

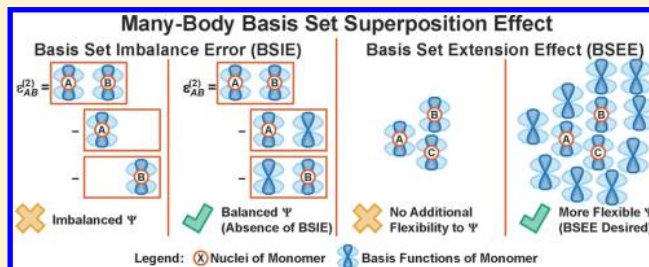
Many-Body Basis Set Superposition Effect

John F. Ouyang and Ryan P. A. Bettens*

Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543

S Supporting Information

ABSTRACT: The basis set superposition effect (BSSE) arises in electronic structure calculations of molecular clusters when questions relating to interactions between monomers within the larger cluster are asked. The binding energy, or total energy, of the cluster may be broken down into many smaller subcluster calculations and the energies of these subsystems linearly combined to, hopefully, produce the desired quantity of interest. Unfortunately, BSSE can plague these smaller fragment calculations. In this work, we carefully examine the major sources of error associated with reproducing the binding energy and total energy of a molecular cluster. In order to do so, we decompose these energies in terms of a many-body expansion (MBE), where a “body” here refers to the monomers that make up the cluster. In our analysis, we found it necessary to introduce something we designate here as a many-ghost many-body expansion (MGMBE). The work presented here produces some surprising results, but perhaps the most significant of all is that BSSE effects up to the order of truncation in a MBE of the total energy cancel exactly. In the case of the binding energy, the only BSSE correction terms remaining arise from the removal of the one-body monomer total energies. Nevertheless, our earlier work indicated that BSSE effects continued to remain in the total energy of the cluster up to very high truncation order in the MBE. We show in this work that the vast majority of these high-order many-body effects arise from BSSE associated with the one-body monomer total energies. Also, we found that, remarkably, the complete basis set limit values for the three-body and four-body interactions differed very little from that at the MP2/aug-cc-pVDZ level for the respective subclusters embedded within a larger cluster.



1. INTRODUCTION

A major barrier to the theoretical study of large chemical systems is the fact that the computational effort of electronic structure methods increases drastically with system size. To circumvent this, one can look to fragmentation methods^{1–4} where a large chemical system is broken up into numerous small subsystems. From there, only certain important interactions between these subsystems are considered for electronic structure calculations so as to recover the total energy of the system. Beyond a specified distance cutoff, each of the subsystems can be treated as multipoles (point charges, dipoles, quadrupoles, and so on) and interacted according to perturbation theory⁵ to further reduce computational cost. Fundamentally, fragmentation methods are rooted in the many-body expansion (MBE), which decomposes the total energy of a cluster, E_{tot} , as the sum of one-body total energies, two-body pairwise interactions, three-body interactions, and so on, up to n -body interactions.⁶ The effects of additional bodies are expected to diminish quickly with increasing bodies, allowing for a good estimate of E_{tot} by truncating the MBE at a low order, typically at or before the four-body interaction.^{7–9}

However, the basis set superposition effect¹⁰ (BSSE) comes into play as energy differences are involved in computing the many-body interactions. BSSE arises when monomers within a molecular cluster borrow basis functions from other monomers to compensate for their basis set incompleteness. The same applies to any subcluster within the molecular cluster. Thus,

when the total energies of interacting monomers and their isolated counterparts are compared, there is an imbalance in the computed many-body interactions. To eliminate this basis set imbalance error (BSIE) for a dimer system, Boys and Bernardi proposed the counterpoise (CP) method to compute the binding energy where the monomer energies are calculated in the dimer basis.¹¹ To clarify our use of terminology, we use the term “location basis” to describe the placement of basis functions at the specified location in the cluster. The CP method was extended for many-monomer molecular clusters to give the Site–Site Function Counterpoise (SSFC) method by calculating the monomer energies in the cluster basis.¹² The many-body counterpoise (MBCP) method^{13,14} was proposed later to approximate the expensive calculation of the monomer energies in the cluster basis by performing a MBE-like decomposition of the effects of the ghost functions present. Nonetheless, the consistent use of the cluster basis in the SSFC method allows for a meaningful decomposition of the binding energy into its many-body contributions. The SSFC method is not a unique extension of the CP method.^{12,15} Valiron and Mayer proposed that the many-body interaction of a subcluster can be instead computed using the set of basis functions centered on the subcluster of interest, i.e., the subcluster basis.¹⁶ These many-body interactions can then be summed to give the

Received: April 13, 2015

Published: September 23, 2015



Valiron–Mayer Function Counterpoise (VMFC) corrected binding energy. While both the SSFC and VMFC methods eliminate BSIE through the use of a consistent basis, the use of the cluster basis in the former incorporates an additional basis set extension effect (BSEE) where the monomers surrounding the subcluster of interest can extend their basis functions—functions present in the cluster basis but not the subcluster basis—to improve the quality of the computed many-body interactions.

The crucial point from the aforementioned counterpoise methods is that many-body BSSE can be divided into two components, namely the BSIE and BSEE (see section 2.2 for a detailed description). The BSIE (where the E stands for error) is undesirable, causing pairwise interactions and consequently the binding energy of large clusters to be overstabilizing. On the other hand, the BSEE (where the E stands for effect) is necessary to reproduce the binding energy and total energy of molecular clusters because all the monomers are better electronically described with the additional external basis functions. Earlier, we showed that the use of the cluster basis leads to rapid convergence of the MBE,¹⁷ indicating that the BSEE is indeed present in the total energy. When the subcluster basis is used, we observed that the MBE converged rapidly, but to an “incorrect value.” There is a significant difference between this “incorrect value” and the total energy, which is essentially the BSEE. More importantly, the rapid convergence associated with the subcluster basis suggested that the BSEE diminishes as rapidly as the many-body interactions. This is of relevance as we noticed that the many-body interactions computed using the subcluster basis are commonly employed in the construction of ab initio based potential energy surface (PES) in the literature.^{18–21} Furthermore, perturbation theory is only compatible with many-body interactions computed using the subcluster basis as the former is BSSE-free by definition regardless of the type of BSSE. Thus, it is important to determine if BSEE is significant in these many-body interactions.

In this work, we examine the amount of BSSE, in particular BSEE, present in many-body interactions in order to identify the major sources of error associated with reproducing the binding energy and total energy of a molecular cluster via an MBE. First, we investigate whether the BSEE is significant in the many-body interactions up to the four-body level. Second, we introduce the many-ghost many-body expansion (MGMBE) to precisely and quantitatively account for both the BSIE and BSEE. Remarkably, we found that the oscillatory behavior of the MBE when diffuse functions are involved can be traced to the BSEE in the one-body interactions, i.e., the monomer total energies. Third, with the removal of the monomer total energies and associated BSEE, the MGMBE is able to accurately reproduce the binding energies of molecular clusters using the energies of numerous subclusters that are no larger than four monomers. Notably the utilization of embedded charges, or a coulomb field, is entirely unnecessary to accomplish this.

2. THEORY

Before discussing the theory behind the MBE, many-body BSSE, and MGMBE, we need to define the following terms and quantities which will be constantly used throughout this work. From here on, we denote the molecular cluster of interest simply as the “cluster” while a “subcluster” refers to a collection of monomers taken from the cluster. In the counterpoise

methods, additional basis functions are placed on the locations of nuclei in the cluster, but without the nuclei being present in the electronic structure calculation, and these functions are called “ghost functions.” We also use the term “location basis” to describe the location at which basis functions are placed in the calculation. For example, the cluster basis refers to the placement of basis functions at the locations of all nuclei present in the cluster. Each of the “bodies” in the many-body interactions refers to a monomer from the cluster, which is taken to be an individual water molecule in this work. When discussing the BSEE, we denote a “ghost-body” as the set of ghost functions centered on a monomer surrounding the subcluster of interest. Table 1 summarizes the relevant quantities described in this work.

Table 1. List of Important Quantities Presented in This Work, Followed by a Brief Definition and the Equation in Which It First Appeared

quantity	definition	eq
E_{tot}	total energy of a cluster.	1
$E_{A\cdots K\cdots M}$	total energy of k -mer subcluster $A\cdots K$ calculated in the presence of ghost functions centered on $L\cdots M$	4 ^a
ϵ_{tot}^C	binding energy computed using the cluster basis	7
$E_{\text{ext}}^{(k)}$	basis set extension effect (BSEE) in the total k -body interaction	8
$\xi_{A\cdots K,L\cdots M}$	BSEE from m -ghost-body $L\cdots M$ in the k -body interaction of $A\cdots K$	9
$\epsilon_{A\cdots K\cdots M}$	k -body interaction of $A\cdots K$ computed using total energies calculated with basis functions centered on $A\cdots K\cdots M$	10 ^b

^a $E_{A\cdots K\cdots M}$ is mentioned much earlier in text at the beginning of section 2.2. ^b $\epsilon_{A\cdots K\cdots M}$ is defined and explained much earlier in text at the second paragraph of section 2.3.

2.1. Many-Body Expansion. For a cluster containing n monomers, the MBE allows us to decompose the total energy of the cluster, E_{tot} , into its many-body contributions

$$E_{\text{tot}} = \sum_A \binom{n}{1} \epsilon'_A + \sum_{A<B} \binom{n}{2} \epsilon'_{AB} + \sum_{A<B<C} \binom{n}{3} \epsilon'_{ABC} + \sum_{A<B<C<D} \binom{n}{4} \epsilon'_{ABCD} + \dots + \epsilon'_{A\cdots N} \quad (1)$$

where $\epsilon'_{A\cdots K}$ is the k -body interaction of the k -mer subcluster $A\cdots K$, of which there are $\binom{n}{k}$ of such terms (Figure 1). The prime symbol in $\epsilon'_{A\cdots K}$ indicates that the basis functions are placed exclusively at the location of the nuclei; i.e., no ghost functions are involved. In this work, we truncate the MBE at the four-body level and thus only provide the relevant equations up to the four-body interaction.

$\epsilon'_{A\cdots K}$ is not directly obtainable from electronic structure calculations, which only gives the total energy, $E_{A\cdots K}$, of the k -mer subcluster of interest. Thus, we need to write the many-body interactions in terms of the total energies. $\epsilon'_{A\cdots K}$ is defined recursively using lower-body interactions^{2,6,7,22} and can then be expressed in terms of total energies

$$\epsilon'_A = E_A \quad (2a)$$

$$\epsilon'_{AB} = E_{AB} - (E_A + E_B) \quad (2b)$$

$$\epsilon'_{ABC} = E_{ABC} - (E_{AB} + E_{AC} + E_{BC}) + (E_A + E_B + E_C) \quad (2c)$$

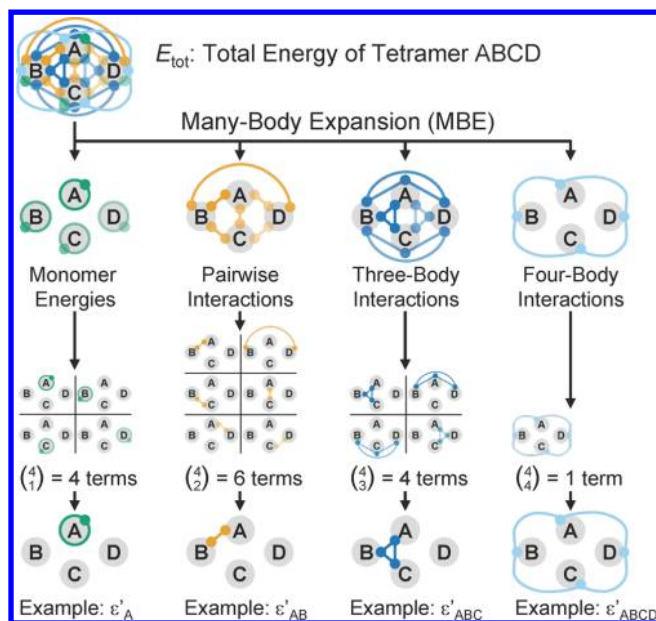


Figure 1. MBE allows us to easily identify the numerous interactions between monomers that is encompassed within the total energy of the cluster. The MBE is illustrated here for a tetramer ABCD ($n = 4$) where the total energy, E_{tot} is decomposed into the total k -body interactions, $k = 1-4$, which comprises $\binom{n}{k}$ individual terms. The explicit formulas for calculating each individual k -body interaction are given in eqs 2a–2d in the text.

$$\begin{aligned} \epsilon'_{ABCD} = & E_{ABCD} - (E_{ABC} + E_{ABD} + E_{ACD} + E_{BCD}) \\ & + (E_{AB} + E_{AC} + E_{AD} + E_{BC} + E_{BD} + E_{CD}) \\ & - (E_A + E_B + E_C + E_D) \end{aligned} \quad (2d)$$

Here, the one-body interaction, ϵ'_A , is the total energy of isolated monomer A while the two-body interaction, ϵ'_{AB} , gives the pairwise interaction between monomers A and B. The three-body interaction, ϵ'_{ABC} , can be understood as the effect of a third monomer C on the interaction between the other two monomers A and B, and the higher-body interactions can be interpreted similarly. In order to employ the MBE, the k -body interaction of individual subclusters presented in eqs 2a–2d has to be collected to give the total k -body interaction. Many total energy terms are repeated during this collection process, and a compact expression of the total k -body interaction is derived and presented in the Supporting Information.

Other than expressing the total energy of the cluster using the MBE, another quantity of interest is the binding energy of the cluster

$$\begin{aligned} \epsilon_{\text{tot}} = & E_{\text{tot}} - \sum_A \binom{n}{1} E_A \\ = & \sum_{A < B} \binom{n}{2} \epsilon'_{AB} + \sum_{A < B < C} \binom{n}{3} \epsilon'_{ABC} + \sum_{A < B < C < D} \binom{n}{4} \epsilon'_{ABCD} + \\ & \dots + \epsilon'_{A \dots N} \end{aligned} \quad (3)$$

2.2. Many-Body Basis Set Superposition Effect. In the electronic structure calculation of a cluster, basis functions from one monomer can be exploited by other monomers to compensate for their basis set incompleteness, and this is

known as BSSE. In many-body systems, this BSSE can be divided into two components, namely the basis set imbalance error (BSIE) and basis set extension effect (BSEE). The distinction between these two components becomes clear when we compare the various placements of basis functions, i.e., the location basis, in computing the many-body interactions (Table 2). So far, the total energy, $E_{A \dots K}$, is written such that it is

Table 2. Comparison of the Choice of Basis in Computing the Many-Body Interactions and Binding Energy, Together with the Name of the Method Reported in the Literature^a

location basis	nuclei-centered	subcluster	cluster
many-body interactions	$\epsilon'_{A \dots K}$	$\epsilon_{A \dots K}$	$\epsilon_{A \dots K \dots N}$
eq for many-body interactions	eqs 2a–2d	eqs 4a–4d	eqs 6a–6d
binding energy	ϵ'_{tot}	ϵ_{tot}^S	ϵ_{tot}^C
eq for binding energy	eq 3	eq 5	eq 7
name in literature	uncorrected ^b	VMFC/ kCBS	SSFC/CP
absence of BSIE? ^c	no	yes	yes
presence of BSEE? ^c	no	no	yes
does MBE converge to E_{tot} ?	yes	no	yes
MBE convergence	slow, oscillatory	rapid	rapid
further remarks	default basis	agree with PT	very expensive

^aFurthermore, the absence of the BSIE and presence of BSEE in the many-body interactions are compared with their effects on the convergence of the MBE of the total energy. ^bThere is no formal name for the nuclei-centered basis as it is the default basis in electronic structure calculations. ^cAs summarized toward the end of section 2.2, the absence of BSIE and the presence of BSEE is desirable, as in the case of the cluster basis.

determined by the identity of the k -mer subcluster or, more specifically, the location of the nuclei constituting the subcluster. In the context of molecular orbital based electronic structure calculations, the total energy also depends on the placement of basis functions. For example, the total energy of monomer A calculated in the nuclei-centered basis centered on A alone would be different from that using the set of basis functions centered on the cluster A...N. From here on, we use the more general notation, $E_{A \dots K \dots M}$. The overline in the subscript indicates the presence of ghost functions centered on L...M in the electronic structure calculation.

The most straightforward method to compute the many-body interactions is to place basis functions exclusively at the locations of nuclei in the electronic structure calculations. This is the usual way of calculating an electronic energy of a molecule. Thus, the k -body interaction computed using the nuclei-centered basis follows eqs 2a–2d presented earlier. We emphasize again that the prime symbol in the many-body interaction, $\epsilon'_{A \dots K}$, indicates the nuclei-centered basis where the number of basis functions are different across the different total energy terms. This is in contrast to the “consistent” subcluster and cluster basis which will be introduced shortly. Similarly, the binding energy computed using the nuclei-centered basis, ϵ_{tot} , follows eq 3. The ϵ_{tot} is often called the uncorrected binding energy for reasons that will be obvious in the following discussion. In computing ϵ'_{AB} using eq 2b, it is clear that E_{AB} is calculated using more basis functions as compared to E_A and E_B . In calculating E_{AB} , monomer A can utilize basis functions centered on monomer B to improve the description of its wave function and vice versa. This is obviously absent in the

calculation of E_A and E_B . This imbalance in the number of basis functions used in the three different calculations of the total energy is the origin of the BSIE. The same BSIE manifests in higher-body interactions in eqs 2c and 2d and the binding energy in eq 3. For the two-body interactions, the BSIE leads to the interactions being overstabilizing. Previously, we also found that MBEs using $\epsilon'_{A\cdots K}$ exhibit slow and oscillatory convergence, especially when diffuse basis functions are present.¹⁷

To remove the BSIE in many-body interactions, we need to ensure that there is a common set of basis functions employed in each of the total energy calculation. The smallest common set is one that is centered on the subcluster for which the many-body interaction is computed. We denote this as the subcluster basis and the k -body interaction can be written as

$$\epsilon_A = E_A \quad (4a)$$

$$\epsilon_{AB} = E_{AB} - (E_{\overline{AB}} + E_{\overline{BA}}) \quad (4b)$$

$$\epsilon_{ABC} = E_{ABC} - (E_{\overline{ABC}} + E_{\overline{ACB}} + E_{\overline{BCA}}) + (E_{\overline{ABC}} + E_{\overline{BAC}} + E_{\overline{CAB}}) \quad (4c)$$

$$\epsilon_{ABCD} = E_{ABCD} - (E_{\overline{ABCD}} + E_{\overline{ABDC}} + E_{\overline{ACDB}} + E_{\overline{BCDA}}) + (E_{\overline{ABCD}} + E_{\overline{ACBD}} + E_{\overline{ADBC}} + E_{\overline{BCAD}} + E_{\overline{BDAC}} + E_{\overline{CDAB}}) - (E_{\overline{ABCD}} + E_{\overline{BACD}} + E_{\overline{CABD}} + E_{\overline{DABC}}) \quad (4d)$$

where the total energy terms containing the same number of monomers are grouped together to reduce clutter. For each of the many-body interactions, the same set of basis functions centered on the subcluster of interest is employed for all the total energy calculations and this introduces ghost functions denoted by the overline in the subscript of total energy terms. Note that, unlike the nuclei-centered basis, the prime symbol is not present here as there is a consistent number of basis functions in each total energy calculation. The binding energy computed using the subcluster basis is

$$\epsilon_{\text{tot}}^S = \sum_{A<B} \binom{n}{2} \epsilon_{AB} + \sum_{A<B<C} \binom{n}{3} \epsilon_{ABC} + \sum_{A<B<C<D} \binom{n}{4} \epsilon_{ABCD} + \dots + \epsilon_{A\cdots N} \quad (5)$$

The ϵ_{tot}^S is known as the Valiron–Mayer function counterpoise (VMFC) corrected binding energy¹⁶ in the literature. In our previous study,¹⁷ we referred to the subcluster basis as the k CBS method as named by Góra et al.²³ The subcluster basis is the standard way of predicting many-body interactions in the construction of ab initio based PES as it is free of BSIE.^{18–21,24} Furthermore, these many-body interactions are reproduced with high accuracy using multipoles and perturbation theory⁵—which are BSSE-free by definition—at intermediate to long intermolecular separations. Unlike the ϵ_{tot} in eq 3, we cannot express ϵ_{tot}^S as the difference between the total energy and the monomer total energies. This is because the sum of the $\epsilon_{A\cdots K}$ does not add up to the total energy; i.e., eq 1 does not hold true here. This is related to the fact that total energies between different $\epsilon_{A\cdots K}$'s cannot be reused. For example, the different E_A and $E_{\overline{AB}}$ are involved in computing ϵ_A and ϵ_{AB} , respectively, whereas the nuclei-centered counterpart would only require the same E_A in both cases. Thus, when the $\epsilon_{A\cdots K}$'s are summed in a MBE according to eq 1, the total energy terms do not cancel to

give the exact total energy eventually. This implies that there are some effects present in the total energy that are not accounted for in the subcluster basis. In fact, this is due to the second component of BSSE—the BSEE.

Apart from the subcluster basis, another common set of basis functions that remove BSIE is one that is centered on the cluster. We denote this as the cluster basis and the k -body interaction can be written as

$$\epsilon_A = E_{\overline{AB\cdots N}} \quad (6a)$$

$$\epsilon_{AB} = E_{\overline{ABC\cdots N}} - (E_{\overline{ABC\cdots N}} + E_{\overline{BAC\cdots N}}) \quad (6b)$$

$$\epsilon_{ABC} = E_{\overline{ABCD\cdots N}} - (E_{\overline{ABCD\cdots N}} + E_{\overline{ACBD\cdots N}} + E_{\overline{BCAD\cdots N}}) + (E_{\overline{ABCD\cdots N}} + E_{\overline{BACD\cdots N}} + E_{\overline{CABD\cdots N}}) \quad (6c)$$

$$\epsilon_{ABCD} = E_{\overline{ABCDE\cdots N}} - (E_{\overline{ABCDE\cdots N}} + E_{\overline{ABDC\cdots N}} + E_{\overline{ACDB\cdots N}} + E_{\overline{BCDA\cdots N}} + E_{\overline{ACDB\cdots N}} + E_{\overline{BCDA\cdots N}}) + (E_{\overline{ABCDE\cdots N}} + E_{\overline{ACBDE\cdots N}} + E_{\overline{ADBCE\cdots N}} + E_{\overline{BCADE\cdots N}} + E_{\overline{BDACE\cdots N}} + E_{\overline{CDABE\cdots N}}) - (E_{\overline{ABCDE\cdots N}} + E_{\overline{BACDE\cdots N}} + E_{\overline{CABDE\cdots N}} + E_{\overline{DABCE\cdots N}}) \quad (6d)$$

For all the total energy calculations, the set of basis functions centered on the entire cluster is employed. Thus, basis functions centered on other monomers surrounding the subcluster of interest are involved, indicated by the overline in the many-body interaction. For example, computing $\epsilon_{\overline{ABC\cdots N}}$ in eq 6b requires total energies involving basis functions centered on $C\cdots N$ surrounding the subcluster AB. The binding energy computed using the cluster basis is

$$\begin{aligned} \epsilon_{\text{tot}}^C &= E_{\text{tot}} - \sum_A \binom{n}{1} E_{\overline{AB\cdots N}} \\ &= \sum_{A<B} \binom{n}{2} \epsilon'_{\overline{ABC\cdots N}} + \sum_{A<B<C} \binom{n}{3} \epsilon'_{\overline{ABCD\cdots N}} \\ &\quad + \sum_{A<B<C<D} \binom{n}{4} \epsilon'_{\overline{ABCDE\cdots N}} + \dots + \epsilon'_{\overline{A\cdots N}} \end{aligned} \quad (7)$$

The ϵ_{tot}^C is named the site–site function counterpoise (SSFC) corrected binding energy.¹² It is commonly referred to simply as the counterpoise (CP) method as the binding energy is a direct generalization of the CP method for a dimer system.¹¹ The cluster basis ensures that a common set of basis functions is employed in each of the total energy calculations, removing the undesirable BSIE. Furthermore, all the total energy terms employ the same basis, allowing for the sum of these many-body interactions to add up to the total energy according to eq 1. Comparing the subcluster basis and cluster basis, there is an additional effect in the latter where the ghost functions surrounding the subcluster, e.g. functions centered on $C\cdots N$ in the case of $\epsilon_{\overline{ABC\cdots N}}$, improve the many-body interaction associated with the subcluster. This is the BSEE. Mathematically, we define the BSEE in the total k -body interaction, $E_{\text{ext}}^{(k)}$, as the difference between the total k -body interaction computed using the cluster basis and subcluster basis

$$E_{\text{ext.}}^{(1)} = \sum_A \binom{n}{1} (\varepsilon_{\overline{AB\cdots N}} - \varepsilon_A) \quad (8a)$$

$$E_{\text{ext.}}^{(2)} = \sum_{A < B} \binom{n}{2} (\varepsilon_{\overline{ABC\cdots N}} - \varepsilon_{AB}) \quad (8b)$$

$$E_{\text{ext.}}^{(3)} = \sum_{A < B < C} \binom{n}{3} (\varepsilon_{\overline{ABCD\cdots N}} - \varepsilon_{ABC}) \quad (8c)$$

$$E_{\text{ext.}}^{(4)} = \sum_{A < B < C < D} \binom{n}{4} (\varepsilon_{\overline{ABCDE\cdots N}} - \varepsilon_{ABCD}) \quad (8d)$$

Unlike the BSIE, the BSEE is important in reproducing the total energy of a cluster. We have previously shown that the MBEs using $\varepsilon_{\overline{A\cdots KL\cdots N}}$ exhibit rapid convergence to the total energy by the four-body interaction.¹⁷ This indicates that the total energy contains the BSEE as part of the variational optimization and/or the perturbative treatment of electron correlation in the electronic structure calculation of the total energy. The borrowing of basis functions from other monomers surrounding the subcluster does improve the flexibility of the wave function of the subcluster and consequently the quality of the many-body interaction computed. On a side note, this is likewise true in valence-bonded systems where the bonding between atoms can be improved by the basis functions from other surrounding atoms. This importance of BSEE also applies to the binding energy where the BSEE should be incorporated into the many-body interactions used to compute the binding energy. Therefore, we consider $\varepsilon_{\text{tot}}^C$ and not $\varepsilon_{\text{tot}}^S$ to be the best estimate of the binding energy at a given level of theory and basis set.

To summarize many-body BSSE, there are two components, namely the BSIE and BSEE. The first component is undesirable, arising from an imbalance in the number of basis functions when computing energy differences in the many-body interactions and the binding energy. This BSIE can be removed by using a common set of basis functions in each of the total energy calculations which can be fulfilled by the use of the subcluster basis or the cluster basis. The second component originates from the extension of the subcluster basis due to the presence of monomers surrounding the subcluster in the cluster. This BSEE is necessary to reproduce the binding energy and total energy of the cluster and can be accounted for using the cluster basis. However, computing the many-body interactions in the cluster basis is very expensive and defeats the usefulness of the MBE in decomposing a large many-body system into manageable few-body subsystems. Thus, we wish to analyze the amount of many-body BSSE present so as to accurately yet cheaply reproduce the binding energy and total energy.

2.3. Many-Ghost Many-Body Expansion. To account for both the BSIE and BSEE (section 2.2), we introduce the many-ghost many-body expansion (MGMBE). The MGMBE defines these two components of many-body BSSE up to the order of truncation of the many-body interactions, allowing us to establish the amount of many-body BSSE present in the many-body interactions. The MGMBE performs a two-dimensional many-body decomposition with each decomposition accounting for one component of many-body BSSE (Figure 2). The

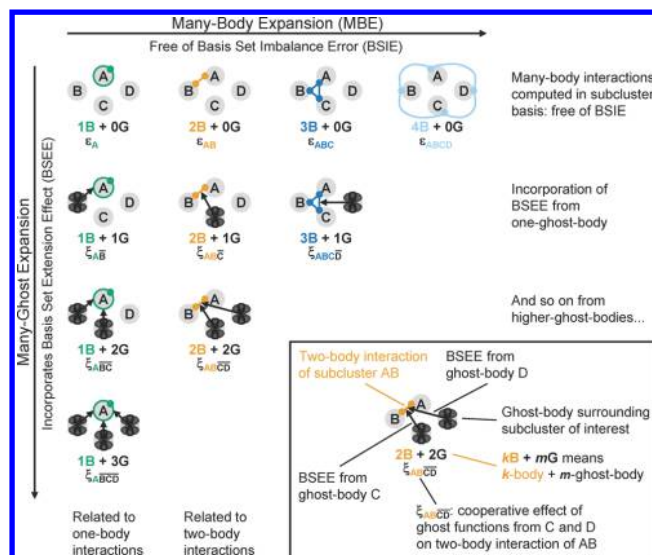


Figure 2. MGMBE performs a two-dimensional many-body decomposition with the first being an MBE (left to right) up to the k -body interaction computed using the subcluster basis, $\varepsilon_{\overline{A\cdots K}}$, using eqs 4a–4d. It is important to note that the BSIE is removed by performing this calculation. A second many-ghost expansion (top to bottom) then decomposes the BSEE from the cluster basis into contributions from m -ghost-body, represented by the black lobes. These BSEE terms, $\xi_{\overline{A\cdots KL\cdots M}}$, are shown here for $k + m \leq 4$, which can be computed using eqs 10a–10f. The inset explains the various symbols in the figure while the comments at the right and bottom edge summarizes the components of the MGMBE along each row and column, respectively. We also note that along the diagonal where $k + m$ is constant, the BSIE and BSEE cancel as these terms share the same basis functions.

first decomposition involves the MBE (section 2.1) using the many-body interactions computed using the subcluster basis, $\varepsilon_{\overline{A\cdots K}}$, ensuring that these interactions are free of the BSIE. The second decomposition, denoted the many-ghost expansion, breaks down the BSEE present in the cluster basis into contributions from one ghost body, two ghost bodies, and so on, up to $(n - k)$ ghost bodies. To reiterate, a ghost body refers to the set of ghost functions centered on a monomer surrounding the subcluster of interest. Both decompositions can be truncated at a low order to hopefully reproduce the binding energy and total energy of the cluster at a low computational cost. We note that the MGMBE is a logical extension of the earlier many-body counterpoise (MBCP) method.^{13,14} The MBCP method seeks to cheaply approximate the $\varepsilon_{\text{tot}}^C$ in eq 7 by performing two separate many-body decompositions on both the E_{tot} and $E_{\overline{AB\cdots N}}$. The former decomposition is an MBE computed using the nuclei-centered basis while the latter decomposition is essentially the many-ghost expansion performed on the monomer total energies. In the MGMBE, we extend the many-ghost expansion for any arbitrary k -body interaction to identify the BSEE present.

Previously, we mentioned that the total energy of a subcluster does not depend solely on the identity of the subcluster of interest but also on the placement of basis functions. To recap, we denoted $E_{\overline{A\cdots KL\cdots M}}$ as the total energy of subcluster $A\cdots K$ calculated in the presence of additional ghost functions centered on $L\cdots M$. The same situation applies to the many-body interactions, evident from the discussion on the nuclei-centered, subcluster and cluster basis in section 2.2. Here, we denote $\varepsilon_{\overline{A\cdots KL\cdots M}}$ as the k -body interaction of the k -

mer subcluster $A\cdots K$ computed using total energies calculated with the set of basis functions centered on $A\cdots KL\cdots M$. The overline in the subscript denotes the ghost bodies, namely the set of ghost functions centered on monomers $L\cdots M$ surrounding the subcluster. For example, $\varepsilon_{AB\overline{CD}} = E_{AB\overline{CD}} - (E_{A\overline{BCD}} + E_{B\overline{ACD}})$. In particular, the many-body interaction computed using the subcluster basis, $\varepsilon_{A\cdots K}$, and the cluster basis, $\varepsilon_{A\cdots KL\cdots N}$, are specific cases of this general notation. In the former, there are no ghost bodies involved, while the entire cluster (excluding the subcluster of interest) constitutes all the ghost bodies in the latter case. Now, we can write the MGMBE of the total energy as

$$E_{\text{tot}} = \sum_A \binom{n}{1} \varepsilon_A + \sum_{A<B} \binom{n}{2} \varepsilon_{AB} + \sum_{A<B<C} \binom{n}{3} \varepsilon_{ABC} + \sum_{A<B<C<D} \binom{n}{4} \varepsilon_{ABCD} + \dots + \sum_{A,B} \binom{n}{1} \binom{n-1}{1} \xi_{A\overline{B}} + \sum_{A<B,C} \binom{n}{2} \binom{n-2}{1} \xi_{A\overline{BC}} + \sum_{A<B<C,D} \binom{n}{3} \binom{n-3}{1} \xi_{A\overline{BCD}} + \dots + \sum_{A,B<C} \binom{n}{1} \binom{n-1}{1} \xi_{A\overline{BC}} + \sum_{A<B,C,D} \binom{n}{2} \binom{n-2}{2} \xi_{A\overline{BCD}} + \dots + \sum_{A,B<C,D} \binom{n}{1} \binom{n-1}{3} \xi_{A\overline{BCD}} + \dots + \xi_{\text{higher}} \quad (9)$$

where $\xi_{A\cdots KL\cdots M}$ is the BSEE from m -ghost-body $L\cdots M$ in the k -body interaction of k -mer subcluster $A\cdots K$, of which there are $\binom{n}{k} \cdot \binom{n-m}{m-k}$ of such terms. The first line in eq 9 gives the MBE using many-body interactions computed using the subcluster basis using eqs 4a–4d). While these many-body interactions are free of BSIE, they lack the important BSEE. These missing BSEE terms are added in the following lines with each line introducing contributions from a different number of ghost bodies. For cases where $k+m \leq 4$, $\xi_{A\cdots KL\cdots M}$ can be expressed as

$$\xi_{A\overline{B}} = \varepsilon_{A\overline{B}} - \varepsilon_A \quad (10a)$$

$$\xi_{A\overline{BC}} = \varepsilon_{A\overline{BC}} - \varepsilon_{AB} \quad (10b)$$

$$\xi_{A\overline{BCD}} = \varepsilon_{A\overline{BCD}} - \varepsilon_{ABC} \quad (10c)$$

$$\xi_{A\overline{BC}} = \varepsilon_{A\overline{BC}} - (\varepsilon_{A\overline{B}} + \varepsilon_{A\overline{C}}) + \varepsilon_A \quad (10d)$$

$$\xi_{A\overline{BCD}} = \varepsilon_{A\overline{BCD}} - (\varepsilon_{A\overline{BC}} + \varepsilon_{A\overline{BD}}) + \varepsilon_{AB} \quad (10e)$$

$$\xi_{A\overline{BCD}} = \varepsilon_{A\overline{BCD}} - (\varepsilon_{A\overline{BC}} + \varepsilon_{A\overline{BD}} + \varepsilon_{A\overline{CD}}) + (\varepsilon_{A\overline{B}} + \varepsilon_{A\overline{C}} + \varepsilon_{A\overline{D}}) - \varepsilon_A \quad (10f)$$

The meaning of these terms can be better understood by looking at specific examples. For example, $\xi_{A\overline{BCD}}$ in eq 10c quantifies the amount by which the ghost functions centered on D affect the three-body interaction of ABC, i.e., the BSEE from D on ε_{ABC} . Likewise, $\xi_{A\overline{BCD}}$ in eq 10e gives the cooperative effect of the ghost functions centered on both C and D on the two-body interaction of AB, and higher-ghost-body BSEEs can be interpreted similarly. These terms represent the many-body decomposition of the BSEE present in the cluster basis. As such, eqs 10a–10c, eqs 10d and 10e, and eq 10f resemble eq 2a, eq 2b, and eq 2c, respectively. Upon comparison between the two sets of equations, there is an additional $\varepsilon_{A\cdots C}$ term (last

term in each equation) in eq 10. This is the 0-ghost-body term where there is no BSEE and the equivalent in a MBE corresponds to a 0-body interaction which is zero and thus omitted in the many-body interaction expressions. In order to compute the $\xi_{A\cdots KL\cdots M}$ terms, all the $\varepsilon_{A\cdots KL\cdots M}$ terms have to be expressed in terms of total energies that can be readily obtained from electronic structure calculations. Here, we give an example where we express $\xi_{A\overline{BCD}}$ in terms of total energies

$$\begin{aligned} \xi_{A\overline{BCD}} &= \varepsilon_{A\overline{BCD}} - \varepsilon_{A\overline{BC}} - \varepsilon_{A\overline{BD}} + \varepsilon_{AB} \\ &= (E_{A\overline{BCD}} - E_{A\overline{BCD}} - E_{B\overline{ACD}}) \\ &\quad - (E_{A\overline{BC}} - E_{A\overline{BC}} - E_{B\overline{AC}}) \\ &\quad - (E_{A\overline{BD}} - E_{A\overline{BD}} - E_{B\overline{AD}}) + (E_{AB} - E_{AB} - E_{B\overline{A}}) \end{aligned} \quad (11)$$

From eq 11, we observe that the maximum number of basis functions is limited to that of four monomers in computing $\xi_{A\overline{BCD}}$. In fact, the maximum number of basis functions is limited to $(k+m)$ monomers in computing $\xi_{A\cdots KL\cdots M}$. It is also obvious that rewriting the $\xi_{A\cdots KL\cdots M}$ terms in terms of total energies can lead to cumbersome expressions. Fortunately, many total energy terms are repeated across different $\xi_{A\cdots KL\cdots M}$'s which can be collected to give a more compact expression when all the $\xi_{A\cdots KL\cdots M}$ terms are summed. The derivation of these working equations is presented in the Supporting Information.

The two many-body decompositions in the MGMBE can be truncated at a low order to hopefully reproduce the binding energy and total energy of a cluster. Given that the two decompositions are independent, the BSEE present in each of the k -body interactions can be truncated at a different m ghost body. A prudent choice would be to truncate at order (k, m) such that $k+m = \alpha$, keeping the maximum number of basis functions in each electronic structure calculation to that of α monomers. For example, truncating the MGMBE at $\alpha = 2$ would include the ε_A , ε_{AB} , and $\xi_{A\overline{B}}$ terms while truncation at $\alpha = 3$ includes the previously mentioned terms as well as ε_{ABC} , $\xi_{A\overline{BC}}$, and $\xi_{A\overline{BC}}$ terms.

A surprising result surfaced when the truncation order of the MGMBE is such that $k+m = \alpha$. Careful analysis of the working equations in the Supporting Information revealed that all the total energies involving any ghost functions vanish when we sum the $\varepsilon_{A\cdots K}$ and $\xi_{A\cdots KL\cdots M}$ terms with $k+m = \alpha$, where α is a constant. Consequently, we obtain the many-body interactions computed using the nuclei-centered basis from this summation. This point is worth emphasizing, which is the inclusion of BSEE terms in the MGMBE cancels exactly all of the total energies involving any ghost functions in eq 9. This implies that an MBE using the nuclei-centered basis truncated at α bodies incorporates some BSEE, in particular contributions from up to $m = (\alpha - k)$ ghost bodies in each of the k -body interactions. We stress that this surprising result only occurs when the MGMBE terms are summed across different k number of interacting bodies to obtain either the binding energy or total energy. To illustrate this cancellation, let us consider the sum of $\xi_{A\overline{B}}$, $\xi_{B\overline{A}}$, and ε_{AB} . The first two terms would be $\xi_{A\overline{B}} = E_{A\overline{B}} - E_A$ and $\xi_{B\overline{A}} = E_{B\overline{A}} - E_B$, respectively, and the total energies involving ghost functions would be eliminated when we include the $\varepsilon_{AB} = E_{AB} - E_{A\overline{B}} - E_{B\overline{A}}$. Thus, we are left with the $\varepsilon'_{AB} = E_{AB} - E_A - E_B$. In essence, the BSEE in $\xi_{A\cdots KL\cdots M}$ replaces the total energy terms involving ghost functions in the $\varepsilon_{A\cdots K}$ with corresponding ghost-free terms, “transforming” it into the nuclei-centered counterpart, $\varepsilon'_{A\cdots K}$. Expressed alternatively—the

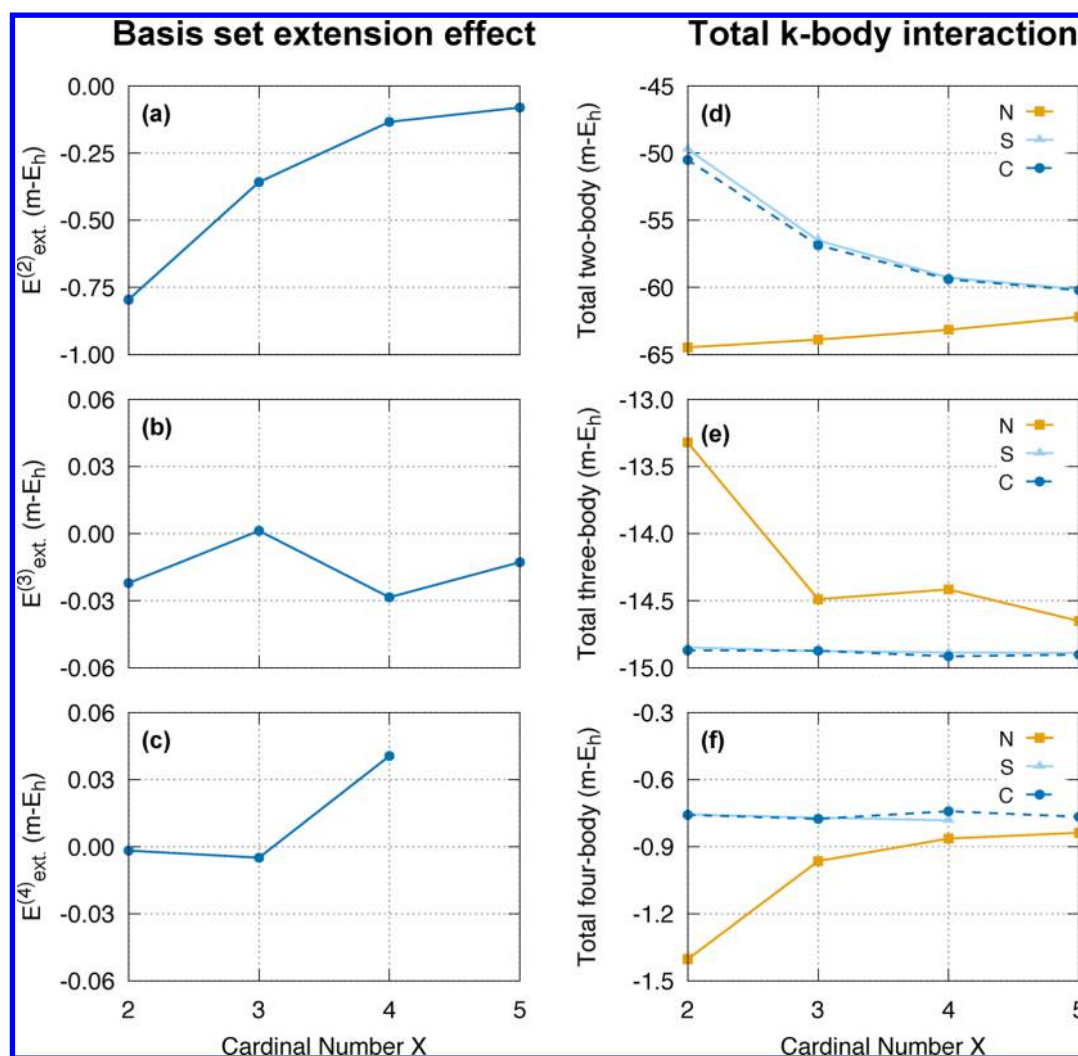


Figure 3. (a–c) BSEE in the total k -body interaction, $E_{\text{ext}}^{(k)}$, as defined in eq 8 and the (d–f) total k -body interaction for the cage isomer of $(\text{H}_2\text{O})_6$ with increasing basis set quality at MP2/AVXZ. The total k -body interactions are computed using various *location* bases, namely the nuclei-centered (N), subcluster (S), and cluster (C) basis described in section 2.2 to determine the effects of many-body BSSE on the many-body interactions. In particular, the lines for the cluster basis are dashed to show clearly the similarities between that and the subcluster basis results. The $E_{\text{ext}}^{(4)}$ and total four-body interaction computed using the subcluster basis at MP2/AVSZ are omitted due to steep computational cost.

BSIE for a higher-body interaction (something that must be subtracted) is an BSEE for a lower-body interaction (something that must be added)—and the two effects cancel each other exactly!

3. RESULTS AND DISCUSSION

All quantum chemical calculations were performed using the MOLPRO suite of programs.²⁵ Calculations were carried out at the second-order Møller–Plesset perturbation (MP2) level of theory using the aug-cc-pVnZ, labeled AVnZ, $n = \text{D, T, Q, 5}$, Dunning correlation-consistent basis sets augmented with diffuse functions.^{26,27} The explicitly correlated MP2 (MP2-F12) level of theory²⁸ was employed with the AVDZ basis set.

3.1. Basis Set Extension Effect in Many-Body Interactions. In our previous work, we observed rapid convergence in the MBEs using either the subcluster or cluster basis. This indirectly suggests that the difference between these two MBEs—the BSEE—should converge rapidly with the number of bodies. We computed the $E_{\text{ext}}^{(k)}$ and total k -body interaction for the $(\text{H}_2\text{O})_6$ cage and prism isomers up to the four-body term with increasing basis set quality. Both isomers

are taken from Richard et al.¹⁴ and showed similar trends. Thus, the data for the cage isomer are shown in Figure 3, while the prism isomer counterparts are in Figure S1 of the Supporting Information. Similar studies exist in the literature^{29,30} but are performed on small trimer and tetramer clusters^{29,30} or focused on the binding energy.^{14,31} Instead, we choose to separately examine the BSEE in each of the total k -body interactions, especially between the two-body and the three-and-higher-body interactions, because they are dominated by different intermolecular interactions.⁵

At the two-body level (Figure 3a), the $E_{\text{ext}}^{(2)}$ is always negative, indicating that the additional ghost functions in the cluster basis help to lower the two-body interactions. As expected, increasing the quality of the basis set decreases this borrowing of basis functions to improve the two-body interactions. These BSEEs are generally small, below 1 $m-E_h$, because the additional basis functions in the cluster basis are not centered on the nuclei or on regions between nuclei where the interaction occurs. This is in contrast to the use of midbond functions where the placement of basis functions at regions between interacting molecules improves the description of the interaction.³² We

point out that the $E_{\text{ext}}^{(k)}$ also serves as an error indicator of how well the many-body interactions computed using the subcluster basis can be used in place of the cluster basis counterpart to reproduce the binding energy or total energy. Thus, the $E_{\text{ext}}^{(2)}$ can still be substantial if very high accuracy is demanded. For the higher-body interactions (Figure 3b,c), the $E_{\text{ext}}^{(3)}$ and $E_{\text{ext}}^{(4)}$ are minuscule—smaller than 0.045 m-E_h —and we can treat the many-body interactions computed in both the subcluster and cluster basis to be practically the same. This is of comfort as the use of the subcluster basis renders the construction of MBE-based ab initio water potentials^{24,33} possible. The reduction in dimensionality from applying the MBE is preserved unlike the cluster basis which depends on the geometry of the cluster. Indeed, the many-body interactions computed using the subcluster basis were used to construct ab initio based PES to study large water clusters and bulk water.^{18–21}

The tiny $E_{\text{ext}}^{(3)}$ and $E_{\text{ext}}^{(4)}$ bring us to an unrelated but important result. At the complete basis set (CBS) limit, there is no BSSE, i.e., $E_{\text{ext}}^{(k)} = 0$. While the converse is not necessarily true, it is worthwhile to investigate if the CBS limit can be approximated using moderate-sized basis sets. Clearly, this is true for the three-body (Figure 3e) and four-body interactions (Figure 3f). Both the total three-body and four-body interaction computed using the subcluster or cluster basis (light and dark blue lines) appear to have converged, presumably to the CBS limit, varying by $0.005\text{--}0.015 \text{ m-E}_h$. This was mentioned in passing recently in the construction of an ab initio water PES where the three-body interactions computed using the subcluster basis at CCSD(T)/AVTZ are very similar to the CBS limit values.²⁰ With the removal of BSIE, we only require an AVDZ basis set to obtain CBS limit three-body and four-body interactions. This result implies that primarily the convergence of the total energies with increasing basis set quality comes from changes in the one-body and two-body interactions. Thus, we can obtain the total energies of water clusters with increasing basis set quality by recalculating the one-body and two-body interactions at the respective basis sets. The fact that three-and-higher-body interactions do not require a large basis set to achieve the CBS limit eliminates the need for extrapolation or ad hoc measures as commonly employed for two-body interactions. The ad hoc methods involve taking a fraction of the two-body interaction computed using the subcluster and nuclei-centered basis,^{34–37} motivated by the well-documented trend^{34,36–38} that these two quantities converge to the CBS limit from above and below, respectively (Figure 3d).

As with our previous work,¹⁷ there is no guarantee that the observations made on the small $(\text{H}_2\text{O})_6$ clusters still hold true when larger clusters are studied. To this end, we computed the BSEE up to the four-body interaction for a homologous series of optimized $(\text{H}_2\text{O})_{8-16}$ clusters taken from Maheshwary et al.,³⁹ which is presented together with the hexamer results (Figure 4). Since various cluster sizes are involved, all the energies reported henceforth will be on a per monomer basis. Calculations were performed at MP2/AVDZ and MP2/AVTZ. The explicitly correlated MP2-F12 theory model²⁸ was also employed with the AVDZ basis set as this combination typically yielded results of MP2/AVQZ quality,⁴⁰ complementing the MP2/AVDZ and MP2/AVTZ results.

As mentioned earlier, $E_{\text{ext}}^{(k)}$ serves as an error indicator of how well the cheaper subcluster basis can be used in place of the more expensive cluster basis. Here, we wish to clarify what we deem to be an acceptable value for $E_{\text{ext}}^{(k)}$. Studies on atomization energies and reaction enthalpies often require calculations to

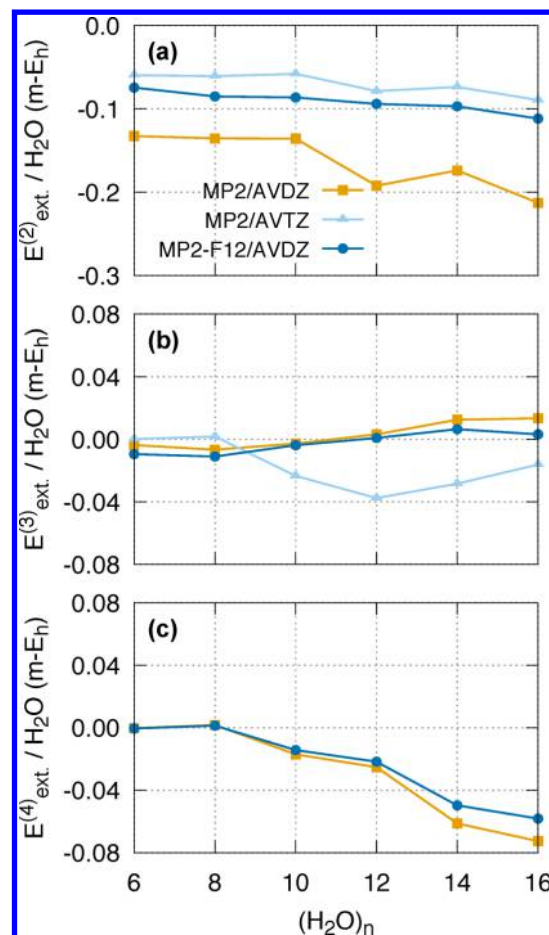


Figure 4. BSEE in the total k -body interaction, $E_{\text{ext}}^{(k)}$, per (H_2O) monomer for water clusters of increasing size, $(\text{H}_2\text{O})_{6-16}$, computed at MP2 levels of theory with various basis sets. The results for $E_{\text{ext}}^{(4)}$ are not available at the MP2/AVTZ level due to the steep computational cost involved in computing the four-body interactions.

agree with experiments within chemical accuracy, which is 4.2 kJ mol^{-1} or 1.6 m-E_h .⁴¹ However, the MBE is often used to study the dynamical evolution of large molecular clusters and do not involve bond breaking. As such, we introduce the “dynamical accuracy” where the error for large clusters is computed on a per monomer basis as the properties derived from dynamical simulations are intensive in nature. A suitable dynamical accuracy might be 10% of the average molecular kinetic energy at room temperature, $(3/2)kT$, which is about 0.14 m-E_h or 0.37 kJ mol^{-1} .

From Figure 4a, we again observe that the $E_{\text{ext}}^{(2)}$ decreases with increasing basis set quality where the use of the higher quality MP2/AVTZ (light blue line) or the explicitly correlated MP2-F12/AVDZ (dark blue line) halved the small $E_{\text{ext}}^{(2)}$ present in MP2/AVDZ (orange line). While the $E_{\text{ext}}^{(2)}$ per monomer at MP2/AVTZ falls within dynamical accuracy, the BSEE exhibits a slow increase with increasing cluster size. Fortunately, due to the small system size, this small $E_{\text{ext}}^{(2)}$ can be practically eliminated through the use of larger basis sets or CBS extrapolation. Indeed, CBS extrapolation is routinely applied to two-body interactions employed in ab initio two-body water potentials.^{21,42,43} At the higher-body level, we confirmed that the $E_{\text{ext}}^{(3)}$ and $E_{\text{ext}}^{(4)}$ are, if not negligible, then acceptable. The $E_{\text{ext}}^{(3)}$ is insignificant, always below 0.040 m-E_h per monomer (Figure 4b). The $E_{\text{ext}}^{(4)}$ shows an increasing trend with increasing cluster

size (Figure 4c). Nonetheless, the value is quite small (<0.080 m- E_h per monomer) and would be even smaller if a larger basis set such as AVTZ is used. Furthermore, there would be some partial cancellation of the BSEE when the three-body and four-body interactions are summed. Therefore, we conclude that the cheaper subcluster basis can be employed in computing three-body and four-body interactions in place of their more expensive cluster basis counterpart.

Next, we overlaid the total k -body interaction computed using the subcluster basis at different basis set quality (Figure 5). The cluster basis counterpart shows identical trends and is

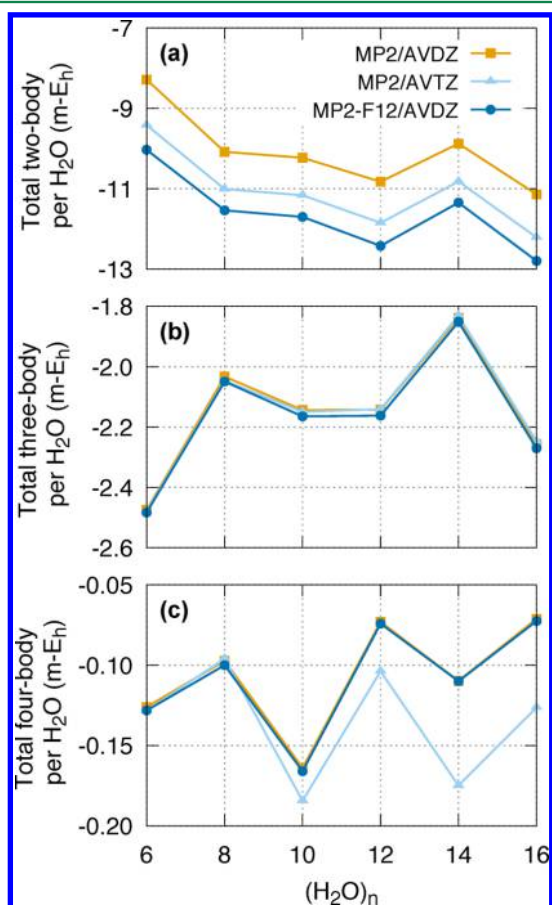


Figure 5. Comparison of the total k -body interaction computed using the subcluster basis per (H_2O) monomer for water clusters of increasing size, $(\text{H}_2\text{O})_{6-16}$, computed at various levels of theory and basis sets.

presented in Figure S2 of the Supporting Information. It is clear that the total three-body and four-body interactions remain the same regardless of the basis set used (Figure 5b,c). The total four-body interaction at MP2/AVTZ (light blue line) appears to be different due to the scale of the energy axis, which exaggerates the small difference (<0.055 m- E_h) between the MP2/AVTZ and MP2/AVDZ values. The total two-body interaction (Figure 5a) becomes more stabilizing with increasing basis set quality, echoing the hexamer results.

In summary, we made three key observations: (i) the $E_{\text{ext}}^{(2)}$ is small but significant and diminishes with increasing basis set quality, (ii) the $E_{\text{ext}}^{(3)}$ and $E_{\text{ext}}^{(4)}$ are much smaller, supporting the use of the cheaper subcluster basis to compute the three-body and four-body interactions, and (iii) the three-body and four-

body interactions computed using the subcluster basis have converged to the CBS limit using an AVDZ basis set.

3.2. Many-Ghost Many-Body Expansion of the Total Energy. CBS extrapolation at the two-body level would eliminate the $E_{\text{ext}}^{(2)}$, which can be coupled with AVDZ-quality three-body and four-body interactions to yield binding energies of CBS quality without using the expensive cluster basis. Apart from that, one may also be interested in reproducing the total energy at a particular basis set. This is useful in assessing the accuracy of fragmentation methods,¹⁻⁴ such as the Combined Fragmentation Method,⁴⁴⁻⁴⁶ where small groups of adjacent atoms are treated as bodies and selected many-body interactions are computed to approximate the total energies of large chemical systems. We employed the MGMBE truncated at different order in an attempt to reproduce the total energy (Table 3). In this way, we can determine whether the omission of certain BSEEs affects the accuracy of the predicted total energy.

From Table 3, including the BSEE from a one-ghost-body into the one-body interaction decreased the error by 1 order of magnitude as seen in the entry $\{1, 0, 0, 0\}$. This suggests that the one-body interaction is very sensitive to the BSEE. This is not surprising as the one-body interaction constitutes the majority ($\approx 99.98\%$) of the total energy. The error decreased again when more BSEE was incorporated (entry $\{2, 1, 0, 0\}$). However, from entry $\{2, 1, 0, 0\}$ to $\{3, 2, 1, 0\}$, further inclusion of BSEE resulted in a larger error. Hypothesizing that this could be due to the BSEE in the one-body interaction, we varied the truncation order of the BSEE in the one-body interaction (entry $\{1, 2, 1, 0\}$, $\{2, 2, 1, 0\}$, and $\{3, 2, 1, 0\}$) and observed a fluctuation in the error. While the data are not shown here, the error actually oscillates wildly, changing in sign from positive (entry $\{1, 2, 1, 0\}$) to negative (entry $\{2, 2, 1, 0\}$) and back to positive again (entry $\{3, 2, 1, 0\}$). Recall the surprising result in section 2.3 that the MBE using the nuclei-centered basis truncated at the α -body term contains the BSEE from up to $m = (\alpha - k)$ ghost-bodies in each of the k -body interactions. This suggests that the similar oscillatory behavior reported previously¹⁷ in the MBEs using the nuclei-centered basis could be due to the BSEE present in the one-body interaction. To determine if the two oscillatory behaviours are related, we compared the convergence of the MBE using the nuclei-centred basis to the total energy of the cluster with that of the MGMBE of the one-body interaction to the total one-body interaction in the cluster basis (Figure 6).

It is clear from Figure 6 that the two many-body decompositions are practically identical except for the first two data points. It appears to be the case that the poor convergence of the MBE using the nuclei-centered basis is almost completely caused by the BSEE in the one-body interaction. The differences in the first two data points is because the MBE (Figure 6a) includes the actual many-body interactions together with the BSEE. In the first two data points, there are additional errors in the MBE associated with neglecting these many-body interactions. From the four-body term onward, the majority of the many-body interactions are accounted for, and virtually all the remaining error is apparently due to BSEE in the one-body interaction.

The errors in the MBE associated with the BSEE not only applies to the brute force computation of all the $\binom{n}{k}$ individual k -body interactions but also to “internally consistent” selected many-body interactions virtually always employed in fragmen-

Table 3. Root Mean Square of the Error per (H₂O) Monomer (RMSE, $m-E_h$) and Maximum Absolute Error per (H₂O) Monomer (MxAE, $m-E_h$), in Reproducing the Total Energy for a Series of Optimized Water Clusters from Figure 4 Calculated at MP2/AVDZ, MP2/AVTZ and MP2-F12/AVDZ^a

m -ghost-body in k -body ^b				MP2/AVDZ		MP2/AVTZ		MP2-F12/AVDZ	
$k = 1$	$k = 2$	$k = 3$	$k = 4$	RMSE	MxAE	RMSE	MxAE	RMSE	MxAE
0	0	0	0	2.505	2.779	1.286	1.402	0.811	0.868
1	0	0	0	0.122	0.162	0.065	0.110	0.431	0.518
2	1	0	0	0.101	0.150	0.032	0.060	0.267	0.343
1	2	1	0	0.280	0.324	0.091	0.105	0.510	0.595
2	2	1	0	0.145	0.197	0.066	0.095	0.338	0.461
3	2	1	0	0.320	0.463	0.088	0.141	0.298	0.462
all	all	all	all	0.044	0.075	— ^c	— ^c	0.037	0.062

^aThe MGMBE includes up to the four-body interaction ($k = 1-4$) of which the BSEEs are truncated at different m -ghost bodies. Error here is defined as the total energy of the cluster minus the MGMBE-predicted total energy. The error per (H₂O) monomer is first obtained before the RMS or maximum is taken. ^bThe digits give the highest number of ghost bodies, m , that are incorporated into the k -body interaction using the MGMBE, and “all” refers to the cluster basis which includes all the BSEEs. For example, the second entry, {1, 0, 0, 0}, indicates that the BSEE from up to one ghost body is incorporated in the one-body interactions, and there are no BSEEs included for the two-to-four-body interactions. ^cAs the four-body interaction computed using the cluster basis is computationally expensive at MP2/AVTZ, an estimate of the total energy is unavailable.

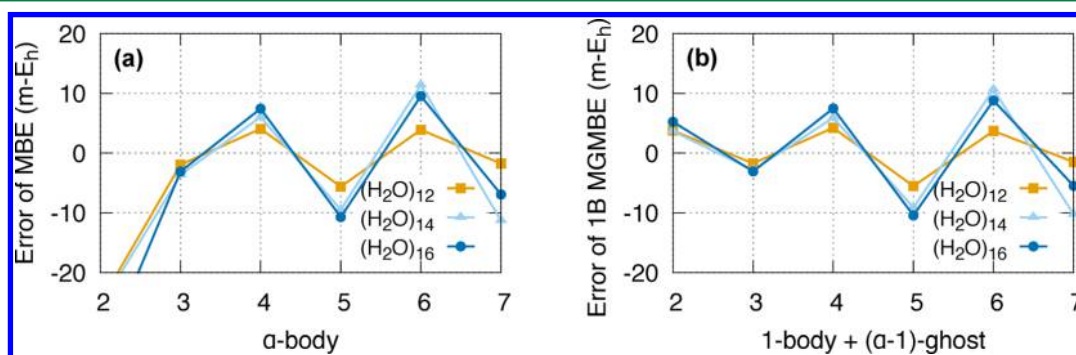


Figure 6. (a) The error of the MBE of the total energy using the nuclei-centered basis truncated at the α -body term follows an almost identical trend as (b) the error of the MGMBE of the total one-body interaction in the cluster basis truncated at the $(\alpha - 1)$ -ghost-body term. The calculations were performed at MP2/AVDZ for (H₂O)₁₂, (H₂O)₁₄, and (H₂O)₁₆ from Figure 4. The error of the MBE is defined as the difference between the total energy, E_{tot} , and the sum of the total one-body interaction up to the total α -body interaction. Similarly, the error of the MGMBE is defined as the difference between the one-body interaction computed using the cluster basis, $E_{A \text{ B-N}}$, and the sum of the BSEE from up to $(\alpha - 1)$ ghost bodies summed across all the monomers.

Table 4. Root Mean Square of the Error per (H₂O) Monomer (RMSE, $m-E_h$) and Maximum Absolute Error per (H₂O) Monomer (MxAE, $m-E_h$), in Reproducing the ϵ_{tot}^C for the Same Water Clusters in Table 3^a

m -ghost-body in k -body ^b			MP2/AVDZ		MP2/AVTZ		MP2-F12/AVDZ	
$k = 2$	$k = 3$	$k = 4$	RMSE	MxAE	RMSE	MxAE	RMSE	MxAE
0	0	0	0.160	0.197	0.046	0.058	0.090	0.104
1	0	0	0.046	0.070	0.073	0.114	0.060	0.090
2	1	0	0.009	0.015	0.019	0.043	0.015	0.028
all	all	all	0.044	0.075	— ^c	— ^c	0.037	0.062

^aThe MGMBE includes up to the four-body interaction ($k = 2-4$) of which the BSEEs are truncated at different m ghost bodies. Error here is defined as the ϵ_{tot}^C of the cluster minus the MGMBE-predicted ϵ_{tot}^C . The error per (H₂O) monomer is first obtained before the RMS or maximum is taken. ^bThe digits give the highest number of ghost bodies, m , that are incorporated into the k -body interaction using the MGMBE, and “all” refers to the cluster basis which includes all the BSEE. For example, the second entry, {1,0,0}, indicates that the BSEE from up to one ghost body is incorporated in the two-body interactions, and there are no BSEEs included for the three-body and four-body interactions. ^cAs the four-body interaction computed using the cluster basis is computationally expensive at MP2/AVTZ, an estimate of the ϵ_{tot}^C is unavailable.

tation methods. These interactions are “internally consistent” in a sense that the many-body interactions of the selected fragments (interacting groups of atoms) and their constituent lower-body interactions are included and only included once. This allows for the BSIE and BSEE to cancel. The poor convergence of the MBE/MGMBE allows us to explain certain observations in fragmentation methods. “Grafting” is employed in some fragmentation methods^{47–49} where the total energy of the system is calculated at a lower level of theory or basis set to

serve as a correction to the predicted total energy. Such grafting approaches not only correct for missing important many-body interactions but also account for the BSEE to a large extent, explaining the low errors associated with these methods. Since the BSEE converges poorly with respect to the number of ghost bodies, the expensive one-body interactions computed using the cluster basis are required to accurately reproduce the total energy. Future investigations to develop cheaper alternatives to the cluster basis will be undertaken. One possibility includes the

omission of certain basis functions from the basis set, in particular the tight valence-type functions (i.e., not diffuse functions), on ghost bodies that are far away from the monomer of interest. Thus, only the contributing diffuse functions remain.

3.3. Many-Ghost Many-Body Expansion of the Binding Energy. We have shown that the poor convergence of the MBE using the nuclei-centered basis is caused by the BSEE in the one-body interactions. With the removal of the one-body interactions and its associated BSEE, we would expect the remaining energy to converge rapidly with the number of bodies. This remaining energy is the binding energy, $\epsilon_{\text{tot}}^{\text{C}}$ and the accuracy of the MGMBE is evaluated in Table 4.

From Table 4, the incorporation of the BSEE greatly reduces the error in reproducing the $\epsilon_{\text{tot}}^{\text{C}}$ eventually giving a tiny error per monomer of below 0.015–0.043 m- E_{h} (entry {2, 1, 0}), which is well within dynamical accuracy. In fact, entry {2, 1, 0} gives a lower error than entry {all, all, all} which incorporates all the BSEE in the two-to-four-body interactions. This can be attributed to a reversal in the sign of the error. In entry {0, 0, 0}, the absence of BSEE which stabilizes the binding energy results in negative errors. On the other hand, the errors from the incorporation of all the BSEE up to the four-body interaction in entry {all, all, all} are positive due to the neglect of higher-than-four-body interactions. Thus, there is some form of error cancellation between the two factors when the majority of the BSEE is accounted for in entry {2, 1, 0}. Furthermore, the maximum number of basis functions ever employed in any total energy calculation is limited to that of four monomers in entry {2, 1, 0}, originating from either the interacting bodies or ghost bodies. This allows for expensive theoretical models such as the Coupled Cluster Singles and Doubles with perturbative Triples [CCSD(T)] to be applied to obtain highly accurate $\epsilon_{\text{tot}}^{\text{C}}$ for large clusters or even bulk-water simulations. It should be emphasized that no charge embedding scheme^{9,50,51} was used, although they are commonly applied to water clusters. The use of such schemes is prevalent in the literature due to the belief that the water–water interactions are highly many-body in nature. However, our results indicate that we only require up to the four-body interactions. It is likely that any apparent higher-than-four-body effects are caused by the BSEE in the one-body interactions which we have shown to be highly many-body in nature (Figure 6).

Notably, the calculations involved in entry {2, 1, 0} are equivalent to that in a MBCEP(4) calculation.^{13,14} A MBCEP(4) calculation would involve a MBE using the nuclei-centered basis truncated at the four-body term minus the one-body interactions with the BSEE truncated at the $(4 - 1) = 3$ -ghost-body level. This is equivalent to a “{3, 2, 1, 0}” MGMBE calculation of the total energy minus the one-body interactions and its associated BSEE, i.e., entry {2, 1, 0} in Table 4. Thus, an “{ $\alpha - 2$, $\alpha - 1$, ..., 0}” MGMBE calculation of the $\epsilon_{\text{tot}}^{\text{C}}$ is identical to an MBCEP(α) calculation. Note that for the MBCEP method, the MGMBE has to be truncated at one order less than that of the MBE. This is important to ensure that all the BSEE in the one-body interactions are properly removed, and this requirement only becomes obvious with the analysis of the BSEE using the MGMBE presented in this work.

4. CONCLUSION

Through a systematic study of water clusters with improving basis set and increasing cluster size, we concluded that one has to account for many-body BSSE in order to reproduce the

many-body interactions computed using the cluster basis. There are two distinct components to the many-body BSSE. The first arises due to an imbalance in the number of basis functions used to compute a particular k -body interaction. In this case, the k -body total energy calculation utilizes many more basis functions than does the lower-body counterparts which are necessary to extract the k -body interaction. The second arises due to the fact that a k -body within a much larger cluster is further stabilized by the basis functions of the surrounding bodies denoted as the BSEE. If one wants to reproduce the binding energy and/or the total energy through a many-body approach, the first BSIE is undesirable as it leads to erroneous many-body interactions. However, the BSEE is important as these extension effects improve the quality of the total energy or binding energy by maximizing the flexibility of the wave function at the given basis set. Thus, the best estimate of the binding energy at a given basis set would be the total energy minus the one-body intramolecular interactions computed using the cluster basis.

We found that both components of the many-body BSSE are accounted for in the three-body and four-body interactions computed using the subcluster basis and that these interactions appear to have converged to the CBS limit at the AVDZ level. For the two-body interactions, and particularly for the one-body intramolecular interactions, important BSEEs are significant and have to be accounted for, thus making the use of the subcluster basis insufficient. To account for both the BSIE and the BSEE, we introduce the MGMBE in this work. The MGMBE performs a two-dimensional many-body decomposition with each decomposition accounting for one component of many-body BSSE. Through the MGMBE of the total energy, we found that the oscillatory behavior encountered in MBEs using diffuse functions is caused by the BSEE in the one-body interactions. With the adequate removal of the one-body interactions and the associated BSEE, the MGMBE successfully reproduces the binding energies of clusters using numerous small calculations that involve no more than four monomers.

Despite the utility of decomposing a large cluster into small subsystems, the MBE and the MGMBE come with a limitation. The number of four-body calculations increases quartically with the cluster size, substantially hindering the scalability of these methods. To circumvent this, a forthcoming publication will establish a rigorous criterion to select out all potentially significant many-body interactions.

■ ASSOCIATED CONTENT

● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00343.

The derivation of the working equations for the MBE (section 2.1) and the MGMBE (section 2.3); Figures S1 and S2, showing the results for the prism isomer counterpart to Figure 3 and the cluster basis counterpart to Figure 5 respectively; and Cartesian coordinates for the water clusters studied in this work (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +65 6516 2846. Fax: +65 6779 1691. E-mail: ryan.pa.bettens@nus.edu.sg.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the National University of Singapore's support from the Academic Research Fund, grant number R-143-000-549-112. The authors also thank the Centre for Computational Science and Engineering for the use of their computers.

REFERENCES

- (1) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632–672.
- (2) Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.
- (3) Collins, M. A.; Bettens, R. P. A. *Chem. Rev.* **2015**, *115*, 5607–5642.
- (4) Raghavachari, K.; Saha, A. *Chem. Rev.* **2015**, *115*, 5643–5677.
- (5) Stone, A. J. *The Theory of Intermolecular Forces*, 2nd ed.; Oxford University Press: Oxford, England, 2013.
- (6) Kaplan, I. G.; Santamaria, R.; Novaro, O. *Mol. Phys.* **1995**, *84*, 105–114.
- (7) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523–7534.
- (8) Pedulla, J.; Kim, K.; Jordan, K. *Chem. Phys. Lett.* **1998**, *291*, 78–84.
- (9) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.
- (10) Liu, B.; McLean, A. D. *J. Chem. Phys.* **1973**, *59*, 4557–4558.
- (11) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (12) Wells, B. H.; Wilson, S. *Chem. Phys. Lett.* **1983**, *101*, 429–434.
- (13) Richard, R. M.; Lao, K. U.; Herbert, J. M. *J. Phys. Chem. Lett.* **2013**, *4*, 2674–2680.
- (14) Richard, R. M.; Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2013**, *139*, 224102.
- (15) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1993**, *97*, 2488–2490.
- (16) Valiron, P.; Mayer, I. *Chem. Phys. Lett.* **1997**, *275*, 46–55.
- (17) Ouyang, J. F.; Cvitkovic, M. W.; Bettens, R. P. A. *J. Chem. Theory Comput.* **2014**, *10*, 3699–3707.
- (18) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2011**, *134*, 094509.
- (19) Medders, G. R.; Babin, V.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 1103–1114.
- (20) Babin, V.; Medders, G. R.; Paesani, F. *J. Chem. Theory Comput.* **2014**, *10*, 1599–1607.
- (21) Góra, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. *J. Chem. Phys.* **2014**, *140*, 194101.
- (22) Hermann, A.; Krawczyk, R. P.; Lein, M.; Schwerdtfeger, P. *Phys. Rev. A: At., Mol., Opt. Phys.* **2007**, *76*, 013202.
- (23) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. *J. Chem. Phys.* **2011**, *135*, 224102.
- (24) Szalewicz, K.; Leforestier, C.; van der Avoird, A. *Chem. Phys. Lett.* **2009**, *482*, 1–14.
- (25) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *MOLPRO*, version 2012.1; Cardiff University: Cardiff, U. K.; Universität Stuttgart: Stuttgart, Germany, 2012. See <http://www.molpro.net>.
- (26) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (27) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (28) Werner, H.-J.; Adler, T. B.; Manby, F. R. *J. Chem. Phys.* **2007**, *126*, 164102.
- (29) Lendvay, G.; Mayer, I. *Chem. Phys. Lett.* **1998**, *297*, 365–373.
- (30) Salvador, P.; Szczyński, M. M. *J. Chem. Phys.* **2003**, *118*, 537–549.
- (31) Kamiya, M.; Hirata, S.; Valiev, M. *J. Chem. Phys.* **2008**, *128*, 074103.
- (32) Williams, H. L.; Mas, E. M.; Szalewicz, K.; Jeziorski, B. *J. Chem. Phys.* **1995**, *103*, 7374–7391.
- (33) Ouyang, J. F.; Bettens, R. P. A. *Chimia* **2015**, *69*, 104–111.
- (34) Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157–9167.
- (35) Shields, R. M.; Temelso, B.; Archer, K. A.; Morrell, T. E.; Shields, G. C. *J. Phys. Chem. A* **2010**, *114*, 11725–11737.
- (36) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Theory Comput.* **2014**, *10*, 49–57.
- (37) Brauer, B.; Kesharwani, M. K.; Martin, J. M. L. *J. Chem. Theory Comput.* **2014**, *10*, 3791–3799.
- (38) Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Nielsen, I. M. B.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150–157.
- (39) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. *J. Phys. Chem. A* **2001**, *105*, 10525–10537.
- (40) Werner, H.-J.; Adler, T. B.; Knizia, G.; Manby, F. R. In *Recent Progress in Coupled Cluster Methods*; Čársky, P., Paldus, J., Pittner, J., Eds.; Springer: Dordrecht, Holland, 2010; pp 573–619.
- (41) Helgaker, T.; Ruden, T. A.; Jørgensen, P.; Olsen, J.; Klopper, W. *J. Phys. Org. Chem.* **2004**, *17*, 913–933.
- (42) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *J. Chem. Phys.* **2008**, *128*, 094313.
- (43) Babin, V.; Leforestier, C.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 5395–5403.
- (44) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. *J. Chem. Theory Comput.* **2012**, *8*, 469–478.
- (45) Tan, H.-J.; Bettens, R. P. A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7541–7547.
- (46) Collins, M. A.; Cvitkovic, M. W.; Bettens, R. P. A. *Acc. Chem. Res.* **2014**, *47*, 2776–2785.
- (47) Furtado, J. P.; Rahalkar, A. P.; Shanker, S.; Bandyopadhyay, P.; Gadre, S. R. *J. Phys. Chem. Lett.* **2012**, *3*, 2253–2258.
- (48) Sahu, N.; Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2013**, *138*, 104101.
- (49) Saha, A.; Raghavachari, K. *J. Chem. Theory Comput.* **2014**, *10*, 58–67.
- (50) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112.
- (51) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.