

Trouble with the Many-Body Expansion

John F. Ouyang, Milan W. Cvitkovic, and Ryan P. A. Bettens*

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

S Supporting Information

ABSTRACT: Longstanding conventional wisdom dictates that the widely used Many-Body Expansion (MBE) converges rapidly by the four-body term when applied to large chemical systems. We have found, however, that this is not true for calculations using many common, moderate-sized basis sets such as 6-311++G** and aug-cc-pVDZ. Energy calculations performed on water clusters using these basis sets showed a deceptively small error when the MBE was truncated at the three-body level, while inclusion of four- and five-body contributions drastically increased the error. Moreover, the error per monomer increases with system size, showing that the MBE is unsuitable to apply to large chemical systems when using these basis sets. Through a systematic study, we identified the cause of the poor MBE convergence to be a many-body basis set superposition effect exacerbated by diffuse functions. This was verified by analysis of MO coefficients and the behavior of the MBE with increasing monomer–monomer separation. We also found poor convergence of the MBE when applied to valence-bonded systems, which has implications for molecular fragmentation methods. The findings in this work suggest that calculations involving the MBE must be performed using the full-cluster basis set, using basis sets without diffuse functions, or using a basis set of at least aug-cc-pVTZ quality.



1. INTRODUCTION

The many-body expansion (MBE) is a useful and ubiquitous formalism in the theoretical study of large chemical systems.^{1–12} The MBE expresses the total energy, E_{tot} of an n -body system as the sum of one-body, two-body, etc., up to n -body energy contributions (see next section for a detailed description). Calculating E_{tot} directly for large systems is often computationally unaffordable. The benefit of the MBE is that for many systems E_{tot} can be well approximated by truncating the expansion to just the first few terms. Truncated MBEs have found especially widespread use in the study of water clusters, in which most intermolecular interactions are assumed to be pairwise additive (i.e., completely captured in an MBE truncated after the two-body term). The remaining (mostly inductive) interaction energy is accounted for by the rest of the terms in the MBE.

A longstanding and crucial question in modeling aqueous systems is how many terms of the MBE are necessary to adequately approximate the total energy. The earliest studies addressing this question were performed on water dimers, trimers, and tetramers. They found that three-body effects accounted for about 10% of the interaction energy, and that four- and higher-body effects were negligible.^{13,14} Subsequent work on slightly larger clusters agreed that four- and higher-body energy contributions were minute.^{15–20} The most thorough examination of many-body effects was performed on water hexamers by Xantheas in 1994, in which he found “the contribution from four-body and higher terms to be negligible for these systems.”²¹

The results from these studies eventually coalesced into an oft-cited piece of conventional wisdom: that the many-body expansion for water converges rapidly by the four-body term and in a well behaved manner.^{22–24} Indeed, most current ab-initio-based simulation models use MBEs truncated at three or occasionally four bodies.^{25–28}

Despite this, in the course of refining our group’s Combined Fragmentation Method (CFM)^{11,12} for use with noncovalent systems, we recently decided to verify the rapid convergence of the MBE for a few water clusters. We calculated all the terms in the MBEs of four (H₂O)₁₆ clusters with HF/6-31++G**, expecting, per conventional wisdom, to observe convergence to the true cluster energy by at most the five-body term (convergence herein defined as consistently having an error less than 1 m- E_h). Instead, not only did these MBEs not converge by anywhere near the five-body term, the convergence was notably erratic (Figure 1). Particularly worrying was that while truncating the many-body expansion at the four-body term led to a decent result (3.2–4.2 m- E_h error for 4444-a,c1b,cie), inclusion of the five-body term *increased* the error (4.6–4.7 m- E_h error for 4444-a,c1b,cie) rather than further converging the MBE toward the true cluster energy.

We are by no means the first to observe problems with the MBE.^{19,31–33} To our knowledge, however, there has been no thorough examination of under what circumstances the MBE fails to converge rapidly. In the following sections, we will demonstrate the wide extent of the MBE convergence problem

Received: May 7, 2014

Published: June 12, 2014

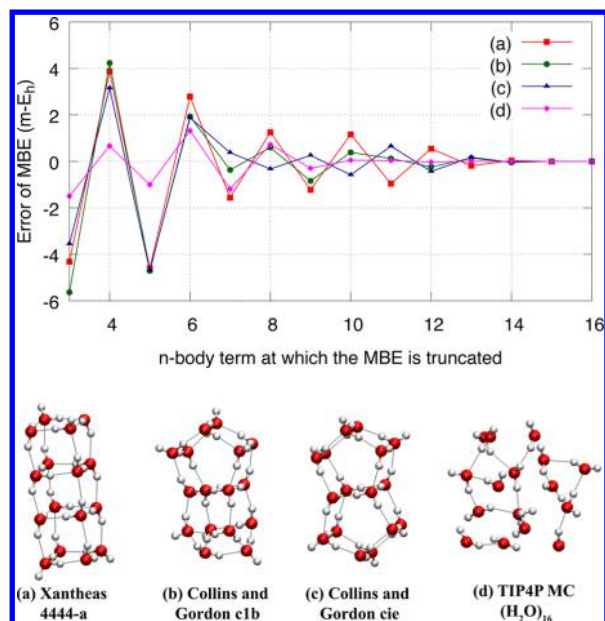


Figure 1. Slow, erratic convergence of the MBE toward the full-cluster energy for four $(\text{H}_2\text{O})_{16}$ clusters calculated at HF/6-31++G**. The linear fused cube 4444-a (a) was obtained from Yoo et al.,²⁹ while the fused pentameric structures c1b (b) and cie (c) were obtained from Pruitt et al.³⁰ The TIP4P MC structure (d) was obtained by taking a random fragment of 16 water molecules from a TIP4P Monte Carlo simulation of 400 water molecules.

and show that a many-body Basis Set Superposition Effect (BSSE) is its cause. We stress that we do not question the accuracy or validity of the many methods that use many-body expansions, or question previous studies of many-body effects in water. Our purpose in this work is simply to refine the conventional wisdom about the MBE.

2. COMPUTATIONAL DETAILS

For a chemical system comprising n monomers, the MBE of the total energy of the system, E_{tot} is the finite sum

$$E_{\text{tot}} = \sum_{k=1}^n \epsilon^{(n,k)} \quad (1)$$

where $\epsilon^{(n,k)}$ is the total k -body energy of the system. The total k -body energy of the system is the component of the total energy due to all k -body effects, so $\epsilon^{(n,1)}$ is the sum of the energies of all isolated monomers; $\epsilon^{(n,2)}$ is the sum of the energies of all dimers minus the energies of the monomers they comprise, i.e., it is the sum of all the pairwise interaction energies. Thus, the total interaction energy of the system is given by

$$\epsilon_{\text{tot}\setminus 1} = E_{\text{tot}} - \epsilon^{(n,1)} \quad (2)$$

For example, consider a system of three molecules, A, B, and C. The one-body energy is

$$\begin{aligned} \epsilon^{(3,1)} &= \epsilon_A^{(3,1)} + \epsilon_B^{(3,1)} + \epsilon_C^{(3,1)} \\ &= E_A + E_B + E_C \end{aligned} \quad (3)$$

the two-body energy is

$$\begin{aligned} \epsilon^{(3,2)} &= \epsilon_{AB}^{(3,2)} + \epsilon_{AC}^{(3,2)} + \epsilon_{BC}^{(3,2)} \\ &= E_{AB} - (E_A + E_B) + E_{AC} - (E_A + E_C) + E_{BC} \\ &\quad - (E_B + E_C) \end{aligned} \quad (4)$$

and the three-body energy is

$$\begin{aligned} \epsilon^{(3,3)} &= \epsilon_{ABC}^{(3,3)} \\ &= E_{\text{tot}} - (E_{AB} - (E_A + E_B) + E_{AC} - (E_A + E_C) \\ &\quad + E_{BC} - (E_B + E_C)) - (E_A + E_B + E_C) \end{aligned} \quad (5)$$

where in all the above, $\epsilon_a^{(3,k)}$ and E_a are the k -body energy and the total energy respectively of the subsystem a of the trimer. For a more detailed explanation of the MBE, see Xantheas.²¹

$\epsilon^{(n,k)}$ can also be expressed recursively in terms of the total energy and lower-body energies³⁴

$$\epsilon^{(n,k)} = \sum_{\alpha} E_{\alpha} - \sum_{i=1}^{k-1} \left[\frac{(n-i)!}{(n-k)!(k-i)!} \right] \epsilon^{(n,i)} \quad (6)$$

where E_{α} is the total energy of the k -mer subsystem α of which there are $\binom{n}{k}$.

We also wish to clarify our use of the term basis set superposition effect. When small basis sets are used in ab initio calculations of molecular clusters, basis functions from one molecule can be utilized by other molecules to compensate for the incompleteness of their basis set. This results in an improved description of the wave function of all molecules in the cluster, which leads to a lowering of the total energy known as the BSSE. One way of quantifying the BSSE is the counterpoise (CP) method.³⁵ In the CP method, the familiar expression for the BSSE in the interaction energy of two molecules, A and B, is given as

$$E_{\text{BSSE}} = (E_A(\mathbf{ab}) - E_A(\mathbf{a})) + (E_B(\mathbf{ab}) - E_B(\mathbf{b})) \quad (7)$$

where \mathbf{a} , \mathbf{b} , and \mathbf{ab} are the basis sets of molecule A, molecule B, and the cluster AB, respectively. Applying the CP method to the interaction energy of A and B, the BSSE-free interaction energy is

$$\begin{aligned} \epsilon_{\text{tot}\setminus 1}^{\text{CP}} &= \epsilon_{\text{tot}\setminus 1} - E_{\text{BSSE}} \\ &= E_{AB}(\mathbf{ab}) - (E_A(\mathbf{a}) + E_B(\mathbf{b})) - E_{\text{BSSE}} \\ &= E_{AB}(\mathbf{ab}) - (E_A(\mathbf{ab}) + E_B(\mathbf{ab})) \end{aligned} \quad (8)$$

Now, all the quantities are calculated consistently in the same basis, namely the basis set of the cluster AB. The brilliance of the CP method lies in that it does not try to remove the lowering of energy in the total energy of the cluster AB due to sharing of basis functions, which is a natural consequence of the variational principle. Instead, it does the opposite where the constituents A and B are calculated in the basis set of the cluster AB so as to achieve a similar lowering of energy.

In the spirit of the CP method, we define the BSSE for the sum of the total energies of all $\binom{n}{k}$ k -mers in the cluster containing n monomers as

$$E_{\text{BSSE}}^{(n,k)} = E^{(n,k)}(\mathbf{n\text{-mer}}) - E^{(n,k)}(\mathbf{k\text{-mer}}) \quad (9)$$

where $\mathbf{n\text{-mer}}$ and $\mathbf{k\text{-mer}}$ are the basis set of the full molecular cluster (cluster basis) and the basis set of the k monomers considered (k -mer basis), respectively. Using our definition, the

removal of the BSSE in the total energies when performing a many-body energy decomposition (such as eqs 4 and 5) will result in all the quantities being calculated consistently in the cluster basis, ensuring that the MBE remains formally exact. Indeed, the use of a consistent cluster basis has been employed previously^{21,36,37} to obtain BSSE-free many-body energies. From eq 9, BSSE can be seen as a lowering of the total energy of the k -mer in the cluster due to the sharing of basis functions from the remaining $n - k$ monomers. Thus, when $k = n$, there is no BSSE, i.e., $E_{\text{BSSE}}^{(n,n)} = 0$. Notably, this definition of BSSE reduces to the familiar expression in eq 7 in the context of the interaction energy of a cluster where $E_{\text{BSSE}} = E_{\text{BSSE}}^{(n,1)}$.

All quantum chemical calculations were performed using the Gaussian 09 package³⁸ or the MOLPRO suite of programs³⁹ at the Hartree–Fock (HF) or second-order Møller–Plesset perturbation (MP2) level of theory. A variety of Pople split-valence basis sets were used, along with the series of Dunning correlation-consistent cc-pVXZ basis sets, $X = 2-4$, labeled VDZ, VTZ, and VQZ. An “A” or “dA” prepended to these basis sets indicate they are augmented or doubly augmented, respectively, with diffuse functions.

3. RESULTS AND DISCUSSION

3.1. Extent of the Poor Convergence of the MBE.

Prompted by our initial results (Figure 1), we attempted to ascertain the extent of the MBE convergence problem. We calculated MBEs up to the five-body term for a variety of $(\text{H}_2\text{O})_n$, n between 6 and 57, geometries (Table 1). Some geometries are optimized structures from the literature; others

Table 1. Error ($m-E_h$) in the Total Energy of Water Clusters $(\text{H}_2\text{O})_n$, $n = 6-57$, As Approximated by an MBE Truncated after the Two- through Five-Body Term^a

$(\text{H}_2\text{O})_n$	two-body	three-body	four-body	five-body
6 ^b	−9.769	−0.488	−0.006	0.017
8 ^b	−16.905	−0.426	0.491	−0.518
10 ^b	−23.227	−1.444	1.092	−1.099
12 ^b	−26.745	−0.399	1.519	−2.541
14 ^b	−27.613	−1.302	2.478	−3.563
16 ^b	−37.045	−0.631	2.940	−5.200
18 ^b	−37.060	−0.563	3.227	−5.848
20 ^b	−46.844	−0.741	4.219	−8.101
24 ^c	−106.289	−10.762	3.252	
32 ^d	−130.794	−6.612	16.895	
57 ^e	−97.126	−5.401	27.334	
6 ^f	−1.674	−0.024	0.022	−0.009
8 ^f	−2.591	−0.072	0.072	−0.023
10 ^f	−1.704	0.041	0.157	−0.098
12 ^f	0.189	0.266	0.274	−0.344
14 ^f	−6.870	−0.003	0.391	−0.495
16 ^f	−11.364	−0.378	1.153	−1.555
18 ^f	−6.524	−0.191	1.382	−2.009
20 ^f	−14.251	−0.721	1.692	−2.910
45 ^f	−36.931	−2.117	8.373	

^aAll calculations done at HF/AVDZ level of theory. ^bOptimized water clusters obtained from Maheshwary et al.⁴⁰ ^cOptimized water clusters obtained from Gora et al.²⁴ ^dOptimized water clusters obtained from Pruitt et al.³⁰ ^eOptimized water clusters obtained from Richard and Herbert.⁴¹ ^fDisordered random fragments of $(\text{H}_2\text{O})_n$ obtained from a TIP4P Monte Carlo simulations. For these disordered fragments, the MBE truncation errors were averaged over four different random fragments for $n = 6 - 20$.

were taken from TIP4P Monte Carlo simulations. These latter structures were included as they are representative of geometries encountered in simulations using truncated-MBE-based water models. All calculations were performed using the AVDZ basis set, which is a better yet still computationally manageable basis set compared to the 6-31++G** basis used in Figure 1. The MBEs in Table 1 were only calculated to at most the five-body term due to the steep computational cost of calculating high-body terms for large clusters: the number of additional calculations required to obtain the k -body energy of an n -body system is $(n!)/(k!(n - k)!)$.

Table 1 shows that the MBEs of very small clusters ($n = 6-8$) indeed converge by the three-body term, as shown in previous studies.²¹ But for larger clusters, while MBEs truncated at the three-body term appear converged, inclusion of four- and five-body energies *increases* the error in the MBE. A notable oscillatory behavior also occurs wherein the error changes sign when three- and higher-body energies are included.

More alarmingly, Figure 2 shows that for the clusters studied in Table 1, the four- and five-body MBE truncation errors per

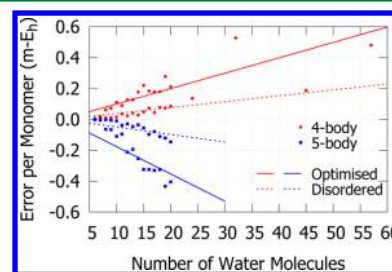


Figure 2. Error-per-monomer of the MBE truncated at the four-body level (red solid circles) and at the five-body level (blue solid squares) for optimized water clusters (solid lines) and disordered water clusters (dashed lines) from Table 1.

monomer increase with system size. This was also noticed previously in a smaller sample of water clusters by Kulkarni et al., who called for further examination.¹⁹ This is concerning as the error-per-monomer should be an intensive, not extensive, property. Otherwise, the scalability of truncated-MBE-based computational methods, such as fragmentation methods and bulk material simulations, becomes questionable.

3.2. Cause of the Poor Convergence of the MBE.

Having found the MBE convergence problem to be widespread, we sought to determine its cause. Initially, we thought the poor convergence was due to the MBE not properly capturing the many-body induction energy. Since induction energy is closely related to the polarizability of the molecules in a system, we performed MBE calculations for the optimized clusters studied in Table 1 using a series of basis sets of increasing polarizability (Table 2). If the poor MBE convergence were due to induction, the error for truncating the MBE should worsen with increasing polarizability. It is clear from Table 2 that this is not the case.

Instead, poor MBE convergence only occurred when using small, incomplete basis sets augmented with diffuse functions, namely 6-311++G** (P3) and AVDZ (D). This led us to suspect that the convergence problem was due to BSSE. As a preliminary test of this, the MBEs in Table 2 were recalculated using the cluster basis, as opposed to the usual k -mer basis, in all k -mer calculations. This eliminated the BSSE in the MBE calculations as explained in section 2. As shown in Table 3, the poor MBE convergence observed in Table 2 disappeared when

Table 2. Error in Approximating the Total Energy of Optimized Water Clusters $(\text{H}_2\text{O})_n$, $n = 8\text{--}20$, with an MBE Truncated at the Four-Body (4B) and Five-Body (5B) Term Using Basis Sets of Increasing Isotropic Dipole–Dipole Polarizability $\bar{\alpha}$

		basis set					
		P1 ^a	P2 ^a	P3 ^a	D ^b	T ^b	Q ^b
$\bar{\alpha}$ (a.u.)		3.74	4.87	6.50	7.97	8.23	8.30
$(\text{H}_2\text{O})_n$		error ($m\text{-}E_h$)					
8	4B	0.01	0.07	1.24	0.49	0.05	0.01
	5B	0.02	0.01	−0.67	−0.52	−0.04	0.00
10	4B	−0.06	−0.01	0.91	1.09	0.04	−0.05
	5B	0.04	0.01	−0.65	−1.10	−0.09	−0.02
12	4B	0.42	0.21	1.13	1.52	0.01	
	5B	0.01	−0.02	−0.55	−2.54	−0.36	
14	4B	0.30	0.25	1.57	2.48	0.19	
	5B	0.06	0.02	−1.12	−3.56	−0.47	
16	4B	0.47	0.32	2.82	2.94	0.00	
	5B	0.11	0.02	−1.75	−5.20	−0.66	
18	4B	0.82	0.38	1.32	3.23		
	5B	0.03	−0.02	−0.66	−5.85		
20	4B	0.98	0.48	2.00	4.22		
	5B	0.03	−0.04	−1.15	−8.10		

^aAll calculations performed at the HF level. Water geometries are from Maheshwary et al.⁴⁰ (the same geometries as used in Table 1). Pople basis set P1: 3-21G, P2: 6-31G**, and P3: 6-311+G(2d,p). ^bDunning basis set AVXZ where X = D or T or Q.

Table 3. Error for Truncating the MBE at the Four-Body (4B) and Five-Body (5B) Term Using the Cluster Basis, As Opposed to the k -mer Basis Used in Table 2, in All k -Body Calculations^a

		basis set				
		P1 ^b	P2 ^b	P3 ^b	D ^c	T ^c
$\bar{\alpha}$ (a.u.)		3.74	4.87	6.50	7.97	8.23
$(\text{H}_2\text{O})_n$		error ($m\text{-}E_h$)				
8	4B	−0.07	−0.02	0.01	0.04	0.04
	5B	0.01	0.01	0.00	0.00	0.00
10	4B	−0.05	−0.03	−0.02	−0.02	−0.01
	5B	0.00	0.00	−0.01	−0.01	−0.01
12	4B	−0.01	0.00	−0.01	0.00	
	5B	0.00	−0.01	−0.02	−0.03	
14	4B	0.01	0.02	0.03	0.05	
	5B	0.00	0.00	−0.01	−0.01	
16	4B	−0.10	−0.05	−0.02	0.02	
	5B	0.03	0.02	0.00	−0.01	
18	4B	−0.03				
	5B	0.00				
20	4B	−0.03				
	5B	0.00				

^aResults shown for a series of optimized water clusters $(\text{H}_2\text{O})_n$, $n = 8\text{--}20$, at the HF level using various basis set of increasing isotropic dipole–dipole polarizability $\bar{\alpha}$. ^bPople basis set P1: 3-21G, P2: 6-31G**, and P3: 6-311+G(2d,p). ^cDunning basis set AVXZ where X = D or T.

BSSE was removed. (Note that not all terms were recalculated due to the computational cost of using the full-cluster basis.) In fact, the BSSE present in the many-body energies can be easily computed as the difference between the errors in both tables.

To verify that BSSE was the cause of the poor MBE convergence, we calculated the full MBEs for two $(\text{H}_2\text{O})_{10}$ clusters, 10PP and 10OB (Figure 3), using basis sets of

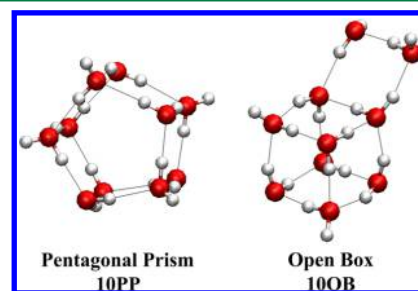


Figure 3. $(\text{H}_2\text{O})_{10}$ clusters chosen for a more detailed study on the cause of poor MBE convergence. Both the pentagonal prism (10PP) and open box (10OB) were obtained from Maheshwary et al.⁴⁰

increasing quality and diffusiveness and using both k -mer bases and cluster bases. MP2 calculations were also performed for the 6-31G** and the 6-311G** series to investigate the effects of electron correlation on the MBE convergence. As the results were similar for both 10PP and 10OB clusters, only the errors of the MBE for 10PP are presented in Figure 4. Results for 10OB are in the Supporting Information.

It is clear from Figure 4 that when the k -mer basis is used (solid lines), the more diffuse functions that are present, the worse the MBE convergence (red and orange solid lines), with the exception of AVTZ and AVQZ. This is all precisely what one would expect if BSSE was the cause of the poor MBE convergence: more diffuse functions lead to more overlap of basis functions between water molecules, increasing BSSE, except for basis sets like AVTZ and AVQZ which are so complete that water monomers need not rely on diffuse functions from their neighbors to describe their wave functions. Indeed, Truhlar and co-workers have made a similar observation by examining the effects of increasing augmentation in the Dunning basis sets.^{42–44} It should be noted that there are still tiny oscillations in the MBE truncation errors for AVTZ and AVQZ in the range of 10–50 $\mu\text{-}E_h$, which are hard to see in the figure. Moreover, when the cluster basis is used (dashed lines) and BSSE is eliminated, the MBE converges by the four-body term regardless of the presence of diffuse functions. Figure 4c–f further show that when electron correlation is included, the MBE errors are amplified. This can be attributed to additional BSSE associated with electron correlation—it is known that correlation energy converges more slowly toward the complete basis set limit than the SCF energy.⁴⁵

We also examined the MO coefficients in these calculations to specifically check whether the BSSE originated from the diffuse functions. HF calculations were performed using VDZ, AVDZ, and AVTZ basis sets on an arbitrarily chosen monomer from 10PP with the ghost basis functions of all other waters in the cluster present. (The choice of monomer does not significantly affect the results due to the symmetry of the cluster.) The distribution of MO coefficients for the occupied MOs is shown in Figure 5. By performing calculations on a single monomer in the cluster basis, all observations are solely due to BSSE and not physical interaction between molecules. If diffuse functions were causing the BSSE—that is, if water molecules were using diffuse basis functions centered on other molecules to improve the description of their own wave

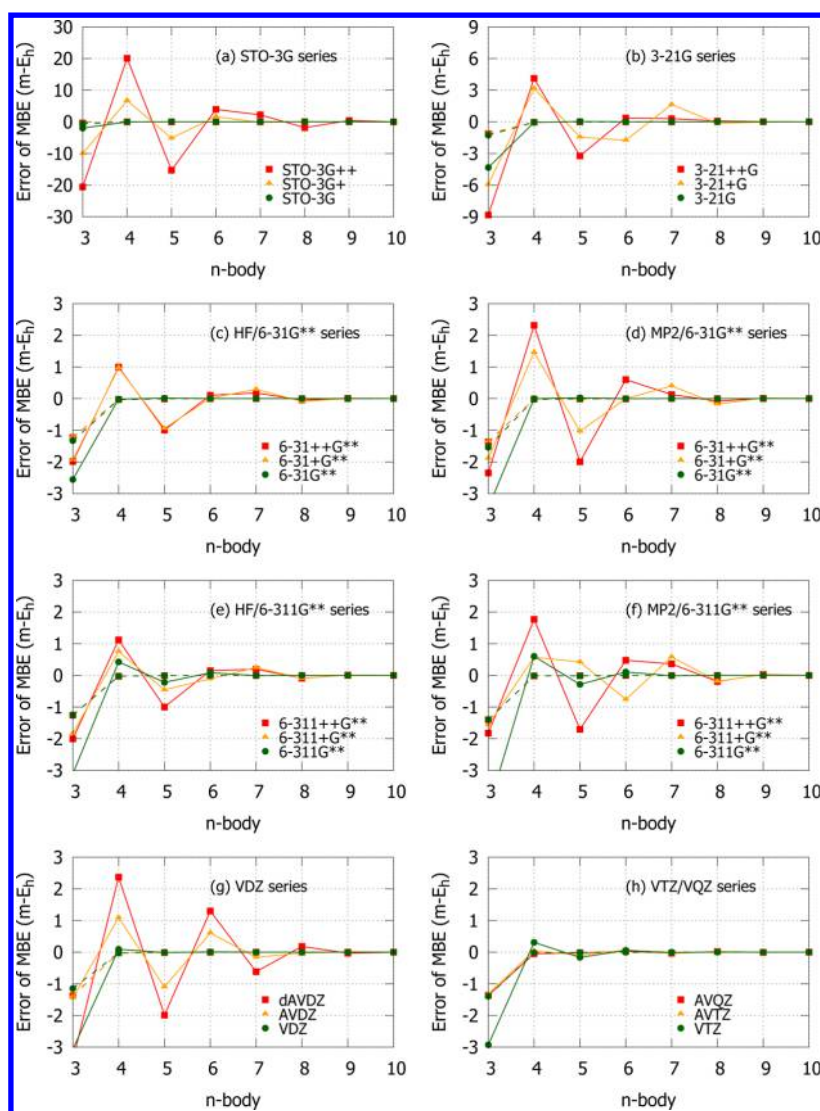


Figure 4. Convergence of the MBE for 10PP using various basis sets: (a) STO-3G series, (b) 3-21G series, (c) HF/6-31G** series, (d) MP2/6-31G** series, (e) HF/6-311G** series, (f) MP2/6-311G** series, (g) VDZ series, and (h) VTZ/VQZ series. Solid lines represent MBE calculated using the k -mer basis, while dashed lines represent MBEs calculated using the cluster basis. It should be noted that the diffuse functions of the 6-31++G** basis set were used as the diffuse functions for the STO-3G and 3-21G basis sets, as these basis sets have no defined diffuse functions.

function—then there should be significant MO coefficients for basis functions centered on the ghost molecules. Similarly, a many-body BSSE effect can be inferred if there are significant nonzero MO coefficients arising from *many* of these ghost molecules simultaneously.

For VDZ (Figure 5a), significant nonzero MO coefficients, represented by red or blue colored regions, are only found for a few water molecules' basis functions. In contrast, the AVDZ basis set (Figure 5b) has significant nonzero MO coefficients on all the water molecules' basis functions. As the ghost water molecules in Figure 5 are ordered by their proximity to the monomer under study, the colored regions become fainter across the horizontal axis due to decreasing overlap of the basis functions from more distant ghosts. Nonetheless, the nonzero coefficients imply that the BSSE is many-body in nature, with contributions from all monomers in the system. The contributions come primarily from the diffuse functions (denoted D in the figure) of both oxygen and hydrogen, again implicating diffuse functions in causing the BSSE. The

MO coefficient distribution for AVTZ (Figure 5c) also shows contributions from diffuse functions, but less so than those for AVDZ. Again, this is due to AVTZ being a more complete basis set: the wave function of the monomer can be described using its own core, valence, and diffuse basis functions without the need for the basis functions of its neighbors. In fact, AVTZ's BSSE contribution to the total energy of the monomer is low at 2.7 ppm of the total energy of the monomer (-0.203 m-E_h) in contrast to the higher contribution from both VDZ (68 ppm, -5.197 m-E_h) and AVDZ (10 ppm, -0.768 m-E_h).

As a final test of our hypothesis, we investigated how MBE convergence is affected by the average nearest-neighbor distances of water molecules in a cluster. HF/AVDZ MBE calculations using the k -mer basis were performed on a series of progressively expanded structures derived from 10PP (Figure 6). The expanded structures were constructed by scaling the distance between the center-of-mass of each water and the center-of-mass of the entire 10PP cluster. This ensures that the nearest-neighbor distances of all water molecules are increased

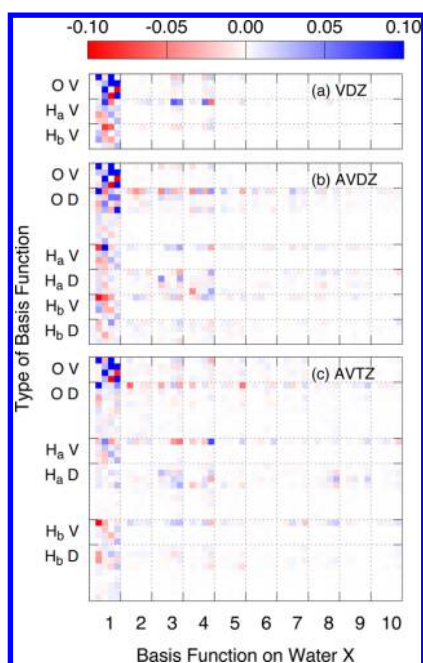


Figure 5. Distribution of MO coefficients of an arbitrarily chosen monomer of **10PP** calculated with the cluster basis using (a) VDZ, (b) AVDZ, and (c) AVTZ basis sets. The vertical axis shows the basis functions arranged according to the nuclei (O or H). Only valence functions (denoted V), functions with the smallest exponent, and diffuse functions (denoted D) are shown. The horizontal axis shows the basis molecule the basis functions are centered on: Water 1 is the monomer under study, and the rest are ghost molecules.

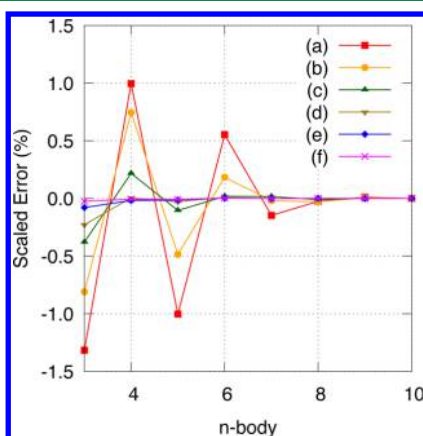


Figure 6. Convergence of the MBE for expanded structures derived from **10PP** with the following mean nearest-neighbor distance: (a) $5.25 a_0$, (b) $5.67 a_0$, (c) $6.61 a_0$, (d) $7.56 a_0$, (e) $8.50 a_0$, and (f) $9.45 a_0$. The MBE truncation error has been normalized to the one-body error (i.e., the interaction energy of the cluster).

by the same factor. As the mean nearest-neighbor distance increases, the oscillations in the MBE error gradually disappear. This is because when the waters are farther apart, the overlap between diffuse functions on different waters decreases exponentially and so does the BSSE. This can be seen explicitly in Figure 7, where an exponential fit captures the decay of BSSE with increasing interwater-molecule distance.

The curious reader may wonder why the error oscillates from positive to negative in nearly all the poorly convergent MBEs we have shown. Our best explanation is that this behavior is related to the inclusion/exclusion principle inherent in the

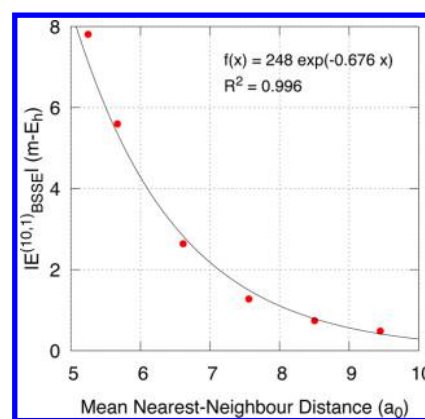


Figure 7. Magnitude of the BSSE in the interaction energy $|E_{\text{BSSE}}^{(10,1)}|$ of the expanded structures derived from **10PP** shown in Figure 6 shows an exponential decay with increasing mean nearest-neighbor distance.

MBE. To obtain a system's k -body energy, $\epsilon^{(k)}$, the total energy of each k -mer in the system has the total energy of all its constituent $(k-1)$ -mers subtracted from it. But this results in oversubtraction of $(k-2)$ -body energies, so the $(k-2)$ -mer total energies have to be added back, and so on (see again eq 4). When subsequently obtaining the $(k+1)$ -body energy, the signs of the terms in the expression switch: k -body energies are subtracted where they were previously added etc. (in addition to there being many more terms in the calculation). So if a particular k -body energy is underestimated—perhaps due to an inadequate basis set—it will contribute to the overestimation of the $(k+1)$ -body energy and then the underestimation of the $(k+2)$ -body energy and so on, leading to oscillation in the error. Of course, the MBE by definition converges by the final term. And each subsequent term in the MBE contains less BSSE-derived error, $E_{\text{BSSE}}^{(n,k+1)} < E_{\text{BSSE}}^{(n,k)}$, since the more water monomers that are in a calculation, the closer the basis set is to the correct, full-cluster basis. But in the early MBE terms, there is much utilization of neighboring waters' diffuse basis functions, and thus more error in each calculated energy, and thus oscillations which diminish as more terms are added.

3.3. Methods to Improve MBE Convergence. We have shown that rapid convergence of the MBE can be guaranteed by performing calculations in the cluster basis or with a high-quality basis set. These are, of course, fairly dispiriting solutions as they greatly increase computational cost, so we have examined several alternatives.

Since basis function overlap is distance-dependent, we tested a distance-cutoff basis (d-c basis) which includes ghost functions only from waters within a specified cutoff distance of the water molecules in a calculation. Testing this d-c basis on the $(\text{H}_2\text{O})_{16}$ 4444-a cluster using various cutoff distances gave poor results (Table 4), however. This is likely due to the small MBE truncation errors involved in the $m-E_h$ range. With slight changes in cutoff distances, one more or one fewer water's ghost functions might be included in a calculation, which could lead to significant changes in calculated energies, drastically affecting the MBE truncation errors. That said, we think the distance-cutoff basis might work with a larger cutoff distance, but in those cases it would be more economical to use a high-quality basis set instead.

Another workaround that has been proposed, albeit for a different problem, is the k -mer-centered basis set (k CBS) approach of Gora et al.²⁴ The k CBS approach attempts to remove BSSE in an MBE calculation by calculating each k -mer

Table 4. Error for Truncating the MBE up to the Five-Body Term, Performed Using Various Methods to Improve MBE Convergence^a

method	error (m-E _h)			
	two-body	three-body	four-body	five-body
k-mer basis	−42.418	−4.763	4.588	−5.563
cluster basis	−56.170	−1.217	−0.043	−0.056
d-c basis, 3 Å	−35.325	−28.202	10.527	9.549
d-c basis, 4 Å	−41.468	−46.804	85.769	−108.259
d-c basis, 5 Å	−48.275	−34.555	87.929	−161.926
kCBS	−50.464	4.779	6.268	6.279
charge field	−12.466	−0.449	2.073	−2.054

^aCalculations were performed at HF/6-31++G** on the (H₂O)₁₆ 4444-a cluster shown in Figure 1. k-mer basis and cluster basis are shown for reference.

and all its subcalculations using the k-mer's basis set. That is, a dimer's two-body contribution would be computed as total energy of the dimer minus the total energy of its constituent monomers, all calculated with the dimer basis set. This results in substantially more calculations to compute the MBE since calculations from previous terms cannot be reused, but it does mean each calculation has no BSSE. We applied the kCBS approach to the (H₂O)₁₆ 4444-a cluster. From Table 4, we see that the MBE does converge rapidly, but to an incorrect value. This is likely because the kCBS approximation is not formally exact: the terms in the MBE do not cancel due to the different numbers of basis functions used in each term's calculations. The kCBS approach certainly does converge correctly when a high-quality basis set is used, as has been demonstrated in the literature, but this seems to be due to the high quality of the basis set, not the kCBS method.

Strategies unrelated to BSSE for improving MBE convergence are widely used. Many truncated-MBE-based computational methods incorporate a charge field to approximate higher-order many-body effects, typically by interacting the one- or two-body fragments with a charge field representing the rest of the system.^{2–4,6,10,46,47} While we have not done a thorough analysis, preliminary results using embedded charges from Stone's distributed multipole analysis⁴⁸ indicate that embedded charges dampen, but do not remove, the oscillatory MBE behavior (Table 4). This is not surprising as the embedded charges only serve to approximate the physical higher-order many-body effects arising from induction and thus do not remove the many-body BSSE.

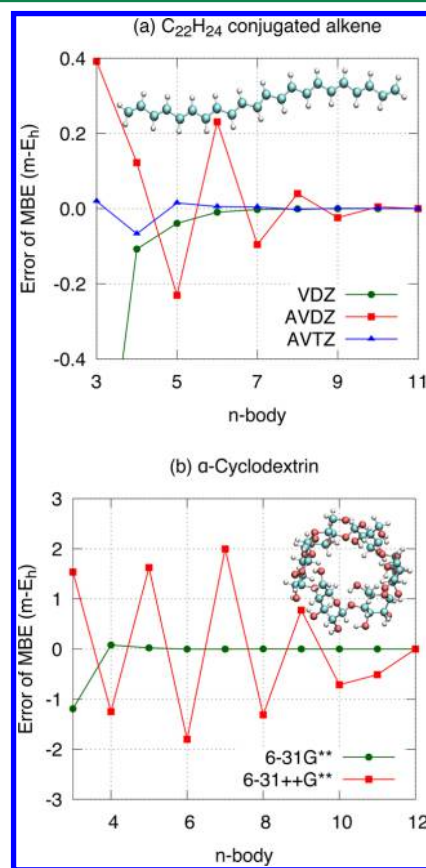
Other methods include high-order many-body effects by performing a low-level ab initio calculation on the full system.^{3,4,49,50} Such methods capture many-body effects far better than methods using only a truncated MBE.⁵¹ The full-system calculations in these methods are not susceptible to BSSE-based MBE convergence issues since they use full-system basis, but lower-body calculations performed using only the k-mer basis are still susceptible.

Thus, unhappily, we have found no alternative for avoiding poor MBE convergence that is more efficient than using the full-cluster basis or a high-quality basis set. As the use of the cluster basis is computationally prohibitive, our recommendation is to use a high-quality basis set for MBE calculations; our results indicate that at least AVTZ-quality is prudent.

3.4. Extension to Valence-Bonded Systems. So far we have only presented data for noncovalent water clusters, but MBE convergence problems also arise in valence-bonded

systems. This has great implications for fragmentation methods, which in most cases use a truncated MBE, or something analogous to it, to approximate the total energies of large chemical systems.^{41,52–54}

In fragmentation methods, small groups of adjacent atoms are treated as bodies. Using our CFM algorithm¹¹ to define groups/bodies, we calculated the MBEs for a 22-carbon C₂₂H₂₄ conjugated alkene and α-cyclodextrin (Figure 8). Slow MBE

**Figure 8.** Convergence of the MBE for the total energy of (a) C₂₂H₂₄ conjugated alkene and (b) α-cyclodextrin at HF level of theory for various basis sets. Here, the CFM algorithm was used to define the bodies in the MBE. The inset shows the structures of the molecules.

convergence is observed in both systems when incomplete basis sets with diffuse functions are used, as seen in the case of AVDZ for C₂₂H₂₄ and 6-31++G** for α-cyclodextrin. This is no surprise as the same borrowing of basis functions from adjacent groups that causes poor convergence in water clusters occurs in these valence-bonded systems. The errors in the C₂₂H₂₄ MBE are small (even negligible) because the molecule's linear shape minimizes basis function overlap. Compare this to the MBE of the more compact α-cyclodextrin, where the errors are beyond chemical accuracy until the inclusion of the nine-body term. Since fragmentation methods rarely include five-or-higher-body effects, it seems likely that fragmentation calculations using BSSE-prone basis sets are liable to, and have in the past been afflicted by, preventable, BSSE-based errors. On an interesting related point, due to the above-mentioned affects, we expect that any calculation that is performed in order to predict bond-breaking energies would be overestimated.

It should be noted, though, that poor MBE convergence in a valence-bonded system depends on the definition of “body”. Another type of MBE that our group has examined is to treat distortions in the internal degrees-of-freedom of a molecule as bodies. Using the equilibrium geometry as a reference, an MBE can be used to calculate the distortion energy of a molecule. We demonstrate a proof-of-concept using the methanol molecule (Figure 9). We distorted the molecule randomly in all 12

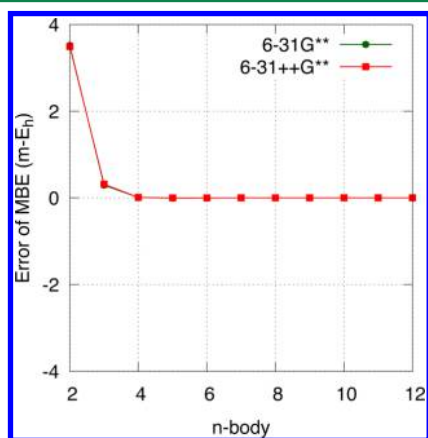


Figure 9. Convergence of the MBE for the distortion energy of a methanol molecule at the HF level of theory for various basis sets. Here, distortions in the 12 internal degrees of freedom are treated as bodies in the MBE. Convergence is clearly independent of the choice of basis set—the results from both basis sets overlap almost perfectly.

degrees of freedom, yielding a total distortion energy of about 140 m- E_h . The degree-of-freedom MBE converges by the four-body term, even when a BSSE-prone basis set is used. This is expected as a consistent basis set is used in all calculations, essentially equivalent to the use of a full-cluster basis. Apart from intramolecular degrees-of-freedom, intermolecular degrees-of-freedom or a combination of both could be treated in the same manner. The utility of such an approach is obvious. A high-dimensional system is broken down into to numerous, completely independent (and thus highly parallelizable) much lower-dimensional function evaluations. Future work will explore how degree-of-freedom MBEs can be used to construct accurate, high-dimensional potential energy surfaces from many lower-dimensional surfaces.

4. CONCLUSIONS

There is no question that the many-body expansion is a theoretically sound and extremely useful formalism in the study of large molecular systems. But it is likewise clear from our observations that care must be taken in its implementation. Rapid convergence at the four-body term of the MBE cannot be assumed, even when convergence appears to have occurred. Incautious use of MBEs with systems and levels of theory susceptible to BSSE is liable to yield errors well beyond chemical accuracy. Moreover, the error per monomer worsens extensively with system size. Such concerns are relevant in valence-bonded and noncovalent systems alike. We conclude that the use of a consistent basis set, either in the form of the full-cluster basis or a high-quality basis set (at least AVTZ quality), is necessary to avoid poor MBE convergence due to BSSE.

■ ASSOCIATED CONTENT

Supporting Information

The graphs for the convergence of the MBE for **10OB** using the basis set in Figure 4 are given in section S1. The Cartesian coordinates for the water clusters studied in this work are given in section S2. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

■ 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, and the Fulbright U.S. Student Program. The authors also thank the Centre for Computational Science and Engineering for the use of their computers.

■ REFERENCES

- (1) Dahlke, E. E.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 10595–10601.
- (2) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.
- (3) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.
- (4) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33–41.
- (5) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (6) Addicoat, M. A.; Collins, M. A. *J. Chem. Phys.* **2009**, *131*, 104103.
- (7) Řezáč, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2009**, *6*, 91–99.
- (8) Weiss, S. N.; Huang, L.; Massa, L. *J. Comput. Chem.* **2010**, *31*, 2889.
- (9) Beran, G. J. O.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480–3487.
- (10) Wang, X.; Liu, J.; Zhang, J. Z. H.; He, X. *J. Phys. Chem. A* **2013**, *117*, 7149–7161.
- (11) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. *J. Chem. Theory Comput.* **2011**, *8*, 469–478.
- (12) Tan, H.-J.; Bettens, R. P. A. *J. Phys. Chem. Chem. Phys.* **2013**, *15*, 7541–7547.
- (13) Elrod, M. J.; Saykally, R. J. *Chem. Rev.* **1994**, *94*, 1975–1997.
- (14) Hermansson, K. *J. Chem. Phys.* **1988**, *89*, 2149.
- (15) Xantheas, S. S. *Philos. Mag. B* **1996**, *73*, 107–115.
- (16) Chen, W.; Gordon, M. S. *J. Phys. Chem.* **1996**, *100*, 14316–14328.
- (17) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163–9168.
- (18) Xantheas, S. S. *Chem. Phys.* **2000**, *258*, 225–231.
- (19) Kulkarni, A. D.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2004**, *121*, 5043–5050.
- (20) Xantheas, S. S. *Struct. Bonding (Berlin, Ger.)* **2005**, *116*, 119–148.
- (21) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523–7534.
- (22) Christie, R. A.; Jordan, K. D. *Struct. Bonding (Berlin, Ger.)* **2005**, *116*, 27–41.
- (23) Szalewicz, K.; Leforestier, C.; van der Avoird, A. *Chem. Phys. Lett.* **2009**, *482*, 1–14.
- (24) Gora, U.; Podeszwa, F.; Cencek, W.; Szalewicz, K. *J. Chem. Phys.* **2011**, *135*, 224102.
- (25) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2011**, *134*, 094509.

- (26) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 014309.
- (27) Medders, G. R.; Babin, V.; Paesani, F. *J. Chem. Theory. Comput.* **2013**, *9*, 1103–1114.
- (28) Kumar, R.; Keyes, T. *Theor. Chem. Acc.* **2012**, *131*, 1197.
- (29) Yoo, S.; Aprá, E.; Zeng, X. C.; Xantheas, S. S. *J. Phys. Chem. Lett.* **2010**, *1*, 3122–3127.
- (30) Pruitt, S. R.; Addicoat, M. A.; Collins, M. A.; Gordon, M. S. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7752–7764.
- (31) Hermann, A.; Krawczyk, R. P.; Lein, M.; Schwerdtfeger, P. *Phys. Rev. A* **2007**, *76*, 013202.
- (32) Riega, H.; Almeida, R.; Rincon, L. *Rev. Mex. Fis. S* **2006**, *52*, 204–207.
- (33) Cui, J.; Liu, H.; Jordan, K. D. *J. Phys. Chem. B* **2006**, *110*, 18872–18878.
- (34) Kaplan, I. G.; Santamaria, R.; Novaro, O. *Mol. Phys.* **1995**, *84*, 105–114.
- (35) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (36) Valiron, P.; Mayer, I. *Chem. Phys. Lett.* **1997**, *275*, 46–55.
- (37) Milet, A.; Moszynski, R.; Wormer, P. E. S.; van der Avoird, A. *J. Phys. Chem. A* **1999**, *103*, 6811–6819.
- (38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision C.01; Gaussian Inc.: Wallingford, CT, 2009.
- (39) 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>.
- (40) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. *J. Phys. Chem. A* **2001**, *105*, 10525–10537.
- (41) Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.
- (42) Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1197–1202.
- (43) Papajak, E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 10–18.
- (44) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.
- (45) Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Nielsen, I. M. B.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150–157.
- (46) Beran, G. J. O. *J. Chem. Phys.* **2009**, *130*, 164115.
- (47) Le, H.-A.; Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2009**, *113*, 10527–10533.
- (48) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.
- (49) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. *J. Chem. Phys.* **2011**, *135*, 044123.
- (50) Richard, R. M.; Lao, K. U.; Herbert, J. M. *J. Phys. Chem. Lett.* **2013**, *4*, 2674–2680.
- (51) Mata, R.; Stoll, H. *Chem. Phys. Lett.* **2008**, *465*, 136–141.
- (52) Hua, D.; Leverentz, H. R.; Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 251–255.
- (53) Kurbanov, E. K.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2012**, *8*, 1–5.
- (54) Kurbanov, E. K.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2013**, *9*, 2617–2628.