THE JOURNAL OF CHEMICAL PHYSICS **139**, 224102 (2013)

Approaching the complete-basis limit with a truncated many-body expansion

Ryan M. Richard, Ka Un Lao, and John M. Herbert[a)]
Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

(Received 10 October 2013; accepted 15 November 2013; published online 9 December 2013)

High-accuracy electronic structure calculations with correlated wave functions demand the use of large basis sets and complete-basis extrapolation, but the accuracy of fragment-based quantum chemistry methods has most often been evaluated using double-ζ basis sets, with errors evaluated relative to a supersystem calculation using the same basis set. Here, we examine the convergence towards the basis-set limit of two- and three-body expansions of the energy, for water clusters and ion–water clusters, focusing on calculations at the level of second-order Møller-Plesset perturbation theory (MP2). Several different corrections for basis-set superposition error (BSSE), each consistent with a truncated many-body expansion, are examined as well. We present a careful analysis of how the interplay of errors (from all sources) influences the accuracy of the results. We conclude that fragment-based methods often benefit from error cancellation wherein BSSE offsets both incompleteness of the basis set as well as higher-order many-body effects that are neglected in a truncated many-body expansion. An n-body counterpoise correction facilitates smooth extrapolation to the MP2 basis-set limit, and at $n = 3$ affords accurate results while requiring calculations in subsystems no larger than trimers.

[http://dx.doi.org/10.1063/1.4836637]

I. INTRODUCTION

Two factors control the accuracy of *ab initio* calculations: the level of theory used to describe electron correlation, and the size of the one-particle basis set. Convergence with respect to the latter, which is the topic of the present work, is generally not an option and one must resort to extrapolation procedures to estimate the complete basis-set (CBS) limit. The fact that the basis set is not complete implies that the finite-basis solution to Schrödinger's equation lacks sufficient flexibility to model the true wave function, leading to *basis-set incompleteness error* (BSIE). It is worth noting that BSIE can sometimes compensate for inadequacies in the treatment of electron correlation; to unmask such errors, it is important to be able to perform CBS extrapolations for the systems of interest.

An additional complication arises for energy differences, namely, the presence of *basis-set superposition error* (BSSE). Consider the canonical case of computing a dimer ($A\cdots B$) binding energy. Here, the dimer (AB) basis set is larger and more flexible than either monomer basis set, hence a naïve calculation of the binding energy according to the equation

$$\Delta E = E_{AB} - E_A - E_B, \tag{1.1}$$

in any finite basis set, represents an unbalanced approximation. BSSE is a reflection of this fact. The familiar Boys-Bernardi counterpoise (CP) correction[1] amounts to adding extra basis functions in calculating E_A and E_B, in such a way as to correct for the unbalanced nature of the binding energy in Eq. (1.1).

BSSE and BSIE are intimately related,[2,3] which makes it difficult to disentangle them or quantitatively define either one.[4] Whereas Eq. (1.1) suggests that BSSE is solely an intermolecular phenomenon, *intramolecular* BSSE may exist as well, especially in larger molecules, and may affect the relative energies of, e.g., a compact versus an extended structure of a polypeptide.[5–10]

Unfortunately, the separation between inter- and intramolecular BSSE is somewhat arbitrary in a macromolecule, depending upon the semantics of defining boundaries between "molecular" units that sever covalent bonds, although intramolecular CP corrections along these lines have been proposed.[11–14] The present study sidesteps this issue by focusing on clusters of small molecules that should thus contain little intramolecular BSSE. We therefore make no attempt to correct for intramolecular BSSE except via extrapolation to the CBS limit.

The Boys-Bernardi CP correction and its many-body generalizations, which are discussed in Sec. II, provide a means to correct for finite-basis inadequacies. Many of these corrections, however, rely on a series of calculations in the full (supersystem) basis set, which can become burdensome for large clusters of molecules, or where a high level of theory is employed. For sufficiently large clusters, even non-BSSE-corrected calculations become intractable, and in such cases a popular avenue is fragment-based quantum chemistry. (See Ref. 15 for a recent review and Ref. 16 for additional discussion of how some of these methods are related to one another.) Fragment-based methods assume that supersystem properties are obtainable from isolated subsystem properties, allowing for a host of smaller calculations, that are embarrassingly parallelizable and can be combined to give an approximate value for the supersystem quantity of interest.

[a)] herbert@chemistry.ohio-state.edu

In the interest of obtaining high-level benchmarks using fragment-based methods, we wish to explore their use with correlated wave functions and large, correlation-consistent basis sets. To this end, two different BSSE corrections have been introduced that are compatible with a truncated (at order n) many-body expansion: the many-body counterpoise correction, MBCP(n),[17] and the Valiron-Mayer function counterpoise correction, VMFC(n).[18] A preliminary comparison of these two approaches was presented in Ref. 17, and the present study provides a more complete follow-up, with particular emphasis on extrapolation to the CBS limit.

At the outset, we note that several previous studies[17,19,20] have demonstrated that a two- or three-body expansion in a triple-ζ basis set is sufficient to obtain a high accuracy "triples" correction,

$$\Delta E_{\text{CCSD(T)}} = E_{\text{CCSD(T)}} - E_{\text{MP2}}. \tag{1.2}$$

Thus, by understanding how errors from a truncated n-body expansion (with or without CP corrections) compare to MP2/CBS results, we learn how such errors affect CCSD(T)/CBS results, where CCSD(T) denotes coupled-cluster theory including single and double excitations, with perturbative triple excitations.

The remainder of the paper is organized as follows. In Sec. II, we briefly recount the formalism underlying the truncated many-body expansion and, especially, the approximate CP-type corrections that are consistent with the n-body expansion. Then, in Sec. IV, we apply these methods to two systems: $(H_2O)_6$ and $(H_2O)_{10}F^-$.

II. THEORY

A. Many-body expansion

The many-body expansion (MBE) for a system of N monomers is given by

$$E = \sum_{I=1}^{N} E_I + \sum_{I<J} \Delta E_{IJ} + \sum_{I<J<K} \Delta E_{IJK} + \cdots + \Delta E_{IJK\cdots N}, \tag{2.1}$$

where $I, J, \ldots$ index monomers with energies $E_I, E_J, \ldots$. The higher-order correction terms in Eq. (2.1) can be conveniently defined via a recursive formula. The n-body correction $\Delta E_{IJK\cdots n}$ is

$$\Delta E_{IJK\cdots n} = E_{IJK\cdots n} - \sum_{IJ\cdots(n-1)} \Delta E_{IJ\cdots n-1} - \cdots - \sum_{I=1}^{N} E_I, \tag{2.2}$$

where the first term on the right, $E_{IJK\cdots n}$, represents the energy of the system composed of monomers $I, J, K, \ldots, n$. It is important to note that the summations in Eq. (2.2) are restricted to monomer indices present in $\Delta E_{IJK\cdots n}$, which according to Eq. (2.1) means that $I < J < K < \cdots < n$.

In the present study, we are interested in the cases $n = 2$ and $n = 3$, for which Eq. (2.2) becomes

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J, \tag{2.3a}$$

and

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK} - E_I - E_J - E_K. \tag{2.3b}$$

In a large system and/or a large basis set, these truncated expressions may dramatically reduce the amount of computer time that is required to compute the energy. Owing to the highly nonlinear scaling of *ab initio* quantum chemistry, this may be true even before the embarrassingly parallelizable nature of the independent monomer, dimer, and trimer calculations is exploited, i.e., even if all calculations are run in serial on a single processor.[17]

B. "Exact" BSSE corrections

In the spirit of Boys-Bernardi CP correction, many-body BSSE corrections will involve "ghost atom" basis functions on some monomers. We will use a notation with subscript monomer indices, as above, but where superscript indices indicate which monomers contribute basis functions. For example, E_I^{IJ} represents the energy of the "real" monomer I computed in the basis of dimer IJ, and E_{IJ}^{IJK} denotes the energy of the dimer IJ computed in the basis of the trimer IJK.

1. Boys-Bernardi counterpoise correction

The most popular BSSE correction method, the Boys-Bernardi CP correction,[1] uses the full cluster basis set for each monomer calculation. The CP-corrected binding energy, $E_{\text{B.E.}}$, is therefore given by Refs. 17 and 21,

$$E_{\text{B.E.}} = E_{IJK\cdots N}^{IJK\cdots N} - \sum_I E_I^{IJK\cdots N}. \tag{2.4}$$

This formula represents the natural N-body generalization of the method introduced by Boys and Bernardi for dimers,[4,22] and we will refer to it simply as the CP correction. Equation (2.4) has also been called the "site-site function CP method."[23]

It is convenient to add and subtract the energy of each monomer, in its own basis set, to arrive at an additive correction, $\mathcal{E}^{\text{CP}}$, to the uncorrected energy:

$$E_{\text{B.E.}} = E_{\text{B.E.}}^{\text{uncorrected}} + \mathcal{E}^{\text{CP}}. \tag{2.5}$$

The uncorrected binding energy is computed by generalizing the naïve formula of Eq. (1.1),

$$E_{\text{B.E.}}^{\text{uncorrected}} = E_{IJK\cdots N}^{IJK\cdots N} - \sum_I E_I^I, \tag{2.6}$$

and therefore, the CP correction is

$$\mathcal{E}^{\text{CP}} = \sum_I \left(E_I^I - E_I^{IJK\cdots N} \right). \tag{2.7}$$

The convenience of this additive correction is that it facilitates a correction for *intramolecular* BSSE by fragmenting a molecule across covalent bonds.[13]

2. Valiron-Mayer function counterpoise correction

In the literature concerning BSSE, there is a prevalent belief that the CP scheme, in one fashion or another, does not properly account for BSSE, though many of these early claims were later refuted.[4] Perhaps the most familiar criticism of the Boys-Bernardi scheme is that it neglects "higher-order" BSSE effects.[24] Valiron and Mayer suggest that for a cluster of N monomers, the proper way to remove all BSSE is to remove BSSE from all pairs of monomers, from all trimers of monomers, and so on. Using this reasoning, it is straightforward to arrive at a BSSE-corrected binding energy of the form:[24]

$$\begin{aligned} E_{\text{B.E.}} = E_{\text{B.E.}}^{\text{uncorrected}} &+ \sum_I \left(E_I^I - E_I^{IJK\cdots N}\right) \\ &+ \sum_{I<J} \left(\Delta E_{IJ}^{IJ} - \Delta E_{IJ}^{IJ\cdots N}\right) \\ &+ \sum_{I<J<K} \left(\Delta E_{IJK}^{IJK} - \Delta E_{IJK}^{IJ\cdots N}\right) + \cdots . \end{aligned} \tag{2.8}$$

Of particular importance here is the fact that the first correction to the binding energy is a term equal to the standard Boys-Bernardi CP correction, which lends support to the claim that the latter procedure neglects higher-order BSSE effects. The BSSE correction $\mathcal{E}^{\text{VMFC}}$ that is suggested by Eq. (2.8) is known as the VMFC correction.

3. CP versus VMFC

The VMFC method has seen little practical use because its cost grows rapidly with the number of monomers. This is easily seen by considering the number of individual calculations each method requires. For an N monomer cluster, the VMFC correction requires $2^N - 1$ calculations in the supersystem basis set, whereas the Boys-Bernardi CP correction requires only N. Furthermore, various studies have shown that the difference in VMFC and CP binding energies is typically less than 1.0 kcal/mol.[18,25,26] A literature search reveals only one exception, namely, $(H_2O)_6$ binding energies computed at the MP2/cc-pVDZ level of theory, for which a difference of about 4 kcal/mol is obtained.[18] However, this sizable difference is reduced to only 0.3 kcal/mol by the addition of diffuse basis functions.[18]

The claim that CP is negligent naturally poses the question of which BSSE method is best. Currently this question is unanswered and it seems likely that debate will continue for some time because any attempt to resolve this issue must contend with the fact that the difference between the CBS result and any truncated basis set result is equal to the sum of BSIE and BSSE. Without an explicit formula for either it is impossible to determine whether CP, VMFC, or some other correction (such as the so-called chemical Hamiltonian approach[11,12]) best eliminates the "true" BSSE. Here, we focus instead on each method's ability to replicate benchmark results.

TABLE I. Total number of distinct electronic structure calculations required in order to evaluate various BSSE corrections.

Name	Full basis[a]	m-mer basis[b]	Total[c]
CP	N	0	N
VMFC	$2^N - 2$	$(2^m - 2)_N C_m$	$\sum_{i=2}^{N} 2^i\, {}_N C_i$
VMFC(n)	0	$(2^m - 2)_N C_m$	$\sum_{i=2}^{n} (2^i - 2)_N C_i$
MBCP(n)	0	$N_{N-1} C_{m-1}$	$N \sum_{i=2}^{n} {}_{N-1} C_{i-1}$

[a] Number of calculations involving the supersystem basis.
[b] Number of calculations involving a subsystem basis with functions on m monomers ($m < N$), where $jC_k = j!/k!(j-k)!$.
[c] Total number of distinct electronic structure calculations.

C. Approximate BSSE corrections

The number of terms required to apply the BSSE corrections discussed above is given in Table I. For VMFC, the number of calculations in the supersystem basis set grows exponentially with N. That fact, combined with an unwieldily number of smaller calculations as well, means that the VMFC approach is feasible only for very small clusters. The CP approach is more tractable but also becomes expensive for large systems, particularly those likely to be investigated using fragment-based methods. To apply these two BSSE corrections to larger systems, approximate versions based on n-body truncation have been proposed.[17,18] We next discuss these approximate methods.

1. MBCP(n)

Our group recently proposed an approximate form of the Boys-Bernardi CP correction that we called the *many-body counterpoise correction*,[17] abbreviated as MBCP(n) when the MBE is truncated at order n. The idea behind MBCP(n) is that each one of the $E_I^{IJK\cdots N}$ terms in the normal CP correction [Eq. (2.4)] can be fragmented using a MBE. This gives a hierarchy of methods, each corresponding to a particular value of n. Here, we focus on the cases $n = 2$ and $n = 3$, for which the MBCP(n) corrections are

$$\mathcal{E}^{\text{MBCP(2)}} = \sum_I \sum_{J \neq I} \left(E_I^{IJ} - E_I^I\right), \tag{2.9a}$$

$$\begin{aligned} \mathcal{E}^{\text{MBCP(3)}} = \sum_I \sum_{J \neq I} \sum_{\substack{K>J \\ K \neq I}} \Big[E_I^{IJK} - \left(E_I^{IJ} - E_I^I\right) \\ - \left(E_I^{IK} - E_I^I\right) - E_I^I \Big]. \end{aligned} \tag{2.9b}$$

The many-body nature of these equations is immediately apparent when compared to Eqs. (2.3a) and (2.3b). The final BSSE-corrected binding energy, at the MBCP(3) level for example, would be written as

$$\begin{aligned} E_{\text{B.E.}}^{\text{MBCP(3)}} = E_{\text{B.E.}}^{\text{uncorrected}} &+ \sum_I E_I^I \\ &+ \mathcal{E}^{\text{MBCP(2)}} + \mathcal{E}^{\text{MBCP(3)}}. \end{aligned} \tag{2.10}$$

There are several points worth making at this time. For MBCP(n) with $n < N$, there will be no supersystem basis set calculations, in fact it is readily seen that the largest basis set needed is that of n monomers at a time. Furthermore, no calculation needed for MBCP(n) will have more than one "real" monomer at a time, e.g., E_I^{IJK} in Eq. (2.9b) consists of one real monomer and two ghost monomers. Finally, because each term in the correction is independent of the other terms, the MBCP(n) procedure is highly parallelizable.

2. *VMFC(n)*

Kamiya *et al.*[18] have proposed an approximate form of the VMFC that we will call VMFC(n). This approach is motivated by the realization that Eq. (2.8) has terms involving monomer basis sets, dimer basis sets, and so on, up to the full supersystem basis set. Neglecting all terms in Eq. (2.8) that involve basis functions on more than n monomers affords

$$E_{\text{B.E.}} = E_{\text{B.E.}}^{\text{uncorrected}} + \sum_{I=1}^{N} E_I^I + \sum_{I<J}^{{}_NC_2} \Delta E_{IJ}^{IJ} + \sum_{I<J<K}^{{}_NC_3} \Delta E_{IJK}^{IJK} + \cdots + \sum_{I<J<K<\cdots<n}^{{}_NC_n} \Delta E_{IJK\cdots n}^{IJK\cdots n}. \quad (2.11)$$

Upon relabeling the terms, we can rewrite this equation as

$$E_{\text{B.E.}} = E_{\text{B.E.}}^{\text{uncorrected}} + \sum_{I=1}^{N} E_I^I + \mathcal{E}^{\text{VMFC(2)}} + E^{\text{VMFC(3)}} + \cdots + \mathcal{E}^{\text{VMFC}(n)}, \quad (2.12)$$

which defines the various n-body BSSE corrections, $\mathcal{E}^{\text{VMFC}(n)}$.

For a given truncation order, n, VMFC(n) includes all the terms that MBCP(n) does, as well as terms arising from dimers in the n-body basis, trimers in the n-body basis, and so on, up to n-mers in the n-body basis. As such, VMFC(n) is far more expensive than MBCP(n), except for $n = 2$, where the two methods are equivalent. It is reasonable to ask at this point if these "extra" terms afford additional accuracy, but as discussed in Sec. II B 3, the answer is complicated by BSIE, as well as by the error in truncating the MBEs that give rise to VMFC(n) and MBCP(n). These errors are difficult to disentangle. The results presented below, along with previous studies,[18,25,26] seem to suggest that these terms are not worth the increased computational time.

III. COMPUTATIONAL DETAILS

In previous tests of our MBCP(n) method,[17] we used eight isomers of $(H_2O)_6$ for which binding energies computed at the MP2-R12/K2–level of theory,[27–29] along with geometries optimized at the MP2/haTZ level, are available in Ref. 30. Here, "haTZ" is a "heavy augmented" basis, consisting of cc-pVTZ for hydrogen and aug-cc-pVTZ for other atoms. The K2–basis set comes from Ref. 28, and consists of aug-cc-pV5Z with the two highest angular momentum shells removed from each atom. The MP2-R12 method contains terms that depend linearly on the inter-electron distance, and should converge rapidly with respect to the one-particle basis set. As such, these energies are expected to be converged to the CBS result without the need for extrapolation,[30] making them ideal benchmarks for comparing MBCP(n) and VMFC(n).

We also consider a set of ten $(H_2O)_{10}F^-$ clusters, using unrelaxed structures from a molecular dynamics simulation as reported in Ref. 16. Energetics were computed at the resolution-of-identity (RI) MP2 level,[31] using standard auxiliary basis sets.[32]

CBS-quality results were obtained via separate extrapolations for the Hartree-Fock (HF) energy and the correlation energy.[33] For the HF energies a three-point fit was used,[34]

$$E_\zeta^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + ae^{-b\zeta}, \quad (3.1)$$

where a and b are fitting parameters and $\zeta = 2, 3, 4$ for the aug-cc-pVDZ, aug-cc-pVTZ, aug-cc-pVQZ basis sets. (We will refer to these as the aug-cc-pVζZ sequence of basis sets, or "aζZ" for short.) For the correlation energy we used a two-point fit ($\zeta = 3, 4$),[35,36]

$$E_\zeta^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + c\zeta^{-3}, \quad (3.2)$$

where c is a fitting parameter. When the highest basis set used is aug-cc-pVζZ, both of these extrapolation schemes have been shown to be capable of replicating benchmark results of at least aug-cc-pV(ζ + 1)Z quality.[37,38] This means that our current CBS extrapolation procedure should adequately replicate the quintuple-ζ benchmarks. Coordinates and energies for these benchmark clusters are available in the supplementary material.[39]

Binding energies are reported with respect to relaxed monomers. This adds an additional complication from a BSSE correction standpoint, because BSSE corrections are performed at the supersystem geometry. It therefore becomes necessary to account for the relaxation via an additive correction, E^{relax}.[22,25] The form of the relaxation energy,

$$E^{\text{relax}} = \sum_{I=1}^{N} \left(E_I^I - E_I^{\text{opt}}\right), \quad (3.3)$$

is independent of the type of BSSE correction. Here, E_I^I denotes monomer I's energy computed in its own basis at the supersystem geometry, and E_I^{opt} is the same quantity computed at the minimum-energy monomer geometry.

All electronic structure calculations were performed using Q-CHEM.[40] Creation of Q-CHEM inputs for the MBE calculations was carried out using our program FRAGME∩T,[16,41] which creates the appropriate input files when the user inputs the desired fragmentation, embedding, and capping method. (In the present study, each fragment consists of either a single H_2O molecule or F^- ion, so no capping is necessary.) FRAGME∩T also handles the creation of all input files required to compute the BSSE corrections. The MBE is truncated at either $n = 2$ or $n = 3$; we refer to these two methods as "2B" and "3B," respectively. The EE-2B and EE-3B methods are the corresponding electrostatically embedded (EE) methods. We use a point-charge embedding along the lines of that used by Dahlke and Truhlar,[42–44] in which Mulliken charges (computed at the B3LYP/6-31G* level) are used to represent those atoms not included in a monomer, dimer, or trimer electronic structure calculation. The monomer geometries used to compute these point charges are those appropriate for the cluster geometry.

IV. RESULTS

A. CBS extrapolation

The extrapolation *ansätze* in Eqs. (3.1) and (3.2) are rooted in the observed exponential convergence of Hartree-Fock energies with respect to ζ,[33,34] and the provable ζ^{-3} convergence of correlation energies (at least for atoms at the level of second-order perturbation theory; see Ref. 36 and references therein). However, there is no reason to believe that the parameters a, b, and c in Eqs. (3.1) and (3.2) should be the same for the total cluster energy as they are, e.g., for the monomer energies. Stated differently, the results of the CBS extrapolation will depend upon whether the various components of the binding energy are added together before or after extrapolation.

In an attempt to ensure that our forthcoming results are not biased unfairly towards one BSSE correction or another, we briefly consider the impact of several different ways of calculating the CBS-extrapolated binding energy. The relaxed binding energy, $\widetilde{E}^{X}_{\text{B.E.}}$, where the tilde indicates that the relaxation energy is included and X indicates a particular flavor of BSSE correction, e.g., X = CP, can be extrapolated in eight unique ways:

$$\widetilde{E}^{X}_{\text{B.E.}} = \left(E^{\text{uncorrected}}_{\text{B.E.}} + E^{\text{relax}}\right) + (\mathcal{E}^{X}), \tag{4.1a}$$

$$\widetilde{E}^{X}_{\text{B.E.}} = \left(E^{\text{uncorrected}}_{\text{B.E.}}\right) + (E^{\text{relax}} + \mathcal{E}^{X}), \tag{4.1b}$$

$$\widetilde{E}^{X}_{\text{B.E.}} = \left(E^{\text{uncorrected}}_{\text{B.E.}}\right) + (E^{\text{relax}}) + (\mathcal{E}^{X}), \tag{4.1c}$$

$$\widetilde{E}^{X}_{\text{B.E.}} = (E_{(\text{H}_2\text{O})_6} + E^{\text{relax}} + \mathcal{E}^{X*}), \tag{4.1d}$$

$$\widetilde{E}^{X}_{\text{B.E.}} = (E_{(\text{H}_2\text{O})_6} + E^{\text{relax}}) + (\mathcal{E}^{X*}), \tag{4.1e}$$

$$\widetilde{E}^{X}_{\text{B.E.}} = (E_{(\text{H}_2\text{O})_6} + \mathcal{E}^{X*}) + (E^{\text{relax}}), \tag{4.1f}$$

$$\widetilde{E}^{X}_{\text{B.E.}} = (E_{(\text{H}_2\text{O})_6}) + (E^{\text{relax}} + \mathcal{E}^{X*}), \tag{4.1g}$$

$$\widetilde{E}^{X}_{\text{B.E.}} = (E_{(\text{H}_2\text{O})_6}) + (E^{\text{relax}}) + (\mathcal{E}^{X*}). \tag{4.1h}$$

In each of these equations, the terms grouped in parentheses are added together prior to extrapolation, so that in Eq. (4.1d), for example, only a single extrapolation is performed, whereas in Eq. (4.1h) all three components of the binding energy are extrapolated separately. As will become evident shortly, $(E^{\text{uncorrected}}_{\text{B.E.}} + E^{\text{relax}} + \mathcal{E}^{X})$ is equivalent to Eq. (4.1d) and $(E^{\text{uncorrected}}_{\text{B.E.}} + \mathcal{E}^{X}) + (E^{\text{relax}})$ is equivalent to Eq. (4.1f), so the equations above represent all unique permutations.

For convenience, the various energetic components in these equations are defined in Table II, but we call attention to two components in particular. For a particular BSSE correction method, X, we can define

$$\mathcal{E}^{X} = E^{X}_{\text{B.E.}} - E^{\text{uncorrected}}_{\text{B.E.}}, \tag{4.2}$$

and a formula for $\mathcal{E}^{X}$ can be worked out from Eq. (2.5), Eq. (2.10), or Eq. (2.12), for X = CP, MBCP(n), and VMFC(n), respectively. The definition of $\mathcal{E}^{X*}$ is best given mathematically:

$$\mathcal{E}^{X*} = \mathcal{E}^{X} - \sum_{I=1}^{N} E^{I}_{I}. \tag{4.3}$$

TABLE II. Energetic quantities used in the CBS extrapolations in Eqs. (4.1a)–(4.1h).

Quantity	Description
$\widetilde{E}^{X}_{\text{B.E.}}$	Relaxed, BSSE-corrected binding energy using BSSE correction method X
$E^{X}_{\text{B.E.}}$	BSSE-corrected binding energy using correction method X
$E^{\text{uncorrected}}_{\text{B.E.}}$	Binding energy of the system, not corrected for BSSE
$E_{(\text{H}_2\text{O})_6}$	The energy of the $(\text{H}_2\text{O})_6$ supersystem
E^{relax}	Sum of the individual monomer relaxation energies
$\mathcal{E}^{X}$	BSSE correction for BSSE method X
$\mathcal{E}^{X*}$	Same as $\mathcal{E}^{X}$, except that the $\sum_I E^I_I$ terms in $\mathcal{E}^{X}$ and $E^{\text{uncorrected}}_{\text{B.E.}}$ have been cancelled

For the eight $(\text{H}_2\text{O})_6$ clusters, we have computed the MP2/CBS value of $\widetilde{E}^{X}_{\text{B.E.}}$ for each of the five BSSE correction methods, X, that are considered in this work (including no BSSE correction at all), using each of the eight extrapolations suggested in Eqs. (4.1a)–(4.1h). Mean unsigned errors (MUEs) for each extrapolation, relative to the benchmark results in Ref. 30, are listed in Table III. It is immediately apparent that the eight schemes do not predict results of equal quality.

In light of the results in Table III, we will use Eq. (4.1e) to compute CBS binding energies for the rest of the manuscript. Aside from being the most accurate partitioning scheme [alongside Eq. (4.1g)], Eq. (4.1e) has the added benefit that the terms in the first set of parenthesis only depend on how the supersystem's energy is computed and the second set contains all of the dependence on the particular choice of BSSE correction. Consequentially, error considerations in the rest of the paper will be facilitated by allowing us to break the final error into contributions from fragmenting the supersystem and those from approximate BSSE corrections. Because the BSSE corrections are additive, the errors will be as well. It is worth noting that Eq. (4.1d) represents the "normal" way of performing a CBS extrapolation, because it

TABLE III. Errors (relative to MP2/CBS benchmarks in Ref. 30) in MP2 binding energies of $(\text{H}_2\text{O})_6$ obtained by extrapolating results from five different BSSE correction schemes using the eight different extrapolation methods in Eqs. (4.1a)–(4.1h).

	Mean unsigned error kcal mol^{-1}							
	(4.1a)	(4.1b)	(4.1c)	(4.1d)	(4.1e)	(4.1f)	(4.1g)	(4.1h)
None	1.60	1.65	1.60	0.73	0.75	1.07	0.75	1.11
CP	0.88	0.13	0.88	0.09	0.02	0.09	0.02	0.38
MBCP(2)	0.67	0.10	0.67	0.33	0.19	0.32	0.19	0.17
MBCP(3)	0.89	0.16	0.89	0.10	0.04	0.10	0.04	0.40
VMFC(3)	0.90	0.12	0.90	0.09	0.05	0.09	0.05	0.41

is more in line with the original physical motivations of the extrapolation procedures (see Ref. 36, for example).

B. $(H_2O)_6$ clusters

1. Extrapolation without counterpoise correction

Historically, fragment methods have been used as an approximation to a particular level of theory, and their merit has been judged by their ability to reproduce supersystem properties computed at the same level of theory. Here, we repeat this exercise for the relaxed binding energies of the water hexamers. The mean signed errors (MSEs), where

$$\text{error} = E_{\mathrm{X/a}\zeta\mathrm{Z}} - E_{\mathrm{MP2/a}\zeta\mathrm{Z}}, \qquad (4.4)$$

for X = 3B, EE-2B, and EE-3B (recall that aζZ $\equiv$ aug-cc-pVζZ) are shown in Fig. 1. Based on previous work by Dahlke and Truhlar,[42–44] albeit in a more limited selection of basis sets, it is possible to give an empirical accuracy ordering of

$$2\mathrm{B} < 3\mathrm{B} \approx \text{EE-2B} < \text{EE-3B}$$

for the fragment methods considered here. To a large extent, the data in Fig. 1 are in agreement with this trend. (MSEs for 2B/aζZ are not shown in Fig. 1, but are on the order of 10 kcal/mol.) We do note a tendency for the fragment methods to underbind the cluster, and we suggest that this arises from a neglect of stabilizing higher-order interactions; similar underbinding has been seen by other researchers.[18,42–44]

The ability for fragment methods to accurately reproduce a given level of theory is a nice result, but perhaps the more interesting result from Fig. 1 is that the errors generally decrease as ζ increases, with the sole exception being the fortuitously accurate EE-2B/aDZ results. The latter method thus represents a "Pauling point,"[45] in the sense that both addition of three-body terms (EE-3B/aDZ) and/or extension of the basis set (EE-2B/aTZ or EE-3B/aTZ) have deleterious effects on the agreement with supersystem benchmarks. In view of the popularity of double-ζ basis sets in fragment-based approaches, often at the MP2 level,[18,42–44,46–49] this error cancellation should be borne carefully in mind. In fact, there has been little systematic analysis of the errors in fragment-based methods as a function of basis size. In the few studies where a sequence of aug-cc-pVζZ basis sets has been used,[50–52] the conclusions are mixed. Often (but not always), the results improve as ζ increases,[51,52] but the fragment molecular orbital (FMO) method, as applied to $(H_2O)_{16}$, exhibits errors that clearly grow larger as a function of ζ.[50]

At some level, the error statistics in Fig. 1 serve to reiterate the conclusions of Dahlke and Truhlar,[42–44] namely, that two- and three-body expansions offer reasonable accuracy for clusters of polar molecules if—and only if—electrostatic embedding is employed. For $(H_2O)_6$, all such methods afford errors <0.6 kcal/mol, whereas non-embedded three-body results err by >1.2 kcal/mol. However, we are interested in pushing the limits of such approaches in terms of their ability to reproduce high-level benchmarks, and to this end we have recomputed the MSEs for $(H_2O)_6$, except that now we use MP2/CBS values[30] as the benchmark for all fragment-based methods, regardless of basis set. In other words, we replace $E_{\mathrm{MP2/a}\zeta\mathrm{Z}}$ in Eq. (4.4) by $E_{\mathrm{MP2/CBS}}$, in order to define the error. These results are plotted in Fig. 2.

It is important to note that most previous benchmark studies of fragment-based methods have not examined errors with respect to large- or complete-basis results. As such, Fig. 2 reveals an important point that is perhaps worth stating explicitly: errors less than "chemical accuracy" (loosely defined as $\lesssim$1 kcal/mol) are no cause for celebration if the corresponding supersystem benchmark calculations (used to quantify the error) are performed at a low level of theory.

Interestingly, of the various methods examined in Fig. 2, the full (supersystem) MP2 results are *less* accurate, as compared to MP2/CBS benchmarks, than the truncated MBEs. (The exception is the two-body method without electrostatic embedding, which exhibits MSEs of ~10 kcal/mol that are not shown in Fig. 2.) As such, the more faithfully that a fragment method reproduces MP2/aζZ results, the larger is its MSE with respect to MP2/CBS results. Of course, the difference in the MP2/aζZ and MP2/CBS binding energies is equal to the sum of the BSSE and BSIE and so Fig. 2 is really proving that neither of these errors is negligible. Moreover, we have noted previously[17] that extrapolation of non-BSSE-corrected MP2/aζZ energies for $(H_2O)_6$ fails to reproduce the

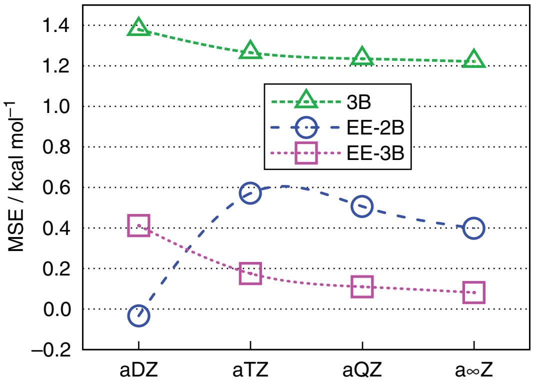

FIG. 1. Mean signed error (MSE) in the uncorrected MP2/aζZ binding energies for eight isomers of $(H_2O)_6$ computed at the MP2/aζZ level. The MSE is computed with respect to supersystem MP2 results in the same basis set. The "a∞Z" values represent CBS extrapolations.

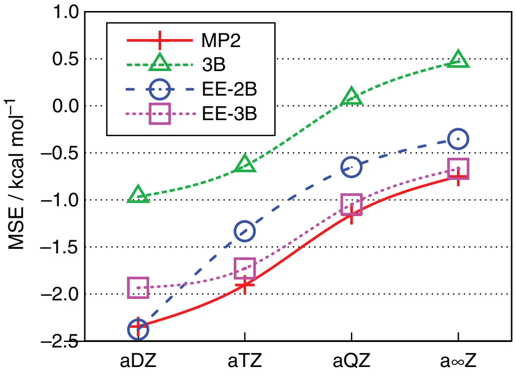

FIG. 2. Convergence (to the CBS limit) of the uncorrected MP2/aζZ binding energies for isomers of $(H_2O)_6$. The MSE is computed with respect to MP2/CBS benchmarks from Ref. 30.

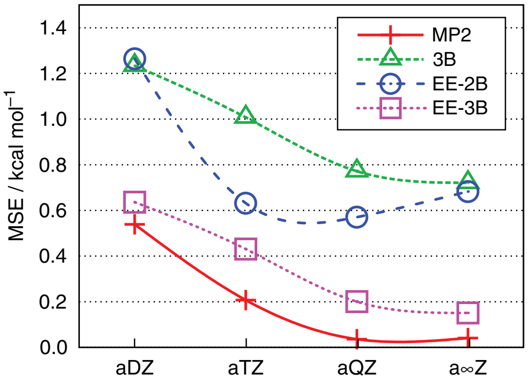

FIG. 3. Convergence (to the CBS limit) of the relative energies of $(H_2O)_6$ isomers. The mean signed errors (MSEs) are computed with respect to MP2/CBS benchmarks from Ref. 30.

consensus value[17,30,53] of the MP2/CBS limit (at least when basis sets no larger than aQZ are employed in the extrapolation), and this is evident in Fig. 2. By contrast, extrapolation of CP-corrected energies through MP2/aQZ *does* obtain the correct limit.[17] The BSSE will be dealt with below.

In chemistry we are often interested in *relative* energies rather than total binding energies, and if two conformations are of approximately the same shape and size, then it is possible that the two might exhibit comparable BSSE, such that a BSSE correction would not be necessary in order to predict the relative energy. To this end, we have calculated the relative stabilities of the eight $(H_2O)_6$ isomers and compared them again to MP2/CBS results from Ref. 30. MSEs obtained from this analysis are plotted as a function of basis set size in Fig. 3.

Unlike the total binding energies, which fail to converge to the CBS limit even following extrapolation, owing to residual BSSE even in the aQZ basis set, the relative energies computed using supersystem MP2 calculations *do* converge to the correct limit. These energies are within 0.2 kcal/mol of that limit already at the MP2/aTZ level. This supports our hypothesis that the BSSE for these systems is similar and thus cancels in relative energy calculations. Further confirmation is obtained by computing the CP corrections to these relative energies. For the aDZ, aTZ, and aQZ basis sets, these CP corrections are 0.5, 0.3, 0.1 kcal/mol, respectively, and <0.1 kcal/mol for the CBS-extrapolated result. Comparing Figs. 2 and 3, where the benchmark is MP2/CBS in both cases, we see that BSSE cancellation leads to relative energies that are significantly more accurate as compared to total binding energies, and various truncated MBEs achieve an accuracy of <1 kcal/mol with respect to relative energies computed at the MP2/CBS level.

That said, the success of this cancellation hinges on the various isomers being of approximately the same size and shape, and cannot hold globally across the potential energy surface. For example, an energy calculation on six sufficiently well-separated water monomers will afford zero BSSE, and from this point of view, a calculation of the total binding energy is just an extreme example of a relative energy calculation. Moreover, compact $(H_2O)_6$ structures are smaller than the radial extent of the most diffuse atomic orbitals in the aTZ and aQZ basis sets, suggesting that all isomers may exhibit roughly comparable BSSE. Calculations performed on a linear (chain) isomer of $(H_2O)_6$ afford CP corrections that are 3.6, 1.8, and 0.83 kcal/mol different from those obtained for the prism isomer, in the aDZ, aTZ, and aQZ basis sets, respectively, and 0.24 kcal/mol different for the CBS-extrapolated result. These values are considerably larger than the average CP corrections quoted above for the compact $(H_2O)_6$ structures, reflecting the difference in cluster shape. This underscores the importance of CP corrections in exploring the *global* potential energy surface, where the BSSE may not cancel from one basin to the next.

In fact, even for stable, hydrogen-bonded cluster isomers, different regions of the potential surface may be affected differently by BSSE, which can be seen by examining larger clusters where a greater variety of hydrogen-bonding morphologies are available. As an example, we have computed dual-basis[54] RIMP2/CBS energetics (based on a two-point cc-pVTZ and cc-pVQZ extrapolation) for 80 different isomers of $(H_2O)_{20}$, corresponding to the 20 lowest-energy structures from each of four families of isomers, as determined by extensive basin-hopping Monte Carlo simulations[55,56] on the TIP4P[57] potential surface. Examples of these structural motifs are depicted in Fig. 4, where we also plot the relative energies of all 80 isomers. Both CP-corrected and uncorrected results are shown, and the difference between the two is uniformly small ($\lesssim$0.2 kcal/mol) for three of the four classes of isomers. Apparently, the BSSE is approximately the same for all 60 of these isomers. For the dodecahedral isomers, however, the CP correction shifts the energies by about 1 kcal/mol relative to the energies of the other three families of isomers, meaning that the BSSE must be about 1 kcal/mol different in the dodecahedral regions of the potential surface.

2. BSSE-corrected results

In view of these discrepancies for $(H_2O)_{20}$, we next discuss the use of truncated MBEs in conjunction with BSSE corrections. As discussed above, BSSE and BSIE are intimately intertwined; nevertheless, it must be the case that the difference between the BSSE-free MP2/CBS benchmarks in Ref. 30 and the uncorrected MP2 energy computed in any finite basis set is equal to the total basis set error (BSE). That is,

$$\text{BSE} = \text{BSSE} + \text{BSIE}. \tag{4.5}$$

BSSE can be reduced (if not eliminated) by means of the CP- and VMFC-type corrections discussed in Sec. II, while BSIE is eliminated only by extrapolation to the CBS limit.

Figure 5 plots the MSEs for BSSE-corrected, MP2/aζZ binding energies of $(H_2O)_6$ isomers, with respect to MP2/CBS benchmarks. We identify the difference between the MP2/aζZ and the MP2/CBS result as the BSE, which is well-defined once the CBS limit is established. On the other hand, we do not consider that any of the BSSE corrections discussed in Sec. II rigorously *defines* BSSE, hence only the sum BSSE + BSIE is well-defined. The BSIE is independent of the BSSE correction that is used, so its net effect is to shift each one of the errors by some amount. Thus, the relative difference

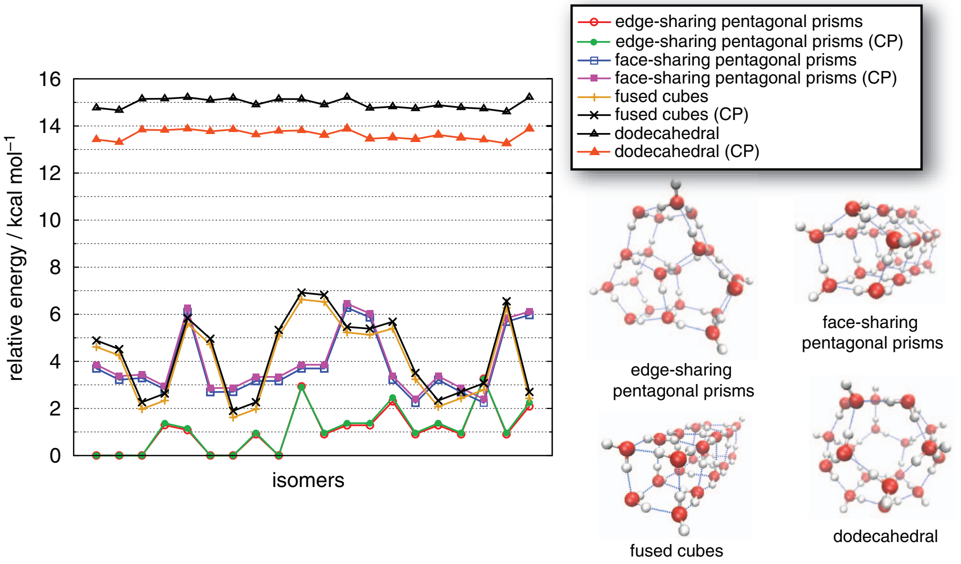

FIG. 4. Relative energies of twenty isomers from each of four structural motifs[55] of $(H_2O)_{20}$. Energies were computed at the dual-basis[54] RIMP2/cc-pVXZ level (for X = T and Q), either with or without CP correction, and then extrapolated to the complete-basis limit using an X^{-3} extrapolation formula. For three of the four motifs, CP correction alters the relative energies by no more than 0.2 kcal/mol. For the dodecahedral isomers, however, the CP-corrected and uncorrected results differ by ≈1 kcal/mol.

between errors in Fig. 5 evaluated for the same basis set but for different BSSE corrections is indicative of the error inherent to a given BSSE correction. These relative differences are quite small compared to the total residual BSE, so we cannot say conclusively, based on this one system, which BSSE correction is the most accurate. However, the MBCP(2) [≡ VMFC(2)] method is likely the *least* accurate, as it provides results that are significantly different from the other three BSSE corrections, and upon extrapolation, misses the established MP2/CBS benchmark by almost 0.2 kcal/mol. The other three BSSE corrections predict the correct limit to within 0.05 kcal/mol.

For a given fragmentation method, basis set, and BSSE correction method, the final error is the sum of the error incurred by using a truncated MBE (Fig. 1), which is usually positive (indicative of underbinding), plus the residual BSE (Fig. 5), which is also positive. Figure 2 shows us that in general all of these methods overbound, when BSSE is not accounted for, and so error cancellation occurs, particularly between the BSSE and the fragment approximation making the uncorrected fragment calculations better estimates of the benchmark value than the full supersystem calculation. Similar error cancellation was observed in an earlier study of the VMFC(n) method.[18]

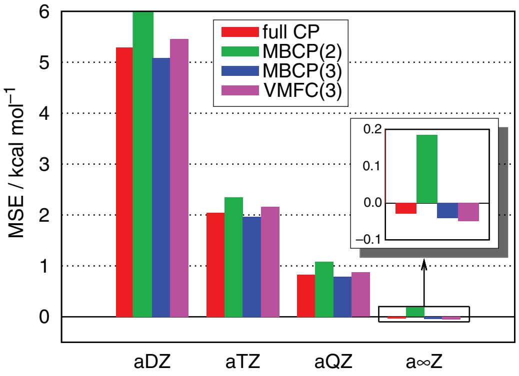

FIG. 5. MSEs in MP2/aζZ binding energies for isomers of $(H_2O)_6$ using four different BSSE corrections in various basis sets. The error, computed with respect to MP2/CBS benchmarks, may be identified as the residual basis set error (BSSE + BSIE).

C. $(H_2O)_{10}F^-$ clusters

In Sec. IV B, we examined various contributions to the error in a truncated many-body approximation to the binding energy of $(H_2O)_6$, both with and without corrections for BSSE. Here, we examine a larger and more strongly bound system, namely, isomers of $(H_2O)_{10}F^-$ that we have previously shown to be much more challenging for fragment-based quantum chemistry, as compared to water clusters.[16,41,58–60] As will be evident from the timing data presented below, computing CP-corrected MP2/CBS binding energies using traditional supersystem methods is quite challenging for systems of this size, and we cannot push the benchmarks too much larger without considerable effort. In particular, we again use Eq. (4.1e) to extrapolate to the CBS limit, but for $(H_2O)_{10}F^-$, large-basis MP2-R12 calculations are not available, and RIMP2/aQZ is the highest level of theory that is feasible on the hardware available to us. Thus we have no independent means to validate the MP2/CBS extrapolations for this system.

The most straightforward error is that associated with truncation of the MBE, which we can assess in the same way that we did for the $(H_2O)_6$ clusters, namely, by comparing the uncorrected supersystem RIMP2/aζZ binding energy to that obtained using a truncated MBE at the same level of theory. The MSEs for the 3B and EE-3B methods are shown in Fig. 6. As with the $(H_2O)_6$ clusters we again see a decrease in

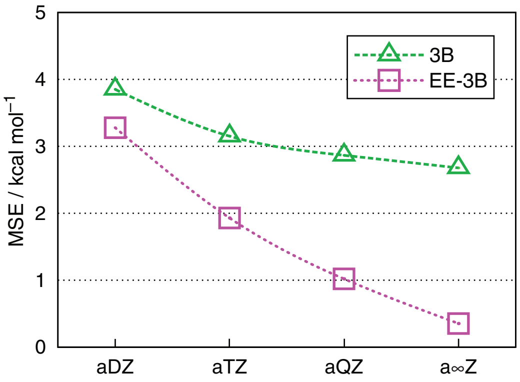

FIG. 6. Convergence of uncorrected RIMP2/aζZ binding energies to the CBS limit, as measured by the mean signed error (MSE) for ten $(H_2O)_{10}F^-$ cluster isomers. The MSE is measured relative to the supersystem RIMP2 energy in the same basis set.

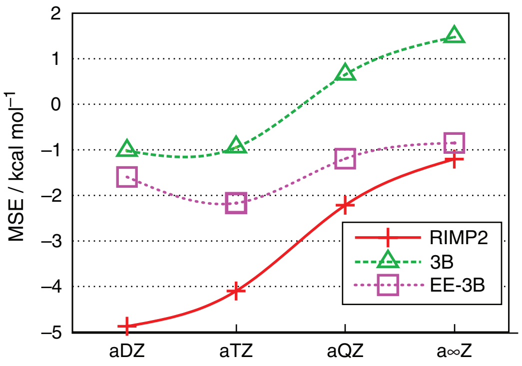

FIG. 7. Convergence of the uncorrected RIMP2/aζZ binding energies to the CBS limit, for ten $(H_2O)_{10}F^-$ clusters. Convergence is measured using the mean signed error (MSE) with respect to an RIMP2/CBS benchmark computed as a weighted average of CP-corrected and uncorrected RIMP2 results, each extrapolated to the CBS limit, as described in the text.

error as a function of ζ, in contrast to FMO results for $(H_2O)_{16}$ using cc-pVζZ basis sets.[50] Our $(H_2O)_{10}F^-$ system is closer in size to $(H_2O)_{16}$, as compared to the water hexamers examined above, which suggests that system size does not explain the deterioration in the FMO results with increasing basis size. The authors of Ref. 50 suggest that the smaller polarization function exponents in the larger correlation-consistent basis sets are a problem for the FMO method, perhaps due to (unspecified) problems with the self-consistent embedding scheme when basis functions centered on one monomer strongly overlap another monomer. Our own (unpublished) experience with self-consistent "XPol" embedding[59] suggests that larger basis sets can be a problem when proper monomer densities, as opposed to point charges, are used to compute the intermonomer Coulomb interactions in a self-consistent embedding scheme. In any case, the present results show that this is not a problem for the point-charge embedding used here, which is stable with respect to expansion of the basis set, even in the presence of diffuse functions.

Further analysis requires some estimate of the benchmark quantity. To this end, we assume that the CBS-extrapolated, CP-corrected RIMP2 binding energies lie closer to the true RIMP2/CBS values as compared to extrapolation *without* BSSE correction, which is the case for all eight of the $(H_2O)_6$ isomers.[17] We furthermore assume that the CP-corrected RIMP2 binding energies converge to the CBS limit from above and that the uncorrected results converge from below, which is again true of all the $(H_2O)_6$ isomers[17] as well as other systems that we have studied, such as tryptamine–water.[61] Indeed, it is this empirical observation that has led to the average of CP-corrected and uncorrected results, each extrapolated to the basis-set limit, being used as an estimate of the true CBS result.[53,61] For $(H_2O)_6$, extrapolation of the CP-corrected MP2 results affords very good agreement with an independently verifiable MP2/CBS limit, but for $(H_2O)_{10}F^-$ we have no such independent benchmark, and extrapolation using only aDZ, aTZ, and aQZ results may or may not achieve the true CBS limit. As such, we will use a 3:1 weighted average of the CP-corrected and uncorrected extrapolations as a benchmark for the MP2/CBS limit. (Thus, the chosen benchmark lies closer to the CP-corrected extrapolation.) Error bars on this MP2/CBS result will be established by assuming that the unweighted (1:1) average of the extrapolated CP-RIMP2/CBS and RIMP2/CBS results is a lower bound to the true MP2/CBS result, and that the CP-RIMP2/CBS extrapolation provides an upper bound. Based on these assumptions, we conclude that our RIMP2/CBS benchmarks should be accurate to within about 0.4 kcal/mol.

Figure 7 plots MSEs in the uncorrected RIMP2/aζZ binding energies as a function of basis size, using the weighted average just described as the RIMP2/CBS benchmark. Based on Fig. 7, it is apparent that a cancellation of errors (with BSSE compensating for truncation of the MBE) is again at work, as EE-3B/aζZ results lie closer to the benchmark as compared to RIMP2/aζZ results. (As a result, reproducing the full RIMP2/aζZ result in a given basis set is not the best measure of how well a given fragment method reproduces the binding energy in the CBS limit.) This shows that the approximate cancellation between BSSE and higher-order many-body effects continues to hold in clusters larger than $(H_2O)_6$; the magnitude of both is increasing with system size.

As in the case of the water hexamer, aQZ results—possibly in conjunction with CBS extrapolation—are required to get within ~1 kcal/mol of the benchmark CBS value of the total binding energy. Accurate prediction of relative energies is less demanding, as shown in Fig. 8. Errors in EE-3B/aζZ relative energies are <0.6 kcal/mol in magnitude, as compared to RIMP2/CBS benchmarks, even in the aDZ basis set, and with a triple-ζ basis the error is <0.2 kcal/mol. As in the case of $(H_2O)_6$ clusters, calculation of the BSSE corrections for each $(H_2O)_{10}F^-$ isomer confirms that they are of similar magnitude and, therefore, cancel in the evaluation of relative energies. Somewhat surprisingly, these corrections are similar in magnitude to the values obtained for $(H_2O)_6$ isomers, despite the increased system size.

We next apply the BSSE corrections, and in Fig. 9 we plot the residual BSE across the aug-cc-pVζZ sequence of basis sets, using several different BSSE corrections. The difference in accuracy between the MBCP(3) and VMFC(3) methods is about 0.2 kcal/mol, smaller than the ≈0.4 kcal/mol error bar that was established above for the benchmark values. (Recall, however, that the MBCP(3) method is less expensive

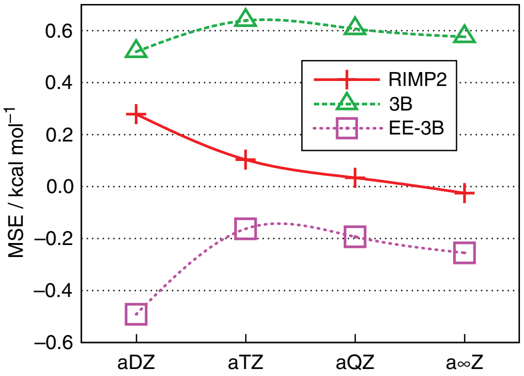

FIG. 8. Convergence of the relative energies of $(H_2O)_{10}F^-$ cluster isomers to the CBS limit. MSEs are measured relative to the RIMP2/CBS benchmarks that are discussed in the text.

than VMFC(3).) Combined with the similar results for the $(H_2O)_6$ systems, this suggests that both the VMFC(n) and the MBCP(n) corrections converge rapidly and to nearly the same value. This implies that basis-set extension effects, which are absent in the MBCP(n) approach, are negligible.

Given the very similar performance of VMFC(3) and MBCP(3), we prefer the latter for reasons of cost, as the MBCP(3) approach requires only a proper subset of the electronic structure calculations that are required for VMFC(3). Timing data for a $(H_2O)_{10}F^-$ cluster are shown in Fig. 10. These timings represent the total, aggregate computer time required to compute the CP-corrected [Eq. (2.4)] RIMP2/aζZ binding energy in each aug-cc-pVζZ basis set, along with the relative speed-up engendered by using the MBCP(2), MBCP(3), or VMFC(3) correction. For the purposes of this comparison, all calculations are run in serial, meaning that the "embarrassingly parallelizable" nature of the fragment-based corrections is *not* exploited. Wall times for the fragment-based calculations could therefore be trivially and dramatically reduced by running individual jobs on separate processors; already, however, the MBCP(3) method reduces the cost by ∼70%, without significant loss of accuracy.

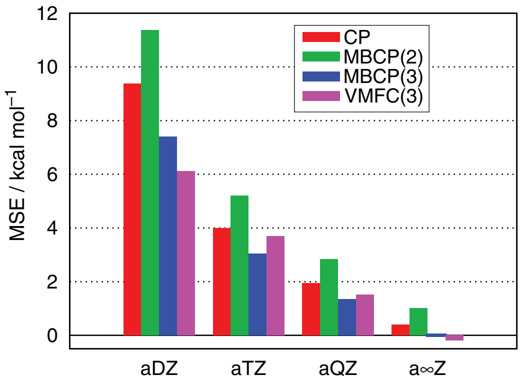

FIG. 9. Residual basis set error for ten isomers of $(H_2O)_{10}F^-$, computed at the RIMP2 level using various basis sets. The MSE is evaluated with respect to the RIMP2/CBS benchmarks that are discussed in the text.

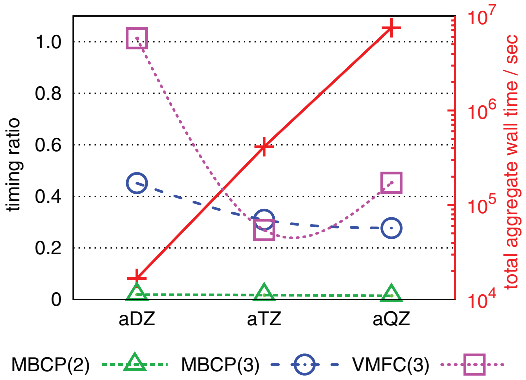

FIG. 10. Total aggregate computer time (red crosses, to be read from the logarithmic axis on the right) required to perform CP-corrected RIMP2/aζZ calculations on a single $(H_2O)_{10}F^-$ cluster geometry. Other symbols depict the relative times required to evaluate various low-order BSSE corrections (to be read from the linear axis on the left). All calculations are performed in serial on a 2.5 GHz Opteron processor.

V. CONCLUSIONS

The present study has two aims. First and foremost, we continue our efforts[16,17,41] to understand the strengths and weaknesses of fragment-based methods, particularly for high-accuracy calculations. Considering that a large number of previous studies using fragment-based quantum chemistry have focused on results obtained using double-ζ basis sets, we have explored the ability of truncated many-body expansions to replicate MP2/CBS results rather than, say, MP2/aug-cc-pVDZ results. Accuracy of <1 kcal/mol in total binding energies (with respect to MP2/CBS results) is obtainable by means of electrostatically embedded three-body expansions, but requires the use of quadruple-ζ basis sets and/or CBS extrapolation. We also find that the three-body result in a given, finite basis set often lies closer to the MP2/CBS benchmark as compared to full (supersystem) MP2 in the same basis set. This is ultimately fortuitous, resulting from a cancellation of errors wherein BSSE partially offsets errors engendered by neglecting higher-order many-body effects. It is worth noting all the same, as it implies that evaluating the accuracy of fragment-based methods in small basis sets may not provide a full picture of how such methods perform in the basis-set limit.

A second focus of this work is on the effects of low-cost BSSE corrections based on low-order many-body expansions, including the MBCP(n) method of Ref. 17 and the VMFC(n) method of Ref. 18. The intertwined nature of BSSE and basis incompleteness makes it difficult to ascertain which approach is most accurate, but at the $n = 3$ level the difference between the two is comparable to, or smaller than, the intrinsic error in establishing the CBS limit. On the one hand, this implies that so-called "basis-set extension" effects, which are not corrected by the MBCP(n) approach, must be quite small, at least for the $(H_2O)_6$ and $(H_2O)_{10}F^-$ examples considered here. At the same time, it means that there is no reason *not* to use the lower-cost MBCP(n) approach, which results in significant computational savings and enables smooth extrapolation of MP2/aug-cc-pVζZ results (ζ = D, T, Q) to the MP2/CBS limit, by means of electronic structure calculations on systems no larger than trimers.

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center under project PAA0003. J.M.H. is an Arthur P. Sloan Foundation Fellow and a Camille Dreyfus Teacher-Scholar.

[1] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[2] F. Jensen, Chem. Phys. Lett. **261**, 633 (1996).
[3] R. M. Balabin, J. Chem. Phys. **132**, 211103 (2010).
[4] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).
[5] R. A. DiStasio, Jr., R. P. Steele, Y. M. Rhee, Y. Shao, and M. Head-Gordon, J. Comput. Chem. **28**, 839 (2007).
[6] A. E. Shields and T. van Mourik, J. Phys. Chem. A **111**, 13272 (2007).
[7] T. van Mourik, J. Phys. Chem. A **112**, 11017 (2008).
[8] T. van Mourik, Chem. Phys. Lett. **473**, 206 (2009).
[9] J. Cao and T. van Mourik, Chem. Phys. Lett. **485**, 40 (2010).
[10] D. Toroz and T. van Mourik, Phys. Chem. Chem. Phys. **12**, 3463 (2010).
[11] I. Mayer, Int. J. Quantum Chem. **23**, 341 (1983).
[12] I. Mayer, Mol. Phys. **92**, 503 (1997).
[13] F. Jensen, J. Chem. Theory Comput. **6**, 100 (2010).
[14] H. Kruse and S. Grimme, J. Chem. Phys. **136**, 154101 (2012).
[15] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).
[16] R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113 (2012).
[17] R. M. Richard, K. U. Lao, and J. M. Herbert, J. Phys. Chem. Lett. **4**, 2674 (2013).
[18] M. Kamiya, S. Hirata, and M. Valiev, J. Chem. Phys. **128**, 074103 (2008).
[19] D. M. Bates, J. R. Smith, T. Janowski, and G. S. Tschumper, J. Chem. Phys. **135**, 044123 (2011).
[20] K. U. Lao and J. M. Herbert, J. Chem. Phys. **139**, 034107 (2013).
[21] G. S. Tschumper, "Reliable electronic structure computations for weak noncovalent interactions in clusters," in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari (Wiley-VCH, 2009), Vol. 26, Chap. 2, pp. 39–90.
[22] L. Turi and J. J. Dannenberg, J. Phys. Chem. **97**, 2488 (1993).
[23] B. H. Wells and S. Wilson, Chem. Phys. Lett. **101**, 429 (1983).
[24] P. Valiron and I. Mayer, Chem. Phys. Lett. **275**, 46 (1997).
[25] G. Lendvay and I. Mayer, Chem. Phys. Lett. **297**, 365 (1998).
[26] P. Salvador and M. M. Szczęśniak, J. Chem. Phys. **118**, 537 (2003).
[27] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991).
[28] W. Klopper, J. Chem. Phys. **102**, 6168 (1995).
[29] E. Valeev, Chem. Phys. Lett. **395**, 190 (2004).
[30] D. M. Bates and G. S. Tschumper, J. Phys. Chem. A **113**, 3555 (2009).
[31] M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).
[32] F. Weigend, M. Häser, J. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[33] F. Jensen, J. Chem. Phys. **110**, 6601 (1999).
[34] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).
[35] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).
[36] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).
[37] D. Feller, K. A. Peterson, and J. G. Hill, J. Chem. Phys. **135**, 044102 (2011).
[38] D. Feller, J. Chem. Phys. **138**, 074103 (2013).
[39] See supplementary material at http://dx.doi.org/10.1063/1.4836637 for cluster coordinates and benchmark energies.
[40] 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).
[41] R. M. Richard and J. M. Herbert, J. Chem. Theory Comput. **9**, 1408 (2013).
[42] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).
[43] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).
[44] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **4**, 1 (2008).
[45] P.-O. Löwdin, Int. J. Quantum Chem., Symp. **28**, 19 (1986).
[46] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).
[47] A. D. Kulkarni, V. Ganesh, and S. R. Gadre, J. Chem. Phys. **121**, 5043 (2004).
[48] A. P. Rahalkar, M. Katouda, S. R. Gadre, and S. Nagase, J. Comput. Chem. **31**, 2405 (2010).
[49] N. Sahu, S. D. Yeole, and S. R. Gadre, J. Chem. Phys. **138**, 104101 (2013).
[50] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **123**, 134103 (2005).
[51] G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).
[52] A. P. Rahalkar, B. K. Mishra, V. Ramanathan, and S. R. Gadre, Theor. Chem. Acc. **130**, 491 (2011).
[53] S. S. Xantheas, C. J. Burnham, and R. J. Harrison, J. Chem. Phys. **116**, 1493 (2002).
[54] R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong, and M. Head-Gordon, J. Chem. Phys. **125**, 074108 (2006).
[55] D. J. Wales and M. P. Hodges, Chem. Phys. Lett. **286**, 65 (1998).
[56] TIP4P geometries for the $(H_2O)_{20}$ clusters (from Ref. 55) were downloaded from the Cambridge Cluster Database, see http://www-wales.ch.cam.ac.uk/CCD.html.
[57] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Imprey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[58] K. U. Lao and J. M. Herbert, J. Phys. Chem. A **116**, 3042 (2012).
[59] J. M. Herbert, L. D. Jacobson, K. U. Lao, and M. A. Rohrdanz, Phys. Chem. Chem. Phys. **14**, 7679 (2012).
[60] K. U. Lao and J. M. Herbert, J. Phys. Chem. Lett. **3**, 3241 (2012).
[61] J. R. Clarkson, J. M. Herbert, and T. S. Zwier, J. Chem. Phys. **126**, 134306 (2007).