

# Unbiased auxiliary basis sets for accurate two-electron integral approximations

Cite as: J. Chem. Phys. **127**, 114107 (2007); <https://doi.org/10.1063/1.2777146>

Submitted: 28 June 2007 • Accepted: 06 August 2007 • Published Online: 19 September 2007

Francesco Aquilante, Roland Lindh and Thomas Bondo Pedersen



View Online



Export Citation

## ARTICLES YOU MAY BE INTERESTED IN

**Atomic Cholesky decompositions: A route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency**

The Journal of Chemical Physics **130**, 154107 (2009); <https://doi.org/10.1063/1.3116784>

**Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen**

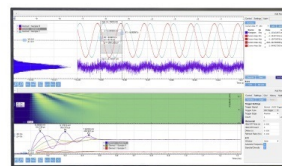
The Journal of Chemical Physics **90**, 1007 (1989); <https://doi.org/10.1063/1.456153>

**Modern quantum chemistry with [Open]Molcas**

The Journal of Chemical Physics **152**, 214117 (2020); <https://doi.org/10.1063/5.0004835>

Challenge us.

What are your needs for  
periodic signal detection?



Zurich  
Instruments

# Unbiased auxiliary basis sets for accurate two-electron integral approximations

Francesco Aquilante,<sup>a)</sup> Roland Lindh,<sup>b)</sup> and Thomas Bondo Pedersen<sup>c)</sup>

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

(Received 28 June 2007; accepted 6 August 2007; published online 19 September 2007)

We propose Cholesky decomposition (CD) of the atomic two-electron integral matrix as a robust and general technique for generating auxiliary basis sets for the density fitting approximation. The atomic CD (aCD) auxiliary basis set is calculated on the fly and is not biased toward a particular quantum chemical method. Moreover, the accuracy of the aCD basis set can be controlled with a single parameter. © 2007 American Institute of Physics. [DOI: 10.1063/1.2777146]

## I. INTRODUCTION

The Cholesky decomposition (CD) of electron repulsion integrals<sup>1</sup> and the density fitting (DF) approximation<sup>2</sup> are two efficient ways to significantly reduce the computational expense due to the two-electron integrals in *ab initio* and density functional theory (DFT) calculations. Although they may seem different, CD and DF are, in fact, closely related. This relation was already recognized by Beebe and Linderberg<sup>1</sup> in their original work on CD of the atomic orbital (AO) electron repulsion integral matrix. They explicitly demonstrated that the Cholesky decomposition is nothing but an inner projection of the Coulomb operator onto a specifically constructed basis. Similarly, Vahtras *et al.*<sup>3</sup> recognized the formal analogy between DF and CD. (In Ref. 3, as in the majority of subsequent work, the DF approximation is called the “resolution of identity” (RI) approximation. In this paper, however, we exclusively use the term density fitting.)

The importance of the DF approximation for the development of fast *ab initio* and DFT methods is well documented in the literature. Fast DF algorithms<sup>4–7</sup> are available for evaluating the Coulomb Fock matrix. More recently,<sup>8–10</sup> efforts in the difficult task of accelerating the computation of the Hartree-Fock (HF) exchange have led to fast DF-based algorithms for HF and hybrid DFT. The computational advantages of using the DF approximation in correlated calculations are also remarkable.<sup>11–13</sup> In particular, Ahlrichs and co-workers have made DF a method of large impact in the quantum chemistry community, as witnessed by the frequent use of this approach for theoretical studies of large systems of interest, especially in biochemistry. Their development of analytic gradients<sup>14,15</sup> and Hessians<sup>16,17</sup> within the DF approximation was instrumental in this respect. Ahlrichs and co-workers,<sup>4,18</sup> Weigend and Häser,<sup>14</sup> and Weigend *et al.*<sup>19</sup> have in a number of papers also benchmarked the method and have published various method-dependent auxiliary ba-

sis sets to be used in combination with standard Gaussian AO basis sets.

After a long period of limited interest,<sup>20–22</sup> CD has been used for speeding up correlated calculations by one of the authors.<sup>23–28</sup> Other applications of the CD approach have been limited to only a few recent studies,<sup>29–31</sup> possibly due to the lack of analytic derivatives. The CD approach has focused on an accuracy of the total energy which is several orders of magnitude better than the accepted error in DF calculations. The increased accuracy is accompanied by an increased computational cost compared to DF. Still, the Cholesky formulation of second-order coupled cluster linear response theory is at least one order of magnitude faster than the conventional algorithm.<sup>24</sup>

The limitations of the DF and CD approaches can be summarized as follows. The DF approximation suffers from the fact that externally designed auxiliary basis sets are required. Moreover, each auxiliary basis set is biased towards a specific quantum chemical method and there is no continuous way to improve the accuracy of the results if needed. The CD suffers from the lack of an analytic formulation of geometrical derivatives, and perhaps, from the fact that it has only been exploited in the high accuracy regime where it is not competitive in speed with the DF approach. The major advantages are as follows: The DF approach is very efficient and expressions for analytic gradients and Hessians have been derived and implemented; the CD approach does not require any preoptimized auxiliary basis set, and the accuracy of the method is fully controlled with a single parameter.

In this study, we will exploit the observation that DF and CD are equivalent to (1) introduce the one-center approximation to CD, making it more competitive with the DF approach, and (2) suggest a procedure for on-the-fly construction of unbiased DF auxiliary basis sets whose accuracy can be controlled by a single parameter.

## II. THEORY

The DF approach consists in expanding the real AO products  $\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})$  on a set of auxiliary real Gaussian basis functions  $\Phi_\alpha(\mathbf{r})$ , i.e.,

<sup>a)</sup>Electronic mail: francesco.aquilante@teokem.lu.se

<sup>b)</sup>Electronic mail: roland.lindh@chemphys.lu.se

<sup>c)</sup>Present address: Atomistix A/S, c/o Niels Bohr Institute, Rockefeller Complex, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark. Electronic mail: tpb@atomistix.com

$$\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}) \approx \sum_Q C_{\mu\nu}^Q \Phi_Q(\mathbf{r}). \quad (1)$$

With a given Hermitian positive definite metric  $\hat{g}(\mathbf{r}, \mathbf{r}')$ , minimizing the “distance” between the approximate and exact distributions,

$$\begin{aligned} f_{\mu\nu} &= \int d\mathbf{r} \int d\mathbf{r}' \left( \chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}) - \sum_P C_{\mu\nu}^P \Phi_P(\mathbf{r}) \right) \hat{g}(\mathbf{r}, \mathbf{r}') \\ &\quad \times \left( \chi_\mu(\mathbf{r}')\chi_\nu(\mathbf{r}') - \sum_Q C_{\mu\nu}^Q \Phi_Q(\mathbf{r}') \right) \\ &= \langle \mu\nu | \mu\nu \rangle_g - 2 \sum_P C_{\mu\nu}^P \langle \mu\nu | P \rangle_g + \sum_{PQ} C_{\mu\nu}^P C_{\mu\nu}^Q \langle P | Q \rangle_g, \end{aligned} \quad (2)$$

determines the expansion coefficients through the linear equations,

$$\sum_P \langle P | Q \rangle_g C_{\mu\nu}^P = \langle \mu\nu | Q \rangle_g. \quad (3)$$

The value of the function  $f_{\mu\nu}$  corresponding to the solution of Eq. (3) is given by

$$f_{\mu\nu}^{\min} = \langle \mu\nu | \mu\nu \rangle_g - \sum_{PQ} C_{\mu\nu}^P \langle P | Q \rangle_g C_{\mu\nu}^Q, \quad (4)$$

and thus measures the error in representing the diagonal element  $\langle \mu\nu | \mu\nu \rangle_g$ . The optimum choice of auxiliary basis set is therefore one for which  $f_{\mu\nu}^{\min}=0$  for all AO products  $\chi_\mu\chi_\nu$ .

Although other metrics have been investigated (see, e.g., Refs. 3 and 32), the Coulomb metric  $\hat{g}=|\mathbf{r}-\mathbf{r}'|^{-1}$  is usually employed in quantum chemistry. With this choice, the optimum auxiliary basis set is defined such that the two-electron integral matrix is represented exactly. However, the auxiliary basis functions are normally optimized for each AO basis set by minimization of the total energy error (with respect to conventional calculations) in a set of molecular calculations with no direct constraint regarding the representation of the integrals.<sup>4,8,19,33</sup> Not surprisingly, the accuracy of the resulting auxiliary basis sets are biased toward the quantum chemical method used in the optimization procedure. In order to obtain unbiased auxiliary basis sets for the DF approximation, we therefore seek an optimization procedure that explicitly takes into account the integral accuracy.

It is well known that the accuracy of the Cholesky integral representation can be controlled with a single parameter, the decomposition threshold  $\delta \geq 0$ .<sup>1,23,24</sup> The absolute error in each two-electron integral is bounded from above by  $\delta$ , leading to an integral representation which has no bias toward any particular quantum chemical method.<sup>1,23,24</sup> As discussed by Beebe and Linderberg,<sup>1</sup> the CD approach is equivalent to a Gram-Schmidt orthonormalization of the AO product functions in the Coulomb metric. Thus, the CD procedure can be viewed as a generation of a linearly independent (in the Coulomb metric) set of AO product functions which we here

refer to as the “Cholesky basis.” The Cholesky basis is close to the optimum choice of auxiliary basis set, as

$$f_{\mu\nu}^{\min} \leq \delta, \quad (5)$$

by construction.<sup>1,23</sup> It should be stressed that using the Cholesky basis as auxiliary basis set makes the DF and CD approaches completely equivalent.

For computational efficiency, the auxiliary basis functions are normally chosen to be centered on the nuclei. In general, however, the Cholesky basis contains product functions in which the AOs are centered on different nuclei. As a consequence, using the Cholesky basis as auxiliary basis set will require evaluation of four-center two-electron integrals, making the approach significantly more expensive than standard DF calculations where only three-center integrals need to be evaluated. Here, we investigate the accuracy of the CD procedure when only one-center product functions (i.e., both AOs of the product function are centered on the same nucleus) are allowed to enter the Cholesky basis. In this one-center CD (1C-CD) approach, the overall integral error is not bounded by  $\delta$ . Even with  $\delta=0$ , the 1C-CD procedure will not, in general, reproduce the four-center part of the integral matrix. On the other hand, as the number of potential Cholesky basis functions in 1C-CD is lower than that of full CD, the 1C-CD is substantially faster.

Even though the 1C-CD approach involves only nucleus-centered auxiliary basis functions, it still requires decomposition of an integral matrix whose dimension scales quadratically with the size of the molecular system. To bring down the computational cost to a minimum, we also investigate the Cholesky basis obtained from a CD of the atomic two-electron integral matrix. This atomic CD (aCD) auxiliary basis set needs not be stored in one of the usual basis set library formats but is constructed on-the-fly from the given AO basis set. As we shall see, the quality of the aCD auxiliary basis set is controlled by the aCD decomposition threshold. The computational overhead of performing the aCD is insignificant, as the decomposition needs only be done for each unique atom/AO basis set pair. As an example, provided that the same AO basis set is used on all carbon and hydrogen atoms, only two decompositions are needed to construct the full aCD auxiliary basis set for an alkane  $C_nH_{2n+2}$  regardless of the value of  $n$ .

The aCD approach is similar to the linear combination of atomic electron distributions (RI-LCAD) method of Ten-no and Iwata,<sup>34,35</sup> who constructed the auxiliary basis set through diagonalization of the atomic two-electron integral matrix. Eigenvectors corresponding to eigenvalues below a given threshold are removed, while the remaining ones constitute the auxiliary basis set. The number of nonredundant orbital products selected in this way grows only linearly with the size of the valence basis set. This feature is expected for the aCD as well. One difference with our approach is that the RI-LCAD basis sets are linear combination of all AO product functions  $\chi_\mu\chi_\nu$  while the aCD basis sets contain only a (non-redundant) subset of these. Although insignificant in most practical applications, we finally note that the need for a

matrix diagonalization makes the RI-LCAD more expensive than the aCD.

### III. RESULTS AND DISCUSSION

In this section, we compare the accuracy of the CD, 1C-CD, and DF approaches using both standard and aCD auxiliary basis sets for the latter. We compare both absolute and relative energies for the chemical reactions used by Baker and Chan<sup>36</sup> to test their procedure for locating transition states. All DFT calculations are performed on 20 of the 25 reactions of the test set, while HF and second-order Møller-Plesset (MP2) data are obtained for a subset of these (the 17 closed-shell species). For all selected reactions we have reoptimized the transition states and the products or reactants at the B3LYP (Becke's three-parameter functional<sup>37</sup> with the Lee-Yang-Parr correlation functional)<sup>38,39</sup> level of theory with the 6-31G\* basis set. Structures of all 40 molecular systems are provided in the supporting information.<sup>40</sup> All calculations are performed with a development version of the MOLCAS quantum chemistry software.<sup>41,42</sup>

#### A. Comparison of CD and conventional DF approaches

In this section, we compare the CD approach at a range of decomposition thresholds to the DF approach with auxiliary basis sets especially designed for nonhybrid DFT energies (RI-J) and MP2 correlation energy corrections (RI-C). All calculations are performed with the split-valence polarized (SVP) basis set.<sup>43</sup> For the DF calculations, the associated RI-J (Ref. 44) and RI-C (Ref. 19) auxiliary basis sets of Beglmann *et al.* are used. Calculations are performed at the SVWN5 (functional V of Ref. 45), B3LYP, HF, and MP2 levels of theory.

Table I reports the maximum absolute error and the root-mean-square (rms) error of the approximate calculations relative to conventional calculations on the test set by Baker and Chan. From Table I we first note that regardless of decomposition threshold, the CD approximation gives an accuracy which is as good as or much better than that of the DF approach. Furthermore, we observe that the CD is not biased toward any specific method, as opposed to the combination of the DF approach and RI-J or RI-C auxiliary basis sets. In particular, there seems to be an intrinsic deficiency of such basis sets in producing acceptable approximations for the HF-type exchange contributions. For both RI-J and RI-C, we observe relatively large errors for the B3LYP total energies, while the deviations in MP2 energies are no worse than the HF ones. This indicates that the major problem lies in a poor approximation of the HF-type exchange. It is worth noticing that this failure is not simply related to the number of functions in the auxiliary basis sets. In fact, although the RI-C sets are considerably larger than the RI-J ones, the B3LYP errors are roughly the same.

In Fig. 1 we report the average number of Cholesky vectors  $M$  used in the calculations of Table I. We observe that this number is larger than the number of RI-J or RI-C auxiliary functions which, by construction, is in the range

TABLE I. Test set of Baker and Chan. Absolute deviations with respect to conventional calculations of total and activation energies at the SVWN5, B3LYP, HF, and MP2 levels of theory with the SVP basis set. Density fitting with the RI-J and RI-C auxiliary basis sets compared to the CD and 1C-CD approximations for three different decomposition thresholds ( $\delta$ ).

Method	Auxiliary basis	Total energy		Activation energy	
		rms error	Max error	rms error	Max error
		(kcal mol <sup>-1</sup> )		(kcal mol <sup>-1</sup> )	
SVWN5	RI-J	0.288	0.504	0.079	0.144
	RI-C	122.732	538.655	1.057	4.500
	CD $\delta=10^{-4}$	0.136	0.260	0.047	0.090
	CD $\delta=10^{-5}$	0.032	0.057	0.021	0.036
	CD $\delta=10^{-6}$	0.003	0.008	0.001	0.004
	1C-CD $\delta=10^{-4}$	0.032	0.086	0.017	0.047
	1C-CD $\delta=10^{-5}$	0.019	0.062	0.018	0.062
	1C-CD $\delta=10^{-6}$	0.015	0.062	0.017	0.062
B3LYP	RI-J	193.031	491.443	1.476	4.782
	RI-C	131.597	396.156	0.938	3.937
	CD $\delta=10^{-4}$	0.115	0.214	0.049	0.088
	CD $\delta=10^{-5}$	0.012	0.053	0.021	0.035
	CD $\delta=10^{-6}$	0.002	0.007	0.001	0.004
	1C-CD $\delta=10^{-4}$	0.021	0.062	0.020	0.062
	1C-CD $\delta=10^{-5}$	0.117	0.219	0.042	0.094
	1C-CD $\delta=10^{-6}$	0.117	0.218	0.042	0.093
HF	RI-J	1041.693	2446.430	7.879	23.656
	RI-C	80.808	184.879	0.479	1.469
	CD $\delta=10^{-4}$	0.063	0.015	0.047	0.109
	CD $\delta=10^{-5}$	0.013	0.031	0.016	0.031
	CD $\delta=10^{-6}$	0.008	0.031	0.012	0.031
	1C-CD $\delta=10^{-4}$	0.110	0.281	0.032	0.094
	1C-CD $\delta=10^{-5}$	0.084	0.250	0.028	0.062
	1C-CD $\delta=10^{-6}$	0.076	0.250	0.027	0.094
MP2 <sup>a</sup>	RI-J	1048.610	2448.816	7.941	23.750
	RI-C	87.574	184.844	0.513	1.656
	CD $\delta=10^{-4}$	0.063	0.154	0.052	0.093
	CD $\delta=10^{-5}$	0.013	0.020	0.020	0.027
	CD $\delta=10^{-6}$	0.002	0.007	0.001	0.003
	1C-CD $\delta=10^{-4}$	0.129	0.281	0.035	0.094
	1C-CD $\delta=10^{-5}$	0.095	0.250	0.025	0.062
	1C-CD $\delta=10^{-6}$	0.088	0.250	0.024	0.062

<sup>a</sup>On top of DF-HF employing the same auxiliary basis.

from two to four times the number of AO basis functions  $N$ . The average number of Cholesky vectors increases roughly linearly with the logarithm of the decomposition threshold. Finally, the results of Table I clearly indicate that the CD method gives reliable and balanced results even with rather high thresholds ( $\delta=10^{-4}$ ).

Comparing the CD and 1C-CD results in Table I we find that the latter, too, is a balanced approximation: The 1C-CD errors are only slightly larger than the CD errors. However, while the CD errors can be rigorously controlled by the decomposition threshold, the accuracy of the 1C-CD approach cannot be improved beyond a certain limit. This is a consequence of the fact that at some decomposition threshold, all one-center product functions are employed as auxiliary functions and lowering the threshold therefore can not improve



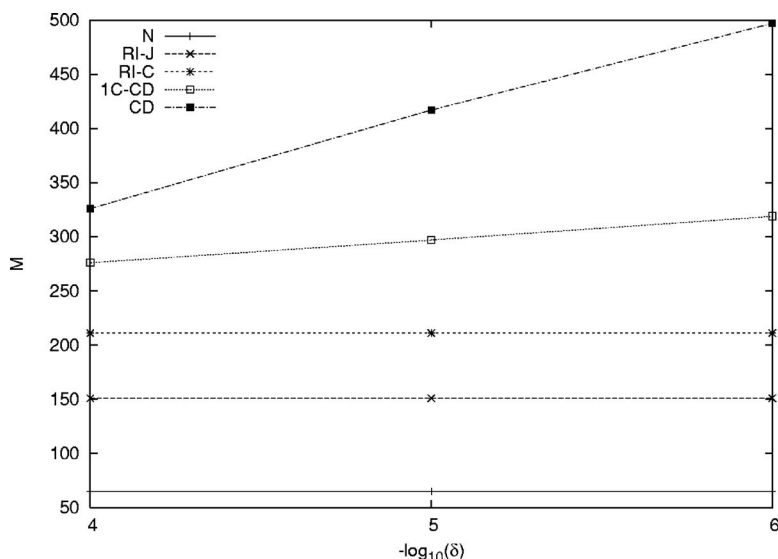


FIG. 1. Test set of Baker and Chan. Average number of Cholesky vectors needed at a given decomposition threshold. The average number of auxiliary RI-J and RI-C basis functions and of SVP basis functions ( $N$ ) is reported for comparison.

the accuracy. In other words, the complete one-center Cholesky basis is not able to span the full product space beyond a given accuracy. Nevertheless, 1C-CD retains the unbiased nature of the full CD approach and, to some limited extent, its error control. Finally, we note that even for  $\delta = 10^{-4}$ , 1C-CD provides better accuracy than the RI-J and RI-C auxiliary basis sets.

Due to the reduction in the number of Cholesky vectors to be generated (see Fig. 1) for 1C-CD compared to full CD, we have observed speedups up to 30% in the decomposition for large systems. The reduction of computational cost for the subsequent use of the Cholesky vectors is on the order of 20%. This implies that the 1C-CD is more competitive with standard DF approaches than the full CD. On the average, the 1C-CD ( $\delta = 10^{-4}$ ) requires only 30% more functions than the RI-C auxiliary basis set. At this cost, 1C-CD produces an unbiased auxiliary basis set which is more accurate than RI-C.

## B. Hierarchies of aCD auxiliary basis sets

From the Gaussian product theorem<sup>46</sup> it follows that the aCD sets consist of Gaussian functions not necessarily partitioned into shells of a given angular momentum. Rather, the aCD sets represent a subset of the cartesian components spanned by the full orbital product distributions. For a more practical generation and subsequent use of the aCD sets, we use a spherical harmonic representation. We thus use complete angular shells corresponding to the exponents and angular types of the Gaussian product functions selected by the aCD procedure. This technical aspect has no consequences for the accuracy and performance of the aCD auxiliary basis sets.

Hierarchies of increasingly accurate aCD auxiliary basis sets are easily generated by decreasing the decomposition threshold. As demonstrated above for the 1C-CD approach, however, there is a limit to the accuracy that can be achieved using only one-center product functions. Such a limit must be expected for the aCD auxiliary basis sets as well. In order to assess the accuracy of the aCD auxiliary basis sets, we start by analyzing the data in Table II. The total energy of the

benzene molecule is computed at different levels of theory and the absolute deviation from the results of conventional calculations are reported for each DF-based approach with different choices of the auxiliary basis set. The notation aCD- $n^*$  refers to the auxiliary basis set resulting from decomposition of the atomic two-electron integral matrices with decomposition threshold  $\delta = 10^{-n}$ . Noting that standard auxiliary basis sets, such as RI-J and RI-C, often do not contain the highest possible angular momentum orbitals spanned by the AO product functions, we also introduce the aCD- $n$  sets which are obtained from the aCD- $n^*$  sets by removing the highest angular momentum orbitals.

While the accuracy of the RI-J and RI-C sets is biased, the aCD- $n^*$  sets produce a nearly uniform error for all methods investigated. In order to understand the reasons behind this significant difference, we have computed the value of

TABLE II. Benzene molecule. Absolute deviations with respect to conventional calculations of the total energies at the BLYP (Becke's exchange functional (Ref. 47) with Lee-Yang-Parr correlation functional (Ref. 38 and 39), B3LYP, HF, and MP2 levels of theory with the SVP basis set. Density fitting with the RI-J, RI-C, and aCD auxiliary basis sets.

Auxiliary basis	$\Delta E/mE_h$			
	BLYP	B3LYP	HF	MP2 <sup>a</sup>
RI-J	0.866	19.679	95.238	101.147
RI-C	8.964	7.714	3.122	2.534
aCD-2*	5.318	1.889	11.718	12.542
aCD-3*	3.857	3.681	3.059	2.652
aCD-4*	0.014	0.016	0.151	0.208
aCD-5*	0.013	0.010	0.117	0.168
aCD-8*	0.012	0.001	0.107	0.156
aCD-2	81.708	74.928	47.712	41.683
aCD-3	27.690	27.573	26.910	24.210
aCD-4	0.487	0.285	0.257	0.932
aCD-5	0.473	0.291	0.199	0.835
aCD-8	0.116	0.009	0.417	1.015

<sup>a</sup>On top of DF-HF employing the same auxiliary basis.

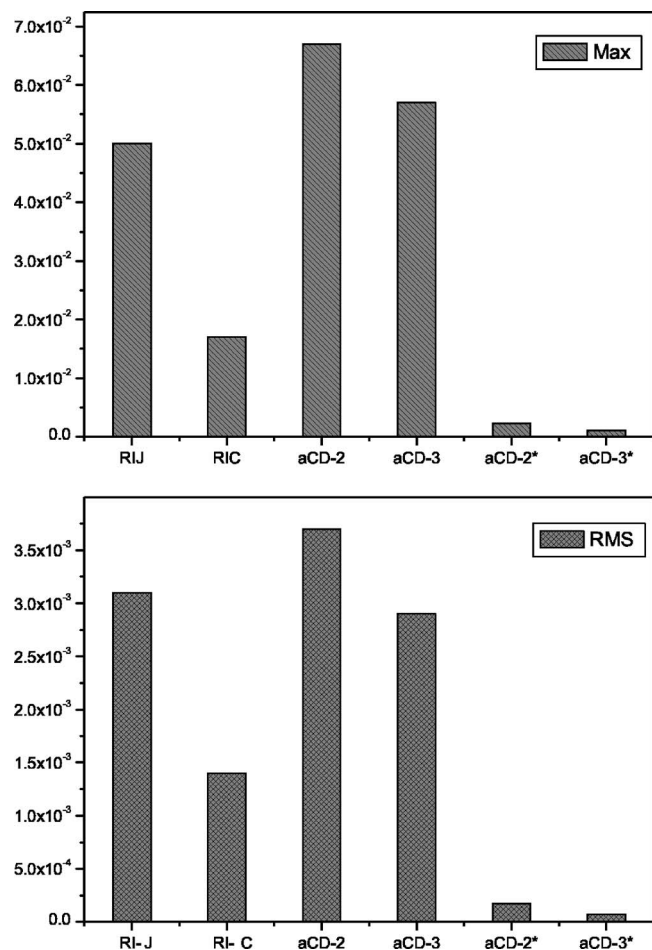


FIG. 2. Benzene molecule. Maximum and rms error in the representation of the diagonal AO two-electron integrals employing different auxiliary basis sets for the SVP valence basis set.

$f_{\mu\nu}^{\min}$  [Eq. (4)] for the benzene molecule. Figure 2 displays the maximum and rms values over the orbital distributions  $\chi_{\mu}\chi_{\nu}$ . The aCD- $n^*$  sets provide an accuracy which is one order of magnitude better than that of the standard auxiliary basis sets. Due to the positive definite nature of the two-electron integral matrix, the maximum errors are also upper bounds to the errors in the representation of any of the off-diagonal two-electron integrals. The aCD- $n$  sets, on the other hand, give a poor representation of the exact AO integrals, resulting in decreased accuracy of the energy. We also note that, as  $n$  increases, the change in accuracy is less smooth for the aCD- $n$  sets than for the aCD- $n^*$  ones. The  $f_{\mu\nu}^{\min}$  analysis reveals that even at a very tight threshold, such as  $n=8$ , the maximum error in the diagonal integrals is on the order of  $10^{-2}$  with the aCD- $n$  sets. Hence, there is no real improvement in going from aCD-2 to aCD-3, as confirmed also by the results in Table II.

The accuracy of computed energies, however, is not simply related to the accuracy of the integrals. For instance, how can the RI-J set produce accurate non-hybrid DFT energies although it fails to accurately reproduce the integral matrix? The results in Fig. 2 suggest that the opposite should happen. The answer is that, by construction, the RI-J set is able to reproduce just those integrals that most significantly contribute to the evaluation of the Coulomb energy, at least for

TABLE III. Test set of Baker and Chan. Absolute deviations with respect to conventional calculations of total and activation energies at the SVWN5, B3LYP, HF, and MP2 levels of theory with the SVP basis set. Density fitting with aCD auxiliary basis sets at four different thresholds.

Method	Auxiliary basis	Total energy		Activation energy	
		rms error	Max error	rms error	Max error
		(kcal mol <sup>-1</sup> )		(kcal mol <sup>-1</sup> )	
SVWN5	aCD-2 <sup>*</sup>	2.812	5.906	0.256	0.641
	aCD-3 <sup>*</sup>	1.945	4.500	0.166	0.469
	aCD-4 <sup>*</sup>	0.036	0.094	0.021	0.062
	aCD-5 <sup>*</sup>	0.025	0.062	0.019	0.062
	aCD-2	47.940	115.507	4.421	17.406
	aCD-3	12.558	29.533	1.445	4.437
	aCD-4	0.636	2.375	0.112	0.375
	aCD-5	0.211	0.391	0.052	0.156
B3LYP	aCD-2 <sup>*</sup>	0.839	1.719	0.178	0.453
	aCD-3 <sup>*</sup>	0.836	1.516	0.129	0.301
	aCD-4 <sup>*</sup>	0.021	0.062	0.013	0.031
	aCD-5 <sup>*</sup>	0.018	0.062	0.018	0.062
	aCD-2	43.641	105.565	4.202	16.531
	aCD-3	12.068	28.222	1.405	4.312
	aCD-4	0.582	2.281	0.116	0.406
	aCD-5	0.164	0.344	0.060	0.141
HF	aCD-2 <sup>*</sup>	0.844	17.000	0.382	1.312
	aCD-3 <sup>*</sup>	0.619	1.344	0.127	0.316
	aCD-4 <sup>*</sup>	0.093	0.359	0.084	0.336
	aCD-5 <sup>*</sup>	0.088	0.355	0.083	0.336
	aCD-2	25.829	59.668	3.631	12.156
	aCD-3	10.860	22.971	1.420	4.000
	aCD-4	0.435	1.840	0.125	0.343
	aCD-5	0.159	0.469	0.088	0.187
MP2 <sup>a</sup>	aCD-2 <sup>*</sup>	8.956	18.000	0.371	1.125
	aCD-3 <sup>*</sup>	0.590	1.406	0.115	0.293
	aCD-4 <sup>*</sup>	0.084	0.187	0.024	0.062
	aCD-5 <sup>*</sup>	0.082	0.187	0.022	0.047
	aCD-2	22.167	51.030	3.594	11.875
	aCD-3	9.728	20.872	1.334	3.781
	aCD-4	0.459	1.785	0.121	0.344
	aCD-5	0.284	0.562	0.071	0.148

<sup>a</sup>On top of DF-HF employing the same auxiliary basis.

ground state calculations. Our analysis is based on the full set of  $f_{\mu\nu}^{\min}$  values and, therefore, is not completely applicable to the type of auxiliary basis sets obtained by data fitting. The aCD sets mimic the full CD which is the key to their high accuracy for all methods investigated.

Our conclusions concerning the unbiased nature of the aCD sets are confirmed by the results in Table III. Compared with Table I, we note the similarity between the 1C-CD approximation and the aCD. This analogy is also seen in Fig. 3. In fact, the aCD and 1C-CD results are virtually identical for very tight thresholds. Moreover, the similarity of the two approaches is manifested also in the size of the two auxiliary basis sets, as clearly seen from the curves in Figs. 1 and 4. The average number of auxiliary basis functions employed in the DF calculations of the test set by Baker and Chan is shown in Fig. 4 for various aCD approximations. Clearly, the

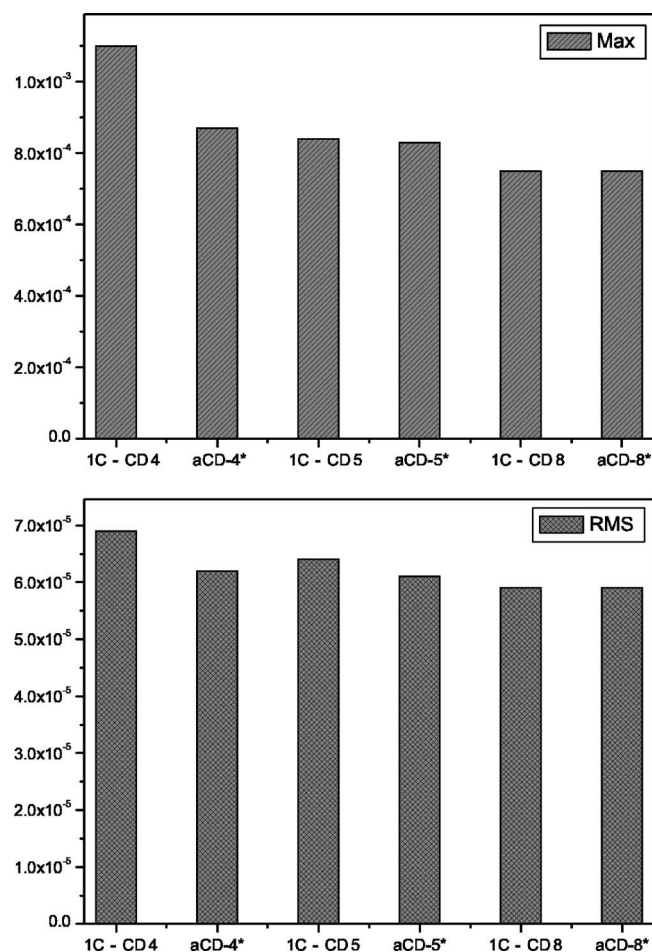


FIG. 3. Benzene molecule. Maximum and rms error in the representation of the diagonal AO two-electron integrals employing various 1C-CD and aCD approximations for the SVP valence basis set.

aCD- $n^*$  and aCD- $n$  sets are somewhat larger than the corresponding RI-C ones. We find that this as a reasonable price to pay for an unbiased and highly accurate auxiliary basis set.

Table IV reports errors obtained with the aCD- $n$  sets generated from correlation consistent AO basis sets. The

TABLE IV. Test set of Baker and Chan. Absolute deviations with respect to conventional calculations of total and activation energies at the HF and MP2 levels of theory with correlation consistent basis sets. Density fitting with aCD auxiliary basis sets at three different thresholds.

		Total energy		Activation energy	
Method	Auxiliary basis	rms error	Max error	rms error	Max error
		(kcal mol <sup>-1</sup> )		(kcal mol <sup>-1</sup> )	
cc-pVDZ					
HF	aCD-3	0.289	0.969	0.077	0.156
	aCD-4	0.222	0.719	0.081	0.141
	aCD-5	0.201	0.594	0.050	0.086
MP2 <sup>a</sup>	aCD-3	0.908	2.656	0.098	0.234
	aCD-4	0.363	0.750	0.071	0.219
	aCD-5	0.316	0.625	0.055	0.094
cc-pVTZ					
HF	aCD-3	0.340	0.609	0.049	0.074
	aCD-4	0.126	0.234	0.051	0.078
	aCD-5	0.128	2.341	0.039	0.062
MP <sup>a</sup>	aCD-3	0.964	1.609	0.372	0.641
	aCD-4	0.752	1.484	0.350	0.609
	aCD-5	0.558	0.984	0.218	0.344

<sup>a</sup>On top of DF-HF employing the same auxiliary basis.

small errors indicate that the aCD approach for generating auxiliary basis sets can indeed be applied with any AO basis set.

## IV. SUMMARY

Based on the close relationship between DF and CD, we advocate the use of atomic CD for on-the-fly generation of unbiased auxiliary basis sets. The accuracy of these auxiliary basis sets can be controlled by a single parameter, the decomposition threshold. We have also analyzed the performance of a modified CD approach, 1C-CD, in which only one-center AO product distributions are allowed to contrib-

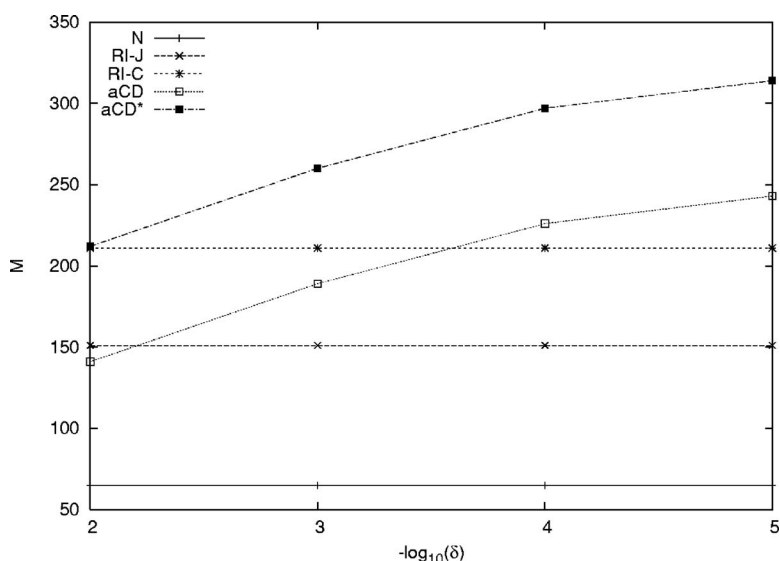


FIG. 4. Test set of Baker and Chan. Average number of aCD auxiliary basis functions needed at a given atomic decomposition threshold. The average number of auxiliary RI-J and RI-C basis functions and of SVP basis functions ( $N$ ) is reported for comparison.

ute. While the 1C-CD approach is significantly faster than full CD, it does not downgrade the accuracy to any significant degree.

*Note added in proof:* Yang, Rendell, and Frisch<sup>48</sup> have recently proposed another on-the-fly generation of auxiliary basis sets. This class of auxiliary basis sets is, however, designed for Coulomb fitting only.

## ACKNOWLEDGMENTS

Funding from the Swedish Research Council (VR) is gratefully acknowledged. Part of this work was carried out under the HPC-EUROPA project (RII3-CT-2003-506079), with the support of the European Community -Research Infrastructure Action of the FP6. We thank M. Sc. Laura Muresan for help with the graphics in this article.

<sup>1</sup>N. H. F. Beebe and J. Linderberg, *Int. J. Quantum Chem.* **7**, 683 (1977).

<sup>2</sup>J. L. Whitten, *J. Chem. Phys.* **58**, 4496 (1973).

<sup>3</sup>O. Vahtras, J. Almlöf, and M. Feyereisen, *Chem. Phys. Lett.* **213**, 514 (1993).

<sup>4</sup>K. Eichkorn, O. Treutler, H. Öm, M. Häser, and R. Ahlrichs, *Chem. Phys. Lett.* **240**, 283 (1995).

<sup>5</sup>F. R. Manby, P. J. Knowles, and A. W. Lloyd, *J. Chem. Phys.* **115**, 9144 (2001).

<sup>6</sup>M. Sierka, A. Hogeckamp, and R. Ahlrichs, *J. Chem. Phys.* **118**, 9136 (2003).

<sup>7</sup>A. Sodt, J. E. Subotnik, and M. Head-Gordon, *J. Chem. Phys.* **125**, 194109 (2006).

<sup>8</sup>F. Weigend, *Phys. Chem. Chem. Phys.* **4**, 4285 (2002).

<sup>9</sup>R. Polly, H. J. Werner, F. R. Manby, and P. J. Knowles, *Mol. Phys.* **102**, 2311 (2004).

<sup>10</sup>F. Aquilante, T. B. Pedersen, and R. Lindh, *J. Chem. Phys.* **126**, 194106 (2007).

<sup>11</sup>A. P. Rendell and T. J. Lee, *J. Chem. Phys.* **101**, 400 (1994).

<sup>12</sup>C. Hättig and F. Weigend, *J. Chem. Phys.* **113**, 5154 (2000).

<sup>13</sup>H. J. Werner, F. R. Manby, and P. J. Knowles, *J. Chem. Phys.* **118**, 8149 (2003).

<sup>14</sup>F. Weigend and M. Häser, *Theor. Chim. Acta* **97**, 331 (1997).

<sup>15</sup>D. Rappoport and F. Furche, *J. Chem. Phys.* **122**, 064105 (2005).

<sup>16</sup>P. Deglmann, F. Furche, and R. Ahlrichs, *Chem. Phys. Lett.* **362**, 511 (2002).

<sup>17</sup>P. Deglmann, K. May, F. Furche, and R. Ahlrichs, *Chem. Phys. Lett.* **384**, 103 (2004).

<sup>18</sup>K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, *Theor. Chim. Acta*

**97**, 119 (1997).

<sup>19</sup>F. Weigend, A. Köhn, and C. Hättig, *J. Chem. Phys.* **116**, 3175 (2002).

<sup>20</sup>I. Røeggen and E. Wisløff-Nielsen, *Chem. Phys. Lett.* **132**, 154 (1986).

<sup>21</sup>D. W. O'Neal and J. Simons, *Int. J. Quantum Chem.* **36**, 673 (1989).

<sup>22</sup>S. Wilson, *Comput. Phys. Commun.* **58**, 71 (1990).

<sup>23</sup>H. Koch, A. Sánchez de Merás, and T. B. Pedersen, *J. Chem. Phys.* **118**, 9481 (2003).

<sup>24</sup>T. B. Pedersen, A. M. J. Sánchez de Merás, and H. Koch, *J. Chem. Phys.* **120**, 8887 (2004).

<sup>25</sup>I. García Cuesta, T. B. Pedersen, H. Koch, and A. M. J. Sánchez de Merás, *Chem. Phys. Lett.* **390**, 170 (2004).

<sup>26</sup>T. B. Pedersen, H. Koch, L. Boman, and A. M. J. Sánchez de Merás, *Chem. Phys. Lett.* **393**, 319 (2004).

<sup>27</sup>I. García Cuesta, T. B. Pedersen, H. Koch, and A. Sánchez de Merás, *ChemPhysChem* **7**, 2503 (2006).

<sup>28</sup>B. Fernández, T. B. Pedersen, A. Sánchez de Merás, and H. Koch, *Chem. Phys. Lett.* **441**, 332 (2007).

<sup>29</sup>I. Røeggen, *J. Chem. Phys.* **124**, 184502 (2006).

<sup>30</sup>F. Aquilante, T. B. Pedersen, H. Koch, and A. Sánchez de Merás, *J. Chem. Phys.* **125**, 174101 (2006).

<sup>31</sup>A. Öhrn and F. Aquilante, *Phys. Chem. Chem. Phys.* **9**, 470 (2007).

<sup>32</sup>Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6692 (2005).

<sup>33</sup>D. E. Bernholdt and R. J. Harrison, *J. Chem. Phys.* **109**, 1593 (1998).

<sup>34</sup>S. Ten-no and S. Iwata, *Chem. Phys. Lett.* **240**, 578 (1995).

<sup>35</sup>S. Ten-no and S. Iwata, *J. Chem. Phys.* **105**, 3604 (1996).

<sup>36</sup>J. Baker and F. Chan, *J. Comput. Chem.* **17**, 888 (1996).

<sup>37</sup>A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).

<sup>38</sup>C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).

<sup>39</sup>B. Miehl, A. Savin, H. Stoll, and H. Preuss, *Chem. Phys. Lett.* **157**, 200 (1989).

<sup>40</sup>See EPAPS Document No. E-JCPSA6-127-308735 for details of the structures of the reactants and transition states selected from the Baker test set of reactions. This document can be reached through a direct link in the online article's HTML reference section or via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>).

<sup>41</sup>MOLCAS 6, University of Lund, Sweden, 2005 (see <http://www.teokem.lu.se/molcas/>).

<sup>42</sup>G. Karlström, R. Lindh, P. Malmqvist *et al.*, *Comput. Mater. Sci.* **28**, 222 (2003).

<sup>43</sup>A. Schäfer, H. Horn, and R. Ahlrichs, *J. Chem. Phys.* **97**, 2571 (1992).

<sup>44</sup>F. Weigend, *Phys. Chem. Chem. Phys.* **8**, 1057 (2006).

<sup>45</sup>S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).

<sup>46</sup>T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2000).

<sup>47</sup>A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).

<sup>48</sup>R. Yang, A. P. Rendell, and M. J. Frisch, *J. Chem. Phys.* **127**, 074102 (2007).