Coupled Cluster and Møller–Plesset Perturbation Theory Calculations of Noncovalent Intermolecular Interactions using Density Fitting with Auxiliary Basis Sets from Cholesky Decompositions

Jonas Boström,[†] Michal Pitoňák,[‡] Francesco Aquilante,[§] Pavel Neogrády,[∥] Thomas Bondo Pedersen,*,[⊥] and Roland Lindh*,[§]

[†]Department of Theoretical Chemistry, Chemical Center, Lund University, P.O. Box 124 S-221 00 Lund, Sweden

[‡]Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, Mlynská Dolina, SK-84215 Bratislava, Slovak Republic, and Computing Center of the Slovak Academy of Sciences, Dúbravská cesta č. 9, SK-84535 Bratislava, Slovak Republic

[§]Department of Chemistry—Ångström, the Theoretical Chemistry Programme, Uppsala University, P.O. Box 518, SE-751 20 Uppsala, Sweden

[∥]Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, Mlynská Dolina, SK-84215 Bratislava, Slovak Republic

[⊥]Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway

ABSTRACT: We compute noncovalent intermolecular interaction energies for the S22 test set [*Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993] of molecules at the Møller−Plesset and coupled cluster levels of supermolecular theory using density fitting (DF) to approximate all two-electron integrals. The error due to the DF approximation is analyzed for a range of auxiliary basis sets derived from Cholesky decomposition (CD) in conjunction with correlation consistent and atomic natural orbital valence basis sets. A Cholesky decomposition threshold of $10^{-4}E_h$ for full molecular CD and its one-center approximation (1C-CD) generally yields errors below 0.03 kcal/mol, whereas $10^{-3}E_h$ is sufficient to obtain the same level of accuracy or better with the atomic CD (aCD) and atomic compact CD (acCD) auxiliary basis sets. Comparing to commonly used predefined auxiliary basis sets, we find that while the aCD and acCD sets are larger by a factor of 2−4 with triple-ζ AO basis sets, they provide results 1−2 orders of magnitude more accurate.

1. INTRODUCTION

Density fitting (DF; or resolution of the identity, RI)[1−8] has become a standard technique for speeding up molecular orbital (MO)-based electronic structure calculations. The speed-up is achieved by expanding products of atomic orbitals (AOs) in an atom-centered auxiliary basis set, thus replacing four-center two-electron integrals with two- and three-center integrals involving AOs and auxiliary basis functions. Auxiliary basis sets are normally preoptimized and tailored to reproduce specific energy contributions with sufficient accuracy. A commonly accepted definition of the term *sufficient accuracy* is that the error due to the DF approximation is at least 1 order of magnitude smaller than the error arising from the incompleteness of the AO basis set, which means that an auxiliary basis set typically has been constructed for each AO basis set and quantum chemical method in order to minimize the number of auxiliary basis functions (normally 2−4 times the number of AO basis functions). Recently, auxiliary basis sets have been developed for Coulomb (RI-J2)[9] and Coulomb and Hartree−Fock (HF) exchange (RI-JK)[10,11] intended for use with any AO basis set up to a certain cardinal number. The RI-C[12,13] auxiliary basis sets were individually optimized for MP2 calculations with a specific AO basis set but have been applied

to coupled cluster methods as well. These auxiliary basis sets thus are designed for applications of nonhybrid density functional theory (DFT), hybrid DFT and HF, and second-order correlated theory of the electronic ground state. Cholesky decomposition (CD) of two-electron integrals[14,15] offers an alternative approach to the construction of auxiliary basis sets, as described in more detail in the reviews by Pedersen et al.[16] and Aquilante et al.[17] In the CD-based approach, auxiliary basis sets are constructed on-the-fly without using a definition of sufficient accuracy in energy terms. Instead, the accuracy of individual two-electron integrals, and hence the accuracy of the underlying finite basis Hamiltonian, is controlled by the decomposition threshold. The decomposition threshold thus provides indirect control of the accuracy of total ground and excited state energies as well as other molecular properties through a single parameter. Focusing on individual integrals rather than on ground state energy contributions, the CD-based auxiliary basis sets are unbiased in the sense that they are not designed with a specific quantum chemical method or property

in mind. For this reason, we sometimes use the term *ab initio* to characterize this family of auxiliary basis sets.

The CD-based auxiliary basis sets have been extensively benchmarked in two recent studies[18,19] that emphasize the unbiased nature of these sets and document the correlation between decomposition threshold and the accuracy of total and relative energies. The accuracy of total energies and dipole moments of small molecules as a function of the decomposition threshold was documented by Boström et al.[18] using HF theory, hybrid and nonhybrid DFT, and second-order Møller–Plesset perturbation theory (MP2), thus covering Coulomb, exchange, and dynamical correlation contributions to the total energy. Vertical electronic excitation energies computed at the complete active space self-consistent field (CASSCF) and multiconfigurational second-order perturbation (CASPT2) levels of theory were the subject of a subsequent study by Boström et al.[19] These studies cover total and relative energies in individual molecules but do not cover interactions between molecules, which is the subject of the present work.

The HF and DFT methods are often insufficient for describing intermolecular interactions, especially when it comes to noncovalent complexes where dispersion becomes increasingly important. It is therefore common to use dynamical correlation methods like MP2[20] or coupled cluster theory[21] for an accurate description. It is reasonable to expect that highly correlated theories require higher accuracy in the individual two-electron integrals than HF and DFT, since the occupied and virtual spaces couple and cross-terms need to be described in a proper way. It thus is fairly obvious that an accurate HF reference function is a prerequisite for a correlated calculation. While a few computational studies of weak intermolecular interactions using a CD-based implementation have appeared,[22−24] no systematic study of the accuracy of the full range of CD-based auxiliary basis sets is recorded in the literature.

It is the purpose of the present work to study the dependence on the decomposition threshold of the accuracy of intermolecular interaction energies computed at different levels of correlated theory such that recommended values of the decomposition threshold to achieve sufficient accuracy can be established. *Sufficient accuracy* needs to be defined, of course. Jurečka et al.[25] compared DF-MP2 with preoptimized auxiliary basis sets to conventional MP2 results for the interaction energies in hydrogen-bonded and stacked DNA base pairs and found that the error due to the DF approximation was below 0.03 kcal/mol, which we will adopt as our primary target accuracy. Since coupled cluster theory is a high-precision method, however, we will also investigate the auxiliary basis set requirements for obtaining results an order of magnitude closer to the conventional results; i.e., we use 0.003 kcal/mol as a secondary target accuracy.

We start in section 2 by a brief outline of the underlying theory of CD-based auxiliary basis sets. Section 3 contains computational details. Results are discussed in section 4, and our conclusions are given in section 5.

2. THEORETICAL BACKGROUND

The central feature of DF is the expansion of products of AOs in a set of auxiliary functions.[1−8] Given an AO basis set $\{\mu(\mathbf{r})\}$ and an auxiliary basis set $\{J(\mathbf{r})\}$, the DF approximation consists of the expansion

$$\mu(\mathbf{r})\,\nu(\mathbf{r}) \approx \widetilde{\mu(\mathbf{r})\,\nu(\mathbf{r})} = \sum_J C_{\mu\nu}^J J(\mathbf{r}) \tag{1}$$

The expansion coefficients in this work are determined by least-squares fitting in the Coulomb metric. That is, we minimize the norm in the Coulomb metric, $(\Delta_{\mu\nu}|\Delta_{\mu\nu})$ in Mulliken notation, of the fitting error $\Delta_{\mu\nu}(\mathbf{r}) = \mu(\mathbf{r})\,\nu(\mathbf{r}) - \widetilde{\mu(\mathbf{r})\nu}(\mathbf{r})$. This leads to the fitting equations $(\Delta_{\mu\nu}|J) = 0$ or, equivalently,

$$\sum_K C_{\mu\nu}^K (K|J) = (\mu\nu|J) \tag{2}$$

The two-electron integrals may then be represented as

$$(\mu\nu|\kappa\lambda) \approx (\widetilde{\mu\nu}|\widetilde{\kappa\lambda}) = \sum_{JK} C_{\mu\nu}^J (J|K) C_{\kappa\lambda}^K = \sum_J L_{\mu\nu}^J L_{\kappa\lambda}^J \tag{3}$$

The vectors appearing in the last expression, which is used in all parts of our implementation (i.e., we never use and never explicitly compute the fitting coefficients),[26] are defined by

$$L_{\mu\nu}^J = \sum_K (\mu\nu|K) R_K^J \tag{4}$$

where $\sum_I R_J^I R_K^I = (G^{-1})_{JK}$ with $G_{JK} = (J|K)$. The absolute value of the error of the DF integral representation

$$(\mu\nu|\kappa\lambda) - (\widetilde{\mu\nu}|\widetilde{\kappa\lambda}) = (\Delta_{\mu\nu}\Delta_{\kappa\lambda}) \tag{5}$$

is bounded from above by the fitting error norms minimized in the fitting procedure:

$$|(\Delta_{\mu\nu}|\Delta_{\kappa\lambda})| \leq \sqrt{(\Delta_{\mu\nu}|\Delta_{\mu\nu})(\Delta_{\kappa\lambda}|\Delta_{\kappa\lambda})} \tag{6}$$

While $(\Delta_{\mu\nu}|\Delta_{\mu\nu})$ is minimized by solving eq 2, its value is determined solely by the quality of the auxiliary basis set, and errors due to the DF approximation thus can be traced back to the choice of auxiliary basis set. The upper bound can be made arbitrarily small by selecting an auxiliary basis set that nearly spans the same space as the products of AOs. The usefulness of the DF approximation stems from the fact that the number of auxiliary basis functions needed to produce a fit with reasonable accuracy is far less than the total number of products due to linear or near-linear dependence. Moreover, choosing atom-centered auxiliary functions, we only need to compute two- and three-center two-electron integrals. The auxiliary basis sets benchmarked in this work are constructed on-the-fly using different CD schemes.

With a given decomposition threshold δ, the incomplete CD procedure yields a set of M Cholesky vectors $\{\mathbf{L}^J, J = 1, 2, ..., M\}$ such that all elements of the positive semidefinite residual matrix[14]

$$D_{\mu\nu,\kappa\lambda}^{(M)} = (\mu\nu|\kappa\lambda) - \sum_{J=1}^M L_{\mu\nu}^J L_{\kappa\lambda}^J \tag{7}$$

lie in the interval $[-\delta,\delta]$. That is, every integral included in the original integral matrix has an accuracy of at least δ. The Cholesky vectors are computed in a recursive scheme according to[14]

$$L_{\mu\nu}^J = [D_{[\kappa\lambda]_J,[\kappa\lambda]_J}^{(J-1)}]^{-1/2} D_{\mu\nu,[\kappa\lambda]_J}^{(J-1)} \tag{8}$$

where the recursion stops when the largest diagonal element of the residual matrix is below (or equal to) δ. Here, $D_{\mu\nu,\kappa\lambda}^{(0)} = (\mu\nu|\kappa\lambda)$ and $[\kappa\lambda]_J$ is the index of the largest residual diagonal element at the $(J - 1)$th recursion; i.e., the AO product that gives rise to the Jth Cholesky vector. We refer to these parent products as the Cholesky basis. The DF and CD approaches are completely equivalent when the Cholesky basis is chosen as the

auxiliary basis and the fitting error norms become identical to the diagonal elements of the residual matrix.[16,17] For more details on the implementation of two-electron integral CD, we refer to Aquilante et al.[17,26]

In full CD (FCD),[14,15] the auxiliary basis set is constructed by performing an incomplete CD with a given decomposition threshold δ of the full molecular two-electron integral matrix such that $(\Delta_{\mu\nu}|\Delta_{\mu\nu}) \leq \delta$, $\forall$ $\mu\nu$. While FCD thus provides complete error control through δ, the FCD auxiliary basis set is composed of both atom-centered functions and two-center (product) functions. The one-center CD (1C-CD)[27] scheme is identical to FCD except that two-center functions are not allowed to enter the auxiliary basis set. The strict error control of FCD is, however, lost as the inequality $(\Delta_{\mu\nu}|\Delta_{\mu\nu}) \leq \delta$ only holds if μ and ν are centered on the same atom.

In atomic CD (aCD),[27] the auxiliary basis set is constructed by CD of the atomic two-electron integral matrix for each unique atom in the molecule. Like 1C-CD, the aCD auxiliary basis sets contain atom-centered functions only, and $(\Delta_{\mu\nu}|\Delta_{\mu\nu}) \leq \delta$ only holds if μ and ν are centered on the same atom. Unlike FCD and 1C-CD, the aCD auxiliary basis set contains more functions than the products entering the Cholesky basis; for computational reasons, functions are added to complete the shells of the auxiliary basis.[27] Finally, the atomic compact CD (acCD)[28] scheme is a modification of aCD where linear dependence among the primitive Gaussians is removed by means of a second CD.

The benefits of the CD-based schemes can be summarized as follows:

- constructed on-the-fly
- defined for any AO basis set
- applicable in conjunction with any quantum chemical method for the calculation of general molecular properties, including ground and excited state energies (unbiased)
- accuracy controlled by a single parameter, the decomposition threshold δ

There are also some drawbacks compared to preoptimized auxiliary basis sets. First, there is an overhead involved in constructing the auxiliary basis set on-the-fly. While the overhead is negligible for aCD and acCD, it can be significant for FCD and 1C-CD where the entire molecular integral matrix is Cholesky decomposed. Second, exponents and contractions are inherited from the AO products singled out by the CD procedure and therefore not optimized for maximum computational efficiency. Third, since the CD-based auxiliary basis sets are constructed to preserve the properties of the Hamiltonian and thus not specifically tailored to a quantum chemical method, they tend to be larger and therefore more computationally demanding than the preoptimized sets. Note, however, that this is not always the case. In a recent study by Hohenstein and Sherill[24] on wave-function-based symmetry-adapted perturbation theory (SAPT), for example, the 1C-CD scheme leads to smaller auxiliary basis sets than the corresponding preoptimized ones for comparable levels of accuracy. It should be mentioned that the CD philosophy can be adapted to construct extremely small auxiliary basis sets designed for specific energy contributions, so-called method-specific CD (MSCD),[17,29] which we will not investigate in this work.

3. COMPUTATIONAL DETAILS

All calculations are performed using development version 7.7 of the quantum chemical program package MOLCAS[26] at the

MP2,[20] coupled cluster singles and doubles (CCSD),[30] and CCSD with perturbative triples corrections [CCSD(T)][31] levels of theory. Implementation details can be found in Aquilante et al.,[17,26] Pitoňák et al.,[32] and references therein. We only note here that besides an implementation of second-order coupled cluster linear response theory by Pedersen et al.,[33] the coupled cluster implementation in MOLCAS[26,32] is the first one based entirely on the DF integral representation. As discussed in the implementation papers,[26,32] the main advantages of the DF approximation in canonical CCSD and CCSD(T) are reduction of input/output (I/O) and simplification of the parallel algorithm. In practice, therefore, the operation count (and hence CPU time requirement) is approximately the same for conventional and DF-based CCSD and CCSD(T), whereas the I/O bottleneck (and hence wall time) is significantly reduced.

The CCSD(T) model is generally accepted as the "gold standard" for calculating noncovalent intermolecular interactions, whereas the MP2 method is known to deliver quite accurate results in most cases (hydrogen bonded or saturated dispersion dominated complexes) but to fail for π–π stacking.[34] The performance of third-order correlation methods, like CCSD, for noncovalent interactions is rather unsatisfactory, failing mostly for the same types of complexes as MP2. Nevertheless, the third-order correlation correction was recently found to be very useful in empirical schemes, like the MP2.5 model of Pitoňák et al.,[35] which turned out to be capable of dealing with the problematic cases.

A suitable set of molecules for assessing the accuracy of the DF-based correlation methods is the S22 test set of Jurečka et al.,[36] which consists of 22 model complexes. We have performed calculations on the full S22 test suite with the HF and MP2 methods. Since the DF-based CCSD and CCSD(T) implementation is as demanding as the conventional one in terms of CPU time (see above), we have excluded half of the molecules[37] in the S22 set for these methods. Using a smaller subset allows us to be consistent in the choice of basis sets and to afford to study the impact of several parameters. The S22 set is divided into hydrogen bonded complexes, complexes with predominant dispersion contribution, and mixed complexes, and we have chosen molecules from all three categories for our limited test suite.[37]

Even if we only consider rather small molecules in this work, we substitute the reference (conventional) results with the ones obtained by the FCD method with a restrictive ($10^{-10}E_h$) decomposition threshold. By doing so, we are able to take advantage of the efficient parallel implementation of our new CCSD(T) module in MOLCAS.[26,32] Calculations have been performed with the four types of auxiliary basis sets (FCD, 1C-CD, aCD, and acCD) with decomposition thresholds of 10^{-3}, 10^{-4}, 10^{-5}, and $10^{-6}E_h$. Three types of valence basis sets at the double- and triple-ζ levels are used, namely, cc-pVXZ,[38] ANO-L-VXZP,[39] and aug-cc-pVXZ[40] (with X = D, T). The frozen core approximation is used in all correlated calculations, and the interaction energies are corrected for basis set superposition errors using the counterpoise method.[41] For comparison, we also perform HF and MP2 calculations of the S22 test suite using preoptimized auxiliary basis sets. Weigend's[11] RI-JK set (standard contracted) is used for HF, and the RI-C set of Weigend et al.[13] is used for MP2 (based on the RI-JK-HF reference). Weigend[11] constructed the RI-JK set for use with triple- as well as quadruple-ζ AO basis sets. We here use the RI-JK set with the double- and triple-ζ AO basis sets mentioned
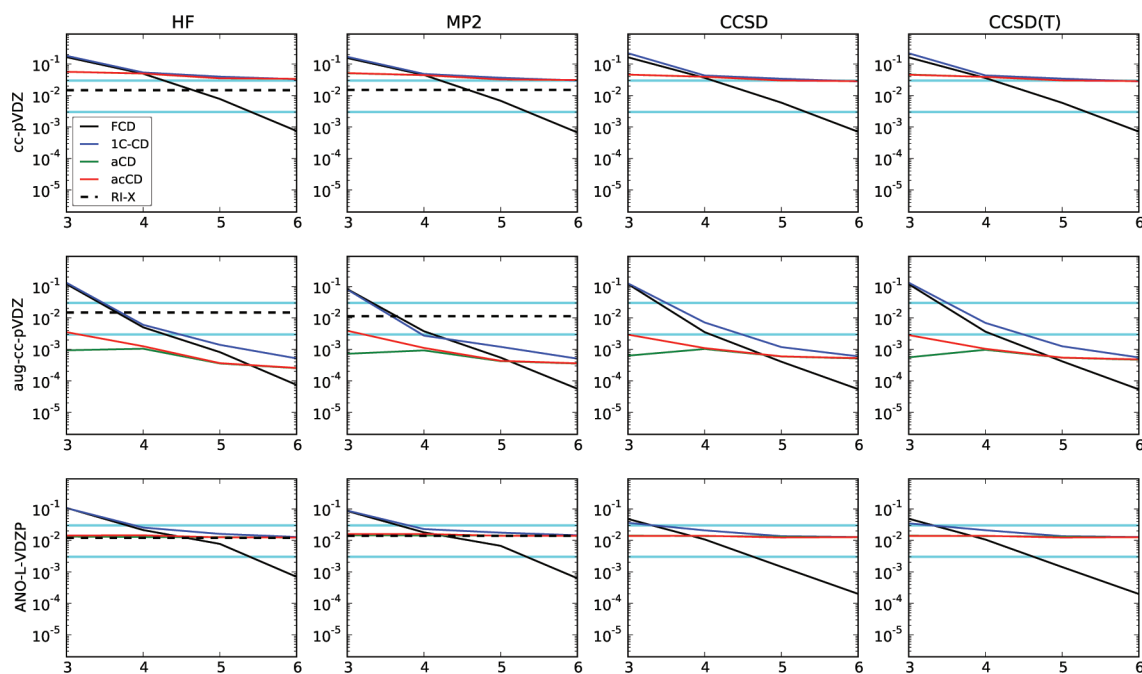
Figure 1. The 90th percentile (in kcal/mol) of the distribution of absolute errors in interaction energies due to the DF approximation as a function of decomposition threshold. Each row of panels corresponds to a specific double-ζ basis set and each column to a specific quantum chemical method, as indicated. In each panel, the negative logarithm of the decomposition threshold is the abscissa, and the absolute error is the ordinate (logarithmic scale). Horizontal lines mark the target accuracies 0.003 and 0.03 kcal/mol. The label "RI-X" is used to denote RI-JK for HF and RI-C for MP2.

above without modifications: no additional diffuse functions are added to the RI-JK set when used in conjunction with the aug-cc-pVXZ AO basis sets, and no functions are removed when used in conjunction with double-ζ AO basis sets. With the ANO-L-VXZP sets, we use the same RI-JK and RI-C sets as with the analogous cc-pVXZ sets.

4. RESULTS AND DISCUSSION

We will use the short-hand notation FCD-n, 1C-CD-n, aCD-n, and acCD-n to denote an auxiliary basis set obtained with a decomposition threshold of $10^{-n}E_h$. The term *highly accurate* means that the absolute interaction energy error due to the DF approximation is below the secondary target accuracy 0.003 kcal/mol (approximately 1.0 cm^{-1} or 4.8 μE_h), whereas *accurate* is used for absolute errors below 0.03 kcal/mol (approximately 10 cm^{-1} or 48 μE_h).

4.1. Overall Accuracy. Rather than average value and standard deviation, we use the less optimistic 90th percentile to characterize the computed absolute interaction energy error distributions. Among the commonly used definitions of percentiles, we employ a linear interpolation formula as follows. Given a set of N absolute interaction energy errors $\{\Delta E_j, j = 1, 2, 3, ..., N\}$ sorted in ascending order, the Pth percentile, $0 \leq P \leq 100$, is given by

$$\rho_P = \begin{cases} \Delta E_1 & r = 1 \\ \Delta E_i + d(\Delta E_{i+1} - \Delta E_i) & 1 < r < N \\ \Delta E_N & r = N \end{cases} \quad (9)$$

where i and d are the integer and decimal parts, respectively, of the rank

$$r = i + d = 1 + \frac{P(N-1)}{100} \quad (10)$$

The 90th percentiles obtained with double-ζ basis sets are plotted in Figure 1 as a function of the decomposition threshold. In accordance with the theoretical discussion above, only FCD provides complete error control through the decomposition threshold. A saturation effect due to the lack of two-center functions is observed for 1C-CD when the threshold is reduced below a certain level depending on the choice of basis set. The aCD and acCD auxiliary basis sets provide nearly constant accuracies at any decomposition threshold from $10^{-3}E_h$ and below. The same level of accuracy is observed regardless of quantum chemical method with a given basis set and auxiliary basis set. While the quality of an auxiliary basis set thus is unbiased with respect to the quantum chemical method, it does depend quite strongly on the quality of the valence basis set. The aCD and acCD auxiliary sets produce results that are barely accurate with cc-pVDZ, accurate results with ANO-L-VDZP, and highly accurate results with aug-cc-pVDZ. The coupling between basis set quality and auxiliary basis set quality can also be seen for FCD. For example, FCD-4 gives nearly accurate results with cc-pVDZ and accurate results with ANO-L-VDZP, and with aug-cc-pVDZ we obtain results close to the 0.003 kcal/mol target. The 1C-CD sets yield results with virtually the same accuracy as FCD with higher thresholds, whereas lower thresholds lead to results comparable to the aCD and acCD sets.

Several of these trends are also observed for the absolute interaction energy errors computed with triple-ζ basis sets, see Figure 2. Only FCD provides complete error control. Although 1C-CD appears to provide complete error control to the same extent as FCD, this is solely due to the saturation effect being postponed to lower decomposition thresholds (i.e., two-center functions are not needed in FCD with higher decomposition thresholds). Note that with cc-pVTZ we do observe the saturation effect going from $\delta = 10^{-5}E_h$ to $10^{-6}E_h$. The higher
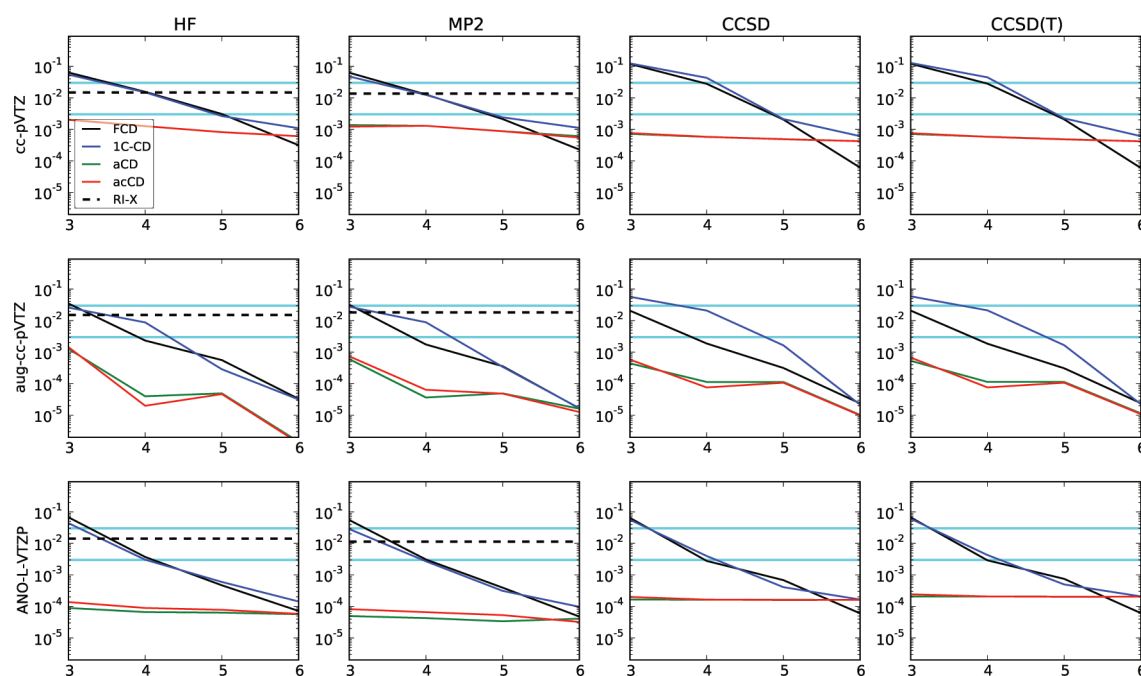
Figure 2. The 90th percentile (in kcal/mol) of the distribution of absolute errors in interaction energies due to the DF approximation as a function of decomposition threshold. Each row of panels corresponds to a specific triple-ζ basis set and each column to a specific quantum chemical method as indicated. In each panel, the negative logarithm of the decomposition threshold is the abscissa, and the absolute error is the ordinate (logarithmic scale). Horizontal lines mark the target accuracies 0.003 and 0.03 kcal/mol. The label "RI-X" is used to denote RI-JK for HF and RI-C for MP2.

quality of triple-ζ basis sets compared to double-ζ ones is reflected in the quality of the auxiliary basis sets. In particular, the aCD and acCD sets provide highly accurate results with a decomposition threshold of $10^{-3}E_h$. We also observe an improvement for FCD and 1C-CD compared to the analogous double-ζ results: FCD-4 and 1C-CD-4 provide accurate or highly accurate results with any of the triple-ζ basis sets.

In our earlier study on total energy accuracy,[18] we found that a decomposition threshold of $10^{-4}E_h$ provides an average total energy error of at most 0.01 kcal/(mol·electron). The complexes of the S22 set contain between 20 and 136 electrons, corresponding to average errors in total ground state energies between 0.2 and 1.36 kcal/mol. The significantly smaller interaction energy errors thus are ascribable to error cancellation when subtracting monomer energies from the dimer energy in the counterpoise correction scheme. Similar error cancellations were observed for excitation energies.[19]

As expected from the results reported by Jurečka et al.,[25] the RI-JK and RI-C absolute errors are below the 0.03 kcal/mol target and at roughly the same level regardless of AO basis set, see Figures 1 and 2. With the cc-pVDZ basis set, the RI-X (generic notation for RI-JK or RI-C as appropriate) results are even somewhat more accurate than the CD-based ones, except for FCD with threshold $10^{-5}E_h$ and below. This result is perhaps somewhat surprising at first sight, but one has to recall that the RI-JK set is designed for triple- and quadruple-ζ AO basis sets and thus more accurate than one would normally expect from a preoptimized auxiliary set at the double-ζ level. This is confirmed by running RI-C-MP2 and acCD-3-MP2 with a conventional HF reference function: the 90th percentiles then become 0.01255 and 0.00364 kcal/mol, respectively. Note that these calculations also indicate that the accuracy of correlated acCD interaction energies is limited by the accuracy of the reference HF function, suggesting that one might benefit from a hybrid approach where the HF reference is determined with

high accuracy (conventional or FCD with threshold below $10^{-5}E_h$), while the correlation correction is computed using acCD-3. With the ANO-L-VDZP basis set, RI-X and aCD/acCD yield nearly identical errors, whereas aCD and acCD are about an order of magnitude more accurate than RI-X with the aug-cc-pVDZ basis set. With triple-ζ basis sets the aCD and acCD sets are at least 1−2 orders of magnitude more accurate than RI-X. The greater accuracy of aCD/acCD compared to RI-X comes at a computational cost, as we shall discuss in more detail below.

4.2. Auxiliary Basis Set Pruning. Deriving auxiliary basis sets by means of CD implies that the number and nature of auxiliary functions be determined indirectly through the decomposition threshold. In our implementation of HF and MP2, the dimension of the auxiliary basis set determines the computational cost. The ratio of the number of auxiliary functions to the number of AO basis functions is plotted as a function of decomposition threshold in Figure 3. The ratios are averaged over the molecules of the S22 test set. For the molecular auxiliary sets (FCD and 1C-CD), we observe a nearly linear increase with the (negative logarithm of the) decomposition threshold with the slope of the FCD lines being larger than the slope of the corresponding 1C-CD lines. The average ratios range from about 3 to approximately 8 for FCD and from 3 to 6 for 1C-CD. The aCD and acCD auxiliary sets contain the same number of functions by construction, and hence, only aCD is plotted in Figure 3. The aCD ratios increase with a much smaller slope than FCD and 1C-CD as the decomposition threshold is decreased. For double-ζ basis sets, the aCD ratio ranges from 4 to 6, whereas for triple-ζ basis sets the range is from 6 to 8. For aug-cc-pVTZ and ANO-L-VTZP, the aCD ratio is larger than the FCD and 1C-CD ratios with any of the decomposition thresholds used in this work.

We saw above that a decomposition threshold of $10^{-3}E_h$ provides highly accurate results for aCD as well as acCD with
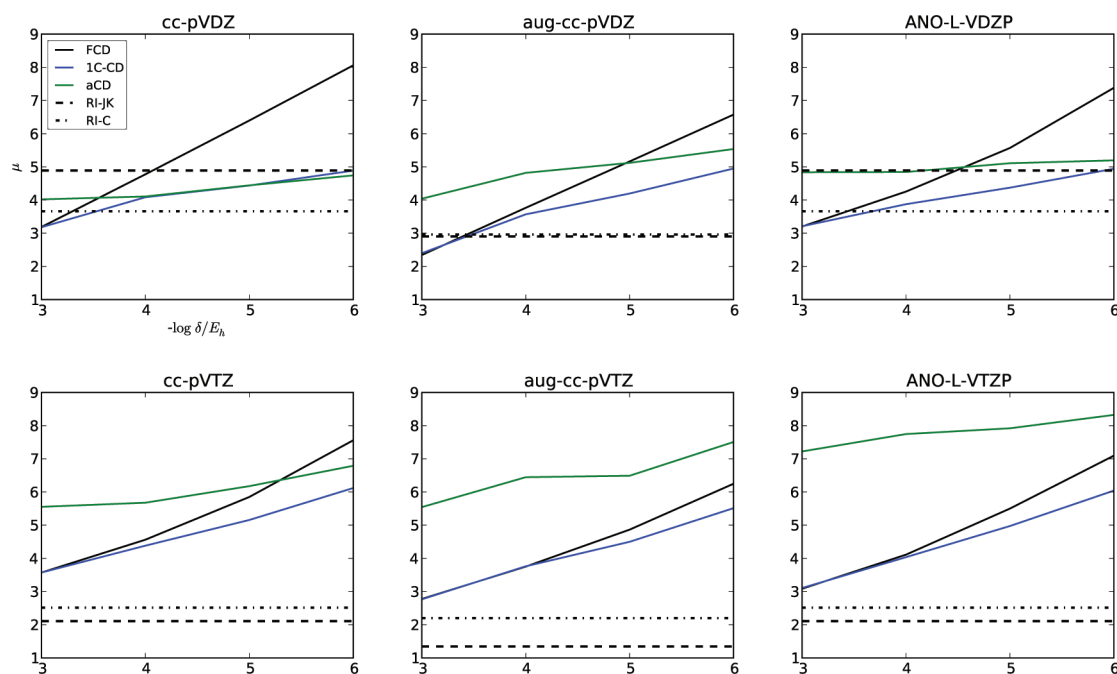
Figure 3. Average ratio (μ) of the number of auxiliary basis functions to the number of AO basis functions as a function of decomposition threshold for each basis set employed in this work. The acCD ratios are identical to the aCD ones and therefore not plotted.
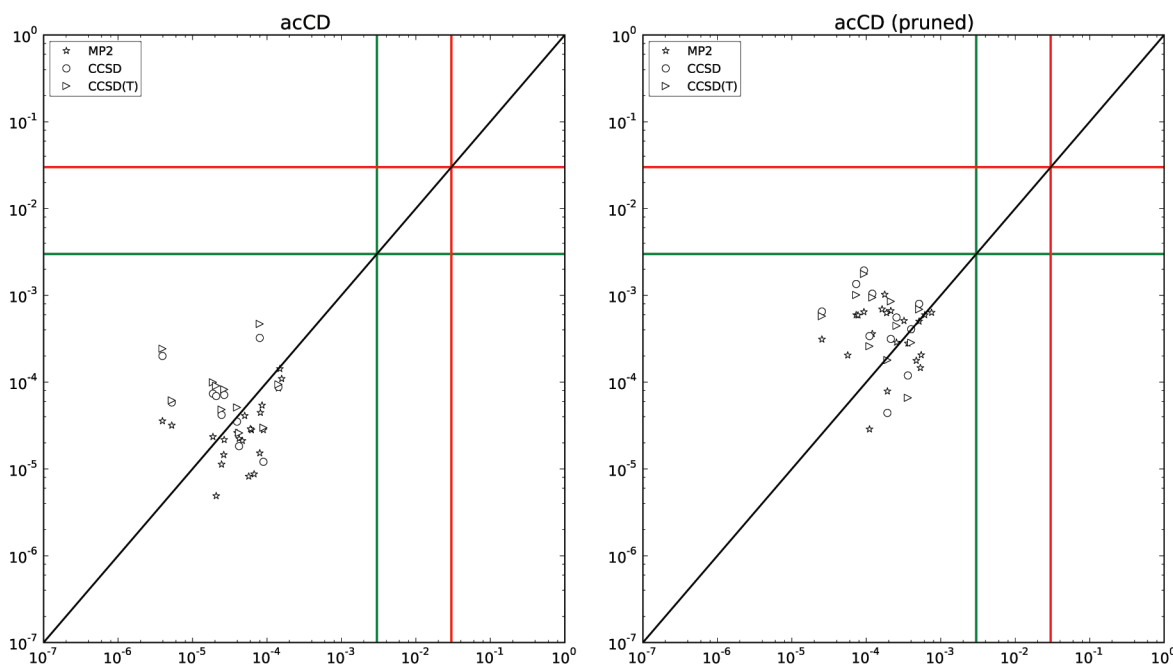

Figure 4. Absolute errors in interaction energies (in kcal/mol) due to the unpruned and pruned acCD-3 auxiliary basis sets with the ANO-L-VTZP AO basis set. Each point represents a pair of HF and correlated absolute errors, with HF absolute error as the abscissa and correlated absolute error as the ordinate. The green and red vertical and horizontal lines mark the target accuracies 0.003 and 0.03 kcal/mol, respectively. The diagonal line is a visual aid: points above this line indicate that the correlated interaction energy is less accurate than the HF interaction energy and *vice versa* for points below the line.

triple-ζ basis sets, but the dimension of aCD-3 and acCD-3 is comparable to FCD-6. This is mainly due to many higher angular momentum auxiliary basis functions. We have previously shown that the highest angular momentum components can be neglected with a minor accuracy loss in total and excitation energies.[18,19,27]

We here investigate this auxiliary basis set pruning for acCD-3 with the ANO-L-VTZP basis set for which the auxiliary-to-AO

ratio is largest. Employing pruning, the average auxiliary-to-AO ratio is reduced by 31% from 7.22 to 4.95, which is comparable to the average ratio obtained with double-ζ basis sets. While the computational cost is reduced by essentially the same amount (34%) for HF and MP2, Figure 4 shows that the interaction energies remain highly accurate even if the errors are about 1 order of magnitude larger than those of the (unpruned) acCD-3 set.

Figure 3 reveals that in general the unpruned CD-based auxiliary basis sets are larger than the RI-X ones by up to a factor of about 4. In our implementation, this implies that the RI-X calculations are up to a factor of 4 faster than the CD-based ones. The much improved accuracy thus carries a computational penalty. It is of interest to compare the auxiliary-to-AO ratios at a decomposition threshold where the RI-X and acCD sets yield comparable accuracy. The 90th percentiles obtained with RI-X and the ANO-L-VTZP basis set are 0.014 and 0.011 kcal/mol for HF and MP2, respectively, while those obtained with pruned acCD-2 are 0.015 and 0.014 kcal/mol. The average auxiliary-to-AO ratios are 2.11 for RI-JK, 2.51 for RI-C, and 3.48 for pruned acCD-2. Hence, the computational penalty for the unbiased nature of pruned acCD-2 is less than a factor of 2 for comparable accuracy. Decreasing the decomposition threshold a bit to obtain pruned acCD-2.5, the computational penalty remains at a factor of about 2 (average auxiliary-to-AO ratio: 4.58), while the accuracy comes close to the 0.003 kcal/mol target (0.004 kcal/mol for both HF and MP2).

5. CONCLUSIONS

On the basis of the results presented above, we recommend that a decomposition threshold of $10^{-4}E_h$ be used for FCD and 1C-CD. This threshold generally results in interaction energy errors below 0.03 kcal/mol and improving with basis set quality—with aug-cc-pVTZ, for instance, the error is found to be below 0.003 kcal/mol with this threshold. It is advisable to check the results obtained with this threshold in conjunction with double-ζ basis sets (without diffuse functions added), e.g., by running a few calculations with FCD-5 or FCD-6.

A decomposition threshold of $10^{-3}E_h$ is found to be sufficient for aCD and acCD auxiliary basis sets to produce errors below 0.003 kcal/mol, except with double-ζ basis sets (without diffuse functions added) where the error may be close to 0.03 kcal/mol. Test calculations using the ANO-L-VTZP basis set indicate that further increasing the decomposition threshold as well as pruning may be employed to speed up calculations while maintaining an accuracy close to or below 0.003 kcal/mol.

The RI-JK and RI-C auxiliary basis sets yield HF and MP2 errors below 0.03 kcal/mol but well above 0.003 kcal/mol. For triple-ζ AO basis sets, the RI-C auxiliary basis set is about 4 times smaller than the (unpruned) aCD and acCD ones, but the errors are at least 1–2 orders of magnitude larger.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: thomas.b.pedersen@gmail.com; roland.lindh@kvac.uu.se.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Boys, S. F.; Shavitt, I. *A Fundamental Calculation of the Energy Surface for the System of Three Hydrogen Atoms*; University of Wisconsin Technical Report WIS-AF-13, 1959.

(2) Billingsley, F. P.; Bloor, J. E. *J. Chem. Phys.* **1971**, *55*, 5178–5190.

(3) Whitten, J. L. *J. Chem. Phys.* **1973**, *58*, 4496–4501.

(4) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41–51.

(5) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 3396–3402.

(6) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 4993–4999.

(7) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

(8) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

(9) Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

(10) Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(11) Weigend, F. *J. Comput. Chem.* **2008**, *29*, 167–175.

(12) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(13) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

(14) Beebe, N. H. F.; Linderberg, J. *Int. J. Quantum Chem.* **1977**, *12*, 683–705.

(15) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. *J. Chem. Phys.* **2003**, *118*, 9481–9484.

(16) Pedersen, T. B.; Aquilante, F.; Lindh, R. *Theor. Chem. Acc.* **2009**, *124*, 1–10.

(17) Aquilante, F.; Boman, L.; Boström, J.; Koch, H.; Lindh, R.; Sánchez de Merás, A.; Pedersen, T. B. Cholesky Decomposition Techniques in Electronic Structure Theory. In *Linear-Scaling Techniques in Computational Chemistry and Physics. Methods and Applications*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: New York, 2011; Challenges and Advances in Computational Chemistry and Physics, Vol. *13*, Chapter 13, pp 301–343.

(18) Boström, J.; Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Theory Comput.* **2009**, *5*, 1545–1553.

(19) Boström, J.; Delcey, M. G.; Aquilante, F.; Serrano-Andrés, L.; Pedersen, T. B.; Lindh, R. *J. Chem. Theory Comput.* **2010**, *6*, 747–754.

(20) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

(21) Bartlett, R.; Musial, M. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

(22) García Cuesta, I.; Pedersen, T. B.; Koch, H.; Sánchez de Merás, A. M. J. *ChemPhysChem* **2006**, *7*, 2503–2507.

(23) Fernández, B.; Pedersen, T. B.; Sánchez de Merás, A.; Koch, H. *Chem. Phys. Lett.* **2007**, *441*, 332–335.

(24) Hohenstein, E. G.; Sherill, C. D. *J. Chem. Phys.* **2010**, *132*, 184111.

(25) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578–4582.

(26) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224–247.

(27) Aquilante, F.; Lindh, R.; Pedersen, T. B. *J. Chem. Phys.* **2007**, *127*, 114107.

(28) Aquilante, F.; Gagliardi, L.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2009**, *130*, 154107.

(29) Boman, L.; Koch, H.; Sánchez de Merás, A. *J. Chem. Phys.* **2008**, *129*, 134107.

(30) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910–1918.

(31) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(32) Pitoňák, M.; Aquilante, F.; Hobza, P.; Neogrády, P.; Noga, J.; Urban, M. *Collect. Czech. Chem. Commun.* **2011**, *76*, 713–742.

(33) Pedersen, T. B.; Sánchez de Merás, A. M. J.; Koch, H. *J. Chem. Phys.* **2004**, *120*, 8887–8897.

(34) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023–5063.

(35) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10*, 282−289.

(36) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.

(37) The included molecules are ammonia dimer, water dimer, formic acid dimer and formamide dimer from hydrogen bonded complexes, methane dimer, ethene dimer and benzene−methane from complexes with predominant dispersion contribution and ethene−ethine, benzene−water, benzene-ammonia, and benzene−HCN from mixed complexes, 11 complexes altogether.

(38) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(39) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. *Theor. Chim. Acta* **1990**, *77*, 291−306.

(40) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(41) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.