Convergence of the many-body expansion with respect to distance cutoffs in crystals of polar molecules: Acetic acid, formamide, and imidazole

Philip M. Nelson and C. David Sherrill[a)]

Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, USA

The many-body expansion (MBE), where one computes the total energy of a supersystem as the sum of the dimer, trimer, tetramer, etc. subsystems, provides a convenient approach to compute the lattice energies of molecular crystals. We investigate approximate methods for computing the non-additive three-body contributions to the crystal lattice energy of the polar molecules acetic acid, imidazole, and formamide, comparing to coupled-cluster singles, doubles, and perturbative triples [CCSD(T)] level benchmarks. Second-order Møller–Plesset perturbation theory MP2, if combined with a properly damped Axilrod-Teller-Muto (ATM) dispersion potential, displays excellent agreement with CCSD(T) at a substantially reduced cost. Errors between dispersion-corrected MP2 and CCSD(T) are less than 1 kJ mol^{-1} for all three crystals. However, the three-body energy requires quite large distance cutoffs to converge, up to 20 Å or more.

[a)]Electronic mail: sherrill@gatech.edu

I. INTRODUCTION

Drug molecules can crystallize in different polymorphic structures, each with different properties and stabilities.[1,2] Thus, successfully predicting and identifying the most stable polymorphs of a molecule is an important task in drug development.[3,4]

Recently, our group and others have applied the many-body expansion (MBE),[5–7] to compute the lattice energy of molecular crystals.[8–12] The MBE is commonly used to calculate the energy of molecular clusters; the energy of a molecular system is decomposed into a sum of the energies of all the constituent monomers, dimers, trimers, etc.:

$$E = \sum_i E_i + \sum_{i<j} \Delta E_{ij}^{(2)} + \sum_{i<j<k} \Delta E_{ijk}^{(3)} + \dots . \tag{1}$$

$\Delta E_{ij}^{(2)}$ is the two-body interaction energy,

$$\Delta E_{ij}^{(2)} = E_{ij} - E_i - E_j, \tag{2}$$

and $\Delta E_{ijk}^{(3)}$ is the non-additive three-body interaction energy,

$$\Delta E_{ijk}^{(3)} = E_{ijk} - (\Delta E_{ij} + \Delta E_{ik} + \Delta E_{jk}) - (E_i + E_j + E_k). \tag{3}$$

The expansion is exact if taken to completion, but in practice it is truncated at some lower order, typically trimers or tetramers, taking advantage of the observation that the higher-order terms are usually negligible. The MBE offers two computational advantages: (1) truncating the expansion at lower orders allows one to use high-scaling wavefunction methods that would be intractable on the whole supersystem and (2) computing the n-body terms is pleasantly parallel and can be distributed easily across computing resources. The MBE is thus a promising method to compute coupled-cluster energies of large supermolecular systems.

As a crystal is a periodic system, to apply the MBE to a crystal, one must first construct a finite supercell of the crystal lattice. The lattice energy of a crystal with one unique monomer in the asymmetric unit ($Z = 1$) can be computed through a modified MBE as a sum of the deformation energy of the unique monomer plus (non-additive) two-body, three-body, etc., interaction energies involving one fixed reference monomer i. The interaction energies are divided by appropriate factors to yield final results on a per-reference-molecule (or per mole of reference molecules) basis:[13]

$$E_{\text{lattice}} = \Delta E_{\text{def}} + \frac{1}{2}\sum_{j} E_{ij}^{(2)} + \frac{1}{3}\sum_{j<k} E_{ijk}^{(3)} + \ldots . \tag{4}$$

As many crystal systems have several low-energy polymorphs, all lying within a few kJ mol^{-1} of each other,[4,14] the crystal lattice energy must be computed as accurately as possible. The so-called "chemical accuracy" of 1 kcal mol^{-1} or 4 kJ mol^{-1} seems to be the bare minimum target accuracy for studies of crystal polymorphs, and indeed accuracy of $\sim$1 kJ mol^{-1} or better should be greatly preferable. To achieve such accuracy levels seems challenging, but may be possible with the MBE as discussed below. Certainly one requirement is that any constituent computations must be carried out at a high level of quantum mechanical theory. Another requirement is that the MBE treatment must be carried to sufficiently high order to converge the approximation within the desired target accuracy.

Successfully converging the many-body contributions also requires careful consideration of the cutoff distances used to generate the crystal supercell, which may need to be quite large to converge to the bulk, periodic limit. Recent studies on the benzene crystal[15,16] have established the success of using the MBE to compute coupled-cluster singles, doubles, and perturbative triples [CCSD(T)][17] lattice energies of molecular crystals up to sub-kJ mol^{-1} accuracy. For benzene, truncating the MBE at the trimer level was sufficient to converge the CCSD(T) lattice energy, which appears to be within 1 kJ mol^{-1} accuracy or better of estimates of the experimental 0 K lattice energy, when all appropriate corrections are applied, such as the zero-point vibrational energy, finite-temperature enthalpy corrections, and the change in geometry from 0 K to the experimental crystal structure at 138 K.[16]

However, the steep cost of coupled cluster methods prevents the application of said approach to larger molecules. In addition, the combinatorics of the many-body expansion, which can give a huge number of trimer configurations in a supercell, make obtaining converged, high-accuracy MBE results a nearly insurmountable task without a careful selection of approximate methods and screening protocols for computation of the higher-order bodies in the MBE. One approach is to screen out contributions to the MBE that are estimated to be negligible, for example dimers or trimers with large distances between the constituent monomers.[18–20] Another strategy to reduce the computational cost is to use faster, more approximate methods for the higher-body terms. Both approaches are particularly salient for

the higher-order terms in the MBE, which dominate the computational cost unless special steps are taken to mitigate the number of higher-body clusters included.

We have previously identified approximate methods to compute the two-body contributions accurately and efficiently, comparing to CCSD(T) benchmarks for the X23 dataset.[21] We have also studied the role of three-body dispersion in benzene, carbon dioxide, and triazine;[22] all three of these systems have monomers with zero dipole moment, such that one would expect three-body dispersion to dominate over three-body polarization due to the lack of dipole-driven polarization terms. We found that the Axilrod-Teller-Muto (ATM)[23,24] three-body dispersion formula performed best in recouping the three-body dispersion missing in MP2, compared to CCSD(T) benchmark values, provided an appropriate damping function is used; the Tang-Toennies[25] damping form worked particularly well. In benzene, the three-body dispersion [estimated as the difference between CCSD(T) and MP2] was 3.12 kJ mol^{-1}, compared to 0.51 kJ mol^{-1} for the three-body induction (estimated as the MP2 non-additive three-body contribution). The total three-body energy in all systems was converged by a maximum cutoff distance of about 12 Å for the closest interatomic distance between two monomers.

Klimeš and coworkers recently investigated lattice energies crystalline ethane, ethene, and acetylene using the MBE including dimers, trimers, and tetramers, comparing MP2 and the random phase approximation (RPA) to CCSD(T) benchmarks.[26] The non-additive three-body energies were converged to within a few tenths of a kJ mol^{-1} by 20 Å. However, neither MP2 nor RPA accurately reproduced CCSD(T) values for the non-additive three- and four-body energies. Other recent approaches to computing crystal lattice energies include periodic MP2,[27] periodic diffusion Monte Carlo,[28] and MBE results embedded in periodic boundary result.[29]

In this work, expanding upon our previous work on non-polar molecules,[22] we investigate approximate methods to compute the non-additive three-body contributions to the lattice energies of crystalline systems of three polar molecules: formamide, acetic acid, and imidazole. In contrast to the dispersion-dominated systems studied previously, the three-body energies in these systems will contain significant contributions from three-body polarization in addition to three-body dispersion. This, in turn, may mean that three-body contributions may converge much more slowly with respect to intermolecular distance than in the previous studies of non-polar molecules.

Here we present CCSD(T) benchmark values for the non-additive three-body contribution to the lattice energies of the molecules studied. There have been previous studies of various approximate methods vs CCSD(T) benchmark values for diverse sets of van der Waals trimers. Beran and co-workers presented[30] the 3B-69 test set, which includes three trimers from each of the crystals in the X23 data set.[31] Gordon and co-workers have presented the S22(3) test set,[32] which includes trimers generated from the S22 test set.[33] Ochieng and Patkowski have presented[34] a test set of 20 trimer geometries including various types of interactions such as π-π, anion-π, cation-π, etc.

Here our focus is on examining the convergence of the three-body contribution to the lattice energy with respect to the distances between monomers, and on the ability of approximate methods to reproduce the three-body energy accurately. In the MBE framework, this requires not only accurate estimates of the three-body energy in each constituent trimer, but also the absence of a systematic bias towards over- or under-binding, which would accumulate and spoil the accuracy of the overall three-body energy.

Non-additive polarization and exchange-repulsion effects are already captured with simple Hartree–Fock theory, and MP2 allows for corrections to these effects due to electron correlation.[35] However, MP2 does not include any modeling of non-additive three-body dispersion,[36] which in the Møller-Plesset perturbation theory framework requires at least MP3.[35,37] Thus, here we consider two different alternatives to CCSD(T) that are still capable of including three-body dispersion: MP2.5,[38] and Axilrod-Teller-Muto (ATM) models.[23,24] MP2.5 is a simple average of MP2 and MP3, and has been found to be more accurate than standard MP3 for intermolecular interactions.[30,38] In particular, MP2.5 shows good accuracy for non-additive three-body interactions in van der Waals trimers.[30] Alternatively, ATM models have a simple functional form (see below) based on triple-dipole dispersion in atom trimers, and have been found effective in modeling three-body interactions in van der Waals trimers when paired with supermolecular MP2.[16,22,39]

As expected, we do find slower convergence of the three-body contributions to the lattice energy with respect to the distances between the monomers, compared to our earlier study of non-polar molecules.[22] This slower convergence means that we need to include trimers with larger intermolecular separations to achieve any given target level of accuracy, and this leads to an explosion in the number of unique trimers to be included in the computation. Thus, we also explore strategies for screening out trimers with negligible contributions to

the three-body energy.

II. METHODS

Generation of Trimer Geometries

A development version of the CrystaLattE[40] program was used to generate all trimer geometries. CrystaLattE generates a supercell of the crystal geometry, then identifies all symmetry-unique dimers, trimers, etc., containing a reference monomer and satisfying a cutoff distance criterion.

The crystal structures for acetic acid (ACETAC01),[41] formamide (FORMAM02),[42] and imidazole (IMAZOL04)[43] were obtained from the Cambridge Structural Database (CSD).[44] For each structure, CrystaLattE was used to generate all symmetry-unique trimer geometries within a specified maximum monomer separation (15 or 20 Å) from a central, reference monomer. MP2 and MP2.5 results for all three crystals systems and CCSD(T) results for formamide were obtained out to the 20 Å cutoff. There were 48046, 30546, and 26039 symmetry-unique trimers generated for formamide, acetic acid, and imidazole, respectively using the 20 Å cutoff value. As the full CCSD(T) results for acetic acid and imidazole at these cutoff values were deemed unfeasible due to computational limitations, we generated all the symmetry-unique trimers within a 15 Å maximum monomer separation cutoff: 6253 for acetic acid and 5572 for imidazole.

The trimers can be classified using a number of geometric parameters. Here, we use the maximum intermonomer separation distance (which we refer to as R_{max} throughout), the minimum intermonomer separation distance (R_{min} throughout), as well as the geometric and harmonic means of the intermonomer separation distances. The intermonomer separation distance is defined as the closest interatomic distance between two monomers.

Wavefunction Methods

All computations used Psi4[45] version 1.8 and the aug-cc-pVXZ, X $\in$ {D,T,Q}, basis sets of Dunning and coworkers.[46,47] The density fitting and the frozen core approximations were utilized in all computations. SCF computations used the aug-cc-pVXZ-JKFIT auxiliary basis sets.[48] MPn and CCSD(T) computations used the aug-cc-pVXZ-RI auxiliary basis

sets.[49] SCF results were tightly converged to 10^{-10} Hartree.

The MP2 correlation energies were computed by using the two-point formula of Helgaker and coworkers[50] to extrapolate aug-cc-pVTZ and aug-cc-pVQZ correlation energies to the complete basis set (CBS) limit. The CCSD(T) results were obtained by using a focal-point approach[51] utilizing the difference in MP2 and CCSD(T) correlation energies in a smaller basis set, in this case aug-cc-pVDZ:

$$E_{\mathrm{CCSD(T)/CBS}} \approx E_{\mathrm{MP2/CBS}} + \left[E_{\mathrm{CCSD(T)/aDZ}} - E_{\mathrm{MP2/aDZ}}\right]. \tag{5}$$

Such approximations are known to be particularly effective for computing interaction energies.[52]

The MP2.5 correlation energy is defined as the mean of the MP2 and MP3 correlation energies.[38] The MP2.5 energies were extrapolated in a similar focal-point approach, using the same basis sets:

$$E_{\mathrm{MP2.5/CBS}} \approx E_{\mathrm{MP2/CBS}} + \left[E_{\mathrm{MP2.5/aDZ}} - E_{\mathrm{MP2/aDZ}}\right]. \tag{6}$$

The Boys-Bernardi counterpoise correction[53] was used in all computations, with the trimer basis set used to calculate all fragment energies for each trimer, as suggested by Wells and Wilson.[54]

Axilrod-Teller-Muto Dispersion

While MP2 is known to effectively treat many-body polarization effects, it lacks inclusion of many-body dispersion effects.[35] The Axilrod-Teller-Muto (ATM) three-body potential[23,24] may be used to correct the MP2 results for three-body dispersion:

$$U_{3b} = \sum_{A<B<C} C_9^{ABC} \frac{\left(1 + 3\cos\hat{A}\cos\hat{B}\cos\hat{C}\right)}{R_{AB}^3 R_{BC}^3 R_{AC}^3}, \tag{7}$$

where the indices A, B, C run over atoms, C_9^{ABC} are the dipole-dipole-dipole polarizability coefficients, R are the distances between atom centers, and $\hat{A}$ are the angles between atom centers. The C_9 coefficients can be approximated using the pairwise C_6 coefficients:[55]

$$C_9^{ABC} = \sqrt{C_6^{AB} C_6^{AC} C_6^{BC}}. \tag{8}$$

The ATM potential diverges at short interatomic distances and must be damped to ensure proper behavior. We used a Tang-Toennies (TT) damping function,[25] specifically, a product of three two-body TT damping functions appropriate for damping C_6 coefficients. In our previous tests of three-body dispersion in crystals of non-polar molecules, we found that this damping function performed best of those considered.[22]

$$f_9^{abc}(\beta) = f_6^{ab}(R_{ab}, \beta)\, f_6^{bc}(R_{bc}, \beta)\, f_6^{ca}(R_{ca}, \beta), \tag{9}$$

where $f_6^{ab}(R_{ab}, \beta)$ is a two-body Tang-Toennies damping function,

$$f_6(R, \beta) = 1 - \sum_{n=0}^{6} \left(\frac{(\beta R)^k}{k!} \right) \exp -\beta R, \tag{10}$$

and the β damping coefficient is given by:

$$\beta = -0.31(r_a^{\mathrm{vdW}} + r_b^{\mathrm{vdW}}) + 3.43. \tag{11}$$

The C_6 coefficients used to approximate the C_9 coefficients in the ATM potential were obtained using the DFTD4 program of Grimme and coworkers.[56,57] The damping coefficients for the Tang-Toennies functions were obtained from the work of von Lillienfeld and Tkatchenko, based on ab initio computations of vdW radii.[58]

For imidazole, the largest molecule considered here, an example CCSD(T)/CBS trimer computation takes 18.3 hours, while MP2/CBS takes 1.4 hours, on a compute node with an Intel Xeon 6226 processor (using 8 out of 24 cores to accommodate multiple jobs on a single node), a RAID0 array scratch disk, and 32 GB of RAM used.

III. RESULTS

The convergence of the three-body energy in the formamide crystal with respect to two geometric parameters, R_{max} and R_{min} discussed above, is shown in Figure 1a for MP2, MP2 corrected with undamped (MP2+ATM) and Tang-Toennies damped ATM dispersion [MP2+ATM(TT)], MP2.5, and CCSD(T). The MP2+ATM(TT) results are in excellent agreement with the CCSD(T) benchmarks, with a cumulative three-body error of only 0.05 kJ mol^{-1} at the maximum monomer separation of 20 Å. The R_{min} results show that the three-body energy is converged by a minimum monomer separation of about 12 Å. In contrast,

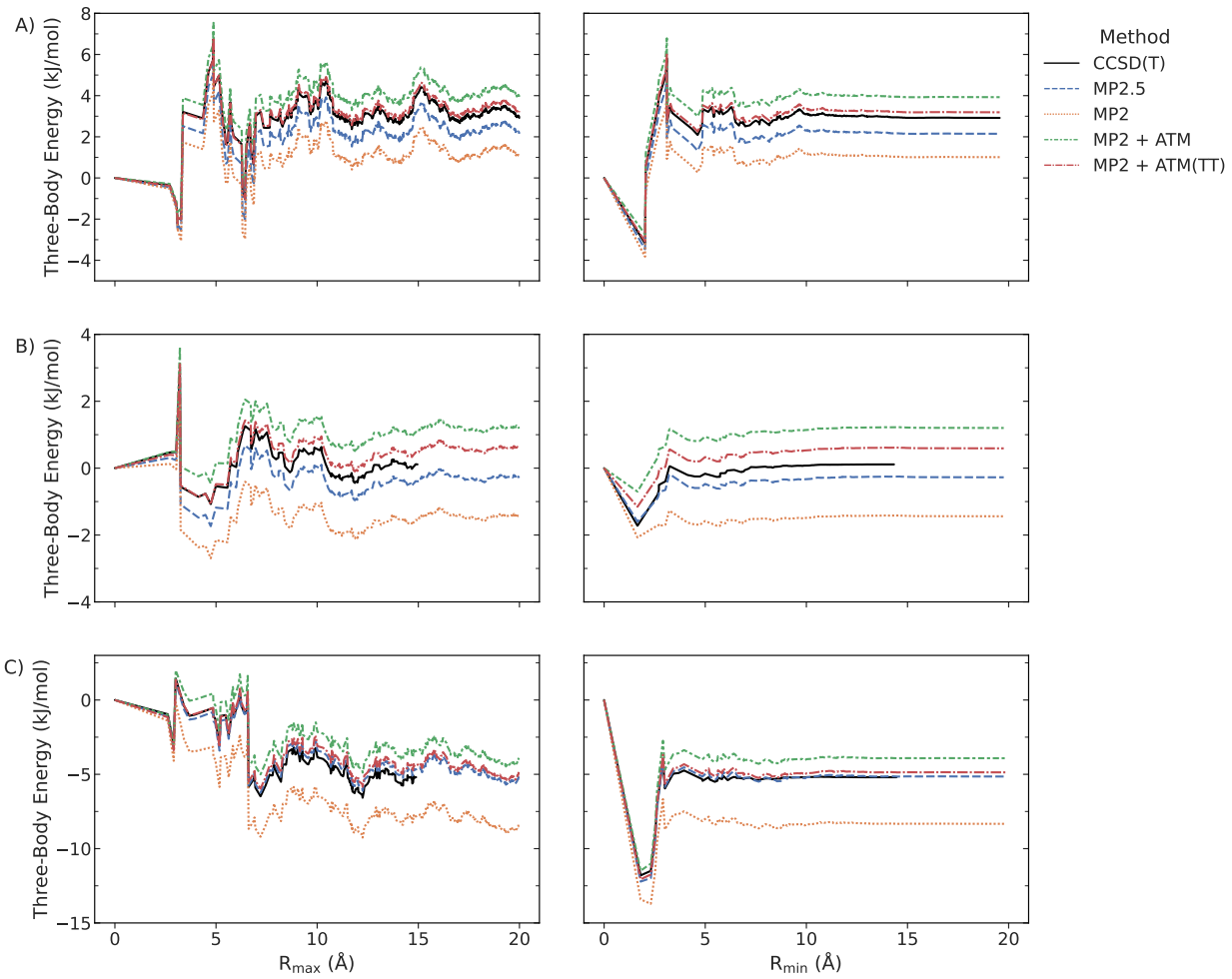

FIG. 1. Convergence of the non-additive three-body energy with respect to maximum monomer separation (R_{max}, left panels) and minimum monomer separation (R_{min}, right panels) in (a) formamide, (b) acetic acid, and (c) imidazole.

a large maximum monomer separation is needed to converge the total three-body energy, as can be seen in the R_{max} results. At most distances, the overall three-body contribution is positive for most methods. This is consistent with what one might expect for three-body dispersion, which tends to be positive. However, at shorter distances, the three-body contributions are often negative. This is due to favorable three-body polarization terms. Although two-body polarization is always negative, three-body polarization can be positive or negative. MP2, which includes three-body polarization but not three-body dispersion, shows both positive and negative three-body energies, depending on the cutoff distance and criterion used (R_{min} or R_{max}). In the right-hand panel for R_{min}, the cumulative three-

body contributions for MP2 are slightly positive for intermediate and larger cutoff distances (around 1-2 kJ mol^{-1}). When three-body dispersion is included via CCSD(T) or ATM, the three-body energies are shifted higher on the graph compared to the MP2 curves.

The convergence of the three-body energy in the acetic acid crystal is shown in Figure 1b. For acetic acid and imidazole (below), CCSD(T) trimers beyond an $R_{\max}$ cutoff of 15 Å were not computed due to the prohibitive cost. MP2+ATM(TT) results are again in good agreement with the CCSD(T) benchmarks, with a total error of 0.36 kJ mol^{-1} at the 15 Å cutoff.

The convergence of the three-body energy in the imidazole crystal is shown in Figure 1c. At the 15 Å cutoff distance, the error for MP2+ATM(TT) versus the CCSD(T) benchmark was 0.81 kJ mol^{-1}. The larger error correlates with the larger magnitude of the three-body energy in imidazole, around -4.87 kJ mol^{-1} vs 2.92 and 0.59 kJ mol^{-1} in formamide and acetic acid, respectively. As in the case of formamide, the cumulative three-body energy plots for acetic acid and imidazole tend to track the behavior of MP2, but they are shifted to higher (more positive or less negative) values, reflecting their inclusion of three-body dispersion absent in MP2. However, in the case of the latter two crystals, the MP2 three-body energies are uniformly negative.

As expected, without any ATM three-body dispersion correction, the MP2 three-body energies are significantly more negative than CCSD(T). MP2 with the undamped ATM dispersion correction overcorrects (too positive) as the dispersion energy becomes too repulsive for very close contacts without correctly damping the ATM potential. However, using the Tang-Toennies damped three-body dispersion results in excellent agreement with CCSD(T) results.

For the formamide and acetic acid crystals, the MP2.5 results underestimate the three-body contribution to the lattice energy, resulting in errors larger than those for MP2+ATM(TT). However, for imidazole, MP2.5 very slightly outperforms the MP2+ATM(TT) method. Given that MP2+ATM(TT) is a much faster method than MP2.5, and gives more accurate results for two out of three crystal systems here, it remains the preferred approximate method for computing three-body energies.

The frustratingly slow convergence of the three-body contribution to the many body energy for the systems presented here, substantially slower than the non-polar crystals studied in Ref. 22, highlights the importance of establishing approximate methods and screening

procedures for computing crystal lattice energies of polar molecules using the MBE. In contrast to the benzene, carbon dioxide, and triazine crystals,[22] the polar molecules studied here require longer ranges in both $R_{\max}$ and $R_{\min}$ to converge the three-body energy. For benzene for example, the three-body energy was converged by a $R_{\max}$ cutoff value of about 12 Å and an $R_{\min}$ cutoff value of 7 Å, while formamide is apparently still not well converged by an $R_{\max}$ cutoff of 20 Å. This can be attributed to the three-body polarization persisting at longer ranges than the three-body dispersion. Consistent with this interpretation, it is clear that convergence with respect to $R_{\min}$ and $R_{\max}$ is essentially the same for MP2 (lacking three-body dispersion but including three-body polarization) as it is for other methods.

It is obvious that one can apply a much stricter (shorter) $R_{\min}$ cutoff to the trimer configurations than $R_{\max}$. Unfortunately, however, screening out trimers by using an appropriate $R_{\min}$ cutoff does not eliminate enough to make large $R_{\max}$ cutoffs feasible. The configurations with a small $R_{\min}$ and large $R_{\max}$, corresponding to trimers with two close monomers and one far away, contribute non-negligibly to the total three-body energy. This results in a large number of configurations that cannot be screened out, and one must extend the supercell to include large $R_{\max}$ in order to converge the MBE energy. Similarly large cutoff distances were required to converge the two-body energy contributions for the same systems. The two-body energies for formamide, acetic acid, and imidazole were converged to within 0.5 kJ mol^{-1} of their asymptotic limit by 29.5, 17.9, and 22.0 Å, respectively.[13]

In Figure 2a, the formamide trimers are classified based on their $R_{\max}$ and $R_{\min}$ distances. For each bin, the number of trimers and contribution to the three-body energy are shown, using the CCSD(T) values. It can be seen that the closest trimers ($R_{\min} < 4$ Å and $R_{\max} < 10$ Å) contribute by far the most to the three-body energy, despite their comparatively very small number. The dramatic increase in trimers for large $R_{\max}$ values can clearly be seen. A cutoff value to minimize the number of trimers while keeping errors as small as possible thus requires a dual cutoff, leveraging both $R_{\min}$ and $R_{\max}$ values, so that the short-long trimers ($R_{\min} < 4$ Å and $R_{\max} > 10$ Å) are not neglected, while keeping the total number of long range trimers under control.

Figure 2a also shows the errors of MP2+ATM(TT) versus CCSD(T) benchmarks classified based on their geometric parameters. The errors versus the CCSD(T) benchmarks are largest for the trimers with $R_{\min} < 4$ Å and $R_{\max} < 10$ Å. In fact, the accumulated errors for all other trimers are essentially negligible. The former number on the order of just a few

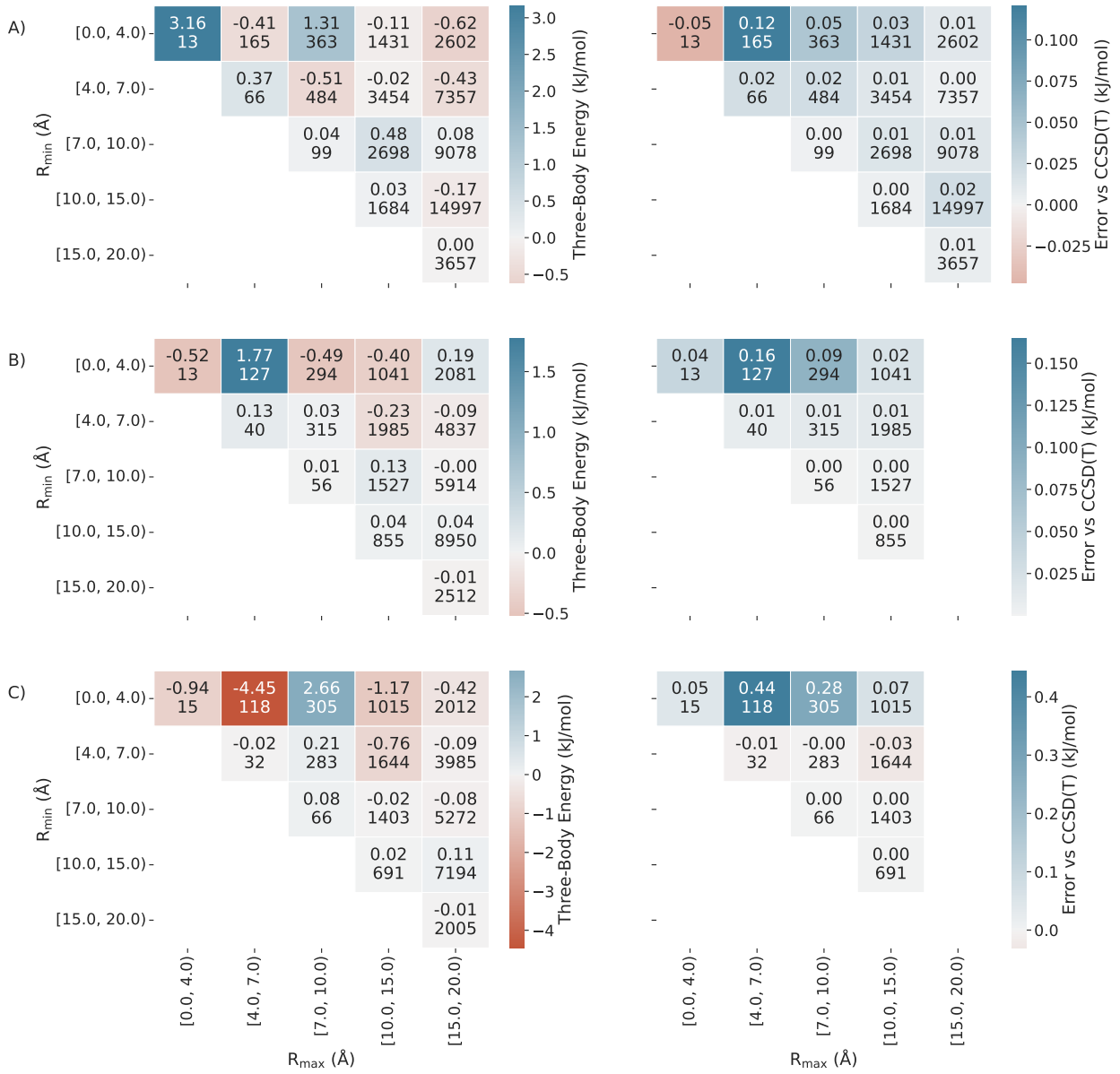

FIG. 2. Left side: contribution to the non-additive three-body lattice energy (top numbers) and number of unique trimers (bottom numbers) for each bin of trimers, partitioned according to the values of R_{min} and R_{max}. Right side: Similar figure, but energy values are now MP2+ATM(TT) error vs CCSD(T) benchmarks. Results provided for A) formamide, B) acetic acid, and C) imidazole. For acetic acid and imidazole, values on the left are CCSD(T) where available, otherwise MP2+ATM(TT). Energies in kJ mol^{-1}.

hundred trimers, making a multi-level approach, where one computes this subset of trimers with CCSD(T) and the remaining with MP2+ATM(TT), a feasible option for obtaining the

three-body energy of crystal systems within benchmark level accuracy. However, due to favorable error cancellation, in this particular case, even if all trimers are computed with MP2+ATM(TT), the total error is only 0.05 kJ mol^{-1}.

The geometric binning of trimers in the acetic acid and imidazole crystals is shown in Figures 2b and 2c. For both, the energy values shown are CCSD(T) where available ($R_{\max} < 15$ Å) and MP2+ATM(TT) elsewhere. The imidazole and acetic acid results are similar to those of formamide. The three-body energy is dominated by the closest trimers, yet close-long geometries still contribute significantly. However, the errors versus CCSD(T) are driven by the closest geometries.

If one computes the closest ($R_{\min} < 4$ Å and $R_{\max} < 10$ Å) geometries with CCSD(T) and the rest with MP2+ATM(TT), the resulting errors compared to the full CCSD(T) results are just 0.05 and 0.03 kJ mol^{-1} for acetic acid and imidazole, requiring only 434 out of 30546 and 438 out of 26039 trimers for each. While some error remains unknown in acetic acid and imidazole due to missing CCSD(T) benchmark values at longer ranges ($R_{\max} > 15$ Å), we expect these errors to be minimal.

A. Comparison of long-range screening methods

As mentioned above, a multi-level approach can help reduce computational cost while minimizing errors. However, as observed above, there are a large number of long-range trimers whose contributions to the three-body lattice energy are not negligible. We find it advisable to screen out the trimer configurations with negligible contributions to reduce the overall cost, which is overwhelmingly dominated by the growth of long range trimer configurations for large supercell sizes. In addition to the $R_{\min}$ parameter discussed above, we consider here using the geometric mean of the intermonomer distances R_{geom}, and the harmonic mean of the intermonomer distances R_{harm} as screening parameters. The convergence plot of the three-body energies of formamide, acetic acid, and imidazole with respect to the geometric mean R_{geom} and harmonic mean R_{harm} can be found in the Supplemental Information.

Figure 3 shows the errors incurred versus the number of trimers screened out for the three geometric screening parameters considered here: $R_{\min}$, R_{geom}, and R_{harm}. For all three crystals, a substantial number of trimers in the supercell can be screened out while keeping

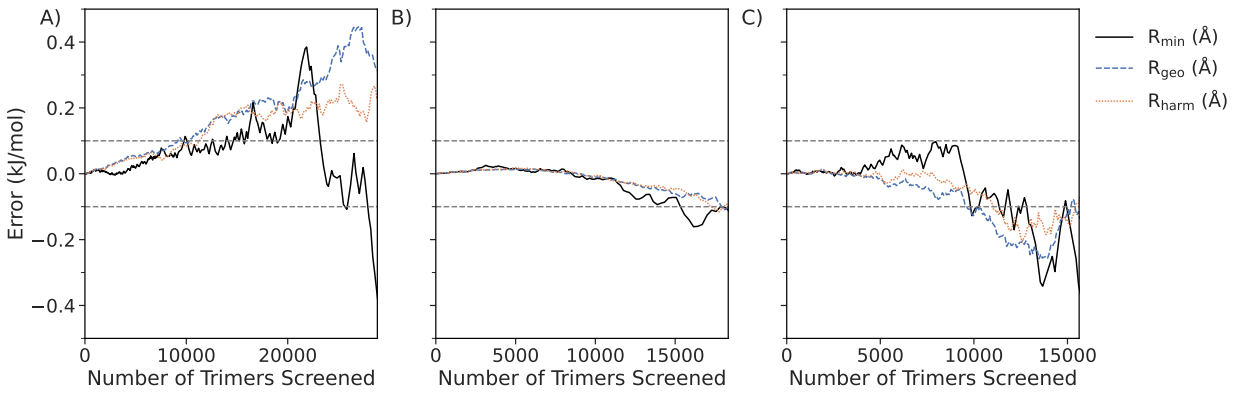

FIG. 3. Errors incurred versus number of trimers screened out for the R_{min}, R_{geom}, and R_{harm} screening parameters for the (a) formamide, (b) acetic acid, and (c) imidazole crystals. Errors are in kJ mol^{-1} at the MP2+ATM(TT) level of theory versus unscreened results at the same level of theory.

the errors negligible (under 0.1 kJ mol^{-1}). Until a large fraction of the trimers are discarded, all three parameters perform similarly. While the R_{geom} and R_{harm} parameters incur larger errors in the formamide crystal for similar numbers of trimer geometries discarded, they appear to give a more stable error with respect to number of trimers screened in all three crystals, suggesting they may be preferable over R_{min} as a general screening parameter.

B. Best estimates of the crystal lattice energies

Table I shows the best estimates of the two-body and three-body energies for the acetic acid, formamide, and imidazole crystals, in addition to the three non-polar crystals studied in Ref. 22, benzene, carbon dioxide, and triazine. The two-body energies are CCSD(T)/CBS level results from Ref. 21. The best estimate of the three-body energies is CCSD(T)/CBS where available, otherwise MP2+ATM(TT) results were used. To discriminate among polymorphs, which often lie very close to each other in energy, it is evidently necessary to include three-body terms in the lattice energies, which range in importance from 0.59 to -4.87 kJ mol^{-1} for crystals of the three polar molecules considered here, and 1.43 to 4.75 kJ mol^{-1} for the dispersion bound crystals. Computing the three-body energies with MP2 results in substantial errors in all cases, up to 3.46 kJ mol^{-1} for the imidazole crystal.

TABLE I. Best estimates for the two-body and three-body contributions to the crystal lattice energies of acetic acid, formamide, and imidazole, in kJ mol^{-1}. Two-body energies are CCSD(T)/CBS results from Ref. 21 and three-body energies for benzene, carbon dioxide, and triazine are from Ref. 22. Three-body energies are CCSD(T)/CBS where available, otherwise MP2+ATM(TT).

Crystal	2-body energy	3-body energy	
	Best estimate	Best estimate	MP2
Acetic acid	−75.51	0.59 (-0.8 %)	−1.44
Formamide	−76.76	2.92 (-4.0 %)	1.01
Imidazole	−93.30	-4.87 (5.0 %)	−8.33
Benzene	−57.99	3.63 (-6.7 %)	0.51
Carbon dioxide	−30.11	1.43 (-5.0 %)	0.42
Triazine	−58.36	4.75 (-8.9 %)	2.16

IV. CONCLUSION

We have computed the non-additive three-body crystal lattice contributions for crystalline formamide, acetic acid, and imidazole using CCSD(T), MP2, dispersion-corrected MP2, and MP2.5. Our results show that MP2 with Axilrod-Teller-Muto three-body dispersion can achieve excellent agreement with CCSD(T) benchmarks for the three-body contributions to the crystal lattice energy. The MP2.5 results, however, have larger errors compared to the CCSD(T) benchmarks, despite the greater computational cost.

Importantly, we find contributions from long-range trimer geometries beyond even 15 Å are apparently non-negligible for crystals of polar molecules. However, a substantial portion of the long-range contributions can be screened out while keeping errors small. A combination of long-range screening and approximate methods like MP2 corrected with a damped ATM dispersion potential can drastically reduce the computational cost of obtaining high accuracy three-body contributions to the crystal lattice energies. In addition, or alternatively, one could perform an inexpensive periodic boundary condition computation to fully capture many-body polarization effects, and the correct this baseline by higher-level computations on select dimers and trimers.

SUPPLEMENTAL MATERIAL

The supplemental material contains Cartesian coordinates and energetics for all trimers in the study, at each level of theory considered. Also provided is a figure of the convergence of the cumulative three-body energies with respect to geometric and harmonic mean cutoff criteria.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the U.S. National Science Foundation through Grant No. CHE-1955940.

DATA AVAILABILITY

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

REFERENCES

[1]J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, "Ritonavir: An extraordinary example of conformational polymorphism," Pharm. Res. **18**, 859–866 (2001).

[2]R. Censi and P. Di Martino, "Polymorph impact on the bioavailability and stability of poorly soluble drugs," Molecules **20**, 18759–18776 (2015).

[3]S. L. Price, "Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism," Acc. Chem. Res. **42**, 117–126 (2009).

[4]A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein, "Facts and fictions about polymorphism," Chem. Soc. Rev. **44**, 8619–8635 (2015).

[5]D. Hankins, J. W. Moskowitz, and F. H. Stillinger, "Water molecule interactions," J. Chem. Phys. **53**, 4544–4554 (1970).

[6]S. S. Xantheas, "Ab-initio studies of cyclic water clusters $(H_2O)_n$, $n = 1 - 6$. II. Analysis of many-body interactions," J. Chem. Phys. **100**, 7523–7534 (1994).

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

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

[9]P. J. Bygrave, N. L. Allan, and F. R. Manby, "The embedded many-body expansion for energetics of molecular crystals," J. Chem. Phys. **137**, 164102 (2012).

[10]M. J. Gillan, D. Alfe, P. J. Bygrave, C. R. Taylor, and F. R. Manby, "Energy benchmarks for water clusters and ice structures from an embedded many-body expansion," J. Chem. Phys. **139**, 114101 (2013).

[11]S. Hirata, K. Gilliard, X. He, J. Li, and O. Sode, "Ab initio molecular crystal structures, spectra, and phase diagrams," Acc. Chem. Res. **47**, 2721–2730 (2014).

[12]G. J. O. Beran, "Modeling polymorphic molecular crystals with electronic structure theory," Chem. Rev. **116**, 5567–5613 (2016).

[13]D. P. Metcalf, A. Smith, Z. L. Glick, and C. D. Sherrill, "Range-dependence of two-body intermolecular interactions and their energy components in molecular crystals," J. Chem. Phys. **157**, 084503 (2022).

[14]J. Nyman and G. M. Day, "Static and lattice vibrational energy differences between polymorphs," CrystEngComm **17**, 5154–5165 (2015).

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

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

[17]K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A 5th-order perturbation comparison of electron correlation theories," Chem. Phys. Lett. **157**, 479–483 (1989).

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

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

[20]D. R. Broderick and J. M. Herbert, "Scalable generalized screening for high-order terms in the many-body expansion: Algorithm, open-source implementation, and demonstration," J. Chem. Phys. **159**, 174801 (2023).

[21]C. T. Sargent, D. P. Metcalf, Z. L. Glick, C. H. Borca, and C. D. Sherrill, "Benchmarking two-body contributions to crystal lattice energies and a range-dependent assessment of approximate methods," J. Chem. Phys. **158**, 054112 (2023).

[22]Y. Xie, Z. L. Glick, and C. D. Sherrill, "Assessment of three-body dispersion models against coupled-cluster benchmarks for crystalline benzene, carbon dioxide, and triazine," J. Chem. Phys. **158**, 094110 (2023).

[23]B. M. Axilrod and E. Teller, "Interaction of the van der Waals type between three atoms," J. Chem. Phys. **11**, 299–300 (1943).

[24]Y. Muto, "Force between nonpolar molecules," J. Phys. Math. Soc. Japan **17**, 629–631 (1943).

[25]K. T. Tang and J. P. Toennies, "An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients," J. Chem. Phys. **80**, 3726–3741 (1984).

[26]K. N. Pham, M. Modrzejewski, and J. Klimeš, "Assessment of random phase approximation and second-order Møller–Plesset perturbation theory for many-body interactions in solid ethane, ethylene, and acetylene," J. Chem. Phys. **158**, 144119 (2023).

[27]Y. H. Liang, H. Ye, and T. C. Berkelbach, "Can spin-component scaled MP2 achieve kJ/mol accuracy for cohesive energies of molecular crystals?" J. Phys. Chem. Lett. **14**, 10435–10441 (2023).

[28]F. D. Pia, A. Zen, D. Alfè, and A. Michaelides, "How accurate are simulations and experiments for the lattice energies of molecular crystals?" (2024), arXiv:2402.13059 [cond-mat.mtrl-sci].

[29]J. Hoja, A. List, and A. D. Boese, "Multimer embedding approach for molecular crystals up to harmonic vibrational properties," J. Chem. Theory and Comput. **20**, 357–367 (2024).

[30]J. Rezac, Y. Huang, P. Hobza, and G. J. Beran, "Benchmark calculations of three-body intermolecular interactions and the performance of low-cost electronic structure methods," J. Chem. Theory Comput. **11**, 3065–3079 (2015).

[31]A. M. Reilly and A. Tkatchenko, "Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals," J. Chem. Phys. **139**, 024705 (2013).

[32]M. Alkan, P. Xu, and M. S. Gordon, "Many-body dispersion in molecular clusters," J. Phys. Chem. A **123**, 8406–8416 (2019).

[33]P. Jurečka, J. Šponer, J. Černý, and P. Hobza, "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," Phys. Chem. Chem. Phys. **8**, 1985–1993 (2006).

[34]S. A. Ochieng and K. Patkowski, "Accurate three-body noncovalent interactions: the insights from energy decomposition," Phys. Chem. Chem. Phys. **25**, 28621–28637 (2023).

[35]G. Chałasinski, M. Szczęśniak, and S. M. Cybulski, "Calculations of nonadditive effects by means of supermolecular møller-plesset perturbation theory approach: Ar_3 and Ar_4," J. Chem. Phys. **92**, 2481–2487 (1990).

[36]P. Xu, M. Alkan, and M. S. Gordon, "Many-body dispersion," Chem. Rev. **120**, 12343–12356 (2020).

[37]G. Chałasinski, M. M. Szczęśniak, and R. A. Kendall, "Supermolecular approach to many-body dispersion interactions in weak van der waals complexes: He, Ne, and Ar trimers," J. Chem. Phys. **101**, 8860–8869 (1994).

[38]M. Pitoňák, P. Neogrády, J. Černý, S. Grimme, and P. Hobza, "Scaled MP3 non-covalent interaction energies agree closely with accurate CCSD(T) benchmark data," ChemPhysChem **10**, 282–289 (2009).

[39]Y. Huang and G. J. O. Beran, "Reliable prediction of three-body intermolecular interactions using dispersion-corrected second-order møller-plesset perturbation theory," J. Chem. Phys. **143**, 044113 (2015).

[40]C. H. Borca, B. W. Bakr, L. A. Burns, and C. D. Sherrill, "CrystaLattE: Automated computation of lattice energies of organic crystals exploiting the many-body expansion to achieve dual-level parallelism," J. Chem. Phys. **151**, 144103 (2019).

[41]I. Nahringbauer, "Reinvestigation of the crystal structure of acetic acid (at +5°C and –190°C)," Acta Chem. Scand. **24**, 453–462 (1970).

[42]E. D. Stevens, "Low-temperature experimental electron density distribution of formamide," Acta Crystallogr. B **34**, 544–551 (1978).

[43]B. M. Craven, R. K. McMullan, J. D. Bell, and H. C. Freeman, "The crystal structure of imidazole by neutron diffraction at 20°C and –150°C," Acta Crystallogr. B **33**, 2585–2589 (1977).

[44]F. H. Allen, "The Cambridge Structural Database: a quarter of a million crystal structures and rising," Acta Crystallogr. B **58**, 380–388 (2002).

[45]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, "Psi4 1.4: Open-source software for high-throughput quantum chemistry," J. Chem. Phys. **152**, 184108 (2020).

[46]T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007–1023 (1989).

[47]R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," J. Chem. Phys. **96**, 6796–6806 (1992).

[48]F. Weigend, "A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency," Phys. Chem. Chem. Phys. **4**, 4285–4291

(2002).

[49]F. Weigend, A. Köhn, and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," J. Chem. Phys. **116**, 3175–3183 (2002).

[50]A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, "Basis-set convergence in correlated calculations on Ne, N_2, and H_2O," Chem. Phys. Lett. **286**, 243–252 (1998).

[51]A. L. L. East and W. D. Allen, "The heat of formation of NCO," J. Chem. Phys. **99**, 4638–4650 (1993).

[52]L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches," J. Chem. Phys. **141**, 234111 (2014).

[53]S. F. Boys and F. Bernardi, "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," Mol. Phys. **19**, 553–566 (1970).

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

[55]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," J. Chem. Phys. **132**, 154104 (2010).

[56]E. Caldeweyher, C. Bannwarth, and S. Grimme, "Extension of the D3 dispersion coefficient model," J. Chem. Phys. **147**, 034112 (2017).

[57]E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, "A generally applicable atomic-charge dependent london dispersion correction," J. Chem. Phys. **150**, 154122 (2019).

[58]O. A. von Lilienfeld and A. Tkatchenko, "Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids," J. Chem. Phys. **132**, 234109 (2010).