The Journal of Chemical Physics

AIP Publishing

RESEARCH ARTICLE | JUNE 15 2023

Benchmark coupled-cluster lattice energy of crystalline benzene and assessment of multi-level approximations in the many-body expansion

Carlos H. Borca ; Zachary L. Glick ; Derek P. Metcalf ; Lori A. Burns ; C. David Sherrill

J. Chem. Phys. 158, 234102 (2023)
https://doi.org/10.1063/5.0159410

Benchmark coupled-cluster lattice energy of crystalline benzene and assessment of multi-level approximations in the many-body expansion

Cite as: J. Chem. Phys. 158, 234102 (2023); doi: 10.1063/5.0159410
Submitted: 23 May 2023 • Accepted: 25 May 2023 •
Published Online: 15 June 2023

Carlos H. Borca,[1,2] Zachary L. Glick,[1] Derek P. Metcalf,[1] Lori A. Burns,[1] and C. David Sherrill[1,3,a)]

AFFILIATIONS

[1] Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, USA
[2] Department of Chemical and Biological Engineering, School of Engineering and Applied Science, Princeton University, 41 Olden Street, Princeton, New Jersey 08544, USA
[3] School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

[a)] **Author to whom correspondence should be addressed:** sherrill@gatech.edu

ABSTRACT

The many-body expansion (MBE) is promising for the efficient, parallel computation of lattice energies in organic crystals. Very high accuracy should be achievable by employing coupled-cluster singles, doubles, and perturbative triples at the complete basis set limit [CCSD(T)/CBS] for the dimers, trimers, and potentially tetramers resulting from the MBE, but such a brute-force approach seems impractical for crystals of all but the smallest molecules. Here, we investigate hybrid or multi-level approaches that employ CCSD(T)/CBS only for the closest dimers and trimers and utilize much faster methods like Møller–Plesset perturbation theory (MP2) for more distant dimers and trimers. For trimers, MP2 is supplemented with the Axilrod–Teller–Muto (ATM) model of three-body dispersion. MP2(+ATM) is shown to be a very effective replacement for CCSD(T)/CBS for all but the closest dimers and trimers. A limited investigation of tetramers using CCSD(T)/CBS suggests that the four-body contribution is entirely negligible. The large set of CCSD(T)/CBS dimer and trimer data should be valuable in benchmarking approximate methods for molecular crystals and allows us to see that a literature estimate of the core-valence contribution of the closest dimers to the lattice energy using just MP2 was overbinding by 0.5 kJ mol^{-1}, and an estimate of the three-body contribution from the closest trimers using the T0 approximation in local CCSD(T) was underbinding by 0.7 kJ mol^{-1}. Our CCSD(T)/CBS best estimate of the 0 K lattice energy is −54.01 kJ mol^{-1}, compared to an estimated experimental value of −55.3 ± 2.2 kJ mol^{-1}.

I. INTRODUCTION

In 2006, Schweizer and Dunitz[1] computed the lattice energy of crystalline benzene using a many-body expansion (MBE).[2,3] In this approach, one accumulates the lattice energy as a sum of contributions from dimers, trimers, tetramers, etc. within the crystal. Schweizer and Dunitz stopped at the level of dimers, which neglects non-pairwise-additive contributions like higher-order polarization and three-body dispersion. Nevertheless, because benzene lacks a permanent dipole with which to polarize other monomers, one would have expected such higher-order effects to be minor. Disappointingly, Schweizer and Dunitz obtained a very poor agreement with the experiment. They employed second-order Møller–Plesset Perturbation Theory (MP2) and observed that "MP2 calculations give more than double the correct value and are clearly not very useful. The counterpoise-corrected values yield erratic values for the lattice energy, and it is not obvious how any larger basis set or improved BSSE (basis set superposition error) correction would influence the result." This led them to conclude that "the goal of calculating an unconditionally reliable *ab initio* value for the lattice energy of benzene may still seem remote, but it is on the horizon and should be attainable."

As demonstrated by Ringer and Sherrill in 2008,[4] the main problem was not with the MBE or with neglecting many-body contributions but with stopping at the MP2 level. Although this approach is considered by some to be an "accurate" quantum chemistry method, it can strongly overestimate intermolecular π–π interactions.[5–7] However, using the "gold standard" of quantum chemistry, coupled-cluster through perturbative triple excitations [CCSD(T)],[8] and estimating results at the complete-basis-set (CBS) limit, yielded results apparently within "chemical accuracy" (±1 kcal mol^{-1}).[4] Later, several studies indicated that the three-body effects were significant (e.g., 14% of the lattice energy).[9–11] In 2014, CCSD(T)/CBS estimates for a large number of trimers from the crystal indicated that the total three-body effect is closer to half of this value,[12] which nevertheless remains significant for high-accuracy estimates. In this same year, Chan and co-workers reported a *tour de force* study of crystalline benzene aimed at determining the lattice energy within 1 kJ mol^{-1} accuracy.[13] That study, including partial accounting for beyond-trimer contributions, four-electron contributions, and other small corrections, demonstrated convincingly that the many-body expansion is entirely capable of providing very-high-accuracy results, at least for some molecular solids. Other studies since 2008 have also demonstrated the reliability of the MBE for computations of molecular crystals.[14–20]

These examples suggest that the MBE, coupled with *ab initio* quantum chemistry methods, can compute lattice energies very accurately, more accurately than is currently possible with standard force-field methods or density functional theory methods. High-accuracy computations of lattice energies would be very beneficial in energetically ranking polymorphs, which are often very close in energy. The high computational cost of the most accurate approaches, like CCSD(T)/CBS, is mitigated by the fact that the individual N-mer computations in the MBE are totally independent and thus can be run in a parallel fashion. Nevertheless, remaining obstacles to the routine application of a CCSD(T)/CBS MBE have been (a) the lack of commonly available software to automate the construction and running of the individual N-mer computations and their reconstruction into a crystal lattice energy, and (b) thorough benchmark studies demonstrating when CCSD(T)/CBS is truly necessary and when lower levels of theory may be substituted for computational efficiency. To help address obstacle (a), we recently introduced CrystaLattE,[21] an open-source program that takes a crystal structure in crystallographic information file (CIF), automatically creates a series of non-symmetry-redundant N-mers and associated quantum chemistry input files for PSI4,[22] and processes the results into the lattice energy. The present work is aimed at gaining a better understanding of issue (b), initially for the test case of crystalline benzene. Our expectation based on prior work (e.g., Refs. 7 and 12) is that, at long range, approximations to CCSD(T)/CBS using smaller basis sets and/or less complete electron correlation models should allow much faster computations at only a modest reduction in accuracy. Therefore, we imagine a "multi-layer" approach in which expensive CCSD(T)/CBS or other high-accuracy computations are performed on only a handful of the most important dimers and trimers, and faster methods are used for the remainder.

Here, we obtain CCSD(T)/CBS results for symmetry-unique dimers, *trimers, and even a few tetramers* within crystalline benzene, up to large cutoff distances (for a total of 420 dimers, 1977 trimers, and 24 tetramers). This dramatically expands upon our 2014 study of trimers, which involved only 366 (symmetry-redundant) trimers and only used the aug-cc-pVDZ basis set. It also improves upon the CCSD(T) approach utilized by Chan and co-workers,[13] which involved local correlation approximations whose effects on interaction energies have not been completely characterized. For dimers, we have also assessed the effect of the core-valence correlation. These CCSD(T)/CBS dimer and trimer benchmark results should be valuable for testing more approximate methods for noncovalent interactions and lattice energy computations. Compared to these benchmark results, we assess the reliability of CCSD(T)/CBS estimates obtained using smaller basis sets, the MP2 method, and Axilrod–Teller–Muto (ATM)[23,24] estimates of three-body dispersion. Our results indicate that core-valence contributions have a minor effect and are only important at very short ranges, that smaller basis sets remain effective for computing CCSD(T)/CBS estimates beyond very short ranges, and that MP2 and/or ATM can be effective approximations at long ranges, saving a huge fraction of the computational time. Computations on tetramers indicate that four-body effects make a negligible contribution (0.1 kJ mol^{-1} or less).

II. THEORY

In the MBE,[2,3] the total energy of a molecular cluster is given as

$$E = \sum_{I}^{N} E_I + \sum_{I<J}^{N} \Delta E_{IJ}^{(2)} + \sum_{I<J<K}^{N} \Delta E_{IJK}^{(3)} + \sum_{I<J<K<L}^{N} \Delta E_{IJKL}^{(4)} + \cdots . \tag{1}$$

Here, N is the number of monomers in the cluster, E_I is the energy of monomer I, and $\Delta E_{IJ^{(2)}}$ is the interaction energy of the dimer formed by monomers I and J,

$$\Delta E_{IJ}^{(2)} = E_{IJ} - E_I - E_J, \tag{2}$$

ΔE_{IJK} is the non-additive three-body energy of the trimer formed by monomers I, J, and K,

$$\Delta E_{IJK}^{(3)} = E_{IJK} - E_I - E_J - E_K - \Delta E_{IJ}^{(2)} - \Delta E_{IK}^{(2)} - \Delta E_{JK}^{(2)}, \tag{3}$$

and $\Delta E_{IJKL}^{(4)}$ is the non-additive four-body energy of the tetramer formed by monomers I, J, K, and L, which is likewise computed by subtracting all monomer energies and all dimer and non-additive trimer interaction energies from E_{IJKL}.

Equation (1) is exact when terms up to the non-additive N-body contribution for a cluster of N molecules are included. Truncating at a lower order, typically at the level of two- or three-body interactions, provides an efficient way to approximate the total energy at a substantially reduced computational cost.

The lattice energy of a crystal is defined as the energy required to construct the lattice starting from a state where all its molecules are infinitely separated. In this study, geometries are extracted from an experimentally-determined crystalline structure. The monomer geometries are assumed to be rigid, and any monomer deformation terms are, therefore, neglected. This should be a good approximation for a rigid molecule like benzene. Hence, the lattice energy is the sum of the N-body interaction energies, $\Delta E_{IJ}^{(2)}$, $\Delta E_{IJK}^{(3)}$, etc., up through the chosen truncation level. However, if the crystal is modeled as an infinitely extended solid, then there are an infinite number of dimers, trimers, etc. Therefore, the lattice energy is computed *per monomer*, or per mole of monomers, to obtain a finite result. For this purpose, a *reference* monomer is picked, and that monomer will appear in all

N-mers of the MBE. Then all dimers, trimers, etc. that satisfy various filtering criteria are selected.

For each N-mer retained, its N-body interaction energy is computed. This energy is divided by N, the number of monomers in the N-mer, to obtain the contribution *per monomer*. Each symmetry-unique N-mer needs to be computed only once, and its energy per monomer should be multiplied by the number of symmetry-equivalent N-mers that contain the same reference monomer. The contribution, $\mathcal{C}_{N-\text{mer}}$, of each N-mer to the crystal lattice energy is thus computed as

$$\mathcal{C}_{N-\text{mer}} = \mathcal{R}_{N-\text{mer}} \times \frac{\Delta E^{(N)}_{I\ldots N}}{N}, \tag{4}$$

where N is the number of monomers and $\mathcal{R}_{N-\text{mer}}$ is the number of replicas of each unique N-mer. For dimers, $\Delta E^{(N)}_{I\ldots N}$ is the dimer interaction energy, as defined in Eq. (2); for trimers and higher-order N-mers, $\Delta E^{(N)}_{I\ldots N}$ corresponds to the non-additive portion of the many-body energy, e.g., as defined for a trimer in Eq. (3).

III. COMPUTATIONAL DETAILS AND METHODS

The computational procedure was split into two parts. First, a set of symmetry-unique dimer, trimer, and tetramer configurations was extracted from a model of crystalline benzene using the CrystaLattE code.[21] Then, computations were performed using the Psi4 electronic structure program.[25] These computations were distributed and run in parallel among available resources on the clusters of the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology.

A. CrystaLattE setup

An experimentally-determined structure of crystalline benzene at 138 K that has been used in previous computational studies by our group and others[4,12,13,21,26–28] was obtained in CIF format from the Cambridge Structural Database (CSD). The structure corresponds to CSD code `BENZEN01`, and it is the same employed in our demonstration application of CrystaLattE.[21]

CrystaLattE generates a supercell by replicating the unit cell described by the CIF file. It then selects dimers, trimers, etc., each containing the "reference monomer" and satisfying the appropriate *N-mer cutoff*. The N-mer cutoff computes the closest atom–atom distance between each pair of monomers and then ensures that the longest of these is shorter than a given cutoff distance. The user can specify different cutoffs for different N-mer orders, i.e., different dimer, trimer, and tetramer cutoffs. For this work, we updated CrystaLattE to automatically compute the size of the supercell required to achieve the requested N-mer cutoffs. A threshold of 1.2 times the sum of the van der Waals radii was employed to automatically detect monomers in the supercell through the breadth-first search algorithm.

Here, we employed a dimer cutoff distance of 30 Å, although the very highest levels of theory considered here were only employed out to 8 Å (at this distance, slightly more approximate high-level methods performed very well and were utilized as benchmarks for the remaining dimers, as described below). A trimer cutoff distance of 15 Å and a tetramer cutoff distance of 10 Å were employed to keep the number of three- and four-body interactions computed manageable. The symmetry-uniqueness of the configurations was determined using a chemical-space descriptor.[21] 420 dimer, 1977 trimer, and 1332 tetramer configurations were generated using these cutoffs. The MBE was truncated at the level of tetramers.

B. Psi4 setup

All calculations were executed using Psi4 1.2.1,[25] except in cases where core-valence correlation corrections were required. Those calculations were run with a nightly-build, development version, 1.4a2.dev313.[22] Calculations were executed with two methods: (1) coupled cluster with single and double excitations and perturbative triple excitations, or CCSD(T),[8] and (2) second-order Møller–Plesset perturbation theory, or MP2. The self-consistent field (SCF) procedure and the correlation energies were computed using the density-fitting approximation for both CCSD(T) and MP2 calculations.[22,29–31] Core orbitals were constrained to remain doubly occupied except in computations meant to assess core-valence correlation, which were performed with core-valence basis sets. The Boys and Bernardi counterpoise correction (CP)[32,33] for basis set superposition error treatment was used in all the calculations. For example, when computing three-body contributions for a trimer, the energies of the constituent monomers and dimers were obtained using the basis functions of all three monomers.

The frozen natural orbital (FNO) approximation, with the default occupation number cutoff of 10^{-6} electrons, was employed in CCSD(T) calculations of dimers.[31] In our initial trimer computations, we noticed some very small errors in the non-additive three-body interaction energies due to the FNO approximation, which appeared to be at least partially systematic; that is, the errors tended to have the same sign and thus could accumulate when summing large numbers of trimer contributions. Therefore, we regenerated the trimer CCSD(T)/CBS data without the use of the FNO approximation. With the default FNO cutoff of 10^{-6} electrons, the total CCSD(T)/CBS three-body contribution to the lattice energy was underestimated by 0.24 kJ mol^{-1} vs results without the FNO approximation.

1. Two-body interactions

MP2 and CCSD(T) calculations were performed to analyze how the total two-body contribution to the crystal lattice energy depends upon the choice of method, basis set (or basis set extrapolation), and core-valence correlation effects. MP2 calculations were performed using the aug-cc-pVDZ basis set, dubbed aDZ for brevity, and the aug-cc-pVTZ basis, labeled aTZ.

CCSD(T) calculations employed a focal-point scheme to extrapolate the results to the complete basis set (CBS),[34,35] which works well for noncovalent interactions when using correlation-consistent basis sets.[6,36–38] Smaller basis sets often capture higher-order electron correlation effects well. Therefore, CCSD(T) in a large basis set may be approximated by

$$E(\text{CCSD(T)/Large}) \approx E(\text{MP2/Large}) + \delta^{\text{CCSD(T)}}_{\text{MP2}}/\text{Small}, \tag{5}$$

where large and small refer to the relative size of the basis set and the difference, or δ, term is defined as

$$\delta^{\text{CCSD(T)}}_{\text{MP2}}/\text{Small} = E(\text{CCSD(T)/Small}) - E(\text{MP2/Small}). \tag{6}$$

The large basis in Eq. (5) can be substituted by an estimate of the CBS limit, and it has been shown that a two-point extrapolation of the correlation energy is usually sufficient for this purpose.[39] Following the focal-point approach, in this study, a DF-FNO-CCSD(T)/aug-cc-pVXZ computation was accompanied by a two-point CBS extrapolation of the DF-MP2 correlation energy using aug-cc-pV(X+1)Z and aug-cc-pV(X+2)Z basis sets. The Hartree–Fock reference contribution to the focal-point energy is taken as the Hartree–Fock energy in the largest basis set considered, i.e., aug-cc-pV(X+2)Z. To analyze the impact of the basis set, we generated CCSD(T)/CBS focal-point estimates using both double-zeta (X = 2) and triple-zeta (X = 3) basis sets for the coupled-cluster part of the estimate. To analyze core-valence correlation effects, we performed analogous focal-point CCSD(T)/CBS procedures using Dunning's weighted core-valence (cc-pwCVXZ) basis sets for both the CCSD(T) and MP2 computations, with unfreezing of the core electrons.

To make the text more readable, we will use a shorthand notation for the focal-point procedures. For example, the method dubbed CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) indicates a focal-point computation estimating the CCSD(T)/CBS limit according to Eqs. (5) and (6), using DF-FNO-CCSD(T)/aug-cc-pwCVTZ together with an estimate of the MP2/CBS limit based on DF-MP2/aug-cc-pwCVQZ and DF-MP2/aug-cc-pwCV5Z calculations.

For two-body computations, we utilized an energy convergence criterion of 10^{-10} a.u. except for focal-point computations at higher levels of theory than CCSD(T)/CBS(a[TQ]Z; δ:aDZ), where a cutoff of 10^{-9} a.u. was used.

We expect our highest level of theory used here, CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ), to be very accurate and a suitable high-level reference for two-body interactions. The only potentially significant effect that it omits is correlated motion between four electrons simultaneously, modeled by quadruple excitations in the coupled-cluster ansatz. Unfortunately, coupled-cluster through quadruple excitations, CCSDTQ, is extremely expensive computationally, and perturbative approximations to it remain very expensive. Moreover, as shown by Smith and Patkowski,[40] small double-ζ basis sets do not seem sufficient to accurately estimate the effect of quadruple excitations on intermolecular interactions. Chan and co-workers[13] attempted to estimate the effect of quadruple excitations on the two-body energy using a perturbative estimate of quadruples, i.e., the CCSDT(Q) model. Unfortunately, due to computational expense, they were forced to use the very small 6-31G basis set and freeze 56 electrons in the computation (there are only 24 core electrons in the benzene dimer). They then attempted to correct for these approximations by a simple scaling of the quadruples contribution by the ratio of the (T) contribution in a large basis set with 24 frozen electrons vs the (T) correction in the 6-31G basis with 56 frozen electrons. This procedure (actually an average of two similar procedures) yields an estimate of +0.36 kJ mol^{-1} for the two-body contribution of quadruples. Due to the significant approximations employed, we consider this estimate to carry substantial uncertainty, but we are unable to perform more accurate estimates of the quadruple effects here due to computational expense. Nevertheless, we are reassured that the quadruples effect appears to be quite small, and smaller than some of the other effects that we explore here.

2. Three-body interactions

To analyze how the total three-body contribution to the computed crystal lattice energy depends upon the choice of method and basis set size, MP2/aDZ, MP2/aTZ, and CCSD(T)/CBS(a[TQ]Z;δ:aDZ) computations were performed on trimers up to a trimer cutoff of 15 Å, yielding 1977 symmetry-unique trimers. In addition, MP2 results were subject to the addition of an undamped molecular Axilrod–Teller–Muto (ATM) correction[23,24] to approximate the three-body dispersion missing in MP2. This correction is computed using a molecular ATM constant of 82 657.65 a.u. taken from previous estimations using Density Functional Theory-based Symmetry-Adapted Perturbation Theory [SAPT(DFT)].[9] For trimer computations, we utilized an energy convergence criterion of 10^{-10} a.u. For the four closest trimers only, we performed higher-level CCSD(T)/CBS(a[Q5]Z; δ:aTZ) computations. These used the FNO approximation (with a cutoff 10^{-8} electrons for weakly occupied NO's, or 100 times tighter than the normal conservative cutoff) and an energy convergence criterion of 10^{-8} a.u.

3. Four-body interactions

We analyzed the four-body contribution to the lattice energy by performing MP2/aDZ and MP2/aTZ computations on 1332 tetramers. These computations should suffice to capture any four-body induction/polarization or exchange-repulsion effects that might be important. In addition, 24 CCSD(T)/CBS(a[TQ]Z; δ:aDZ) tetramer computations were performed to analyze individual four-body contributions to the computed crystal lattice energy and their dependence upon the choice of method and basis set size. For these tetramer calculations, the energy convergence criterion was 10^{-10} a.u.

C. Numerical precision

As previously pointed out by Richard, Lao, and Herbert,[41] one must be careful in applications of the many-body expansion because the addition of large numbers of energies can cause numerical precision problems in the final result if care is not taken. The electronic energy for each monomer, dimer, etc., is converged to some cutoff value, $\varepsilon_{\text{Conv}}^{(N)}$, where the superscript (N) indicates that we might choose the cutoff value differently when computing two-body, three-body, or four-body contributions. The cutoff value on the convergence of the energy means that the energy will have an uncertainty of approximately $\varepsilon_{\text{Conv}}^{(N)}$. In our tests, the errors caused by these cutoff values (vs much more tightly converged values) appear to be random rather than systematic. Therefore, the uncertainty in the sums and differences of these energies will just be the square root of the sum of the squares of the individual uncertainties. For a dimer interaction energy, $\Delta E_{IJ}^{(2)} = E_{IJ} - E_I - E_J$, the uncertainty in the final result is just $\sqrt{\sigma(E_{IJ})^2 + \sigma(E_I)^2 + \sigma(E_J)^2}$. If all energies are computed with the same two-body convergence criterion $\varepsilon_{\text{Conv}}^{(2)}$, this simplifies to $\varepsilon_{\text{Conv}}^{(2)}\sqrt{3}$. Because the contribution of this dimer's interaction energy to the overall two-body lattice energy is $(\mathcal{R}_{IJ}\Delta E_{IJ}^{(2)})/2$ according to Eq. (4), the uncertainty in this contribution is $\varepsilon^{(2)}_{\mathrm{Conv}}\mathcal{R}_{IJ}\sqrt{3}/2$. In general, an individual N-mer's contribution to the overall non-additive N-body contribution to the lattice energy is $\varepsilon^{(N)}_{\mathrm{Conv}}\mathcal{R}_{N\mathrm{-mer}}\sqrt{n^{(N)}}/N$, where $n^{(N)}$ is the number of component energies required to compute the non-additive N-body contribution of that N-mer (3 for a dimer, 7 for a trimer, and 15 for a tetramer). When the focal-point approximation is used, additional energy computations are involved [e.g., each CCSD(T)/CBS(a[TQ]Z; δ:aDZ) single-point energy requires three MP2 energies and one CCSD(T) energy]. If we introduce the multiplier $\lambda^{(N)}$ to account for any such additional energy computations (here, it is 1 for MP2 and 4 for all focal-point procedures considered), the uncertainty in an N-mer's lattice energy contribution becomes $\varepsilon^{(N)}_{\mathrm{Conv}}\mathcal{R}_{N\mathrm{-mer}}\sqrt{\lambda^{(N)}n^{(N)}}/N$.

We wish the uncertainty in the *overall* N-body contribution to the lattice energy, which is just a simple sum of the energy contributions from each individual N-mer. Each N-mer may have a different number of replicas $\mathcal{R}_{N\mathrm{-mer}}$. However, we are concerned here with order-of-magnitude estimates of the uncertainty, so we may replace this factor with the *average* number of replicas, which is just $\mathcal{M}^{(N)}_{\mathrm{Redundant}}/\mathcal{M}^{(N)}_{\mathrm{Unique}}$, where $\mathcal{M}^{(N)}_{\mathrm{Redundant/Unique}}$ is the number of redundant or symmetry-unique N-mers in the computation. Doing this allows us to estimate the overall uncertainty for the N-body contribution as

$$\sigma\left(\Delta E^{(N)}\right) = \varepsilon^{(N)}_{\mathrm{Conv}}\left[\frac{\mathcal{M}^{(N)}_{\mathrm{Redundant}}}{\mathcal{M}^{(N)}_{\mathrm{Unique}}}\right]\frac{\sqrt{\lambda^{(N)}\,n^{(N)}\,\mathcal{M}^{(N)}_{\mathrm{Unique}}}}{N}. \tag{7}$$

Using the Psi4 convergence criteria noted earlier (nearly all energies converged to 10^{-10} a.u.) and the number of N-mers considered, we estimate that all of our N-body contributions to the lattice energy have a numerical uncertainty of under 0.001 kJ mol^{-1} due to the addition of large numbers of individual contributions.

D. Hardware details

Computations were performed on the Hive and Phoenix clusters at Georgia Tech, utilizing nodes with dual Intel Xeon processors (for Hive, E5-2640 v4 processors with ten cores each, and for Phoenix, Xeon Gold 6226 processors with 12 cores each). Computations utilized locally attached scratch disks (either a RAID0 array of hard disk drives or NVMe storage). Nodes contained 256 GB of RAM (Hive) or 192 GB of RAM (Phoenix). On Hive, five MP2 jobs would usually be accommodated simultaneously on each of the nodes by using four cores for each calculation. For CCSD(T) calculations, given the higher memory and scratch demand, only two 10 or 12 core jobs, at most, could be executed on each node. When utilizing Hive, we were often able to utilize as many as 800 cores simultaneously.

IV. RESULTS AND DISCUSSION

In this section, we first explore the computational cost of conducting this benchmark study. Then, we present the results for two-, three-, and four-body interactions. We focus on the convergence of the MBE, how the methods, basis sets, and cutoffs affect the accuracy of the calculated crystal lattice energy, and on a proposed multi-level scheme to achieve maximum efficiency in computing highly accurate lattice energies.

A. Timings

Table I summarizes the average computational cost (in core hours) of computing one N-mer with the various *ab initio* methods employed in this study. As mentioned earlier, we used different numbers of cores for CCSD(T) vs MP2 computations.

For MP2, increasing basis set size from double-to triple-zeta quadruples the average cost of calculating a dimer or a trimer and causes a six-fold increase in cost for a tetramer. For focal-point coupled-cluster (CC) methods, increasing the size of the basis set is extremely expensive. On average, the cost increases more than 18-fold when switching from CCSD(T)/CBS(a[TQ]Z; δ:aDZ) to CCSD(T)/CBS(a[Q5]Z; δ:aTZ) and about 20-fold when going from CCSD(T)/CBS(aC[TQ]Z; δ:aCDZ) to CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) for benzene dimer. This result is not surprising given the steep $\mathcal{O}(o^3v^4)$ scaling of the CCSD(T) portion of the focal-point procedure, but quantifying the computational cost increases is helpful in assessing cost/accuracy tradeoffs.

Incorporating core-valence correlation effects more than triples the average cost of focal-point CCSD(T)/CBS computation on a benzene dimer. The cost increase in this case is due to the extra determinants generated when unfreezing the core electrons and also due to the introduction of the weighted core-valence (cc-pwCVXZ) basis sets, which are larger than standard valence basis sets.

The time required to compute a trimer using the CCSD(T)/CBS(a[TQ]Z; δ:aDZ) approach is comparable to the time required to compute a dimer using a similar approach but with increased basis sets, CCSD(T)/CBS(a[Q5]Z; δ:aTZ). There are many more trimers to be computed than dimers, so the computational cost per N-mer becomes a much more serious concern for trimers (or tetramers).

Table II shows the average computational cost of running a whole N-mer set with the various levels of theory and cutoffs employed here. The times reported assume that jobs can be distributed simultaneously on 800 cores, as was routinely achievable for us on a campus cluster. In considering Table II, it is important to recall that we used quite large cutoffs so that we could carefully

TABLE I. Comparison of the computational cost, in core-hours, of calculating one N-mer with the different methods employed in this study.

	Cost [core hours][a]		
Level of theory	2-mer	3-mer	4-mer
CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ)	1245.83	N/A	N/A
CCSD(T)/CBS(a[Q5]Z; δ:aTZ)	347.64	N/A	N/A
CCSD(T)/CBS(aC[TQ]Z; δ:aCDZ)	60.99	N/A	N/A
CCSD(T)/CBS(a[TQ]Z; δ:aDZ)	18.68	612.37[b]	2155.54
MP2/aTZ	0.36	2.25	25.67
MP2/aDZ	0.09	0.55	4.49

[a] On the Hive cluster, except where otherwise noted. Nodes were running multiple jobs simultaneously (see text).

[b] On the Phoenix cluster and without the benefit of FNO approximations.

TABLE II. Comparison of the estimated effective computational cost, in wall time hours when utilizing 800 cores simultaneously, of calculating a set of *N*-mers with the different methods and cutoffs employed in this benchmark study.

	Effective walltime (h)		
Level of theory	2-mers	3-mers	4-mers
CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ)	747.5	N/A	N/A
CCSD(T)/CBS(a[Q5]Z; δ:aTZ)	208.6	N/A	N/A
CCSD(T)/CBS(aC[TQ]Z; δ:aCDZ)	36.6	N/A	N/A
CCSD(T)/CBS(a[TQ]Z; δ:aDZ)	11.2	1530.9	3664.4
MP2/aTZ	0.3	5.6	44.9
MP2/aDZ	0.1	1.4	7.9
N-mer cutoff (Å)	20	15	10
Set size	420	1977	1332

examine convergence with respect to cutoff distance. Additionally, as we will discuss below, CCSD(T)-accuracy computations are only necessary for close *N*-mers, allowing for a dramatic reduction in computation time compared to the brute force timings reported in Table II.

Table II clearly demonstrates that CCSD(T) is too computationally costly for routine use in computing three-body (or higher) contributions to the lattice energies of molecular crystals if the molecule is as large as benzene or larger. More computationally efficient methods need to be explored for trimers and tetramers (we explore this question below). On the other hand, the two-body terms should dominate the lattice energy, and one sees that the full set of dimers is attainable at the rather accurate CCSD(T)/CBS(a[TQ]Z; δ:aDZ) level of theory within half a day. MP2 computations are dramatically faster than the CCSD(T)/CBS estimates, and the full set of dimers can be obtained using MP2/aTZ or MP2/aDZ in only a few minutes. The speedups of MP2 over CCSD(T) are even more dramatic for the trimers. This motivates us to explore the accuracy of MP2 and MP2-based approaches below.

B. Two-body interactions

We now turn to the energy analysis of the two-body interactions. Figure 1 shows how the total two-body contribution to the lattice energy accumulates when adding a growing number of dimer configurations. Contributions are summed in increasing order according to the closest contact distance between monomers in each dimer, the *dimer cutoff*. To facilitate the analysis, the data in Fig. 1 have been split into three domains with respect to the dimer cutoff: the short- (<4 Å), mid- (4–8 Å), and long-range (>8 Å). The shaded area, delimiting the mid-range domain, helps distinguish these three regions. Table III presents the accumulated lattice energy contribution for each method and distance domain and further distinguishes between long-range (8–20 Å) and very-long-range (20–30 Å). The highest level of theory considered here for dimers is CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ), i.e., a focal-point approach using a Hartree–Fock/aug-cc-pwCV5Z reference energy, a two-point extrapolation of the MP2 correlation energy using the aug-cc-pwCVQZ and aug-cc-pwC5Z basis sets, and a coupled-cluster correction computed as CCSD(T)/aug-cc-pwCVTZ - MP2/aug-cc-pwCVTZ. These computations include core-valence electron correlation. This time-consuming method is used for $R \leq 8$. In tests of the range $8 < R \leq 14$, we found that the faster (valence-only) focal-point approach CCSD(T)/CBS(a[TQ]Z; δ:aDZ) agrees within 0.01 kJ mol^{-1} of our best results for the accumulated lattice energy contribution, and so that method is adopted as the reference method for $8 < R \leq 30$ Å.

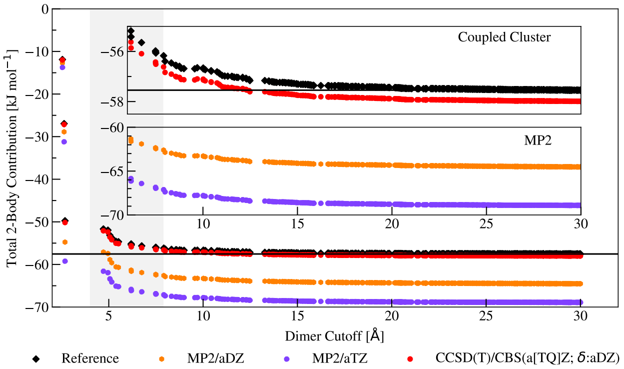

FIG. 1. Accumulation of two-body contributions to the lattice energy as a function of the closest contact distance between monomers. The mid-range region (4–8 Å) is shaded. The first three dimers, whose monomers are separated by less than 2.7 Å, account for most of the total two-body contribution. Reference values are CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) for ≤8 Å and CCSD(T)/CBS(a[TQ]Z; δ:aDZ) for >8 Å (see text). All curves become nearly identical (but shift) in the mid and long ranges, indicating that differences arise mainly from short- or early mid-range contacts.

The two-body component of the lattice energy converges rapidly with distance, showing a quasi-exponential decay, regardless of the level of theory. The individual dimer interaction energies in the long-range domain basically agree, with a few exceptions, across different levels of theory. Figure 1 and Table III demonstrate a stark contrast between the proportion of the total two-body contribution originating from each distance domain. The short-range set, which includes the three coordination-shell dimers, accounts for about 86% of the total reference two-body contribution; the intermediate-range, including 14 dimers, accounts for about 12%; the long-range, which includes 125 dimers, accounts for 2%; and finally, the 278 dimers in our very-long-range bin account for only 0.2% of the total two-body contribution. This suggests that higher levels of theory only need to be applied to the closer dimers, an idea we return to below.

As shown in Table III, among the four CCSD(T)/CBS methods considered, the largest difference among the summed contributions of the three closest trimers is 0.78 kJ mol^{-1}. A variation of this size means that the choice of CCSD(T)/CBS estimation scheme matters, at least for the closest dimers, if ~1 kJ mol^{-1} accuracy is desired for the total lattice energy. However, in the medium range, the CCSD(T)/CBS estimates differ by at most 0.06 kJ mol^{-1} for the lattice energy contribution, and we noted earlier the essentially exact agreement between CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) and CCSD(T)/CBS(a[TQ]Z; δ:aDZ) in the range $8 < R \leq 14$ Å. Figure 2 further illustrates these points, displaying errors in the dimer contributions to the lattice energy [Eq. (4)] vs the highest level of theory considered here, CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ), for increasingly distant dimers across different levels of theory. Errors for the other

TABLE III. Summary of total two-body lattice energy contributions, split by separation domain, for each method employed in this study. All energies are in kJ mol^{-1}. Cutoff distances are expressed in Å. The highest level of theory used here, CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ), is used as the reference value for assessing more approximate methods in the range $R \leq 8$. In the range $8 < R \leq 14$, CCSD(T)/CBS(a[TQ]Z; δ:aDZ) matches the reference method within 0.01 kJ mol^{-1} for the total lattice energy contribution and is, therefore, adopted as the new reference method for $8 < R \leq 30$. Percentages reflect the contribution to the overall two-body energy computed at the corresponding level of theory. Energy differences with respect to the best reference values for each range are indicated in parentheses.

Range	Total $R \leq 30$	Short $R \leq 4$	Medium $4 < R \leq 8$	Long $8 < R \leq 20$	V. Long $20 < R \leq 30$
CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ)		−49.75 Ref.	−6.64 Ref.		
CCSD(T)/CBS(a[Q5]Z; δ:aTZ)		−49.66 (0.08)	−6.65 (−0.01)		
CCSD(T)/CBS(aC[TQ]Z; δ:aCDZ)		−50.44 (−0.69)	−6.70 (−0.06)		
CCSD(T)/CBS(a[TQ]Z; δ:aDZ)	−57.99 100.00% (−0.44)	−50.16 86.5% (−0.41)	−6.67 11.5% (−0.03)	−1.05 1.8% Ref.	−0.11 0.2% Ref.
CCSD(T)/CBS Ref.	−57.55 100.00% Ref.	−49.75 86.4% Ref.	−6.64 11.5% Ref.	−1.05 1.8% Ref.	−0.11 0.2% Ref.
MP2/aTZ	−68.91 100.0% (−11.36)	−59.22 85.9% (−9.47)	−8.14 11.8% (−1.50)	−1.40 2.0% (−0.35)	−0.14 0.2% (−0.03)
MP2/aDZ	−64.53 100.0% (−6.98)	−54.76 84.9% (−5.01)	−8.10 12.6% (−1.46)	−1.46 2.3% (−0.41)	−0.21 0.3% (−0.10)
Number of dimers	420 100.0%	3 0.7%	14 3.3%	125 29.8%	278 66.2%

approximate CCSD(T)/CBS methods are only discernible for the first three dimers. Therefore, we see that *basis set effects beyond* CCSD(T)/CBS(a[TQ]Z; δaDZ) *and core-valence effect are relevant for the first three dimers only*. Errors in MP2 persist somewhat longer but also decrease rapidly with increasing dimer separations.

As seen in Table III, for the three closest dimers, the CCSD(T)/CBS(a[TQ]Z; δ:aDZ) summed lattice energy contribution is −50.16 kJ mol^{-1}. If we increase the size of the basis set used for both the MP2 and the CCSD(T) portions of the focal-point approach to obtain CCSD(T)/CBS(a[Q5]Z; δ:aTZ), the lattice energy is reduced in magnitude to −49.66 (a basis set correction of +0.50 kJ mol^{-1}). If, instead, we include core-valence correlation by unfreezing the core electrons and moving to the core-valence basis sets, this yields the CCSD(T)/CBS(aC[TQ]Z; δ:aCDZ) approach, with a short-range lattice energy contribution of −50.44 kJ mol^{-1} or a core-valence correction of −0.28 kJ mol^{-1} when compared to the baseline CCSD(T)/CBS(a[TQ]Z; δ:aDZ) method. Frequently, quantum chemists add separate corrections for valence basis set extensions and for core-valence effects. If we used such an approach here, we would simply add the previous two corrections, +0.50 and −0.28 kJ mol^{-1}, to expect a total correction of +0.22 kJ mol^{-1}. Here we were able to compute CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) explicitly, which accounts for both effects simultaneously. The short-range lattice energy contribution for this method, −49.75 kJ mol^{-1}, indicates that the CCSD(T)/CBS(a[TQ]Z; δ:aDZ) estimate of −50.16 requires a total correction of +0.41 kJ mol^{-1}. Therefore, attempting to add the valence basis correction and core-valence correction separately would introduce an error of 0.22 − 0.41 = −0.19 kJ mol^{-1} compared to the explicit CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) computation. The assumption of additivity of valence basis extensions and core-valence effects is not so accurate in this case and recovers only about half of the full correction. On the other hand, a total error of −0.19 kJ mol^{-1} for these effects may be perfectly acceptable for all but the most demanding of applications, and the sum of the time required for CCSD(T)/CBS(a[TQ]Z; δ:aDZ), CCSD(T)/CBS(a[Q5]Z; δ:aTZ), and CCSD(T)/CBS(aC[TQ]Z; δ:aCDZ) computations would still be less than the time of the full CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) computation, as seen from Table I.

Turning now to the MP2 results, we see substantial differences (~1–5 kJ mol^{-1}) between MP2 and the reference CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) values for the first three dimers in Fig. 2. MP2 overbinds these dimers, as is typical for MP2 applied to π–π interactions. The larger-basis MP2/aTZ results exhibit larger overbinding, again as is typical.[4,12] The MP2 results exhibit a much larger basis set dependence than the CCSD(T)/CBS focal-point estimates, which is only natural because of the large basis sets and extrapolation used in the MP2 portion of the CCSD(T)/CBS focal-point approach. However, note that the MP2/aDZ and MP2/aTZ dimer energies closely agree with each other in Fig. 2 after the first three dimers, and moreover, their errors vs the reference

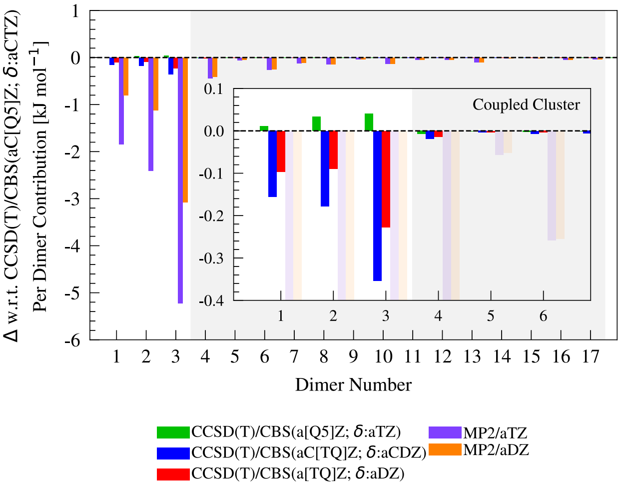

FIG. 2. Energy difference with respect to CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) per dimer contribution. The first three dimers, those belonging to the first coordination shell, account for the largest differences per dimer contribution with respect to the reference method. For perturbation theory, the differences per dimer contribution decrease to less than 0.5 kJ mol^{-1} after the third structure. For coupled cluster, those differences are less than 0.02 kJ mol^{-1} after the third structure. Coordination-shell dimers are the only relevant contributors to basis set effects beyond CCSD(T)/CBS(a[TQ]Z; δ:aDZ) and to core-valence correlation.

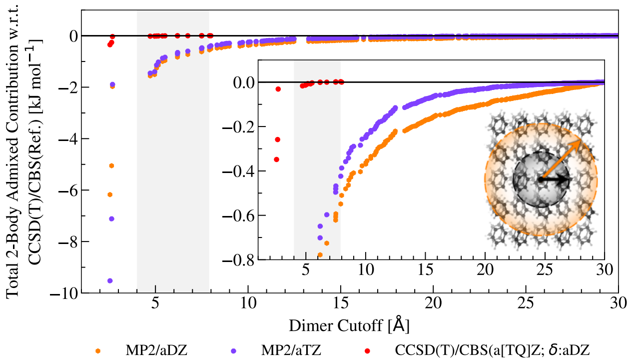

FIG. 3. Errors in the total two-body contribution to the lattice energy when using a two-layer approximation in which a high level of theory (here, the reference data) is used for dimers whose closest intermonomer contact distance is below a switching cutoff (given on the x axis), and a lower level of theory [MP2/aDZ, MP2/aTZ, or CCSD(T)/CBS(a[TQ]Z; δ:aDZ)] is used for the remaining dimers (up to 30 Å). The inset figure illustrates the use of a high level of theory for the closer dimers (black area) and a lower level of theory for the remaining dimers (orange area). As shown here and discussed in the text, CCSD(T)/CBS(a[TQ]Z; δ:aDZ) is essentially an exact match to our highest-level data from CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) by about 6 Å and is, therefore, used as the reference method for dimers beyond 8 Å. The graph illustrates that MP2 is an effective substitute for CCSD(T)/CBS at long range and that accuracy vs computational cost can be effectively tuned via the switching cutoff.

CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) values rapidly diminish with distance. This indicates that MP2 with either basis set may become a suitable replacement for CCSD(T)/CBS in the evaluation of distant dimer contributions. Indeed, as seen in Table III, the total MP2 lattice energy contribution of the medium-range dimers is −8.10 (aDZ) or −8.14 (aTZ) kJ mol^{-1}, compared to a reference value of −6.64 kJ mol^{-1}, or an error of about −1.5 kJ mol^{-1}. Such an error may be entirely acceptable in many applications (and could be reduced by redefining "short range" to a slightly larger cutoff). For the long-range contributions, MP2/aTZ becomes slightly more accurate than MP2/aDZ, yielding −1.40 and −1.46 kJ mol^{-1}, respectively, vs a reference value of −1.05 kJ mol^{-1} (errors under one-half kJ mol^{-1}). In the very long range, the reference CCSD(T)/CBS value is only −0.11 kJ mol^{-1}, and the MP2/aTZ value of −0.14 kJ mol^{-1} agrees very well. The MP2/aDZ estimate of −0.21 kJ mol^{-1} agrees well in an absolute sense, but the relative error is disappointing over this range. The sum of long range and very long range exhibits an error of −0.38 kJ mol^{-1} for MP2/aTZ and −0.51 kJ mol^{-1} for MP2/aDZ.

Given the rapid decrease in the magnitude of the lattice energy contributions as a function of intermolecular distance and the fairly good quality of MP2 estimates, at least at medium range or longer, we turn to consider possible multi-layered approaches aimed at reducing the cost of computing highly accurate crystal lattice energies. Figure 3 explores a multi-layered approach using a cutoff scheme to switch from a highly accurate level of theory, used for the short-range, to a more efficient one, employed for the long-range. This is illustrated by the sketch in the figure, with the black sphere representing the accurate-method domain and the orange shell depicting the efficient-method domain. The radius of the black sphere corresponds to the x axis in Fig. 3, whereas the radius of the orange sphere is fixed and would correspond to the largest *dimer cutoff* considered here, 30 Å.

In Fig. 3, the horizontal black-dotted line represents the total two-body contribution as computed using the best reference CCSD(T)/CBS data we have for the full set of dimers, i.e., CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) for $R \leq 8$ Å, and CCSD(T)/CBS(a[TQ]Z; δ:aDZ) $8 < R \leq 30$ Å. All other point successions in Fig. 3 represent the error incurred by computing the total two-body contribution as an admixture composed by taking short-range contributions from our best CCSD(T)/CBS reference estimates until a certain switching cutoff distance and employing a more efficient method for the remaining dimers. Therefore, for example, the first orange circle from the left represents the total two-body error when no interactions are computed with CCSD(T)/CBS and the full set of 420 dimers is computed with MP2/aDZ. In that case, the radius of the black sphere is minimal and includes no dimer.

The figure demonstrates the utility of a multi-layered approach for computing the two-body contributions to the lattice energy. Therefore, as long as CCSD(T)/CBS estimates are used for short-range contacts, accurate lattice energies can be obtained even when MP2 is used for medium- and long-range contacts. The accuracy of the hybrid approach can be tuned by changing the switchover distance, after which CCSD(T)/CBS is replaced by MP2. Past about 5 Å, errors in this composite approach reduce to less than 1 kJ mol^{-1}, and by 8 Å, the errors become −0.5 kJ mol^{-1} (aDZ) or −0.4 kJ mol^{-1} (aTZ) (as also seen in Table III). Errors of a few tenths of 1 kJ mol^{-1} should be negligible in many applications, and they may be further reduced by increasing the switchover distance. Switching to MP2 beyond 8 Å would result in only 17 out of 420 dimers being computed with CCSD(T)/CBS (see Table III), drastically reducing computational costs.

C. Three-body interactions

Table IV presents an analysis of the three-body interactions, split by distance domain, similar to that of Table III for two-body interactions. During analysis, we found that the energetic contributions of the trimers correlated somewhat better with the geometric mean of the three closest-contact intermolecular distances than with the "trimer cutoff" (the maximum of those three distances), so our analysis in this section uses the geometric mean instead. Note that trimers that would have a geometric mean intermolecular distance of >15 Å but a trimer cutoff distance of 15 Å (e.g., a trimer forming a triangle with one short side and two long sides of >15 Å) are not included in our analysis.

For trimers, the energy differences are calculated with respect to CCSD(T)/CBS(a[TQ]Z; δ:aDZ), the most accurate method generally employed here for trimers (however, see Sec. IV E below for even higher-level computations restricted to the first four trimers only). Although MP2 provides reasonable results for dimers (especially at long range), it is not as reliable for trimers, as it lacks any account of three-body dispersion (which would require the correlated motion of three electrons—MP2 includes only two-electron correlation). Therefore, we have considered molecular Axilrod–Teller–Muto (ATM) corrections to MP2 to approximately account for three-body dispersion.

The total three-body contribution to the lattice energy is an order of magnitude smaller than that of two-body interactions, and it is positive (making the crystal lattice slightly less stable compared to the two-body estimates). The much smaller importance of the three-body contribution and the much larger number of trimers compared to dimers (1977 vs 420, for our choices of cutoff distances) highlight the desirability of multi-level approaches that use different levels of theory for different N-mers and/or distances.

The CCSD(T)/CBS(a[TQ]Z; δ:aDZ) results in Table IV indicate that the short-range set (0–3 Å), which includes the four coordination-shell trimers, accounts for 2.66 kJ mol^{-1} (73%) of the total three-body contribution; the intermediate-range (3–6 Å), including 186 trimers, accounts for 0.76 kJ mol^{-1} (21%); and the

TABLE IV. Summary of total three-body contributions, split by separation domain, for each method employed in this study. All energies are in kJ mol^{-1}. Cutoff distances, which are the geometric mean of the three closest-contact intermolecular separations, are expressed in Å. Only trimers with a "trimer cutoff" of <15 Å are included in the data (see text). Percentages reflect the contribution to the overall three-body energy computed at the corresponding level of theory. Accumulated energy differences with respect to the reference method for three-body interactions, CCSD(T)/CBS(a[TQ]Z; δ:aDZ), are indicated in parentheses for each range.

Range	Total $R \le 15$	Short $R \le 3$	Medium $3 < R \le 6$	Long $6 < R \le 15$
CCSD(T)/CBS(a[TQ]Z; δ:aDZ)	3.63 100.00% Ref.	2.66 73.3% Ref.	0.76 21.1% Ref.	0.20 5.6% Ref.
MP2/aTZ	0.53 100.0% (−3.10)	0.26 48.9% (−2.40)	0.24 44.7% (−0.53)	0.03 6.4% (−0.17)
MP2/aDZ	0.43 100.0% (−3.19)	0.16 37.1% (−2.50)	0.26 59.7% (−0.51)	0.01 3.2% (−0.19)
MP2/aTZ + ATM	3.33 100.0% (−0.29)	1.68 50.4% (−0.98)	1.33 40.0% (0.57)	0.32 9.6% (0.12)
MP2/aDZ + ATM	3.24 100.0% (−0.39)	1.58 48.9% (−1.08)	1.35 41.9% (0.59)	0.30 9.2% (0.10)
ATM	2.80 100.0% (−0.82)	1.42 50.7% (−1.24)	1.10 39.1% (0.33)	0.29 10.2% (0.08)
Number of unique trimers	1977 100.0%	4 0.2%	186 9.4%	1787 90.4%

long-range (6–15 Å), which includes the remaining 1787 trimers, accounts for only 0.20 kJ mol^{-1} (6%) of the total three-body contribution.

Table IV demonstrates the unsuitability of plain MP2 for computing the three-body contributions; MP2 misses more than 3 kJ mol^{-1} of the three-body term, which is only 3.63 kJ mol^{-1} in size [CCSD(T)/CBS]. Most of the error comes from the four short-range trimers, but MP2 misses 0.70 kJ mol^{-1} from the remaining trimers. In all cases, the MP2 values for the three-body contribution are not positive enough (three-body dispersion, missing in MP2, tends to be positive). MP2 estimates of the three-body term become essentially zero for the long-range trimers, indicating that three-body induction contributions have died out at this range. However, CCSD(T)/CBS(a[TQ]Z; δ:aDZ) results provide another 0.20 kJ mol^{-1} for the long-range bin, indicating that three-body dispersion remains a small contributor at these distances. Pure ATM is able to provide a good estimate for the long-range bin (0.29 kJ mol^{-1}), and MP2+ATM is also good at these distances because the MP2 contribution here is very nearly zero. However, in the intermediate range, ATM overestimates the three-body dispersion (1.10 kJ mol^{-1} vs the reference value of 0.76 kJ mol^{-1}), and the three-body induction contributions captured by MP2 lend about another quarter of one kJ mol^{-1}, making MP2+ATM overestimate the CCSD(T)/CBS result by about 0.6 kJ mol^{-1} for the intermediate range. For the short-range bin, as just mentioned, the MP2 values are significantly underestimated vs the benchmark values due to the lack of three-body dispersion. MP2/aDZ and MP2/aTZ values are in nearly perfect agreement with each other across the distance bins in Table IV, except for the short-range bin, where they differ by a minor 0.1 kJ mol^{-1}.

ATM underestimates the three-body dispersion contribution at a short range, resulting in MP2+ATM underestimating the CCSD(T)/CBS contribution at a short range by about 1 kJ mol^{-1}. Overall, the underestimation of three-body dispersion at short range and the overestimation of three-body dispersion at medium and long range by ATM mean that MP2+ATM provides fortuitously accurate results overall, 3.24 kJ mol^{-1} (MP2/aDZ+ATM) or 3.33 kJ mol^{-1} (MP2/aTZ+ATM), vs the CCSD(T)/CBS value of 3.63 kJ mol^{-1}. Plain ATM is not quite as good, at 2.80 kJ mol^{-1}, due to its lack of three-body induction and short-range three-body exchange-repulsion (in sum, worth about 0.4–0.5 kJ mol^{-1} according to MP2).

Figure 4 presents how the total three-body contribution to the lattice energy accumulates when adding an increasing number of trimers (similarly to Fig. 1 for the two-body contribution). Figure 4 demonstrates the convergence of the three-body energy within half a kJ mol^{-1} of the asymptotic value by a trimer geometric mean of 5 Å and to a quarter of a kJ mol^{-1} by 6.5 Å for all methods. This suggests that approximate treatments could truncate the list of trimers to include them well before the 15 Å trimer cutoff employed here. Of course, longer cutoffs would probably be required for crystals of small organic molecules with net dipole moments, as the induction contribution would then be expected to have a larger range.[42]

Figure 4 again illustrates that in crystalline benzene, three-body dispersion effects are much more important than three-body induction effects because the pure ATM model is a much better match to the benchmark CCSD(T)/CBS data than MP2. Because there is no double-counting among effects between ATM and MP2, the MP2+ATM results are an even closer match to the benchmark CCSD(T)/CBS data (MP2+ATM includes a treatment of both three-body induction and three-body dispersion). However, note that the very close asymptotic agreement between MP2+ATM and CCSD(T)/CBS results from a closing of the gap between the MP2+ATM and CCSD(T)/CBS curves as one progresses to more distant trimers. In principle, these curves should remain separated but parallel if MP2+ATM becomes nearly exact for long-range trimers. The gap closes because ATM (with the parameters used) overestimates three-body dispersion at long range (as previously noted in the discussion of Table IV).

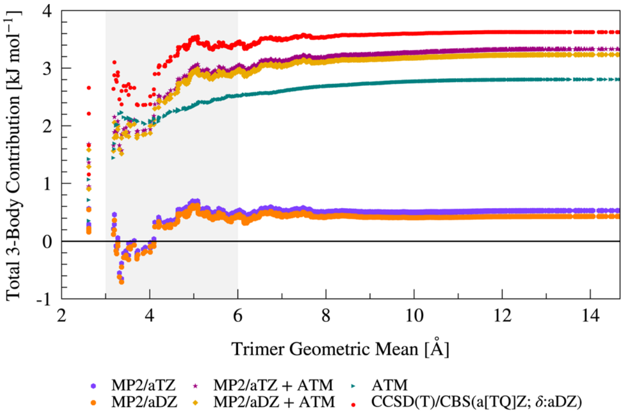

FIG. 4. Accumulation of three-body contributions to the lattice energy as a function of the geometric mean of the three closest-contact inter-monomer distances. The three-body contribution is very small compared to the two-body contribution for benzene. Nonetheless, three-body dispersion effects are clearly reflected in the difference between CCSD(T) and MP2, which is about 3 kJ mol^{-1} and can be approximately accounted for by using an Axilrod–Teller–Muto correction. Only trimers with a "trimer cutoff" of 15 Å are included in the data (see text).

We now turn to examine the biggest individual contributions and their errors. Figure 5 presents an analysis of the contribution of each trimer across different levels of theory. The figure includes the first 26 trimers sorted by the "trimer cutoff" distance. The results are plotted as errors vs the CCSD(T)/CBS(a[TQ]Z; δ:aDZ) reference data. For an individual trimer, the MP2/aDZ and MP2/aTZ results are nearly identical, suggesting very weak basis set dependence. This is consistent with the previous observation of very similar total three-body energies from MP2/aDZ and MP2/aTZ. The first four trimers, which constitute the short-range domain, show larger energy differences across all levels of theory compared to the remaining trimers, analogous to what was observed for dimers. These four configurations are the main source of errors (see also Table IV), with an overall overbinding tendency.

The convergence of the lattice energy contribution errors in Fig. 5 is remarkably quick for MP2. Most mid-range trimers show negligible contribution errors for MP2, with only a few trimers with larger errors in the ~0.1–0.2 kJ mol^{-1} range. The individual errors have mixed signs, which reduces the additive error. The molecular ATM correction helps improve the accuracy of the individual MP2 contributions in all cases but one (by a negligible margin).

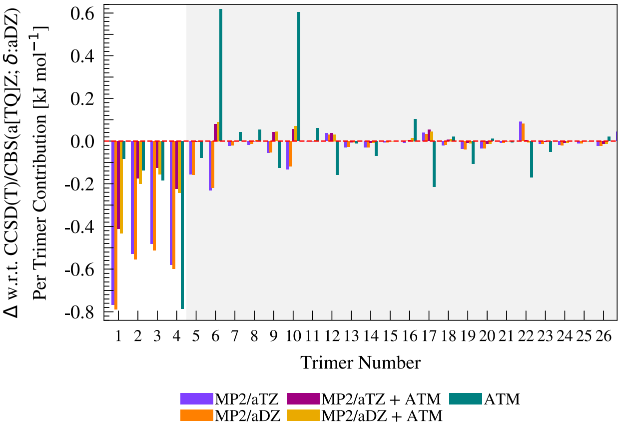

FIG. 5. Energy difference with respect to CCSD(T)/CBS(a[TQ]Z; δ:aDZ) per trimer contribution for the first 26 trimers sorted by the "trimer cutoff" (longest of the three intermolecular closest-contact distances). The first ten trimers account for the largest differences per trimer contribution with respect to the reference method. The differences per trimer contribution decrease to less than 0.1 kJ mol^{-1} after the tenth structure. In addition, the ATM correction cuts the energy differences of the first ten trimers by about half or more.

We stated previously that the molecular ATM (teal) achieves a better overall representation of three-body interactions than pure MP2. Figure 5 provides additional insight and shows that, on a trimer-by-trimer basis, the ATM contributions do not match those of CCSD(T)/CBS any better than MP2. Indeed, the ATM produces a few notable outliers. The simple molecular ATM model used here is suitable at long range but is not expected to be as reliable at shorter range. One reason is that it depends only on intermolecular COM distances and angles and, therefore, cannot distinguish between trimers with different monomer orientations but equivalent COM distances. There are such trimers in crystalline benzene, as discussed in some detail in the CrystaLattE implementation article.[21] Figure 5 and Table IV suggest that the molecular ATM model becomes more appropriate at larger distances. Another reason is that the simple ATM approach used here was not damped at short range. In a concurrent study,[43] we examined ATM corrections for benzene, carbon dioxide, and triazine and found that MP2 plus atomic ATM corrections with Tang–Toennies damping works more effectively for short-range trimers (although errors can remain a significant fraction of one kJ mol^{-1} for the accumulated short-range contribution).

We switch now to the investigation of multi-layered approaches to improve the efficiency of accurate crystal lattice energy calculations. In a similar fashion to the analysis of mixed approaches to the two-body energy in Fig. 3, Fig. 6 shows the errors in the three-body energy vs our benchmark value if the benchmark method is used for closer trimers but a more approximate method (MP2, ATM, or MP2+ATM) is used for more distant trimers. Figure 6 supports the idea that multi-level approaches enable dramatic efficiency improvements with little compromise in accuracy. Even when we switch to approximate methods immediately after the closest four trimers are treated with the reference method, the error is already within 0.7 kJ mol^{-1} (as also seen in Table IV).

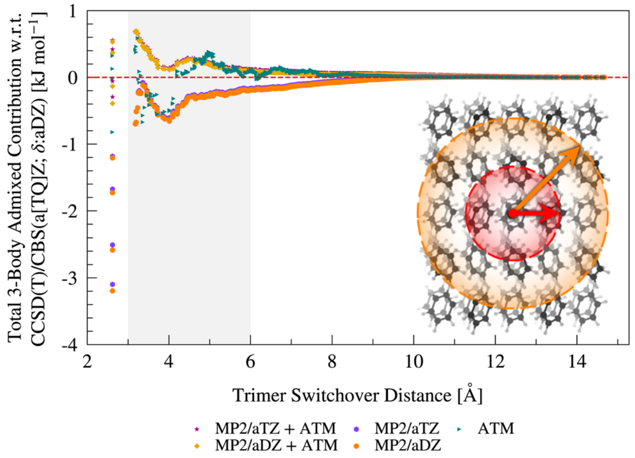

FIG. 6. Errors in the total three-body contribution to the lattice energy when using a two-layer approximation employing two different methods. A high level of theory [here, CCSD(T)/CBS(a[TQ]Z; δ:aDZ)] is used when the geometric mean of the three inter-monomer distances is below a certain switching cutoff (given on the x axis), and a lower level of theory (MP2, ATM, or MP2+ATM) is used for the remaining trimers (up to 15 Å). The inset figure illustrates the use of a high level of theory for the closer trimers (red area) and a lower level of theory for the more distant trimers (orange area). The graph illustrates that only the closer trimers need to be computed using CCSD(T)/CBS, and more distant trimers may be computed using the approximate methods considered, with errors of only about 0.5 kJ mol^{-1} after a switchover distance of 4 Å.

Due to the particulars of the error cancellation, admixing CCSD(T)/CBS with ATM-corrected MP2 results in slightly larger errors than the admixture of CCSD(T)/CBS with plain MP2 for switchover distances around 3 Å, but MP2+ATM becomes a superior method for the more distant trimers after a switchover distance of around 3.5 Å. As the switchover distance is increased, the MP2+ATM errors continue to decrease to below 0.2 kJ mol^{-1} a little before 4 Å, briefly rise back to about 0.3 kJ mol^{-1} around 4.5 Å, and then smoothly decrease toward zero. Using plain MP2 for the more distant trimers causes increasing errors from 3.5 to 4 Å (where they reach about −0.6 kJ mol^{-1}) and then decreasing errors thereafter. Using only ATM for the more distant trimers yields more erratic error behavior as a function of the switchover distance.

Overall, Fig. 6 supports the notion that a model like MP2+ATM, which can approximately describe three-body induction, exchange, and dispersion, can be a very effective replacement for the much more time-consuming CCSD(T) method for more distant trimers. For a switchover distance of 4 Å or greater, the errors induced by replacing CCSD(T) with MP2+ATM are very small, 0.3 kJ mol^{-1} or less, and if necessary, they can be reduced by using larger switchover distances.

D. Four-body interactions

To our knowledge, four-body interactions in crystalline benzene have only been studied before by Chan and co-workers.[13] Their study performed explicitly-correlated local CCSD(T) computations (including the T0 approximation for triples)[44] on the single tetramer with the closest intermolecular contacts (this symmetry-unique tetramer has eight symmetry-redundant replicas). This tetramer

TABLE V. Summary of total four-body contributions, split by separation domain, for each method employed in this study. All energies are in kJ mol^{-1}. Cutoff distances are expressed in Å. Percentages reflect the contribution to the overall four-body energy computed at the corresponding level of theory.

Range	Total $R \leq 10$	Short $R \leq 4$	Medium $4 < R \leq 8$	Long $8 < R \leq 10$
CCSD(T)/CBS (a[TQ]Z; δ:aDZ)	N/A	−0.11	N/A	N/A
MP2/aTZ	0.14	−0.05	0.20	−0.01
MP2/aDZ	0.14	−0.05	0.19	−0.01
Number of tetramers	1332 100.0%	1 0.1%	459 34.5%	872 65.5%

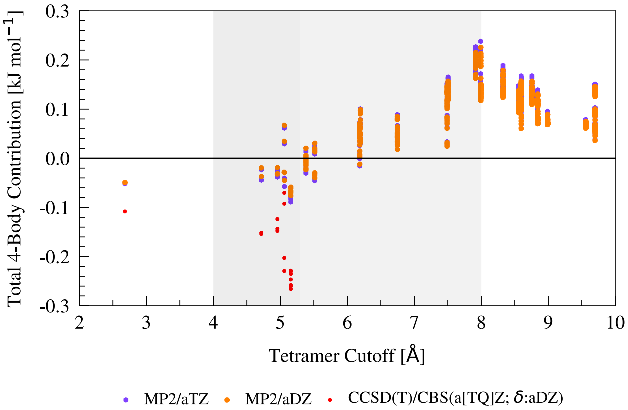

FIG. 7. Accumulation of four-body contributions to the lattice energy as a function of the longest of the four monomer–monomer closest-contact distances (the *tetramer cutoff*). Due to crystal symmetry, several tetramers have identical tetramer cutoffs, leading to vertical stacks of points. The additional, darker shaded area in this figure helps delimit the portion of the mid-range region fully covered with CCSD(T)/CBS(a[TQ]Z; δ:aDZ) (<5.3 Å) from that where only spot checks were performed with this level of theory. The asymptotic value, computed with MP2, fluctuates in the range depicted by the plot, which is practically negligible. The total four-body contribution is extremely small compared to the total three-body contribution and vanishes when compared to the total two-body contribution for benzene. Tetramers containing stacked conformations of benzene are slightly more negative when computed CCSD(T) vs MP2.

makes a four-body contribution of −0.38 kJ mol^{-1} to the crystal lattice energy. This contribution was almost canceled out by their estimate of four-electron correlation effects in the crystal, +0.36 kJ mol^{-1}. This suggests that both of these effects could be neglected without incurring significant errors. Nonetheless, the extent to which four-body interactions could contribute remains a bit uncertain without performing post-Hartree–Fock computations on more than the nearest-neighbor tetramer and discarding the T0 approximation for triple excitations. We address both of these issues here.

In Table V, we report total four-body contributions, computed with MP2, up to a tetramer cutoff of 10 Å (resulting in 1332 tetramers). There is only one tetramer that falls into the short-range domain. The mid-range contribution is of the opposite sign and comparable in order of magnitude to the short-range. The long range contribution is negligible.

Due to the prohibitive cost of computing the total four-body contribution with CCSD(T)/CBS(a[TQ]Z; δ:aDZ), the tetramer cutoff employed for this level of theory was reduced to 5.3 Å. Some additional tetramers were computed as spot checks at longer separations. We computed the four-body contribution of 24 *symmetry-unique* tetramers (which account for 180 symmetry-redundant structures) with CCSD(T)/CBS(a[TQ]Z; δ:aDZ).

Figure 7 presents the accumulation of the total four-body contribution to the lattice energy when considering an increasing value for the tetramer cutoff (similarly to Fig. 1). The scale of the y axis in this figure is an order of magnitude smaller than that of trimers (Fig. 4), highlighting how remarkably small the total four-body contribution is. According to these results, the effect of the basis set size is minimal. Additionally, the variations of the accumulated four-body contribution at several cutoffs (the span of the distribution of points at a given cutoff) show that, at 10 Å, the total energy is converged to a variation smaller than 0.1 kJ mol^{-1}.

In lieu of a cost-prohibitive comprehensive analysis of errors vs CCSD(T)/CBS according to distance domain, in Fig. 8, we present MP2/aDZ and MP2/aTZ errors for the lattice energy contributions of several selected tetramers. The figure includes the first 17 tetramers, spanning the short-range domain and a little over a quarter of the mid-range domain (up to a tetramer cutoff of 5.3 Å), and seven additional randomly chosen tetramers, which are distributed among the mid- and long-range regions (5.3–10 Å). The results are plotted using CCSD(T)/CBS(a[TQ]Z; δ:aDZ) as references (red dots in Fig. 7). For the closest tetramer (the only one falling in the short range domain), MP2 in either basis set gives errors of less than −0.06 kJ mol^{-1}, which agrees very well in an absolute sense to our CCSD(T) estimate, and the errors rapidly decrease with increasing distance. As was the case for trimers, basis set size effects for individual tetramer contributions are minimal, with the triple-ζ

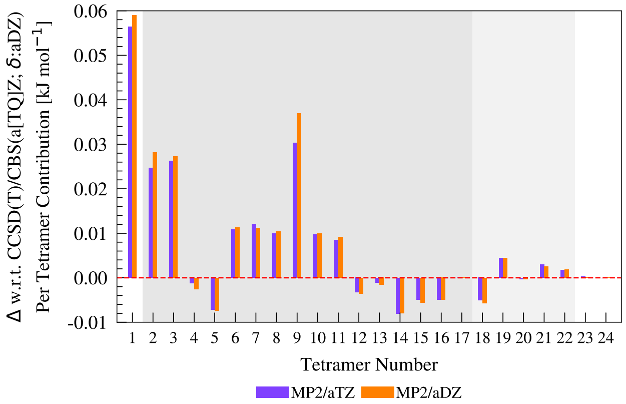

FIG. 8. Energy difference with respect to CCSD(T)/CBS(a[TQ]Z; δ:aDZ) per tetramer contribution. The figure includes the first 17 tetramers up to a tetramer cutoff of 5.3 Å and seven additional tetramers distributed along the mid- and long-range, from 5.3 to 10 Å. The differences per tetramer contribution decrease to extraordinarily small numbers at longer ranges. The only coordination-shell tetramer shows the largest energy difference, almost twice that of the next largest one. The differences per tetramer contribution are of mixed signs and one order of magnitude smaller than those of trimers.

basis appearing slightly better in most cases. Even among these 24 tetramers, individual errors of mixed signs can be found, suggesting that the overall accumulated error in the four-body contribution to the lattice energy is small.

Due to the limited tetramer cutoff employed for CCSD(T)/CBS(a[TQ]Z; δ:aDZ), we omit an admixed contribution plot, like Figs. 3 and 6. However, if we repeat the multi-layered scheme, adding up the individual contributions of the 17 tetramers at the CCSD(T)/CBS(a[TQ]Z; δ:aDZ) level of theory up to a switching cutoff of 5.3 Å and appending the remaining 1315 contributions, up to 10 Å, computed with MP2/aTZ, we obtain a remarkably small total four-body admixed contribution of -0.03 kJ mol^{-1}. If instead of MP2/aTZ, we use MP2/aDZ, that admixed contribution would be an essentially identical -0.02 kJ mol^{-1}. Our estimate is very small in magnitude and suggests that tetramers do not contribute significantly to the lattice energy of crystalline benzene, even if one desires a high-accuracy result.

E. Coupled cluster lattice energy of crystalline benzene

We present the benchmark-level results for the two-, three-, and four-body contributions to the lattice energy of crystalline benzene in Table VI. Compared to the three-body results discussed earlier, here we have replaced values for the four closest trimers with an even higher level of theory, CCSD(T)/CBS(a[Q5]Z; δ:aTZ) (using FNO's with a tight occupation number cutoff of 10^{-8}). The updated contribution from these four trimers is 2.61 kJ mol^{-1}, compared to 2.66 kJ mol^{-1} at the CCSD(T)/CBS(a[TQ]Z; δ:aDZ) level of theory (see Table IV). This very minor change is pleasing and validates the use of CCSD(T)/CBS(a[TQ]Z; δ:aDZ) for the remaining trimers. By analogy to our analysis of two-body terms, we expect CCSD(T)/CBS(a[TQ]Z; δ:aDZ) to be an even better approximation to CCSD(T)/CBS(a[Q5]Z; δ:aTZ) for the more distant trimers.

For benzene, the N-body expansion converges very rapidly, with each N-body contribution decreasing by an order of magnitude (or more) from the previous one. While the great majority of the lattice energy comes from two-body interactions, three-body terms are not totally negligible for high-accuracy studies: they contribute 3.57 kJ mol^{-1}, or 7% of the total lattice energy. These findings seem reasonable given that benzene lacks a permanent dipole moment, so one would expect polarization terms to be small, and these would normally be a large contributor to three- and four-body terms. However, the crystal packs closely enough that three-body dispersion is not negligible. Four-body effects, estimated at -0.03 kJ mol^{-1}, are negligible for crystalline benzene.

With these very accurate results in hand, it is worthwhile to compare them to some of the best previous results from the literature. To our knowledge, the most accurate previous two-body energy comes from the study of Chan and co-workers,[13] who estimated the Hartree–Fock CBS limit and added electron correlation via explicitly-correlated CCSD(T)-F12a and an aug-cc-pVTZ basis set. The total two-body HF/CBS contribution was 21.36 kJ mol^{-1}. Electron correlation contributions were computed for dimers up to a center-of-mass separation of 11 Å, contributing -77.30 kJ mol^{-1}. A core correlation correction, computed for nearest-neighbor dimers using MP2/aug-cc-pwCVQZ, yielded -0.58 kJ mol^{-1}. As mentioned previously, they also attempted to estimate the effect of quadruple excitations, although this estimate involved the too-small 6-31G basis set, a large number of frozen orbitals, and simple scaling schemes that attempted to correct for the small basis set and large frozen core; this yielded an additional 0.36 kJ mol^{-1}. Despite these approximations, we are unable to improve upon this quadruple excitation estimate in the current work. Finally, they utilized existing asymptotic long-range corrections (from Ref. 9) to estimate dispersion contributions from dimers beyond 11 Å, based on dynamic polarizabilities computed via density functional theory; this provided an additional -1.48 kJ mol^{-1}. The sum of all these two-body contributions is -57.64 kJ mol^{-1}. Our best two-body estimate,

TABLE VI. Benchmark-level results for the lattice energy of crystalline benzene using a many-body approach. Contributions to the lattice energy are given for each non-additive N-body term. Only symmetry-unique structures are counted. Cutoff distances are in Å, and energies are in kJ mol^{-1}.[a]

N	Level of theory	Cutoff	Structures	Total contribution
2	CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ)	$R \le 8$	17	−56.39
2	CCSD(T)/CBS(a[TQ]Z; δ:aDZ)	$8 < R \le 30$	403	−1.16
2	Total	$R \le 30$	420	−57.55
3	CCSD(T)/CBS(a[Q5]Z; δ:aTZ)	$R \le 3$	4	2.61
3	CCSD(T)/CBS(a[TQ]Z; δ:aDZ)	$3 < R \le 15$	1973	0.96
3	Total	$R \le 15$	1977	3.57
4	CCSD(T)/CBS(a[TQ]Z; δ:aDZ)	$R \le 5.3$	17	−0.23
	MP2/aTZ	$5.3 < R \le 10$	1315	0.20
4	Total	$R \le 10$	1332	−0.03
	Total			−54.01

[a] For dimers and tetramers, the cutoff distances refer to the "dimer cutoff" and "tetramer cutoff" criteria, respectively. For trimers, we used the geometric mean of the intermolecular closest-contact distances as in the trimer analysis.

−57.55 kJ mol^{-1}, is in excellent agreement with this previous best result, especially given numerous technical differences between our theoretical approaches. Because we did not include a four-electron correction, a more direct comparison would be through three-electron corrections only; that comparison is −57.55 (present value) vs −58.00 kJ mol^{-1} (Chan's value without four-electron correction), or a difference of 0.45 kJ mol^{-1}. This difference of about a half of 1 kJ mol^{-1} remains quite reasonable and is consistent with Chan's goal of computing the lattice energy with an accuracy of less than 1 kJ mol^{-1}. Nevertheless, given that this is half of the desired error bars, it is perhaps worthwhile to trace the origins of the difference.

To do so, we compared our CCSD(T)/CBS(a[Q5]Z; δ:aTZ) focal-point dimer energies to the HF/CBS + CCSD(T)-F12a/aug-cc-pVTZ correlation energies for all dimers computed by Chan and co-workers (up to a COM distance of 11 Å). These approaches should be quite comparable.[45] There are a couple of worrying discrepancies of ~0.4 kJ mol^{-1} among the first four dimers, but perhaps fortuitously, the differences appear with opposite signs and thus cancel. Indeed, all the small differences cancel by the time all dimers up to 11 Å (COM separation) are considered, and the two-body energies are −55.97 kJ mol^{-1} for CCSD(T)/CBS(a[Q5]Z; δ:aTZ) and −55.96 kJ mol^{-1} for HF/CBS + CCSD(T)-F12a/aug-cc-pVTZ correlation. Therefore, the discrepancy comes from other sources than the base electronic structure method. One discrepancy arises from the estimate of core-valence electron correlation. In the present study, we promoted the theoretical approach from CCSD(T)/CBS(a[Q5]Z; δ:aTZ) to CCSD(T)/CBS(aC[Q5]Z; δ:aCTZ) to account for core-valence correlation (i.e., moved from the aug-cc-pVXZ to the aug-cc-pwCVXZ basis set and unfroze core electrons). This gives us a total contribution of −56.04 kJ mol^{-1} for the dimers up to 11 Å COM distance, or a core-valence correction of −56.04 − (−55.97) = −0.07 kJ mol^{-1}. By contrast, Chan and co-workers computed core-valence corrections using only the MP2 level of theory (with an aug-cc-pwCVQZ basis) and only for the first four dimers. Indeed, we have found that only the first three or four dimers are relevant for the core-valence correlation, but apparently, MP2 significantly overestimates the effect: Chan and co-workers obtained a value of −0.58 kJ mol^{-1} for the core-valence lattice energy contribution, for an error of −0.51 kJ mol^{-1}. This is quite close to the previously noted difference of 0.45 kJ mol^{-1} between our final two-body lattice energies, and so the primary source of discrepancy is the treatment of the core-valence correlation. The MP2 approximation used by Chan and co-workers to estimate the core-valence contribution—a fairly standard approach—is not particularly reliable in the case of crystalline benzene.

We may also directly compare our results for longer-range dimers. Chan and co-workers treated only 15 dimers with coupled-cluster theory (up to 11 Å COM distance), whereas we included 420 (up to minimum intermonomer separations of 30 Å). On the other hand, Chan and co-workers used an asymptotic formula from Ref. 9 to estimate the dispersion for a very large number of dimers beyond 11 Å COM separation. These two rather different approaches give essentially the same results for dimer contributions beyond the 11 Å COM distance. We obtain −57.55 (total reference CCSD(T) result) − (−56.04) (reference results up to 11 Å) = −1.51 kJ mol^{-1} for the two-body contributions beyond 11 Å COM. Chan's result is −1.48 kJ mol^{-1}. This excellent agreement may be partly fortuitous, but otherwise, it suggests that the asymptotic estimates from Szalewicz and co-workers[9] are very effective for their purpose and that they work quite well as close as 11 Å COM separation.

For three-body terms, the present study improves substantially on the methodology used in our 2014 examination.[12] In that work, we used CCSD(T)/aug-cc-pVDZ for only the 366 closest (symmetry-redundant) trimers, and then we used MP2/aug-cc-pVDZ for an additional 7750 (symmetry-redundant) trimers, plus the asymptotic limit for ATM corrections for all trimers beyond the first 366 treated by CCSD(T). Here, we improve the electronic structure method to focal-point CCSD(T)/CBS(a[TQ]Z; δ:aDZ), which should be able to capture induction effects more accurately thanks to the MP2/CBS extrapolation inherent in the focal-point scheme used. We also go out to 1787 *symmetry-unique* trimers using this CCSD(T)/CBS approach, a dramatic expansion in the number of trimers treated at a high level. Our earlier estimate of the three-body contribution to the lattice energy was 0.89 kcal mol^{-1}, or 3.72 kJ mol^{-1}.[12] The present result of 3.57 kJ mol^{-1} is in remarkably good agreement (likely somewhat fortuitously so) and supports our earlier assertion[12] that MP2+ATM is a good substitute for CCSD(T) beyond about the first 366 symmetry-redundant trimers. The closer trimers have a degeneracy factor of 6, so these 366 symmetry-redundant trimers would equate to about 61 symmetry-unique trimers. Consulting Table IV, we see that this would put us about one third of the way through the 186 trimers in the "medium" ($3 < R \le 6$) range, in which MP2/aDZ+ATM exhibits an accumulated error of 0.59 kJ mol^{-1}. Therefore, the very good agreement between the present three-body estimate and our 2014 estimate suggests that the majority of the error in the "medium" range in Table IV comes from the first portion of the range and could be eliminated by employing CCSD(T)/CBS partway into the medium range trimers. Indeed, this is exactly what was shown in the discussion of Fig. 6.

We may also compare our result to the study by Chan and co-workers for three-body terms.[13] For three-body terms, that study used estimates of the Hartree–Fock CBS limit and computed correlation energies using an orbital-specific virtual (OSV) local explicitly-correlated CCSD(T)-F12a approach with the T0 approximation for triples,[44] i.e., OSV-LCCSD(T0)-F12a.[46,47] This approach was used for 96 symmetry-unique trimers with an average center-of-mass distance of less than 9 Å. Compared to canonical CCSD-F12a, Chan and co-workers found the largest local correlation error among the four closest trimers to be −0.13 kJ mol^{-1}. Unfortunately, due to computational expense, the effect of the T0 approximation was not tested. The nearest-neighbor trimers contributed 1.95 kJ mol^{-1}, and the total for 96 trimers was 3.09 kJ mol^{-1}. The core correlation was assessed via MP2/aug-cc-pwCVQZ and was found to contribute −0.01 kJ mol^{-1} for the nearest neighbor trimers. A very large number of trimers beyond the average COM distance of 9 Å were modeled by an ATM model of three-body dispersion taken from Ref. 9 and contributed an additional 0.35 kJ mol^{-1} for an overall three-body contribution of 3.43 kJ mol^{-1} to the lattice energy.

For the four closest trimers, as noted earlier, we computed three-body energies using the computationally very expensive CCSD(T)/CBS(a[Q5]Z; δ:aTZ) approach, which should be similar in quality to the OSV-LCCSD(T0)-F12a/aug-cc-pVTZ approach of Chan and co-workers, but having the advantage that we do not invoke the T0 approximation. In both approaches, the effect of triple excitations is evaluated in the aug-cc-pVTZ basis set. For the four closest trimers, we obtain a three-body contribution of 2.61 kJ mol^{-1}, which is 0.66 kJ mol^{-1} greater than the 1.95 kJ mol^{-1} value of Chan and co-workers, suggesting that the T0 approximation can cause errors in three-body energies that are significant if one wishes to obtain sub-kJ mol^{-1} accuracy. If we sum all the trimers within 9 Å COM distance, using CCSD(T)/CBS(a[TQ]Z; δ:aDZ) beyond the first four, we obtain 3.41 vs 3.09 kJ mol^{-1} for Chan and co-workers. Our trimers beyond 9 Å COM distance provide an additional 0.29 kJ mol^{-1}, vs the long-range asymptotic estimate of 0.35 kJ mol^{-1} of Chan and co-workers. Our total three-body value, 3.57 kJ mol^{-1}, agrees very well with the final value of 3.43 kJ mol^{-1} from Chan and co-workers due to a cancellation of errors between the nearest-neighbor trimers and the remaining short-range trimers.

Finally, for the four-body terms, we can again compare them to the work of Chan and co-workers.[13] Their study considered only the closest tetramer. They used an approach similar to the one they used for trimers described earlier, except that electron correlation was computed using the non-augmented cc-pVTZ basis set. The total four-body lattice energy contribution of this tetramer was −0.11 kJ mol^{-1} for Hartree–Fock, plus a correlation contribution of −0.28 kJ mol^{-1}, for a total four-body energy of −0.38 kJ mol^{-1}. For the closest tetramer, we obtain a four-body contribution to the lattice energy of −0.11 kJ mol^{-1} at the focal-point CCSD(T)/CBS(a[TQ]Z; δ:aDZ) level of theory. It is not immediately obvious which result should be more accurate for the tetramer. Our approach has the advantage of using diffuse functions and not the T0 approximation, whose effects on four-body energies are unknown. On the other hand, the approach of Chan and co-workers utilizes a triple-ζ basis for the triples contribution (compared to the double-ζ basis for ours) and explicit correlation techniques for singles and doubles (although our [TQ] extrapolation of MP2 should be about as effective at capturing these effects). The two results are in fairly good agreement in an absolute sense (they both agree that the contribution is very small). Our results on further trimers indicate that the contributions become slightly positive at larger ranges, and then they die off quickly so that the overall four-body sum is negligible (−0.03 kJ mol^{-1}). By considering only the first tetramer, it appears that the study of Chan and co-workers has slightly overestimated the magnitude of four-body contributions by a few tenths of a kJ mol^{-1}.

Overall, our best estimate of the overall crystal lattice energy of benzene using the 138 K crystal structure is −54.01 kJ mol^{-1}, which compares to the final estimate of −54.58 ± 0.76 kJ mol^{-1} at this same structure by Chan and co-workers. As just discussed earlier, our approaches are overall similar but differ in several details, such as the number of N-mers treated directly with quantum mechanics, the treatment of core-valence correlation, and the use of the T0 approximation for three- and four-body terms by Chan and co-workers. Therefore, the overall level of agreement is quite good. Directly comparin to experiment is challenging, as discussed by Chan and co-workers[13] and others.[27,48,49] The experiment measures sublimation energies rather than lattice energies, so to estimate the lattice energy from the experiment, one must correct for contributions from zero point energy as well as finite-temperature thermal effects from phonon modes. Taking into account these zero-point and finite-temperature enthalpy corrections, Chan and co-workers estimated a 0 K experimentally deduced lattice energy of −55.3 ± 2.2 kJ mol^{-1}. If we utilize their estimate of a −1.32 ± 0.1 kJ mol^{-1} correction for changing the geometry from the 138 K structure to the hypothetical 0 K structure, our resulting lattice energy of −55.33 kJ mol^{-1} is in excellent agreement. We expect the largest source of error in our computations to be the lack of quadruple excitations in the coupled-cluster treatment, which should make a small contribution to the closest dimers and perhaps trimers.

V. CONCLUSIONS

The primary objective of the current study was to investigate "multi-level" schemes for accurate yet efficient computation of lattice energies of molecular crystals using wavefunction methods and the molecular many-body expansion, with benzene as the initial case study. Based on our previous work on non-covalent interactions in general and crystalline benzene in particular,[4,12,21] we hypothesized that CCSD(T) complete-basis-set (CBS) estimates for short-range dimers and trimers, plus MP2 estimates for long-range dimers and trimers (possibly supplemented by Axilrod–Teller–Muto corrections for three-body dispersion), would yield a very accurate lattice energy while providing a tremendous savings in computational cost compared to CCSD(T)/CBS computations on all significant dimers and trimers. To carefully assess such schemes, we performed CCSD(T)/CBS computations on many more dimers, trimers, and tetramers from the crystal than had previously been considered at this level. We investigated the convergence of two- and three-body terms with respect to intermolecular distances to assess what distance cutoffs would be appropriate for various target accuracies. We also investigated at what distances MP2 or MP2+ATM corrections became reliable replacements for the full CCSD(T)/CBS estimates. The basis set dependence of the CCSD(T) and MP2 values and the importance of core-valence contributions were also considered.

Our final best estimate of the CCSD(T)/CBS lattice energy of crystalline benzene at the 138 K experimental crystal structure, −54.01 kJ mol^{-1}, agrees very well with the lattice energy of −54.58 kJ mol^{-1} reported by Chan and co-workers[13] using overall similar methods that differ in several details; this agreement is partly fortuitous and results from several partially canceling differences. When adjusted for the difference in crystal structure between 138 K and the hypothetical 0 K structure (a correction of −1.32 kJ mol^{-1}),[13] our estimate of −55.33 kJ mol^{-1} is in excellent agreement with the best estimate[13] of the 0 K experimental lattice energy, −55.3 ± 2.2 kJ mol^{-1}. The agreement is partially fortuitous because we were unable to add any corrections for quadruple excitations in the coupled-cluster treatment.

As hypothesized, MP2(+ATM) does appear to be a viable replacement for CCSD(T)/CBS for medium- to long-range two- and three-body contributions. The MP2 computations are dramatically faster, yielding approximately CCSD(T)/CBS-quality results for the lattice energy at a greatly reduced computational cost. For medium- and long-range interactions, MP2 provides nearly identical results whether an aDZ or aTZ basis is used, so we recommend aDZ for additional computational speed. If we replace CCSD(T)/CBS with MP2/aDZ for all dimers with a closest-contact separation of 4 Å or more and with MP2/aDZ+ATM for all trimers where the geometric mean of three closest-contact intermolecular distances is more than

3 Å, we incur an error vs our best estimate of −1.3 kJ mol^{-1}, which is quite accurate (well within "chemical accuracy") and would require CCSD(T)/CBS computations on only three closest dimers and four closest trimers; however, we also note that this result benefits from some error cancellation between positive three-body errors and negative two-body errors. This error level is likely acceptable for most applications, but it could be further reduced by moving the switching cutoff to larger values. Including more dimers with CCSD(T) is more beneficial for reducing the error than including more trimers with CCSD(T), simply because the two-body contribution is much larger than the three-body contribution.

CCSD(T)/CBS was also used to evaluate several close tetramers, and many additional tetramers were evaluated with MP2. MP2 tends to underestimate the lattice energy contribution of the tetramers compared to CCSD(T), but nevertheless, the total four-body contribution appears to be negligible.

Although approximate methods appear to be very suitable for more distant dimers and trimers, we were surprised to see that nailing down the lattice energy contributions of the very closest dimers and trimers is apparently even more difficult than previously realized.

Our baseline benchmark method in this study is a focal-point approach based on MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ computations to extrapolate to the MP2/CBS limit, with a correction for higher-order electron correlation computed as the difference between CCSD(T) and MP2 in the smaller aug-cc-pVDZ basis. This composite approach was denoted CCSD(T)/CBS(a[TQ]Z; δ:aDZ) for short. For dimers at short and medium ranges, we considered even larger basis sets for the MP2 and CCSD(T) computations and the importance of core-valence correlation. Using the aug-cc-pwCVXZ basis sets, core-valence computations analogous to our baseline benchmark level demonstrate a core-valence correction of −0.28 kJ mol^{-1} for the lattice energy contribution of the closest three dimers. If we instead increase all the basis sets by one ζ level, we estimate a basis set incompleteness correction of +0.50 kJ mol^{-1}. Adding these two effects lets us estimate the simultaneous effect of basis set incompleteness and core-valence correlation, yielding a combined correction of +0.22 kJ mol^{-1}. However, we were able to explicitly compute results using core-valence correlation and larger-ζ basis sets, yielding an actual combined lattice energy correction of +0.41 kJ mol^{-1}, demonstrating that these two effects are not quite addictive.

Fortunately, we found that the baseline method CCSD(T)/CBS(a[TQ]Z; δ:aDZ) is extremely accurate for all but the closest three dimers and for computing the three-body contributions of trimers at all distances when compared to more accurate CCSD(T)/CBS estimates.

In the course of this work, we have created a very large number of dimer, trimer, and tetramer CCSD(T)/CBS computations that can serve as excellent benchmarks for evaluating other strategies for the accurate and efficient computation of lattice energies. In particular, the trimers should be valuable in evaluating methods meant to capture three-body dispersion.

This study has also produced sufficient high-quality data to improve upon previous theoretical estimates of the lattice energy of crystalline benzene and its two-, three-, and four-body components. Comparing the literature values,[13] we discovered that the common practice of estimating core-valence effects using the simpler MP2 method is not very accurate for the lattice energy of crystalline benzene and can lead to errors of about −0.5 kJ mol^{-1}. In addition, the popular T0 approximation[44] in local treatments of electron correlation appears to give errors of ~0.7 kJ mol^{-1} for the three-body contribution to the lattice energy of the first four trimers. The present estimate of the four-body contribution includes an expanded number of tetramers at the CCSD(T)/CBS level and additional tetramers using MP2 and is even smaller (−0.03 kJ mol^{-1}) compared to the previous literature value (−0.38 kJ mol^{-1}).

Although the present work demonstrates that MP2(+ATM) can effectively replace CCSD(T)/CBS for more distant dimers and trimers in a many-body expansion for crystalline benzene, it remains to be seen if other approximate methods are even more effective for this purpose and how lattice energies converge and how multi-level approximations behave for different molecular crystals. In particular, crystals of molecules with permanent dipole moments are expected to have larger three- and four-body contributions and to require longer cutoff distances. Such questions are part of an ongoing investigation in our laboratory, and initial results are reported in Refs. 43 and 50.

SUPPLEMENTARY MATERIAL

See the supplementary material for non-additive n-body interaction energies and lattice energy contributions of the dimers, trimers, and tetramers for crystalline benzene with all the levels of theory employed in this study. For trimers, we found faster convergence of the three-body contribution of the lattice energy with respect to the geometric mean of the intermolecular separations than the "trimer cutoff." However, the supplementary information also provides alternative versions of Figs. 4 and 6 based on the trimer cutoff criterion. The supplementary material also includes input files containing the geometries of all dimers and trimers considered in atomic units.

ACKNOWLEDGMENTS

C.H.B. wants to acknowledge Dr. Thomas Darden for the enriching discussions about crystal lattice energies and *ab initio* methods, Dr. Mehmet Belgin and Dr. Semir Sarajlic for their dedicated support with high-performance computing, and Dr. Dominic A. Sirianni for his assistance with hardware management. This work was supported in part by the U.S. National Science Foundation through Grant No. CHE-1955940. The initial development of the CrystaLattE software was supported by a grant from the U.S. Department of Defense HPCMP Applications Software Initiative (HASI) and by the U.S. National Science Foundation through a Sustainable Software Infrastructure grant (Grant No. ACI-1449723). This research is supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology and by the Georgia Tech Hive cluster, funded by the U.S. National Science Foundation through Grant No. MRI-1828187.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Carlos H. Borca: Data curation (lead); Formal analysis (lead); Investigation (lead); Software (lead); Visualization (lead); Writing – original draft (lead). **Zachary L. Glick**: Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Visualization (supporting). **Derek P. Metcalf**: Data curation (supporting); Investigation (supporting). **Lori A. Burns**: Methodology (supporting); Software (supporting). **C. David Sherrill**: Conceptualization (lead); Data curation (supporting); Formal analysis (supporting); Funding acquisition (lead); Investigation (supporting); Methodology (lead); Project administration (lead); Software (supporting); Supervision (lead); Writing – original draft (supporting); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available with the article and its supplementary material.

REFERENCES

[1] W. B. Schweizer and J. D. Dunitz, J. Chem. Theory Comput. **2**, 288 (2006).

[2] D. Hankins, J. W. Moskowitz, and F. H. Stillinger, J. Chem. Phys. **53**, 4544 (1970).

[3] H. Stoll and H. Preuß, Theor. Chim. Acta **46**, 11 (1977).

[4] A. L. Ringer and C. D. Sherrill, Chem. - Eur. J. **14**, 2542 (2008).

[5] P. Hobza, H. L. Selzle, and E. W. Schlag, J. Phys. Chem. **100**, 18790 (1996).

[6] M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill, J. Am. Chem. Soc. **124**, 10887 (2002).

[7] C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009).

[8] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[9] R. Podeszwa, B. M. Rice, and K. Szalewicz, Phys. Rev. Lett. **101**, 115503 (2008).

[10] S. Wen and G. J. O. Beran, J. Chem. Theory Comput. **7**, 3733 (2011).

[11] O. A. von Lilienfeld and A. Tkatchenko, J. Chem. Phys. **132**, 234109 (2010).

[12] M. R. Kennedy, A. R. McDonald, A. E. DePrince, M. S. Marshall, R. Podeszwa, and C. D. Sherrill, J. Chem. Phys. **140**, 121104 (2014).

[13] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, Science **345**, 640 (2014).

[14] S. Hirata, J. Chem. Phys. **129**, 204104 (2008).

[15] P. J. Bygrave, N. L. Allan, and F. R. Manby, J. Chem. Phys. **137**, 164102 (2012).

[16] M. J. Gillan, D. Alfè, P. J. Bygrave, C. R. Taylor, and F. R. Manby, J. Chem. Phys. **139**, 114101 (2013).

[17] O. V. Shishkin, R. I. Zubatyuk, A. V. Maleev, and R. Boese, Struct. Chem. **25**, 1547 (2014).

[18] S. Hirata, K. Gilliard, X. He, J. Li, and O. Sode, Acc. Chem. Res. **47**, 2721 (2014).

[19] G. J. O. Beran, Chem. Rev. **116**, 5567 (2016).

[20] S. Mattsson, B. Paulus, F. A. Redeker, H. Beckers, S. Riedel, and C. Müller, Chem. - Eur. J. **25**, 3318–3324 (2019).

[21] C. H. Borca, B. W. Bakr, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **151**, 144103 (2019).

[22] D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Phys. **152**, 184108 (2020).

[23] B. M. Axilrod and E. Teller, J. Chem. Phys. **11**, 299 (1943).

[24] Y. Muto, J. Phys. Soc. Jpn. **17**, 629 (1943).

[25] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Theory Comput. **13**, 3185 (2017).

[26] G. E. Bacon, N. A. Curry, and S. A. Wilson, Proc. R. Soc. London, Ser. A **279**, 98 (1964).

[27] A. Otero-de-la-Roza and E. R. Johnson, J. Chem. Phys. **137**, 054103 (2012).

[28] A. L. P. Nguyen, T. G. Mason, B. D. Freeman, and E. I. Izgorodina, J. Comput. Chem. **42**, 248 (2020).

[29] M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).

[30] D. E. Bernholdt and R. J. Harrison, Chem. Phys. Lett. **250**, 477 (1996).

[31] A. E. DePrince and C. D. Sherrill, J. Chem. Theory Comput. **9**, 2687 (2013).

[32] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).

[33] D. Feller, J. Chem. Phys. **96**, 6104 (1992).

[34] A. L. L. East and W. D. Allen, J. Chem. Phys. **99**, 4638 (1993).

[35] A. G. Császár, W. D. Allen, and H. F. Schaefer, J. Chem. Phys. **108**, 9751 (1998).

[36] M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102 (2011).

[37] L. A. Burns, M. S. Marshall, and C. D. Sherrill, J. Chem. Theory Comput. **10**, 49 (2013).

[38] C. E. Warden, D. G. A. Smith, L. A. Burns, U. Bozkaya, and C. D. Sherrill, J. Chem. Phys. **152**, 124109 (2020).

[39] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).

[40] D. G. A. Smith, P. Jankowski, M. Slawik, H. A. Witek, and K. Patkowski, J. Chem. Theory Comput. **10**, 3140 (2014).

[41] R. M. Richard, K. U. Lao, and J. M. Herbert, Acc. Chem. Res. **47**, 2828 (2014).

[42] D. P. Metcalf, A. Smith, Z. L. Glick, and C. D. Sherrill, J. Chem. Phys. **157**, 084503 (2022).

[43] Y. Xie, Z. L. Glick, and C. D. Sherrill, J. Chem. Phys. **158**, 094110 (2023).

[44] M. Schütz and H.-J. Werner, Chem. Phys. Lett. **318**, 370 (2000).

[45] D. A. Sirianni, L. A. Burns, and C. D. Sherrill, J. Chem. Theory Comput. **13**, 86 (2017).

[46] J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, J. Chem. Phys. **136**, 144105 (2012).

[47] M. Schütz, J. Yang, G. K.-L. Chan, F. R. Manby, and H.-J. Werner, J. Chem. Phys. **138**, 054109 (2013).

[48] A. M. Reilly and A. Tkatchenko, J. Chem. Phys. **139**, 024705 (2013).

[49] G. A. Dolgonos, J. Hoja, and A. D. Boese, Phys. Chem. Chem. Phys. **21**, 24333 (2019).

[50] C. T. Sargent, D. P. Metcalf, Z. L. Glick, C. H. Borca, and C. D. Sherrill, J. Chem. Phys. **158**, 054112 (2023).