

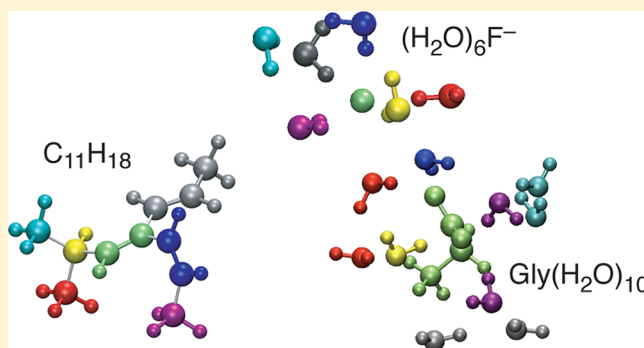
# Many-Body Expansion with Overlapping Fragments: Analysis of Two Approaches

Ryan M. Richard and John M. Herbert\*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

## S Supporting Information

**ABSTRACT:** The traditional many-body expansion—in which a system's energy is expressed in terms of the energies of its constituent monomers, dimers, trimers, etc.—has recently been generalized to the case where the “monomers” (subsystems, or “fragments”) overlap. Two such generalizations have been proposed, and here, we compare the two, both formally and numerically. We conclude that the two approaches are distinct, although in many cases they yield comparable and accurate results when truncated at the level of dimers. However, tests on fluoride–water clusters suggest that the approach that we have previously called the “generalized many-body expansion” (GMBE) [*J. Chem. Phys.* **137**, 064113 (2012)] is more robust, with respect to the choice of fragments, as compared to an alternative “many overlapping body expansion” [*J. Chem. Theory Comput.* **8**, 2669 (2012)]. A more detailed justification for the GMBE is also provided here.



## I. BACKGROUND

Fragment-based quantum chemistry methods,<sup>1,2</sup> in which the energy of a large molecule or cluster is approximated in terms of numerous, smaller calculations, can largely be grouped into two categories depending upon whether the fragments are disjoint (containing no nuclei in common) or not. For a collection of  $N$  disjoint fragments, the ground-state energy can be expressed in terms of the *many-body expansion* (MBE),

$$\begin{aligned} E_0 &= \sum_{I=1}^N E_I^{(1)} + \sum_{I<J=1}^{N C_2} (E_{IJ}^{(1)} - E_I^{(1)} - E_J^{(1)}) + \dots \\ &= \sum_{I=1}^N E_I^{(1)} + \sum_{I<J=1}^{N C_2} \Delta E_{IJ}^{(1)} + \dots \end{aligned} \quad (1.1)$$

Here,  $E_I^{(1)}$  is the energy of the  $I$ th fragment,  $N C_2 = N(N-1)/2$  is the number of unique dimers, and  $E_{IJ}^{(2)}$  is the energy of the dimer created from the union of fragments  $I$  and  $J$ . Equation 1.1 represents a departure from the usual notation for the MBE,<sup>2–6</sup> but will be useful in the forthcoming discussion. The meaning of the superscript “(1)” is discussed later. When the MBE is truncated at the  $n$ -body level, it can be written recursively as

$$E_0^{(n)} = E^{(1)} + E^{(2)} + \dots + E^{(n)} \quad (1.2)$$

where<sup>7</sup>

$$E^{(n)} = \sum_{I<J<\dots<n=1}^{N C_n} E_{I \cup J \cup \dots \cup n}^{(1)} - \sum_{k=1}^{n-1} N_{-k} C_{N-n} E^{(k)} \quad (1.3)$$

This provides an explicit formula for the ground-state energy in terms of the energies of dimers, trimers, etc., up to  $n$ -mers. Of particular interest is the  $n = 2$  case, for which these equations reduce to the expression<sup>5</sup>

$$E_0^{(2)} = \sum_{I<J=1}^{N C_2} E_{IJ}^{(1)} - (N-2) \sum_{I=1}^N E_I^{(1)} \quad (1.4)$$

In contrast, for fragments that share nuclei in common, several groups have recognized that the ground-state energy can be approximated according to the expression<sup>1,8–11</sup>

$$\begin{aligned} E_0 &\approx \sum_{I=1}^N E_I^{(1)} - \sum_{I<J=1}^{N C_2} E_{IJ}^{(1)} + \sum_{I<J<K=1}^{N C_3} E_{I \cap J \cap K}^{(1)} + \dots \\ &\quad + (-1)^{N+1} E_{I \cap J \cap \dots \cap N}^{(1)} \end{aligned} \quad (1.5)$$

Here,  $E_{I \cap J}^{(1)}$  is the energy of the fragment formed from the intersection of fragments  $I$  and  $J$ . Equation 1.5 has been called the *cardinality-guided approximation*,<sup>8–10</sup> owing to its similarity to the inclusion/exclusion principle (IEP) for the cardinality of a set, expressed in terms of the cardinalities of a collection of (possibly nondisjoint) subsets. If the fragments happen to be disjoint, then eq 1.5 reduces to a one-body approximation, namely, eq 1.1 truncated after the first summation. This suggests that it might be possible to generalize eq 1.5, to obtain something that resembles the MBE but which is valid for overlapping fragments.

Received: November 10, 2012

Published: February 7, 2013



In fact, two such generalizations have been proposed recently.<sup>1,12</sup> The first of these was the so-called *many-overlapping-body expansion* (MOBE) proposed by Mayhall and Raghavachari.<sup>12</sup> This expansion, which is in the spirit of eq 1.1, takes the form

$$E_0 = \sum_{I=1}^M c_I E_I^{(1)} + \sum_{I < J=1}^{M C_2} c_I c_J \Delta E_{IJ}^{(1)} + \sum_{I < J < K=1}^{M C_3} c_I c_J c_K \Delta E_{IJK}^{(1)} + \dots \quad (1.6)$$

The coefficients  $c_I$  are discussed below, and  $M$  denotes the total number of fragments that are considered. (We denote this as  $M$  rather than  $N$ , for reasons that will become clear.) The remaining terms in eq 1.6 are defined as

$$\Delta E_{IJ}^{(1)} = E_{I \cup J}^{(1)} - E_I^{(1)} - E_J^{(1)} + E_{I \cap J}^{(1)} \quad (1.7)$$

and

$$\begin{aligned} \Delta E_{IJK}^{(1)} = & E_{I \cup J \cup K}^{(1)} - \Delta E_{IJ}^{(1)} - \Delta E_{IK}^{(1)} - \Delta E_{JK}^{(1)} - E_I^{(1)} - E_J^{(1)} \\ & - E_K^{(1)} + E_{I \cap J}^{(1)} + E_{I \cap K}^{(1)} + E_{J \cap K}^{(1)} - E_{I \cap J \cap K}^{(1)} \end{aligned} \quad (1.8)$$

To comprehend the MOBE, eq 1.6, one must understand how traditional fragmentation ideas are modified in order to use it. First, as with all fragment-based methods, a large system is partitioned into  $N$  subsystems (fragments), capping severed valencies if fragmentation cuts across covalent bonds. Then, according to the MOBE procedure,<sup>12</sup> one next adds  $M - N$  additional fragments to this set, constructed from all possible intersections involving 2, 3, ...,  $N$  fragments out of the original  $N$ . In general, this means that one must consider  $M = 2^N - 1$  fragments in eq 1.6. This also explains the origin of the coefficients in eq 1.6, since more than one intersection might give rise to the same fragment. The coefficient  $c_I$  equals the number of times that fragment  $I$  appears in the summation, with a sign equal to  $(-1)^{\alpha+1}$  where  $\alpha$  is the number of unique fragments whose intersection gave rise to fragment  $I$ . (By convention, or since  $I \cap I = I$ , each of the original  $N$  fragments gets a coefficient  $c_I = 1$ .)

Independently (and unaware of the work in ref 12), the present authors proposed a *generalized many-body expansion* (GMBE),<sup>1</sup> in the spirit of eq 1.5 rather than eq 1.1. The GMBE can be expressed as

$$E_0 \approx \sum_{I=1}^{N C_n} E_I^{(n)} - \sum_{I < J=1}^{N C_n C_2} E_{I \cap J}^{(n)} + \dots + (-1)^{N C_n + 1} E_{I \cap J \cap K \cap \dots \cap N C_n}^{(n)} \quad (1.9)$$

Again, a description of the fragmentation method helps to clarify. We first create  $N$  fragments, then decide the order (one-body, two-body, etc.) at which to truncate the GMBE. Suppose that we truncate at the  $n$ -body level. In that case, we construct all possible unions of  $n$  fragments (“ $n$ -mers”), then apply eq 1.9 to this set of objects. Because writing out the indices for all fragments involved in the unions of an  $n$ -mer is tedious and cumbersome, we employ a notation in which a superscript denotes the cardinality of the  $n$ -mer. For example,  $E_I^{(1)}$  denotes the energy of the  $I$ th monomer ( $I = 1, \dots, N$ ) and  $E_I^{(2)}$  is the energy of the  $I$ th dimer ( $I = 1, \dots, N C_2$ ).

Preliminary tests of the MOBE (eq 1.6) and the GMBE (eq 1.9) suggest that both approaches, when truncated at the two-body level, afford accurate approximations to the supersystem energy.<sup>1,12</sup> From our point of view, each of these methods

constitutes a certain set-theoretical expression for how to approximate the supersystem energy, given a set of overlapping fragments that might be generated in a number of ways.<sup>1,2</sup> Our goal is to compare these two energy expressions at both a formal level and a numerical level.

## II. THEORY

**A. Generalized Many-Body Expansion (GMBE).** In an effort to elucidate the underlying physics of the GMBE, we furnish an argument for how one might write down this expansion a priori. The argument stems from the same line of reasoning that is used to derive the IEP, a proof of which can be found in many textbooks on probability theory. The IEP is simply a counting argument to ensure that no subset of elements is over- or undercounted, so that the cardinality of a (super)set can be expressed in terms of the cardinalities of a collection of nondisjoint subsets. This principle has sometimes been invoked<sup>8</sup> as a justification for eq 1.5, but this is somewhat misleading as it appears to imply an equivalence between cardinality and energy. Below, we clarify the connection between the IEP and the GMBE, expanding upon arguments in ref 1.

To begin, consider a system of particles partitioned into  $N$  nondisjoint sets, with two stipulations: each particle must appear in at least one set and each *pair* of particles must appear in at least one set. The latter stipulation stems from the fact that the Hamiltonian involves two-particle operators, and we aim to use the IEP to recover the exact Hamiltonian after fragmentation. The condition that each pair of particles appear in at least one set is trivially satisfied at levels  $n = 2$  and above; for  $n = 1$ , some of the equalities that appear below are actually approximations.

Should we choose to place the particles into sets based on spatial proximity (i.e., to preserve the local chemical environment), then we would term them *fragments* and denote the  $I$ th one as  $F_I^{(1)}$ , where the superscript has the same meaning as above. In practice, one is usually interested in the set of  $n$ -mers that arise by taking all  $n$ -tuple unions of fragments (of which there are  $m = N C_n$ ), with the  $I$ th  $n$ -tuple denoted  $F_I^{(n)}$ . We denote the set of all particles (the universe) as  $\mathcal{U}$ .

To derive a general energy expression that will hold for any  $m$ , and which does not overcount particle–particle interactions, we use set theory to ensure that fragmentation preserves the supersystem Hamiltonian operator. As shown in Appendix A, the Hamiltonian  $\hat{H}(\mathcal{U})$  that includes all pairwise interactions among the particles in  $\mathcal{U}$  can be expressed as

$$\begin{aligned} \hat{H}(\mathcal{U}) = & \sum_{I=1}^m \hat{H}(F_I^{(n)}) - \sum_{I < J=1}^{m C_2} \hat{H}(F_I^{(n)} \cap F_J^{(n)}) + \dots \\ & + (-1)^{m+1} \hat{H}\left(\bigcap_{I=1}^m F_I^{(n)}\right) \end{aligned} \quad (2.1)$$

We use the notation  $\hat{H}(S)$  to indicate the Hamiltonian corresponding to pairwise interactions of particles in the set  $S \subseteq \mathcal{U}$ .

An energy expression follows from the ground-state expectation value,  $E_0 = \langle \Psi_0 | \hat{H}(\mathcal{U}) | \Psi_0 \rangle$ :

$$E_0 = \sum_{I=1}^m \langle \Psi_0 | \hat{H}(F_I^{(n)}) | \Psi_0 \rangle - \sum_{I < J=1}^{mC_2} \langle \Psi_0 | \hat{H}(F_I^{(n)} \cap F_J^{(n)}) | \Psi_0 \rangle + \dots + (-1)^{m+1} \langle \Psi_0 | \hat{H}(\bigcap_{I=1}^m F_I^{(n)}) | \Psi_0 \rangle \quad (2.2)$$

This can be rewritten as

$$E_0 = \sum_{I=1}^m E_I^{(n)} - \sum_{I < J=1}^{mC_2} E_{I \cap J}^{(n)} + \dots + (-1)^{m+1} E_{I \cap J \cap \dots \cap m}^{(n)} \quad (2.3)$$

where the energies are to be interpreted as the expectation values of the individual fragment Hamiltonians, with respect to the ground-state supersystem wave function. In practical calculations of interest,  $|\Psi_0\rangle$  is not available so we approximate each expectation value in eq 2.3 with a fragment electronic structure calculation. This approximation is essentially an appeal to Kohn's principle of the "locality of electronic matter".<sup>13,14</sup> Note also that the same motivation works if point charges are included as part of each "subset" Hamiltonian,  $\hat{H}(S)$ , as part of a charge-embedding scheme designed to better approximate the electronic structure of each local region of the system.

It is also useful to define *intersection-corrected energies*,

$$\mathcal{E}_I^{(n)} = E_I^{(n)} - \sum_{J=I+1}^m E_{I \cap J}^{(n)} + \sum_{J < K=I+1}^{mC_2} E_{I \cap J \cap K}^{(n)} - \dots + (-1)^{m-I} E_{J < K < \dots < (m-I)}^{(n)} \quad (2.4)$$

In terms of these quantities, eq 2.3 assumes a compact form analogous to the MBE in eq 1.2:

$$E_0 \approx \sum_{I=1}^m \mathcal{E}_I^{(n)} \quad (2.5)$$

We have called this the *generalized many-body expansion* (GMBE), truncated at order  $n$ .

To the best of our knowledge, no one has worked out a closed form for eq 1.2 to arbitrary order, but closed forms for  $n = 2$  and  $n = 3$  are available.<sup>5,7</sup> Here, we are interested in  $n = 2$  exclusively, and we next show that the two-body GMBE is equivalent to the traditional two-body expansion, in the case that the monomers do not intersect. For the  $n = 2$  case of eq 2.5, assuming disjoint monomers, we obtain the following using eq 2.4:

$$E_0 \approx \sum_{I=1}^m E_I^{(2)} - \sum_{I < J=1}^{mC_2} E_{I \cap J}^{(2)} + \dots + \sum_{I < J < \dots < N-1=1}^{mC_{N-1}} (-1)^N E_{I \cap J \cap \dots \cap (N-1)}^{(2)} \quad (2.6)$$

We have simplified this expression somewhat by noting that at most  $N - 1$  dimers may intersect. If the monomers are disjoint, then we can evaluate the intersections in terms of the monomers by realizing that the same monomer must appear in all dimers or else the intersection is an empty set. Taking all unions of monomers that do not result in empty sets, and making use of the identity

$$\sum_{\alpha=0}^n (-1)^\alpha {}_nC_\alpha = 0 \quad (2.7)$$

eq 2.6 reduces to eq 1.4, the traditional two-body energy formula expressed in closed form.

Finally, let us highlight two important properties of the GMBE. First, it is easy to verify that when  $n = N$ , this expansion affords the exact ground-state energy,  $E_0$ . (See eq A1 in Appendix A and also ref 1.) Second, when using the GMBE, it is not necessary to include any  $n$ -mer that is a subset of another  $n$ -mer. The latter fact was suggested by example in ref 1. More generally, if  $A \subseteq B$ , then one can show that

$$E_0 = \sum_{I=1}^{m-1} E_I^{(n)} - \sum_{I < J=1}^{m-1C_2} E_{I \cap J}^{(n)} + \dots + (-1)^{m-1} \sum_{I < J < \dots < \gamma=1}^{m-1C_\gamma} E_{I \cap J \cap \dots \cap \gamma}^{(n)} + \dots + (-1)^m \sum_{I < J < \dots < (m-1)=1}^{m-1C_{m-1}} E_{I \cap J \cap \dots \cap (m-1)}^{(n)} \quad (2.8)$$

This equation says that one obtains the same energy if  $A$  is not considered at all.

**B. Many-Overlapping-Body Expansion (MOBE).** Owing to the difficulty in treating intersections as fragments and keeping them distinct, it is difficult to analyze the MOBE in an analogous manner to the GMBE. However, some properties of the MOBE can be deduced from a simple example that demonstrates the utility of this approach. Consider a system composed of three fragments labeled 1, 2, and 3. According to the prescription of Mayhall and Raghavachari,<sup>12</sup> the one-body energy for this system is given by

$$E_{\text{MOBE}}^{(1)} = \sum_{I=1}^3 E_I^{(1)} - E_{1 \cap 2}^{(1)} - E_{1 \cap 3}^{(1)} - E_{2 \cap 3}^{(1)} + E_{1 \cap 2 \cap 3}^{(1)} \quad (2.9)$$

Intersections appearing in this expression become fragments when evaluating dimers and higher order  $n$ -mers.<sup>12</sup> Referring to these fragments (including intersections) collectively as "monomers", the total number of monomers is  $2^N - 1$  and the number of dimers is  $2^{N-1}C_2$ , although not all of these need be unique.

For our three-body example, the seven terms in eq 2.9 lead to  ${}_\gamma C_2 = 21$  corrections of the form  $\Delta E_{IJ}^{(1)}$  (eq 1.7). Naively, then, the MOBE in eq 1.6, when truncated at the two-body level, consists of  $7 + 4 \times 21 = 91$  separate electronic structure calculations but, in fact, many of these are redundant. First, note that if  $A \supset B$  then  $\Delta E_{AB}^{(1)} = 0$  according to eq 1.7. The identities

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (2.10a)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (2.10b)$$

can then be used to reduce these 91 separate terms down to just 13. A completely general two-body MOBE energy expression for a three-fragment system is

$$E_{\text{MOBE}}^{(2)} = E_{1\cup 2}^{(1)} + E_{1\cup 3}^{(1)} + E_{2\cup 3}^{(1)} - E_{(1\cup 2)\cap(1\cup 3)}^{(1)} \\ - E_{(1\cup 2)\cap(2\cup 3)}^{(1)} - E_{(1\cup 3)\cap(2\cup 3)}^{(1)} + E_{(1\cap 2)\cup(1\cap 3)}^{(1)} \\ + E_{(1\cap 2)\cup(2\cap 3)}^{(1)} + E_{(1\cap 3)\cup(2\cap 3)}^{(1)} - E_{1\cap 2}^{(1)} - E_{1\cap 3}^{(1)} \\ - E_{2\cap 3}^{(1)} + E_{1\cap 2\cap 3}^{(1)} \quad (2.11)$$

Equation 2.11 demonstrates that, when using the MOBE, one *does* need to consider  $n$ -mers that are subsets of other  $n$ -mers. To see this, consider the seventh, eighth, and ninth terms in eq 2.11, which represent the dimers formed as unions of monomer intersections. Each of the resulting dimers is a proper subset of the sets represented by the first three terms in eq 2.11. These terms do not cancel out as they would in the GMBE.

This lack of cancellation has an important consequence, namely, that given a set of  $N$  monomers, an  $N$ -body approximation to the MOBE does *not* afford the exact ground-state energy, as it would in the case of the traditional MBE or the GMBE. One can understand this because the MOBE potentially generates  $2^N - 1$  monomers, so that the exact energy is recovered only at order  $2^N - 1$ . If the monomers are disjoint, this issue goes away and the MOBE reduces to the traditional many-body expansion, eq 1.1.

**C. Comparison of the GMBE and the MOBE.** A two-body truncation of the GMBE, for the system of three monomers considered in Section IIB, affords an energy expression

$$E_{\text{GMBE}}^{(2)} = E_{1\cup 2}^{(1)} + E_{1\cup 3}^{(1)} + E_{2\cup 3}^{(1)} - E_{(1\cup 2)\cap(1\cup 3)}^{(1)} \\ - E_{(1\cup 2)\cap(2\cup 3)}^{(1)} - E_{(1\cup 3)\cap(2\cup 3)}^{(1)} \\ + E_{(1\cup 2)\cap(1\cup 3)\cap(2\cup 3)}^{(1)} \quad (2.12)$$

Note that the final “triple intersection” term does not appear in  $E_{\text{MOBE}}^{(2)}$ , eq 2.11, but the latter does include some terms that are absent from  $E_{\text{GMBE}}^{(2)}$ . This demonstrates conclusively that the MOBE and the GMBE are not equivalent; nevertheless, for many of the systems considered in Section III, we find that  $E_{\text{GMBE}}^{(2)} \approx E_{\text{MOBE}}^{(2)}$ . This suggests that

$$E_{(1\cup 2)\cap(1\cup 3)\cap(2\cup 3)}^{(1)} \approx E_{(1\cap 2)\cup(1\cap 3)}^{(1)} + E_{(1\cap 2)\cup(2\cap 3)}^{(1)} \\ + E_{(1\cap 3)\cup(2\cap 3)}^{(1)} - E_{1\cap 2}^{(1)} - E_{1\cap 3}^{(1)} - E_{2\cap 3}^{(1)} \\ + E_{1\cap 2\cap 3}^{(1)} \quad (2.13)$$

This suggests that perhaps higher-order intersections such as  $(1 \cup 2) \cap (1 \cup 3) \cap (2 \cup 3)$  can be approximated in terms of lower-order intersections, which might be exploited in order to reduce the number of intersections that appear in the GMBE. Results in section III, however, reveal that this would have to be done with care, as we are able to find fragmentation schemes for which  $E_{\text{GMBE}}^{(2)}$  is a significantly better approximation to the supersystem energy than is  $E_{\text{MOBE}}^{(2)}$ .

By way of comparison, it is also worth noting that, for  $N$  monomers, the maximum number of  $n$ -mers generated by the GMBE is  $m = {}_N C_n$  (though some may not be unique), and thus, the final energy expression consists of at most  $2^m - 1$  terms. (In practical applications, many of these may be redundant; the redundant terms are not generated by our fragmentation code.)

In contrast, the MOBE generates  $2^N - 1$  monomers and thus the maximum number of  $n$ -mers is  $m = 2^{N-1} C_n$ . The final MOBE energy expression is a linear combination of all of these  $n$ -mers, along with all  $(n - 1)$ -mers,  $(n - 2)$ -mers, etc., on down to monomers. For  $n > 1$  and  $N > 1$ , the GMBE will have fewer terms to evaluate, as compared to the MOBE. For example, in the case considered above ( $N = 3$  and  $n = 2$ ),  $E_{\text{GMBE}}^{(2)}$  contains 7 terms as compared to 13 terms for  $E_{\text{MOBE}}^{(2)}$ .

As a final point of comparison, we consider a system composed of seven groups (indestructible units).<sup>1</sup> Let us assign  $N = 3$  fragments as follows:

$$F_1^{(1)} = \{1, 2, 3, 4\} \quad (2.14a)$$

$$F_2^{(1)} = \{1, 3, 5, 7\} \quad (2.14b)$$

$$F_3^{(1)} = \{1, 4, 6, 7\} \quad (2.14c)$$

These particular assignments were chosen so that, at the  $n = 1$  level, no intersection is empty and at the  $n = 2$  level, no dimer constitutes the entire system. Given these fragments, the two-body MOBE is

$$E_{\text{MOBE}}^{(2)} = E_{123457}^{(1)} + E_{123467}^{(1)} + E_{134567}^{(1)} - E_{12347}^{(1)} - E_{13457}^{(1)} \\ - E_{13467}^{(1)} + E_{134}^{(1)} + E_{137}^{(1)} + E_{147}^{(1)} - E_{13}^{(1)} - E_{14}^{(1)} \\ - E_{17}^{(1)} + E_1^{(1)} \quad (2.15)$$

where the subscript indices refer to groups.

Relative to the GMBE, the MOBE omits the final term in eq 2.12, involving the intersection of three dimers, or at best approximates this term as in eq 2.13. Since we know that the GMBE precisely accounts for each term in the supersystem Hamiltonian (see Appendix A), let us examine this triple intersection term to determine whether the corresponding terms in the Hamiltonian appear in the MOBE. For the fragments selected in eq 2.14, the triple intersection  $(F_1^{(1)} \cup F_2^{(1)}) \cap (F_1^{(1)} \cup F_3^{(1)}) \cap (F_2^{(1)} \cup F_3^{(1)})$  evaluates to  $\{1, 3, 4, 7\}$ . The simplest type of interaction that could appear in the Hamiltonian is the appearance of a single group, and it is easily verified that each of the seven groups appears only once in the MOBE. A similar exercise shows that all interactions between pairs of groups also appear only once. However, when we repeat this check for interactions among three groups, we see that any interaction that is a net result of the mutual interaction between groups 3, 4, and 7 is counted zero times. Similarly, the interaction among all four groups is not counted either. Since these interactions were present in the initial dimers ( $F_1^{(1)} \cup F_2^{(1)}$ , etc.), they ought to show up in the final energy expression.

### III. NUMERICAL RESULTS

**A. Computational Details.** We next make a numerical comparison of the GMBE and the MOBE, based on the fragmentation scheme suggested in eq 2.14. That is, each system considered here is first divided into seven groups, from which we construct three overlapping fragments as suggested in eq 2.14.

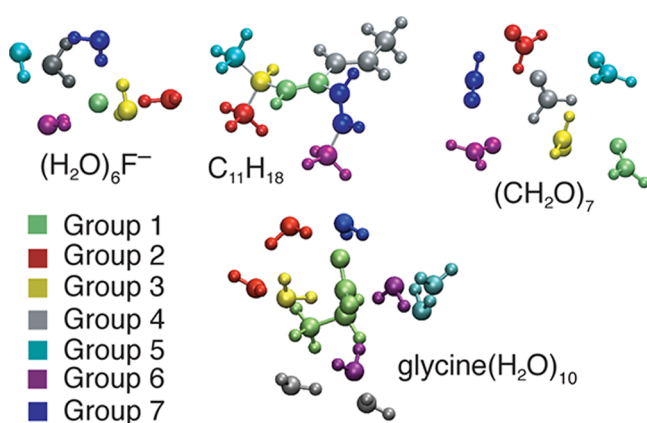
We have written a program, called FRAGMENT, that can fragment a given system according to any one of several possible fragmentation methods. Various embedding and capping methods are also available, as described in ref 1. The FRAGMENT program generates all required electronic structure input files, calls an electronic structure package, and finally



computes the energy according to the GMBE. Electronic structure calculations were performed using Q-Chem, v. 4.0.<sup>15</sup> The MOBE has not been implemented in our code, so a Linux script was written to generate the appropriate Q-Chem input files in this case.

Previously, we introduced a nomenclature for classifying fragmentation schemes.<sup>1</sup> According to this nomenclature, the calculations presented here employ a user-defined fragmentation method at the  $n = 2$  level, a capping method equivalent to that used in the systematic molecular fragmentation method,<sup>16</sup> no embedding method, and a single layer of theory. All energies are computed at the MP2/6-311++G(d,p) level.

Four systems are investigated here, and these are depicted in Figure 1 along with group assignments for each. (Cartesian



**Figure 1.** Group assignments for the four systems considered here.

coordinates for these systems are available in the Supporting Information.) These systems include one polyatomic example,  $C_{11}H_{18}$ , where fragmentation across covalent bonds is necessary, plus three different noncovalent clusters:  $(CH_2O)_7$ ,  $(H_2O)_6F^-$ , and  $Gly \cdot (H_2O)_{10}$ . The  $(H_2O)_6F^-$  cluster, in particular, was selected because previous work by our group suggests that fluoride–water clusters are challenging test cases for monomer-based methods.<sup>1,17–19</sup> Since  $(H_2O)_6F^-$  represents the first test of the MOBE for ionic systems, we have also included for comparison the charge-neutral  $Gly \cdot (H_2O)_{10}$  system, where glycine (Gly) exists in its zwitterionic tautomer,  $^+NH_3CH_2CO_2^-$ .

In the interest of fairness, it should be noted that the MOBE has previously been applied only to macromolecular systems, not clusters, using a chemically-motivated fragmentation scheme based on the covalent-bond topology of the macromolecule.<sup>12</sup> Nevertheless, the MOBE does lend itself to a set-theoretical presentation, and a direct comparison of this approach to the deeply set-theoretical GMBE proves to be informative.

**B. Comparison of GMBE and MOBE.** Unsigned errors for the two-body GMBE and MOBE methods, relative to the corresponding supersystem calculation, are listed in Table 1 for each of the systems examined here. Before analyzing these results, it bears pointing out that these are somewhat contrived fragmentation calculations, not least because each of the three dimers common to both the GMBE and the MOBE are one group shy of being the entire system. Nevertheless, some significant differences between these two approaches will be noted below.

**Table 1.** Unsigned Errors in Two-Body Expansions, Relative to Supersystem MP2/6-311++G(d,p) Calculations

system	group	error/kcal mol <sup>-1</sup>	
	labels	GMBE	MOBE
$(CH_2O)_7$	Figure 1 <sup>a</sup>	0.03	<0.01
$C_{11}H_{18}$	Figure 1 <sup>a</sup>	<0.01	0.23
$Gly \cdot (H_2O)_{10}$	Figure 1 <sup>a</sup>	0.11	0.13
$Gly \cdot (H_2O)_{10}$	$1 \leftrightarrow 3^b$	0.11	0.18
$Gly \cdot (H_2O)_{10}$	$1 \leftrightarrow 4^b$	0.11	0.85
$Gly \cdot (H_2O)_{10}$	$1 \leftrightarrow 7^b$	0.11	0.54
$Gly \cdot (H_2O)_{10}$	$2 \leftrightarrow 7^b$	0.14	0.52
$Gly \cdot (H_2O)_{10}$	$5 \leftrightarrow 7^b$	0.15	0.14
$Gly \cdot (H_2O)_{10}$	$6 \leftrightarrow 7^b$	0.02	0.48
$(H_2O)_6F^-$	Figure 1 <sup>a</sup>	0.12	0.17
$(H_2O)_6F^-$	$1 \leftrightarrow 3^b$	0.13	2.64
$(H_2O)_6F^-$	$1 \leftrightarrow 4^b$	0.13	2.47
$(H_2O)_6F^-$	$1 \leftrightarrow 7^b$	0.12	1.93
$(H_2O)_6F^-$	$2 \leftrightarrow 7^b$	0.47	0.71
$(H_2O)_6F^-$	$5 \leftrightarrow 7^b$	0.22	0.26
$(H_2O)_6F^-$	$6 \leftrightarrow 7^b$	0.17	0.42

<sup>a</sup>Partitioned as shown in Figure 1. <sup>b</sup>Swapping two group labels, relative to Figure 1.

Of the three systems considered here, one might expect  $(H_2O)_6F^-$  clusters to exhibit a relatively large contribution from the “triple intersection” term in eq 2.12, owing to three-body polarization effects. Our analysis will therefore focus on  $(H_2O)_6F^-$ . Using the group assignment for  $(H_2O)_6F^-$  that is depicted in Figure 1, both the two-body GMBE and the two-body MOBE exhibit similar errors of 0.1–0.2 kcal/mol, seemingly suggesting that the three-fragment intersection is not important or is adequately approximated by eq 2.13. Careful examination of Figure 1, however, reveals that the term in question—the intersection of groups 3, 4, and 7—involves three water molecules but not the fluoride ion. For this reason, we also consider several alternative partitions in which we simply relabel which group is defined as group 7. In particular, we swap group label 7 with that of group 1, 2, 5, or 6. (Swapping  $3 \leftrightarrow 7$  or  $4 \leftrightarrow 7$  does not change the original calculation.) Results for these alternative numbering schemes are also listed in Table 1, and we find one case ( $1 \leftrightarrow 7$ ) for which the two-body MOBE exhibits an error of almost 2 kcal/mol while the GMBE is accurate to  $\approx 0.1$  kcal/mol. Further swapping of group labels identified two other cases ( $1 \leftrightarrow 3$  and  $1 \leftrightarrow 4$ ) for which the two-body MOBE exhibits errors  $\geq 2.5$  kcal/mol. These are huge errors for a system where one expects these methods to be nearly exact, in view of the fact that dimers encompass nearly the entire system. In contrast, none of the group labeling schemes results in a two-body GMBE error larger than 0.5 kcal/mol.

Focusing on the case where the group labels differ by  $1 \leftrightarrow 7$  interchange relative to Figure 1, we can understand the large difference between  $E_{MOBE}^{(2)}$  and  $E_{GMBE}^{(2)}$  based on the fact that, for this particular case, the 3–4–7 triple intersection includes the  $F^-$  ion, so polarization effects are expected to be quite important for this term. Table 2 decomposes the 3–4–7 intersection in the GMBE in terms of all of the interactions that go into it. As discussed in Section IIC, the two-body MOBE properly accounts for both single groups and pairwise group interactions appearing in the Hamiltonian. Subtracting these contributions from the interaction energy for the 3–4–7 intersection, we conclude that the three-group interaction

**Table 2.** Energy Decomposition Analysis of  $E_{\text{GMBE}}^{(2)}$  for  $(\text{H}_2\text{O})_6\text{F}^-$ , with Groups 1  $\leftrightarrow$  7 Exchanged Relative to Figure 1

interaction	energy/hartree
$E_{347}^{(1)}$	−252.356265
$E_3^{(1)}$	−76.2938278
$E_4^{(1)}$	−76.2937145
$E_7^{(1)}$	−99.6975948
$\Delta E_{34}^{(1)}$	0.001223
$\Delta E_{37}^{(1)}$	−0.036402
$\Delta E_{47}^{(1)}$	−0.039012
$\Delta E_{347}^{(1)}$	−0.0030632

energy is −0.0030632 hartree or 1.92 kcal/mol, which is essentially identical to the error in the two-body MOBE for this particular partition of  $(\text{H}_2\text{O})_6\text{F}^-$ .

Of the four “ $n \leftrightarrow 7$ ” cases in Table 1, the  $1 \leftrightarrow 7$  interchange is unique in that it results in the fluoride ion appearing in two fragments rather than three. The  $1 \leftrightarrow 3$  and  $1 \leftrightarrow 4$  interchanges, which also lead to large errors in the two-body MOBE, have the fluoride ion appearing in the 3–4–7 intersection but not in all three of these monomers. That the GMBE is largely insensitive to this imbalance in the fragmentation patterns bodes well for its use with automated, “black box” fragmentation schemes based on distance thresholds rather than covalent bond connectivity. Such approaches are presently being pursued in our group.

The MOBE’s sensitivity with respect to the choice of fragments can be seen in the  $\text{Gly} \cdot (\text{H}_2\text{O})_{10}$  results as well, where we performed the same swaps among fragment labels as in the case of  $(\text{H}_2\text{O})_6\text{F}^-$ . For  $\text{Gly} \cdot (\text{H}_2\text{O})_{10}$ , errors in the two-body MOBE range from 0.13–0.85 kcal/mol (see Table 1), which is not as large as the variations seen in the case of  $(\text{H}_2\text{O})_6\text{F}^-$ , perhaps because the water–glycine interaction energy is smaller than that of water–fluoride. Nevertheless, errors for the two-body GMBE are  $\leq 0.15$  kcal/mol for each  $\text{Gly} \cdot (\text{H}_2\text{O})_{10}$  fragmentation scheme.

**C. Timings and Accuracy for a Larger System.** In the calculations reported above, we considered only small systems in order to make an incisive, term-by-term comparison of the MOBE and the GMBE. Ultimately, fragment-based methods are intended for larger systems, so here, we present some results for  $(\text{H}_2\text{O})_{57}$  clusters. The accuracy of various GMBE-based methods, relative to supersystem calculations performed at the Hartree–Fock (HF)/6-31G(d) and B3LYP/6-31+G(d) level are shown in Table 3. As in our previous work,<sup>1</sup> we use the notation “XYZ( $n$ )” to denote an  $n$ -body truncation of the GMBE (where  $n = 1$  or 2 here), in combination with fragmentation method XYZ. The notation “EE-XYZ( $n$ )” denotes the use of a point-charge electrostatic embedding. The two fragmentation methods examined in Table 3 are the *systematic molecular fragmentation* (SMF) method of Deev and Collins,<sup>16</sup> which is based on bond connectivity and for these calculations amounts to one  $\text{H}_2\text{O}$  monomer per fragment, and the *generalized energy-based fragmentation* (GEBF) method of Li et al.,<sup>20</sup> which is based on a 3 Å distance threshold, which in  $(\text{H}_2\text{O})_{57}$  amounts to 3–4 monomers per fragment. Five isomers of  $(\text{H}_2\text{O})_{57}$  were considered in ref 1; structures and additional details can be found there.

In previous calculations on  $(\text{H}_2\text{O})_{57}$ , we were forced to resort to highly restrictive cutoffs for forming dimers of fragments, due to limitations in our fragmentation algorithm at that time.<sup>1</sup>

**Table 3.** Mean Absolute Errors per  $\text{H}_2\text{O}$  Monomer in the Energies of Five Different Isomers of  $(\text{H}_2\text{O})_{57}$ , Relative to Supersystem Calculations Performed at the HF/6-31G(d) or B3LYP/6-31+G(d) Level

method	MAE/kcal mol <sup>−1</sup>	
	HF	B3LYP
SMF(2)	4.87	3.74
EE-SMF(2)	0.45	0.36
GEBF(1)	3.80	2.95
EE-GEBF(1)	0.22	0.40
GEBF(2) <sup>a</sup>	2.69	2.07
EE-GEBF(2) <sup>a</sup>	0.13	0.31
GEBF(2) <sup>b</sup>	0.02	0.02
EE-GEBF(2) <sup>b</sup>	0.04	<0.01

<sup>a</sup>Limited to dimers formed from intersecting fragments, from ref 1.

<sup>b</sup>Full GEBF(2) results.

The GEBF(2) calculations for this system that were reported in ref 1 were restricted to dimers formed from intersecting (essentially, nearest-neighbor) fragments, but in Table 3, we report calculations in which this restriction has been lifted, owing to improvements in our algorithm for computing intersections. Comparison of these “full” GEBF(2) results to previous ones reveals that the restriction to “overlapping dimers” was a serious one, at least in the absence of electrostatic embedding. The discrepancies are less pronounced when electrostatic embedding is employed, presumably because the point charges mimic the polarization effects of distant water molecules, which are the sorts of dimers that were neglected in previous GEBF(2) calculations on  $(\text{H}_2\text{O})_{57}$ . This observation may ultimately prove useful to reduce the number of independent electronic structure calculations that are required at the GEBF(2) level, based on some kind of distance- or intersection-based thresholding criteria. In any case, we observe that GEBF(2) calculations—either with or without electrostatic embedding—reproduce supersystem energies to an accuracy of  $\leq 0.04$  kcal/mol per  $\text{H}_2\text{O}$  monomer.

The next reasonable question is, “at what cost?”. To answer this, timings for the three most accurate methods, averaged over five isomers of  $(\text{H}_2\text{O})_{57}$ , are presented in Table 4.

**Table 4.** CPU Time Required<sup>a</sup> for a Single-Point Energy Calculation on  $(\text{H}_2\text{O})_{57}$  at the HF/6-31G(d) or B3LYP/6-31+G(d) Level

method	CPU time/sec	
	HF	B3LYP
supersystem	397	6414
EE-GEBF(2) <sup>b</sup>	326	1566
GEBF(2) <sup>c</sup>	10 241	89 197
EE-GEBF(2) <sup>c</sup>	10 250	87 969

<sup>a</sup>Calculations performed using Q-Chem<sup>15</sup> running on a single 2.5 GHz Opteron processor. <sup>b</sup>Limited to dimers formed from intersecting fragments. <sup>c</sup>Full GEBF(2) results.

Although the individual electronic structure calculations required for GEBF(2) are small ( $\leq 8$   $\text{H}_2\text{O}$  molecules in this example), the number of such calculations is quite large in the case of a *full* GEBF(2) calculation, and the total wall time required for GEBF(2) actually significantly exceeds that required for a supersystem calculation. Several points are worth noting, however. First, had we chosen a correlated wave

function method scaling as  $O(N^5)$  or worse, or had we employed a significantly larger basis set that might be beset by numerical linear dependencies in a large system, then the supersystem calculation might not even be feasible. Second, the real advantage of GMBE-based methods lies in their “embarrassingly parallel” nature. Assuming, as a rough estimate, that the use of  $p$  processors would reduce the total wall time for GEBF calculations by a factor of  $p$ , then we can guess that only  $p \approx 26$  processors [HF/6-31G(d) level] or  $p \approx 14$  processors [B3LYP/6-31+G(d) level] are required in order to render the fragment-based calculation faster (in terms of wall time) than the supersystem calculation, at least when the latter is run in serial mode. This is a fairly modest level of parallelism.

Finally, the timings in Table 4 highlight the importance of dimer cutoffs, based on overlap criteria, which can avoid the factorial growth in the number of independent electronic structure calculations required for GMBE(2) calculations.<sup>1</sup> EE-GEBF(2) calculations in which all nonintersecting dimers are discarded reproduce supersystem energies to within  $\approx 0.3$  kcal/mol (Table 3), yet are faster than the corresponding supersystem calculations even when no parallelism is exploited at all. For the MP2/6-311++G(d,p) calculations on Gly·(H<sub>2</sub>O)<sub>10</sub> that were reported in Section IIIB, the total CPU time for the MOBE calculations is only about 5% more than that required for the corresponding GMBE calculation, hence, these comments regarding timings should carry over to MOBE calculations as well.

#### IV. CONCLUSIONS

Two generalizations of the many-body expansion to the case of intersecting (overlapping) fragments have been examined here. Of the two, only the method that we call the GMBE counts all interactions appearing in the supersystem Hamiltonian exactly once. This result, which is proved herein, lends theoretical justification to existing fragment-based methods<sup>8,16,20</sup> that are based on either an exact or an approximate one-body energy formula with overlapping fragments, as detailed in our previous work.<sup>1</sup> In contrast, the GMBE is extended here to include two-body terms, yielding very high accuracy. As an alternative to the GMBE, the MOBE energy expression,<sup>12</sup> when truncated at the  $n$ -body level, does not directly evaluate the energy of subsystems constructed from the intersection of  $n$  fragments, whereas such terms are included explicitly in the GMBE.

Numerical calculations on (H<sub>2</sub>O)<sub>6</sub>F<sup>−</sup> and Gly·(H<sub>2</sub>O)<sub>10</sub> clusters demonstrate that both the two-body GMBE and the two-body MOBE exhibit good accuracy for certain fragmentation schemes, but the omission of “triple intersection” terms ( $F_1^{(1)} \cap F_2^{(1)} \cap F_3^{(1)}$ ) in the MOBE sometimes leads to large errors for certain fragmentation patterns, e.g., errors of 1.9–2.6 kcal/mol for (H<sub>2</sub>O)<sub>6</sub>F<sup>−</sup> and 0.48–0.85 kcal/mol for Gly·(H<sub>2</sub>O)<sub>10</sub>. This suggests that greater care is needed when selecting fragments for ionic systems. GMBE results, on the other hand, are much more robust with respect to the choice of fragments. This feature makes the GMBE a promising approach for use with “black box” fragmentation schemes that do not require user oversight and are not based on covalent bond topology, as well as for application to macromolecular systems with ionic functional groups. Applications to (H<sub>2</sub>O)<sub>57</sub> demonstrate that excellent accuracy (<0.1 kcal/mol) is maintained in this larger system. Significant speed-up, without significant loss of accuracy, is possible by placing thresholds on dimer formation, which is an aspect that we plan to explore systematically in future work.

#### ■ APPENDIX A: SET-THEORETICAL EXPRESSION FOR THE SUPERSYSTEM HAMILTONIAN

In this appendix, we derive eq 2.1. Recall that  $m = {}_N C_n$  denotes the number of  $n$ -mers of fragments. If  $m = 1$ , then the entire system (universe,  $\mathcal{U}$ ) is contained in a single  $n$ -mer, so

$$\hat{H}(\mathcal{U}) = \hat{H}(F_1^{(n)}) \quad (\text{A1})$$

where  $\hat{H}(S)$  denotes the Hamiltonian for the set  $S$  of particles.

For the  $m = 2$  case the universe is divided into two sets,  $F_1^{(n)}$  and  $F_2^{(n)}$ :

$$\mathcal{U} = F_1^{(n)} \cup F_2^{(n)} \quad (\text{A2})$$

However, the system can also be partitioned according to

$$\mathcal{U} = (F_1^{(n)} \setminus F_2^{(n)}) \cup (F_2^{(n)} \setminus F_1^{(n)}) \cup (F_1^{(n)} \cap F_2^{(n)}) \quad (\text{A3})$$

These three sets are, respectively, the part of  $F_1^{(n)}$  that is not in  $F_2^{(n)}$ , the part of  $F_2^{(n)}$  that is not in  $F_1^{(n)}$ , and the part that they share in common. Note that these three sets are disjoint, and since  $\hat{H}$  is additively separable, we can therefore write  $\hat{H}(\mathcal{U})$  as a sum of three Hamiltonians, partitioned according to eq A3. We then obtain

$$\begin{aligned} \hat{H}(\mathcal{U}) &= \hat{H}(F_1^{(n)} \setminus F_2^{(n)}) + \hat{H}(F_2^{(n)} \setminus F_1^{(n)}) + \hat{H}(F_1^{(n)} \cap F_2^{(n)}) \\ &\quad + \hat{H}(F_1^{(n)} \cap F_2^{(n)}) - \hat{H}(F_1^{(n)} \cap F_2^{(n)}) \\ &= \hat{H}[(F_1^{(n)} \setminus F_2^{(n)}) \cup (F_1^{(n)} \cap F_2^{(n)})] \\ &\quad + \hat{H}[(F_2^{(n)} \setminus F_1^{(n)}) \cup (F_1^{(n)} \cap F_2^{(n)})] \\ &\quad - \hat{H}(F_1^{(n)} \cap F_2^{(n)}) \\ &= \hat{H}(F_1^{(n)}) + \hat{H}(F_2^{(n)}) - \hat{H}(F_1^{(n)} \cap F_2^{(n)}) \end{aligned} \quad (\text{A4})$$

From here, we proceed by induction. Let us assume that the following is true for all  $m \leq x$ :

$$\begin{aligned} \hat{H}(\mathcal{U}) &= \sum_{l=1}^x \hat{H}(F_l^{(n)}) - \sum_{l < j=1}^{x-1} \hat{H}(F_l^{(n)} \cap F_j^{(n)}) + \dots \\ &\quad + (-1)^{x-1} \hat{H}(\bigcap_{l=1}^x F_l^{(n)}) \end{aligned} \quad (\text{A5})$$

We have already verified this explicitly for the  $m = 1$  and  $m = 2$  cases. To show that eq A5 actually provides a general expression for  $\hat{H}(\mathcal{U})$ , it suffices to show that this expression is valid for the case of  $m = x + 1$   $n$ -mers. In other words, we need to show that

$$\begin{aligned} \hat{H}(\bigcup_{l=1}^{x+1} F_l^{(n)}) &= \sum_{l=1}^{x+1} \hat{H}(F_l^{(n)}) - \sum_{l < j=1}^{x+1} \hat{H}(F_l^{(n)} \cap F_j^{(n)}) + \dots \\ &\quad + (-1)^x \hat{H}(\bigcap_{l=1}^{x+1} F_l^{(n)}) \end{aligned} \quad (\text{A6})$$

We will show that the left side of the eq A6 can be rearranged to yield the right side. First, note that

$$\begin{aligned} \hat{H}(\bigcup_{l=1}^{x+1} F_l^{(n)}) &= \hat{H}(\bigcup_{l=1}^x F_l^{(n)} \cup F_{x+1}^{(n)}) \\ &= \hat{H}(\bigcup_{l=1}^x F_l^{(n)}) + \hat{H}(F_{x+1}^{(n)}) \\ &\quad - \hat{H}(\bigcup_{l=1}^x F_l^{(n)} \cap F_{x+1}^{(n)}) \end{aligned} \quad (\text{A7})$$

which is similar to the manipulations that we performed in the  $m = 2$  case. The final term in this equation can be rewritten as

$$\begin{aligned} \hat{H}(\bigcup_{I=1}^x F_I^{(n)} \cap F_{x+1}^{(n)}) \\ = \sum_{I=1}^x \hat{H}(F_I^{(n)} \cap F_{x+1}^{(n)}) - \sum_{I < J=1}^{x_2} \hat{H}(F_I^{(n)} \cap F_J^{(n)} \cap F_{x+1}^{(n)}) \\ + \dots + (-1)^{x-1} \hat{H}(\bigcap_{I=1}^x F_I^{(n)} \cap F_{x+1}^{(n)}) \end{aligned} \quad (\text{A8})$$

using the inductive hypothesis, eq A5. Substituting this into eq A7 gives

$$\begin{aligned} \hat{H}(\mathcal{U}) = \hat{H}(F_{x+1}^{(n)}) + \sum_{I=1}^x \hat{H}(F_I^{(n)}) - \sum_{I < J=1}^{x_2} \hat{H}(F_I^{(n)} \cap F_J^{(n)}) \\ + \dots + (-1)^{x-1} \hat{H}(\bigcap_{I=1}^x F_I^{(n)}) \\ - \sum_{I=1}^x \hat{H}(F_I^{(n)} \cap F_{x+1}^{(n)}) \\ + \sum_{I < J=1}^{x_2} \hat{H}(F_I^{(n)} \cap F_J^{(n)} \cap F_{x+1}^{(n)}) - \dots \\ + (-1)^x \hat{H}(\bigcap_{I=1}^x F_I^{(n)} \cap F_{x+1}^{(n)}) \end{aligned} \quad (\text{A9})$$

In deriving this expression, we have once again used the inductive hypothesis for  $x$ , this time to expand the first term on the right in eq A7. Combining terms,

$$\begin{aligned} \hat{H}(\mathcal{U}) = \sum_{I=1}^{x+1} \hat{H}(F_I^{(n)}) - \sum_{I < J=1}^{x+1, x_2} \hat{H}(F_I^{(n)} \cap F_J^{(n)}) + \dots \\ + (-1)^{x-1} \hat{H}(\bigcap_{I=1}^x F_I^{(n)}) \\ + \sum_{I < J=1}^{x_2} \hat{H}(F_I^{(n)} \cap F_J^{(n)} \cap F_{x+1}^{(n)}) - \dots \\ + (-1)^x \hat{H}(\bigcap_{I=1}^{x+1} F_I^{(n)}) \end{aligned} \quad (\text{A10})$$

At this point, we need to generalize this procedure for combining terms and collapsing intersections, in order to deal with the terms represented by ellipses in eq A10. Let us denote by  $T_{1,y}$  a term represented by the first ellipses with an index  $y < x$ :

$$T_{1,y} = (-1)^{y-1} \sum_{I < \dots < y=1}^{x_2} \hat{H}(\bigcap_{I=1}^y F_I^{(n)}) \quad (\text{A11})$$

Within the second set of ellipses there will be a term (call it  $T_{2,z}$ ), indexed by  $z = y - 1$ , of the form

$$T_{2,z} = (-1)^z \sum_{I < \dots < z=1}^{x_2} \hat{H}(\bigcap_{I=1}^z F_I^{(n)} \cap F_{x+1}^{(n)}) \quad (\text{A12})$$

The sum of these two terms is

$$\begin{aligned} T_{1,y} + T_{2,z} = (-1)^{y-1} \left[ \sum_{I < \dots < y=1}^{x_2} \hat{H}(\bigcap_{I=1}^y F_I^{(n)}) \right. \\ \left. + \sum_{I < \dots < y-1=1}^{x_2} \hat{H}(\bigcap_{I=1}^{y-1} F_I^{(n)} \cap F_{x+1}^{(n)}) \right] \end{aligned} \quad (\text{A13})$$

The first term on the right generates all Hamiltonians involving  $y$  fragments, in which none of the  $y$  fragments are  $F_{x+1}^{(n)}$ . The second term will produce all Hamiltonians involving the intersection of  $y$  fragments, and  $F_{x+1}^{(n)}$  is included. Together, these two terms are the sum over all Hamiltonians resulting from the intersection of  $y$  fragments, if there are  $x + 1$  fragments to choose from. Thus, the two terms can be combined to afford

$$T_{1,y} + T_{2,z} = (-1)^{y-1} \sum_{I < \dots < y=1}^{x+1, x_2} \hat{H}(\bigcap_{I=1}^y F_I^{(n)}) \quad (\text{A14})$$

This result demonstrates that the  $y$ th term in the first set of ellipses (eq A10) will combine with the  $z$ th term in the second set of ellipses, to afford a term of the form appearing in eq A14. In deriving eq A10, we have already combined the  $y = 1$  term with the  $z = 0$  term and the  $y = 2$  term with the  $z = 1$  term, as these appeared explicitly in eq A9. The remaining terms can be combined, as outlined above, to afford

$$\begin{aligned} \hat{H}(\mathcal{U}) = \sum_{I=1}^{x+1} \hat{H}(F_I^{(n)}) - \sum_{I < J=1}^{x+1, x_2} \hat{H}(F_I^{(n)} \cap F_J^{(n)}) + \dots \\ + (-1)^x \hat{H}(\bigcap_{I=1}^{x+1} F_I^{(n)}) \end{aligned} \quad (\text{A15})$$

This is the desired result for the case  $m = x + 1$  case. By induction, the form of  $\hat{H}(\mathcal{U})$  given in eqs A5 and A15 must be valid for all  $m$ , that is, for an arbitrary number of  $n$ -mers of fragments.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Cartesian coordinates for the systems considered here. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [herbert@chemistry.ohio-state.edu](mailto:herbert@chemistry.ohio-state.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center under projects PAS0291 and PAA0003. J.M.H. is an Arthur P. Sloan Foundation Fellow and a Camille Dreyfus Teacher-Scholar.

## ■ REFERENCES

- (1) Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137* (064113), 1–17.



- (2) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2011**, *112*, 632–672.
- (3) Kulkarni, A. D.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2004**, *121*, 5043–5050.
- (4) Cui, J.; Liu, H.; Jordan, K. D. *J. Phys. Chem. B* **2006**, *110*, 18872–18878.
- (5) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1–6.
- (6) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7579–7590.
- (7) Kaplan, I. G.; Santamaria, R.; Novaro, O. *Mol. Phys.* **1995**, *84*, 105–114.
- (8) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125* (104109), 1–10.
- (9) Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2008**, *129* (234101), 1–7.
- (10) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405–2418.
- (11) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.
- (12) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *8*, 2669–2675.
- (13) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168–3171.
- (14) Prodan, E.; Kohn, W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11635–11638.
- (15) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P., Jr.; R., A. D.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (16) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122* (154102), 1–12.
- (17) Lao, K. U.; Herbert, J. M. *J. Phys. Chem. A* **2012**, *116*, 3042–3047.
- (18) Lao, K. U.; Herbert, J. M. *J. Phys. Chem. Lett.* **2012**, *3*, 3241–3248.
- (19) Herbert, J. M.; Jacobson, L. D.; Lao, K. U.; Rohrdanz, M. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7679–7699.
- (20) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193–2199.