JCTC Journal of Chemical Theory and Computation

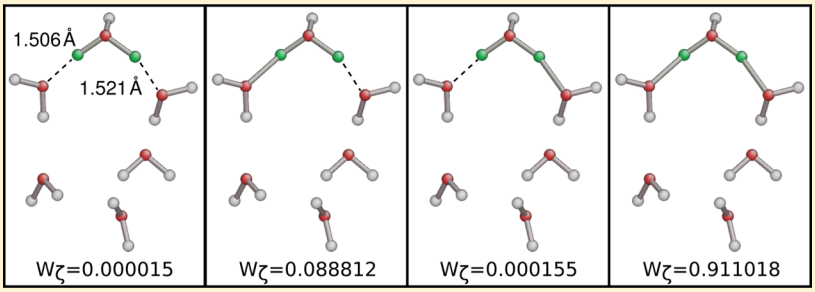
Article

pubs.acs.org/JCTC

Reactive Many-Body Expansion for a Protonated Water Cluster

Peter Pinski*,[†,‡] and Gábor Csányi[†]

[†]Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom



ABSTRACT: We generalize the standard many-body expansion technique that is used to approximate the total energy of a molecular system to enable the treatment of chemical reactions by quantum chemical techniques. By considering all possible assignments of atoms to monomer units of the many-body expansion and associating suitable weights with each, we construct a potential energy surface that is a smooth function of the nuclear positions. We derive expressions for this reactive many-body expansion energy and describe an algorithm for its evaluation, which scales polynomially with system size, and therefore will make the method feasible for future condensed phase simulations. We demonstrate the accuracy and smoothness of the resulting potential energy surface on a molecular dynamics trajectory of the protonated water hexamer, using the Hartree−Fock method for the many-body term and Møller−Plesset theory for the low order terms of the many-body expansion.

1. INTRODUCTION

Correlated quantum chemistry methods such as coupled cluster theory represent the state of the art in solving the Schödinger equation to obtain the total energy of a group of atoms with the electrons in their ground state.[1] However, using a level of theory that is sufficiently accurate in describing common systems of interest leads to very unfavorable scaling of the computational cost with system size (e.g., $O(N^7)$ for CCSD-(T)) and coupled with the necessity of using large basis sets, limits the direct applicability of these methods to isolated small molecules.

Nevertheless, for molecular liquids and solids, breaking up the total energy into a combination of monomer, dimer, trimer, etc. contributions offers a way to compute the total energy very accurately. In these so-called many-body expansion (MBE) methods, the total energy of M distinct monomers is written using the following expansion:

$$E^{(1)} = \sum_{i}^{M} E_i$$

$$E^{(2)} = \sum_{i<j}^{M} E_{ij} - E_i - E_j$$

$$\begin{aligned} E^{(3)} &= \sum_{i<j<k}^{M} E_{ijk} - (E_{ij} - E_i - E_j) - (E_{ik} - E_i - E_k) \\ &\quad - (E_{jk} - E_j - E_k) - E_i - E_j - E_k \\ &= \sum_{i<j<k}^{M} E_{ijk} - E_{ij} - E_{ik} - E_{jk} + E_i + E_j + E_k \\ &\vdots \\ E^{\text{total}} &= \sum_{n}^{M} E^{(n)} \end{aligned} \tag{1}$$

where E^{total} is the energy of the entire system, the one-body energy $E^{(1)}$ is a sum of just the energies of the isolated monomers, $E^{(2)}$ corresponds to a sum of pair interaction energies, and $E^{(3)}$ to the three-body interactions, etc. Equation 1 is an identity, and the number of terms scales factorially with M, so it is not yet useful. The key insight is that high order terms can be computed very accurately by approximate means, often just using electrostatics without treating individual electrons explicitly. Indicating such an approximate model by a tilde, the MBE energy of order n becomes

$$E^{\text{MBE}(n)} = \tilde{E}^{\text{total}} - \tilde{E}^{(1)} - \ldots - \tilde{E}^{(n)} + E^{(1)} + \ldots + E^{(n)} \tag{2}$$

Received: June 10, 2013
Published: December 16, 2013

   dx.doi.org/10.1021/ct400488x | J. Chem. Theory Comput. 2014, 10, 68−75

The MBE energy thus defined converges very rapidly for molecular systems, and useful models are obtained by treating only the monomers and dimers with the quantum chemical approach. Careful examination demonstrated that it is sometimes advantageous to use different approximate methods at each level of the MBE, depending on the desired accuracy.[2]

Many-body expansion approaches are familiar from a multitude of fragmentation methods (reviewed by Gordon et al.[3]). On the other hand, recent developments include the point charge embedding method by Dahlke and Truhlar,[4−6] which was applied to clusters containing water, ammonia, and sulfuric acid[7] or zinc complexes.[8] Another embedding scheme, which includes approximations for Coulomb interaction and exchange-repulsion, was used for solid carbon dioxide, hydrogen fluoride, and ice.[9] Liquid water was simulated in a Monte Carlo scheme on the basis of a many-body expansion.[10] The systematic treatment of many-body interactions by the MBE method has led to a number of force fields developments for water.[11−13] The MBE approach has also been used recently to introduce a two-body CCSD(T)-correction to density functional theory by means of machine learning, recovering a large fraction of the coupled-cluster energy of a wide range of phases of water.[14]

The requirement of assigning atoms to distinct monomers makes the MBE approach, as presented, unsuitable for the treatment of chemical reactions. Using the example of an $H_{13}O_6^+$ cluster, we develop a new scheme, which allows the molecular dynamics simulation of reactive species in solution by dynamically reassigning atoms to monomers and averaging the energies of the resulting partitionings using suitable weighting factors. While the concept of multiple partitionings of atoms to molecules is familiar in the context of multistate empirical valence bond formalism,[15] to the best of our knowledge our scheme is the first that reproduces the quantum mechanical MBE potential energy surface in a smooth and controllable way.

We stress that the present approach is distinct from the more common idea of fitting a potential energy surface—or force field—to quantum chemistry data.[14,16−19] In the latter case, the potential energy corresponding to a given nuclear configuration is written as a linear combination of terms, each of which is a function of the distance (measured in configuration space) between the given configuration and those in a database. In contrast, the present approach defines the potential energy as a weighted sum over *partitionings*, whose terms correspond to identical nuclear positions, but differ in how the atoms are grouped into monomers for the purposes of the MBE expansion.

In this work, the total energy is computed as a sum of the Hartree−Fock energy for the full system and a many-body expansion of the Møller−Plesset correlation energy (MP2). Compared to an expansion of the full MP2 energy, the resulting typical error for a given MBE order is an order of magnitude smaller.[5] This work aims to illustrate the principle of our method and we expect that the method is applicable in conjunction with other MBE models.

2. THEORY

2.1. Assignment of Monomers. In principle, the monomer units of the MBE need not map onto covalently bonded groups of atoms. However, once a mapping between atoms and monomers is given, the MBE energy of a given order is well-defined by eq 2. Naturally, the rate of convergence of the MBE will strongly depend on the mapping. In the following, we generalize the concept of the "monomer", taking it to mean a group of atoms which is treated as a single unit for the purposes of the MBE in eq 1. Excess protons in water form strong bonds with surrounding molecules, which nevertheless fluctuate as diffusion takes place.[20] It is to be expected, that if, for any given configuration of nuclear positions, a sufficient number of water molecules are assigned, together with the proton, to a single monomer, then the resulting MBE energies will be as accurate as for pure water. Monomers need to be assigned dynamically so that they are large enough to maintain accuracy but sufficiently small to be treated efficiently.

It is conceptually useful to introduce a fixed cutoff picture. In this case, if the distance between two atoms is smaller than a predefined cutoff they are assigned to the same monomer. Conversely, if all possible paths between two atoms contain at least one interatomic distance larger than the cutoff, the atoms belong to different monomers. Assuming that all monomers have a closed-shell electronic structure, their charges can be computed as $n_H - 2n_O$, with the numbers of hydrogen and oxygen atoms given by n_H and n_O, respectively. Clearly, the resulting MBE potential energy surface exhibits discontinuities wherever monomers are reassigned as the positions of the nuclei are varied.

2.2. Smooth Assignment of Monomers. Define Z to be the set of all possible monomer partitionings of a system of N atoms, irrespective of the nuclear positions. Each partitioning ζ is assigned a weight $W_\zeta(\mathbf{r}^N)$ that depends on the nuclear coordinates $\mathbf{r}^N$, and the total energy of this *reactive many-body expansion*, RMBE, is defined as

$$E = \sum_{\zeta \in Z} W_\zeta(\mathbf{r}^N) E_\zeta^{\mathrm{MBE}}(\mathbf{r}^N) \tag{3}$$

and the weights sum to unity,

$$\sum_{\zeta \in Z} W_\zeta(\mathbf{r}^N) = 1 \tag{4}$$

The RMBE potential energy surface will be smooth provided that both the MBE energy and the corresponding weighting function are smooth for all possible partitionings. In the following, we distinguish between connectivities and partitionings:

- A *connectivity* is a graph whose vertices are the N nuclei, defined by connecting or disconnecting each of the $N(N-1)/2$ pairs of atoms.
- A *partitioning* is a specific assignment of atoms to monomers.

For a system containing N atoms, a complete set Λ of $2^{N(N-1)/2}$ different possible connectivities can be defined by either connecting or disconnecting each of the $N(N-1)/2$ pairs of atoms. A connectivity λ is mapped onto a partitioning ζ by assigning each subset of connected atoms to the same monomer. Note that there is a many-to-one correspondence between connectivities and partitionings, as illustrated in Figure 1: several connectivities are mapped onto the same partitioning. Likewise, for any possible partitioning of a system there is at least one corresponding connectivity. We define as $C(\zeta)$ the set of all connectivities $\lambda \in C(\zeta)$ that map onto the partitioning ζ.

To assign a weight for a given connectivity we consider functions $f(d_{\alpha\beta})$ of interatomic distances $d_{\alpha\beta}$ with $f(d) = 1$ for $d < a$ and $f(d) = 0$ for $d > b$, where a and b are lower and upper cutoffs. If a pair of atoms corresponds to vertices with an edge

dx.doi.org/10.1021/ct400488x | J. Chem. Theory Comput. 2014, 10, 68−75

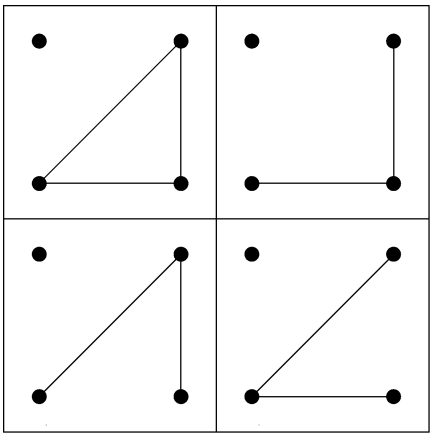


Figure 1. Four of the $2^{4(4-1)/2} = 64$ connectivities for a system of four vertices. They all correspond to the same partitioning: The three connected dots are always the same "monomer", and the top left dot is another monomer of its own.

between them, the weight associated with that pair is $v^{\lambda}_{\alpha\beta} = f(d_{\alpha\beta})$; otherwise the weight is $v^{\lambda}_{\alpha\beta} = 1 - f(d_{\alpha\beta})$. The total weight of a connectivity λ is defined as the product of all atom pair weights:

$$W_{\lambda}(\mathbf{r}^N) \equiv \prod_{\alpha<\beta}^{N} v^{\lambda}_{\alpha\beta} \tag{5}$$

This allows us to define the weight of a partitioning as the sum of weights of all connectivities that map onto that partitioning:

$$W_{\zeta}(\mathbf{r}^N) = \sum_{\lambda\in C(\zeta)} W_{\lambda}(\mathbf{r}^N) \tag{6}$$

For each connectivity in which a particular bond is connected, there is one other connectivity in which the same bond is disconnected, but all other bonds are left unchanged. Given a configuration $\mathbf{r}^N$, the sum over all connectivity weights is therefore unity:

$$\sum_{\lambda\in\Lambda} W_{\lambda}(\mathbf{r}^N) = \prod_{\alpha<\beta}^{N} \{f(d_{\alpha\beta}) + 1 - f(d_{\alpha\beta})\} = 1 \tag{7}$$

As every connectivity maps onto one partitioning, eq 4 is satisfied: $\Sigma_{\zeta\in Z} W_{\zeta}(\mathbf{r}^N) = \Sigma_{\lambda\in\Lambda} W_{\lambda}(\mathbf{r}^N) = 1$. We can rewrite eq 3 as the sum over connectivities:

$$E = \sum_{\lambda\in\Lambda} W_{\lambda}(\mathbf{r}^N) E^{\mathrm{MBE}}_{\zeta(\lambda)}(\mathbf{r}^N) \tag{8}$$

Figure 2 illustrates partitionings and their weights for a geometry generated from a molecular dynamics simulation of $H_{13}O_6^+$. (Molecular images generated from AtomEye.[22])

In principle, the number of all possible connectivities for a system is $2^{N(N-1)/2}$, which would make this approach intractable for all but the very smallest systems. Fortunately, most interatomic distances $d_{\alpha\beta}$ will not be in the cutoff range (a,b), so that $f(d_{\alpha\beta})$ is either exactly one or zero. Any connectivity in which atoms α and β are disconnected for $d_{\alpha\beta} < a$, or connected for $d_{\alpha\beta} > b$, has a weight of zero and can be discarded, leaving nonzero weights only for configurations in which interatomic bond distances are in the range (a,b):[21]

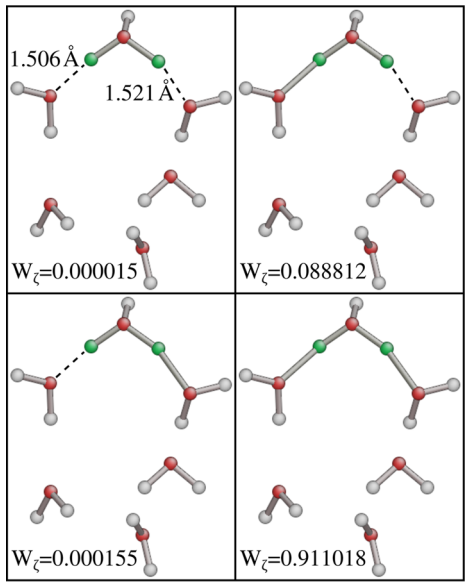


Figure 2. Illustration of partitionings and their associated weights with a = 1.5 Å, b = 1.6 Å, and the weighting function given in eq 19. The nuclear positions in the four panels are exactly the same, and the partitionings shown are the only ones with nonzero weight. The dashed lines indicate bonds whose length is between the limits a and b but are not included in the given partitioning. The four panels represent the four possible ways of the two bonds being included or not.

$$W_{\lambda}(\mathbf{r}^N) = \prod_{\{\alpha<\beta | a\le d_{\alpha\beta}\le b\}}^{N} v^{\lambda}_{\alpha\beta} \tag{9}$$

With an MBE expansion up to n-body interactions, the number of different energies needed is expected to scale approximately as the n-th power of system size. In practice, a better scaling can be achieved, by approximating or truncating the interactions between distant monomers, which is common practice when using the MBE approach.[12−14] If the number of interatomic distances between a and b is m, then the number of different connectivities is 2^m. For sufficiently large systems, m scales linearly with system size, and managing the partitionings would replace individual subsystem energy calculations as the bottleneck. However, dealing with the full connectivities is unnecessary as the next section shows.

2.3. Subsystem-Based Evaluation. Let us consider a partitioning ζ^* of the system defined by the fixed upper cutoff b. If two atoms belong to different monomers in this partitioning, then this is also true for all other partitionings defined by the various connectivities in the cutoff range (a,b). This allows us to break down the system into subsystems, each corresponding to a monomer in ζ^*. In Figure 2, the subsystems are the $H_7O_3^+$ unit and the three remaining water molecules.

First, consider the one-body energy term:

$$E^{(1)} = \sum_{\lambda\in\Lambda} W_{\lambda} \sum_{i\in M(\lambda)} E_i \tag{10}$$

where $M(\lambda)$ is the set of all monomers in the partitioning $\zeta(\lambda)$. If the atoms are enumerated from 1 to N, then the set P of all possible monomers in the system is given by the power set of $\{1...N\}$, that is, the set of all subsets, excluding the empty set: $P = \mathcal{P}(\{1...N\})\backslash\varnothing$. Define the discrete function δ as follows:

dx.doi.org/10.1021/ct400488x | J. Chem. Theory Comput. 2014, 10, 68−75

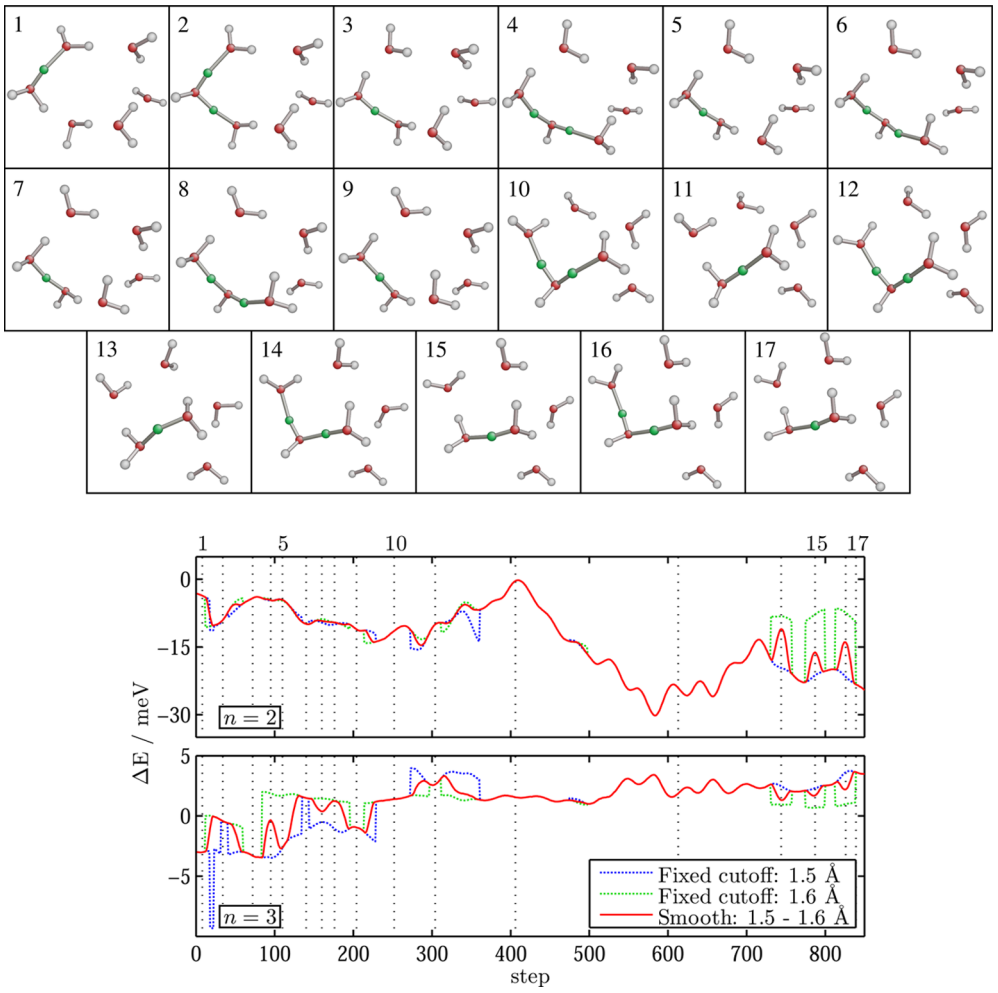


Figure 3. Top: Partitionings occurring along the trajectory with a fixed cutoff of 1.55 Å, doubly coordinated hydrogens are colored green. The graphs show the energy errors for two- and three-body expansions of the MP2 correlation energy. Note the discontinuities in the energy using fixed cutoffs, as opposed to the smooth RMBE curve. The vertical dotted lines mark the positions of the images along the trajectory.

$$\delta(i, M(\lambda)) = \begin{cases} 1 & i \in M(\lambda) \\ 0 & i \notin M(\lambda) \end{cases} \quad \text{where} \quad i \in P \tag{11}$$

which allows us to rewrite eq 10:

$$E^{(1)} = \sum_{\lambda \in \Lambda} \sum_{i \in P} \delta(i, M(\lambda)) W_\lambda E_i = \sum_{i \in P} E_i \sum_{M(\lambda) \ni i} W_\lambda \tag{12}$$

In the last step, the following shorthand notation is used to sum over all connectivities containing the monomer i: $\Sigma_{M(\lambda)\ni i} \equiv \Sigma_{\{\lambda | i \in M(\lambda)\}}$.

The weight of each connectivity is a product of the interatomic weights. From eq 9, it is clear that any nonzero weights can only arise from connections between atoms within the same subsystem, and so, we can define connectivities λ_s, which are restricted to a subsystem s. Then, the connectivity weight becomes a product of weights associated with subsystem connectivities:

$$W_\lambda = \prod_{s \in M(\zeta^*)} w_{\lambda_s} \tag{13}$$

Consider subsystem s, which contains the monomer i, and the set Λ_s of all connectivities within that subsystem. Then,

$$\sum_{M(\lambda) \ni i} W_\lambda = \sum_{M(\lambda_s) \ni i} w_{\lambda_s} \prod_{\{t \in M(\zeta^*) | t \neq s\}} \sum_{\lambda_t \in \Lambda_t} w_{\lambda_t} = \sum_{M(\lambda_s) \ni i} w_{\lambda_s} \tag{14}$$

meaning that the summation over weights W_λ of all connectivities λ, which contain monomer i can be replaced with a summation over weights of subsystem connectivities w_{λ_s}. Using this simplification, with $M(s)$ referring to the set of all distinct monomers in any connectivities of s, eq 12 becomes

$$\begin{aligned} E^{(1)} &= \sum_{s \in M(\zeta^*)} \sum_{i \in M(s)} E_i \sum_{M(\lambda) \ni i} W_\lambda \\ &= \sum_{s \in M(\zeta^*)} \sum_{i \in M(s)} E_i \sum_{M(\lambda_s) \ni i} w_{\lambda_s} \\ &= \sum_{s \in M(\zeta^*)} \sum_{\lambda_s \in \Lambda_s} w_{\lambda_s} \sum_{i \in M(\lambda_s)} E_i \end{aligned} \tag{15}$$

Exploiting the many-to-one relation between connectivities and partitionings, the sum of all weights w_{λ_s} for connectivities that contain identical sets of monomers $M(\lambda_s)$ can be replaced by the single partitioning weight w_{ζ_s}. This allows us to express the previous equation in terms of the sets Z_s of subsystem partitionings:

dx.doi.org/10.1021/ct400488x | J. Chem. Theory Comput. 2014, 10, 68−75

$$E^{(1)} = \sum_{s\in M(\zeta^*)} \sum_{\zeta_s\in Z_s} w_{\zeta_s} \sum_{i\in M(\zeta_s)} E_i \tag{16}$$

This shows that the one-body energy can be computed from a sum over expressions that are derived from energies and interatomic distances within the respective subsystems only. For large systems, the computational effort to compute one-body energies is expected to scale linearly with system size.

Analogously, the expressions for the two- and three-body energies become

$$E^{(2)} = \sum_{s} \sum_{\zeta_s} w_{\zeta_s} \sum_{\{i,j\}\in D(\zeta_s)} \varepsilon_{ij} + \sum_{s<t} \sum_{\zeta_s,\zeta_t} w_{\zeta_s} w_{\zeta_t} \sum_{i\in M(\zeta_s)} \sum_{j\in M(\zeta_t)} \varepsilon_{ij}$$

$$E^{(3)} = \sum_{s} \sum_{\zeta_s} w_{\zeta_s} \sum_{\{i,j,k\}\in T(\zeta_s)} \varepsilon_{ijk} + \sum_{s\neq t} \sum_{\zeta_s,\zeta_t} w_{\zeta_s} w_{\zeta_t} \sum_{i\in M(\zeta_s)} \sum_{\{j,k\}\in D(\zeta_t)} \varepsilon_{ijk} + \sum_{s<t<u} \sum_{\zeta_s,\zeta_t,\zeta_u} w_{\zeta_s} w_{\zeta_t} w_{\zeta_u} \sum_{i\in M(\zeta_s)} \sum_{j\in M(\zeta_t)} \sum_{k\in M(\zeta_u)} \varepsilon_{ijk} \tag{17}$$

where s, t and u range over the monomers in ζ^*, and $M(\zeta)$, $D(\zeta)$, and $T(\zeta)$ refer to the respective sets of distinct monomers, dimers, and trimers in ζ. We used the following definitions:

$$\varepsilon_{ij} = E_{ij} - E_i - E_j$$

$$\varepsilon_{ijk} = E_{ijk} - E_{ij} - E_{ik} - E_{jk} + E_i + E_j + E_k \tag{18}$$

This analysis shows that, without making any approximations, the exponential complexity can be restricted to very small subsystems. As a result, the computational effort for this method with an MBE expansion to order n would scale with the n-th power of the system size, or better.

3. RESULTS

3.1. Methodology. The RMBE method was tested by calculating energies and forces for geometries from a previously computed molecular dynamics (MD) trajectory for a $H_{13}O_6^+$ cluster in ring form. The QUIP code[23] was used for MD with a time step of 0.25 fs at a temperature of 325 K. Energies and gradients for the MD simulation were obtained at each time step from MP2 calculations with a 6-31G* basis set, carried out with Molpro.[24] From this, a continuous trajectory of 850 steps was selected, ensuring that it contained a few proton transfers.

We implemented RMBE in Python, again with Molpro as the quantum chemistry back end. Hartree−Fock calculations were performed for the full cluster, and only the MP2 correlation energies expanded using MBE. Energies and forces were then compared to the MP2 reference results for the full cluster. Both the MBE and the reference calculations employed the aug-cc-pVDZ basis set. The following function was used for interatomic weighting, which we chose to ensure that a sufficient number of derivatives are zero at a and b:

$$f(d) = \begin{cases} 1 & x \le -1 \\ \dfrac{1}{1+\exp\left(x + \frac{1}{1-x} - \frac{1}{1+x}\right)} & |x| < 1 \\ 0 & x \ge 1 \end{cases} \quad \text{with}$$

$$x = \frac{2d - (a+b)}{b-a} \tag{19}$$

3.2. Energies and Forces. Figure 3 compares energy errors of RMBE, with smooth weighting in the range a = 1.5 Å and b = 1.6 Å, with results employing fixed cutoffs at a and b. As expected, using the latter leads to discontinuities, which are comparable in size with the total error. On the other hand, the smooth weighting of RMBE leads to a smooth evolution of the energy. Figure 4 and Table 1 show that in going from the one-body to the three-body expansion of the MP2 correlation energy, the error decreases by about an order of magnitude with each expansion level. Recent ab initio potential energy fits[13,14] with errors on the scale of 1−10 meV have shown significant improvements in the prediction of structural and dynamical properties of water, which suggests that, for molecular materials, MBE expansions truncated at second or third order would be sufficiently accurate for many purposes.

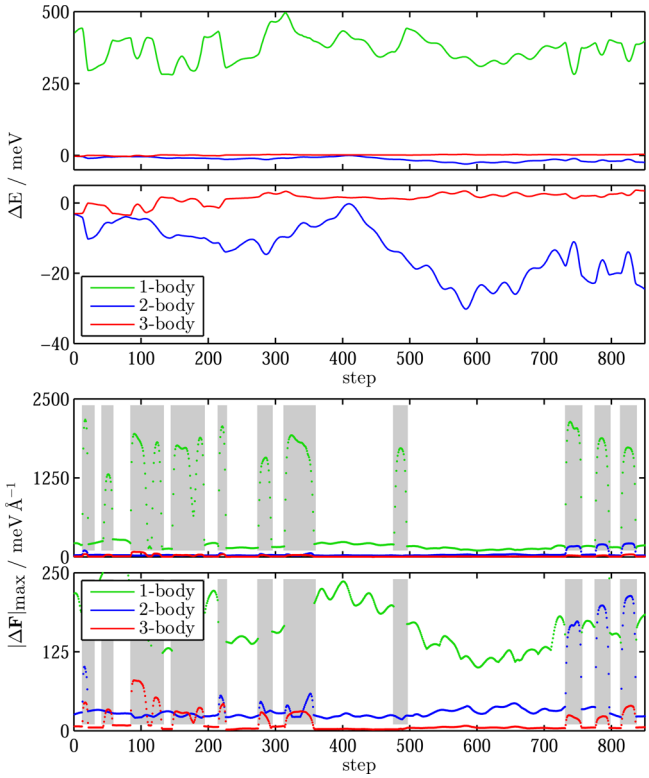


Figure 4. Comparison of energy errors (two top panels) and largest force error magnitudes (two bottom panels) for one- (green), two- (blue), and three-body (red) expansions of the MP2 correlation energy with smooth weighting in the range a = 1.5 Å and b = 1.6 Å. The second and fourth panels are views of the first and third panels, respectively, with expanded vertical axes to highlight the differences between the two- and three-body errors. The gray areas indicate the presence of at least one distance between an H and an O atom in the range (a,b).

Table 1. RMS and Largest Errors in Energies and Individual Atomic Forces for a Weighted Cutoff Range with a = 1.5 Å, b = 1.6 Å

order of MBE	n = 1	n = 2	n = 3
ΔE_{RMS}/meV	372.8	15.3	2.1
ΔE_{max} meV	496.1	30.2	3.7
$\lvert\Delta F\rvert_{RMS}$/meV Å^{-1}	300	21	7
$\lvert\Delta F\rvert_{max}$/meV Å^{-1}	2172	213	80

Figure 4 also shows the magnitudes of the largest force errors on any atom. The errors exhibit distinctive maxima, which typically occur when an interatomic distance is in the (a,b)

dx.doi.org/10.1021/ct400488x | J. Chem. Theory Comput. 2014, 10, 68−75

cutoff range. Their origin can be understood from the force error expression:

$$\begin{aligned}\Delta F &= -(\nabla E^{\mathrm{MBE}} - \nabla E_{\mathrm{exact}}) \\ &= -(\sum_{\zeta\in Z} [E_\zeta^{\mathrm{MBE}} \nabla W_\zeta + W_\zeta \nabla E_\zeta^{\mathrm{MBE}}] - \nabla E_{\mathrm{exact}}) \\ &= -\sum_{\zeta\in Z} [E_\zeta^{\mathrm{MBE}} \nabla W_\zeta + W_\zeta (\nabla E_\zeta^{\mathrm{MBE}} - \nabla E_{\mathrm{exact}})] \end{aligned} \quad (20)$$

When the distance between a pair of atoms is in the (a,b) range, the gradients of the weights W_ζ are nonzero and the force error expression is dominated by the first term, $-\Sigma_\zeta E_\zeta^{\mathrm{MBE}} \nabla W_\zeta$. Since $\Sigma_\zeta \nabla W_\zeta = 0$ (due to the sum rule $\Sigma_\zeta W_\zeta = 1$), the peaks in the error arise from energy differences between partitionings. This energy difference will decrease as the order of the MBE is increased. This error also scales with the size of the weight gradients. A fixed cutoff partitioning of the system can be represented by using step functions for interatomic weighting, and in this case, the weight gradients are infinite and therefore δ-peaks appear in the force error whenever monomers are reassigned. Conversely, a wider cutoff range (a,b) should reduce the magnitudes of weight gradients and therefore also the height of the force error peaks. The second term in the force error in eq 20 is proportional to the force error of the MBE itself.

In order to demonstrate further that there are no discontinuities in the forces, Figure 5 shows the x-, y-, and z-components of the forces on selected atoms computed with a one-body expansion of the correlation energy. Since at this level of the MBE the force errors are almost as large as the total forces, any discontinuity due to the partitioning switches would be easy to see.

The largest errors in the energies and forces that we observed during the course of the trajectory are shown in Figure 6 for different choices of a and b in the range 1.3−1.6 Å. Energies become more accurate with increasing a and b, but since the largest errors in the forces stem from the peaks associated with partitioning switches, they become smaller as the interval width $b-a$ increases. The computational effort increases with the upper cutoff b and the interval width. From our exploration of the small protonated water cluster, good choices for this system are 1.45−1.50 Å for a and 1.60 Å for b.

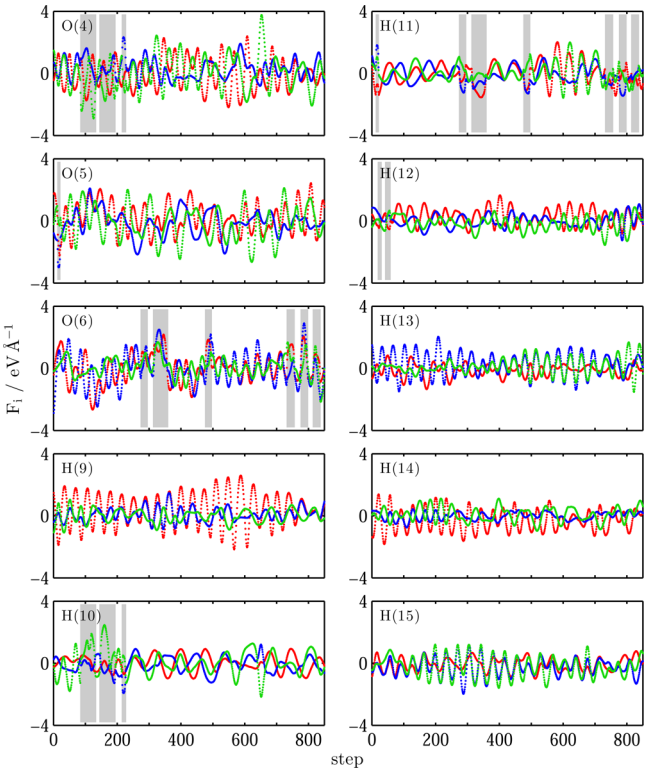


Figure 5. x- (red), y- (blue) and z-components (green) of forces on individual atoms computed with a one-body expansion of the MP2 correlation energy to examine their smoothness. The atoms that comprise the $H_7O_3^+$ monomer visible in the fourth frame of Figure 3 are shown. The gray areas indicate where the atoms have an interatomic distance in the interval a = 1.5 Å, b = 1.6 Å to an atom of a different type.

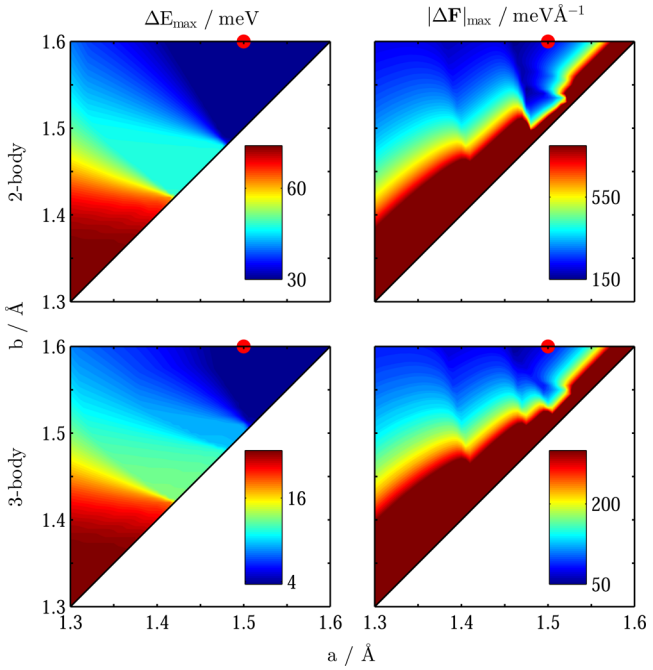


Figure 6. Largest errors encountered along the entire trajectory for expansions of the MP2 correlation energy and the associated forces. Top left: two-body energy errors. Top right: two-body force errors. Bottom left: three-body energy errors. Bottom right: three-body force errors. Red markers: cutoff choice a = 1.5 Å, b = 1.6 Å. The largest force error is expected to diverge when approaching the $a = b$ diagonal in the thermodynamical limit.

3.3. Computational Efficiency of MBE. As discussed in the Introduction, the computational cost of quantum chemistry methods scales with relatively high powers of the system size (theoretically 4 in case of Hartree−Fock and 5 for MP2), whereas the cost of the many body expansion truncated beyond n-body terms scales with the nth power without any distance cutoff, and linearly, if a finite cutoff is employed. These scalings are the same for our reactive many body expansion. To demonstrate that not only the scaling is favorable, but that MBE calculations on condensed systems are beginning to be feasible on very large parallel computers, Table 2 shows the timings we measured on a single CPU core for clusters of one to six water molecules (without an extra proton). Of course, for such small clusters, the MBE calculation takes several times longer than the full MP2 calculation. The idea behind the MBE is that it is to be used for large systems, and so, the table shows

dx.doi.org/10.1021/ct400488x | J. Chem. Theory Comput. 2014, 10, 68−75

Table 2. Timings for Calculating the Potential Energy and Its Gradient on a Single CPU Core Using Hartree−Fock and MP2, the Latter Shown Both on the Full System, and Using an MBE Expansion Truncated beyond Two-Body (MBE2) and Three-Body (MBE3) Terms[a]

N	AVDZ				AVTZ			
H_2O	HF	MP2	MBE2	MBE3	HF	MP2	MBE2	MBE3
1	6	10	10	10	11	17	17	17
2	10	16	35	35	59	110	144	144
3	20	38	76	114	320	636	381	1017
4	45	105	134	284	864	2103	728	3272
5	98	279	207	582	1921	5629	1186	7545
6	208	668	297	1047	5414	16131	1754	14472
32	8×10^4	1×10^6	5×10^3	6×10^4	4×10^6	5×10^7	3×10^4	9×10^5
64	1×10^6	3×10^7	9×10^3	1×10^5	6×10^7	2×10^9	6×10^4	2×10^6
128	1×10^7	7×10^8	2×10^4	2×10^5	1×10^9	5×10^{10}	1×10^5	4×10^6

[a]The timings for 1−6 water molecules are measured directly, while those for 32, 64, and 128 molecules are estimated using the measured scaling laws. A distance cutoff of 5 Å was assumed when estimating the required number of two- and three-body terms. All times are in seconds.

the extrapolated estimates of the time it would take to perform the full quantum chemistry and the MBE calculation on 32, 64, and 128 molecules. The extrapolation was carried out using scaling laws fit to our measure execution times, which for an aug-cc-pVDZ (AVDZ) basis set yielded an exponent of 3.1 for HF and 4.7 for MP2, and for the aug-cc-pVTZ (AVTZ) basis 3.9 for HF and 5.0 for MP2. The gain of MBE truncated beyond two-body terms over the full MP2 calculation is more than 3 orders of magnitude for 32 molecules, and more than 4 orders of magnitude for 128 molecules. It is also clear that the MBE cost is about the same as the baseline HF calculation for 32 molecules, and the latter dominates for larger systems. Therefore, practical MBE dynamics would use a less computationally demanding method than HF for its many body term, either density functional theory with a nonhybrid functional or polarizable electrostatics.

4. DISCUSSION

We have shown how to generalize the many body expansion technique to allow the treatment of chemical reactions. The new reactive many body expansion (RMBE) is a sum of the traditional many-body expansion energy over all possible assignments of atoms to monomer units, each multiplied by a suitable weighting function. The key step for the construction of smooth weight functions was to consider all possible connectivities that lead to a given partitioning. Assigning weights to the connectivities directly ensures that the weight approaches zero smoothly as the distance between any pair of atoms in a partitioning approaches a predetermined cutoff, and also that for any given configuration of the nuclei, the weights over all partitionings sum to unity. We have shown how to evaluate the sum of the energy expression over all possible partitionings using an algorithm that scales polynomially in the order of the many-body expansion.

The practical use of RMBE to study chemical reactions in condensed phase systems with explicitly correlated quantum chemical calculations is rather challenging, even for a second-order many-body expansion. We envisage the RMBE to be used mainly in conjunction with parametrized fits of the potential energy surfaces of the small clusters that appear in the MBE expansion. An increasing number of accurate parametrizations are available,[12,13,25] and progress is being made to increase the automation in generating these parametrizations.[14]

AUTHOR INFORMATION

Corresponding Author

*E-mail: peter.pinski@cantab.net.

Present Address

‡Max Planck Institute for Chemical Energy Conversion, Stiftstr. 34−36, 45470 Mülheim an der Ruhr, Germany

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful to M. J. Gillan and A. P. Bartók for a careful reading of the manuscript.

REFERENCES

(1) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: Mineola, 1996.

(2) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. *J. Chem. Phys.* **2011**, *135*, 224102.

(3) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632−672.

(4) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46−53.

(5) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342−1348.

(6) Speetzen, E. D.; Leverentz, H. R.; Lin, H.; Truhlar, D. G. Electrostatically embedded many-body expansion for large systems. In *Accurate Condensed-Phase Quantum Chemistry*; Manby, F. R., Ed.; CRC Press: Boca Raton, 2010; pp 105−128.

(7) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573−1584.

(8) Kurbanov, E. K.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2012**, *8*, 1−5.

(9) Bygrave, P. J.; Allan, N. L.; Manby, F. R. *J. Chem. Phys.* **2012**, *137*, 164102.

(10) O'Neill, D. P.; Allan, N. L.; Manby, F. R. In *Accurate Condensed-Phase Quantum Chemistry*; Manby, F. R., Ed.; CRC Press: Boca Raton, 2010; Chapter Ab initio Monte Carlo simulations of liquid water, pp 163−194.

(11) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *Science* **2007**, *315*, 1249−1252.

(12) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2011**, *134*, 094509.

(13) Medders, G. R.; Babin, V.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 1103−1114.

(14) Bartók, A. P.; Gillan, M. J.; Manby, F. R.; Csányi, G. *Phys. Rev. B* **2013**, *88*, 054104.

(15) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547−5551.
(16) Ischtwan, J.; Collins, M. A. *J. Chem. Phys.* **1994**, *100*, 8080.
(17) Manzhos, S.; Carrington, T. *J. Chem. Phys.* **2006**, *125*, 194105.
(18) Braams, B.; Bowman, J. *Int. Rev. in Phys. Chem.* **2009**, *28*, 577−606.
(19) Behler, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930.
(20) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. *Nature* **1999**, *397*, 601−604.
(21) The empty product with no terms is understood to equal 1, as usual.
(22) Li, J. *Model. Simul. Mater. Sci. Eng.* **2003**, *11*, 173.
(23) Bartók-Pártay, A.; Cereda, S.; Csányi, G.; Kermode, J.; Solt, I.; Szlachta, W.; Várnai, C.; Winfield, S. QUIP. see http://www.libatoms.org (accessed Nov. 15, 2013).
(24) 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.; Dob-byn, 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.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteins-son, T.; Wang, M.; Wolf, A. *MOLPRO*, version 2010.2, a package of ab initio programs. 2010; see http://www.molpro.net (accessed Nov. 15, 2013).
(25) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 014309.