90]P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[91]T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **132**, 144104 (2010).
[92]D. J. Tozer and N. C. Handy, J. Chem. Phys. **109**, 10180 (1998).
[93]M. Grüning, V. Gritsenko, S. J. A. van Gisbergen, and E. J. Baerends, J. Chem. Phys. **114**, 652 (2001).
[94]E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **133**, 014101 (2010).
[95]A. J. Cohen, P. Mori-Sánchez, and W. Yang, Science **321**, 792 (2008).
[96]C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009).
[97]E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **132**, 184111 (2010).
[98]C. J. Burnham and S. S. Xantheas, J. Chem. Phys. **116**, 1479 (2002).
[99]A. Halkier, P. Jorgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).
[100]T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, New York, 2000).
[101]K. Wolinski and P. Pulay, J. Chem. Phys. **118**, 9497 (2003).
[102]R. Z. Khalliulin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon, J. Phys. Chem. A **111**, 8753 (2007).
[103]S. S. Xantheas, J. Chem. Phys. **104**, 8821 (1996).
[104]S. S. Xantheas, C. J. Burnham, and R. J. Harrison, J. Chem. Phys. **116**, 1493 (2002).
[105]S. S. Xantheas and E. Aprà, J. Chem. Phys. **120**, 823 (2004).
[106]G. S. Fanourgakis, E. Aprà, and S. S. Xantheas, J. Chem. Phys. **121**, 2655 (2004).
[107]S. Bulusu, S. Yoo, E. Aprà, S. S. Xantheas, and X. C. Zeng, J. Phys. Chem. A **110**, 11781 (2006).
[108]S. Yoo, E. Aprà, X. C. Zeng, and S. S. Xantheas, J. Phys. Chem. Lett. **1**, 3122 (2010).