

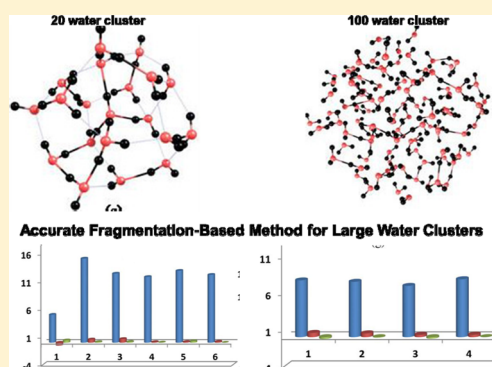
Dimers of Dimers (DOD): A New Fragment-Based Method Applied to Large Water Clusters

Arjun Saha and Krishnan Raghavachari*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

S Supporting Information

ABSTRACT: We have developed a dimer-based two-body interaction model, denoted “dimers of dimers” (DOD), for the study of water clusters using a fragmentation approach. By including a second layer at a lower level of theory (e.g., Hartree–Fock), our DOD2 model accurately reproduces the total energies of water clusters ranging in size from 16 to 100 molecules using the MP2 method with a variety of basis sets. The mean absolute error in the calculated DOD2 total energy for a broad range of calculations is only 0.83 kcal/mol. The size of the fragments is independent of the cluster size, and the number of fragments grows fairly slowly with the cluster size. Our results suggest that our method should be applicable for the study of large water clusters.



INTRODUCTION

Application of high level *ab initio* electronic structure calculations on large molecules (e.g., polypeptides, large water clusters, materials systems, etc.) has long been a bottleneck in computational chemistry. The steep computational scaling of high level electronic structure methods seriously limits the applicability of quantum chemical methods. For example, the computational scaling of the accurate CCSD(T)¹ method (coupled cluster theory with single and double excitations including a perturbative treatment of triple excitations; frequently called the “gold standard of quantum chemistry”) is $O(N^7)$, restricting the application of such electron correlated methods to relatively small systems (typically smaller than 25 heavy atoms). Although the rapid evolution of relatively inexpensive density functional theory (DFT) methods has made them most widely used electronic structure methods in computational chemistry, it is important to continue developments in traditional *ab initio* methods whose accuracy can be systematically improved. However, the steep computational scaling precludes brute force applicability for large molecules using faster workstations or computer clusters. Consistent efforts² are thus being made to reduce the computational expense associated with high-level quantum mechanical calculation. For example, substantial research has been focused on reducing the scaling (with the ultimate goal of attaining linear scaling) using local electronic structure treatments for large molecules.

In an alternative approach, many promising methods are being developed where attempts are made to fragment the whole system into smaller pieces.^{3–6} With the goal of avoiding high-level electronic structure calculations on the full (unfragmented) system, smaller and relatively cheap calculations are executed independently on the fragments (sub-

systems) of the whole system. Finally, the energy and other desirable properties of the unfragmented large molecule (supersystem) can be predicted by efficient accumulation of the numerous fragment calculations.

The inherent advantage of fragment-based methods lies in the fact that each of the small fragment calculations is not only cheap but can also be massively parallelizable owing to their complete independence. In an ideal case of dividing the supersystem into equal-sized fragments, the individual subsystem calculations are independent of the size of the supersystem, and the computational cost depends on the number of fragments (NF).⁷ The past decade has seen an intense activity in the field of fragment-based methods. Numerous groups^{3–6} have developed diverse ideas, retaining the local electronic environment of the fragments with different link atoms or capping methods, assigning self-consistently evaluated charge to each fragment to derive long-range interactions, etc. Although many fragment-based methods have been developed, the accuracy of a single black-box method (not requiring user intervention) for a broad class of large systems (bonded as well as nonbonded systems, biomolecules, clusters, extended systems, etc.) is not yet shown. It is beyond the scope of this report to discuss all the relevant literature, hence the authors recommend an article⁸ as a detailed and consolidated review on this field.

We have recently developed a general fragment-based approach (molecules-in-molecules, MIM)⁹ for the accurate evaluation of the energies of large molecules. In this method, we use a multilayer partitioning technique with multiple levels of theory using a generalized hybrid energy expression, similar

Received: June 6, 2013

Published: December 12, 2013

in spirit to the ONIOM¹⁰ methodology. The most important part of our proposed method is the use of a second layer (or more layers) using a cheaper method for the entire molecule (or larger fragments) to capture the long-range effects that improve the accuracy substantially. The basic equation for two-layer scheme used is

$$E(\text{real}) = E_{\text{low}}(\text{real}) + \sum_{i=1}^n [E_{\text{high}}(i) - E_{\text{low}}(i)]$$

where, $E_{\text{low}}(\text{real})$ denotes full system calculation at the low level of theory, $E_{\text{high}}(i)$ and $E_{\text{low}}(i)$ are energies of fragment i at the high and low levels of theory, and n is number of fragments that is determined by the distance threshold or effective radius. As in most fragment-based methods, the overcounting in the calculations due to overlapping fragments has to be carefully taken into account. This is done automatically through the “inclusion–exclusion principle” via the formation of primary and derivative subsystems in our method.⁹ While the MIM approach worked successfully⁹ for systems with a large network of covalent bonds, the applicability and accuracy for systems where different ranges of nonbonded interactions predominate (like clusters) has not yet been shown.

Many-body expansion is a well-known and popularly used method. The energy of a system (containing N fragments) in terms of any many-body expanded fragmentation method is given as

$$E = \sum_i E_i + \sum_{i<j} \Delta E_{ij} + \sum_{i<j<k} \Delta E_{ijk} + \dots \quad (1)$$

Here, E is the energy of the supersystem, E_i denotes the energy of subsystem i , ΔE_{ij} is the two-body interaction term ($E_{ij} - E_i - E_j$) between i and j , etc. The subsystems i and j are generally assumed not to overlap (different water molecules, for example). In most of the existing fragmentation methods, this is usually truncated at two-body interactions for practical reasons. Depending on the computational resources and the size of the system in hand, the costly three-body interaction may be incorporated to achieve better accuracy. This has been successfully carried out using the FMO method, though the authors do not suggest this as a practical approach for large molecules.¹¹ It is clear that direct introduction of four-body or higher terms will be formidable in the case of large systems.

In a recent paper, we proposed using a novel two-body expansion approach involving overlapping subsystems (many-overlapping body, MOB).¹² The inclusion of two-body interactions between overlapping fragments was suggested as a way of effectively including the critical many body terms to gain accuracy at a tractable cost. The applicability of MOB was demonstrated for bonded systems such as large polypeptides and alkane chains.¹² Similar ideas involving overlapping subsystems have also been suggested by Herbert and co-workers⁷ who have explored their applicability for nonbonded systems. While it will be interesting to explore a two-body expansion involving overlapping fragments for water clusters, determining the intersecting fragments in the case of a large system (e.g., $(\text{H}_2\text{O})_{200}$, vide infra) appears to generate a large number of subsystem calculations that may be a bottleneck in this approach.

In this article, we report the performance of a new many-body expanded fragmentation method on large water clusters. With an intention to achieve accuracy comparable to that including a four-body expansion, one can choose a more clever

fragmentation approach or choose a larger “effective monomer” size. We pursued the latter idea in this report. However, for simplicity, we decided to choose a model involving “non-overlapping effective monomers”. In particular, by choosing a water dimer as an effective monomer unit, we consider two-body interactions between pairs of such units. The basic idea relies on the simple fact that interaction between four entities can be reformulated in terms of interactions between two bodies by considering two entities at a time. We derive such a rescaled two-body approach in this paper and assess its performance for a variety of water clusters.

METHOD

Molecular clusters are stabilized by large numbers of non-bonded interactions and are challenging systems for successful application of fragment-based methods.¹³ To estimate the full electronic energy of a large molecular cluster with high accuracy, one must capture all the nonbonded interactions involved in the system. The conventional treatment of the two-body problem in fragment-based methods fails to approximate molecular clusters accurately. We suggest that successful execution of this task can be carried out by considering traditional pairwise interactions between water dimer units (vide infra). Our method differs from the “traditional two-body problem” only in the fact that instead of choosing one molecule at a time, we choose two. To be careful with the nomenclature, we termed our methods as “dimers of dimers” (DOD), while the conventional two-body problem is termed as “dimers of monomers” (DOM). In principle, this approach can be extended to give rise to a hierarchy where one can carefully expand the “monomer unit” size to multiple entities (i.e., the building blocks of the molecular clusters) followed by a conventional two-body treatment. Additionally, we decided to adopt a nonoverlapping “effective monomer” approach (i.e., each water molecule is present only in one “monomer unit”). It is important to note here that selection of building blocks (in a jungle of three-dimensional clusters) for use as a single entity or effective monomer unit in the two-body treatment is the most crucial step. In DOD, we define our “effective monomer units” (based on H-bond topology)¹⁴ as two building blocks (e.g., water molecules) based on their nonbonded interactions in that local environment.

The range of OH distances involved in H-bonding in our clusters is very broad. For the optimized geometries of the smaller water clusters (16, 17, 20, and 25, vide infra), the H-bond distances are typically in the range of 1.7–2.0 Å. However, for the cluster geometries of the larger clusters (40, 60, and 100) that are selected as snapshots from a molecular dynamics simulation (i.e., not optimized, vide infra), they occur over a much broader range: the relatively stronger H-bonds being in the range of 1.7–2.3 Å, while some of the comparatively weaker H-bond distances being in the range of 2.6–2.9 Å. The assignment of dimers in our model is done according to the following algorithm. The procedure that we follow is adapted from the available algorithms for the “pairing problem” in the field of graph theory and network optimization¹⁵ in computer science. A noteworthy aspect of the current algorithm is that the identified dimers are independent of the order in which the water molecules or the constituent atoms are specified. Our method of choosing the dimer building blocks is illustrated below.

- (1) To start with, all water molecules (monomers) are “available” to form dimers. Starting from the initial coordinates, intermolecular distances are computed. The shortest H-bonded distances between available monomers are used to assign “dimers”. Since each molecule is part of only one dimer, once a dimer is formed, the two associated monomers are removed from the “available” list. The procedure is followed until all “strongly bound dimers” are assigned using a cutoff distance. A cutoff O...H distance of 2.3 Å is sufficient for all the optimized systems, while a longer cutoff distance of 2.85 Å is used for the nonoptimized structures.
- (2) This procedure is not guaranteed to find optimum dimers or even assign each molecule to a dimer. In general, each water monomer will be fairly close to (i.e., within H-bonding distance of) several other monomers in our water clusters. In addition to the weaker H-bonds discussed above, there are many cases where a water molecule is not assigned to a dimer because its “partner” is already assigned to a different dimer. A simple example is a set of 4 water molecules in a row (1...2...3...4) where the shortest H-bonded distance corresponds to that between molecules 2 and 3. The algorithm will find only one dimer since the distance between molecules 1 and 4 is not appropriate for H-bonding. It is clear that finding two dimers (1...2 and 3...4) will be a better starting point to include the major interactions in the system. Thus we use the “augmenting path” procedure to identify other dimers.
- (3) In the augmenting path procedure, we start with a pair of closest monomers (as measured by the O...O distance) that are “available”. We use the same O...H cutoff distance as above, though a different cutoff criterion can be used if necessary, e.g., to identify weaker H-bonds. Using this distance criterion, the two monomers may be “H-bonded” directly to each other or via one dimer as in the example above or via more dimers. We have found that the use of a network augmenting path length up to three intervening dimers is sufficient to assign most water molecules to a distinct dimer in all our clusters. If no path is found for the selected pair of monomers, we search with the next closest pair of water molecules. If multiple paths are found, the optimal path is chosen based on a distance criterion to have the smallest difference between the H-bonded distances that are formed along a path and those that are broken along the path. Once a path is found, the connectivity of the dimers is redefined to be consistent with the new path, and the pair of monomers is removed from the available list. This is repeated until all monomers are assigned to distinct dimers or until no more dimers are found.
- (4) At the end of the procedure, the matching is considered “perfect” if all monomers are assigned to a dimer for even-numbered clusters or only one monomer in unmatched for odd-numbered clusters. This is achieved for all isomers of the smaller clusters 16, 17, 20, and 25 and some isomers of the larger clusters. However, in some isomers of the larger clusters (40, 60, and 100) there are a small number of monomers that cannot be assigned to a dimer. They are each connected to the nearest dimer to form a small number of trimers (up to four in the case of the larger clusters).

- (5) Each dimer, along with the small number of trimers, serves as an “effective monomer” in the evaluation of the many-body effects.

Once the effective monomers are generated, we perform a conventional two-body treatment based on a distance cutoff. The distance cutoff defines the effective radius of the sphere within which all the desired two-body correction terms are included (*vide infra*). On top of this approach, we have also employed a two-layer theory strategy (can also be extended to multilayers if needed) as with our MIM scheme.⁹ We find that coupling of “effective monomer generation” and two layers of theory leads to a very high level of accuracy. Generation of DODs and subsequent two-body treatment are executed by scripts containing a series of perl modules. All calculations were performed using the Gaussian Development Version program package.¹⁶

There are several previous approaches that have also used ideas similar to our two layer strategy to treat long-range interactions. In Stoll’s approach,¹⁷ the correlation energy is determined by an “incremental” method based on the local correlational energy. The incremental scheme involves localized atom-centered and/or bond orbitals obtained from a standard SCF calculation. Two-body interactions are then determined involving these selected localized orbitals. In “DOD”, localized orbitals are not needed, and each calculation deals with the total energy of individual subsystems that are assembled to derive the total energy of the supersystem. The multicentered ONIOM technique¹⁸ introduced by Tschumper is another popular fragmentation method. It is also a technique where inclusion–exclusion principle is used to blend different fragmentation approaches with ONIOM. While we also use two layers and inclusion–exclusion principle, our idea of using disjoint dimers as “effective monomers” is quite different. Incorporation of background molecular charges in fragment-based methodologies is another popular strategy. For example, Truhlar and co-workers^{3t} combine many-body expansion with electrostatic embedding through the use of point charges. In “DOD”, we have not used electrostatic embedding though longer range effects are included via the second layer.

Assessment of DOD. The treatment of water clusters using fragment-based methods has attracted enormous attention in the past few years. They are simple but important molecules to investigate H-bonding in chemical systems. Different research groups are using a variety of fragment-based approaches on different sizes of water clusters to either probe or calibrate their methods.¹⁹ In this section, we discuss the ability of DOD to approximate unfragmented *ab initio* results on a variety of small to large water clusters. Eight different sizes of water clusters have been selected for the current project ((H₂O)_n with *n* = 16, 17, 20, 25, 40, 60, 100, and 200). These systems are challenging because they represent a great range of structures and binding energies. We have selected MP2 method to illustrate the applicability of the method. Similar performance is expected for more highly correlated methods such as CCSD(T) (*vide infra*).

Table 1 shows absolute energy errors for all the systems investigated. DOM1 and DOD1 denote regular two-body and our proposed two-body method, respectively, employing only one level of theory. Note that DOM1 results are reported without a distance cutoff, while DOD1 results use a distance cutoff of 5 Å. We use the closest approach scheme (i.e., the closest distance between any of the atoms in the two systems) to apply the 5 Å threshold. Similarly, DOM2 and DOD2

Table 1. Absolute Energy Errors (kcal/mol) for All of the Systems^a

size	system	DOM1	DOM2	DOD1	DOD2
16	I	9.03	2.03	2.15	0.65
	II	−34.34	−1.24	−2.50	0.16
	III	−33.47	−0.82	−0.42	0.26
	IV	−32.68	−0.69	8.96	0.89
	V	−32.06	−0.50	7.7	1.00
	VI	20.70	1.92	−4.00	0.92
17	I	−26.60	1.23	9.71	2.03
	II	−37.72	−0.89	9.82	0.15
	III	−47.01	−1.79	5.59	1.00
20	I	−14.68	1.57	5.71	1.21
	II	−45.65	−1.75	−2.19	0.38
	III	−37.05	−2.03	−0.58	0.48
	IV	−35.73	−0.38	−6.04	0.13
	V	−39.54	−0.48	7.88	0.80
	VI	−36.96	−0.55	−5.33	0.01
25	I	−51.77	−1.57	6.03	−0.88
	II	−52.78	−1.35	−15.62	−0.40
	III	−61.59	−1.92	−1.47	0.27
	IV	−51.02	−1.05	0.72	0.12
40	I	−15.23	4.24	−9.47	1.38
	II	−17.40	5.54	4.11	0.76
	III	−12.81	3.73	−10.93	0.94
	IV	−7.25	5.14	−4.68	3.37
	V	−23.65	3.27	−6.27	2.13
60	I	−33.18	−1.35	−0.35	1.62
	II	−14.08	−1.29	−20.43	0.51
	III	−35.99	−2.82	−19.58	0.72
	IV	−37.76	−1.89	−43.69	−0.18
	V	−11.39	−0.65	−10.22	0.71
100	I	−58.05	−4.38	−49.97	−1.93
	II	−56.98	−3.97	−62.60	−0.51
	III	−56.93	−2.69	−58.13	−1.62
	IV	−51.30	−2.26	−56.92	0.11

^aSmall clusters (16, 17, 20, 25, and 40 water clusters) were treated with MP2/6-311G(d,p):HF/6-311G(d,p). Large clusters were treated with MP2/6-31G:HF/6-31G. All the DOM2 and DOD2 methods were performed with a distance cut-off of 5 Å.

indicate the inclusion of a second layer where a cheaper level of theory is used to account for the missing long-range interactions. The geometries for 16, 17, 20, and 25 clusters (Figure 1) have been taken from literature refs 19 and 20 and represent calculated minima. The rest of them are generated through a molecular dynamics package (NAMD). Their geometries have not been optimized at any quantum chemical level, and they may provide different and more challenging environments. The geometries of all the clusters are included in the Supporting Information. We use MP2 for the high level of theory and HF for the low level of theory. For simplicity, we use the same basis set in both high and low levels of theory. Since HF calculations are subsets of MP2, this provides some efficiency in avoiding duplication. In Table 1, we use the 6-311G(d,p) basis set, for all the small clusters (ranging from 16 to 25) for calibration purposes. The 40-mer cluster is also treated with the 6-311G(d,p) basis set, while the remaining larger clusters have been treated with 6-31G basis set to show the successful performance of DOD. Thus, the smaller clusters are investigated with MP2/6-311G(d,p):HF/6-311G(d,p), while the larger systems are investigated at MP2/6-31G:HF/

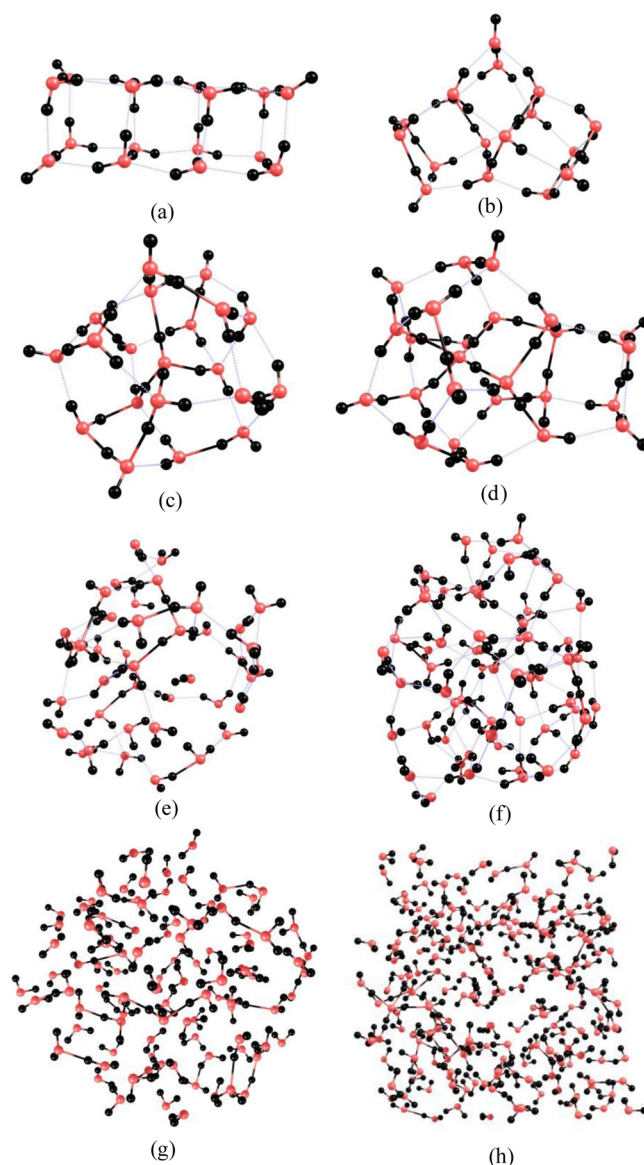


Figure 1. Each figure represents one isomer from different sizes of clusters, i.e., (a) 16, (b) 17, (c) 20, (d) 25, (e) 40, (f) 60, (g) 100, and (h) 200 water clusters studied.

6-31G, where “:” is used to separate the high and low levels of theory.

Table 1 also shows the importance of the idea of adding a “second layer” in fragment-based methods. On going from DOM1 to DOM2 (similarly DOD1 to DOD2), the error improves remarkably regardless of the size of the system. This clearly suggests that adding a cheaper level of theory as a second layer in fragment-based methods (if computationally feasible) can be more effective than clever fragmentation schemes. Surprisingly, DOM2 performs quite well for all the clusters studied with a mean absolute deviation of 2.03 kcal/mol. However, the error for 11 species is larger than 2 kcal/mol. DOD2 reduces the mean absolute deviation further to 0.83 kcal/mol with only 3 species having an error >2 kcal/mol. This is impressive for the total energy and establishes the applicability of DOD2 for water clusters of different sizes.

Table 2 shows the average absolute energy errors averaged over all the isomers for different clusters. We observe that DOD2 result significantly improves over DOM2. The results

Table 2. Average Absolute Energy Errors (kcal/mol) for All of the Systems^a

size	DOM1	DOM2	DOD1	DOD2
(16) ₆	27.04	1.20	4.28	0.64
(17) ₃	37.11	1.31	8.37	1.06
(20) ₆	34.93	1.13	4.62	0.50
(25) ₄	54.29	1.47	5.96	0.41
(40) ₅	15.14	4.38	7.10	1.71
(60) ₅	26.48	1.60	18.85	0.75
(100) ₄	55.81	3.33	56.90	1.04

^aSmall clusters (16, 17, 20, 25, and 40 water clusters) were treated with MP2/6-311G(d,p):HF/6-311G(d,p). Large clusters were treated with MP2/6-31G:HF/6-31G. The number of isomers for each cluster size is shown as a subscript.

for the different cluster systems are illustrated graphically in Figure 2. It is noteworthy to point out here that the dependence of absolute energy errors on cluster size does not exhibit any trend for small clusters, whereas it increases with the cluster size (Figure 2h) in case of large clusters. In

addition, to demonstrate the applicability of DOD for larger systems, we have also applied DOD2 on a water cluster containing 200 molecules (*vide infra*).

Relative Energy and Binding Energy. One of the effective ways to judge a fragment method is to find out how accurately it can reproduce the relative energy profiles of different isomers. Since energy differences are important in electronic structure methods, we carried out relative energy analysis on small cluster models with our fragment method. While absolute energy errors may be difficult to converge, relative energies are frequently easier to calculate since systematic errors due to fragmentation get automatically canceled. Figure 3 shows relative energy profiles for all the small clusters. We observe that while DOM2 performs reasonably well, there are some instances where the relative energy orderings are not reproduced. However, DOD2 curves track the full calculations quite closely in almost all clusters. We also evaluated the binding energies for all systems with both DOM and DOD. Table 3 shows the performance of all the methods in calculating MP2 binding energies. Table 4 lists the MP2 binding energies averaged over the different isomers for

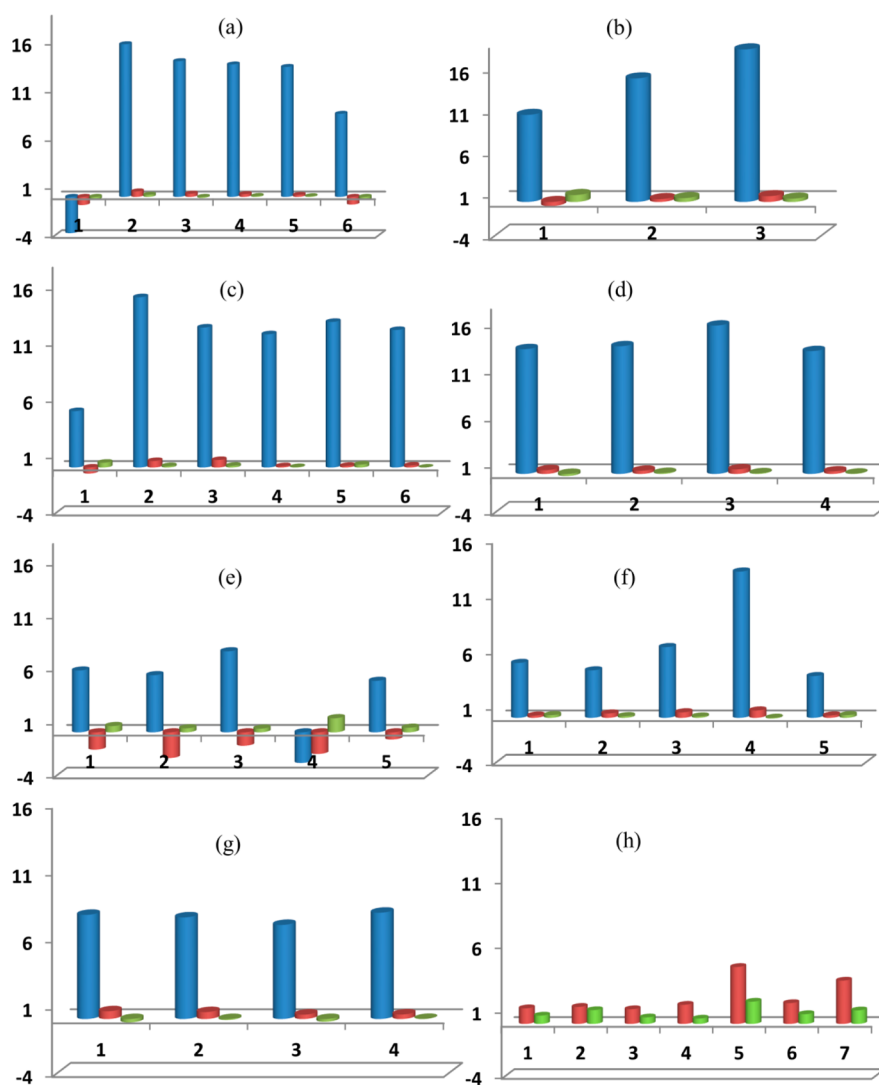


Figure 2. Percent error in MP2 binding energies for (a) 16, (b) 17, (c) 20, (d) 25, (e) 40, (f) 60, and (g) 100 water clusters. (h) Average absolute energy errors for all clusters (clusters are arranged in increasing size along the *x*-axis). Blue, red, and green denote errors associated with DOM1, DOM2, DOD2, respectively.

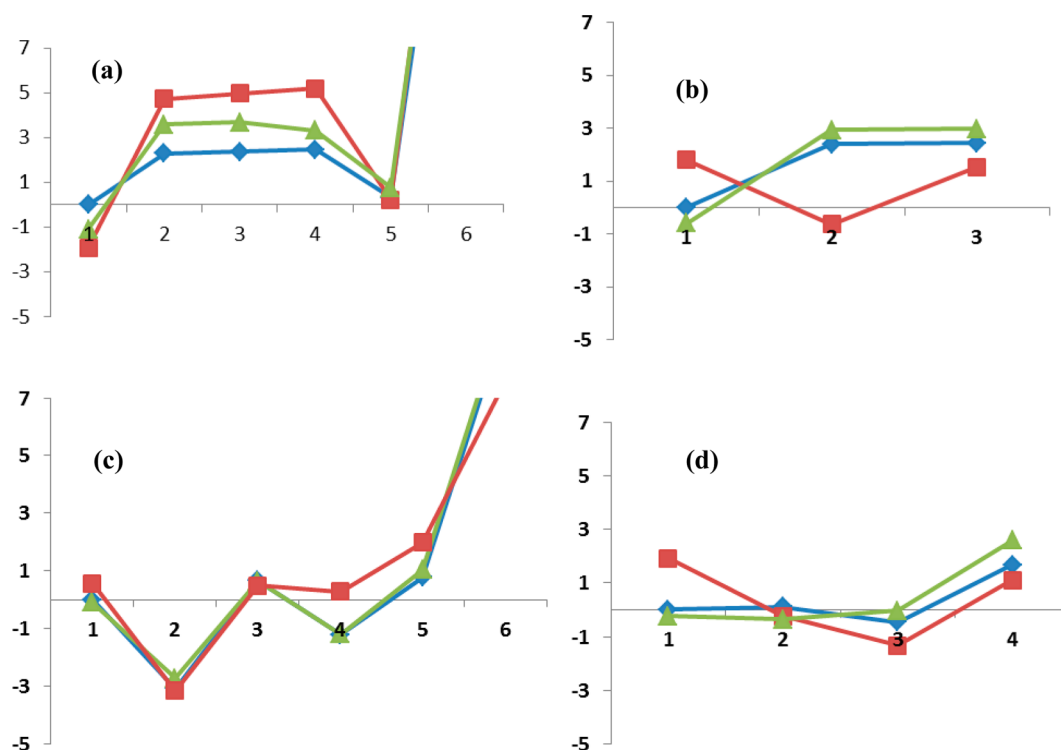


Figure 3. Relative energy profiles for (a) 16, (b) 17, (c) 20, and (d) 25 water clusters. Blue, red, and green indicate unfragmented MP2, DOM2, and DOD2 results, respectively. Clusters were treated with MP2/6-311G(d,p):HF/6-311G(d,p). The highest-energy isomer (the last point) was not accommodated in the specified range of y-axis in case of (a) and (c), so they are invisible. But the trend of the relative energy profiles for different methods (the main focus here) is still clearly illustrated.

each cluster size. We observe the same trend as in the case of absolute energy errors. In fact DOD2 can reproduce the total binding energies accurately regardless of the cluster size. Figure 2 compares percent errors in binding energies associated with all the methods proposed here.

Basis Set Dependence. We have calibrated our method with both smaller and larger basis sets on all the small clusters. It is important to point out here is that DOD is highly robust with respect to both methods and basis sets. Tables 5 and 6 show the performance of both DOM and DOD with Dunning-style augmented basis sets (aug-cc-pVDZ and aug-cc-pVTZ). One can easily notice that with the use of larger basis sets, absolute energy errors are well within 1 kcal/mol. It is interesting that the results for the larger basis sets appear to be consistently better than the results with the smaller basis sets. The same trend of accuracy level is preserved in estimating binding energies and relative energy profiles. The performance in reproducing relative energy profiles is shown in Figure 4. It is clear that DOD2 outperforms DOM2 to estimate the relative energies for such water clusters.

Distance Cutoff. Since our goal is to perform a very high level of ab initio calculation within a given amount of time, we have carried out a rigorous analysis within the DOD framework to find out the optimal procedure (i.e., computationally most efficient) to calculate the energies. It is important to point out here that the scaling of any fragment-based method depends on the number of fragments as well as the size of each fragment. In DOD, we have focused on both of these factors to reduce the scaling. The most important point is that the largest MP2 calculation needed is independent of the size of the molecule. This is a consequence of our redefinition of “monomer units” as appropriately chosen water dimers. Thus even for $(\text{H}_2\text{O})_{200}$,

the largest MP2 calculation involves only water tetramers (dimer + dimer) or pentamers (dimer + trimer if any trimer is present). Thus the overall scaling of the MP2 part only depends on the number of fragments (scales as $O(N^2)$ without a distance cutoff and better than $O(N^2)$ due to the 5 Å cutoff, vide infra). Note, however, that the HF calculation on the whole molecule scales formally as $O(N^4)$, though the effective scaling is much better.

Figure 5 shows the performance of DOD with increasing distance parameter. With increasing cutoff distance, the number of independent fragment calculations also increases. To balance the cost and accuracy, we analyze the performance of our method as a function of the distance cutoff. Table 7 shows the effect of variable distance parameter on all the small clusters considered in this report. Based on our analysis, we have used a distance cutoff of 5 Å which provides errors very close to that obtained by using infinite distance cutoff. It is important to point out that the longer range effects that are missing at 5 Å are included indirectly via the second layer. Overall, from both points of view (size and number of independent fragment calculations), DOD offers a highly affordable alternative for high level ab initio calculation.

To illustrate the performance of DOD for larger systems, we performed a MP2/6-311G(d,p):HF/6-311G(d,p) calculation on $(\text{H}_2\text{O})_{200}$. For this system, the entire calculation took <7 h on a desktop i7-based workstation. Of that total, the HF/6-311G(d,p) part of the calculation consumed about 3 h. The complete 6000 basis function MP2/6-311G(d,p) calculation was not feasible on that workstation. Larger calculations (in terms of cluster size as well as basis set size) are clearly possible and will be carried out in the future. In the case of much larger systems where the HF calculation becomes a bottleneck, we can

Table 3. Percent Error in MP2 Binding Energies for All of the Systems^a

size	system	DOM2	DOD2
16	I	−0.84	−0.27
	II	0.57	0.22
	III	0.34	−0.10
	IV	0.29	0.09
	V	0.21	0.07
	VI	−0.79	−0.31
17	I	−0.48	0.84
	II	0.35	0.47
	III	0.70	0.41
20	I	−0.53	0.43
	II	0.58	0.13
	III	0.67	0.16
	IV	0.12	0.04
	V	0.15	0.27
	VI	0.18	0.00
25	I	0.40	−0.23
	II	0.34	0.10
	III	0.49	0.07
	IV	0.27	0.03
40	I	−1.62	0.59
	II	−2.41	0.38
	III	−1.26	0.35
	IV	−2.03	1.33
	V	−0.67	0.43
60	I	0.20	0.24
	II	0.39	0.15
	III	0.50	0.12
	IV	0.66	−0.06
	V	0.21	0.24
100	I	0.58	−0.25
	II	0.52	−0.07
	III	0.33	−0.20
	IV	0.34	0.01

^aSmall clusters (16, 17, 20, 25, and 40 water clusters) were treated with MP2/6-311G(d,p):HF/6-311G(d,p). Large clusters were treated with MP2/6-31G:HF/6-31G. All the DOD2 methods were performed with a distance cut-off of 5 Å. All the DOM2 methods were performed with full distance cut-off.

Table 4. Average Percent Binding Energy Errors (kcal/mol) for All of the Systems^a

size	DOM2	DOD2
(16) ₆	0.51	0.17
(17) ₃	0.51	0.57
(20) ₆	0.37	0.17
(25) ₄	0.38	0.10
(40) ₅	1.60	0.62
(60) ₅	0.40	0.16
(100) ₄	0.45	0.13

^aSmall clusters (16, 17, 20, 25, and 40 water clusters) were treated with MP2/6-311G(d,p):HF/6-311G(d,p). Large clusters were treated with MP2/6-31G:HF/6-31G.

use a similar strategy using a third level of theory to evaluate the HF energy itself using larger fragments. This will be considered in our future work.

Beyond MP2. Finally, while the MP2 procedure was used to illustrate the method, the same DOD scheme can be used with other more expensive methods such as CCSD(T). In

Table 5. Absolute Energy Errors (kcal/mol) and Percent Binding Energy Errors for (H₂O)₁₆, (H₂O)₁₇, and (H₂O)₂₀^a

size	system	absolute energy error (kcal/mol)		percent binding energy error (kcal/mol)	
		DOM2	DOD2	DOM2	DOD2
16	I	1.43	−0.26	−0.97	−0.08
	II	1.53	0.42	−0.88	−0.26
	III	1.22	0.12	−0.71	−0.10
	IV	1.30	0.29	−0.75	−0.18
	V	1.07	0.26	−0.62	−0.19
	VI	1.88	0.90	−1.11	−0.53
17	I	1.35	0.05	0.73	−0.03
	II	1.10	−0.18	0.59	0.07
	III	1.18	−0.44	0.63	0.24
20	I	1.83	0.25	−0.90	−0.12
	II	1.39	0.00	−0.62	0.00
	III	1.56	0.07	−0.70	−0.03
	IV	1.21	−0.22	−0.55	0.10
	V	1.54	−0.28	−0.68	0.12
	VI	1.15	−0.38	−0.52	0.17

^aAll the clusters were treated with MP2/aug-cc-pVDZ: HF/aug-cc-pVDZ. All the DOD2 methods were performed with a distance cut-off of 5 Å. All the DOM2 methods were performed with full distance cut-off.

Table 6. Absolute Energy Errors (kcal/mol) and Percent Binding Energy Errors for (H₂O)₁₆ and (H₂O)₁₇^a

size	system	absolute energy error (kcal/mol)	percent binding energy error (kcal/mol)
16	I	−0.39	0.26
	II	−0.08	0.04
	III	−0.28	0.16
	IV	−0.23	0.13
	V	−0.24	0.14
	VI	0.34	−0.20
17	I	−0.16	0.56
	II	−0.41	0.44
	III	−0.28	0.18

^aAll the clusters were treated with MP2/aug-cc-pVTZ:HF/aug-cc-pVTZ levels using the DOD2 method (with 5 Å distance cut-off).

order to demonstrate the applicability of DOD2 to higher order correlated methods, we have performed CCSD(T)/aug-cc-pVTZ:MP2/aug-cc-pVTZ calculations on an illustrative (H₂O)₁₆ isomer using the large aug-cc-pVTZ basis set. The full benchmark calculations on this system (1472 basis functions) have been reported by Yoo et al.²¹ We selected the isomer labeled 4444-a that has the lowest CCSD(T) energy among the isomers studied by Yoo et al. The CCSD(T)/aug-cc-pVTZ total energy obtained using DOD2 fragmentation using a cutoff distance of 3.5 Å (−1221.75047 hartree) has an error of only 0.41 kcal/mol relative to the exact value (−1221.74981 hartree). Within our model it involves 16 tetramer calculations and 8 dimer calculations and was carried out on an i7-based desktop workstation (~200 h). For comparison, Yoo et al.²¹ have reported that the direct CCSD(T)/aug-cc-pVTZ calculation took 400 000 processor hours, though it was carried out on a massively parallel platform. These results suggest that accurate CCSD(T) calculations on large clusters are possible even with modest workstations, since only tetramers or pentamers have to be treated at the CCSD(T) level. Overall, “DOD” has effective

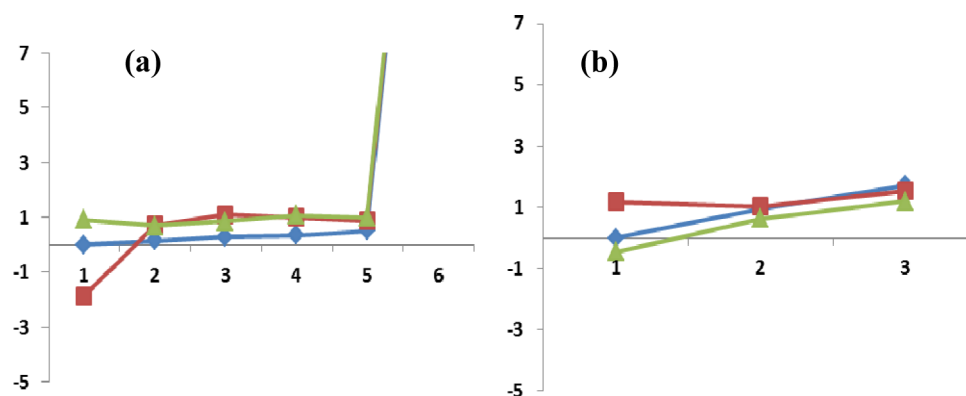


Figure 4. Relative energy profiles for (a) 16 and (b) 17 clusters. Blue, red, and green indicate unfragmented MP2, DOM2, and DOD2 results, respectively. Clusters were treated with MP2/aug-cc-pVTZ:HF/aug-cc-pVTZ.

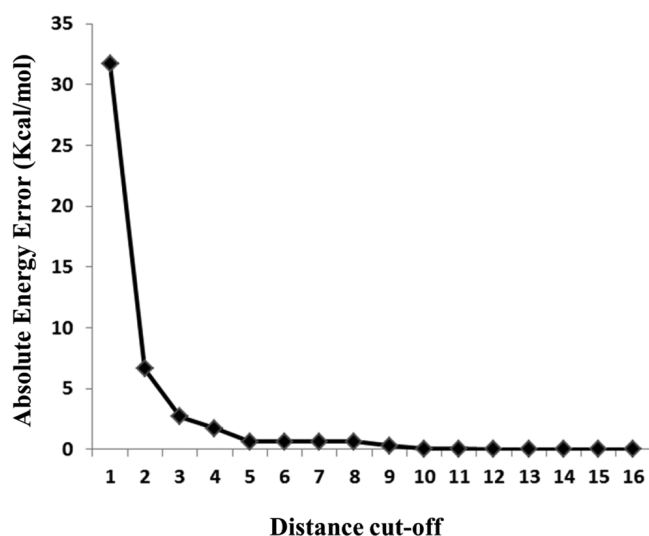


Figure 5. Convergence of absolute energy errors (kcal/mol) with distance cutoff for 20-I water cluster. Cluster was treated with MP2/aug-cc-pVDZ:HF/aug-cc-pVDZ levels in DOD2 at each stationary point.

Table 7. Convergence of Absolute Energy Errors (kcal/mol) with Distance Cut-off^a

size	system	cut-off (3.5 Å)	cut-off (5 Å)	cut-off (7 Å)	cut-off (full)
16	I	-1.49	-0.26	0.12	0.12
	II	-0.09	0.42	0.42	0.45
	III	-1.60	0.12	0.17	0.17
	IV	-1.89	0.29	0.34	0.34
	V	-0.57	0.26	0.33	0.33
	VI	0.38	0.90	0.90	0.93
17	I	-0.60	0.05	0.05	0.05
	II	-0.32	-0.18	-0.13	-0.13
	III	-1.09	-0.44	-0.45	-0.45
20	I	-0.01	0.32	0.57	0.62
	II	-2.02	-0.00	0.10	0.11
	III	-1.87	0.03	0.20	0.21
	IV	-0.79	-0.14	-0.09	-0.10
	V	-1.90	-0.31	0.13	0.13
	VI	-1.34	-0.27	-0.18	-0.16

^aAll the clusters were treated with MP2/aug-cc-pVDZ: HF/aug-cc-pVDZ using DOD2.

scaling to be useful for large water clusters using accurate high level electron correlation methods.

Exploring Potential Energy Surfaces. Thus far, we have demonstrated that the DOD method is capable of yielding accurate total energies of water clusters at fixed geometries. However, several important points need to be considered before the method is useful in more complex applications. The first point involves potential energy surfaces (geometry optimization, for example) where a change in geometry can cause a different selection of dimers using our algorithm. While we have not explored this aspect yet, it may be useful to choose the initial dimers at the starting geometry and optimize the geometry for several steps without redefining the dimers after each displacement. This static treatment will be somewhat akin to molecular mechanics type schemes where bond types are defined. The second point involves the hard distance threshold employed. Slight change in initial geometry can result in an addition or removal of few atoms or molecules inside the specified distance threshold and can lead to discontinuous potential energy surfaces. It will be useful in the future to explore other nonrigid criteria (e.g., “dynamic fragmentation”)²² that can yield smoother potential energy surfaces. For many of the smaller systems, we can easily calibrate the results since the entire molecule can be included in forming the dimers of dimers (i.e., no distance cutoff), or the parameter can be adjusted to incorporate all the important interactions resulting in a smoother potential energy surface. We plan to explore these points in the future. However, we point out that many of these points are common for most fragment-based methods, and “DOD” is not an exception.

CONCLUSION

We have proposed a method which combines effective monomer generation coupled with a second layer of theory (DOD2) within a many-body framework to estimate unfragmented MP2 energies of large clusters. This substantially improves the results over the conventional two-body interaction model. The method has been rigorously tested for water clusters of different sizes. The mean absolute deviation in the calculated MP2 total energies for a broad range of cluster sizes and isomers at the DOD2 level is only 0.83 kcal/mol. Successful application of DOD2 on 100–200 water clusters really shows promise to evaluate high-level ab initio methods on very large clusters (>1000) in near future.

■ ASSOCIATED CONTENT

■ Supporting Information

This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge support from the NSF grant CHE-1266154 at Indiana University. We thank Prof. Balaji Raghavachari (University of Texas in Dallas) for stimulating discussions about the pairing problem and network optimization algorithms. We also thank Dr. Raghunath O. Ramabhadran for several helpful suggestions.

■ REFERENCES

- (1) Raghavachari, K.; Trucks, G. W. *Chem. Phys. Lett.* **1989**, *157*, 6.
- (2) (a) He, X.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 405–411. (b) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 114108. (c) Kobayashi, M.; Imamura, Y.; Nakai, H. *J. Chem. Phys.* **2007**, *127*, 074103. (d) Saravanan, C.; Shao, Y.; Baer, R.; Ross, P. N.; Head-Gordon, M. *J. Comput. Chem.* **2003**, *24*, 618. (e) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51. (f) Burant, Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1996**, *105*, 8969.
- (3) (a) Fujimoto, H.; Koga, N.; Fukui, K. *J. Am. Chem. Soc.* **1981**, *103*, 7452–7457. (b) Fedorov, D. G.; Kitauro, K. *J. Phys. Chem. A* **2007**, *111*, 6904–6914. (c) Pruitt, S. R.; Fedorov, D. G.; Kitauro, K.; Gordon, M. S. *J. Chem. Theory Comput.* **2010**, *6*, 1–5. (d) Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2010**, *132*, 094102. (e) Isegawa, M.; Wang, B.; Truhlar, J. *Chem. Theory Comput.* **2013**, *9*, 1381–1393. (f) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112. (g) Netzloff, H. M.; Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 134113. (h) Beran, G. J. O. *J. Chem. Phys.* **2009**, *130*, 164115. (i) Netzloff, H. M.; Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 134113. (j) Huang, L.; Massa, L.; Karle, J. *Proc. Natl. Acad. Sci.* **2006**, *103*, 1233. (l) Le, H.; Tan, H.; Ouyang, J.; Bettens, P. A. *J. Chem. Theory Comput.* **2012**, *8*, 469–478. (m) Collins, M. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744–7751. (n) Richard, R. M.; Herbert, J. M. *J. Chem. Theory Comput.* **2013**, *9*, 1408–1416. (o) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. *J. chem. phys.* **2011**, *135*, 044123. (p) Vysotskiy, V. P.; Cederbaum, L. S.; Sommerfeld, T.; Voora, V. K.; Jordan, K. D. *J. Chem. Theory Comput.* **2012**, *8*, 893–900. (q) Temelso, B.; Shields, G. C. *J. Chem. Theory Comput.* **2011**, *7*, 2804–2817. (r) Friedrich, J.; Dolg, M. *J. Chem. Theory Comput.* **2009**, *5*, 287–294. (s) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348. (t) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46–53. (u) Huff, E. M.; Pulay, P. *Mol. Phys.* **2009**, *107*, 1197–1207. (v) Friedrich, J. *J. Chem. Theory Comput.* **2012**, *8*, 1597.
- (4) (a) Fedorov, D. G.; Ishida, T.; Kitauro, K. *J. Phys. Chem. A* **2005**, *109*, 2638. (b) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitauro, K. *J. Phys. Chem. A* **2007**, *111*, 2722. (c) Kitauro, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701. (d) Fedorov, D. G.; Slipchenko, L. V.; Kitauro, K. *J. Phys. Chem. A* **2010**, *114*, 8742. (e) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165. (f) Gadre, S. R.; Ganesh, V. *J. Theor. Comput. Chem.* **2006**, 5835. (g) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109. (h) Kavathekar, R.; Khire, S.; Ganesh, V.; Rahalkar, A. P.; Gadre, S. R. *J. Comput. Chem.* **2009**, *30*, 1167. (i) Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2008**, *129*, 234101. (f) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405. (g) Addicoat, M. A.; Collins, M. A. *J. Chem. Phys.* **2009**, *131*, 104103. (h) Collins, M. A. *J. Chem. Phys.* **2006**, *125*, 104104. (i) Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102. (k) Netzloff, H. M.; Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 134113. (l) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293. (m) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276. (n) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33. (o) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1. (p) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573. (q) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 683. (r) Kurbanov, E. K.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2013**, *9*, 2617. (13) (s) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193. (t) Huang, L.; Massa, L.; Karle, J. *Int. J. Quantum Chem.* **2005**, *103*, 808. (u) Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. *J. Phys. Chem. A* **2008**, *112*, 10864. (v) Hua, S. G.; Hua, W. J.; Li, S. H. *J. Phys. Chem. A* **2010**, *114*, 8126.
- (5) (a) (b) Beran, G. J. O.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480. (c) Wen, S.; Nanda, K.; Hunag, Y.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578. (d) Nanda, K.; Beran, G. J. O. *J. Chem. Phys.* **2012**, *137*, 174106.
- (6) (a) Wang, Y.; Sosa, C. P.; Cembran, A.; Truhlar, D. J.; Gao, J. *J. Phys. Chem. B* **2012**, *116*, 6781. (b) Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657. (c) Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346. (d) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890. (e) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108. (f) Song, L.; Han, J.; Lin, Y. L.; Xie, W.; Gao, J. *J. Phys. Chem. A* **2009**, *113*, 11656. (g) Wierchowski, S. J.; Kofke, D. A.; Gao, J. *J. Chem. Phys.* **2003**, *119*, 7365. (h) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459. (i) Cembran, A.; Song, L.; Mo, Y.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 2702. (j) Bates, D. M.; Smith, J. R.; Tschumper, G. S. *J. Chem. Theory Comput.* **2011**, *7*, 2753. (k) Mei, Y.; Ji, C. G.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *125*, 094906. (l) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
- (7) Herbert, J. M.; Richard, M. R. *J. Chem. Phys.* **2012**, *137*, 064113.
- (8) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2011**, *112*, 632–672.
- (9) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.
- (10) Svensson, M.; Humbel, S.; Froese, R.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- (11) (a) Fedorov, D. G.; Kitauro, U. *J. Chem. Phys.* **2004**, *120*, 6832. (b) Pruitt, S. R.; Addicoat, M. A.; Collins, M. A.; Gordon, M. S. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7752–7764. (c) Pruitt, S. R.; Steinmann, C.; Jensen, J. H.; Gordon, M. S. *J. Chem. Theory Comput.* **2013**, *9*, 2235.
- (12) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *8*, 2669–2675.
- (13) Sahu, N.; Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2013**, *138*, 104101.
- (14) (a) Kuo, J.; Singer, S. J. *Phys. Rev.* **2003**, *E67* (1), 016114. (b) Knight, C.; Singer, S. J. *J. Phys. Chem. B* **2005**, *109*, 21040. (c) Knight, C.; Singer, S. J. *J. Phys. Chem. A* **2009**, *113*, 12433.
- (15) (a) Khuller, S.; Raghavachari, B. *Handbook of Combinatorial Optimization*; Springer Science+Business Media: New York, 2013. (b) Lovasz, L.; Plummer, M. D. *Matching Theory*; Elsevier Science Publishers B. V.: New York, 1986.
- (16) 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 DV*, revision H.08; Gaussian, Inc.: Wallingford, CT, 2010.

(17) (a) Stoll, H. *Phys. Rev. B* **1992**, 46, 6700. (b) Stoll, H.; Paulus, B.; Fulde, P. *J. Chem. Phys.* **2005**, 123, 144108.

(18) Hopkins, B. W.; Tschumper, G. S. *J. Comput. Chem.* **2003**, 24, 1563.

(19) Furtado, J. P.; Rahalkar, A. P.; Shanker, S.; Bandyopadhyay, P.; Gadre, S. R. *J. Phys. Chem. Lett.* **2012**, 3, 2253–2258.

(20) Allen, F. H. *Acta Crystallogr.* **2002**, B58, 380–388.

(21) Yoo, S.; Apra, E.; Zeng, X. C.; Xantheas, S. J. *Phys. Chem. Lett.* **2010**, 1, 3122–3127.

(22) Lange, A. W.; Voth, A. G. *J. Chem. Theory Comput.* **2013**, 9, 4018.