

The versatility of the Cholesky decomposition in electronic structure theory

Thomas Bondo Pedersen¹  | Susi Lehtola²  | Ignacio Fdez. Galván³  | Roland Lindh³ 

¹Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, Oslo, Norway

²Department of Chemistry, University of Helsinki, Helsinki, Finland

³Department of Chemistry, BMC, Uppsala University, Uppsala, Sweden

Correspondence

Roland Lindh, Department of Chemistry, BMC, Uppsala University, Uppsala, Sweden.

Email: roland.lindh@kemi.uu.se

Funding information

Academy of Finland, Grant/Award Numbers: 350282, 353749; Norges Forskningsråd, Grant/Award Number: 262695; Vetenskapsrådet, Grant/Award Number: 2020-03182

Edited by: Anna Krylov, Associate Editor and Peter R. Schreiner, Editor-in-Chief

Abstract

The resolution-of-the-identity (RI) or density fitting (DF) approximation for the electron repulsion integrals (ERIs) has become a standard component of accelerated and reduced-scaling implementations of first-principles Gaussian-type orbital electronic-structure methods. The Cholesky decomposition (CD) of the ERIs has also become increasingly deployed across quantum chemistry packages in the last decade, even though its early applications were mostly limited to high-accuracy methods such as coupled-cluster theory and multiconfigurational approaches. Starting with a summary of the basic theory underpinning both the CD and RI/DF approximations, thus underlining the extremely close relation of the CD and RI/DF techniques, we provide a brief and largely chronological review of the evolution of the CD approach from its birth in 1977 to its current state. In addition to being a purely numerical procedure for handling ERIs, thus providing robust and computationally efficient approximations to the exact ERIs that have been found increasingly useful on modern computer platforms, CD also offers highly accurate approaches for generating auxiliary basis sets for the RI/DF approximation on the fly due to the deep mathematical connection between the two approaches. In this review, we aim to provide a concise reference of the main techniques employed in various CD approaches in electronic structure theory, to exemplify the connection between the CD and RI/DF approaches, and to clarify the state of the art to guide new implementations of CD approaches across electronic structure programs.

This article is categorized under:

Electronic Structure Theory > Ab Initio Electronic Structure Methods

Software > Quantum Chemistry

Quantum Computing > Theory Development

KEYWORDS

auxiliary basis sets, Cholesky decomposition, reduced scaling, two-electron integrals

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *WIREs Computational Molecular Science* published by Wiley Periodicals LLC.

1 | INTRODUCTION

The development of ab initio methods over the last 40–50 years has focused on harnessing the elusive locality of quantum effects in molecular systems, as the proper use of this locality enables computational studies of much larger molecular systems with the available computer resources. Almost inevitably, one then faces the question of the compact representation of the electron repulsion integrals (ERIs), which can be achieved by rank-reducing factorizations, for instance.

In a seminal paper published in 1977, Beebe and Linderberg¹ suggested that the symmetric positive semi-definite matrix of the ERIs could be efficiently represented through the use of an *incomplete* Cholesky decomposition (CD),^{2,3} whose accuracy is controlled with a single adjustable parameter. Studying electronic structure calculations in a basis of contracted Gaussian-type orbitals, they demonstrated that the ERIs can be transformed into the molecular orbital (MO) basis faster when the CD approach is employed than when the conventional ERI approach is employed. As many wave function theories such as Møller–Plesset perturbation theory are conventionally written in terms of MO integrals, and as the MO transformation step may dominate the computational cost, the advantages of faster integral transforms can be considerable in many contexts.

However, Beebe and Linderberg also predicted that the explicit use of the ERIs in ab initio methods would be “unnecessary,” as algorithms could be redesigned to take advantage of the CD factorization of the ERIs; indeed, this is a key aspect of modern implementations of electronic structure algorithms employing the CD. Furthermore, Beebe and Linderberg foresaw that the CD procedure could be advantageous also for other types of positive definite matrices in MO theory, such as the one- and two-particle reduced density matrices; many applications of this type can be found in the literature, as well.

This focus article is a chronological account of the use of CD in ab initio theory for the efficient representation of ERIs as well as for other purposes. In addition to the CD, we will also discuss resolution-of-the-identity (RI) or density-fitting (DF) approaches. The two terms—RI and DF—are often used interchangeably in the literature (as will also be done in this work), although the latter is more often used in the context of computing classical Coulomb interactions between two densities.

As we will discuss in this work, there is an extremely close relation between the RI/DF and the CD approaches; in fact, as will be shown in Section 2.1, the CD of the ERIs can be viewed as RI/DF in an auxiliary basis of atomic orbital (AO) products chosen by the pivoted Cholesky decomposition procedure.^{4,5} Accompanied with this connection, our review will focus around the auxiliary basis sets used in the RI/DF approach, and we will especially review the automatic generation of auxiliary basis sets for RI/DF approach using the CD. Despite the kinship between the CD and RI/DF approaches, the considerable literature on the RI/DF technique will not be reviewed in detail in this work due to the present focus on CD techniques.

The layout of this work is as follows. Next, in Section 2, we will discuss the CD approach. The section begins with a presentation of the “vanilla” CD algorithm found in the mathematical literature. However, electronic structure applications typically employ CD for so-called Gram matrices, such as the overlap or Coulomb overlap matrix. As discussed in Section 2.1, the CD of such matrices can be viewed as a Gram–Schmidt orthogonalization of the associated basis functions, and this connection can be used to fashion more efficient CD algorithms. The review then continues with literature applications of the CD on ERI decompositions and other methods in Section 2.2. The RI/DF approach is briefly presented in Section 3, after which we review various approaches to build auxiliary basis sets for RI/DF in Section 3.1. The review ends with a brief conclusion in Section 4. This review can be supplemented with the more technically oriented and verbose report on the subject in Reference [6]. A brief technical description of analytical gradients in association with CD is presented elsewhere.⁷

2 | CHOLESKY DECOMPOSITIONS

The Cholesky decomposition² (CD) is given by

$$\mathbf{M} \approx \mathbf{L}\mathbf{L}^T, \quad (1)$$

where the Cholesky vectors \mathbf{L} form a lower (or upper) triangular matrix. The CD can be constructed for any symmetric positive semi-definite matrix \mathbf{M} ; see Reference [8] and references therein for the mathematical background.

The Cholesky factorization was originally introduced in electronic structure theory by Beebe and Linderberg¹ to decompose the electron repulsion integrals (ERIs). Assuming real AOs or MOs $\{\chi_\mu\}$, the ERIs in chemists' also known as Mulliken notation are given by

$$(\mu\nu|\kappa\lambda) = \int \frac{\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})\chi_\kappa(\mathbf{r}')\chi_\lambda(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}d^3\mathbf{r}', \quad (2)$$

where $\mu, \nu, \kappa, \lambda$ are the AO/MO indices. Importantly, the ERI tensor is positive semi-definite,^{9–12} and can be viewed as a measure of the distance of the orbital products $\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})$ and $\chi_\kappa(\mathbf{r}')\chi_\lambda(\mathbf{r}')$ in the Coulomb metric $g(\mathbf{r}, \mathbf{r}') = |\mathbf{r}-\mathbf{r}'|^{-1}$. Because of the positive semi-definiteness, we can apply Equation (1) to decompose the ERIs as

$$(\mu\nu|\kappa\lambda) \approx \sum_n L_{\mu\nu}^n L_{\kappa\lambda}^n. \quad (3)$$

However, as the real ERIs defined by Equation (2) satisfy symmetries with the interchange of the indices $\mu \leftrightarrow \nu$, $\kappa \leftrightarrow \lambda$, and $\mu\nu \leftrightarrow \kappa\lambda$, one also has to require that the elements of the Cholesky vectors \mathbf{L}^n satisfy $L_{\mu\nu}^n = L_{\nu\mu}^n$, because $\chi_\mu\chi_\nu$ and $\chi_\nu\chi_\mu$ are the same orbital product. This symmetry is commonly employed to describe the pair $\mu\nu$ with a single compound index, which we will denote by i . Choosing $|i\rangle = |\mu_i\nu_i\rangle$ with $\nu_i \leq \mu_i$ (or equivalently with $\mu_i \leq \nu_i$), the symmetry of $L_{\mu\nu}^n = L_{\nu\mu}^n$ is thus explicitly built-in in the mapping $\mu\nu \leftrightarrow i$, and the length of the Cholesky vectors \mathbf{L}^n is decreased by roughly one half. Employing the compound indices, the ERI tensor $(\mu\nu|\kappa\lambda) := (\mathbf{V})_{\mu\nu,\kappa\lambda}$ can now be written as a matrix $(\mathbf{V})_{ij}$, which is also positive semi-definite. We can then write the Cholesky decomposition in matrix $(\mathbf{V} = \mathbf{L}\mathbf{L}^T)$ or component form $(V_{ij} = \sum_{n=1}^N L_{i,n}L_{j,n})$, where N is the size of the orbital product basis.

In practice, the Cholesky vectors $\mathbf{L}^n = (L_{1,n}, L_{2,n}, L_{3,n}, \dots, L_{N,n})^T$ are computed one by one by an iterative procedure. As was already mentioned, a full mathematical analysis can be found in Reference [8]; in the following, we will only outline the main parts of the procedure.

The key to understanding the pivoted Cholesky algorithm is that it is controlled by the errors of the diagonal elements of the matrix to be decomposed (here \mathbf{V}). Because the whole procedure boils around the diagonal remainder, it is stored in memory for the whole procedure as a vector \mathbf{d} .

At the beginning, the error vector is simply given by the diagonal of \mathbf{V} , $\mathbf{d}^{(0)} = \text{diag}(\mathbf{V})$. Whenever new Cholesky vectors $\mathbf{L}^{(n)}$ are computed, the error vector is updated correspondingly $\mathbf{d}^{(n)} = \mathbf{d}^{(n-1)} - \text{diag}[\mathbf{L}^{(n)}(\mathbf{L}^{(n)})^T]$, that is, by

$$d_i = d_i - L_{i,n}^2. \quad (4)$$

Importantly, it is easy to see that the residual matrix $\mathbf{V} - \sum_n \mathbf{L}^{(n)}(\mathbf{L}^{(n)})^T$ remains symmetric positive semi-definite for any n ; this also means that $d_i \geq 0$ in the absence of errors due to finite precision.

As the sum of the diagonal errors $\sum_i d_i$ measures the error of the current Cholesky approximation, an optimally convergent algorithm that achieves maximal error reduction per iteration is obtained by zeroing out the largest error in \mathbf{d} at every iteration.⁸ This is achieved by sorting \mathbf{d} in descending order, and choosing the entry corresponding to the largest (so far unprocessed) entry as the pivot index, that is, as the next Cholesky vector to generate.

What does the algorithm for the generation of the Cholesky vectors look like? In the following, we try to maximize the similarity of the notation with Harbrecht et al.⁸ for clarity. However, as we wish to generate our Cholesky vectors as columns, we define our Cholesky matrix as the transpose of those of Harbrecht et al., $\mathbf{L} = \mathbf{L}^T$.

Following Harbrecht et al., let π denote the row/column indices of the pivoting, so that π_i is the index of the i th pivot function, which denotes the orbital product $\mu_i\nu_i$. The computation of the Cholesky vector for iteration $m \in [1, N]$ starts out by setting the diagonal element¹

$$L_{\pi_m, m} = \sqrt{d_{\pi_m}}; \quad (5)$$

while the off-diagonal elements $i \in [m+1, \dots, N]$ are computed as $L_{i,n} = (\mathbf{V} - \mathbf{L}\mathbf{L}^T)_{i,n} / \sqrt{d_n}$; with the pivoting this reads⁸

¹It is easy to see that when no pivoting is used ($\pi_i = i$), Equations (5) and (6) indeed correspond to diagonal and off-diagonal elements of the Cholesky vector.

$$L_{\pi_i,m} = \frac{1}{L_{\pi_m,m}} \left(V_{\pi_i,\pi_m} - \sum_{j=1}^{m-1} L_{\pi_i,j} L_{\pi_m,j} \right). \quad (6)$$

When the Cholesky vector $\mathbf{L}^n = (\{L_{i,n}\}_{i=1}^N)$ defined by Equations (5) and (6) has been computed, the diagonal error is updated by Equation (4). One can especially observe from Equations (4) and (5) that the resulting $d_{\pi_m} = 0$, that is, the largest error is zeroed out at every iteration, as was already mentioned above. This is the reason for the optimal convergence of the CD algorithm.

After each iteration, the remaining error in the decomposition at iteration m can be computed from

$$\epsilon^{\text{sum}} = \sum_{i=m+1}^N d_{\pi_i}. \quad (7)$$

The procedure of Harbrecht et al.⁸ is stopped when ϵ reaches a predefined tolerance threshold τ , $\epsilon < \tau$. In contrast, quantum chemical implementations of the Cholesky decomposition commonly employ a different termination criterion that is defined by the maximum element of the error vector

$$\epsilon^{\text{max}} = \max_{i > m} d_{\pi_i}. \quad (8)$$

Clearly, both stopping criteria of Equations (7) and (8) lead to the exact decomposition at the limit $\tau \rightarrow 0$. However, the former will lead to larger numbers of Cholesky vectors for a fixed value of τ than the latter criterion.

Having laid out the algorithmic implementation of the procedure, we can already make some general observations about the method. Perhaps the defining property of the Cholesky decomposition behind its popularity in electronic structure theory is exactly that the iterative procedure can be formulated in a “direct”¹³ manner, such that only the diagonal elements \mathbf{d} of the supermatrix \mathbf{V} and the individual Cholesky vectors $\mathbf{L}^{(n)}$ need to be stored. The storage cost is therefore essentially determined by that of the Cholesky vectors, themselves. However, because the Cholesky decomposition efficiently picks out any linear dependencies in the contents of the decomposed matrix—which are formidable in the case of the orbital product basis used in the ERI tensor—the number of resulting Cholesky vectors is often much smaller than the dimensions of the original matrix, and in practice the number of Cholesky vectors N_{CD} scales linearly with the number of orbital basis functions N_{OBF} as $N_{\text{CD}} = \gamma N_{\text{OBF}}$ with typical values of $\gamma \in [3, 10]$, depending entirely on the employed decomposition threshold τ .

We conclude this section by noting that Beebe and Linderberg speculated that a proper implementation of the CD approach would make the computation of the integrals scale cubically with system size and the integral transformation scale quartically with respect to the size of the orbital basis set. This has since been verified by a number of efficient implementations, which brings us naturally over to the next subsection, which discusses the way in which the CD of the ERIs is computed in modern implementations.

2.1 | Connection to orthogonalization methods

An important connection between the Cholesky decomposition discussed above in the general case and matrix orthogonalization methods was originally discussed by Aquilante et al.^{4,5} Namely, when applied to a Gram matrix, that is, an overlap matrix in the usual metric $g(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')$ or a Coulomb overlap matrix in the case of the Coulomb metric $g(\mathbf{r}, \mathbf{r}') = |\mathbf{r} - \mathbf{r}'|^{-1}$ employed in the ERIs, the CD can in fact be seen to be equivalent to a Gram–Schmidt orthogonalization procedure. This connection is therefore not a general property of the CD, but a consequence of the context of its application to matrices that describe overlaps of functions in some function space. Due to its practical importance, we will devote this subsection to elucidating this connection originally pointed out in References [4,5] by rederiving it from the general equations shown above for the CD.

The argument of Aquilante et al.^{4,5} is as follows. Let us assume that we have performed the Cholesky decomposition of the ERIs, which has yielded us a series of pivot functions $|\pi_j\rangle = |\mu_j \nu_j\rangle$. Let us now examine what happens if we apply the Gram–Schmidt orthogonalization technique to this set of pivot functions. The Gram–Schmidt procedure determines

orthonormalized functions $|Q_j\rangle$ that are obtained by first projecting out previously processed vectors from the original functions

$$|\tilde{Q}_j\rangle = |\pi_j\rangle - \sum_{k=1}^{j-1} |Q_k\rangle(Q_k|\pi_j), \quad (9)$$

and then by normalizing the remainder

$$|Q_j\rangle = \left(\tilde{Q}_j|\tilde{Q}_j\right)^{-1/2} |\tilde{Q}_j\rangle. \quad (10)$$

It is now straightforward to make the connection that^{4,5}

$$(\mu\nu|Q_j) = L_{\mu\nu,j}. \quad (11)$$

First, it is obvious from Equations (5) and (6) that the first function generated by the Gram–Schmidt procedure satisfies $(\mu\nu|Q_1) = (\pi_1|\pi_1)^{-1/2}(\mu\nu|\pi_1) = L_{\mu\nu,1}$, as the first orthonormalized function is simply the first pivot function in normalized form.⁴ Having established that Equation (11) holds for the first function, we can prove the connection in general by the use of induction.

Computing $(\mu\nu|\tilde{Q}_m)$ from Equations (9) and (10) we have

$$(\mu\nu|Q_m) = \left(\tilde{Q}_m|\tilde{Q}_m\right)^{-1/2} (\mu\nu|\tilde{Q}_m), \quad (12)$$

$$(\mu\nu|\tilde{Q}_m) = (\mu\nu|\pi_m) - \sum_{k=1}^{m-1} (\mu\nu|Q_k)(Q_k|\pi_m). \quad (13)$$

On the one hand, we can rewrite Equation (13) using the claim of Equation (11) as

$$(\mu\nu|\tilde{Q}_m) = (\mu\nu|\pi_m) - \sum_{k=1}^{m-1} L_{\mu\nu,k}L_{\pi_m,k}. \quad (14)$$

On the other hand, Equation (6) reads in the same notation as

$$L_{\mu\nu,m} = \frac{1}{L_{\pi_m,m}} \left((\mu\nu|\pi_m) - \sum_{j=1}^{m-1} L_{\mu\nu,j}L_{\pi_m,j} \right). \quad (15)$$

The right hand side of Equation (14) is clearly equal to the parenthesis Equation (15). But Equation (15) still contains the “diagonal” element, and to complete the proof we need to show that this factor is the normalization integral in Equation (10). This is also easy to show: according to Equation (9) the normalization integral in Equation (12) is

$$\left(\tilde{Q}_m|\tilde{Q}_m\right) = (\pi_m|\pi_m) - \sum_{k=1}^{m-1} (\pi_m|Q_k)(Q_k|\pi_m), \quad (16)$$

$$= (\pi_m|\pi_m) - \sum_{k=1}^{m-1} L_{\pi_m,k}L_{\pi_m,k}, \quad (17)$$

$$= d_{\pi_m}^{(m)} = L_{\pi_m,m}^2, \quad (18)$$

where we have again used Equation (11) to arrive from Equation (16) to Equation (17), and then applied Equation (4) to identify the diagonal error in $V - LL^T$ at iteration m , which is the “diagonal” element of the Cholesky vector according to Equation (5) in Equation (18), q.e.d.

We have thus reproduced the proof of References [4,5] that when applied to an overlap (a.k.a. Gram) matrix, the pivoted Cholesky decomposition is analogous to performing Gram–Schmidt orthogonalization on the sequence of pivot functions chosen by the CD. Why did we go through this trouble? This connection is of immense practical use. Not only does it give an intuitive understanding of what the Cholesky decomposition does when applied to the ERIs, but also it can be used to greatly simplify the computation of the Cholesky decomposition—as well as that of its derivatives. One can especially see that *the Cholesky decomposition of the ERI is the same as RI/DF in the basis of the orbital products represented by the pivot indices*; this will be discussed in detail in Section 3.

Realizing that the Cholesky decomposition can be performed in two steps—(i) identification of the pivot functions and (ii) formation of the Cholesky vectors—allows the use of a different type of algorithm than the single-step approach presented in Section 2. As we will discuss in the literature review in Section 2.2, modern CD approaches rely on such two-step procedures,⁶ which afford the following advantages.

When Equation (8) is employed as the stopping criterion, the first step can be optimized by throwing out any orbital products i with diagonal errors $d_i < \tau$ at every iteration. Because the orbital product basis contains many linear dependencies, the number of significant elements in the \mathbf{d} vector decreases rapidly, meaning that the consequent pivot indices become faster and faster to determine. Once the pivot indices have been determined, the Cholesky vectors can be efficiently computed from orthogonalized three-center integrals as in the RI/DF technique, which is discussed in Section 3. This also obviates the need to keep the Cholesky vectors in memory, as the vectors can be computed from the pivot indices on-the-fly following established RI/DF methodologies.

2.2 | Literature review

What follows will primarily be a review on developments on the use of CD technology in electronic structure theory. Key contributions specific to RI/DF are also mentioned when relevant to the presentation.

A first necessary step for chemical applications of the CD procedure was to find a way to compute nuclear gradients. Initially, O’Neal and Simons¹⁴ proposed augmenting the ERI matrix with the derivative integrals before carrying out the CD procedure. However, this limits the accuracy of the computed gradients to the Cholesky threshold, requiring the use of extremely small decomposition thresholds τ , which is clearly unsatisfactory. We will shortly discuss later developments that allow exact gradients for CDs determined with rather generous thresholds.

Further development of the CD approach was slow during the 1980s and 1990s.^{14,15} In the meantime, the development of the RI/DF approximation¹⁶ to ERIs offered an attractive and simple way around the somewhat awkward and not too efficient recursive nature of the original version of the CD procedure discussed in Section 2: the RI/DF approach discussed in Section 3 determines a reduced rank representation of the orbital product basis set by the introduction of an auxiliary basis set. The expansion coefficients of the ERIs in the auxiliary basis can be determined variationally, but the strict control over the accuracy of the resulting approach is lost in RI/DF, as it employs a fixed auxiliary basis set instead of choosing the auxiliary functions adaptively as in CD (Section 2.1). However, the loss of guaranteed precision is not an issue for many standard chemical applications, as discussed by Weigend et al.,¹⁷ for instance.

In 2003, Koch et al.¹⁸ developed the first modern CD implementation for Hartree–Fock (HF) and second-order Møller–Plesset (MP2) theories in the Dalton software package,¹⁹ following the original procedure of Section 2. In the following year, they reported implementations for computing dynamic polarizabilities at the approximate coupled-cluster singles-and-doubles (CC2) level of theory.²⁰ Both implementations demonstrated significant reductions in processor time compared with the use of conventional ERI techniques.

These efforts inspired the MolCAS²² developers to implement and investigate the CD approach, as well. First, Aquilante et al.²¹ developed the local-exchange (LK) method for fast evaluation of the exchange contribution to the Fock matrix (see Figure 1), the direct Coulomb contribution being inexpensive and straightforward. Implementations of CD-based scaled-opposite-spin MP2 theory,²³ complete active space self-consistent field (CASSCF)²⁴ theory, and multiconfigurational second-order perturbation theory (CASPT2)²⁵ followed in a rapid sequence.

The derivative issue was also addressed and solved in a different way from the original approach of O’Neal and Simons¹⁴: Aquilante et al.⁴ argued that the proper derivatives are not generated by CD applied to the derivative integrals but rather by differentiating the Cholesky approximation itself, Equation (3). However, the recursive nature of the

Cholesky basis is a significant complication in such an implementation. Yet, as discussed in Section 2.1, Aquilante et al. showed that Equation (3) could equally well be expressed as

$$\begin{aligned}(\mu\nu|\kappa\lambda) &\approx \sum_n L_{\mu\nu}^n L_{\kappa\lambda}^n \\ &= \sum_{mn} (\mu\nu|\pi_m)(\mathbf{V}^{-1})_{mn}(\pi_n|\kappa\lambda),\end{aligned}\quad (19)$$

where $\{\pi_m\}$ again denote orbital products $\mu_m\nu_m$ chosen by the CD as pivot indices, and the Coulomb matrix \mathbf{V} is defined to have the matrix elements $\mathbf{V}_{m,n} = (\pi_m|\pi_n)$. The key discovery here is that the derivatives of Equation (19) are straightforward to compute with minor modifications of the algorithms developed for RI/DF techniques discussed in Section 3, obviating the need to take the derivatives of the iterative CD procedure discussed in Section 2.

Similarly motivated by the efforts by Koch et al. and the MolCAS developers, Røeggen and Johansen²⁶ suggested an efficient CD algorithm for use with family type basis sets that employ the same Gaussian exponents for various angular momenta.

As suggested by Beebe and Linderberg,¹ the CD procedure also turns out to be useful for other quantities than the ERIs. For instance, Aquilante et al.²⁷ used CD to generate localized MOs in a non-iterative fashion from the one-particle density matrix in the AO basis, while Boman et al.²⁸ applied the CD approach in a brilliant work to the matrices occurring in the energy expressions for a given method. For example, the direct Coulomb contribution to the energy in HF, CASSCF, and Kohn–Sham density-functional theory (KS-DFT) can be written as

$$E_C = \sum_{\mu\nu,\kappa\lambda} D_{\mu\nu}(\mu\nu|\kappa\lambda)D_{\kappa\lambda} = \sum_{\mu\nu,\kappa\lambda} M_{\mu\nu,\kappa\lambda},\quad (20)$$

where the CD procedure can now be applied to the matrix \mathbf{M} that describes products of density matrix elements $D_{\mu\nu}$ with the corresponding AOs $\chi_{\mu}\chi_{\nu}$ which together describe the electron density participating in the Coulomb interaction; this yields the *Coulomb decomposition*. Boman et al. also reported three versions adapted for the HF exchange energy, which differ in the ways the contributions are screened. Overall, they found that the length of the CD can be reduced

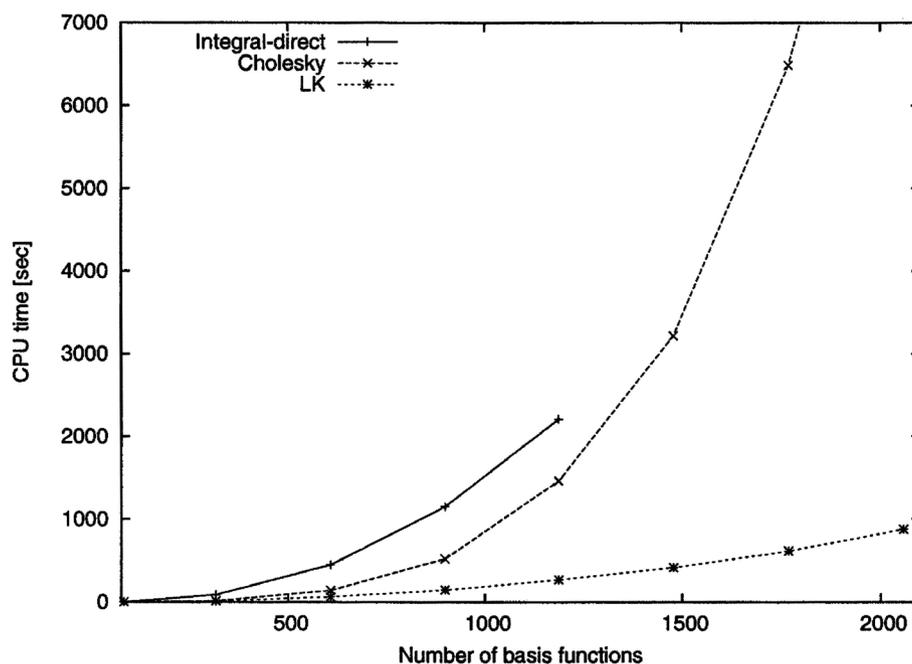


FIGURE 1 Comparison of the scaling of the CPU time for the exchange Fock matrix build for linear alkanes using standard Cholesky, the local exchange (LK) Cholesky, and the conventional integral-direct methods using a CD threshold of $\tau = 10^{-4} E_h$ and the cc-pVTZ basis set. (Reprinted from Reference [21] with permission from AIP Publishing.)

by an order of magnitude or more compared with the standard CD applied to the ERIs, and that the associated reductions in computational time may reach two orders of magnitude.²⁸

Another distinct application of the CD procedure is by Lehtola,²⁹ who followed the techniques discussed later in Section 3 to cure overcompleteness in molecular electronic structure calculations with AO basis sets. Even when the AO basis is orthonormal on a single atom, the overlap matrix may develop pathological singularities from interatomic blocks that make electronic structure calculations numerically ill-conditioned. However, in analogy to the methods discussed in Section 3, the CD approach can be used to identify subsets of orbital basis functions that are able to span the whole set of atomic-orbital basis functions²⁹; per the discussion in Section 2.1, the method of Reference [29] can be seen as an optimal Gram–Schmidt orthogonalization of the AO basis. This automated procedure to remove pathological linear dependencies enables the routine use of extremely diffuse basis functions to describe weakly bound electrons,²⁹ extended basis sets for benchmark quality calculations,³⁰ as well as studying highly repulsive parts of potential energy surfaces where nuclei may even be on top of each other.³¹

The developments of the CD method for ERIs over the last 10–15 years have included implementations of a wide range of ab initio methods, including Green's function methods,³² an adaptive metric for fragment molecular orbital MP2,³³ equation-of-motion MP2,³⁴ and second-order symmetry-adapted perturbation theory.³⁵ Implementations for nuclear gradients and non-adiabatic coupling vectors at the CASSCF and CASPT2 level of theory have been presented by various groups.^{36–40} Carter et al. have used the procedure in linear-scaling and localized multireference configuration interaction (MRCI) theory,^{41,42} while Lehtola et al. have used it for studies on the Perdew–Zunger self-interaction correction^{43,44} and the perfect pairing hierarchy,⁴⁵ which is a family of truncated coupled-cluster (CC) methods. Other types of CC theories have seen use of Cholesky decomposed ERIs by others. An extensive list includes reports on implementations with parallel CC singles-and-doubles with perturbative triples corrections (CCSD(T)),⁴⁶ for the evaluation of equation-of-motion CC (EOM-CC),⁴⁷ in connection with efficient CC singles (CCS) and CC singles-and-doubles (CCSD) implementations on graphics processing units (GPUs),⁴⁸ a combination of reverse Cuthill–McKee transformation in combination with CD for compact two-electron integral representation in combination with MP2 and linear coupled-cluster with doubles l-DCC,⁴⁹ for the evaluation of analytical gradients for the CCSD and EOM-CCSD methods using the explicit derivatives of the iteratively generated Cholesky vectors,⁵⁰ for the CCS, CCSD, and CCSDT models combining CD representation of the ERIs with tensor-reduction schemes for the triples CC amplitudes,⁵¹ and most recently for CCSD gradients⁵² using the approach of Aquilante et al.⁴ to compute the derivative integrals in the parent basis set.

To continue, Bozkaya et al. have implemented over the last few years in a collective effort CD and RI/DF for various orbital-optimized schemes (the OMP2, OMP2.5, and OCCSD methods),^{53–55} quasi-degenerate perturbation theory,⁵⁶ as well as energies and analytic gradients for OCCSD, and EOM-CCSD.^{57,58} These CD approaches are available³⁵ in the Psi4 program package.⁵⁹ Similarly, Hohenstein et al. have presented impressive implementations of CD ERI methods in connection with reduced-rank CC theory on GPUs,⁶⁰ combined with tensor hypercontraction of the doubles amplitudes.⁶¹

In this list of implementations of CD ERIs in ab initio methods one also finds a variational orbital-optimized two-particle reduced-density-matrix method,⁶² a self-consistent-field valence bond (VBSCF) method,⁶³ and an implementation in density matrix renormalization group (DMRG) second-order *N*-electron valence state perturbation theory (NEVPT2).⁶⁴ Piccardo and Soncini⁶⁵ have also formulated a modified version of the procedure that can be applied to both the symmetric ERI matrix and the antisymmetric matrix of the two-electron spin–orbit integrals.

The recently developed e^T software package⁶⁶ for CC methods has been designed to exclusively employ the CD representation of the ERIs. Folkestad et al.⁶⁷ presented a highly efficient implementation of the two-step CD procedure of Aquilante et al.⁶ that only stores the pivot indices from which the integrals can be computed as needed. Folkestad et al.⁶⁷ have used this procedure in e^T in CCSD calculations on systems with up to 80,000 AO basis functions, see Figure 2.

In parallel with the work of Folkestad et al., Lew-Yee et al.⁶⁸ reported two algorithms on the combination of asymmetric DF together with CD ERIs in a second-order propagator implementation. Most recently, Zhang et al.⁶⁹ reported a careful analysis of the computational costs of the two-step procedure, and determined the optimal way to implement CD with respect to memory footprint and number of floating-point operations.

The CD procedure is now starting to be incorporated to its full extent also in other program packages than DALTON,⁷⁰ OPENMOLCAS,⁷¹ e^T ,⁶⁶ and Psi4.⁵⁹ Helmich-Paris et al.⁷² described the implementation of relativistic Cholesky-decomposed density matrix (CDD) second-order Møller–Plesset perturbation theory (MP2) energies in the Dirac program.⁷³ Nottoli et al.⁷⁴ have recently reported several implementations using CD ERIs in the CFOUR software

package,⁷⁵ while Blaschke and Stopkowicz⁷⁰ presented an extension of the CD technique for complex ERIs arising from the use of London orbitals.⁷⁶ Employing these techniques, Nottoli et al. have reported a quadratically convergent SCF (QCSCF) implementation,⁷⁴ a second-order multi-configuration SCF (MCSCF) implementation⁷⁷ that yields a significant reduction in the computational cost and memory requirements and improves parallel performance, and studied nuclear magnetic resonance shielding constants and molecules in strong magnetic fields.^{78,79} Finally, extremely recently, the application of CD and lower-upper (LU) techniques to a spin-free infinite-order two-component relativistic Hamiltonian in a modified version of the GAMESS program has been discussed by Takashima and Nakai.⁸⁰

3 | RI/DF APPROACH

Could the ERIs be computed more efficiently using an intermediary basis set? This question is the defining characteristic of the RI/DF approach and it has been investigated by several research groups.^{11,16,81–88}

The central idea in the RI/DF approach is that the orbital product basis set can be expanded in terms of a smaller *auxiliary* basis set as

$$\chi_{\mu}(\mathbf{r})\chi_{\nu}(\mathbf{r}) \approx \sum_K C_{\mu\nu}^K \theta_K(\mathbf{r}). \quad (21)$$

The expansion of Equation (21) was originally investigated by Whitten¹¹ as a way to get an efficient representation of the electron densities when computing Coulomb potential energies. The expansion coefficients $C_{\mu\nu}^K$ are obtained by maximizing the fit between the input density in the orbital product basis and the auxiliary basis expansion. Vahtras et al.⁸⁹ studied such fits determined within the Coulomb and overlap metrics, and found the former to yield the most accurate results. This is easily understood, since fitting in the Coulomb metric minimizes the error in the interaction energy,

$$\epsilon_{\mu\nu,\kappa\lambda} = \left(\mu\nu - \sum_L C_{\mu\nu}^L \theta_L | \kappa\lambda - \sum_K C_{\kappa\lambda}^K \theta_K \right). \quad (22)$$

Hence, we find the following relation at the minimum

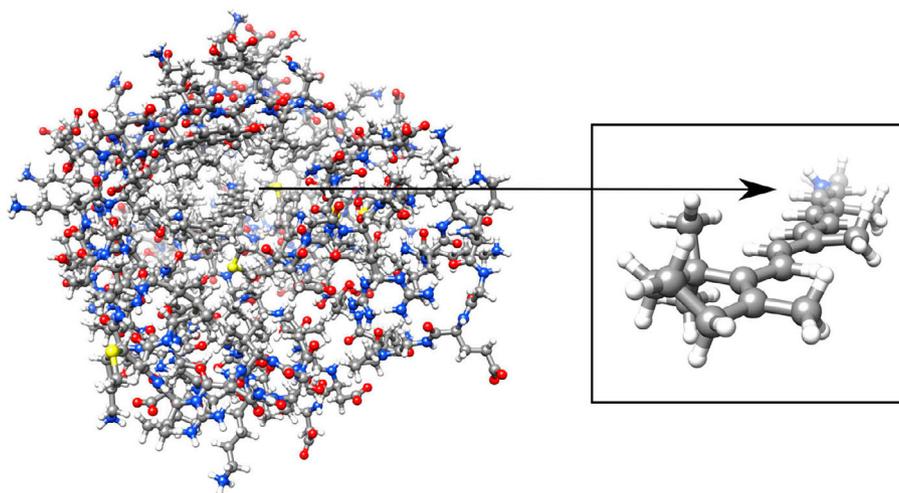


FIGURE 2 Retinal bound to rhodopsin with active retinal was used by Folkestad et al. to exemplify the capacity of their CD implementation in e^T . In this calculation, the one-center version of the CD of the ERIs was used in association with an aug-cc-pVTZ basis set, with $N_{AO} = 79,420$ AO basis functions. (Reprinted from Reference [67] with permission from AIP Publishing.)

$$\frac{\partial \epsilon_{\mu\nu,\kappa\lambda}}{\partial C_{\kappa\lambda}^K} = \left(\mu\nu - \sum_L C_{\mu\nu}^L \theta_L | \theta_K \right) = 0, \quad (23)$$

which yields

$$C_{\mu\nu}^L = \sum_K (\mu\nu|K) (\mathbf{V}^{-1})_{KL}, \quad (24)$$

where K and L are indices used as a shorthand for the auxiliary basis functions θ_K and θ_L , and $(\mu\nu|K)$ and $\mathbf{V}_{KL} = (K|L)$ are the so-called 3- and 2-center ERIs, respectively, which are evaluated over the orbital and auxiliary basis functions. For a recent comparison on the accuracy of various fitting metrics, we refer to the work of Duchemin et al.⁹⁰

The approximation in Equation (21) results in the following approximation for the ERIs:

$$(\mu\nu|\kappa\lambda) \approx \sum_{KL} C_{\mu\nu}^K \mathbf{V}_{KL} C_{\kappa\lambda}^L. \quad (25)$$

When the coefficients are determined by the Coulomb fitting procedure of Equation (24), Equation (25) yields

$$(\mu\nu|\kappa\lambda) \approx \sum_{KL} (\mu\nu|K) (\mathbf{V}^{-1})_{KL} (L|\kappa\lambda). \quad (26)$$

This expression is now seen to be identical to Equation (19): in Equation (19), the pivot functions chosen by the CD take on the role of the chosen auxiliary basis in Equation (26), which enables the reuse of RI/DF techniques within the two-step CD scheme as discussed in Section 2.1.

The connection to the CD technique also allows us to see that the maximum absolute error in the approximate ERIs of the RI/DF procedure is bounded from above by $\max_{\mu\nu,\kappa\lambda} |\epsilon_{\mu\nu,\kappa\lambda}|$, whose value is now determined by the chosen auxiliary basis set. This error can be made small by a judicious choice of the auxiliary basis set: an extended auxiliary basis set will afford small errors.

It has also been recently pointed out that a one-center correction to the RI/DF procedure in which all one-center two-electron integrals are computed exactly can afford reduced RI/DF errors.⁹¹

3.1 | Auxiliary basis sets

We will now review efforts to build accurate auxiliary basis sets along two prongs. The first is to optimize the auxiliary basis sets for a specific purpose, such as for the cost-efficient calculation of the direct Coulomb contribution to the Fock matrix, which is of immense practical importance in standard chemical applications. The second is to try to build general-use basis sets by some automated means, starting from the given orbital basis set. In particular, the CD technique will be seen to guarantee small errors by construction, yielding a *universal* approximation of the ERIs with controlled accuracy.

Once the considerable benefits of the RI/DF approach were discovered, researchers took on the quest of finding optimal auxiliary basis sets. Vahtras et al.⁸⁹ noted that although an exact θ_K in Equation (21) can be found with the Gaussian product theorem, this would not result in any reduction in the required computational effort. In any case, a successful approximation should reproduce the HF energies with suitably small errors, which requires reproducing the correct behavior of isolated atoms, for instance. Hence, Vahtras et al.⁸⁹ stated that “*the expansion basis set should in principle include all one-center products of the original basis set.*” The race was now on to find such auxiliary basis sets.

One way to derive such precomputed auxiliary basis sets was proposed by Ten-no and Iwata. Following Vahtras et al., the linear combination of atomic electron densities (LCAD) approach^{92,93} chooses all single-center orbital products as auxiliary functions. A slight complication of this choice are the angular redundancies (ARs) due to duplicate orbital products of the same total angular momentum, and the Coulombic redundancies (CRs) due to the overcompleteness of the set of orbital products, which are removed in the procedure.^{92,93} Ten-no and Iwata⁹³ demonstrated significant speedups for multiconfigurational calculations with the LCAD scheme with negligible errors (~ 0.1 kcal mol⁻¹).

The group around Ahlrichs and the TURBOMOLE program⁹⁴ did solid work along a different route and produced a preoptimized auxiliary basis set for the standard orbital basis sets in Turbomole at the time, which cover the chemically relevant part (H–Rn) of the periodic table. This RI-J auxiliary basis set was designed to reproduce the direct Coulomb contribution \mathbf{J} to the Fock matrix,^{95,96} significantly speeding up density functional calculations with semilocal functionals. The auxiliary basis was optimized by minimizing the error in the Coulomb energy of isolated atoms and hydrides with a Monte Carlo procedure. The resulting error was found to be around 0.1 mE_h per atom in the early generations of auxiliary basis sets. Equally importantly, the auxiliary basis set was found to be about three times the size of the orbital basis set, $N_{\text{aux}} \approx 3N_{\text{OBF}}$.

The work was extended to the evaluation of the MP2 correlation energy by numerical optimization of exponents and contraction coefficients using an error metric defined by the direct MP2 energy correction,⁹⁷ yielding RI-MP2 auxiliary basis sets. The number of primitives and the contraction pattern of the auxiliary basis were determined based on ad hoc rules. An error of 0.06 mE_h per atom was achieved with a slightly larger auxiliary basis set than the RI-J basis set, $N_{\text{aux}} \approx 4N_{\text{OBF}}$.

It is worth mentioning that the optimization procedures for the RI-J and RI-MP2 auxiliary basis sets employed error metrics associated with total energies, which means that the errors in the values of the individual ERIs was not considered. Reproducing all ERIs generally requires many more auxiliary functions than reproducing just the Coulomb or the MP2 correlation energy. Still, it should be noted that these RI implementations for DFT and MP2 made the Turbomole package one of the few packages at the time that could routinely simulate molecules of a decent size with reasonably sized AO basis sets.

A number of additional publications on auxiliary basis sets adapted to special conditions, other orbital basis sets than the TURBOMOLE ones, and other energy models (for example, both direct Coulomb and exchange contributions), have followed the footsteps outlined by the TURBOMOLE developers.^{98–104} Some families of auxiliary basis sets, such as the OptRI family, have been especially developed for explicitly correlated calculations.^{105–107} However, it is worth noting that no precomputed auxiliary basis sets had been reported for, for example, multiconfigurational electronic structure models at this time (2008). This was possibly a manifestation of the somewhat tedious and ad hoc way in which the preoptimized auxiliary basis sets are constructed: if you had just derived a new orbital basis set or designed a new electronic structure model, you were at best on a waiting list to get the associated auxiliary basis set(s), or you had to optimize them yourself. Furthermore, optimized auxiliary basis sets are almost invariably tailored for reproducing differences in ground state total energies, and may not be accurate for modeling all kinds of molecular properties such as magnetic or core spectroscopies, highlighting the need for alternate avenues.

The first attempt to automate the construction of RI-J auxiliary basis sets was reported by Yang et al.,¹⁰⁸ who developed an ad hoc scheme with 11 steps: the auto-ABS procedure that is controlled by three adjustable parameters. Auto-ABS produces an uncontracted auxiliary basis and employs each exponent over an angular momentum range specific to that exponent. This algorithm was implemented in the Gaussian program and its accuracy in the reproduction of the direct Coulomb energy was compared with that of the so-called “universal” auxiliary functions of Weigend.¹⁰⁹ Yang et al.¹⁰⁸ reported that Auto-ABS reproduces an accuracy similar to that of the auxiliary basis set of Weigend, even though the auto-ABS auxiliary basis can be as much as four times larger than that of Weigend.

When Aquilante and Pedersen worked on implementing the CD method for CASSCF and CASPT2 energy calculations, and Lindh on implementing RI/DF technology for the HF and density functional methods—all three working within the MolCAS package—they started to see similarities between their formulations that were beyond pure coincidence. This should not have been a surprise to anyone, as both techniques are realizations of the inner projections famously explored by Löwdin,^{110,111} as was also elaborated by Vahtras et al. in the educational paper on integral approximations in HF calculations.⁸⁹

Again, as discussed in Sections 2.1 and 3, the only difference between CD and RI/DF ERIs is the choice of the auxiliary basis set. The RI/DF approximation of the ERI can be written in the same form as the CD representation by symmetric splitting of the inverse overlap matrix

$$\begin{aligned}
 & \sum_{KL} (\mu\nu|K) (\mathbf{V}^{-1})_{KL} (L|\kappa\lambda) \\
 &= \sum_{MKL} (\mu\nu|K) (\mathbf{V}^{-1/2})_{KM} (\mathbf{V}^{-1/2})_{ML} (L|\kappa\lambda) \\
 &= \sum_M R_{\mu\nu}^M R_{\kappa\lambda}^M,
 \end{aligned} \tag{27}$$

yielding an expression similar to Equation (3), even when the two-center Coulomb overlap matrix ($K|L$) contains linear dependencies that are removed. Thus, the difference between RI/DF and CD is just *which set* of functions is used as the auxiliary basis: the pretabulated auxiliary basis functions in RI/DF versus the orbital product functions chosen automatically by the CD procedure. Modern implementations of either method work by forming the two-center overlap ($K|L$) and orthogonalizing it by, for example, Löwdin's canonical orthogonalization method,^{112,113} while pathological linear dependencies in the auxiliary basis set can be handled again with CD approaches.^{29,114}

Considering the remarkable success and accuracy of the RI/DF procedure employing a one-center auxiliary basis set, one can ask if the one-center constraint also works within the context of the CD decomposition of the ERIs. The answer to this question is a resounding yes: not only is the resulting one-center CD (1C-CD)¹¹⁵ procedure almost as accurate as the full four-center CD, but also it is about four times faster than full CD. Although a major improvement, the 1C-CD approach still has some potential flaws. It still requires a costly iterative procedure employing the procedure of Section 2 or Section 2.1. Also, even though the auxiliary basis now only contains one-center product functions, the auxiliary basis can change with the molecular geometry, which may result in discontinuities in potential energy surfaces.

Both of these problems can be easily fixed by applying the CD procedure on individual atomic blocks of the ERI supermatrix, thus defining the atomic CD (aCD) procedure.¹¹⁵ Again, a significant reduction in timings is observed compared with full-CD or 1C-CD without a significant loss of accuracy. Of course, neither 1C-CD or aCD provide global ERI error control any more, as only the errors of the one-center ERIs singled out by Vahtras et al.⁸⁹ are controllable.

It is prudent to compare aCD with the LCAD approach of Ten-no and Iwata.^{92,93} The 1C-CD, aCD, and LCAD approaches all work in the original orbital product basis set, but with different selection of the threshold for elimination of (near-)linear dependence. The former two methods typically use a threshold which is four orders of magnitude more loose.

Another important fact to stress is that in contrast to the standard RI/DF auxiliary basis sets that are typically optimized to reproduce total energies for a given method, the full-CD, 1C-CD, LCAD, and aCD families of auxiliary basis sets *approximate the integrals themselves*, meaning that no accidental cancellation of errors is relied upon for the accuracy of the resulting approach, see Figure 3.

However, the three approaches still share an exhaustive list of primitive orbital products, while standard RI/DF basis sets are no more dense than ordinary orbital basis sets. There must be more redundancy that can be removed! Aquilante et al.¹¹⁶ demonstrated that if an additional CD procedure is applied to an atomic ERI list, now expressed in terms of the products of uncontracted, primitive Gaussians instead of products of AOs of contracted Gaussians, significant reductions can indeed be achieved in the primitive Gaussian approximation of the product basis. The resulting atomic compact CD (acCD) auxiliary basis set contains far fewer functions than the original aCD auxiliary basis set, as seen in Figure 4.

Although the acCD auxiliary basis sets are larger than the standard TURBOMOLE auxiliary basis sets, for instance, the acCD basis sets come with several advantages. The acCD basis set can be derived on the fly from any orbital basis set, eliminating the need for pretabulation by a painstaking optimization procedure for each combination of orbital basis, method, and targeted property. The acCD basis is a general-purpose auxiliary basis in that it can be used for any ab initio wave function or density functional model with similar accuracy, and it can also be expected to work well for various molecular properties. Moreover, the accuracy of the approximation is controllable with a single threshold, τ . In short, the CD approach, whether atomic, one-center, or full, can be viewed as a systematic approximation of the two-electron interaction terms of the second-quantized Hamiltonian.

A calibration study of full-CD, 1C-CD, aCD, and acCD thresholds with respect to HF, non-hybrid, and hybrid density functional theory, and MP2 total energies demonstrated satisfactorily small errors of 0.01 kcal mol⁻¹ per electron in conjunction with a CD threshold of $\tau = 10^{-4} E_h$,⁵ compared with the use of the exact ERIs. The same threshold reproduced errors for CASSCF and CASPT2 valence and Rydberg excitation energies of 0.001 eV,¹¹⁷ again compared with the use of exact ERIs. MP2, CCSD, and CCSD(T) interaction energies of weakly bound dimers are found to exhibit errors in the range 0.003 to 0.03 kcal mol⁻¹ compared with the use of conventional ERIs.¹¹⁸ Furthermore, the same benchmark studies show that the error arising from the RI/DF procedure with the aCD and acCD auxiliary basis sets is virtually insignificant for high-quality orbital basis sets. For multiconfigurational methods, such as the CASSCF and CASPT2 models, the use of the acCD auxiliary basis set effectively eliminated the relative timing bottleneck due to the treatment of the ERIs. In a 2009 comparative study of 1C-CD, aCD, and acCD approaches versus the Turbomole RI/DF auxiliary basis sets, Weigend et al.¹⁷ noted that the latter auxiliary basis sets are slimmer but still reproduce energies

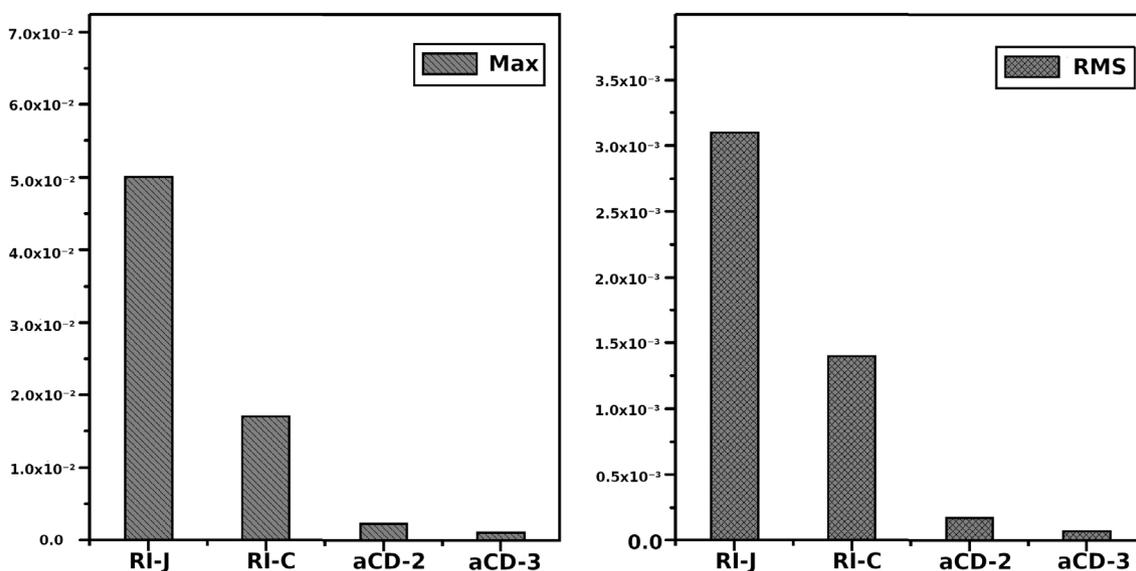


FIGURE 3 A representation from the benzene molecule of the errors in reproducing the ERIs. The maximum (left) and root-mean-square (RMS, right) errors of the AO ERIs are presented employing the RI-J, RI-C, and two aCD auxiliary basis sets for the SVP orbital basis set. For the two aCD basis sets the thresholds of $10^{-2} E_h$ and $10^{-3} E_h$ were used. (Adapted from Reference [115].)

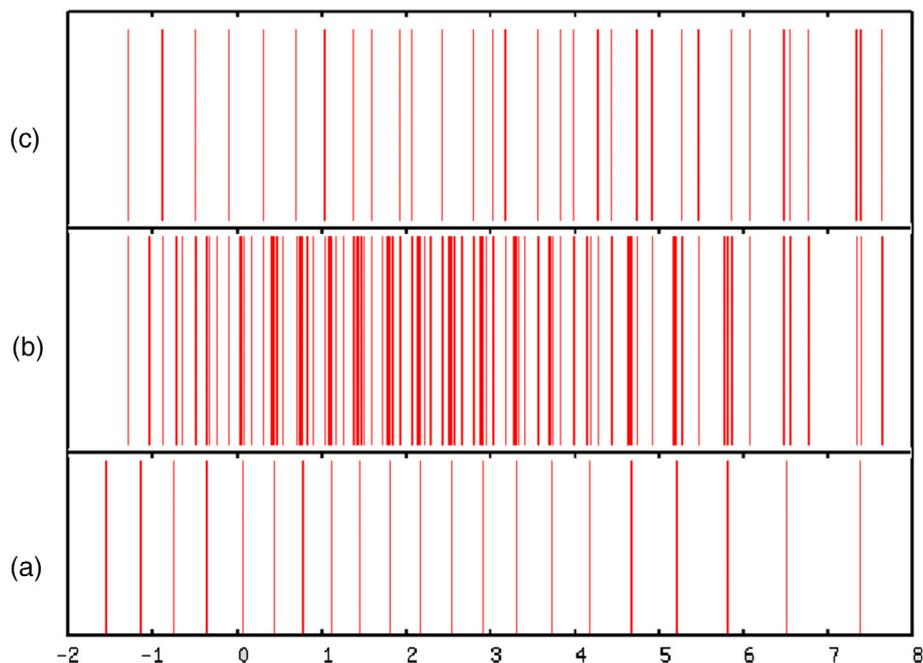


FIGURE 4 Illustration of the sparsity in the primitive Gaussian basis associated to an aCD basis set, compared with that arising from the original orbital basis and the aCD basis sets, exemplified by the s shell of the ANO-RCC basis for the Tc atom. Primitive exponents in (a) the orbital basis (b) the resulting aCD basis ($\tau = 10^{-4} E_h$) and (c) the resulting acCD ($\tau = 10^{-4} E_h$) basis set. Note the use of a logarithmic x axis. (Reprinted from Reference [116], with the permission of AIP Publishing.)

within acceptable error bars. They stated that the acCD auxiliary basis sets could possibly play a role in calculations with larger basis sets.

Over the last 15 years a number of new auxiliary basis sets have been developed and published. Hill et al. have developed RI-MP2 auxiliary basis sets for correlation-consistent basis sets for elements beginning from scandium,^{119–121} while additional auxiliary basis sets for other types of orbital basis sets (weighted core-valence and ECP basis sets,¹²² the 6-31G** and 6-311G** basis sets of Pople et al.,¹²³ lanthanide basis sets,¹²⁴ Turbomole basis sets with diffuse functions,¹²⁵ etc.) have also been developed.

Auxiliary basis sets have also been developed for F12 calculations. Shaw and Hill¹²⁶ reported improved complementary auxiliary basis sets for reaching the HF limit (OptRI+) in F12 calculations, while Kritikou and Hill¹²⁷ reported auxiliary basis sets optimized for reproducing correlation energies computed with F12 second-order Møller–Plesset perturbation theory.

As is evident from this tale, the selection and accuracy of auxiliary basis sets has inherited the complexity and confusion of orbital basis sets. Corrections to auxiliary basis sets for the correlation consistent cc-pV(n+d)Z orbital basis sets have been published as recently as 2023 by Nash et al.¹²⁸

As the literature lacks systematic version control, it can be hard to keep track of the development of orbital basis sets and the corresponding auxiliary basis sets. An example is the recent study of Kermani and Truhlar,¹²⁹ which highlights the ill-known fact that a given orbital basis set such as cc-pVDZ may not be the same in all programs.

Over the last few years researchers have been trying to address this issue by coming up with a simple recipe for generating an accurate auxiliary basis set for the given orbital basis set. Building on the tradition of employing product basis sets in the preceding computational physics literature, Ren et al.¹³⁰ proposed a procedure that forms auxiliary basis sets by forming all possible products and choosing the auxiliary basis functions by Gram–Schmidt orthogonalization. Note that if optimal pivoting is used, according to Section 2.1 this procedure is equivalent to the use of a pivoted Cholesky decomposition. Importantly, the procedure of Ren et al. is independent of the form of the atomic basis set; their implementation used numerical atomic orbitals, which were used to express Gaussian-type orbital basis sets.

Stoychev et al.¹³¹ reported the AutoAux procedure, which is a set of 9 ad hoc rules controlled by 19 adjustable parameters to generate auxiliary basis sets for Gaussian basis sets. The AutoAux basis sets are usually twice as large as standard precomputed auxiliary basis sets but Stoychev et al. claim these to be general-purpose basis sets. Laikov¹³² similarly proposed a method based on ad hoc rules of thumb for optimizing an auxiliary basis set for a given Gaussian orbital basis, which likewise relies on a number of adjustable parameters that control the composition of the resulting auxiliary basis. Semidalas and Martin¹³³ reported an automatic ad hoc procedure to generate complementary auxiliary basis functions for explicitly correlated F12 wave calculations with Gaussian basis sets following an eight-point scheme, which yields an auxiliary basis set with similar quality to the OptRI+ basis set.

On another path, following the footsteps of Aquilante et al.,¹¹⁶ Lehtola¹¹⁴ proposed an atomic CD procedure that employs a pivoted Cholesky decomposition to pick out auxiliary atom-centered basis functions of the standard form composed of a radial function times a spherical harmonic. At variance to the aCD and acCD procedures that employ combinations of Cartesian functions and spherical harmonics in the auxiliary basis that are not supported by most quantum chemistry programs or basis set formats, the procedure of Lehtola produces RI/DF basis sets of the standard form that can be employed without changes in any program that implements RI/DF calculations while simultaneously maintaining the central property of atomic CD of being controlled by a single accuracy threshold. Lehtola's procedure, based on a pivoted Cholesky decomposition, is similar to that of Ren et al.¹³⁰; see discussion above. Moreover, both schemes are similarly applicable to any type of atomic orbital basis set. However, Ren et al. employ the full product orbital basis, while Lehtola preselects the trial products by a pivoted Cholesky decomposition of the ERIs following Aquilante et al. Lehtola compared the new procedure to the AutoAux method of Stoychev et al.¹³¹ and found AutoAux to result in RI/DF errors up to 5 meV in MP2 total energies while the new Cholesky method afforded errors an order of magnitude smaller. However, the generated Cholesky basis sets were found to be significantly larger than those generated by AutoAux.

In extremely recent work Lehtola¹³⁴ discusses a method of contracting the autogenerated Cholesky basis set by the use of a singular value decomposition (SVD) of the three-center integrals, following a previous suggestion of Kállay.¹³⁵ Combining the contraction with a pruning of the high-angular momentum functions, Lehtola¹³⁴ finds that the resulting procedure allows the size of the auxiliary basis to be reduced significantly: HF and MP2 total and atomization energies can be reproduced with some 50% fewer auxiliary basis functions with the new scheme. Lehtola finds that the accuracy of the full auxiliary basis can be captured with $N_{\text{aux}} \approx 5.5N_{\text{OBF}}$, while an accuracy similar to AutoAux is achievable with $N_{\text{aux}} \approx 3.5N_{\text{OBF}}$.

Not minding the complication with the mixing of Cartesians and spherical functions in the aCD basis, Hellmann et al.¹³⁶ applied aCD with only spherical functions to integrals of the long-range part of the range-separated Coulomb operator in time-dependent density functional theory (TD-DFT). Hellmann et al.¹³⁶ observed that the aCD basis set was *consistently smaller* than the recommended general-purpose Coulomb RI basis set by up to one order of magnitude for some molecular systems, without any loss of accuracy to excitation energies. Single-point energies were, however, found to be prone to the errors introduced by the lacking handling of the mixed Cartesian/spherical auxiliary functions used in the aCD scheme, in agreement to the findings of Lehtola.¹¹⁴ However, it is interesting to compare the findings

of Hellmann et al. to the recent work of Zhou et al.,¹³⁷ who found that TDDFT excitation energies of organic molecules can be captured to 0.06 eV error with just a single(!) s-type auxiliary basis function per atom. This suggests that TDDFT calculations are surprisingly insensitive to the accuracy of the auxiliary basis set.

Before concluding this section a word of caution is in order though. Auxiliary basis sets autogenerated from small orbital basis sets will lack high-angular-momentum functions which are necessary to describe products arising from functions on two atoms. As a rule of thumb, the one-center approximations involved in LCAD, 1C-CD and the automated methods for generating auxiliary basis sets discussed above require at least a polarized triple- ζ orbital basis set to afford auxiliary basis sets that are flexible enough to describe two-center orbital products that otherwise lead to large RI/DF errors. It is also known that the high-angular-momentum functions contained in the orbital product basis generated by first-principles automatic algorithms are important for controlling the error in *local* fitting and, therefore, should not be eliminated.¹³⁸ Local fitting is often utilized to accelerate reduced-scaling or linear-scaling algorithms, and accuracy can be improved with the use of robust and variational fitting which corrects the first-order error made in the fit.^{139,140} Robust and variational fitting replaces Equation (25) with

$$\begin{aligned}
 (\mu\nu|\kappa\lambda) \approx & \sum_L (\mu\nu|L) C_{\kappa\lambda}^L + \sum_K C_{\mu\nu}^K (K|\kappa\lambda) \\
 & - \sum_{KL} C_{\mu\nu}^K (K|L) C_{\kappa\lambda}^L.
 \end{aligned}
 \tag{28}$$

It is easy to see that this approximation is quadratic in the fitting error,¹⁴¹ and that it thereby allows flexibility for choosing the fitting coefficients \mathbf{C} , including using other metrics than the Coulomb operator. Equation (28) reduces to the traditional expression of Equation (26) when the coefficients are determined by Coulomb fitting in the full auxiliary basis with Equation (24).

Robust local fitting, however, breaks the positive-definiteness of the ERI tensor and, hence, may lead to *attractive* electron–electron interactions. Moreover, the Hamiltonian may become unbounded from below, leading to variational collapse. This problem was first described by Merlot et al.,¹⁴² and was later discussed in detail by Wirz et al.¹⁴³ to which we refer for the full analysis. In short, Wirz et al. showed that with robust local fitting, Equation (28), the Coulomb term is unbounded from below but the exchange term remains bounded. Conversely, with non-robust local fitting, Equation (26), the Coulomb term is bounded from below, while the exchange term is unbounded. Using either robust or non-robust fitting for both terms may thus lead to variational collapse. On the other hand, as first noted by Manzer et al.,¹⁴⁴ the problem of variational collapse can be circumvented rather easily by using robust fitting for the exchange term only. The Coulomb term can be determined by non-robust fitting with all auxiliary basis functions; this is not an issue as highly efficient algorithms exist for computing the Coulomb matrix.^{145–147}

4 | CONCLUSION

We have reviewed the use of the Cholesky decomposition (CD) in electronic structure theory, focusing on the CD of the electron repulsion integrals that was first discussed by Beebe and Linderberg.¹ Our review was divided into two parts. The first part (Section 2) presented the general CD algorithm, and then proved the connection to Gram–Schmidt orthogonalization that can be used to unite the CD and RI/DF approaches as well as to formulate the efficient two-step CD approach in Section 2.1. We reviewed the development of CD approaches in the literature in Section 2.2: the CD technique enables computationally efficient approaches with reduced memory demands for a large variety of ab initio methods, and we presented several cases in which this technology offers reductions in computational time by at least an order of magnitude. The second part of this work (Section 3) discussed the RI/DF approach, and continued with the review of auxiliary basis set approaches in Section 3.1. The tight relationship between the CD procedure and RI/DF technology has been employed in the literature for automatically generating auxiliary basis sets for RI/DF methods with the CD technique, offering a robust approach that can easily be applied on the fly to any orbital basis set. Indeed, atomic CD procedures are the only way to derive compact auxiliary basis sets with strict error control. It is our view that the developments of the last 20 years demonstrate empirically that auxiliary basis sets which are consistently accurate, robust and of general purpose are best derived in an atomic Cholesky procedure. Moreover, the tight relationship between CD and RI/DF is nowadays obvious also in the literature, as CD and RI/DF are handled in many programs within a single implementation of these two efficient inner projection methods.

It is an interesting coincidence that the previous review⁶ was published in 2011 close to 100 years after the invention of the CD by A.-L. Cholesky in 1910. Similarly, this work also marks almost 100 years since the original publication of the CD method² in 1924.

We conclude this review with another interesting fact. The 2011 review by Aquilante et al.⁶ concluded by noting that the original work of Beebe and Linderberg¹ did not gain many citations in the first 30 years after its publication, and that most of the citations it had accrued by the end of 2010 were from articles published in the preceding 3 years (2007–2010); we note here that the article of Beebe and Linderberg had been cited a total of 97 times by the end of 2010. Aquilante et al. opined that the CD procedure had much more to offer for quantum chemistry in the future. As documented in this article, the CD procedure has been central to the development of efficient and robust computational strategies by many research groups. Today, in September 2023, the paper by Beebe and Linderberg has been cited a total of 398 times according to the Web of Science. We expect this strong growth of the use of CD techniques in quantum chemistry to continue in the upcoming decades.

AUTHOR CONTRIBUTIONS

Thomas Bondo Pedersen: Visualization (equal); writing – review and editing (equal). **Ignacio Fdez. Galván:** Visualization (lead); writing – review and editing (equal). **Roland Lindh:** Writing – original draft (lead); writing – review and editing (equal). **Susi Lehtola:** Writing – review and editing (equal).

ACKNOWLEDGMENTS

TBP acknowledges support from the Research Council of Norway through its Centres of Excellence scheme, project number 262695. SL acknowledges support from the Academy of Finland under project numbers 350282 and 353749. IFG and RL acknowledge the Swedish Research Council (VR, Grant No. 2020-03182) for funding.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Thomas Bondo Pedersen  <https://orcid.org/0000-0001-8967-6055>

Susi Lehtola  <https://orcid.org/0000-0001-6296-8103>

Ignacio Fdez. Galván  <https://orcid.org/0000-0002-0684-7689>

Roland Lindh  <https://orcid.org/0000-0001-7567-8295>

RELATED WIREs ARTICLES

[Challenges in large scale quantum mechanical calculations](#)

REFERENCES

1. Beebe NHF, Linderberg J. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int J Quantum Chem.* 1977;12:683–705. <https://doi.org/10.1002/qua.560120408>
2. Comdt. Benoît. Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés à un système d'équations linéaires en nombre inférieur à celui des inconnues. Application de la méthode à la résolution d'un système défini d'équations linéaires. *Bull Géod.* 1924;2:67–77. <https://doi.org/10.1007/bf03031308>
3. Golub GH, van Loan CF. *Matrix computations.* Baltimore: Johns Hopkins University Press; 1996.
4. Aquilante F, Lindh R, Pedersen TB. Analytic derivatives for the Cholesky representation of the two-electron integrals. *J Chem Phys.* 2008;129:034106. <https://doi.org/10.1063/1.2955755>
5. Pedersen TB, Aquilante F, Lindh R. Density fitting with auxiliary basis sets from Cholesky decompositions. *Theor Chem Acc.* 2009;124: 1–10. <https://doi.org/10.1007/s00214-009-0608-y>
6. Aquilante F, Boman L, Boström J, Koch H, Lindh R, Sánchez de Merás A, et al. Cholesky decomposition techniques in electronic structure theory. In: Leszczynski J, editor. *Challenges and advances in computational chemistry and physics.* Netherlands: Springer; 2011. p. 301–43. https://doi.org/10.1007/978-90-481-2853-2_13
7. Aquilante F, Delcey MG, Pedersen TB, Galván I, Lindh R. Inner projection techniques for the low-cost handling of two-electron integrals in quantum chemistry. *Mol Phys.* 2017;115:2052–64. <https://doi.org/10.1080/00268976.2017.1284354>

8. Harbrecht H, Peters M, Schneider R. On the low-rank approximation by the pivoted Cholesky decomposition. *Appl Numer Math*. 2012; 62:428–40. <https://doi.org/10.1016/j.apnum.2011.10.001>
9. Roothaan CCJ. New developments in molecular orbital theory. *Rev Mod Phys*. 1951;23:69–89. <https://doi.org/10.1103/revmodphys.23.69>
10. Slater JC. *Quantum theory of atomic structure*. New York: McGraw-Hill; 1960.
11. Whitten JL. Coulombic potential energy integrals and approximations. *J Chem Phys*. 1973;58:4496–501. <https://doi.org/10.1063/1.1679012>
12. Power JD, Pitzer RM. Inequalities for electron repulsion integrals. *Chem Phys Lett*. 1974;24:478–83. [https://doi.org/10.1016/0009-2614\(74\)80159-4](https://doi.org/10.1016/0009-2614(74)80159-4)
13. Almlöf J, Faegri K, Korsell K. Principles for a direct SCF approach to LCAOMO *ab-initio* calculations. *J Comput Chem*. 1982;3:385–99. <https://doi.org/10.1002/jcc.540030314>
14. O'Neal DW, Simons J. Application of Cholesky-like matrix decomposition methods to the evaluation of atomic orbital integrals and integral derivatives. *Int J Quantum Chem*. 1989;36:673–88. <https://doi.org/10.1002/qua.560360602>
15. Røeggen I, Wisløff-Nilssen E. On the Beebe–Linderberg two-electron integral approximation. *Chem Phys Lett*. 1986;132:154–60. [https://doi.org/10.1016/0009-2614\(86\)80099-9](https://doi.org/10.1016/0009-2614(86)80099-9)
16. Boys SF, Shavitt I. *A fundamental calculation of the energy surface for the system of three hydrogen atoms*; Technical Report WIS-AF-13. Springfield, VA: University of Wisconsin; 1959.
17. Weigend F, Kattannek M, Ahlrichs R. Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods. *J Chem Phys*. 2009;130:164106. <https://doi.org/10.1063/1.3116103>
18. Koch H, Sánchez de Merás A, Pedersen TB. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J Chem Phys*. 2003;118:9481–4. <https://doi.org/10.1063/1.1578621>
19. Aidas K, Angeli C, Bak KL, Bakken V, Bast R, Boman L, et al. The Dalton quantum chemistry program system. *Wiley Interdiscip Rev: Comput Mol Sci*. 2014;4:269–84. <https://doi.org/10.1002/wcms.1172>
20. Pedersen TB, Sánchez de Merás AMJ, Koch H. Polarizability and optical rotation calculated from the approximate coupled cluster singles and doubles CC2 linear response theory using Cholesky decompositions. *J Chem Phys*. 2004;120:8887–97. <https://doi.org/10.1063/1.1705575>
21. Aquilante F, Pedersen TB, Lindh R. Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. *J Chem Phys*. 2007;126:194106. <https://doi.org/10.1063/1.2736701>
22. Aquilante F, de Vico L, Ferré N, Ghigo G, Malmqvist P-Å, Neogrady P, et al. MOLCAS 7: the next generation. *J Comput Chem*. 2010; 31:224–47. <https://doi.org/10.1002/jcc.21318>
23. Aquilante F, Pedersen TB. Quartic scaling evaluation of canonical scaled opposite spin second-order Møller–Plesset correlation energy using Cholesky decompositions. *Chem Phys Lett*. 2007;449:354–7. <https://doi.org/10.1016/j.cplett.2007.10.087>
24. Aquilante F, Pedersen TB, Lindh R, Roos BO, Sánchez de Merás A, Koch H. Accurate *ab initio* density fitting for multiconfigurational self-consistent field methods. *J Chem Phys*. 2008;129:024113. <https://doi.org/10.1063/1.2953696>
25. Aquilante F, Malmqvist P-Å, Pedersen TB, Ghosh A, Roos BO. Cholesky decomposition-based multiconfiguration second-order perturbation theory (CD-CASPT2): application to the spin-state energetics of CoIII(diiminato)(NPh). *J Chem Theory Comput*. 2008;4:694–702. <https://doi.org/10.1021/ct700263h>
26. Røeggen I, Johansen T. Cholesky decomposition of the two-electron integral matrix in electronic structure calculations. *J Chem Phys*. 2008;128:194107. <https://doi.org/10.1063/1.2925269>
27. Aquilante F, Pedersen TB, Sánchez de Merás A, Koch H. Fast noniterative orbital localization for large molecules. *J Chem Phys*. 2006; 125:174101. <https://doi.org/10.1063/1.2360264>
28. Boman L, Koch H, Sánchez de Merás A. Method specific Cholesky decomposition: coulomb and exchange energies. *J Chem Phys*. 2008;129:134107. <https://doi.org/10.1063/1.2988315>
29. Lehtola S. Curing basis set overcompleteness with pivoted Cholesky decompositions. *J Chem Phys*. 2019;151:241102. <https://doi.org/10.1063/1.5139948>
30. Lehtola S. Polarized Gaussian basis sets from one-electron ions. *J Chem Phys*. 2020;152:134108. <https://doi.org/10.1063/1.5144964>
31. Lehtola S. Accurate reproduction of strongly repulsive interatomic potentials. *Phys Rev A*. 2020;101:032504. <https://doi.org/10.1103/physreva.101.032504>
32. Vysotskiy VP, Cederbaum LS. On the Cholesky decomposition for electron propagator methods: general aspects and application on C60. *J Chem Phys*. 2010;132:044110. <https://doi.org/10.1063/1.3297890>
33. Okiyama Y, Nakano T, Yamashita K, Mochizuki Y, Taguchi N, Tanaka S. Acceleration of fragment molecular orbital calculations with Cholesky decomposition approach. *Chem Phys Lett*. 2010;490:84–9. <https://doi.org/10.1016/j.cplett.2010.03.001>
34. Kumar D, Dutta AK, Manohar PU. Resolution of the identity and Cholesky representation of EOM-MP2 approximation: implementation, accuracy and efficiency. *J Chem Sci*. 2017;129:1611–26. <https://doi.org/10.1007/s12039-017-1378-z>
35. Hohenstein EG, Sherrill CD. Density fitting and Cholesky decomposition approximations in symmetryadapted perturbation theory: implementation and application to probe the nature of π – π interactions in linear acenes. *J Chem Phys*. 2010;132:184111. <https://doi.org/10.1063/1.3426316>
36. Delcey MG, Freitag L, Pedersen TB, Aquilante F, Lindh R, González L. Analytical gradients of complete active space self-consistent field energies using Cholesky decomposition: geometry optimization and spin-state energetics of a ruthenium nitrosyl complex. *J Chem Phys*. 2014;140:174103. <https://doi.org/10.1063/1.4873349>

37. Gyórfy W, Shiozaki T, Knizia G, Werner H-J. Analytical energy gradients for second-order multireference perturbation theory using density fitting. *J Chem Phys.* 2013;138:104104. <https://doi.org/10.1063/1.4793737>
38. Vlaisavljevich B, Shiozaki T. Nuclear energy gradients for internally contracted complete active space second-order perturbation theory: multistate extensions. *J Chem Theory Comput.* 2016;12:3781–7. <https://doi.org/10.1021/acs.jctc.6b00572>
39. Nishimoto Y. Analytic gradients for restricted active space second-order perturbation theory (RASPT2). *J Chem Phys.* 2021;154:194103. <https://doi.org/10.1063/5.0050074>
40. Nishimoto Y, Battaglia S, Lindh R. Analytic first-order derivatives of (X)MS, XDW, and RMS variants of the CASPT2 and RASPT2 methods. *J Chem Theory Comput.* 2022;18:4269–81. <https://doi.org/10.1021/acs.jctc.2c00301>
41. Chwee TS, Carter EA. Cholesky decomposition within local multireference singles and doubles configuration interaction. *J Chem Phys.* 2010;132:074104. <https://doi.org/10.1063/1.3315419>
42. Krisloff DB, Krauter CM, Ricci FJ, Carter EA. Density fitting and Cholesky decomposition of the two-electron integrals in local multireference configuration interaction theory. *J Chem Theory Comput.* 2015;11:5242–51. <https://doi.org/10.1021/acs.jctc.5b00762>
43. Lehtola S, Head-Gordon M, Jónsson H. Complex orbitals, multiple local minima, and symmetry breaking in Perdew–Zunger self-interaction corrected density functional theory calculations. *J Chem Theory Comput.* 2016;12:3195–207. <https://doi.org/10.1021/acs.jctc.6b00347>
44. Lehtola S, Jónsson EÖ, Jónsson H. Effect of complex-valued optimal orbitals on atomization energies with the Perdew–Zunger self-interaction correction to density functional theory. *J Chem Theory Comput.* 2016;12:4296–302. <https://doi.org/10.1021/acs.jctc.6b00622>
45. Lehtola S, Parkhill J, Head-Gordon M. Orbital optimisation in the perfect pairing hierarchy: applications to full-valence calculations on linear polyacenes. *Mol Phys.* 2018;116:547–60. <https://doi.org/10.1080/00268976.2017.1342009>
46. Pitoňák M, Aquilante F, Hobza P, Neogrády P, Noga J, Urban M. Parallelized implementation of the CCSD(T) method in MOLCAS using optimized virtual orbitals space and Cholesky decomposed two-electron integrals. *Collect Czechoslov Chem Commun.* 2011;76:713–42. <https://doi.org/10.1135/cccc2011048>
47. Epifanovsky E, Zuev D, Feng X, Khistyayev K, Shao Y, Krylov AI. General implementation of the resolution-of-the-identity and Cholesky representations of electron repulsion integrals within coupled-cluster and equation-of-motion methods: theory and benchmarks. *J Chem Phys.* 2013;139:134105. <https://doi.org/10.1063/1.4820484>
48. DePrince AE, Kennedy MR, Sumpter BG, Sherrill CD. Densityfitted singles and doubles coupled cluster on graphics processing units. *Mol Phys.* 2014;112:844–52. <https://doi.org/10.1080/00268976.2013.874599>
49. Peng B, Kowalski K. Low-rank factorization of electron integral tensors and its application in electronic structure theory. *Chem Phys Lett.* 2017;672:47–53.
50. Feng X, Epifanovsky E, Gauss J, Krylov AI. Implementation of analytic gradients for CCSD and EOMCCSD using Cholesky decomposition of the electron-repulsion integrals and their derivatives: theory and benchmarks. *J Chem Phys.* 2019;151:014110. <https://doi.org/10.1063/1.5100022>
51. Lesiuk M. Implementation of the coupled-cluster method with single, double, and triple excitations using tensor decompositions. *J Chem Theory Comput.* 2019;16:453–67. <https://doi.org/10.1021/acs.jctc.9b00985>
52. Schnack-Petersen AK, Koch H, Coriani S, Kjørstad EF. Efficient implementation of molecular CCSD gradients with Choleskydecomposed electron repulsion integrals. *J Chem Phys.* 2022;156:244111. <https://doi.org/10.1063/5.0087261>
53. Bozkaya U. Orbital-optimized second-order perturbation theory with density-fitting and Cholesky decomposition approximations: an efficient implementation. *J Chem Theory Comput.* 2014;10:2371–8. <https://doi.org/10.1021/ct500231c>
54. Bozkaya U. Orbital-optimized MP3 and MP2.5 with density-fitting and Cholesky decomposition approximations. *J Chem Theory Comput.* 2016;12:1179–88. <https://doi.org/10.1021/acs.jctc.5b01128>
55. Bozkaya U. Orbital-optimized linearized coupled-cluster doubles with density-fitting and Cholesky decomposition approximations: an efficient implementation. *Phys Chem Chem Phys.* 2016;18:11362–73. <https://doi.org/10.1039/c6cp00164e>
56. Bozkaya U. Efficient implementation of the second-order quasidegenerate perturbation theory with density-fitting and Cholesky decomposition approximations: is it possible to use Hartree–Fock orbitals for a multiconfigurational perturbation theory? *J Chem Theory Comput.* 2019;15:4415–29. <https://doi.org/10.1021/acs.jctc.9b00378>
57. Bozkaya U, Ünal A, Alagöz Y. Energy and analytic gradients for the orbital-optimized coupled-cluster doubles method with the density-fitting approximation: an efficient implementation. *J Chem Phys.* 2020;153:244115. <https://doi.org/10.1063/5.0035811>
58. Ünal A, Bozkaya U. Efficient implementation of equation-of-motion coupled-cluster singles and doubles method with the density-fitting approximation: an enhanced algorithm for the particle–particle ladder term. *J Chem Theory Comput.* 2022;18:1489–500. <https://doi.org/10.1021/acs.jctc.1c01000>
59. Smith DGA, Burns LA, Simmonett AC, Parrish RM, Schieber MC, Galvelis R, et al. Psi4 1.4: open-source software for highthroughput quantum chemistry. *J Chem Phys.* 2020;152:184108. <https://doi.org/10.1063/5.0006002>
60. Hohenstein EG, Martínez TJ. GPU acceleration of rank-reduced coupled-cluster singles and doubles. *J Chem Phys.* 2021;155:184110. <https://doi.org/10.1063/5.0063467>
61. Hohenstein EG, Fales BS, Parrish RM, Martínez TJ. Rankreduced coupled-cluster. III. Tensor hypercontraction of the doubles amplitudes. *J Chem Phys.* 2022;156:054102. <https://doi.org/10.1063/5.0077770>
62. Fosso-Tande J, Nguyen T-S, Gidofalvi G, DePrince AE. Large-scale variational two-electron reduced-density-matrix-driven complete active space self-consistent field methods. *J Chem Theory Comput.* 2016;12:2260–71. <https://doi.org/10.1021/acs.jctc.6b00190>
63. Gong X, Chen Z, Wu W. The application of Cholesky decomposition in valence bond calculation. *J Comput Chem.* 2016;37:2157–62. <https://doi.org/10.1002/jcc.24442>

64. Freitag L, Knecht S, Angeli C, Reiher M. Multireference perturbation theory with Cholesky decomposition for the density matrix renormalization group. *J Chem Theory Comput.* 2017;13:451–9. <https://doi.org/10.1021/acs.jctc.6b00778>
65. Piccardo M, Soncini A. A full-pivoting algorithm for the Cholesky decomposition of two-electron repulsion and spin-orbit coupling integrals. *J Comput Chem.* 2017;38:2775–83. <https://doi.org/10.1002/jcc.25062>
66. Folkestad SD, Kjørnstad EF, Myhre RH, Andersen JH, Balbi A, Coriani S, et al. eT 1.0: an open source electronic structure program with emphasis on coupled cluster and multilevel methods. *J Chem Phys.* 2020;152:184103. <https://doi.org/10.1063/5.0004713>
67. Folkestad SD, Kjørnstad EF, Koch H. An efficient algorithm for Cholesky decomposition of electron repulsion integrals. *J Chem Phys.* 2019;150:194112. <https://doi.org/10.1063/1.5083802>
68. Lew-Yee JFH, Flores-Moreno R, Morales JL, del Campo JM. Asymmetric density fitting with modified Cholesky decomposition applied to second-order electron propagator. *J Chem Theory Comput.* 2020;16:1597–605. <https://doi.org/10.1021/acs.jctc.9b01215>
69. Zhang T, Liu X, Valeev EF, Li X. Toward the minimal floating operation count Cholesky decomposition of electron repulsion integrals. *J Phys Chem A.* 2021;125:4258–65. <https://doi.org/10.1021/acs.jpca.1c02317>
70. Blaschke S, Stopkowicz S. Cholesky decomposition of complex two-electron integrals over GIAOs: efficient MP2 computations for large molecules in strong magnetic fields. *J Chem Phys.* 2022;156:044115. <https://doi.org/10.1063/5.0076588>
71. Galván I, Vacher M, Alavi A, Angeli C, Aquilante F, Autschbach J, et al. OpenMolcas: from source code to insight. *J Chem Theory Comput.* 2019;15:5925–64. <https://doi.org/10.1021/acs.jctc.9b00532>
72. Helmich-Paris B, Repisky M, Visscher L. Relativistic Cholesky decomposed density matrix MP2. *Chem Phys.* 2019;518:38–46. <https://doi.org/10.1016/j.chemphys.2018.11.009>
73. Bast R, Gomes ASP, Saue T, Visscher L, Jensen HJA, Aucar IA, et al. DIRAC, a relativistic ab initio electronic structure program, Release DIRAC23. 2023 <https://doi.org/10.5281/zenodo.7670749> <https://www.diracprogram.org>
74. Nottoli T, Gauss J, Lipparini F. A black-box, general purpose quadratic self-consistent field code with and without Cholesky decomposition of the two-electron integrals. *Mol Phys.* 2021;119:e1974590. <https://doi.org/10.1080/00268976.2021.1974590>
75. Matthews DA, Cheng L, Harding ME, Lipparini F, Stopkowicz S, Jagau T-C, et al. Coupledcluster techniques for computational chemistry: the CFOUR program package. *J Chem Phys.* 2020;152:214108. <https://doi.org/10.1063/5.0004837>
76. London F. Théorie quantique des courants interatomiques dans les combinaisons aromatiques. *J Phys Radium.* 1937;8:397–409. <https://doi.org/10.1051/jphysrad:01937008010039700>
77. Nottoli T, Gauss J, Lipparini F. Second-order CASSCF algorithm with the Cholesky decomposition of the two-electron integrals. *J Chem Theory Comput.* 2021;17:6819–31. <https://doi.org/10.1021/acs.jctc.1c00327>
78. Gauss J, Blaschke S, Burger S, Nottoli T, Lipparini F, Stopkowicz S. Cholesky decomposition of two-electron integrals in quantum-chemical calculations with perturbative or finite magnetic fields using gauge-including atomic orbitals. *Mol Phys.* 2022;121:e2101562. <https://doi.org/10.1080/00268976.2022.2101562>
79. Nottoli T, Burger S, Stopkowicz S, Gauss J, Lipparini F. Computation of NMR shieldings at the CASSCF level using gauge-including atomic orbitals and Cholesky decomposition. *J Chem Phys.* 2022;157:084122. <https://doi.org/10.1063/5.0101838>
80. Takashima C, Nakai H. Matrixdecomposed two-electron integrals in the infinite-order twocomponent Hamiltonian. *Chem Phys Lett.* 2023;828:140714. <https://doi.org/10.1016/j.cplett.2023.140714>
81. Billingsley FP, Bloor JE. Limited expansion of diatomic overlap (LEDO): a near-accurate approximate *ab initio* LCAO MO method. I. Theory and preliminary investigations. *J Chem Phys.* 1971;55:5178–90. <https://doi.org/10.1063/1.1675655>
82. Baerends EJ, Ellis DE, Ros P. Self-consistent molecular Hartree-Fock-Slater calculations I. The computational procedure. *Chem Phys.* 1973;2:41–51. [https://doi.org/10.1016/0301-0104\(73\)80059-x](https://doi.org/10.1016/0301-0104(73)80059-x)
83. Jafri JA, Whitten JL. Electron repulsion integral approximations and error bounds: molecular applications. *J Chem Phys.* 1974;61:2116–21. <https://doi.org/10.1063/1.1682222>
84. Sambe H, Felton RH. A new computational approach to Slater's SCF- $X\alpha$ equation. *J Chem Phys.* 1975;62:1122–6. <https://doi.org/10.1063/1.430555>
85. Dunlap BI, Connolly JWD, Sabin JR. On some approximations in applications of $X\alpha$ theory. *J Chem Phys.* 1979;71:3396–402. <https://doi.org/10.1063/1.438728>
86. Dunlap BI, Connolly JWD, Sabin JR. On first-row diatomic molecules and local density models. *J Chem Phys.* 1979;71:4993–9. <https://doi.org/10.1063/1.438313>
87. van Alsenoy C. Ab initio calculations on large molecules: the multiplicative integral approximation. *J Comput Chem.* 1988;9:620–6. <https://doi.org/10.1002/jcc.540090607>
88. Feyereisen M, Fitzgerald G, Komornicki A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem Phys Lett.* 1993;208:359–63. [https://doi.org/10.1016/0009-2614\(93\)87156-w](https://doi.org/10.1016/0009-2614(93)87156-w)
89. Vahtras O, Almlöf J, Feyereisen MW. Integral approximations for LCAO-SCF calculations. *Chem Phys Lett.* 1993;213:514–8. [https://doi.org/10.1016/0009-2614\(93\)89151-7](https://doi.org/10.1016/0009-2614(93)89151-7)
90. Duchemin I, Li J, Blase X. Hybrid and constrained resolution-of-identity techniques for coulomb integrals. *J Chem Theory Comput.* 2017;13:1199–208. <https://doi.org/10.1021/acs.jctc.6b01215>
91. De Chavez D, Lindh R. One-centre corrected two-electron integrals in inner projection-based integral evaluations. *Mol Phys.* 2023; e2228431.
92. Ten-no S, Iwata S. Three-center expansion of electron repulsion integrals with linear combination of atomic electron distributions. *Chem Phys Lett.* 1995;240:578–84. [https://doi.org/10.1016/0009-2614\(95\)00564-k](https://doi.org/10.1016/0009-2614(95)00564-k)

93. Ten-no S, Iwata S. On approximating electron repulsion integrals with linear combination of atomic-electron distributions. *Int J Quantum Chem.* 1996;60:1319–24.
94. Furche F, Ahlrichs R, Hättig C, Klopper W, Sierka M, Weigend F. Turbomole. *Wiley Interdiscip Rev: Comput Mol Sci.* 2013;4:91–100. <https://doi.org/10.1002/wcms.1162>
95. Eichkorn K, Treutler O, Öhm H, Häser M, Ahlrichs R. Auxiliary basis sets to approximate Coulomb potentials. *Chem Phys Lett.* 1995; 240:283–90. [https://doi.org/10.1016/0009-2614\(95\)00621-a](https://doi.org/10.1016/0009-2614(95)00621-a)
96. Eichkorn K, Weigend F, Treutler O, Ahlrichs R. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor Chem Acc.* 1997;97:119–24. <https://doi.org/10.1007/s002140050244>
97. Weigend F, Häser M, Patzelt H, Ahlrichs R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem Phys Lett.* 1998;294:143–52. [https://doi.org/10.1016/s0009-2614\(98\)00862-8](https://doi.org/10.1016/s0009-2614(98)00862-8)
98. Palanguntikul R, Polly R, Hartke B. Global and local optimization of auxiliary basis sets for RI-MP2 calculations. *Phys Chem Chem Phys.* 2004;6:5456–62. <https://doi.org/10.1039/b412673d>
99. Hamel S, Casida ME, Salahub DR. Assessment of the quality of orbital energies in resolution-of-the-identity Hartree–Fock calculations using de-Mon auxiliary basis sets. *J Chem Phys.* 2001;114:7342–50. <https://doi.org/10.1063/1.1358865>
100. Weigend F. A fully direct RI-HF algorithm: implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys Chem Chem Phys.* 2002;4:4285–91. <https://doi.org/10.1039/b204199p>
101. Weigend F, Köhn A, Hättig C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J Chem Phys.* 2002;116:3175–83. <https://doi.org/10.1063/1.1445115>
102. Hättig C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core–valence and quintuple- ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys Chem Chem Phys.* 2005;7:59–66. <https://doi.org/10.1039/b415208e>
103. Hellweg A, Hättig C, Höfener S, Klopper W. Optimized accurate auxiliary basis sets for RI-MP2 and RI-CC2 calculations for the atoms Rb to Rn. *Theor Chem Acc.* 2007;117:587–97. <https://doi.org/10.1007/s00214-007-0250-5>
104. Weigend F. Hartree–Fock exchange fitting basis sets for H to Rn. *J Comput Chem.* 2008;29:167–75. <https://doi.org/10.1002/jcc.20702>
105. Yousaf KE, Peterson KA. Optimized auxiliary basis sets for explicitly correlated methods. *J Chem Phys.* 2008;129:184108. <https://doi.org/10.1063/1.3009271>
106. Yousaf KE, Peterson KA. Optimized complementary auxiliary basis sets for explicitly correlated methods: aug-cc-pVnZ orbital basis sets. *Chem Phys Lett.* 2009;476:303–7. <https://doi.org/10.1016/j.cplett.2009.06.003>
107. Hill JG, Peterson KA. Correlation consistent basis sets for explicitly correlated wavefunctions: valence and core–valence basis sets for Li, Be, Na, and Mg. *Phys Chem Chem Phys.* 2010;12:10460–8. <https://doi.org/10.1039/c0cp00020e>
108. Yang R, Rendell AP, Frisch MJ. Automatically generated coulomb fitting basis sets: design and accuracy for systems containing H to Kr. *J Chem Phys.* 2007;127:074102. <https://doi.org/10.1063/1.2752807>
109. Weigend F. Accurate coulomb-fitting basis sets for H to Rn. *Phys Chem Chem Phys.* 2006;8:1057–65. <https://doi.org/10.1039/b515623h>
110. Löwdin P-O. Studies in perturbation theory. XI. Lower bounds to energy eigenvalues, ground state, and excited states. *J Chem Phys.* 1965;43:S175–85. <https://doi.org/10.1063/1.1701483>
111. Löwdin P-O. Some properties of inner projections. *Int J Quantum Chem.* 1970;5:231–7. <https://doi.org/10.1002/qua.560050725>
112. Löwdin P-O. Quantum theory of cohesive properties of solids. *Adv Phys.* 1956;5:1–171. <https://doi.org/10.1080/00018735600101155>
113. Löwdin P-O. On the nonorthogonality problem. In: Jenkins S, Kirk SR, Maruani J, Brändas EJ, editors. *Advances in quantum chemistry.* Volume 5. New York: Elsevier; 1970. p. 185–99. [https://doi.org/10.1016/s0065-3276\(08\)60339-1](https://doi.org/10.1016/s0065-3276(08)60339-1)
114. Lehtola S. Straightforward and accurate automatic auxiliary basis set generation for molecular calculations with atomic orbital basis sets. *J Chem Theory Comput.* 2021;17:6886–900. <https://doi.org/10.1021/acs.jctc.1c00607>
115. Aquilante F, Lindh R, Pedersen TB. Unbiased auxiliary basis sets for accurate two-electron integral approximations. *J Chem Phys.* 2007;127:114107. <https://doi.org/10.1063/1.2777146>
116. Aquilante F, Gagliardi L, Pedersen TB, Lindh R. Atomic Cholesky decompositions: a route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency. *J Chem Phys.* 2009;130:154107. <https://doi.org/10.1063/1.3116784>
117. Boström J, Delcey MG, Aquilante F, Serrano-Andrés L, Pedersen TB, Lindh R. Calibration of Cholesky auxiliary basis sets for multi-configurational perturbation theory calculations of excitation energies. *J Chem Theory Comput.* 2010;6:747–54. <https://doi.org/10.1021/ct900612k>
118. Boström J, Pitoňák M, Aquilante F, Neogrády P, Pedersen TB, Lindh R. Coupled cluster and Møller–Plesset perturbation theory calculations of noncovalent intermolecular interactions using density fitting with auxiliary basis sets from Cholesky decompositions. *J Chem Theory Comput.* 2012;8:1921–8. <https://doi.org/10.1021/ct3003018>
119. Hill JG, Platts JA. Auxiliary basis sets for density fitting–MP2 calculations: nonrelativistic triple- ζ all-electron correlation consistent basis sets for the 3d elements Sc–Zn. *J Chem Phys.* 2008;128:044104. <https://doi.org/10.1063/1.2826348>
120. Hill JG. Auxiliary basis sets for density fitting second-order Møller–Plesset perturbation theory: correlation consistent basis sets for the 5d elements Hf–Pt. *J Chem Phys.* 2011;135:044105. <https://doi.org/10.1063/1.3615062>
121. Hill JG. Auxiliary basis sets for density-fitting second-order Møller–Plesset perturbation theory: weighted core-valence correlation consistent basis sets for the 4d elements Y–Pd. *J Comput Chem.* 2013;34:2168–77. <https://doi.org/10.1002/jcc.23372>
122. Hättig C, Schmitz G, Koßmann J. Auxiliary basis sets for density-fitted correlated wavefunction calculations: weighted core-valence and ECP basis sets for post-d elements. *Phys Chem Chem Phys.* 2012;14:6549–55. <https://doi.org/10.1039/c2cp40400a>
123. Tanaka M, Katouda M, Nagase S. Optimization of RI-MP2 auxiliary basis functions for 6-31G** and 6-311G** basis sets for first-, second-, and third-row elements. *J Comput Chem.* 2013;34:2568–75. <https://doi.org/10.1002/jcc.23430>

124. Chmela J, Harding ME. Optimized auxiliary basis sets for density fitted post-Hartree–Fock calculations of lanthanide containing molecules. *Mol Phys*. 2018;116:1523–38. <https://doi.org/10.1080/00268976.2018.1433336>
125. Hellweg A, Rappoport D. Development of new auxiliary basis functions of the Karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RICC calculations. *Phys Chem Chem Phys*. 2015;17:1010–7. <https://doi.org/10.1039/c4cp04286g>
126. Shaw RA, Hill JG. Approaching the Hartree–Fock limit through the complementary auxiliary basis set singles correction and auxiliary basis sets. *J Chem Theory Comput*. 2017;13:1691–8. <https://doi.org/10.1021/acs.jctc.7b00140>
127. Kritikou S, Hill JG. Auxiliary basis sets for density fitting in explicitly correlated calculations: the atoms H–Ar. *J Chem Theory Comput*. 2015;11:5269–76. <https://doi.org/10.1021/acs.jctc.5b00816>
128. Nash HW, Shaw RA, Hill JG. Correlation consistent auxiliary basis sets in density fitting Hartree–Fock: the atoms sodium through argon revisited. *J Comput Chem*. 2023;44:1119–28. <https://doi.org/10.1002/jcc.27069>
129. Kermani MM, Truhlar DG. Reproducibility of calculations on Li species with correlation-consistent basis sets. *Chem Phys Lett*. 2023;825:140575. <https://doi.org/10.1016/j.cplett.2023.140575>
130. Ren X, Rinke P, Blum V, Wieferink J, Tkatchenko A, Sanfilippo A, et al. Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and *GW* with numeric atom-centered orbital basis functions. *New J Phys*. 2012;14:053020. <https://doi.org/10.1088/1367-2630/14/5/053020>
131. Stoychev GL, Auer AA, Neese F. Automatic generation of auxiliary basis sets. *J Chem Theory Comput*. 2017;13:554–62. <https://doi.org/10.1021/acs.jctc.6b01041>
132. Laikov DN. Optimization of atomic density-fitting basis functions for molecular two-electron integral approximations. *J Chem Phys*. 2020;153:114121. <https://doi.org/10.1063/5.0014639>
133. Semidalas E, Martin JML. Automatic generation of complementary auxiliary basis sets for explicitly correlated methods. *J Comput Chem*. 2022;43:1690–700. <https://doi.org/10.1002/jcc.26970>
134. Lehtola S. Automatic generation of accurate and cost-efficient auxiliary basis sets. *J Chem Theory Comput*. 2023;19:6242–54. <https://doi.org/10.1021/acs.jctc.3c00670>
135. Kállay M. A systematic way for the cost reduction of density fitting methods. *J Chem Phys*. 2014;141:244113. <https://doi.org/10.1063/1.4905005>
136. Hellmann L, Tölle J, Niemeyer N, Neugebauer J. Automated generation of optimized auxiliary basis sets for long-range-corrected TDDFT using the Cholesky decomposition. *J Chem Theory Comput*. 2022;18:2959–74. <https://doi.org/10.1021/acs.jctc.2c00131>
137. Zhou Z, Della Sala F, Parker SM. Minimal auxiliary basis set approach for the electronic excitation spectra of organic molecules. *J Phys Chem Lett*. 2023;14:1968–76. <https://doi.org/10.1021/acs.jpcc.2c03698>
138. Ihrig AC, Wieferink J, Zhang IY, Ropo M, Ren X, Rinke P, et al. Accurate localized resolution of identity approach for linear-scaling hybrid density functionals and for many-body perturbation theory. *New J Phys*. 2015;17:093020. <https://doi.org/10.1088/1367-2630/17/9/093020>
139. Dunlap BI. Robust and variational fitting: removing the four-center integrals from center stage in quantum chemistry. *J Mol Struct (Theochem)*. 2000;529:37–40. [https://doi.org/10.1016/s0166-1280\(00\)00528-5](https://doi.org/10.1016/s0166-1280(00)00528-5)
140. Dunlap BI. Robust and variational fitting. *Phys Chem Chem Phys*. 2000;2:2113–6. <https://doi.org/10.1039/b000027m>
141. Reine S, Tellgren E, Krapp A, Kjærgaard T, Helgaker T, Jansik B, et al. Variational and robust density fitting of four-center two-electron integrals in local metrics. *J Chem Phys*. 2008;129:104101. <https://doi.org/10.1063/1.2956507>
142. Merlot P, Kjærgaard T, Helgaker T, Lindh R, Aquilante F, Reine S, et al. Attractive electron–electron interactions within robust local fitting approximations. *J Comput Chem*. 2013;34:1486–96. <https://doi.org/10.1002/jcc.23284>
143. Wirz LN, Reine SS, Pedersen TB. On resolution-of-the-identity electron repulsion integral approximations and variational stability. *J Chem Theory Comput*. 2017;13:4897–906. <https://doi.org/10.1021/acs.jctc.7b00801>
144. Manzer SF, Epifanovsky E, Head-Gordon M. Efficient implementation of the pair atomic resolution of the identity approximation for exact exchange for hybrid and range-separated density functionals. *J Chem Theory Comput*. 2015;11:518–27. <https://doi.org/10.1021/ct5008586>
145. Ahmadi GR, Almlöf J. The coulomb operator in a Gaussian product basis. *Chem Phys Lett*. 1995;246:364–70. [https://doi.org/10.1016/0009-2614\(95\)01127-4](https://doi.org/10.1016/0009-2614(95)01127-4)
146. White CA, Head-Gordon M. A J matrix engine for density functional theory calculations. *J Chem Phys*. 1996;104:2620–9. <https://doi.org/10.1063/1.470986>
147. Neese F. An improvement of the resolution of the identity approximation for the formation of the coulomb matrix. *J Comput Chem*. 2003;24:1740–7. <https://doi.org/10.1002/jcc.10318>

How to cite this article: Pedersen TB, Lehtola S, Fdez. Galván I, Lindh R. The versatility of the Cholesky decomposition in electronic structure theory. *WIREs Comput Mol Sci*. 2024;14(1):e1692. <https://doi.org/10.1002/wcms.1692>