Scalable generalized screening for high-order terms in the many-body expansion: Algorithm, open-source implementation, and demonstration

View Online   Export Citation   CrossMark

Dustin R. Broderick [iD] and John M. Herbert[a) [iD]

AFFILIATIONS

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

Note: This paper is part of the JCP Special Topic on Modular and Interoperable Software for Chemical Physics.
[a)]Author to whom correspondence should be addressed: herbert@chemistry.ohio-state.edu

ABSTRACT

The many-body expansion lies at the heart of numerous fragment-based methods that are intended to sidestep the nonlinear scaling of *ab initio* quantum chemistry, making electronic structure calculations feasible in large systems. In principle, inclusion of higher-order n-body terms ought to improve the accuracy in a controllable way, but unfavorable combinatorics often defeats this in practice and applications with $n \geq 4$ are rare. Here, we outline an algorithm to overcome this combinatorial bottleneck, based on a bottom-up approach to energy-based screening. This is implemented within a new open-source software application ("Fragment"), which is integrated with a lightweight semi-empirical method that is used to cull subsystems, attenuating the combinatorial growth of higher-order terms in the graph that is used to manage the calculations. This facilitates applications of unprecedented size, and we report four-body calculations in $(H_2O)_{64}$ clusters that afford relative energies within 0.1 kcal/mol/monomer of the supersystem result using less than 10% of the unique subsystems. We also report n-body calculations in $(H_2O)_{20}$ clusters up to $n = 8$, at which point the expansion terminates naturally due to screening. These are the largest n-body calculations reported to date using *ab initio* electronic structure theory, and they confirm that high-order n-body terms are mostly artifacts of basis-set superposition error.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0174293

I. INTRODUCTION

Quantum-chemical fragmentation approaches replace an intractable electronic structure calculation with a large number of relatively small subsystem calculations.[1] Numerous methods of this sort have their theoretical basis in the many-body expansion (MBE) or its generalizations.[1–4] The simplest form of the MBE expresses the total energy as n-body corrections to the sum of fragment energies:

$$E = \sum_{I=1}^{N} E_I + \sum_{I=1}^{N} \sum_{J>I}^{N} \Delta E_{IJ} + \sum_{I=1}^{N} \sum_{J>I}^{N} \sum_{K>J}^{N} \Delta E_{IJK} + \cdots \quad (1)$$

where

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \quad (2)$$

is the two-body interaction energy,

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{JK} - \Delta E_{IK} - E_I - E_J - E_K \quad (3)$$

is a three-body correction, etc. Higher-order terms are expressible in a natural way.[5] Truncation of Eq. (1) at some finite order, n, affords an approximation that we will call MBE(n). This machinery has been exploited as a means to fit classical force fields,[6–8] especially for water[9–13] and for ion–water interactions,[14–19] as well as for high-accuracy calculations of crystal lattice geometries[20] and energies.[21–29] MBE(n) approximations have also been used to accelerate molecular dynamics simulations,[30,31] and as a means to limit the complexity of chemical space in applications of machine learning.[32–37] Finally, MBE(n) forms the basis of the fragment molecular orbital method.[38]

For a system with N fragments, MBE(n) generates a combinatorial number of n-body subsystems,

$$_N C_n = \binom{N}{n} \sim N^n. \quad (4)$$

Assuming that N grows linearly with system size, the asymptotic cost of MBE(n) then grows as $\mathcal{O}(N^n)$, absent some kind of screening

approximation to exclude unimportant subsystems.[1,39–41] The combinatorics associated with Eq. (4) can lead to catastrophic loss-of-precision,[5,42,43] e.g., when embedding charges are used,[5] or for $n > 4$.[43] High-order expansions with $n \geq 5$ have seen little use, and mostly in small molecular clusters.[43–58] Very few studies have examined higher-order terms for $N \geq 20$ fragments.[43,56]

Much of the work with high-order MBE(n) methods has focused on the effects of basis-set superposition error (BSSE), which affects the n-body subsystems differently from how it affects the supersystem, leading to artifacts when the supersystem result is used to assess convergence.[40,52,58–62] To address this, many-body versions of the Boys–Bernardi counterpoise correction have been developed.[40,59–63] By means of these corrections, or else simply by using very large basis sets to eliminate BSSE, it has been determined that MBE(n) for neat water clusters converges at $n = 4$,[40,42,56–58] or possibly $n = 5$.[48] Higher-order n-body interactions persist in ion–water clusters[53,54] and in some response properties,[51,52] whereas MBE(n) energies for proteins appear to converge at $n = 3$, using single-residue fragments.[64] These observations suggest that the vast majority of n-body interactions are insignificant so long as BSSE is taken into account.

Distance-based thresholding has often been considered as a means to tame the combinatorics of MBE(n), but for high-order expansions this effectively limits the subsystems to nearest-neighbors only. This may omit cooperative dipolar interactions that can sometimes cause energetically-important terms to persist at longer length scales.[39] Within a distance-based screening approach, these terms must either be omitted (at the expense of accuracy), or else the distance cutoff must be relaxed, which is detrimental to efficiency. Distance-based screening must furthermore be done carefully (by means of smooth cutoffs) in order to avoid introducing discontinuities in potential energy surfaces.[40,65] *Ab initio* molecular dynamics simulations are quite sensitive to small inconsistencies between the energy and forces, which can lead to catastrophic failure to conserve energy.[66]

In previous work,[41] we introduced an energy-screened MBE(3) method in which a polarizable force field was used to screen the three-body terms. The target level of *ab initio* theory was applied to trimer IJK only if

$$\left| \Delta E_{IJK}^{\text{low}} \right| \geq \tau_3, \tag{5}$$

where $\Delta E_{IJK}^{\text{low}}$ indicates the three-body correction evaluated using a low-level method (force field), and the cutoff τ_3 was set to 0.25 kJ/mol following some testing. This approach proves to be far more efficient than distance-based screening, if the cutoffs are set (in either case) so as to achieve ~1 kcal/mol fidelity with respect to a supersystem calculation at the same level of theory. Four-body terms were omitted in our previous work,[41] with the effect that 1 kcal/mol fidelity requires an additional supersystem correction at a low level of theory, in order to capture four-body polarization that is neglected by MBE(3).

Inclusion of higher-order terms may obviate the need for such a correction, thus the present work is an effort to implement energy screening in a general way at all orders of MBE(n). We describe a new "bottom-up" algorithm, by means of which an arbitrary screening procedure may be extended from low-order to high-order n-body subsystems, for arbitrary n, without the need to define bespoke

screening algorithms for different n-body levels. The bottom-up algorithm is demonstrated in applications using MBE(n) up to order $n = 8$, making use of a new software application in which MBE(n) is tightly integrated with the semi-empirical extended tight-binding (xTB) model.[67,68] This replaces the polarizable force field that was used previously as the low-level screening method, and represents a much more flexible and general means to screen the low-order subsystems.

II. METHODS

For a given subsystem $IJK\cdots$, energy corrections $\Delta E_{IJK\cdots}^{\text{low}}$ are computed using a low-level screening method. Here, we use the semi-empirical GFN2-xTB model[67,68] for that purpose. If

$$\left| \Delta E_{IJK\cdots}^{\text{low}} \right| < \tau_n \tag{6}$$

then the corresponding n-body correction $\Delta E_{IJK\cdots}$ is excluded in Eq. (1), and the parameter τ_n serves as a controllable threshold for screening the n-body interactions. In previous work,[41] we used the effective fragment potential[69] as the low-level screening method, but we have transitioned to GFN2-xTB due to its greater input flexibility and user-friendly software interface.

A. Extensible, generalizable many-body expansion

The generalized (G)MBE of Richard and Herbert[1–4] extends the simple MBE of Eq. (1) to handle overlapping fragments, meaning those that may share some atoms in common. Many other fragmentation schemes can be viewed as special cases of (or approximations to) the GMBE,[1–3] recovering the conventional MBE in Eq. (1) when the fragments are disjoint. The fragments used in the present work will be disjoint (specifically, single H_2O molecules in water clusters), yet the set-theoretical formulation of the GMBE is still useful.

Here, we present a variation on the GMBE that can be updated iteratively to include new fragments. We regard each fragment F_A as a set of atoms

$$F_A = \{A_1, A_2, \ldots, A_n\}, \tag{7}$$

and a fragmentation scheme is composed of a unique set of fragments

$$S = \{F_A, F_B, \ldots, F_N\}. \tag{8}$$

Starting with a scheme

$$S_x = \{F_A, F_B, F_C, \ldots, F_N\}, \tag{9}$$

where x represents the scheme's current state in some sequence of additions, the corresponding approximation to the supersystem energy is

$$E^{(x)} = \sum_{i \in S_x} C_{i,x} E_i \tag{10}$$

where E_i is the energy of fragment F_i. Coefficients $C_{i,x}$ arise from the principle of inclusion-exclusion (PIE) in state x, which prevents double-counting.[2,70,71] When a new fragment F_α is added, the scheme S_x is updated according to

$$S_{x+1} = S_x \cup \{F_\alpha\} \cup \{F_\alpha \cap F_i : F_i \in S_x\}. \tag{11}$$

The corresponding energy update is

$$E^{(x+1)} = E^{(x)} + \Delta E^{(x+1)} \qquad (12)$$

where

$$\Delta E^{(x+1)} = E_\alpha - \sum_{i \in S_x} C_{i,x} E_{i \cap \alpha}. \qquad (13)$$

Here, $C_{i,x}$ is the current coefficient for fragment F_i and $E_{i \cap \alpha}$ is the energy of the fragment formed from the intersection $F_i \cap F_\alpha$. For brevity, we refer to fragments using their subscripts ($A := F_A$) and unions of fragments will be denoted using $AB := F_A \cup F_B$.

B. Fragment taxonomy

A fragmentation scheme consists of three types of fragments: (1) *primary fragments* are small and user-defined, and may overlap; (2) *primitive fragments* are the largest non-overlapping fragments derived from the intersections of primary fragments; and (3) *auxiliary fragments* include the primary fragments and all unions thereof. This taxonomy is illustrated in Fig. 1.

Relationships between fragments can be represented using a directed acyclic graph (DAG), wherein each unique fragment is a node. Edges are drawn from fragments subsystems (parents) to larger fragments constructed from them (children). We refer to this structure as a PIE-DAG, although it is also known in set theory as a Hasse diagram.[72,73] PIE-DAGs are constructed using the following heuristics:
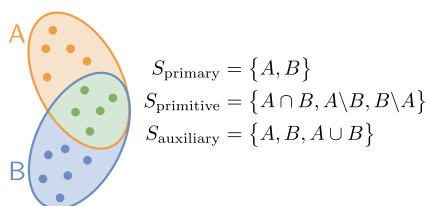
1. Each node represents a fragment (a set of atoms).

$$S_{\text{primary}} = \{A, B\}$$
$$S_{\text{primitive}} = \{A \cap B, A \backslash B, B \backslash A\}$$
$$S_{\text{auxiliary}} = \{A, B, A \cup B\}$$

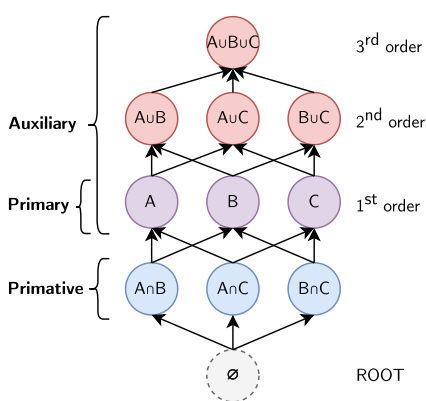**FIG. 1.** Illustration of the fragment taxonomy.

FIG. 2. PIE-DAG for system with disjoint primary fragments $\{A, B, C\}$, meaning that $A \cap B \cap C = \emptyset$.

2. A child node is the union of its parents.
3. A node cannot have direct parents that are subsets of one another.

An example of a PIE-DAG for system with three primary fragments is illustrated in Fig. 2. Note that $A \cap B \cap C = \emptyset$ in this example.

In practice, a fragmentation scheme can be bootstrapped starting from a single auxiliary fragment ($S_0 = \{F_A\}$), followed by sequential additions as in Eq. (11). A screening procedure can be implemented to manage this process. In the present work, the PIE-DAG data structure is used to explore two possible strategies for screening auxiliary fragments, which are described in Sec. IV.

C. Test systems

In Sec. IV, we will evaluate the efficacy of two different energy-based screening algorithms. For demonstration purposes, the screening criterion in Eq. (6) will only be used for $n \geq 3$, with all one- and two-body terms evaluated explicitly. (This also allows us to test screening approximations for the higher-order terms in a manner that is uncontaminated by any approximations introduced at the pairwise level.) As test systems, we will use several sets of $(H_2O)_N$ clusters with $N = 20, 32$, and 64, some of which were taken from previous literature.[74,75] As in previous work,[41,64] an ONIOM-style[1,76] supersystem correction is used to correct for long-range electrostatic and polarization interactions.

Calculations on $(H_2O)_{20}$ and $(H_2O)_{32}$ clusters were performed at the ωB97X-V/def2-TZVPD level with a supersystem correction at the level of Hartree–Fock (HF)/def2-TZVPD. This combination of methods is not a realistic use case for MBE(n) approximations, since both the fragment method and the supersystem correction exhibit the same formal scaling with respect to system size, but this choice facilitates high throughput and is useful for evaluating different algorithms and thresholds. For a more realistic application of fragmentation, we performed calculations on $(H_2O)_{64}$ at the level of second-order Møller–Plesset perturbation theory (MP2), using the jun-cc-pVDZ basis set and a resolution of identity (RI) approximation, in conjunction with a supersystem correction at the HF/jun-cc-pVDZ level.

III. SOFTWARE

The extensible GMBE algorithm described in Sec. II has been implemented in Fragme∩t, a new Python-based application designed for fragmentation calculations.[77] It succeeds a C++ code of the same name that was developed in our group over many years and used in previous (G)MBE studies.[2–5,40–43,59,60,78,79] The new Python application has already been used to obtain converged thermochemical quantities for enzyme-catalyzed reactions,[64] though only non-covalent fragmentation is considered in the present work.

An overview of the fragmentation process is shown in Fig. 3. The Fragme∩t code loads molecular systems, generates primary fragments, and then iteratively generates and screens auxiliary fragments to build the PIE-DAG that was described in Sec. II A. Fragme∩t also includes code capable interfacing with Q-Chem,[80] PySCF,[81] NWChem,[82] ORCA,[83] GFN-xTB,[67,68] and MOPAC.[84] Among these, codes that rely on text input and output files are called as subprocesses and the output files are parsed, whereas codes such
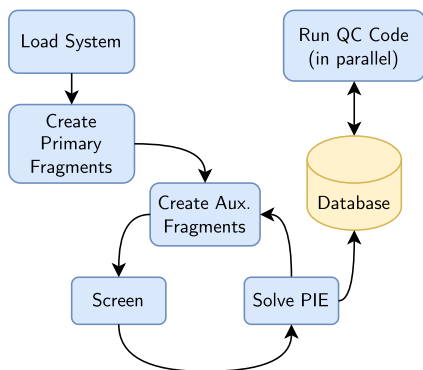
FIG. 3. Fragmentation workflow as implemented in Fragme∩t. Once the supersystem's geometry has been loaded and split into primary fragments, it is passed to Fragme∩t's main loop, which iteratively proposes new auxiliary fragments. These may or may not be added to the PIE-DAG, depending on the outcome of the screening procedure. New fragments and PIE metadata are stored in a SQLite database. Electronic structure calculations are run in parallel and are also stored in the database.

as GFN-xTB and PySCF that can be called as libraries are interfaced directly, to bypass overhead from system calls. Parallel calculations are supported through a prototype HTTP-based architecture capable of running ~60 simultaneous calculations. Calculations reported here were performed by interfacing to Q-Chem for the density functional theory and RIMP2 calculations, and to GFN2-xTB for energy screening.

In developing Fragme∩t, our goal is to provide a framework for rapid implementation and testing of new and existing fragmentation methods. In part, we hope this will facilitate *community validation* of fragmentation methods, by making it relatively easy to perform side-by-side comparisons of different fragmentation methods. (Few such comparisons exist in the literature,[2,3,43,71,85] and several of those that do were performed using Fragme∩t's precursor C++ code.[2,3,43]) To this end, Fragme∩t allows plugins to customize each step of a fragment-based calculation. Python was selected in order to lower the barrier to entry for new developers. Data are saved to a SQLite database, which makes Fragme∩t easy to deploy and allows data to be exchanged between a high-performance computing environment and a local installation, for ease of analysis. The code is presently under rapid development but a pre-release version is available on GitLab.[77]

IV. ALGORITHMS AND TESTS

We next describe implementation and testing of generalized energy-based screening. Screening has directionality within the PIE-DAG paradigm, and we will describe both a top-down and a bottom-up algorithm to implement energy-based screening. Tests presented in this section will demonstrate the superior efficacy of the bottom-up approach.

A. Top-down screening

Top-down screening starts with the high-order nodes on the PIE-DAG and progressively extends this to lower-order nodes. This

is illustrated at the MBE(3) level in Fig. 4(a), for an example consisting of non-overlapping primary fragments $S_{\text{primary}} = \{A, B, C, D\}$. The screening procedure is first applied to the trimers ABC, ABD, ACD, and BCD. Terms that meet the screening criteria serve as initial nodes added to the PIE-DAG, and in this example we assume (for illustrative purposes) that only ABC, ACD, and BCD meet the screening criteria, not ABD. However, the fact that ABD is negligible does not guarantee that any of AB, AD, or BD is negligible, so each of these (and all other dimers) need to be tested. (Again for the sake of illustration, we assume in Fig. 4 that AD and BD are negligible but that AB is not.) Lastly, all one-body (primary) terms are added to the graph, ensuring that it contains all atoms.

As mentioned in Sec. II C, we do not actually screen the two-body interactions in this work, and instead evaluate all of the one-
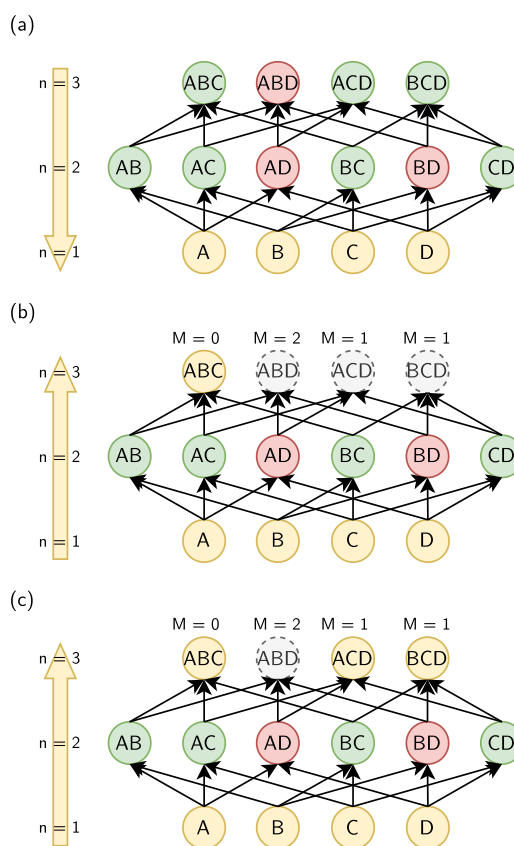

FIG. 4. Schematic view of PIE-DAG fragmentation algorithms based non-overlapping primary fragments $S_{\text{primary}} = \{A, B, C, D\}$, traversed in the direction of the yellow arrows at left. (a) Top-down fragmentation method where all high-order fragments are screened first, followed by successively lower-order fragments. For the sake of example, fragments in red are assumed to be excluded based on screening criteria. Black arrows indicate parent/child relationships. (b) Bottom-up fragmentation where low-order terms are screened and added to the PIE-DAG first (again with fragments in red excluded due to screening), followed by successively higher-order terms. If a high-order term does not have all of its parents, it is not considered for addition, as indicated in gray. (c) Bottom-up screening using the parentage parameter M, with $M = 1$ in this illustration. A high-order term may be considered for addition if it is missing M or fewer parents.

and two-body terms at the target level of theory. (This means one fewer parameter to test in the present work, and two-body screening can easily be incorporated later.) The graph in Fig. 4(a) is intended as a schematic to illustrate that within the top-down approach, any sparsity that may be obtained via screening at the n-body level does not propagate downward, meaning that all of the $(n-1)$-body terms must be screened. This is true for higher-order interactions as well, beyond the $n = 3$ case that is illustrated in Fig. 4(a).

Figure 5 illustrates the results of top-down energy screening applied to a test set of $(H_2O)_{20}$ clusters at the MBE(3) and MBE(4) levels. The top-down method requires a screening procedure at each n-body level, for which we use GFN2-xTB with a three-body threshold τ_3 ranging from $\tau_3 = 0$ (meaning no screening) up to 0.15 kcal/mol. For simplicity, the four-body threshold (τ_4) is set to $\tau_4 = \alpha\tau_3$, for $\alpha = 0.05, 0.10$, or 0.20. Larger values of τ_n indicate more aggressive screening, resulting in fewer subsystems. Throughout this work, error is defined as

$$\text{error} = E_{\text{MBE}} - E_{\text{supersys}} \tag{14}$$

where E_{supersys} is a conventional supramolecular calculation at the "target" level of theory that is used for the MBE(n) terms.

From Fig. 5(a) we find that MBE(3) maintains sub-kcal/mol accuracy provided that $\tau_3 \leq 0.1$ kcal/mol, whereas for larger values there is a rapid increase in the error as the screening becomes too aggressive. This loss of accuracy for $\tau_3 > 0.1$ kcal/mol is substantially mitigated by moving to a four-body expansion. This is true for all values of τ_4 that we test, although the most restrictive value ($\tau_4 = 0.2\tau_3$) does exhibit absolute errors exceeding 1 kcal/mol when $\tau_3 > 0.1$ kcal/mol. Setting $\tau_4 = 0.1\tau_3$ maintains the absolute error within 1 kcal/mol at the expense of increased fragment counts. The latter settings retain most of the trimers, which are generated from the four-body overlaps, and this masks the influence of τ_3. For top-down screening to be effective, both τ_3 and τ_4 must be tuned to strike a balance between accuracy and efficiency, where efficiency

is measured by the fragment counts that appear in the lower panels of Fig. 5.

B. Bottom-up screening

PIE-DAGs can also be constructed upwards from the lowest-order fragments, as illustrated in Fig. 4(b), eliminating higher-order nodes based on whether low-order ones are significant. Somewhat similar procedures have sometimes been used to eliminate higher-order contributions to potential surfaces,[86] which can be couched in a MBE-type formalism,[87] although in that context the expansion is seldom extended to high order. Here, we will develop a general algorithm and show that it can be carried to arbitrary (self-terminating) order in realistic test cases.

The bottom-up process begins by including all primary fragments in the graph, whereupon dimers are added if they meet the screening criteria. In the hypothetical example of Fig. 4(b), dimers AD and BD are determined to be negligible, as they were in the corresponding top-down example of Fig. 4(a). Trimers and higher-order fragments are added subsequently, with the constraint that in order for a new node (representing a non-negligible subsystem) to be added, all of its parents must exist already within the PIE-DAG, meaning that each parent must be non-negligible according to the screening criteria. Under these conditions, the nodes ABD, ACD, and BCD are not considered for addition to the scheme in Fig. 4(b), meaning that they are not even tested with the low-level screening method. The node ABC (depicted in yellow) is considered, and $\Delta E_{ABC}^{\text{low}}$ is evaluated; this trimer will be included in the scheme if and only if it satisfies the three-body screening criteria, meaning Eq. (6) in the present implementation.

Inclusion of a particular trimer IJK depends upon the presence of its parents: IJ, IK, and JK. In the algorithm that we have just described, which is illustrated in Fig. 4(b), each of these parents must be present in order for IJK to be included in the PIE-DAG. We call this the $M = 0$ algorithm because no missing parents are
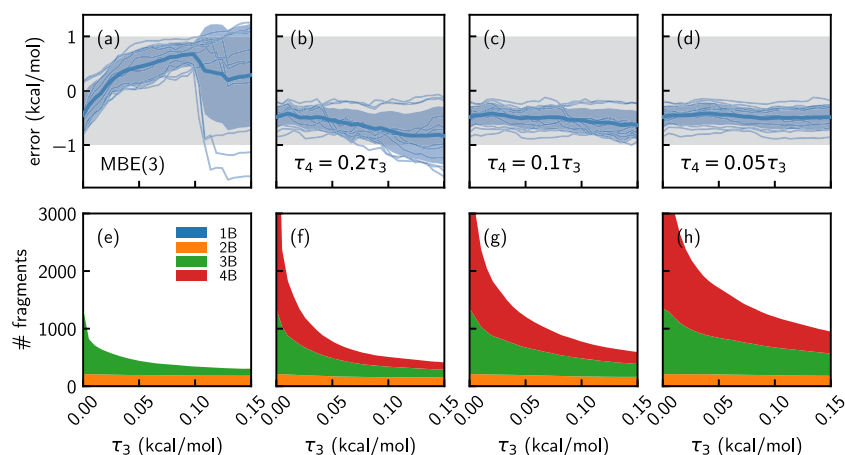

FIG. 5. Top-down energy screening applied to a set of 12 different $(H_2O)_{20}$ clusters at the ωB97X-V/def2-TZVPD level using a HF/def2-TZVPD supersystem correction. Traces in the upper panels show the error for each individual cluster as a function of τ_3, with the mean error in bold, for (a) MBE(3) and (b)–(d) MBE(4) with various choices for τ_4. The blue shaded region delineates one standard deviation around the mean and the gray shaded region indicates the target error of ± 1 kcal/mol. Lower panels (e)–(h) show mean fragment counts at each order in the n-body expansion, in the form of stacked plots.
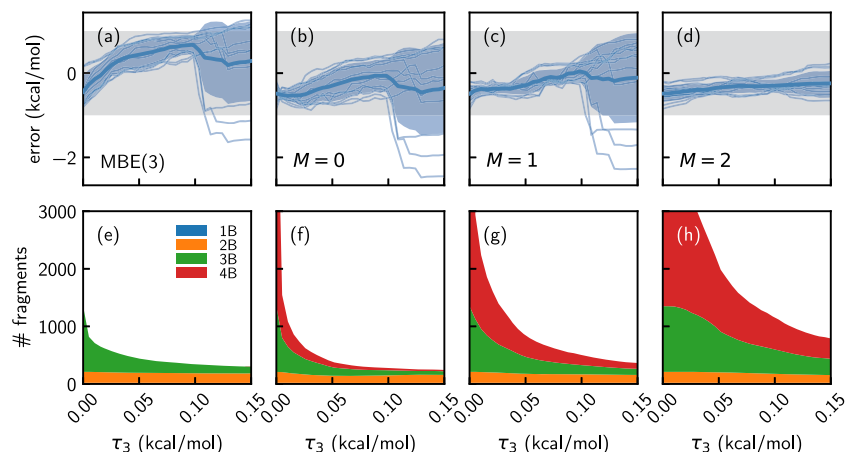
FIG. 6. Bottom-up energy screening applied to the same $(H_2O)_{20}$ clusters that were used in Fig. 5, at the ωB97X-V/def2-TZVPD level with a HF/def2-TZVPD supersystem correction. Traces in the upper panels show the error for each cluster as a function of τ_3, for (a) MBE(3) and (b)–(d) MBE(4) with $M = 0$, 1, or 2. (The screening algorithm is fully specified by τ_3 and M so there is no τ_4 in the bottom-up approach.) Lower panels (e)–(h) show the mean fragment counts in the form of stacked plots.

permitted. Relaxing this constraint, to allow a single missing parent, affords what we call the $M = 1$ algorithm and this is illustrated in Fig. 4(c). Relative to the $M = 0$ algorithm, use of $M = 1$ means considering ACD and BCD, which are thus evaluated using the low-level screening method and included in the PIE-DAG if they meet the screening criteria. Use of $M = 1$ may provide better accuracy in cases where one fragment significantly polarizes two others, such as

an ion located between two water molecules, where the corresponding water dimer might not have been strongly interacting (due to the separation between the water molecules) when considered in the absence of the intervening ion.

Figure 6 shows the bottom-up algorithm applied to the same set of $(H_2O)_{20}$ clusters that were used to test the top-down method. Energy screening is applied *only* to the three-body terms, retaining

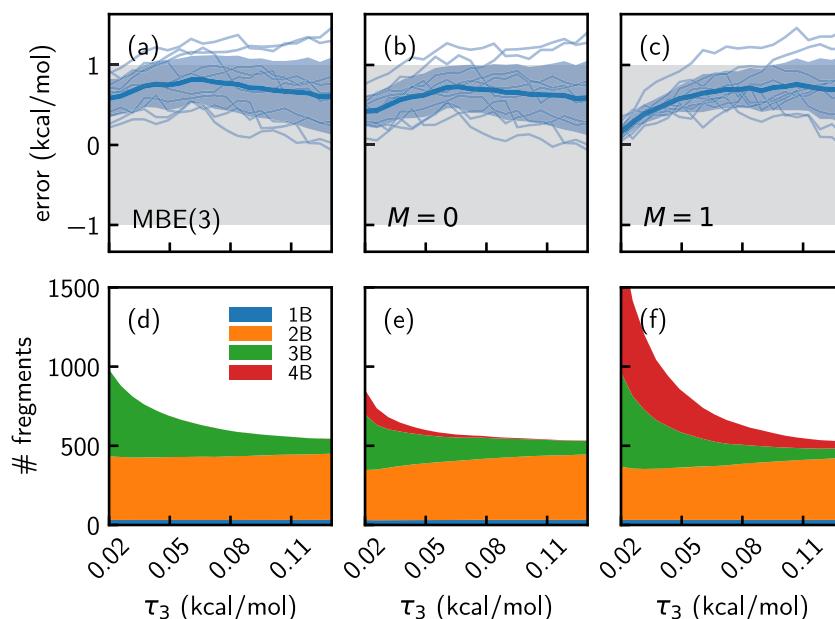**FIG. 7.** Bottom-up energy screening on a collection of 10 $(H_2O)_{32}$ clusters, at the ωB97X-V/def2-TZVPD level with a HF/def2-TZVPD correction. As in Figs. 5 and 6, the traces in the upper panel show the error for each cluster isomer as a function of τ_3, for (a) MBE(3) vs (b)–(c) MBE(4) with either $M = 0$ or 1. Lower panels (d)–(f) show the mean fragment counts.

all one- and two-body terms (as in the top-down example). In this case, four-body terms are accepted or rejected based strictly on parentage ($M = 0$, 1, or 2), rather than using a four-body threshold τ_4.

For all three values of M, the number of fragments decreases as the screening on the lower-level fragments becomes stricter (larger τ_3). A sharp drop in accuracy for MBE(3) is again observed for $\tau_3 > 0.1$ kcal/mol [Fig. 6(a)], which remains evident for MBE(4) using either $M = 0$ [Fig. 6(b)] or $M = 1$ [Fig. 6(c)]. This indicates that addition of four-body terms does not obfuscate the screening that is performed at the three-body level. However, the choice $M = 2$ [Fig. 6(d)] does appear to interfere with the three-body screening, leading to a significant growth in the number of subsystems [Fig. 6(h)]. For $M = 2$, errors do remain tightly clustered even for $\tau_3 > 0.1$ kcal/mol, so in some sense this approach is able to inoculate the method against a poor choice of τ_3, yet that security comes at significant cost. Setting $M = 1$ and $\tau_3 = 0.05$ kcal/mol provides a good balance between accuracy and efficiency.

To evaluate the efficacy of bottom-up energy screening using larger systems, a similar analysis is presented for $(H_2O)_{32}$ clusters in Fig. 7 and for $(H_2O)_{64}$ clusters in Fig. 8. In order to make a direct comparison to results presented above, we use the same levels of theory for $(H_2O)_{32}$ as for the $(H_2O)_{20}$ clusters. However, calculations for $(H_2O)_{64}$ are performed at a level of theory that constitutes a more realistic use case for MBE(n), with subsystem calculations evaluated at the RIMP2/jun-cc-pVDZ level of theory in conjunction with a HF/jun-cc-pVDZ supersystem correction. Error is then measured relative to a supersystem calculation at the RIMP2/jun-cc-pVDZ level. As demonstrated in previous energy-screened MBE(3)

calculations,[41] this results in dramatic speedups without large-scale parallelization, despite the need for a supersystem HF calculation, while maintaining 1 kcal/mol fidelity for relative energies. These $(H_2O)_{64}$ calculations represent the largest application of MBE(4) in electronic structure theory of which we are aware.

As cluster size increases a systematic error emerges, in both the $(H_2O)_{32}$ and $(H_2O)_{64}$ calculations, that is larger than the (relatively constant) error in amongst relative energies of different clusters of the same size. To demonstrate that this systematic error is due to BSSE, we performed MBE(3) and MBE(4) calculations for a single $(H_2O)_{64}$ isomer using the full cluster basis set, meaning that n-body subsystem calculations are carried out using the basis set of the $(H_2O)_{64}$ supersystem. Errors in these cluster-basis calculations are plotted in Fig. 8(c) alongside the corresponding errors obtained using the conventional subsystem basis. Use of the cluster basis reduces the error below 0.025 kcal/mol/monomer for both MBE(3) and MBE(4), whereas the use of the subsystem basis afford errors that grow with τ_3, even when that threshold is set quite conservatively. Although the use of a supersystem basis set may not be practical in large-scale calculations, in future work this can replaced with many-body counterpoise corrections.[59]

V. DISCUSSION

Both top-down and bottom-up screening can achieve good fidelity with respect to the corresponding supersystem calculation at the same level of theory. However, top-down fragmentation has a significant drawback in that it requires a well-defined screening protocol for each n-body order, and lower orders in the expansion do
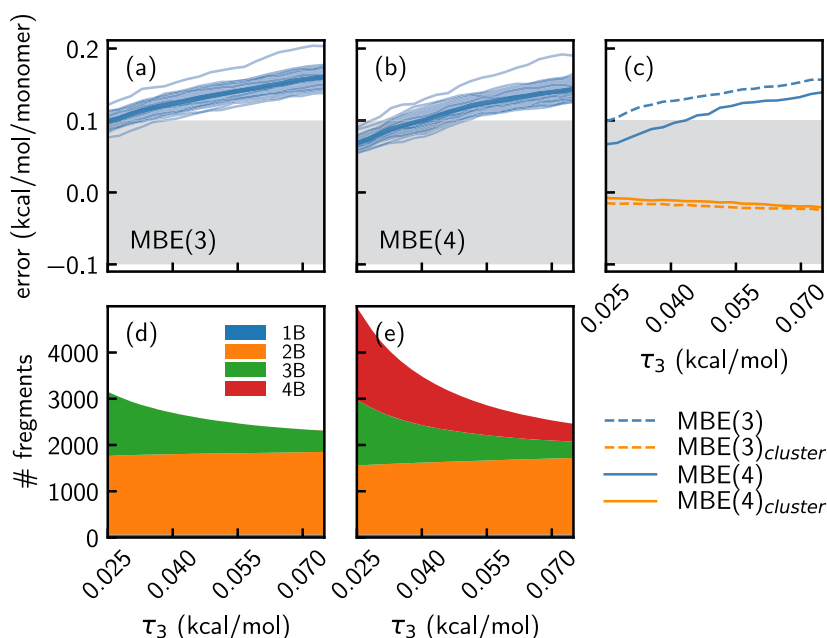
FIG. 8. Bottom-up energy screening for a set of 11 $(H_2O)_{64}$ clusters, at the RIMP2/jun-cc-pVDZ level with a HF/jun-cc-pVDZ supersystem correction. Traces in the upper panel show the errors for (a) MBE(3) and (b) MBE(4) with $M = 1$, both as a function of τ_3 and reported in per-monomer units. (c) Errors for both MBE(3) and MBE(4) for one particular snapshot, in either the usual subsystem basis set (in blue) vs the full cluster basis set (in orange). (d)–(e) Mean fragment counts as a function of τ_3.

not benefit directly from the screening performed at higher orders when it comes to constructing the graph of non-negligible subsystem nodes. Even for a screening protocol as simple as Eq. (6), where n-body terms are excluded based on an energy threshold τ_n, the top-down approach is impractical for large systems when $n > 3$, because every high-order n-body subsystem must be considered individually.

To put this into perspective, and to contrast the efficiency of the bottom-up approach, Fig. 9 presents average timing data to construct the PIE-DAG for $(H_2O)_{20}$, for the calculations reported in Figs. 5 and 6. Cached GFN2-xTB screening data were used to eliminate variations in run times, such that the only cost associated with energy screening is that of a hash lookup for the appropriate energy. When τ_3 is small, both algorithms exhibit a relatively large fragmentation cost because the PIE-DAG is quite dense; however, top-down screening has a high fixed cost even when the screening is more aggressive (larger τ_3). Bottom-up screening is more efficient even when τ_3 is small because it does not attempt to screen fourth-order and higher fragments whose parentage is incomplete. The present implementation can perform a single ΔE_{IJKL}^{low} calculation (using GFN2-xTB) in 0.02 s, and for top-down screening this equates to 4 h just to generate and screen the 679 120 unique subsystems required to apply MBE(4) to $(H_2O)_{64}$. Bottom-up screening can fragment the same system in less than 5 min.

This has significant implications for counterpoise correction schemes that can be used to address BSSE but which have historically been underutilized in fragment-based quantum chemistry, perhaps due to cost. The data in Fig. 8(c) demonstrate that error increases as the number of high-order subsystems is reduced by more aggressive screening, which suggests that most of the high-order terms in MBE(n) contribute nothing more than incomplete error cancellation. Counterpoise corrections for MBE(n) and its generalizations have been developed[40,59,61–63,88–90] but are not yet implemented in our code. These methods introduce an additional combinatorial scaling factor but aggressive energy screening may render these corrections tractable.
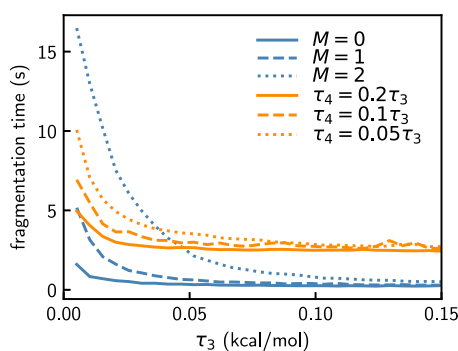
Because the bottom-up algorithm propagates the screening procedure from low orders to high orders, it allows the energy-screened MBE(n) to be extended to arbitrarily high orders. Note that each n-body node in the PIE-DAG has n parents, each of which is a $(n-1)$-body subsystem. Since higher-order subsystems have more parents than lower-order ones, fewer terms need to be considered as n increases, even prior to evaluating $\Delta E_{IJKL...}^{low}$. This natural attenuation is demonstrated in Fig. 10 using $(H_2O)_{20}$ clusters. For each cluster, we ran the bottom-up algorithm up to $n = 20$, with $\tau_3 = 0.05$ kcal/mol and $M = 1$. For 10 out of the 12 clusters in the data set, we found that the procedure terminated naturally (due to a lack of parents) at $n = 7$, and for the remaining two clusters it terminated at $n = 8$.

Figure 10(b) shows the number of unique subsystems generated at each n-body order, excluding terms where $C_{i,x} = 0$ in Eq. (12). An interesting feature of these data is that the number of three-, four-, and five-body terms actually decreases as n increases. In essence, the combination of higher-order n-body terms with energy screening serves to localize inter-fragment interactions into larger domains (characterized by larger n), which are capable of accounting for all inter-domain interactions. The most dramatic examples of this effect are the fused-cube isomers No. 1 and No. 20 (which are provided in the supplementary material), for which the addition of a single eight-body fragment to a MBE(7) scheme reduces the number of subsystems by ≈ 200 for either cluster.

It is commonly asserted that high-order n-body terms are energetically insignificant. In particular, for homogeneous water clusters it has been assumed that the n-body corrections are negligible beyond $n = 4$,[39,40,42,56–58] or possibly $n = 5$.[29,48] (Higher-order terms undoubtedly persist in ion–water clusters.[53,54]) A careful analysis of this assertion in large ($N \gg 20$) water clusters has been hindered by intractable combinatorics, but should now be feasible and we hope to report on this in due course. For now, we note that that terms beyond $n = 5$ in $(H_2O)_{20}$ clusters converge to a constant, nonzero error [Fig. 10(a)]. We ascribe this constant offset to the combined

FIG. 9. Mean fragmentation times for $(H_2O)_{20}$ clusters using bottom-up energy screening (in blue) vs top-down screening (in orange). In both cases, cached data from GFN2-xTB were used for screening, so these timings simply reflect the lookup cost and the cost to construct the PIE-DAG. The top-down method requires data for every tetramer whereas the bottom-up approach does not. The steep cost increase for small values of τ_3 results from construction of a dense PIE-DAG. Benchmarks were performed on a laptop with a 2.6 GHz dual-core Intel Core i5 processor and 16 Gb of memory.
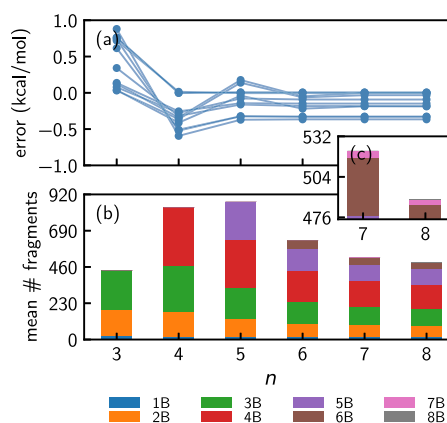

FIG. 10. Bottom-up MBE(n) calculations using the $(H_2O)_{20}$ test set, setting $\tau_3 = 0.05$ kcal/mol and $M = 1$. (a) Errors for all 12 cluster isomers as a function of n. (b) Mean fragment counts for each subsystem size. (c) Close-up view of the mean fragment counts for $n = 7$ and 8. No additional fragments are generated for $n \geq 9$.

effects of BSSE and perhaps some residual error due to the supersystem correction. This analysis suggests that for $N = 20$ monomers, $MBE(n)$ does indeed converge at the $n = 5$ level at worst, and perhaps sooner, if higher-quality basis sets were to be employed. Others have suggested that five-body terms in neat water clusters are BSSE artifacts that disappear in the complete-basis limit.[56] The present results certainly do not preclude that possibility, which we plan to explore in future work.

VI. CONCLUSIONS

We have described a new, open-source implementation of the straightforward $MBE(n)$ method without any kind of electrostatic embedding. The asceticism is by design, as we aim to circumvent complexities arising from self-consistent charge embedding including overly complicated analytic gradients,[66] instead using high-order n-body expansions and low-level supersystem corrections to capture cooperative polarization. Numerical problems arising from unfavorable combinatorics for $n > 3$, which we have previously documented in large-scale $MBE(n)$ calculations,[5,43] are avoided here by means of energy screening. Stability in the presence of diffuse basis functions is achieved by iterating the n-body self-consistent field calculations to convergence.

Using a new "bottom-up" screening algorithm based on the semi-empirical GFN2-xTB method,[67,68] we are able to achieve sub-kcal/mol accuracy for water clusters when $MBE(n)$ is paired with an ONIOM-style supersystem correction. This constitutes a predictive level of accuracy by means of which $MBE(n)$ calculations can reliably rank-order the energies of different hydrogen-bonding networks, which has proven to be a challenging problem for methods that use single-H_2O fragments.[1] Numerically-complete four-body results have been presented for $(H_2O)_{64}$, which represents a reasonable size for the periodic cell in a liquid water simulation. We believe these to be the largest complete application of MBE(4) to date.

Unlike the top-down screening method that we reported previously,[41] which requires energy screening for every individual n-body system, the bottom-up approach propagates screening information upwards. As a result, the screening is much more aggressive (and thus the algorithm more performant) without loss of accuracy, and we find that the number of n-body subsystems naturally attenuates as n increases. As such, $MBE(n)$ can be used without specifying a maximum order n, yet subsystem counts remain manageable. For $(H_2O)_{20}$, terms beyond $n = 5$ appear to contribute only to BSSE. In future work, this will be investigated further using larger basis sets and many-body counterpoise corrections.[59]

The software reported here[77] makes $MBE(n)$ into a robust tool for high-fidelity fragment-based quantum chemistry. Our open-source implementation is easy to interface with various quantum chemistry programs and can be used to prototype new fragmentation strategies. We anticipate that this implementation will lower the entry barrier to using fragmentation methods, and furthermore provide a degree of community validation that has heretofore been lacking in fragment-based quantum chemistry.

SUPPLEMENTARY MATERIAL

Coordinates for all test systems used here.

AUTHOR DECLARATIONS

Conflict of Interest

J.M.H. serves on the board of directors of Q-Chem Inc.

Author Contributions

Dustin R. Broderick: Conceptualization (supporting); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (equal). **John M. Herbert:** Conceptualization (lead); Formal analysis (supporting); Funding acquisition (lead); Methodology (supporting); Project administration (lead); Supervision (lead); Visualization (supporting); Writing – review and editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. Software is available from GitLab.[77]

REFERENCES

[1] J. M. Herbert, "Fantasy versus reality in fragment-based quantum chemistry," J. Chem. Phys. **151**, 170901 (2019).

[2] R. M. Richard and J. M. Herbert, "A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory," J. Chem. Phys. **137**, 064113 (2012).

[3] R. M. Richard and J. M. Herbert, "Many-body expansion with overlapping fragments: Analysis of two approaches," J. Chem. Theory Comput. **9**, 1408–1416 (2013).

[4] L. D. Jacobson, R. M. Richard, K. U. Lao, and J. M. Herbert, "Efficient monomer-based quantum chemistry methods for molecular and ionic clusters," Annu. Rep. Comput. Chem. **9**, 25–58 (2013).

[5] R. M. Richard, K. U. Lao, and J. M. Herbert, "Understanding the many-body expansion for large systems. I. Precision considerations," J. Chem. Phys. **141**, 014108 (2014).

[6] E. F. Bull-Vulpe, M. Riera, A. W. Götz, and F. Paesani, "MB-Fit: Software infrastructure for data-driven many-body potential energy functions," J. Chem. Phys. **155**, 124801 (2021).

[7] M. Riera, C. Knight, E. F. Bull-Vulpe, X. Zhu, H. Agnew, D. G. A. Smith, A. C. Simmonett, and F. Paesani, "MBX: A many-body energy and force calculator for data-driven many-body simulations," J. Chem. Phys. **159**, 054802 (2023).

[8] J. P. Heindel, K. M. Herman, and S. S. Xantheas, "Many-body effects in aqueous systems: Synergies between interaction analysis techniques and force field development," Annu. Rev. Phys. Chem. **74**, 337–360 (2023).

[9] Y. Wang, X. Huang, B. C. Shepler, B. J. Braams, and J. M. Bowman, "Flexible, *ab initio* potential, and dipole moment surfaces for water. I. Tests and applications for clusters up to the 22-mer," J. Chem. Phys. **134**, 094509 (2011).

[10] V. Babin, G. R. Medders, and F. Paesani, "Development of a 'first principles' water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters," J. Chem. Theory Comput. **10**, 1599–1607 (2014).

03 November 2023 10:18:00

[11] F. Paesani, "Getting the right answers for the right reasons: Toward predictive molecular simulations of water with many-body potential energy functions," Acc. Chem. Res. **49**, 1844–1851 (2016).

[12] G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, H. Torabifard, A. P. Bartók, G. Csányi, V. Molinero, and F. Paesani, "Modeling molecular interactions in water: From pairwise to many-body potential energy functions," Chem. Rev. **116**, 7501–7528 (2016).

[13] Q. Yu, C. Qu, P. L. Houston, R. Conte, A. Nandi, and J. M. Bowman, "q-AQUA: A many-body CCSD(T) water potential, including four-body interactions, demonstrates the quantum nature of water from clusters to the liquid phase," J. Phys. Chem. Lett. **13**, 5068–5074 (2022).

[14] D. J. Arismendi-Arrieta, M. Riera, P. Bajaj, R. Prosmiti, and F. Paesani, "i-TTM model for ab $initio$-based ion–water interaction potentials. 1. Halide–water potential energy functions," J. Phys. Chem. B **120**, 1822–1832 (2016).

[15] M. Riera, A. W. Götz, and F. Paesani, "The i-TTM model for ab $initio$-based ion–water interaction potentials. II. Alkali metal ion–water potential energy functions," Phys. Chem. Chem. Phys. **18**, 30334–30343 (2016).

[16] P. Bajaj, A. W. Götz, and F. Paesani, "Toward chemical accuracy in the description of ion–water interactions through many-body representations. I. Halide–water dimer potential energy surfaces," J. Chem. Theory Comput. **12**, 2698–2705 (2016).

[17] J. P. Heindel, Q. Yu, J. M. Bowman, and S. S. Xantheas, "Benchmark electronic structure calculations for $H_3O^+(H_2O)_n$, $n = 0$–5, clusters and tests of an existing 1,2,3-body potential energy surface with a new 4-body correction," J. Chem. Theory Comput. **14**, 4553–4566 (2018).

[18] F. Paesani, P. Bajaj, and M. Riera, "Chemical accuracy in modeling halide ion hydration from many-body representations," Adv. Phys.: X **4**, 1631212 (2019).

[19] A. Caruso and F. Paesani, "Data-driven many-body models enable a quantitative description of chloride hydration from clusters to bulk," J. Chem. Phys. **155**, 064502 (2021).

[20] K. D. Nanda and G. J. O. Beran, "Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations," J. Chem. Phys. **137**, 174106 (2012).

[21] A. L. Ringer and C. D. Sherrill, "First principles computation of lattice energies of organic solids: The benzene crystal," Chem. -Eur. J. **14**, 2542–2547 (2008).

[22] G. J. O. Beran and K. Nanda, "Predicting organic crystal lattice energies with chemical accuracy," J. Phys. Chem. Lett. **1**, 3480–3487 (2010).

[23] C. Müller, D. Usvyat, and H. Stoll, "Local correlation methods for solids: Comparison of incremental and periodic correlation calculations for the argon fcc crystal," Phys. Rev. B **83**, 245136 (2011).

[24] S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran, "Practical quantum mechanics-based fragment methods for predicting molecular crystal properties," Phys. Chem. Chem. Phys. **14**, 7578–7590 (2012).

[25] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, "Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy," Science **345**, 640–643 (2014).

[26] M. R. Kennedy, A. R. McDonald, A. E. DePrince III, M. S. Marshall, R. Podeszwa, and C. D. Sherrill, "Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory," J. Chem. Phys. **140**, 121104 (2014).

[27] G. J. O. Beran, Y. N. Heit, and J. D. Hartman, "Noncovalent interactions in molecular crystals," in *Non-Covalent Interactions in Quantum Chemistry and Physics*, edited by A. O. de la Roza and G. A. DiLabio (Elsevier, Amsterdam, 2017), Chap. 10, pp. 303–331.

[28] C. H. Borca, Z. L. Glick, D. P. Metcalf, L. A. Burns, and C. D. Sherrill, "Benchmark coupled-cluster lattice energy of crystalline benzene and assessment of multi-level approximations in the many-body expansion," J. Chem. Phys. **158**, 234102 (2023).

[29] K. M. Herman and S. S. Xantheas, "A formulation of the many-body expansion (MBE) for periodic systems: Application to several ice phases," J. Phys. Chem. Lett. **14**, 989–999 (2023).

[30] O. Demerdash and T. Head-Gordon, "Convergence of the many-body expansion for energy and forces for classical polarizable models in the condensed phase," J. Chem. Theory Comput. **12**, 3884–3893 (2016).

[31] J. P. Heindel and S. S. Xantheas, "Molecular dynamics driven by the many-body expansion (MBE-MD)," J. Chem. Theory Comput. **17**, 7341–7352 (2021).

[32] K. Yao, J. E. Herr, and J. Parkhill, "The many-body expansion combined with neural networks," J. Chem. Phys. **146**, 014106 (2017).

[33] A. Nandi, C. Qu, P. L. Houston, R. Conte, and J. M. Bowman, "Δ-Machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD(T) level of theory," J. Chem. Phys. **154**, 051102 (2021).

[34] C. Qu, Q. Yu, R. Conte, P. L. Houston, A. Nandi, and J. M. Bowman, "A Δ-machine learning approach for force fields, illustrated by a CCSD(T) 4-body correction to the MB-pol water potential," Digital Discovery **1**, 658–664 (2022).

[35] M. Shiranirad, C. J. Burnham, and N. J. English, "Machine-learning-based many-body energy analysis of argon clusters: Fit for size?," Chem. Phys. **552**, 111347 (2022).

[36] A. M. Maldonado, I. Poltavsky, V. Vassilev-Galindo, A. Tkatchenko, and J. A. Keith, "Modeling molecular ensembles with gradient-domain machine learning force fields," Digital Discovery **2**, 871–880 (2023).

[37] J. M. Bowman, C. Qu, R. Conte, A. Nandi, P. L. Houston, and Q. Yu, "Δ-machine learned potential energy surfaces and force fields," J. Chem. Theory Comput. **19**, 1–17 (2023).

[38] *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, edited by D. G. Fedorov and K. Kitaura (CRC Press, Boca Raton, 2009).

[39] J. F. Ouyang and R. P. A. Bettens, "When are many-body effects significant?," J. Chem. Theory Comput. **12**, 5860–5867 (2016).

[40] K.-Y. Liu and J. M. Herbert, "Understanding the many-body expansion for large systems. III. Critical role of four-body terms, counterpoise corrections, and cutoffs," J. Chem. Phys. **147**, 161729 (2017).

[41] K.-Y. Liu and J. M. Herbert, "Energy-screened many-body expansion: A practical yet accurate fragmentation method for quantum chemistry," J. Chem. Theory Comput. **16**, 475–487 (2020).

[42] K. U. Lao, K.-Y. Liu, R. M. Richard, and J. M. Herbert, "Understanding the many-body expansion for large systems. II. Accuracy considerations," J. Chem. Phys. **144**, 164105 (2016).

[43] R. M. Richard, K. U. Lao, and J. M. Herbert, "Aiming for benchmark accuracy with the many-body expansion," Acc. Chem. Res. **47**, 2828–2836 (2014).

[44] J. M. Pedulla, K. Kim, and K. D. Jordan, "Theoretical study of the n-body interaction energies of the ring, cage and prism forms of $(H_2O)_6$," Chem. Phys. Lett. **291**, 78–84 (1998).

[45] R. A. Christie and K. D. Jordan, "Theoretical investigation of the $H_3O^+(H_2O)_4$ cluster," J. Phys. Chem. A **105**, 7551–7558 (2001).

[46] R. A. Christie and K. D. Jordan, "N-body decomposition approach to the calculation of interaction energies of water clusters," in *Intermolecular Forces and Clusters II*, *Structure and Bonding*, edited by D. Wales and R. A. Christie (Springer-Verlag, Heidelberg, 2005), Vol. 116, pp. 27–41.

[47] B. Paulus, K. Rosciszewski, N. Gaston, P. Schwerdtfeger, and H. Stoll, "Covergence of the ab $initio$ many-body expansion for the cohesive energy of solid mercury," Phys. Rev. B **70**, 165106 (2004).

[48] U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, "Interaction energies of large clusters from many-body expansion," J. Chem. Phys. **135**, 224102 (2011).

[49] C. K. Egan and F. Paesani, "Assessing many-body effects of water self-ions. I: $OH^-(H_2O)_n$ clusters," J. Chem. Theory Comput. **14**, 1982–1997 (2018).

[50] C. K. Egan and F. Paesani, "Assessing many-body effects of water self-ions. II: $H_3O^+(H_2O)_n$ clusters," J. Chem. Theory Comput. **15**, 4816–4833 (2019).

[51] T. J. Mach and T. D. Crawford, "Computing optical rotation via an N-body approach," Theor. Chem. Acc. **133**, 1449 (2014).

[52] B. G. Peyton and T. D. Crawford, "Basis set superposition errors in the many-body expansion of molecular properties," J. Phys. Chem. A **123**, 4500–4511 (2019).

[53] J. P. Heindel and S. S. Xantheas, "The many-body expansion for aqueous systems revisited: II. Alkali metal and halide ion–water interactions," J. Chem. Theory Comput. **17**, 2200–2216 (2021).

[54] K. M. Herman, J. P. Heindel, and S. S. Xantheas, "The many-body expansion for aqueous systems revisited: III. Hofmeister ion–water interactions," Phys. Chem. Chem. Phys. **23**, 11196–11210 (2021).

[55] J. Mato, D. Tzeli, and S. S. Xantheas, "The many-body expansion for metals. I. The alkaline earth metals Be, Mg, and Ca," J. Chem. Phys. **157**, 084313 (2022).

03 November 2023 10:18:00

[56]J. P. Heindel and S. S. Xantheas, "The many-body expansion for aqueous systems revisited: I. Water–water interactions," J. Chem. Theory Comput. **16**, 6843–6855 (2020).

[57]K. Hermansson, "Many-body effects in tetrahedral water clusters," J. Chem. Phys. **89**, 2149–2159 (1988).

[58]J. F. Ouyang, M. W. Cvitkovic, and R. P. A. Bettens, "Trouble with the many-body expansion," J. Chem. Theory Comput. **10**, 3699–3707 (2014).

[59]R. M. Richard, K. U. Lao, and J. M. Herbert, "Achieving the CCSD(T) basis-set limit in sizable molecular clusters: Counterpoise corrections for the many-body expansion," J. Phys. Chem. Lett. **4**, 2674–2680 (2013).

[60]R. M. Richard, K. U. Lao, and J. M. Herbert, "Approaching the complete-basis limit with a truncated many-body expansion," J. Chem. Phys. **139**, 224102 (2013).

[61]J. F. Ouyang and R. P. A. Bettens, "Many-body basis set superposition effect," J. Chem. Theory Comput. **11**, 5132–5143 (2015).

[62]R. M. Richard, B. W. Bakr, and C. D. Sherrill, "Understanding the many-body basis set superposition error: Beyond Boys and Bernardi," J. Chem. Theory Comput. **14**, 2386–2400 (2018).

[63]M. Kamiya, S. Hirata, and M. Valiev, "Fast electron correlation methods for molecular clusters without basis set superposition errors," J. Chem. Phys. **128**, 074103 (2008).

[64]P. E. Bowling, D. R. Broderick, and J. M. Herbert, "Fragment-based calculations of enzymatic thermochemistry require dielectric boundary conditions," J. Phys. Chem. Lett. **14**, 3826–3834 (2023).

[65]J. H. Zhang, T. C. Ricard, C. Haycraft, and S. S. Iyengar, "Weighted-graph-theoretic methods for many-body corrections within ONIOM: Smooth AIMD and the role of high-order many-body terms," J. Chem. Theory Comput. **17**, 2672–2690 (2021).

[66]J. Liu, B. Rana, K.-Y. Liu, and J. M. Herbert, "Variational formulation of the generalized many-body expansion with self-consistent embedding charges: Simple and correct analytic energy gradient for fragment-based *ab initio* molecular dynamics," J. Phys. Chem. Lett. **10**, 3877–3886 (2019).

[67]C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions," J. Chem. Theory Comput. **15**, 1652–1671 (2019).

[68]C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, "Extended tight-binding quantum chemistry methods," WIREs Comput. Mol. Sci. **11**, e1493 (2020).

[69]I. A. Kaliman and L. V. Slipchenko, "LIBEFP: A new parallel implementation of the effective fragment potential method as a portable software library," J. Comput. Chem. **34**, 2284–2292 (2013).

[70]J. Li and S. S. Iyengar, "*Ab initio* molecular dynamics using recursive, spatially separated, overlapping model subsystems mixed within an ONIOM-based fragmentation energy extrapolation technique," J. Chem. Theory Comput. **11**, 3978–3991 (2015).

[71]J. Hellmers and C. König, "A unified and flexible formulation of molecular fragmentation schemes," J. Chem. Phys. **155**, 164105 (2021).

[72]K. A. Baker, P. C. Fishburn, and F. S. Roberts, "Partial orders of dimension 2," Networks **2**, 11–28 (1972).

[73]F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, "Networks beyond pairwise interactions: Structure and dynamics," Phys. Rep. **874**, 1–92 (2020).

[74]D. J. Wales and M. P. Hodges, "Global minima of water clusters $(H_2O)_n$, $n \le 21$, described by an empirical potential," Chem. Phys. Lett. **286**, 65–72 (1998).

[75]D. Yuan, Y. Li, Z. Ni, P. Pulay, W. Li, and S. Li, "Benchmark relative energies for large water clusters with the generalized energy-based fragmentation method," J. Chem. Theory Comput. **13**, 2696–2704 (2017).

[76]L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, and K. Morokuma, "The ONIOM method and its applications," Chem. Rev. **115**, 5678–5796 (2015).

[77]D. R. Broderick, P. E. Bowling, and J. M. Herbert, Fragmℇnt, v. 0.3.0, 2023; https://gitlab.com/john-herbert-group/fragment/-/releases/v0.3.0pre.

[78]J. Liu and J. M. Herbert, "Pair–pair approximation to the generalized many-body expansion: An efficient and accurate alternative to the four-body expansion, with applications to *ab initio* protein energetics," J. Chem. Theory Comput. **12**, 572–584 (2016).

[79]K.-Y. Liu, J. Liu, and J. M. Herbert, "Accuracy of finite-difference harmonic frequencies in density functional theory," J. Comput. Chem. **38**, 1678–1684 (2017).

[80]E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, A. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, A. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince III, R. A. DiStasio, Jr., A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard III, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock III, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, and A. I. Krylov, "Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package," J. Chem. Phys. **155**, 084801 (2021).

[81]Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, "Recent developments in the PySCF program package," J. Chem. Phys. **153**, 024109 (2020).

[82]E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la-Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, Á. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis,

D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. J. Harrison, "NWChem: Past, present, and future," J. Chem. Phys. **152**, 184102 (2020).

[83] F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, "The ORCA quantum chemistry program package," J. Chem. Phys. **152**, 224108 (2020).

[84] R. L. M. Gieseking, "A new release of MOPAC incorporating the INDO/S semiempirical model with CI excited states," J. Comput. Chem. **42**, 365–378 (2020).

[85] S. R. Pruitt, M. A. Addicoat, M. A. Collins, and M. S. Gordon, "The fragment molecular orbital and systematic molecular fragmentation methods applied to water clusters," Phys. Chem. Chem. Phys. **14**, 7752–7764 (2012).

[86] G. Rauhut, "Efficient calculation of potential energy surfaces for the generation of vibrational wave functions," J. Chem. Phys. **121**, 9313–9322 (2004).

[87] A. J. C. Varandas, "Intermolecular and intramolecular potentials: Topographical aspects, calculation, and functional representation via a double many-body expansion method," Adv. Chem. Phys. **74**, 255–338 (1988).

[88] B. H. Wells and S. Wilson, "Van der Waals interaction potentials: Many-body basis set superposition effects," Chem. Phys. Lett. **101**, 429–434 (1983).

[89] P. Valiron and I. Mayer, "Hierarchy of counterpoise corrections for N-body clusters: Generalization of the Boys–Bernardi scheme," Chem. Phys. Lett. **275**, 46–55 (1997).

[90] I. Mayer and I. Bakó, "Many-body energy decomposition with basis set superposition error corrections," J. Chem. Theory Comput. **13**, 1883–1886 (2017).

[91] Ohio Supercomputer Center, http://osc.edu/ark:/19495/f5s1ph73.

03 November 2023 10:18:00