Understanding the many-body expansion for large systems. I. Precision considerations

Ryan M. Richard,[a] Ka Un Lao, and John M. Herbert[b]
Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

(Received 14 January 2014; accepted 17 June 2014; published online 3 July 2014)

Electronic structure methods based on low-order "*n*-body" expansions are an increasingly popular means to defeat the highly nonlinear scaling of *ab initio* quantum chemistry calculations, taking advantage of the inherently distributable nature of the numerous subsystem calculations. Here, we examine how the finite precision of these subsystem calculations manifests in applications to large systems, in this case, a sequence of water clusters ranging in size up to $(H_2O)_{47}$. Using two different computer implementations of the *n*-body expansion, one fully integrated into a quantum chemistry program and the other written as a separate driver routine for the same program, we examine the reproducibility of total binding energies as a function of cluster size. The combinatorial nature of the *n*-body expansion amplifies subtle differences between the two implementations, especially for $n \geq 4$, leading to total energies that differ by as much as several kcal/mol between two implementations of what is ostensibly the same method. This behavior can be understood based on a propagation-of-errors analysis applied to a closed-form expression for the *n*-body expansion, which is derived here for the first time. Discrepancies between the two implementations arise primarily from the Coulomb self-energy correction that is required when electrostatic embedding charges are implemented by means of an external driver program. For reliable results in large systems, our analysis suggests that script- or driver-based implementations should read binary output files from an electronic structure program, in full double precision, or better yet be fully integrated in a way that avoids the need to compute the aforementioned self-energy. Moreover, four-body and higher-order expansions may be too sensitive to numerical thresholds to be of practical use in large systems. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4885846]

I. INTRODUCTION

The time t required to perform an *ab initio* quantum chemistry calculation depends upon both the size of the system, N_A, and the number of basis functions per atom, N_B:

$$t \sim \mathcal{O}\left(N_A^a N_B^b\right). \tag{1.1}$$

Typically, $a \geq 3$ and $2 \leq b \leq 4$. To defeat this highly nonlinear scaling, and thus extend quantum chemistry to large systems, further approximations are required. Fragment-based quantum chemistry methods are an increasingly popular approach in this respect.[1,2] Such methods decompose the (super)system, whose size is N_A, into N_F separate subsystems, the typical size of which we denote as $n_A \ll N_A$. The formal scaling of the total aggregate computer time required is thereby reduced to

$$t \sim N_F \times \mathcal{O}\left(n_A^a N_B^b\right). \tag{1.2}$$

Wall times can be dramatically reduced via coarse-grained parallelization across the N_F fragments. As has been pointed out previously,[3–6] the explanation for why this works is ultimately rooted in the "nearsightedness of electronic matter".[7,8]

Among the litany of fragment-based methods in the literature, we focus here on the many-body expansion (MBE) as applied to non-covalent clusters, which amounts to a decomposition of the cluster (supersystem) energy into a sum of monomer energies, plus corrections for pairwise interactions, three-body interactions, etc. This is the simplest example of a fragment-based method, both conceptually and computationally, and it has the appealing characteristic that the expansion is formally exact if carried to *N*-body interactions for an *N*-body cluster, at least if all calculations are carried out to arbitrary precision. As such, it is often considered that *n*-body expansions (wherein the MBE is truncated at *n*-body interactions, for some $n < N$) represent a sequence of converging approximations to the exact energy. Results presented herein will call that assumption into question, at least when the electronic structure calculations and the MBE are implemented using double-precision arithmetic.

Given the dramatic reduction in wall time that is engendered by the embarrassingly parallelizable nature of the subsystem calculations, such an approach appears to represent a metaphorical "free lunch," in the sense that high-level *ab initio* results might be obtained for large systems, with minimal loss of accuracy and for a fraction of the wall time cost. (Often the total aggregate computer time is significantly reduced as well.[9,10]) These ideas have been applied, for example, to perform MP2 calculations on a system of >24 000 atoms.[11] There is no supersystem MP2 benchmark for a

[a] Present address: School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA.
[b] Author to whom correspondence should be addressed. Electronic mail: herbert@chemistry.ohio-state.edu

system of 24000 atoms, and indeed there have been very few benchmark validation studies of n-body expansions in large systems ($\gtrsim$50–100 atoms, say), and even fewer at respectable levels of theory. We began this work with the intention of establishing such benchmarks, but those comparisons are *not* the purpose of the present work, because in the course of that effort we uncovered serious numerical problems with the MBE as applied to large systems. The present paper describes issues related to loss of precision that stems from the factorial growth in the number of subsystem calculations, as a function of both system size (N) as well as truncation order (n).

II. THEORY

Most fragment-based methods are based at some level on the n-body expansion.[1] If we partition the system into N disjoint fragments (monomers, for the purposes of this work), then the MBE amounts to the following expression for the supersystem energy:

$$E = \sum_I^N E_I + \sum_{I<J}^N \Delta E_{IJ} + \sum_{I<J<K}^N \Delta E_{IJK} + \cdots . \tag{2.1}$$

The quantity E_I represents the energy of monomer I and the quantities $\Delta E_{IJ\ldots}$ are the corrections for n-body interactions, obtained for each $n > 1$ by computing the energy $E_{IJ\ldots}$ of the appropriate n-mer and subtracting all lower-order corrections involving the same monomer units. The first few such corrections are

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J, \tag{2.2a}$$

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK} - E_I - E_J - E_K, \tag{2.2b}$$

$$\begin{aligned}\Delta E_{IJKL} = E_{IJKL} &- \Delta E_{IJK} - \Delta E_{IKL} - \Delta E_{IJL} \\ &- \Delta E_{JKL} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{IL} \\ &- \Delta E_{JK} - \Delta E_{JL} - \Delta E_{KL} \\ &- E_I - E_J - E_K - E_L.\end{aligned} \tag{2.2c}$$

Because each new order in the expansion subtracts out all lower-order terms, Eq. (2.1) is formally exact if carried through order $n = N$, where it just becomes $E = E_{IJK\cdots N}$. In practice, this expansion is always truncated at some $n \ll N$, with the idea that corrections for the mutual interaction of n monomers—that is, genuine n-body interactions that are not captured at lower orders—should become negligible for sufficiently large n. Combined with the formally exact nature of Eq. (2.1), this suggests a sequence of convergent approximations that has led to a widespread belief that the MBE provides a route to arbitrary accuracy, simply by increasing the truncation order, n. For molecular clusters—including clusters composed of polar monomers where many-body induction effects are important—it does generally seem to be the case that the magnitude of the n-body interactions decreases with increasing n,[12–17] although counterexamples exist for molecular polarizabilities,[18] for which an expansion analogous to Eq. (2.1) can be formulated by taking the appropriate derivatives, term-by-term. The situation seems to be somewhat worse in atomic clusters, where the MBE is slowly convergent and oscillatory.[14,19,20] As such, we focus on noncovalent molecular clusters, for which the MBE has seen a resurgence in recent years, both in its own right,[9,10,15–17,21–25] and as the underpinning of the fragment molecular orbital (FMO) method.[1,26,27]

In any case, after deciding upon a truncation order, the n-body approximation to the supersystem energy, $E^{(n)}$, can be manipulated into a minimal closed form. To the best of our knowledge, these closed-form expressions have been published only through $n = 4$:[19]

$$E^{(1)} = \sum_I E_I, \tag{2.3a}$$

$$E^{(2)} = \sum_{I<J} E_{IJ} - (N-2)\sum_I E_I, \tag{2.3b}$$

$$E^{(3)} = \sum_{I<J<K} E_{IJK} - (N-3)\sum_{I<J} E_{IJ} + \tfrac{1}{2}(N-2)(N-3)\sum_I E_I, \tag{2.3c}$$

$$\begin{aligned}E^{(4)} = &\sum_{I<J<K<L} E_{IJKL} - (N-4)\sum_{I<J<K} E_{IJK} \\ &+ \tfrac{1}{2}(N-3)(N-4)\sum_{I<J} E_{IJ} \\ &- \tfrac{1}{6}(N-2)(N-3)(N-4)\sum_I E_I.\end{aligned} \tag{2.3d}$$

In these expressions, the quantity $E_{IJ\ldots n}$ is the energy of the n-mer constructed from the union of monomers I, J, …, n. The four particular cases in Eq. (2.3) are derived in Appendix A, in order to motivate the derivation of more general formulas that appear below.

There are a few examples in the literature where the MBE has been extended beyond four-body terms,[12,14,18–20] with terms up to $n = 7$ considered in Ref. 14 and up to $n = 8$ (for polarizabilities rather than energies) in Ref. 18. In these cases, recourse is usually made to an alternative, recursive formula:[19]

$$E^{(n)} = \sum_{K=1}^{\binom{N}{n}} E_K^{(n)} - \sum_{m=1}^{n-1}\left[\frac{(N-m)!}{(N-n)!(n-m)!}\right] E^{(m)}. \tag{2.4}$$

Here

$$\binom{N}{n} \equiv {}_N C_n = \frac{N}{n!(N-n)!} \tag{2.5}$$

is a binomial coefficient and $E_I^{(n)}$ is the energy of the Ith n-body sub-cluster or "n-mer". Equation (2.4) expresses the n-body approximation to the energy in terms of the n-mer energies (the first term) as well as the lower-order approximations, $E^{(m)}$ ($m < n$).

The recursive nature of Eq. (2.4) is cumbersome, especially in the context of the propagation-of-errors analysis that is presented later in this work, for which it is much more

convenient to express $E^{(n)}$ explicitly in terms of monomer, dimer, trimer, ..., n-mer energies, as in Eq. (2.3). A clear pattern is emerging in that equation, and in Appendix B, we derive the general form of an n-body interaction. Then, in Appendix C, we use this result to show that truncation of the expansion at order n affords the following closed-form expression:

$$E^{(n)} = \sum_{m=0}^{n-1} (-1)^m \binom{N-n-1+m}{m} \sum_{K=1}^{\binom{N}{n-m}} E_K^{(n-m)}. \quad (2.6)$$

This is the arbitrary-order generalization of Eq. (2.3). To the best of our knowledge, this closed-form expression has not been reported previously.

III. COMPUTATIONAL METHODS

Two different computer implementations of the n-body expansion are examined and compared in this work, one that is fully integrated into the Q-CHEM electronic structure program,[28] and another code (FRAGME∩T[5,29]) that we have written as a an independent "driver" to generate and execute Q-CHEM input files, then read results from Q-CHEM's binary scratch files and finally compute $E^{(n)}$. As we will see, the combinatorial nature of the n-body expansion, which leads to factorial growth in the number of terms, with respect to both N and n, makes it highly non-trivial to obtain agreement between these two implementations to a level that would normally be acceptable when comparing different implementations of the same electronic structure algorithm. Ultimately, we will obtain acceptable agreement, but the process of doing so uncovers several issues of paramount importance concerning the use of the MBE for large systems, particularly in regard to the precision and reproducibility of the results.

In what follows, we apply the MBE to a sequence of water clusters ranging in size from $(H_2O)_6$ to $(H_2O)_{47}$, the geometries for which are available in Ref. 30. Each represents the putative global minimum on the TIP4P[31] potential energy surface, as determined by extensive basin-hopping Monte Carlo simulations.[30] All of the calculations reported here are performed at the B3LYP/cc-pVDZ level, using the SG-1 quadrature grid[32] unless otherwise stated. (Higher-quality integration grids are considered in Sec. IV D.) This represents a very modest level of electronic structure theory, but is similar in quality to many other large-scale applications of the MBE, particularly in the context of the FMO approach. FMO calculations on water clusters as large as $(H_2O)_{64}$ have been reported at the Hartree-Fock and B3LYP levels, in small basis sets such as STO-3G, 3-21G, 6-31G*, and 6-31++G**,[33,34] and both two- and three-body expansions for $(H_2O)_{26}$ have been reported at the density-functional level in various small basis sets.[35] In terms of the number of fragments, these studies represent some of the largest benchmark calculations reported to date, but few systematic studies (as a function of either N or n) are available.

A major focus of the present work is the reproducibility (or lack thereof) of n-body energies between the aforementioned two implementations of the MBE, each of which uses the same electronic structure program at its core. Except where noted, the self-consistent field (SCF) convergence criterion is set to $\tau_{\mathrm{SCF}} = 10^{-5}$ a.u., and $\tau_{\mathrm{ints}} = 10^{-9}$ a.u. is used for the integral screening and shell-pair formation threshold. (This is true, in particular, for the calculations presented in Sec. IV A, but in Sec. IV D we examine the effects of systematically tightening both thresholds.) We use fragments consisting of a single H_2O molecule, and use Eq. (2.6) to compute two-body (2B), three-body (3B), and four-body (4B) energies for the aforementioned sequence of water clusters, both with and without electrostatic embedding (EE).

It will emerge that additional complications arise in the case of using EE, so it is worth describing this procedure in some detail. The idea is to accelerate convergence of the MBE by embedding the n-body calculations in some classical representation of the electrostatic potential due to the rest of the system. Often, this embedding consists simply of atom-centered point charges on those monomer units not included in the subsystem electronic structure calculation.[15] In the present work, we use Mulliken charges calculated independently for each monomer unit, at the B3LYP/cc-pVDZ level, for the geometry that the monomer exhibits in the cluster. Arguably, Mulliken charges may not be the best choice for high accuracy, but they actually work surprisingly well in many cases,[15,22,29,36] including for water clusters, and serve here as a representative example of point charges that are derived from the monomer wave functions.

In a script- or driver-based implementation of the electrostatically embedded n-body expansion[15] (EE-nB in our notation), point charges must be added to the electronic structure calculation via the input file. In Q-CHEM and other electronic structure programs, the Coulomb self-interaction of these "external" point charges is (quite sensibly) added into the final SCF energy, but this is undesirable in the present context because the embedding charges are only supposed to polarize the n-mer electronic structure calculations, and should not contribute directly to the energy. Therefore, self-interaction of the embedding charges must be removed, which has dramatic consequences that we shall document.

For an n-body expansion, Li *et al.*[36] have shown that the self-energy $E_{\mathrm{self}}^{(n)}$ that must be subtracted from the total energy, $E^{(n)}$, can be written as the total Coulomb interaction between the embedding charges for all N monomers, E_{Coul}, multiplied by a combinatorial coefficient that counts the number of subsystems of each size. That coefficient is easy to rationalize in light of Eq. (2.6), and the ratio $E_{\mathrm{self}}^{(n)}/E_{\mathrm{Coul}}$ is strictly combinatorial:[36]

$$\frac{E_{\mathrm{self}}^{(n)}}{E_{\mathrm{Coul}}} = -1 + \sum_{m=0}^{n-1} (-1)^m \binom{N}{n-m} \binom{N-n-1+m}{m}. \quad (3.1)$$

This formula has previously been used to remove the Coulomb self-interaction of the embedding charges.[29,36–38] Our FRAGME∩T-based implementation of the n-body expansion computes $E_{\mathrm{self}}^{(n)}$ according to Eq. (3.1) and then subtracts it from $E^{(n)}$ in Eq. (2.6) to obtain the total energy.

IV. RESULTS AND DISCUSSION

A. Reproducibility of *n*-body approximations

In this section, we focus on whether full numerical agreement can be obtained between the Q-CHEM and FRAGME∩T implementations of the MBE. We begin our analysis by computing binding energies of $(H_2O)_N$ clusters, $N = 6$–47, using two- and three-body expansions; signed errors in the binding energies, relative to the supersystem benchmark at the same level of theory, are shown in Fig. 1(a). The data are plotted in this way in order to highlight differences between the two implementations of the MBE, on a scale relevant to the accuracy of either approximation. Plots of the energy difference between the two implementations are available in the supplementary material (see Fig. S1),[39] and indicate that the two implementations agree to within 0.5 kcal/mol for all of these calculations. Thus, the two implementations are indistinguishable on the scale set by the overall errors depicted in Fig. 1(a). These plots serve as a "control experiment" to demonstrate that the two implementations are working and results are reproducible between them.

Figure 1(b) presents an analogous plot using electrostatic embedding for the two- and three-body approximations. (Numerical differences between the two implementations are plotted explicitly in Fig. S2 of the supplementary material.[39]) For $N < 30$, the agreement between the two implementations is quite good, and perhaps unremarkable, although the magnitude of the errors is noteworthy. In much smaller water clusters, the EE-3B method has been shown to afford quite good accuracy,[9,10,15,40,41] whereas for $N > 10$ we see that even this approach affords errors of $\gtrsim 5$ kcal/mol, and for the largest cluster considered here, $(H_2O)_{47}$, errors exceed 20 kcal/mol. As shown below, a significant part of this error arises from neglect of four-body terms, inclusion of which reduces the error to a still-sizable 6 kcal/mol.

Of foremost importance at present is the fact that there is a marked difference between the EE-3B results obtained using the two different implementations of this method. The two implementations differ noticeably starting around $N = 30$, and by $N = 40$ their predictions are dramatically different. To understand why this might be, one must note that the self-energy correction $E^{(n)}_{\rm self}$ can be extremely large when N and/or n is large, e.g., $E^{(4)}_{\rm self} \sim 10^5$ hartree for $(H_2O)_{40}$ described at the four-body level. As such, very small differences can be magnified factorially. As an example, we introduced an error of 10^{-7} a.u. in a Bohr-to-Ångstrom conversion factor that is used to compute $E_{\rm Coul}$ and therefore $E^{(n)}_{\rm self}$ in Eq. (3.1). Figure 2 shows the extent to which this ostensibly small modification changes the EE-*n*B energy. The discrepancy grows sharply as a function of N, exceeding 1 kcal/mol for $(H_2O)_{30}$ at the four-body level. Note also how the sign of the discrepancy oscillates as a function of n, as a consequence of the $(-1)^m$ factor in Eq. (3.1) and the fact that higher-order terms in the n-body expansion carry significant combinatorial coefficients.

A fully integrated implementation of the MBE, in contrast, can avoid calculating the self-energy in the first place, a luxury not presently available to most practitioners because the n-body expansion is not yet widely available in traditional electronic structure program packages. (A pilot implementation is available in Q-CHEM v. 4.2,[42] however.) Even if this method were more widely available, there is an undeniable practical appeal in writing a simple driver program, and thus the consequences of that choice warrant investigation. Section IV B aims to identify the origin of the EE-3B differences between Q-CHEM and FRAGME∩T implementations.

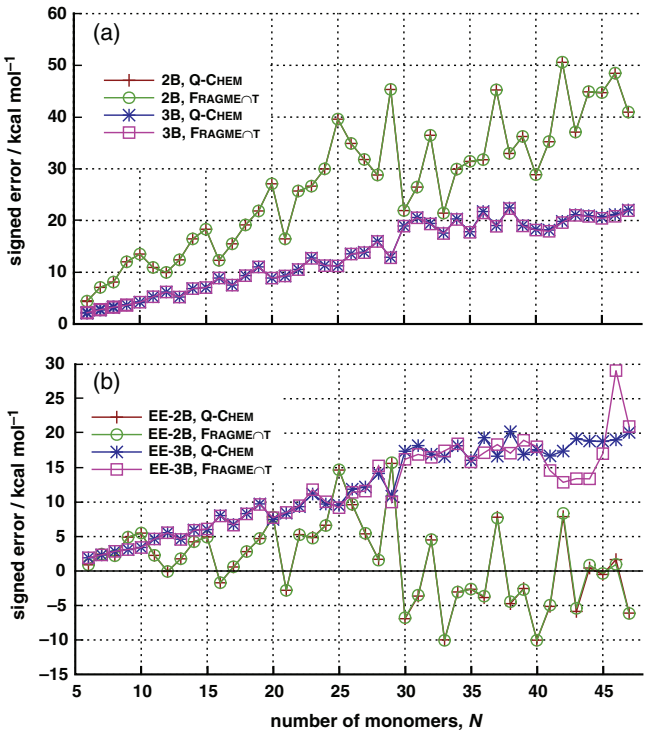

FIG. 1. Signed errors in the total binding energies of water clusters, for (a) two- and three-body approximations and (b) electrostatically embedded 2B and 3B expansions, comparing two different computer implementations of these methods. Errors are measured with respect to a supersystem calculation at the same level of theory (B3LYP/cc-pVDZ). No embedding charges are used in (a), whereas in (b) the embedding consists of non-iterative Mulliken point charges, the values of which were rounded to six decimal places (in a.u.) in calculations using the FRAGME∩T code.

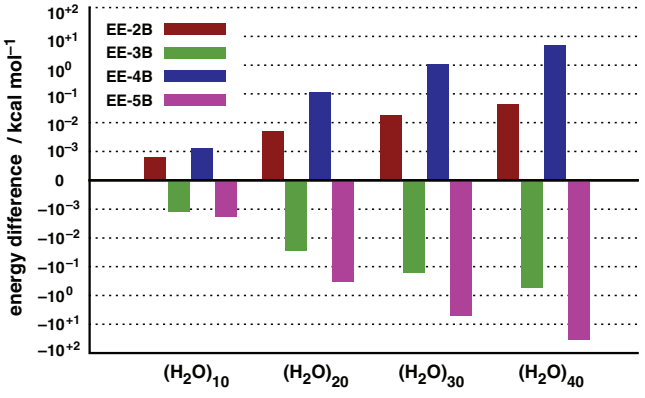

FIG. 2. Energetic discrepancies (on a logarithmic scale) between two different calculations of the FRAGME∩T-based EE-*n*B energy of several water clusters, using two slightly different values of the Bohrs-to-Ångstrom conversion factor in the calculation of $E^{(n)}_{\rm self}$. The two values differ by only $\sim 10^{-7}$ a.u. Electronic structure calculations were performed at the B3LYP/cc-pVDZ level with Mulliken embedding charges, and with $\tau_{\rm SCF} = 10^{-5}$ a.u. and $\tau_{\rm ints} = 10^{-9}$ a.u.

TABLE I. Energetic quantities needed to compute the supersystem energy of $(H_2O)_{42}$, using two different implementations of the EE-3B method. All quantities are computed at the B3LYP/cc-pVDZ level. Subscripted data values are non-significant figures assuming an uncertainty of 10^{-5} hartree (equal to the SCF convergence criterion) in each individual energy calculation.

	Value/hartree	
Quantity	Q-CHEM	FRAGME∩T[a]
$\sum_{IJK} E^{(1)}_{IJK}$	$-2\,632\,377.50025_{49\,854}$	$-2\,646\,723.38546_{20\,224}$
$\sum_{IJ} E^{(1)}_{IJ}$	$-131\,618.29752_{41\,274}$	$-132\,723.52067_{50\,473}$
$(N-3)\sum_{IJ} E^{(1)}_{IJ}$	$-5\,133\,113.6034_{4097}$	$-5176217.3063_{26\,845}$
$\sum_{I} E^{(1)}_{I}$	$-3210.18827_{50\,287}$	$-3265.53410_{72\,462}$
$\frac{1}{2}(N-3)(N-2)\sum_{I} E^{(1)}_{I}$	$-2\,503\,946.85_{4\,522\,386}$	$-2\,547\,116.60_{3\,652\,036}$
$E^{(3)}$	$-3210.75_{13\,364\,022\,598}$	$-17\,622.68_{278\,721\,394}$
E_{Coul}	0.0	-1.35196_{265}
$E^{(3)*}_{\text{Coul}}$ [b]	0.0	-1.35196_{355}
$E^{(3)} - E^{(3)}_{\text{self}}$	$-3210.75_{13\,364\,022\,598}$	$-3210.76_{09\,382\,139\,417}$
$E^{(3)} - E^{(3)*}_{\text{self}}$ [b]	$-3210.75_{13\,364\,022\,598}$	$-3210.75_{13\,442\,139\,407}$

[a]Includes the Coulomb self-energy of the embedding charges.
[b]Value needed to make FRAGME∩T agree with Q-CHEM.

B. Sensitivity to precision

Results in Sec. IV A suggest that the primary difference between the Q-CHEM and FRAGME∩T implementations of the MBE is how electrostatic embedding is treated. As such, we next consider in detail the energetic quantities required for such a calculation, focusing on $(H_2O)_{42}$, which is arguably the smallest cluster in which the difference is appreciable (see Fig. 1(b)). These values are shown in Table I. Bear in mind that energies obtained from the FRAGME∩T implementation contain the point-charge self-interaction, whereas those obtained directly from Q-CHEM do not. Given the excellent agreement between the two implementations for the non-embedded case [Fig. 1(a)], it is reasonable to assume that the disagreement occurs as a result of the self-interaction correction; therefore, Table I also lists the "theoretical" Coulomb interaction value,

$$E^{(3)*}_{\text{Coul}} = \frac{E^{(3)}(\text{FRAGME}\cap\text{T}) - E^{(3)}(\text{Q-CHEM})}{\binom{N}{3} - (N-3)\binom{N}{2} + N\binom{N-2}{2} - 1}, \tag{4.1}$$

defined as the value that exactly cancels the theoretical self-interaction, $E^{(3)*}_{\text{self}}$, which is itself defined as the difference between the Q-CHEM and FRAGME∩T EE-3B binding energies.

Simply comparing E_{Coul} and $E^{(3)*}_{\text{Coul}}$, the 6 kcal/mol difference between the Q-CHEM and FRAGME∩T EE-3B binding energies appears to arise from disagreement in the seventh decimal place (in hartree) of E_{Coul}. This implies that in the process of correcting for the self-energy of the point charges, we have lost approximately four digits of precision. Put another way, if we want our final answer to be precise to within 1 kcal/mol, we need to know E_{Coul} to a precision of at least 10^{-7} hartree.

Although a proper propagation-of-errors analysis of these calculations is presented below in Sec. IV C, the use of significant digits is a ubiquitous shortcut for estimating precision and error propagation, and is thus worth considering here. Within Table I, we have indicated non-significant digits with subscripts, under the assumption that the SCF convergence threshold of 10^{-5} hartree sets a limit of this value on the precision of each individual energy calculation. (The uncertainty in $E^{(3)}_{\text{Coul}}$ is more difficult to determine, so we assume it is the same. However, since $E^{(3)}_{\text{Coul}}$ is a wave function property and the wave function convergences more slowly than the energy, this estimate may actually be slightly optimistic.) Regardless of how the uncertainties are obtained, the following discussion pertains to any set of energies, combined within the MBE, that have an uncertainty of 10^{-5} hartree.

Particularly due to the large prefactors on the one-body term and $E^{(3)}_{\text{self}}$, our significant-figure analysis suggests that the final energy has an uncertainty of 10^{-2} hartree or about 6 kcal/mol. A proper propagation-of-errors analysis (Sec. IV C) actually suggests that the situation is much worse. The point is that even though the individual energies may be computed to high precision, the combinatorial nature of the MBE engenders a significant loss in precision for large systems. Our estimate based on significant digits suggests that it is the loss in precision in E_{Coul} specifically that leads to the 6 kcal/mol difference in the EE-3B binding energies for $(H_2O)_{42}$.

To understand the origin of this loss in precision in the Coulomb energy, note that quite a few implementations of the MBE operate at a script or driver level, creating the necessary input files for the various subsystem calculations, and then calling the binary executable of the electronic structure program for each independent calculation.[15,29,41,43] This step opens up the possibility for precision mismatches between the driver routine and the electronic structure program, unless the driver routine prints and uses all values to the same precision that they are stored internally in the electronic structure program, presumably double precision. Other aspects of the calculations may also inadvertently be different when a driver rather than a fully integrated approach is taken. For example, our integrated implementation makes use of monomer SCF calculations to construct an initial guess for

larger subsystems, whereas our driver-based approach does not.

That said, we suspect that our FRAGME∩T code is more closely interfaced with its electronic structure code than are most driver-based implementations of the MBE, in the sense that FRAGME∩T reads Q-CHEM's binary scratch files in order to obtain the Mulliken embedding charges (and therefore calculate $E^{(n)}_{\rm self}$) in full double precision. Even so, when Q-CHEM input files are generated for the subsystem calculations, these embedding charges *must* be written to the input files in full double precision, else a mismatch occurs. In what follows, we demonstrate that just such a mismatch is the primary cause of the discrepancies between our Q-CHEM and FRAGME∩T implementations of the MBE.

Initially, we generated Q-CHEM input files (via FRAGME∩T) with embedding charges rounded to six decimal places in atomic charge units. Naïvely and *a priori*, this seemed like sufficient precision; the same rounding scheme was used in previous work,[5,9,10,29] and was also used to generate the EE-2B data in Fig. 1(b), for which no significant discrepancy between the Q-CHEM and FRAGME∩T implementations is observed. If instead all significant digits are printed to the input files (about 15 digits, in double precision[44]), then we obtain the EE-3B results plotted in Fig. 3. In this case, excellent agreement between the two implementations is obtained, with a maximum discrepancy of <0.15 kcal/mol, although it is worth noting that the EE-3B discrepancies do generally appear to be growing larger with system size, although not in a strictly monotonic way (see Fig. S3 of the supplementary material[39]).

Although the source of discrepancy for these EE-3B calculations is ultimately traceable to what is arguably an oversight (namely, failure to write the electronic structure input files in full double precision), this example serves as an important cautionary tale. The fact that intermediate rounding in the input files can lead to huge discrepancies in the final energy, while preserving six significant digits in $E_{\rm Coul}$, is a cause for concern. From this point forward, all FRAGME∩T-based

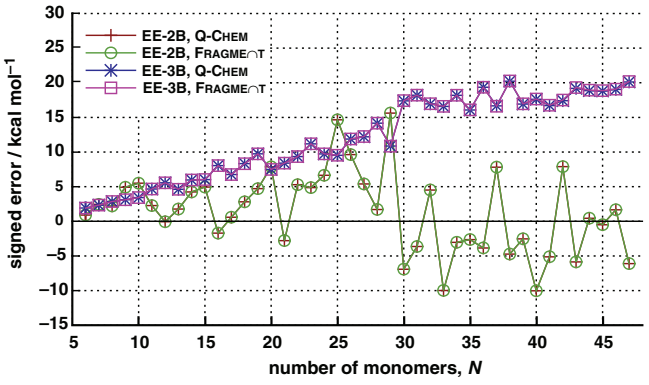

FIG. 3. Signed errors EE-2B and EE-3B approximations to the total binding energies of water clusters, comparing two different computer implementations of these methods. Errors are measured with respect to a supersystem calculation at the same level of theory (B3LYP/cc-pVDZ). Unlike the analogous plot in Fig. 1(b), in which values of the Mulliken embedding charges are rounded to six decimal digits in the FRAGME∩T implementation, here they are used in full double precision.

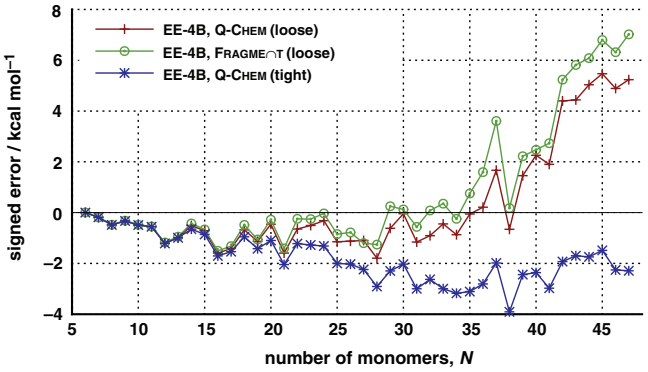

FIG. 4. Signed errors in EE-4B binding energies of $(H_2O)_N$ clusters, relative to supersystem results at the B3LYP/cc-pVDZ level. Two different sets of thresholds are employed: "loose" thresholds ($\tau_{\rm SCF} = 10^{-5}$ a.u. and $\tau_{\rm ints} = 10^{-9}$ a.u.) and "tight" thresholds ($\tau_{\rm SCF} = 10^{-6}$ a.u. and $\tau_{\rm ints} = 10^{-14}$ a.u.).

results will use all meaningful digits in an effort to maximize reproducibility.

At the EE-4B level, even this measure proves insufficient to afford complete agreement between Q-CHEM and FRAGME∩T results, as shown in Fig. 4. For "loose" thresholds of $\tau_{\rm SCF} = 10^{-5}$ a.u. and $\tau_{\rm ints} = 10^{-9}$ a.u., size-dependent errors follow similar trends for both implementations but for $N \gtrsim 30$ there are noticeable quantitative differences. These must be a function of the self-energy correction, as the two approaches use the same subsystem energies. On the other hand, tightening the thresholds to $\tau_{\rm SCF} = 10^{-9}$ a.u. and $\tau_{\rm ints} = 10^{-14}$ a.u. affords (slightly) different subsystem energies, yet very different errors with respect to a supersystem calculation at the same level of theory. We should emphasize that the benchmark value of the supersystem energy is far less affected by this change in the thresholds: e.g., for $(H_2O)_{40}$ the "tight" and "loose" values differ only by 5.5×10^{-4} hartree (=0.35 kcal/mol). This is slightly larger than the looser value of $\tau_{\rm SCF}$, as a result of the fact that $\tau_{\rm ints}$ is modified as well, but "tight" and "loose" EE-4B energies for the same system differ by >4 kcal/mol. This observation underscores the fact that the main issue is propagation of errors in an n-body calculation that requires summing a large number of terms, each multiplied by a binomial coefficient that is growing factorially with respect to both N and n. Error propagation is explored further in Sec. IV C, while the role of the numerical thresholds is examined in more detail in Sec. IV D.

C. Error propagation

The significant-figure analysis discussed above (Table I) represents an approximate way of propagating uncertainties, but more rigorous ways can be envisaged. One such method, used in a previous study of the numerical stability of electronic structure algorithms,[45] is to inject random noise into the calculation. Such an approach could be used in the context of the MBE as well, but as an alternative we will investigate precision issues by means of a propagation-of-errors (PoE) analysis.

Given a function f that depends on a set of independent variables, $\{x_i\}$, with corresponding uncertainties $\{dx_i\}$, the

uncertainty in f is defined as

$$df = \left[\sum_i \left(\frac{df}{dx_i}\right)^2 (dx_i)^2\right]^{1/2}. \tag{4.2}$$

In Sec. II, we provided a closed-form expression for $E^{(n)}$, the n-body approximation to the total energy. If we assume that the energy calculation for each individual subsystem has the same uncertainty (δE), then the uncertainty in the total energy (dE) is

$$dE = \left[\sum_{m=0}^{n-1} \sum_{I=1}^{\binom{N}{n-m}} \left(\frac{dE^{(n)}}{dE_I^{(n-m)}}\right)^2 (\delta E)^2\right]^{1/2}. \tag{4.3}$$

The derivatives appearing in this equation are trivial to evaluate since the total energy is a sum of independent subsystem energies:

$$\frac{dE^{(n)}}{dE_I^{(n-m)}} = (-1)^m \frac{(N-n-1+m)!}{m!(N-n-1)!}. \tag{4.4}$$

As a result, the total uncertainty as estimated by PoE analysis is proportional to the uncertainty δE in each subsystem calculation:

$$dE = \left[\sum_{m=0}^{n-1} \binom{N}{n-m} \left(\frac{(N-n-1+m)!}{m!(N-n-1)!}\right)^2\right]^{1/2} \delta E. \tag{4.5}$$

For an SCF calculation with a convergence threshold of $10^{-\alpha}$ hartree, it seems reasonable to assume that the $(\alpha+1)$st decimal digit in the energy is a random number, hence we anticipate $\delta E \sim 10^{-(\alpha+1)}$ hartree. Taking $\alpha = 5$ (the default SCF convergence criterion in Q-CHEM) and setting $\delta E = 10^{-6}$ hartree, we have used Eq. (4.5) to compute the uncertainty in a truncated n-body expansions as a function of the system size, N, for various values of n. The results are plotted in Fig. 5(a). Note that the PoE analysis does not depend on the identity of the monomers, but for consistency with the water cluster data presented here, we plot PoE results ranging from $N = 6$–47.

Uncertainty in the two-body results remains $\ll 1$ kcal/mol through $N = 47$, but uncertainty in the three-body results is about 4 kcal/mol for $N = 45$, and much larger for the four- and five-body treatments. This is a rather grim result. Insofar as we desire an accuracy of $\lesssim 1$ kcal/mol, which the two-body expansion is fundamentally incapable of delivering (see, e.g., Fig. 3), the precision demands of higher-order expansions will require us to compute each subsystem calculation to a precision better than 10^{-6} hartree even though we do not desire that level of precision with respect to the supersystem energy. This will increase the computational cost of each subsystem calculation.

Perhaps more significantly, these results undercut the long-standing notion that one can obtain arbitrary accuracy simply by extending the n-body expansion to higher orders. In finite-precision arithmetic, the combinatorial nature of the expansion causes errors to accumulate faster (with respect to system size) for larger n. As n increases, more precision is therefore required from the subsystem calculations in order to obtain a specified level of precision in the total energy.

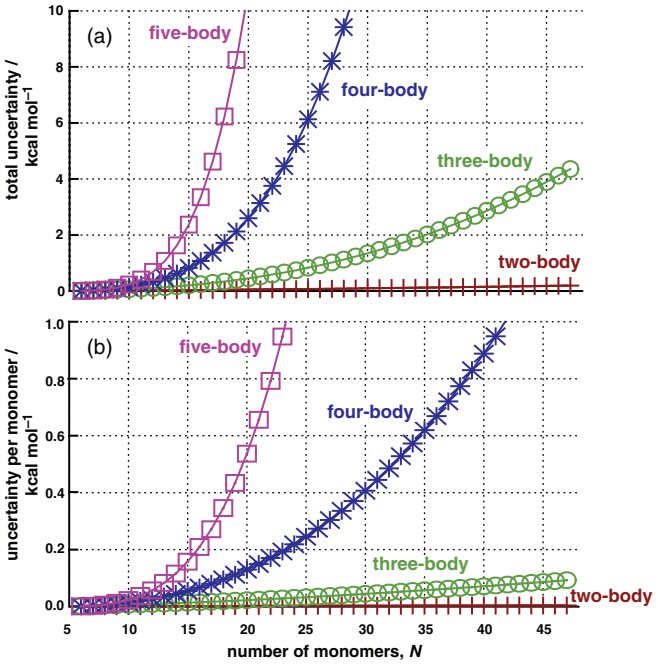

FIG. 5. (a) Total and (b) per-monomer uncertainties in the n-body approximation, $E^{(n)}$, as estimated using propagation-of-errors analysis. Each subsystem energy calculation is taken to have an uncertainty of 10^{-6} hartree.

As shown below, this situation becomes even worse when we consider error propagation in $E^{(n)}_{\text{self}}$.

Previous studies have suggested that errors engendered by the n-body expansion are size-extensive (i.e., linear functions of N), at least as $N \to \infty$.[17,29,41,46] PoE analysis, however, suggests that error accumulation due to finite precision grows nonlinearly as a function of N, and that the rate of growth increases as a function of n. Perhaps due to assumed size-extensivity, it is common to plot errors on a per-monomer basis, although it should be clear from the mathematics that linear growth in the number of monomers will not keep pace with the highly nonlinear growth in the PoE uncertainties. This is underscored by plots of the per-monomer PoE uncertainties in Fig. 5(b). Per-monomer uncertainties rise extremely rapidly at the four- and five-body levels, exceeding 1 kcal/mol/monomer at $N = 42$ and $N = 23$, respectively.

On the other hand, the per-monomer uncertainties remain small (<0.1 kcal/mol/monomer) at the three-body level, and negligible at the two-body level, all the way out to the largest systems considered here ($N = 47$). In conjunction with screening and discarding of distant interactions, as will ultimately be necessary anyway for $N \gg 50$, this may be sufficient to sidestep precision problems at the three-body level, although further testing is needed. The relative insensitivity of these results to system size, at least when measured in per-monomer units, is probably the reason why precision issues have not been carefully considered in previous work on the n-body expansion.

However, the four- and five-body results in Fig. 5 give pause, as even the per-monomer uncertainties are large in these cases, suggesting that precision problems may doom any attempt to use $n > 3$ for large systems. This observation calls into question the assumption that the n-body expansion

is systematically improvable as a function of n. Moreover, the large errors observed in *total* binding energies (e.g., in Fig. 4) suggests that these methods may not afford a uniform description of the potential energy surface. Although one might argue that total binding energies are not a useful figure of merit, insofar as they are probably not experimentally observable except for very small clusters, in another sense the total binding energy is just the relative energy between two very different points on the potential energy surface. As such, the total binding energy provides a measure of how well relative energies are described across the global potential surface.

Up to this point, our PoE analysis has not included the effects of electrostatic embedding. To understand the consequences of embedding, we performed a separate PoE analysis on the $E^{(n)}_{\rm self}$ term, rather than the total energy. The result is

$$dE_{\rm self} = \left| \sum_{m=0}^{n-1} (-1)^m \binom{N}{n-m} \binom{N-n+1+m}{m} \right| \delta E_{\rm Coul}, \tag{4.6}$$

where $\delta E_{\rm Coul}$ is the uncertainty in $E_{\rm Coul}$, the mutual Coulomb interaction of all embedding charges. In Sec. IV B, we showed that the disagreement between FRAGME∩T and Q-CHEM was on the order of 10^{-6} hartree for $E^{(n)}_{\rm Coul}$, so we take $\delta E_{\rm Coul} = 10^{-6}$ hartree and plot the uncertainty in $E^{(n)}_{\rm self}$ in Fig. 6(a). Combining this uncertainty with a finite precision of $\delta E = 10^{-6}$ hartree in the subsystem energies, as above, the total uncertainty in the EE-nB energy is plotted in Fig. 6(b).

As we have already essentially written off the four- and five-body approaches, we focus on EE-3B results. In this case, the total uncertainty rises about twice as quickly as it does for the non-embedded 3B expansion [compare Fig. 5(a) to Fig. 6(b)], reaching almost 10 kcal/mol for $N = 47$. Although this is only about 0.2 kcal/mol/monomer, it does raise questions about the reproducibility of EE-3B results for large systems. That said, PoE analysis does represent something of a worst-case scenario with regard to error cancellation, so we next wish to return to actual electronic structure calculations and investigate the performance of the n-body expansion as a function of various numerical thresholds.

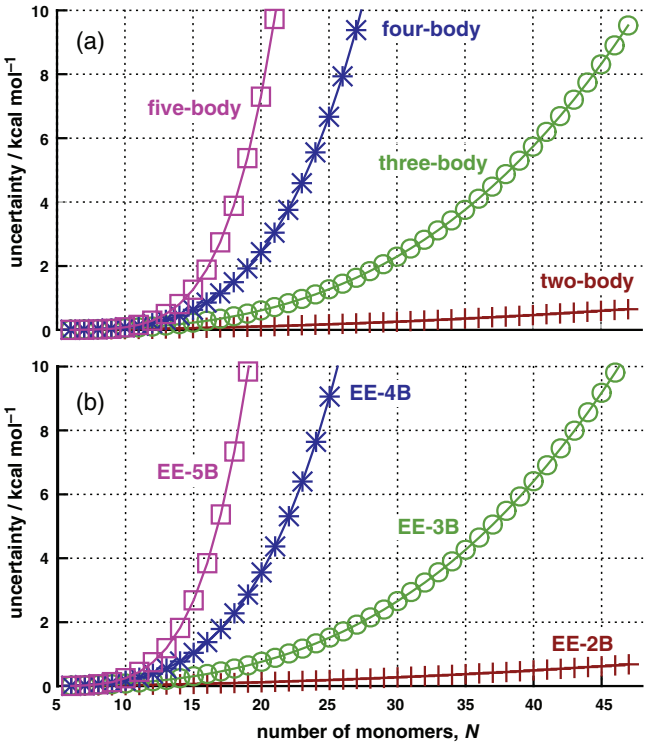

FIG. 6. Total uncertainty in (a) $E^{(n)}_{\rm self}$ and (b) $E^{(n)}$ for EE-nB approximations, assuming an uncertainty of 10^{-6} hartree in $E_{\rm Coul}$. In (b), we also assume an uncertainty of 10^{-6} hartree for each subsystem energy calculation, consistent with the PoE analysis of non-embedded n-body expansions in Fig. 5(a).

D. Numerical thresholds

In the PoE analysis presented above, we assumed an uncertainty of $\delta E = 10^{-6}$ hartree in each subsystem energy calculation, corresponding to an SCF convergence threshold $\tau_{\rm SCF} = 10^{-5}$ hartree. We next examine the effect of systematically tightening this threshold and/or the integral screening threshold, $\tau_{\rm ints}$.

Table II lists errors in EE-nB calculations on $(H_2O)_{40}$ as a function of both thresholds, computed at the B3LYP/cc-pVDZ level. EE-2B results are essentially insensitive to the value of either threshold, at least within the range of thresholds that might reasonably be used in a typical SCF calculation, whereas higher-order expansions are more sensitive.

TABLE II. Errors in EE-nB calculations on $(H_2O)_{40}$ as a function of the thresholds $\tau_{\rm ints}$ and $\tau_{\rm SCF}$, computed at the B3LYP/cc-pVDZ level and compared to a supersystem benchmark computed using $\tau_{\rm ints} = 10^{-14}$ a.u. and $\tau_{\rm SCF} = 10^{-9}$ a.u. Mulliken embedding charges are used in all EE-nB calculations.

$-\log_{10}(\tau/\text{a.u.})$		Error/kcal mol^{-1}		
$\tau_{\rm ints}$	$\tau_{\rm SCF}$	EE-2B	EE-3B	EE-4B
9	5	−10.37	17.34	1.91
10	5	−10.35	18.19	−1.64
11	5	−10.37	18.40	−1.16
12	5	−10.33	18.54	−1.32
13	5	−10.35	18.75	−1.91
14	5	−10.35	18.89	−2.42
9	6	−10.38	17.30	1.62
10	6	−10.35	18.15	−1.93
11	6	−10.37	18.36	−1.45
12	6	−10.33	18.50	−1.61
13	6	−10.35	18.71	−2.20
14	6	−10.36	18.85	−2.71
9	7	−10.38	17.30	1.63
10	7	−10.35	18.14	−1.92
11	7	−10.37	18.36	−1.44
12	7	−10.33	18.50	−1.60
13	7	−10.35	18.71	−2.19
14	7	−10.36	18.85	−2.70
9	8	−10.38	17.30	1.63
10	8	−10.35	18.14	−1.92
11	8	−10.37	18.36	−1.44
12	8	−10.33	18.50	−1.60
13	8	−10.35	18.71	−2.19
14	8	−10.36	18.85	−2.70

Unfortunately, errors in total binding energies remain quite large at the EE-2B level, although the converged EE-2B error of −10.4 kcal/mol is <0.3 kcal/mol/monomer. It remains to be seen whether such errors are achieved consistently across the potential surface (and some reason to think they might not be, owing to differences in basis-set superposition error[10]). However, the insensitivity with regard to thresholds underscores the importance of attempts to develop accurate *two-body* methods for non-covalent clusters.[6,47,48]

For EE-3B and EE-4B calculations, the errors *are* sensitive to thresholds, although for a fixed value of τ_{ints}, any value $\tau_{\text{SCF}} \leq 10^{-6}$ a.u. appears to yield converged errors. It is not clear, however, that the errors have converged with respect to τ_{ints}, although in the EE-3B case the differences between values $\tau_{\text{ints}} \leq 10^{-11}$ a.u. are <0.5 kcal/mol, in a total error whose magnitude exceeds 18 kcal/mol (or about 0.5 kcal/mol/monomer), hence there is no reason to tighten τ_{ints} beyond this point. It is also worth noting that the EE-3B error is actually *larger* than that obtained at the EE-2B level.

Similar convergence of the EE-*n*B errors, with respect to the thresholds τ_{SCF} and τ_{ints}, is observed when TIP3P embedding charges are substituted for Mulliken charges; see Table S1 of the supplementary material.[39] However, in the TIP3P case the errors are more like ∼2 kcal/mol/monomer at the EE-2B level.

At the EE-4B level, noticeable variation in the total binding energy persists even as $\tau_{\text{ints}} \to 10^{-14}$ a.u., which is a hard-coded lower limit in the release version of Q-CHEM. Modifying the code to remove this limitation, we have performed a few calculations with $\tau_{\text{ints}} = 10^{-15}$ a.u. (see Table S3 in the supplementary material[39]), but convergence is still not obtained and there is no justification to decrease τ_{ints} any further, as there are fewer than 16 digits of decimal precision available in double-precision arithmetic.[44] Detailed analysis of the thresholding reveals that the subsystem calculations are too small to be impacted by Cauchy-Schwarz screening,[49] as we have verified by setting the Schwarz screening threshold to 10^{-15} a.u., leading to negligible differences in the *n*-body results. Rather, the role of τ_{ints} in these calculations manifests at the level of forming and screening the shell pairs (see Table S3).

Note from Fig. 4 that the energy differences between tight and loose thresholds at the EE-4B level do not manifest in any significant way until $N > 20$, which emphasizes the importance of large-system benchmarks. However, based on the data in Tables II and S3, we are forced to recommend a value of $\tau_{\text{ints}} = 10^{-15}$ a.u. at the EE-4B level, a choice that will prove extremely costly if a correlated wave function method is used for the subsystem calculations. This represents another way in which four-body calculations for large systems may not be feasible in practice.

For the DFT calculations pursued here, a separate thresholding issue is the choice of DFT quadrature grid. We therefore repeated the B3LYP/cc-pVDZ calculations on $(H_2O)_{40}$ at the EE-4B level, since the results presented above suggest that $n = 4$ is high enough in the *n*-body expansion to see various precision problems emerge, whereas many of these problems are greatly suppressed for $n \leq 3$. Table III shows the EE-4B errors with respect to a supersystem benchmark, using tight thresholds of $\tau_{\text{ints}} = 10^{-14}$ a.u. and $\tau_{\text{SCF}} = 10^{-9}$ a.u. (These calculations use TIP3P embedding charges, where the EE-4B errors are slightly smaller than those obtained using Mulliken charges, although the EE-2B errors are much larger; compare Tables S1 and S2 in the supplementary material.[39]) The quadrature grids are defined in terms of a certain number, N_{rad}, of radial shells per atom, with each radial shell consisting of N_{ang} angular (Lebedev) grid points.[50] Errors with respect to a supersystem calculation span a narrow range from 0.3–0.5 kcal/mol (see Table III), even as the quality of the grid is changed significantly from $(N_{\text{rad}}, N_{\text{ang}}) = (50, 194)$ up to (100, 302).

TABLE III. Errors (in kcal/mol, with respect to a supersystem calculation at the same level of theory) for EE-4B calculations of $(H_2O)_{40}$, computed at the B3LYP/cc-pVDZ level using TIP3P embedding charges with various DFT integration grids. All calculations (including the supersystem benchmark) were performed with $\tau_{\text{ints}} = 10^{-14}$ a.u. and $\tau_{\text{SCF}} = 10^{-9}$ a.u.

	N_{rad}		
N_{ang}	50	75	100
194	0.49	0.45	0.33
230	0.50	0.44	0.32
266	0.51	0.44	0.32
302	0.48	0.41	0.29

For comparison, the SG-1 grid that is used in all other calculations affords an EE-4B error of 0.53 kcal/mol for $(H_2O)_{40}$ with TIP3P embedding charges (see Table S1). The SG-1 grid is constructed starting from $(N_{\text{rad}}, N_{\text{ang}}) = (50, 194)$ but then "pruned" by using a smaller Lebedev grid for some of the radial shells.[32] The fact that the SG-1 error is not significantly larger than the errors obtained with higher-quality grids suggests that the quality of the integration grid is not a major contributor to the precision issues raised in this work. Furthermore, Table S4 in the supplementary material[39] shows that significant size-dependent errors remain at the Hartree-Fock level, where there is no quadrature grid.

E. Floating-point errors

Both Q-CHEM and FRAGME∩T use double-precision floating point arithmetic, and the analysis above suggests that the possibility of floating-point round-off errors needs to be examined, though this turns out not to be a serious issue. To wit, if we take the subsystem energies $E_I^{(n-m)}$ that are needed in Eq. (2.6) as computed by Q-CHEM in double-precision arithmetic, but then compute the binomial coefficients and perform the summation in Eq. (2.6) using octuple-precision arithmetic (≈70 decimal digits of precision[44]), the difference is negligible as compared to a calculation in which these operations are performed in double precision.

To understand this lack of sensitivity, let us consider a simple model. Assume that we compute each subsystem energy to a precision of $10^{-\alpha}$ hartree. Summing the subsystem energies results in a value that, if stored as a floating-point value of precision p (in decimal), has at most p significant digits. In order for the αth decimal place to be stored accurately in memory, we can have at most $p - \alpha$ digits in the mantissa

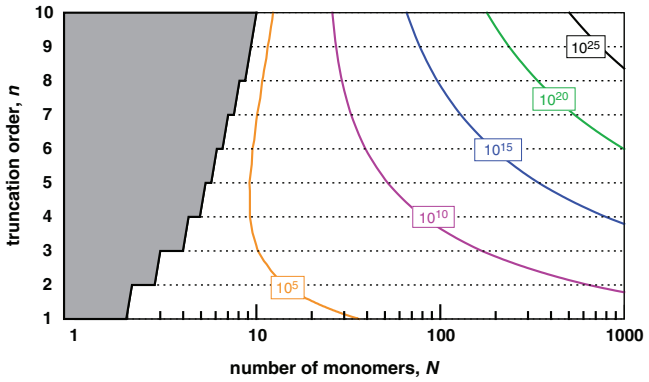

FIG. 7. Contour plot of the total absolute energies (in hartree) of all n-body sub-clusters of a $(H_2O)_N$ cluster, taking the n-body energies to be equal to n times the H_2O monomer energy for illustrative purposes. (The region shaded in gray corresponds to nonsensical combinations of N and n, and should be ignored.) Labeled contours delineate where the estimated sum of n-mer energies equals 10^{+5}, 10^{+10}, 10^{+15}, 10^{+20}, and 10^{+25} hartree.

before one of the α digits will be lost. In other words, the sum of the energies can have at most p significant figures, including the α decimal place. Considering for simplicity only the sum of the n-mer energies (since these are the most numerous subsystems in an n-body expansion), this sum may start to lose precision once

$$p - \alpha \lesssim \log \left| \sum_{I=1}^{\binom{N}{n}} E_I^{(n)} \right|. \tag{4.7}$$

Let us next assume, for simplicity, that the energy of each n-body subsystem of $(H_2O)_N$ is simply n times the energy of a single water monomer. This allows for an order-of-magnitude estimate of the right side of Eq. (4.7), and a contour plot of this approximate energy, as a function of n and N, is presented in Fig. 7. Our default SCF convergence criterion corresponds to $\alpha = 5$, implying that the energetic sum in Eq. (4.7) cannot exceed $\sim 10^{10}$ hartree if we want to preserve five decimal digits (in hartree) of precision. Conversely, for $\alpha = 5$, any combination of n and N that lies below the 10^{+10} hartree contour in Fig. 7 exists in a regime where we *can* store the total energy as a double-precision value, yet preserve five decimal digits of precision (in hartree). Specifically, this simple model suggests that three-, four-, and five-body expansions cannot be extended beyond $N = 150$, $N = 90$, and $N = 50$, respectively, without incurring loss of precision due to round-off error. All of these bounds are greater than the largest system size considered here, explaining the agreement between double- and octuple-precision implementations of Eq. (2.6). Furthermore, according to this argument all combinations of n and N that are shown in Fig. 7 are "safe" if the summation is performed in quadruple precision, for which one can expect almost 34 decimal digits of precision.[44]

V. CONCLUSIONS

The quantum chemistry community has long prided itself on the ability to reproduce the results of a well-defined theoretical model, to extremely high precision, from one software package to the next. Results presented herein suggest that this may be especially difficult in the case of n-body expansion methods, and requires particularly careful attention to issues such as numerical thresholds. Comparing a driver-based implementation of the n-body expansion, which is the most common way to implement this method, to an implementation that is fully integrated within a quantum chemistry program, reveals discrepancies between the two methods, starting at the three-body level, specifically when electrostatic embedding point charges are used in an effort to accelerate convergence of the expansion. These issues are absent at the two-body level, at least for the system sizes consider here ($N \leq 47$).

These discrepancies are ultimately traced to the fact that the driver-based approach requires the calculation and subtraction of the (very large) Coulomb self-energy of the various embedding charges, and this component of the energy is especially susceptible to loss-of-precision problems. At the very least, the driver-based implementation must read and utilize energies and embedding charges from binary scratch files generated by the electronic structure program, in full double precision, rather than reading from a text-based output file where the values of various quantities (including in particular the embedding charges) are often rounded off. Better yet, the n-body expansion can be fully integrated into one's electronic structure code of choice, sidestepping the need to compute the self-energy.

We also find that errors in the n-body expansion (as defined with respect to a supersystem calculation at the same level of theory) can be quite sensitive to the shell-pair screening threshold. Again, the two-body method is largely immune to such problems, and at the three-body level it is a small effect, but starting at the EE-4B level it is not clear that convergence is reached even as this threshold is tightened to 10^{-15} a.u., the smallest meaningful value in a double-precision electronic structure program. The use of such a tight threshold will add significantly to the cost of the individual subsystem calculations, especially for correlated wave function calculations requiring a four-index integral transformation. This, combined with a propagation-of-errors analysis that suggests runaway precision problems at the four-body level, even for relatively modest system sizes, indicates that n-body expansions with $n > 3$ may not be viable in practice except for rather small systems.

This analysis challenges the very dogma of the n-body expansion, namely, the assumption that the supersystem energy can be reproduced to arbitrary accuracy if only the truncation order, n, is increased sufficiently. Results presented here demonstrate that this is clearly *not* the case when the electronic structure calculations are carried out in double precision, as they invariably are in general-purpose electronic structure programs. The loss-of-precision problems documented here amount to a "death by ten thousand (or more) cuts" that actually grows worse at higher orders in the expansion, since the number of subsystems grows exponentially with n. For large systems, one would in practice want to implement distance-based cutoffs designed to limit the number of subsystems,[51] and it is possible that this kind of thresholding may rectify some of the problems encountered herein, or at least delay their onset as a function of system size. Note,

however, that the largest cluster examined here, $(H_2O)_{47}$, is only about 10 Å (five or six water molecules) in diameter, so cutoffs more aggressive than 5 Å would be required to reduce the number of subsystems for this example. As such, distance-based cutoffs are unlikely to be a panacea for loss of precision, so the development of high-accuracy methods that do not require four-body terms is therefore desirable. One promising approach is to use larger, overlapping fragments,[5,6,29,36,43,52–57] in conjunction with a generalized many-body expansion.[5,6,29] Such an approach can provide accurate results at the two-body level, albeit at increased cost per subsystem calculation.

In summary, our recommendations are as follows.

- Script- or driver-based implementations of the n-body expansion need to read and write subsystem energies and (where applicable) embedding charges in full double precision. This likely requires reading binary scratch files generated by the electronic structure program, rather than simply the text-based output file.
- For electrostatically embedded n-body calculations, an implementation that is fully integrated into an electronic structure code is desirable, as it avoids the need to compute and subtract the (very large) Coulomb self-energy of the embedding charges.
- Even for systems of modest size ($N = 20$–30 monomer units), accumulated errors due to finite precision in the subsystem calculations becomes problematic at the EE-4B level, suggesting that only two- or three-body approaches are viable in practice, except for rather small systems.

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 project PAA0003. J.M.H. is an Arthur P. Sloan Foundation Fellow and a Camille Dreyfus Teacher-Scholar. J.M.H. thanks Professor Ryan Steele for discussions about integral thresholding in Q-CHEM.

APPENDIX A: TRUNCATED MBE THROUGH $n = 4$

In this appendix, we verify the 1B, 2B, 3B, and 4B approximations to the energy that are given in Eq. (2.3), starting from the general form of the MBE in Eq. (2.1). To do so, we first need to work out formulas for the trimer and tetramer corrections $\Delta E^{(n)}_{IJ\ldots}$ in terms of lower-order n-mer energies, analogous to the dimer correction formula in Eq. (2.2a). The trimer interaction term is defined by Eq. (2.2b) but upon substituting Eq. (2.2a), it can be simplified to

$$\Delta E^{(1)}_{IJK} = E^{(1)}_{IJK} - E^{(1)}_{IJ} - E^{(1)}_{IK} - E^{(1)}_{JK} + E^{(1)}_{I} + E^{(1)}_{J} + E^{(1)}_{K}. \tag{A1}$$

The tetramer is defined analogously to Eq. (2.2) and can be simplified to afford

$$\begin{aligned}\Delta E^{(1)}_{IJKL} = {} & E^{(1)}_{IJKL} - E^{(1)}_{IJK} - E^{(1)}_{IJL} - E^{(1)}_{IKL} - E^{(1)}_{JKL} \\ & + E^{(1)}_{IJ} + E^{(1)}_{IK} + E^{(1)}_{IL} + E^{(1)}_{JK} + E^{(1)}_{JL} + E^{(1)}_{KL} \\ & - E^{(1)}_{I} - E^{(1)}_{J} - E^{(1)}_{K} - E^{(1)}_{L}. \end{aligned} \tag{A2}$$

The proof of the formulas in Eq. (2.3) now simply consists of summing Eqs. (2.2a), (A1), and (A2) over all dimers, trimers, and tetramers, respectively, and collecting repeated terms. For example, $E^{(2)} = \sum_I E^{(1)}_I + \sum_{I<J}(E^{(1)}_{IJ} - E^{(1)}_I - E^{(1)}_J)$ can be simplified by summing the final two monomer energies over all dimers, recognizing that between $E^{(1)}_I$ and $E^{(1)}_J$, each monomer energy appears $N - 1$ times in the sum over dimers:

$$E^{(2)} = \sum_I E^{(1)}_I + \sum_{I<J} E^{(1)}_{IJ} - (N-1)\sum_I E^{(1)}_I. \tag{A3}$$

This leads directly to Eq. (2.3b).

The three-body case can be simplified using Eqs. (2.3b) and (A1):

$$\begin{aligned} E^{(3)} &= \sum_I \Delta E^{(1)}_I + \sum_{I<J} \Delta E^{(1)}_{IJ} + \sum_{I<J<K} \Delta E^{(1)}_{IJK} \\ &= \sum_{I<J} E^{(1)}_{IJ} - (N-2)\sum_I E^{(1)}_I \\ &\quad + \sum_{I<J<K} \left(E^{(1)}_{IJK} - E^{(1)}_{IJ} - E^{(1)}_{IK} - E^{(1)}_{JK} \right. \\ &\quad \left. + E^{(1)}_I + E^{(1)}_J + E^{(1)}_K\right). \end{aligned} \tag{A4}$$

In the final summation over trimers, the sum of dimer energies can be simplified according to

$$\sum_{I<J<K} (E^{(1)}_{IJ} + E^{(1)}_{IK} + E^{(1)}_{JK}) = (N-2)\sum_{I<J} E^{(1)}_{IJ}, \tag{A5}$$

while the monomer energies summed over trimers affords

$$\sum_{I<J<K} \left(E^{(1)}_I + E^{(1)}_J + E^{(1)}_K\right) = \tfrac{1}{2}(N-1)(N-2)\sum_I E^{(1)}_I. \tag{A6}$$

Substituting Eqs. (A5) and (A6) into Eq. (A4) affords Eq. (2.3c).

The four-body case [Eq. (2.3d)] is derived analogously, using the simplified three-body formula [Eq. (2.3c)] to sum everything through trimers, expanding the $\Delta E^{(1)}_{IJKL}$ term using Eq. (A2), and then summing the resulting one-, two-, and three-body energies over tetramers.

APPENDIX B: n-BODY ENERGY CORRECTION FORMULA

The form of the two-, three-, and four-body energy corrections [Eqs. (2.2a), (A1), and (A2), respectively] suggest a pattern. Generalizing, the energy of an n-body sub-cluster can

be expressed as

$$\Delta E^{(1)}_{IJ\cdots n} = \sum_{m=0}^{n-1}(-1)^m \sum_{\substack{I'<J'<\cdots<(n-m)' \\ \in\{I,J,\ldots,n\}}} E^{(1)}_{I'J'\cdots(n-m)'}, \tag{B1}$$

where the notation in the second summation indicates that one must sum over all unique $(n-m)$-body sub-clusters that can be formed from monomers $I, J, \ldots, n$. This is a cumbersome notation and we introduce the following as a compact alternative:

$$\Delta E^{(n)}_I = \sum_{m=0}^{n-1}(-1)^m \sum_{J\subset I} E^{(n-m)}_J$$
$$= E^{(n)}_I + \sum_{m=1}^{n-1}(-1)^m \sum_{J\subset I} E^{(n-m)}_J. \tag{B2}$$

The notation $\Delta E^{(n)}_I$ employs a superscript to indicate that this is an n-body energy correction, and the subscript I indexes the n-body sub-clusters, $I = 1, \ldots, {}_N C_n$. The notation $J \subset I$ in the second summation of Eq. (B2) indicates that the sum is restricted to $(n-m)$-body sub-clusters formed from the Ith n-body cluster, or in other words, that the monomer indices implicitly contained within J must be a subset of the ones that constitute I.

It is straightforward to verify that the dimer, trimer, and tetramer energies worked out in Appendix A do indeed have the form given in Eq. (B2). Therefore, let us suppose that this equation is valid, and then derive the corresponding result for $\Delta E^{(n+1)}_I$ by induction.

By definition [generalizing Eq. (2.2)], we have

$$\Delta E^{(n+1)}_I = E^{(n+1)}_I - \sum_{m=1}^{n}\sum_{J\subset I} \Delta E^{(n+1-m)}_J. \tag{B3}$$

The inductive hypothesis [Eq. (B2)] can be used to substitute for each $\Delta E^{(n+1-m)}_J$ in this equation, with the result

$$\Delta E^{(n+1)}_I = E^{(n+1)}_I - \sum_{m=1}^{n}\sum_{J\subset I}\left[\sum_{\ell=0}^{n-m}(-1)^\ell \sum_{K\subset J} E^{(n+1-m-\ell)}_K\right]$$
$$= E^{(n+1)}_I - \sum_{m=1}^{n}\sum_{k=1}^{n+1-m}(-1)^{n+1-m-k}\sum_{J\subset I}\sum_{K\subset J} E^{(k)}_K, \tag{B4}$$

where $k = n+1-m-\ell$ has been defined to obtain the second equality. Now the question is, for a given value of m, how many times does $E^{(k)}_K$ appear in Eq. (B4)? Recall that the quantity in square brackets in this equation represents the energy of an $(n+1-m)$-mer; therefore, $E^{(k)}_K$ is the energy of some k-body sub-cluster of this $(n+1-m)$-mer. Different subsystems containing $n+1-m$ monomer units may contain the same *sub*-subsystem with k monomer units, hence the index K need not be unique, and we need to know what amounts to the "degeneracy" of the index K. The answer is equal to the number of ways of choosing $n+1-m$ distinct indices from a set of $n+1$, such that k of the indices are common to each selection. That number is

$$\binom{n+1-k}{n+1-k-m} = \binom{n+1-k}{m}. \tag{B5}$$

Putting this all together,

$$\Delta E^{(n+1)}_I = E^{(n+1)}_I - \sum_{m=1}^{n}\sum_{k=1}^{n+1-m}(-1)^{n+1-m-k}$$
$$\times \binom{n+1-k}{m}\sum_{J\subset I} E^{(k)}_J. \tag{B6}$$

At this point we want to remove the m-dependence from the upper limit in the sum over k. Since ${}_{n+1-k}C_m = 0$ for $k > n+1-m$, we can extend the summation over k all the way to n (equivalent to adding zero), and then reverse the order of the (now independent) summations over k and m:

$$\Delta E^{(n+1)}_I = E^{(n+1)}_I - \sum_{k=1}^{n}\sum_{J\subset I} E^{(k)}_J$$
$$\times \sum_{m=1}^{n}(-1)^{n+1-m-k}\binom{n+1-k}{m}. \tag{B7}$$

The final summation in this equation evaluates to simply $(-1)^{n-k}$, by making use of the identity

$$\sum_{j=0}^{p}(-1)^j\binom{p}{j} = 0. \tag{B8}$$

Thus,

$$\Delta E^{(n+1)}_I = E^{(n+1)}_I + \sum_{k=1}^{n}(-1)^{n+1-k}\sum_{J\subset I} E^{(k)}_J. \tag{B9}$$

Putting $m = n+1-k$ affords the $(n+1)$-body version of Eq. (B2), which therefore proves Eq. (B2) in the general case, by induction.

APPENDIX C: *n*-BODY APPROXIMATION FOR ARBITRARY *n*

We seek to derive the general, closed form expression for $E^{(n)}$ that was given in Eq. (2.6). The proof proceeds by induction, supposing that Eq. (2.6) is correct and then deriving the analogous expression for $E^{(n+1)}$. By definition,

$$E^{(n+1)} = E^{(n)} + \sum_{K}^{\binom{N}{n+1}} \Delta E^{(n+1)}_K$$
$$= \sum_{m=0}^{n-1}(-1)^m\binom{N-n-1+m}{m}\mathcal{E}^{(n-m)}$$
$$+ \sum_K \Delta E^{(n+1)}_K, \tag{C1}$$

where in the second equality we have used the inductive hypothesis for $E^{(n)}$ and introduced the notation

$$\mathcal{E}^{(n)} = \sum_{I=1}^{\binom{N}{n}} E^{(n)}_I. \tag{C2}$$

As in Appendix B, we may replace $\Delta E_K^{(n+1)}$ with a signed summation over the energies of smaller sub-clusters. Performing the summation over K that is required in Eq. (C1), we have

$$\sum_K^{\binom{N}{n+1}} \Delta E_K^{(n+1)} = \sum_K^{\binom{N}{n+1}} \sum_{k=0}^{n} (-1)^k \sum_{J\subset K} E_J^{(n+1-k)} = \sum_{k=0}^{n} (-1)^k \, T_k, \tag{C3}$$

where

$$T_k = \sum_K^{\binom{N}{n+1}} \sum_{J\subset K} E_J^{(n+1-k)}. \tag{C4}$$

Let us carefully consider the two summations that define T_k. The monomer indices that constitute J [an $(n+1-k)$-body cluster] are a subset of those that comprise K [an $(n+1)$-body cluster], thus for a given k, the energy $E_J^{(n+1-k)}$ in question depends on only $n+1-k$ of the monomer indexes. The other k indices can be summed to produce an overall coefficient. For an $(n+1-k)$-body sub-cluster of the N-body supersystem, there are ${}_{N-n-1+k}C_k$ ways of choosing the k indexes that are not common to the $(n+1-k)$-body sub-cluster. Thus,

$$T_k = \binom{N-n-1+k}{k} \mathcal{E}^{(n+1-k)}. \tag{C5}$$

Substituting this result into Eq. (C3) affords

$$\sum_K^{\binom{N}{n+1}} \Delta E_K^{(n+1)} = \sum_{k=0}^{n} (-1)^k \binom{N-n-1+k}{k} \mathcal{E}^{(n+1-k)}. \tag{C6}$$

Comparison to Eq. (C1) shows that the result derived in Eq. (C6) has precisely the form that is needed for the $(n+1)$-body term. This proves Eq. (2.6), by induction.

[1] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).
[2] G. J. O. Beran and S. Hirata, Phys. Chem. Chem. Phys. **14**, 7559 (2012).
[3] S. D. Yeole and S. R. Gadre, J. Chem. Phys. **132**, 094102 (2010).
[4] A. P. Rahalkar and S. R. Gadre, J. Chem. Sci. **124**, 149 (2012).
[5] R. M. Richard and J. M. Herbert, J. Chem. Theory Comput. **9**, 1408 (2013).
[6] L. D. Jacobson, R. M. Richard, K. U. Lao, and J. M. Herbert, Annu. Rep. Comput. Chem. **9**, 25 (2013).
[7] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[8] E. Prodan and W. Kohn, Proc. Natl. Acad. Sci. U.S.A. **102**, 11635 (2005).
[9] R. M. Richard, K. U. Lao, and J. M. Herbert, J. Phys. Chem. Lett. **4**, 2674 (2013).
[10] R. M. Richard, K. U. Lao, and J. M. Herbert, J. Chem. Phys. **139**, 224102 (2013).
[11] T. Sawada, D. G. Fedorov, and K. Kitaura, J. Phys. Chem. B **114**, 15700 (2010).
[12] M. P. Hodges, A. J. Stone, and S. S. Xantheas, J. Phys. Chem. A **101**, 9163 (1997).
[13] J. Cui, H. Liu, and K. D. Jordan, J. Phys. Chem. B **110**, 18872 (2006).
[14] A. Hermann, R. P. Krawczyk, M. Lein, P. Schwerdtfeger, I. P. Hamilton, and J. J. P. Steward, Phys. Rev. A **76**, 013202 (2007).
[15] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).
[16] E. E. Dahlke, H. R. Leverentz, and D. G. Truhlar, J. Chem. Theory Comput. **4**, 33 (2008).
[17] G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).
[18] T. J. Mach and T. D. Crawford, Theor. Chem. Acc. **133**, 1449 (2014).
[19] I. G. Kaplan, R. Santamaria, and O. Novaro, Mol. Phys. **84**, 105 (1995).
[20] B. Paulus, K. Rosciszewski, N. Gaston, P. Schwerdtfeger, and H. Stoll, Phys. Rev. B **70**, 165106 (2004).
[21] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **4**, 1 (2008).
[22] H. R. Leverentz and D. G. Truhlar, J. Chem. Theory Comput. **5**, 1573 (2009).
[23] A. Hermann and P. Schwerdtfeger, J. Chem. Phys. **131**, 244508 (2009).
[24] A. Sebetci and G. J. O. Beran, J. Chem. Theory Comput. **6**, 155 (2010).
[25] K. D. Nanda and G. J. O. Beran, J. Chem. Phys. **137**, 174106 (2012).
[26] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).
[27] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).
[28] A. I. Krylov and P. M. W. Gill, WIREs Comput. Mol. Sci. **3**, 317 (2013).
[29] R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113 (2012).
[30] S. Kazachenko and A. J. Thakkar, J. Chem. Phys. **138**, 194302 (2013).
[31] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Imprey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[32] P. M. W. Gill, B. G. Johnson, and J. A. Pople, Chem. Phys. Lett. **209**, 506 (1993).
[33] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).
[34] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).
[35] J. Friedrich, H. Yu, H. R. Leverentz, P. Bai, J. I. Siepmann, and D. G. Truhlar, J. Phys. Chem. Lett. **5**, 666 (2014).
[36] W. Li, S. Li, and Y. Jiang, J. Phys. Chem. A **111**, 2193 (2007).
[37] X. Wang, J. Liu, J. Z. H. Zhang, and X. He, J. Phys. Chem. A **117**, 7149 (2013).
[38] M. Isegawa, B. Wang, and D. G. Truhlar, J. Chem. Theory Comput. **9**, 1381 (2013).
[39] See supplementary material at http://dx.doi.org/10.1063/1.4885846 for additional data analysis.
[40] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).
[41] M. Kamiya, S. Hirata, and M. Valiev, J. Chem. Phys. **128**, 074103 (2008).
[42] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharmaa, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, "Advances in quantum chemistry contained in the Q-Chem 4 program package" Mol. Phys. (submitted).
[43] V. Ganesh, R. K. Dongare, P. Balanarayan, and S. R. Gadre, J. Chem. Phys. **125**, 104109 (2006).
[44] J.-M. Muller *et al.*, *Handbook of Floating-Point Arithmetic* (Springer, 2009).
[45] G. Knizia, W. Li, S. Simon, and H.-J. Werner, J. Chem. Theory Comput. **7**, 2387 (2011).
[46] L. D. Jacobson and J. M. Herbert, J. Chem. Phys. **134**, 094118 (2011).
[47] K. U. Lao and J. M. Herbert, J. Phys. Chem. Lett. **3**, 3241 (2012).
[48] K. U. Lao and J. M. Herbert, J. Chem. Phys. **139**, 034107 (2013).
[49] M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[50] C. W. Murray, N. C. Handy, and G. J. Laming, Mol. Phys. **78**, 997 (1993).

[51]T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **351**, 475 (2002).
[52]A. P. Rahalkar, V. Ganesh, and S. R. Gadre, J. Chem. Phys. **129**, 234101 (2008).
[53]A. P. Rahalkar, M. Katouda, S. R. Gadre, and S. Nagase, J. Comput. Chem. **31**, 2405 (2010).
[54]N. Sahu, S. D. Yeole, and S. R. Gadre, J. Chem. Phys. **138**, 104101 (2013).
[55]S. Hua, W. Hua, and S. Li, J. Phys. Chem. A **114**, 8126 (2010).
[56]W. Li, W. Hua, T. Fang, and S. Li, "The energy-based fragmentation approach for *ab initio* calculations of large systems," in *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*, edited by J. R. Reimers (Wiley, Hoboken, NJ, 2011), pp. 227–258.
[57]N. J. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **8**, 2669 (2012).