

# Electrostatically Embedded Many-Body Correlation Energy, with Applications to the Calculation of Accurate Second-Order Møller-Plesset Perturbation Theory Energies for Large Water Clusters

Erin E. Dahlke and Donald G. Truhlar\*

*Department of Chemistry and Supercomputing Institute, University of Minnesota,  
Minneapolis, Minnesota 55455-0431*

## **Abstract:**

The electrostatically embedded many-body expansion (EE-MB), previously applied to the total electronic energy, is here applied only to the electronic correlation energy (CE), combined with a Hartree-Fock calculation on the entire system. The separate treatment of the Hartree-Fock and correlation energies provides an efficient way to approximate correlation energy for extended systems. We illustrate this here by calculating accurate Møller-Plesset second-order perturbation theory (MP2) energies for a series of clusters ranging in size from 5 to 20 water molecules. In this new method, called EE-MB-CE, where MB is pairwise additive (PA) or three-body (3B), the full Hartree-Fock energy of a system of  $N$  monomers is calculated (i.e., the many-body expansion is carried out to  $N^{\text{th}}$  order), while the EE-MB method is used to calculate the correlation energy of the system. We find that not only does this new method lead to better energetics than the original EE-MB method, but that one is able to obtain excellent agreement with full MP2 calculations by considering only a two-body expansion of the correlation energy, leading to a considerable savings in computational time as compared to the three-body expansion. Additionally, we propose the use of a cutoff to further reduce the number of two-body terms that must be calculated, and we show that if a cutoff of 6 Å is used one can eliminate up to 44% of the pairs and still calculate energies to within 0.1% of the net interaction energy of the full cluster.

## 1. Introduction

The application of post-Hartree–Fock correlated levels of electronic structure theory (e.g., second-order Møller-Plesset perturbation theory, MP2,<sup>1</sup> coupled cluster theory with single and double excitations, CCSD,<sup>2</sup> or CCSD with quasiperturbative connected triple excitations, CCSD(T)<sup>3</sup>) to systems containing tens to hundreds of atoms provides a grand challenge to the chemical community because of the rapid scaling of the computational cost of such methods with respect to system size. For example, CCSD(T), CCSD, and MP2 scale as  $N^7$ ,  $N^6$ , and  $N^5$ , respectively, where  $N$  is the number of atoms.<sup>4</sup> To meet the challenge of calculating the correlation energy of large systems, there has been considerable interest in trying to develop methods to make the problem more tractable. One approach is to reduce the scaling by using localized molecular orbitals. Such methods include the natural scaling coupled-cluster,<sup>5</sup> divide-and-conquer methods,<sup>6</sup> cluster-in-molecules methods,<sup>7</sup> as well as many others (See, for example, Refs. 8-12 and references within.). Another is to break up a large system into a great number of smaller and more manageable subsystems as in the fragment molecular orbital,<sup>13, 14</sup> systematic molecular fragmentation,<sup>15, 16</sup> and conjugated caps methods.<sup>17-19</sup>

In past work we have presented our own fragmentation-based method, the electrostatically embedded many-body (EE-MB) method,<sup>20</sup> for calculating the energies of large molecular clusters. The EE-MB method calculates the total energy of a large cluster by taking a linear combination of the energies of monomers and dimers (in the case of the EE-PA method, where PA denotes pairwise additive) or monomers, dimers, and trimers (in the case of the EE-3B method, where 3B denotes a three-body approximation), with a key element being that each monomer, dimer, or trimer is

embedded in a field of point charges representing the other  $N - 1$ ,  $N - 2$ , or  $N - 3$  monomers. (A monomer can be defined as a single molecule or as a collection of molecules, and the method can be extended to allow monomers to be portions of large molecules, such as the monomers of a polymer. In the examples discussed in the present paper, a monomer will be a single water molecule.) Using the EE-MB method we were able to reproduce the absolute correlated interaction energy of a cluster of 21 water molecules to within 2%, by using the EE-PA method, and to within 0.2% when the EE-3B method was used.<sup>20</sup> In the present article we present an extension of the electrostatically embedded many-body method, to be called electrostatic embedding of the many-body correlation energy, to predict the MP2 correlation energy for a series of water clusters ranging in size from 5 to 20 water molecules. Because MP2 is the simplest of the correlated methods it provides a good starting point for testing this new method, and since it is the least expensive of the post-Hartree–Fock methods it allows us to compare our results directly to the MP2 energy for clusters containing 10 – 20 water molecules; this would not be possible for the more expensive post-Hartree–Fock methods.

## 2. Theory

By using the electrostatically embedded many-body expansion, the energy of a system of  $N$  interacting particles (monomers) can be written as:

$$V = V_1 + V_2 + V_3 + \cdots + V_N \quad (1)$$

where

$$V_1 = \sum_i E_i \quad (2)$$

$$V_2 = \sum_{i>j} E_{ij} - E_i - E_j \quad (3)$$

$$V_3 = \sum_{i>j>k} [E_{ijk} - E_i - E_j - E_k - (E_{ij} - E_i - E_j) - (E_{ik} - E_i - E_k) - (E_{jk} - E_j - E_k)] \quad (4)$$

and so forth, where  $E_i$ ,  $E_{ij}$ , and  $E_{ijk}$  are the energies of monomers, dimers, and trimers that are embedded in a sea of point charges representing the other  $N - 1$ ,  $N - 2$ , or  $N - 3$  monomers, and  $V_n$  (with  $n > 1$ ) denotes the difference between the  $n$ -body approximation and the  $(n - 1)$ -body approximation. If the series in equation 1 is truncated after the second term one is said to have made the electrostatically embedded pairwise additive approximation, and the total energy of the system can be written as

$$E_{\text{EE-PA}} = \sum_{i>j} E_{ij} - (N - 2) \sum_i E_i \quad (5)$$

where  $N$  is the number of particles in the system, and  $E_i$  and  $E_{ij}$  have the same meaning as above. If no embedding charges are used the subscript on the left side of equation 5 can be changed to PA and one is said to have made the pairwise approximation.<sup>21</sup> If one also considers the three-body terms, the electrostatically embedded three-body energy can be written as

$$E_{\text{EE-3B}} = \sum_{i>j>k} E_{ijk} - (N - 3) \sum_{i>j} E_{ij} + \frac{(N - 3)(N - 2)}{2} \sum_i E_i \quad (6)$$

where  $E_i$ ,  $E_{ij}$ , and  $E_{ijk}$  have the same meanings as in equations 2 – 5. As in the case of the EE-PA energy, if no point charges are used one can write the subscript on the left hand side of equation 6 as 3B, and one is said to have made the three-body approximation.

The electronic energy for any correlated level of electronic structure theory can be written as

$$E_X = E_{\text{HF}} + \Delta E_{\text{corr},X} \quad (7)$$

where  $E_X$  is the electronic energy of correlated method X (X = MP2, CCSD, CCSD(T), etc.),  $E_{\text{HF}}$  is the Hartree–Fock energy of the system, and  $\Delta E_{\text{corr},X}$  is the correlation energy for method X. Since the  $n^{\text{th}}$  term of the many body expansion of equation 1 is simply a linear combination of energies for the 1- to  $n$ -body clusters,  $V_n$  can be rewritten, using equation 7, for any correlated level of theory as

$$V_n = V_{n,\text{HF}} + \Delta V_{n,\text{corr}} \quad (8)$$

As a consequence of equation 8 the total energy of the system can be written as

$$V = (V_{1,\text{HF}} + V_{2,\text{HF}} + V_{3,\text{HF}} + \cdots + V_{N,\text{HF}}) + (\Delta V_{1,\text{corr}} + \Delta V_{2,\text{corr}} + \Delta V_{3,\text{corr}} + \cdots + \Delta V_{N,\text{corr}}) \quad (9)$$

where the first term in parentheses is the many-body expansion of the Hartree–Fock energy, and the second term in parentheses is the many-body expansion of the correlation energy.

The Hartree–Fock energy contains electrostatic and inductive terms that can be long-range (e.g., the electrostatic interaction between dipolar monomers dies off only as  $R^{-3}$ , where  $R$  is the distance between monomers, and the charge–induced dipole interaction dies as  $R^{-4}$ ), whereas the terms due *entirely* to correlation energy are known to decay as  $R^{-6}$ , which is a medium-ranged interaction. However, the inclusion of correlation energy does change the dipole moment of a monomer, leading to changes in the long-range dipole-dipole interactions, which, as mentioned above, die off as  $R^{-3}$ . But, if the change in dipole moment between the correlated level of electronic structure theory and Hartree–Fock theory is small, then this effect is also “small,” despite being long-range in nature. Höffinger et al.<sup>22</sup> have tested the accuracy of a series of electronic structure methods, including both wave function methods and density functional methods

with a variety of basis sets, for predicting dipole moments for a test set of small molecules ( $\text{N}_2$ ,  $\text{CO}_2$ ,  $\text{SO}_2$ ,  $\text{HF}$ ,  $\text{HCl}$ ,  $\text{H}_2\text{O}$ ,  $\text{NH}_3$ ,  $\text{PH}_3$ ). They found that, on average, the dipole moments of these molecules change by 5 – 11% when one goes from Hartree–Fock theory to MP2, with the largest mean percent changes using the aug-cc-pVDZ and aug-cc-pVTZ basis sets (11% and 10%, respectively). If one looks only at the water molecule, it was found that for any of the five basis sets tested the percent change in the dipole moment, as one goes from Hartree–Fock to MP2 theory, is never more than 5%. Moreover, they found that as one considers more highly correlated levels of electronic structure theory (e.g., MP4SDQ or QCISD) the mean percent error changes by at most only an additional 2%. Therefore, since most of the change in the dipole moment due to correlation energy is present at the MP2 level of theory, the use of MP2 theory to test methods such as those presented here should provide good insights into the performance of other post-Hartree–Fock methods.

Given the differing nature of the Hartree–Fock and correlation energies, it is not unreasonable to treat their many-body expansions differently by considering more terms in the many-body expansion of the Hartree–Fock energy (in order to better account for the long-range electrostatic and inductive terms) than in the expansion of the correlation energy. Fortunately, since Hartree–Fock theory formally scales as  $N^4$ , where  $N$  is again the number of atoms, it is less computationally demanding to consider larger clusters with Hartree–Fock theory than it is for the correlated methods. In practice, one can use Hartree–Fock theory for the calculation of moderately sized systems (up to a few hundred atoms) with a large basis set at an affordable cost.<sup>23</sup> Therefore, we propose to calculate the complete Hartree–Fock energy for the system (i.e., to carry out the many-body

expansion to  $N^{\text{th}}$  order) and calculate only the correlation energy of the system by using a truncated many-body series. If the many-body expansion is used for the correlation energy without the presence of point charges, the result is denoted MB-CE, where MB is PA if the first two terms in the series are kept, and MB is 3B if the first three terms are kept. If the electrostatically embedded many-body expansion is used for the correlation energy, the results are denoted EE-MB-CE, where MB has the same subcases as above.

### 3. Computational Methods

In order to test the accuracy of the new methods described in Section 2, a series of water clusters ranging in size from 5 to 20 water molecules were taken from the Cambridge Cluster Database.<sup>24</sup> These clusters are the global-minimum-energy structures at the HF/6-31G(d,p) level of theory.<sup>25</sup> Since water clusters are known to exhibit large many-body effects<sup>26,27</sup> this set of clusters should provide a good test of the different methods described here. Eight different many-body methods were applied these systems: PA, 3B, EE-PA, EE-3B, PA-CE, 3B-CE, EE-PA-CE, and EE-3B-CE, where each method has been described in the previous section. The full cluster calculations were performed using the *Gaussian 03*<sup>28</sup> software package. All many-body calculations were carried out with the MBPAC 07<sup>29</sup> software package, which uses *Gaussian 03* to perform all electronic structure calculations. For the EE-MB and EE-MB-CE calculations, point charges of  $-0.778$  and  $0.389$  were used for the oxygen atoms and hydrogen atoms, respectively, as in Ref. 20.

### 4. Results and Discussion

Table 1 shows the mean errors and mean percent errors for the eight different many-body methods as compared to the full MP2 calculations. One of the striking results

of Table 1 is the improvement of the EE-PA method as compared to the PA approximation. The inclusion of point charges changes the mean unsigned error from 15.95 kcal/mol to only 0.80 kcal/mol, which is consistent with previous results.<sup>20</sup> Considering that the binding energies range from 33.52 to 196.02 kcal/mol with an average of 105.46 kcal/mol, a mean unsigned error of 0.80 kcal/mol is an impressive result. One can also see that inclusion of the three-body terms improves the energy by only 0.5 – 0.6 kcal/mol. Since the trimer calculations are the most numerous and most expensive calculations in the many-body expansion the EE-PA may be sufficient for many applications.

The second significant result is the large reduction in error between the PA and PA-CE methods. The mean unsigned error is reduced by a factor of 72 by including the full Hartree–Fock energy! The other methods show a much smaller change in their mean errors when the full Hartree–Fock energy is included, with changes of a factor of 3, 8, and 1.5 for the 3B, EE-PA, and EE-3B methods, respectively. The EE-PA-CE method is the most accurate method with a mean unsigned error of only 0.10 kcal/mol, which represents a mean percent unsigned error of only 0.09% of the net interaction energies. As mentioned previously, the ability to consistently calculate total energies that are within less than 0.1% of the full cluster calculation by only having to consider two-body terms represents a significant savings in the total computational time needed to carry out the calculation. For example, if one assumes that the time needed to calculate the energy of a water monomer, dimer, and trimer at the MP2/aug'-cc-pVTZ level of theory is 30 seconds, 2 minutes, and 5 minutes respectively, then the total time needed to calculate the MP2 correlation energy for a cluster of 20 water molecules is 6.5 hours with the

EE-PA-CE method as compared to 4.2 days to calculate the correlation energy with the EE-3B-CE method. The fact that good energetics can be determined using only a two-body approximation is consistent with other fragment-based methods that have been proposed in the literature.<sup>13, 18, 30</sup>

The result that the EE-3B-CE method has a larger average error than the EE-PA-CE method is somewhat surprising, as one might expect the EE-3B-CE method to give a smaller average error than the EE-PA-CE method since it contains more terms in the many-body expansion. If the errors for each individual structure are examined, one finds that of the sixteen structures considered, eight have a larger error at the EE-3B-CE level than at the EE-PA-CE level. Similarly, four of the sixteen structures have larger errors at the 3B-CE than at the PA-CE level, despite the 3B-CE method having a lower average error. Since the EE-MB-CE methods use embedding and the MB-CE methods do not, the use of embedding cannot be the sole source of this error. In fact, if the many-body expansion given in equation 1 is not truncated, the result is exact and is independent of whether or not embedding is used. We can also exclude double counting as a source of error. At the EE-PA-CE level we account accurately for the two-body terms and approximate the higher order many-body terms, then, at the EE-3B-CE level, we subtract the two-body terms with approximate three-body effects, and we treat the three-body terms (and lower-order terms) exactly and estimate the higher-order terms. This continues at higher orders so that the method is free of any double counting. The source of the errors for both the EE-3B-CE and 3B-CE methods would be an interesting topic for further study, however, for now we proceed with examining the EE-PA-CE method.

If one is interested in trying to further decrease the cost of the calculation one could consider implementing a cutoff to reduce the number of pairs that one must calculate. Since correlation energy is typically short-ranged as compared to the Hartree–Fock energy (see the Background section for more discussion of this point), it may be reasonable to assume that for a dimer with a large inter-molecular distance the contribution of the correlation energy to the two-body term might be very small. In order to determine a reasonable cutoff for water, we next examine the magnitude of the two-body contribution to the energy ( $V_2$ ) as a function of distance. In order to examine this, the global-minimum-energy structure of the water dimer was optimized at the CCSD(T)/aug-cc-pVTZ level of theory and was separated along the vector connecting the centers of mass from 4 to 10 Å in intervals of 1 Å. For each of these seven geometries, as well as at the optimized geometry, an MP2/aug'-cc-pVTZ single-point calculation was carried out, and the  $V_2$  term was calculated. Figure 1 shows the plot of  $V_2$  as a function of this separation, for both the HF/aug'-cc-pVTZ energy and the MP2/aug'-cc-pVTZ correlation energy. Figure 1 clearly shows that the  $V_2$  term for the MP2 correlation energy goes to zero much more rapidly than for the Hartree–Fock energy. In fact, by  $\sim 4.5$  Å the  $V_2$  term for the MP2 correlation energy is approximately zero. While this plot does not take into account any type of rotational averaging, and any one orientation cannot be fully representative, it does suggest that considering a cutoff between 5 and 6 Å might be a reasonable starting point.

Table 2 compares the mean errors obtained using cutoffs of 5 and 6 Å for the  $V_2$  term to the mean errors if no cutoffs are used. If a cutoff of 5 Å is used, at least one pair can be disregarded in 13 of the 16 structures studied, however, one must go up to clusters

of 10 water molecules before a significant number (5 or more) of pairs can be ignored. If a cutoff of 6 Å is used one must consider structures containing 11 water molecules or more before a significant number of pairs can be ignored; however, by the time 20 water molecules are present only 56% of the total number of pairs need to be considered. Additionally, one can see that for the EE-PA-CE method the use of a cutoff of 6 Å is able to reproduce the accuracy obtained when no cut-off is used. Based on the timing arguments presented in the second paragraph of this section it would take only approximately 3.6 hours to calculate the EE-PA-CE correlation energy for a cluster of 20 water molecules with a cutoff of 6 Å.

Figure 2 shows a plot of the two-body correlation energy versus the center-of-mass separation for each of the 190 dimers in the 20-mer. Based on this figure it is clear that the two-body correlation energy for this cluster goes to zero at approximately 6 Å, as opposed to the gas-phase water dimer, which becomes negligible at about 4.5 Å. As mentioned previously, consideration of the gas-phase water dimer was used as a guide to approximate where an appropriate cutoff might be; however, it is clear from Figure 2 that while such a rudimentary example can give some insight into the choice of cutoff one may still need to consider several cutoffs, or carry out a full analysis on a large cluster to obtain the best possible cutoff for the system of interest.

Figure 3 shows a plot of the electrostatically embedded two-body correlation energy versus the center-of-mass separation of the same cluster as in Figure 2. A comparison of Figures 2 and 3 shows that the addition of the embedded charges does not change the range of the interaction — both decrease to zero at approximately 6 Å; however, it is evident that the addition of the point charges does introduce some of the

higher-order many-body terms as the magnitude of the two-body term in the range of  $\sim 3 - 6 \text{ \AA}$  are noticeably different between the two plots, particularly in the region of the plot from 3.9 to 4.2  $\text{\AA}$ . If one considers the cluster shown to be a series of cubes stacked on top of each other, all the dimers in the region from 3.9 to 4.2  $\text{\AA}$  are pairs of water molecules that form diagonals across the faces of these cubes. The different orientations and distances between the water molecules give rise three clusters of points in Figure 2. However, all of these dimers are a part of larger tetrameric clusters, making up the faces of the cubes, which have cooperative hydrogen bonding around the cycle, leading to large many-body effects. Addition of the embedded point charges helps to mimic these effects, making the series of dimers converge to a smoother envelope of points in Figure 3.

The last issue we would like to discuss is the efficiency with which gradients can be calculated since gradients are necessary for carrying out geometry optimizations or molecular dynamics calculations. In previous work<sup>20</sup> we discussed the linearity of the original EE-MB method and the ease with which gradients could be implemented. For example, the gradient of the EE-PA energy can be written as

$$\nabla E_{\text{EE-PA}} = \sum_{i < j}^N \nabla E_{ij} - (N - 2) \sum_i^N \nabla E_i. \quad (10)$$

where an analytic gradient for the EE-PA method is available for any method that has analytic gradients for the monomer and dimer calculations, provided the program allows for fractionally charged point charges as pseudo-nuclei. For the EE-MB-CE methods presented here, the total energy can still be written as a linear combination of energies. For example, the total energy at the EE-PA-CE level, using correlated method X, is given by

$$\begin{aligned}
E_{\text{EE-PA-CE}} &= E_{\text{HF}} + \Delta E_{\text{EE-PA,corr}} \\
&= E_{\text{HF}} + \sum_{i<j}^N \Delta E_{ij,\text{corr}} - (N-2) \sum_i^N \Delta E_{i,\text{corr}} \\
&= E_{\text{HF}} + \sum_{i<j}^N (E_{ij,X} - E_{ij,\text{HF}}) - (N-2) \sum_i^N (E_{i,X} - E_{i,\text{HF}})
\end{aligned} \tag{11}$$

Because the gradient is a linear operator, the gradient of the energy given in equation 11 can be written as

$$\nabla E_{\text{EE-PA-CE}} = \nabla E_{\text{HF}} + \sum_{i<j}^N (\nabla E_{ij,X} - \nabla E_{ij,\text{HF}}) - (N-2) \sum_i^N (\nabla E_{i,X} - \nabla E_{i,\text{HF}}) \tag{12}$$

and it is again true that the method will have analytic gradients so long as the electronic structure methods used have analytic gradients. A key point here is that the values of our point charges are fixed. Since the magnitude of these charges are fixed, the embedding charges are like fractionally charged nuclei with no basis functions and so the only extension of the usual gradient routines that is required is to allow fractionally charged “nuclei”. This is an important advantage of the present method over some alternative many-body schemes.

## 5. Summary and Conclusions

We have presented here an extension of the electrostatically embedded many-body method that calculates the full Hartree–Fock energy of the system and applies the EE-MB method only to the correlation energy of the system. We have found that for MP2 correlation energies the inclusion of the full Hartree–Fock energy reduces the error of the standard pairwise additive approximation by a factor of 72, and the error of the EE-PA method by a factor of eight. We have also found that one can accurately calculate the energies of clusters containing up to 20 water molecules to within 0.09%, on average, of the net interaction energy by considering only the two-body terms for the correlation

energy. Since the calculations needed to evaluate the three-body terms in many-body expansion are both the most numerous and most expensive, this constitutes a substantial savings in time.

Finally, we have demonstrated that the use of a cutoff for evaluation of the two-body term can reduce the number of dimer terms that need to be calculated substantially, without having a large impact on the accuracy of the EE-PA-CE method. Using a cutoff of 6 Å we are able to reproduce the total energy of a cluster of 20 water molecules to within 0.1% of the net interaction energy by calculating the correlation energy of only 106 of the 190 possible pairs of water molecules. In the future we hope to extend this work both to larger systems and to other levels of correlated electronic structure theory.

**Acknowledgment** The authors thank Ryan Olson for stimulating discussions. This work was supported in part by the National Science Foundation under grant nos. CHE03-49122 and ITR-0428774.

**Supporting Information Available:**

Tables listing the binding energies at the MP2, PA, 3B, EE-PA, EE-3B, PA-CE, 3B-CE, EE-PA-CE, and EE-3B-CE levels of theory for each of the clusters considered in this work. This information is available free of charge via the internet at <http://pubs.acs.org>

## References

- (1) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (2) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35.
- (3) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
- (4) Raghavachari, K.; Anderson, J. B. *J. Phys. Chem.* **1996**, *100*, 12960.
- (5) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935.
- (6) Li, W.; Li, S. *J. Chem. Phys.* **2004**, *121*, 6649.
- (7) Li, S.; Shen, J.; Li, W.; Jiang, Y. *J. Chem. Phys.* **2006**, *125*, 074109.
- (8) Saebø, S.; Pulay, P. *J. Chem. Phys.* **1987**, *86*, 914.
- (9) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330.
- (10) Lee, M. S.; Maslen, P. E.; Head Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592.
- (11) Schutz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661.
- (12) Casassa, S.; Zicovich-Wilson, C. M.; Pisani, C. *Theor. Chem. Acc.* **2005**, *116*, 726.
- (13) Federov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.
- (14) Federov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
- (15) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
- (16) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.
- (17) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2002**, *119*, 3599.
- (18) Li, S.; Li, W.; Fang, T. *J. Amer. Chem. Soc.* **2005**, *127*, 7215.
- (19) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112.
- (20) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory. Comp.* **2007**, *3*, 46.
- (21) Elrod, M. J.; Saykally, R. J. *Chem. Rev.* **1994**, *94*, 1975.
- (22) Höffinger, S.; Wendland, M. *Int. J. Quantum Chem.* **2002**, *86*, 199.
- (23) Polly, R.; Werner, H.-J.; Manby, F.; Knowles, P. *Mol. Phys.* **2004**, *102*, 2311.

- (24) Wales, D. J.; Doye, J. P. K.; Dullweber, A., et al. Cambridge Cluster Database  
<http://www-wales.ch.cam.ac.uk/CCD.html> (accessed March 8, 2006).
- (25) Maheshwary, A.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. *J. Phys. Chem. A* **2001**, *105*, 10525.
- (26) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523.
- (27) Dahlke, E. E.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 10595.
- (28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Robb, G. E. S. M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; S. Clifford; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; P. Piskorz; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; M. A. Al-Laham; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; P. M. W. Gill; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. **2003**, *Gaussian03-version c01*; Gaussian Inc.: Wallingford, CT, **2004**.
- (29) Dahlke, E. E.; Truhlar, D. G. MBPAC 07; University of Minnesota: Minneapolis, **2007**.
- (30) Federov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.

**Table 1: Comparison of Mean Errors<sup>a</sup> (kcal/mol) and Mean Percent Errors<sup>b</sup> (%) for Different Many-Body Methods, as Compared to Full Cluster Calculations<sup>c</sup>**

	MSE	MUE	RMSE	MPSE	MPUE	RMPSE
PA	15.95	15.95	17.55	15.40	15.40	15.47
3B	0.55	0.56	0.71	0.69	0.69	0.88
EE-PA	0.80	0.80	0.84	0.82	0.82	0.83
EE-3B	-0.34	0.35	0.51	-0.24	0.26	0.33
PA-CE	0.22	0.22	0.24	0.22	0.22	0.23
3B-CE	-0.05	0.17	0.24	0.01	0.15	0.18
EE-PA-CE	-0.10	0.10	0.11	-0.09	0.09	0.10
EE-3B-CE	-0.23	0.23	0.34	-0.16	0.17	0.21

<sup>a</sup> MSE, MUE, RMSE denote mean signed, mean unsigned, and root mean squared errors respectively

<sup>b</sup> MPSE, MPUE, RMPSE denote mean percent signed, mean percent unsigned, and root mean percent squared errors respectively.

<sup>c</sup> All calculations correspond to the MP2/aug'-cc-pVTZ level of theory

**Table 2: Comparison of Mean Errors<sup>a</sup> (kcal/mol) for PA-CE and EE-PA-CE methods with  $R_{\text{cut}} = 5 \text{ \AA}$ ,  $R_{\text{cut}} = 6 \text{ \AA}$ , and  $R_{\text{cut}} = \infty$**

	$R_{\text{cut}} = 5 \text{ \AA}$			$R_{\text{cut}} = 6 \text{ \AA}$			$R_{\text{cut}} = \infty$		
	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE
PA-CE	0.58	0.58	0.71	0.38	0.38	0.46	0.22	0.22	0.23
EE-PA-CE	0.31	0.33	0.46	0.02	0.07	0.09	-0.09	0.09	0.10

<sup>a</sup> MSE, MUE, RMSE denote mean signed, mean unsigned, and root mean squared errors respectively

### Figure Captions

Figure 1. Two-body energy versus center-of-mass separation for the water dimer. The solid line is the HF/aug'-cc-pVTZ result, the dashed line is the result for the MP2 correlation energy also using the aug'-cc-pVTZ basis set.

Figure 2. Two-body energy versus center-of-mass separation for the 190 dimers of  $(\text{H}_2\text{O})_{20}$  structure shown. Each circle represents the MP2/aug'-cc-pVTZ two-body correlation energy for one of the dimers. In Figs. 2 and 3, many of the circles cannot be seen because they are obscured by other circles.

Figure 3. Plot of the electrostatically embedded two-body energy versus center-of-mass separation for the 190 dimers of  $(\text{H}_2\text{O})_{20}$  structure shown. Each circle represents the electrostatically embedded MP2/aug'-cc-pVTZ two-body correlation energy for one of the dimers.

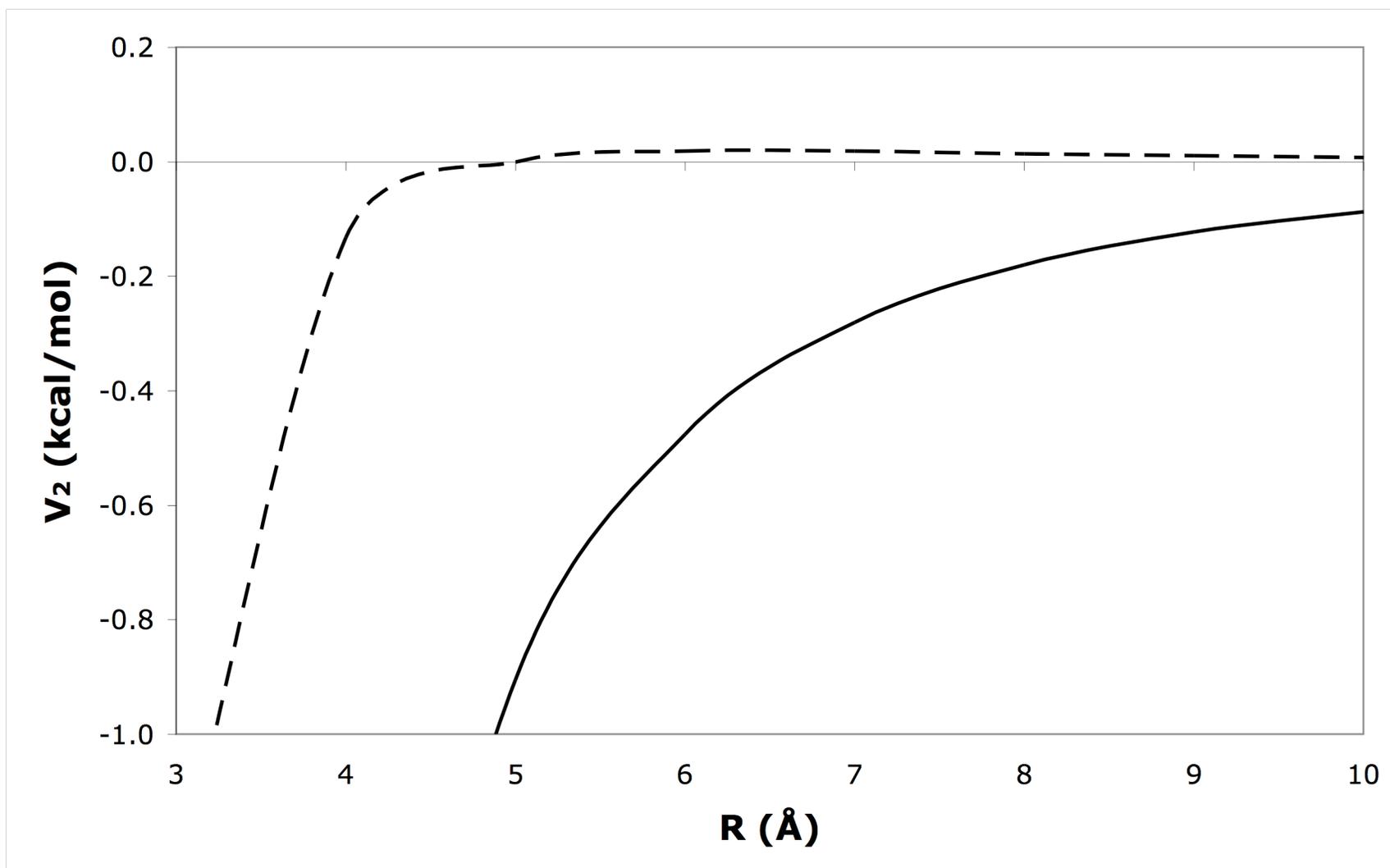


Fig. 1

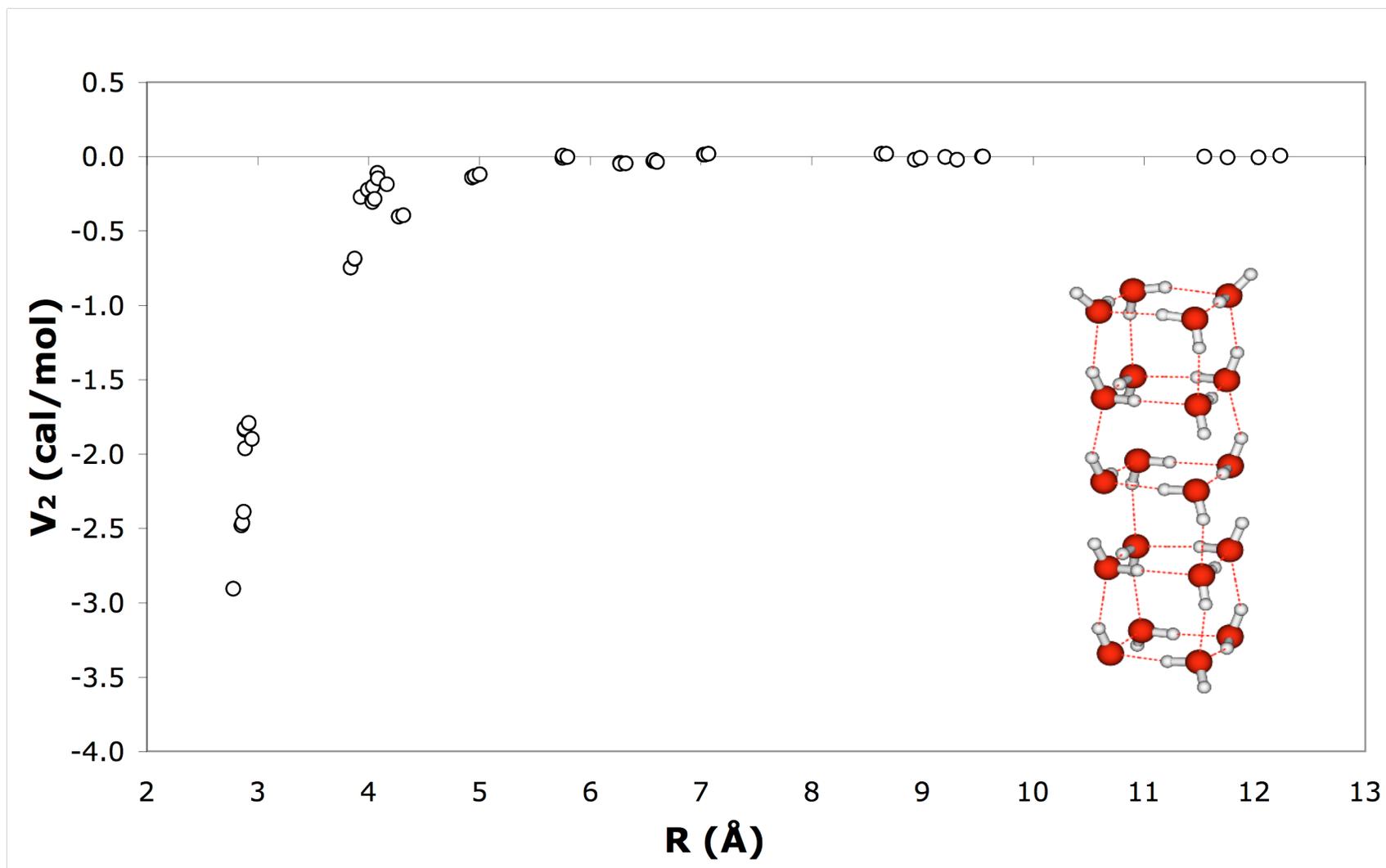


Fig. 2

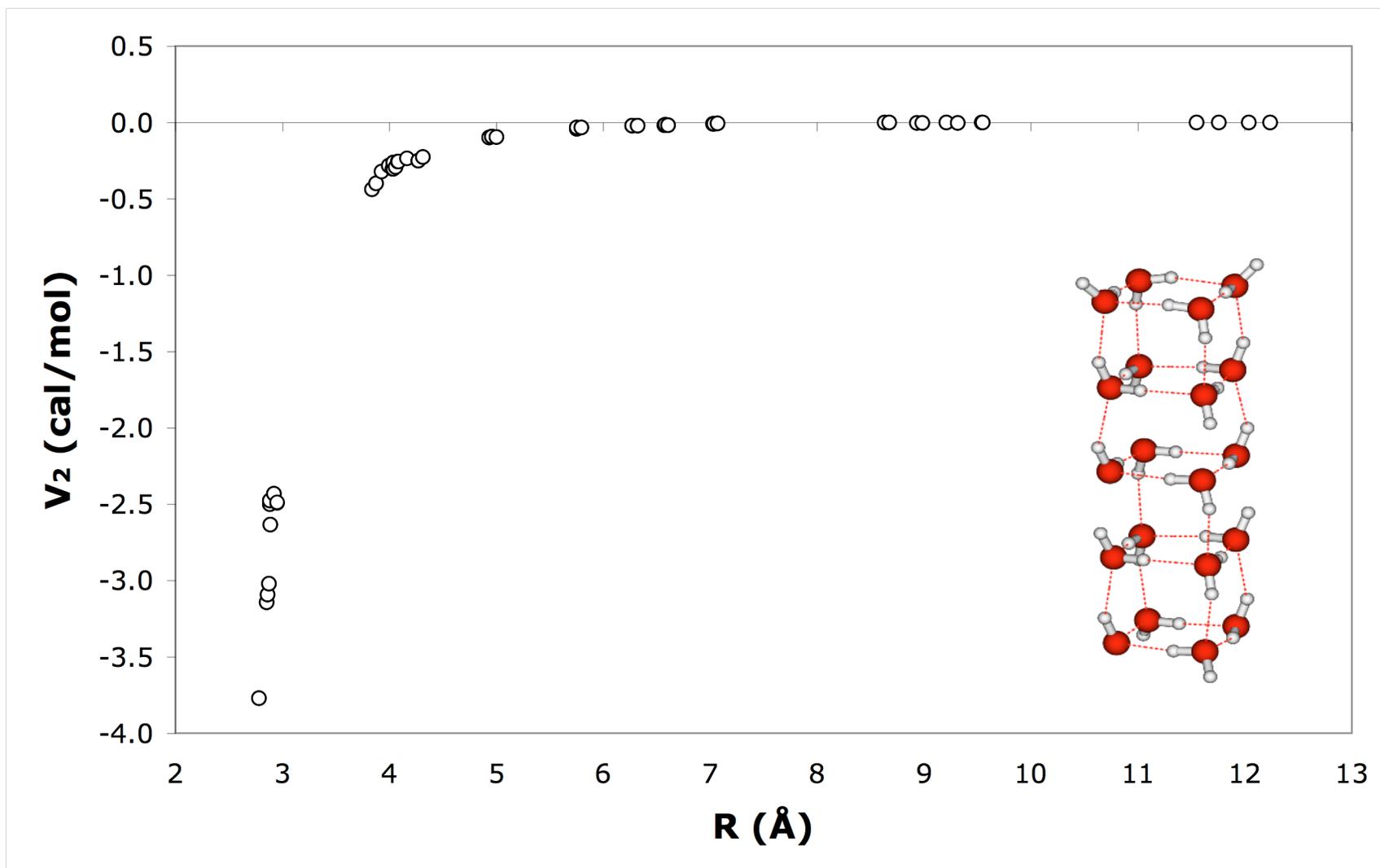


Fig. 3