The Journal of Chemical Physics

An efficient algorithm for Cholesky decomposition of electron repulsion integrals

Cite as: J. Chem. Phys. **150**, 194112 (2019); https://doi.org/10.1063/1.5083802
Submitted: 30 November 2018 • Accepted: 30 April 2019 • Published Online: 20 May 2019

Sarai D. Folkestad, Eirik F. Kjønstad and Henrik Koch

View Online Export Citation CrossMark

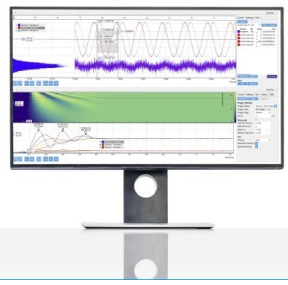
ARTICLES YOU MAY BE INTERESTED IN

Second-order MCSCF optimization revisited. I. Improved algorithms for fast and robust second-order CASSCF convergence
The Journal of Chemical Physics **150**, 194106 (2019); https://doi.org/10.1063/1.5094644

Implementation of analytic gradients for CCSD and EOM-CCSD using Cholesky decomposition of the electron-repulsion integrals and their derivatives: Theory and benchmarks
The Journal of Chemical Physics **151**, 014110 (2019); https://doi.org/10.1063/1.5100022

Adventures in DFT by a wavefunction theorist
The Journal of Chemical Physics **151**, 160901 (2019); https://doi.org/10.1063/1.5116338

Challenge us.
What are your needs for periodic signal detection?
Watch
Zurich Instruments



© 2019 Author(s).

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

An efficient algorithm for Cholesky decomposition of electron repulsion integrals

Cite as: J. Chem. Phys. 150, 194112 (2019); doi: 10.1063/1.5083802
Submitted: 30 November 2018 • Accepted: 30 April 2019 •
Published Online: 20 May 2019

View Online Export Citation CrossMark

Sarai D. Folkestad,[1,2,a)] Eirik F. Kjønstad,[1,2,a)] and Henrik Koch[1,2,b)]

AFFILIATIONS

[1] Department of Chemistry, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[2] Scuola Normale Superiore, Piazza dei Cavaleri 7, 56126 Pisa, Italy

[a)] **Contributions:** S. D. Folkestad and E. F. Kjønstad contributed equally to this work.
[b)] **Electronic mail:** henrik.koch@sns.it

ABSTRACT

Approximating the electron repulsion integrals using inner projections is a well-established approach to reduce the computational demands of electronic structure calculations. Here, we present a two-step Cholesky decomposition algorithm where only the elements of the Cholesky basis (the pivots) are determined in the pivoting procedure. This allows for improved screening, significantly reducing memory usage and computational cost. After the pivots have been determined, the Cholesky vectors are constructed using the inner projection formulation. We also propose a partitioned decomposition approach where the Cholesky basis is chosen from a reduced set generated by decomposing diagonal blocks of the matrix. The algorithm extends the application range of the methodology and is well suited for multilevel methods. We apply the algorithm to systems with up to 80 000 atomic orbitals. The accuracy of the integral approximations is demonstrated for a formaldehyde-water system using a new Cholesky-based CCSD implementation.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5083802

I. INTRODUCTION

The Beebe-Linderberg[1] algorithm for the Cholesky decomposition (CD) of the electron repulsion integral matrix was developed in the 1970s. Beebe and Linderberg observed that, given the rank-deficiency of the matrix, significant computational savings are obtainable through decomposition. Furthermore, they identified the approach as an inner projection in the sense introduced by Löwdin.[2,3] The algorithm was later modified to include screening by Røeggen and Wisløff-Nilssen,[4] who also demonstrated that the numerical rank is proportional to the number of atomic orbitals (AOs), as had already been suggested.[1] An algorithm suited for large-scale applications was first proposed in 2003 by Koch, Sánchez de Merás, and Pedersen.[5] This algorithm was implemented in DALTON[6] and subsequently included in the Molcas program.[7,8] A number of applications based on the Cholesky decomposition of the integrals have since been published.[9–12]

The resolution of identity (RI) method, also called density fitting,[13] is an inner projection technique introduced by Vahtras, Almlöf, and Feyereisen.[14] In RI, the inner projection is onto the space spanned by an auxiliary basis. The use of prefitted auxiliary basis sets in this projection has gained much popularity.[15,16] Cholesky decomposition and RI are equivalent if the auxiliary basis is chosen to be the Cholesky basis. With a Cholesky basis, the error introduced by the inner projection is rigorously controlled by the decomposition threshold. This property is not shared by prefitted basis sets.

One advantage of preoptimized auxiliary bases is that they are usually one-centered, making the integrals at most three-centered and, therefore, computationally cheaper. A Cholesky basis, on the other hand, typically includes many two-center functions. Pedersen and co-workers have advocated the atomic Cholesky decomposition (aCD) and one-center Cholesky decomposition (1C-CD) methods, where the Cholesky basis is restricted to one-center

Published under license by AIP Publishing

functions.[17] These methods necessarily imply a certain loss of accuracy. Nevertheless, the auxiliary basis sets of aCD and 1C-CD are, unlike prefitted bases, not biased toward any method or specific quantity while retaining the advantages of preoptimized RI.[17,18]

Alternatively, the computational cost of a Cholesky decomposition may be reduced by controlling the error in method specific quantities, such as the Coulomb or exchange energies, rather than the electron repulsion integrals. This type of method specific Cholesky decomposition has been shown to substantially reduce the size of the auxiliary basis with no added loss of accuracy in the target quantities.[19] The approach is well suited for multilevel methods, where only subsets of integrals are needed in the correlated treatments.[20–22]

To be generally applicable, an integral approximation scheme must have analytic geometric derivatives. Such derivatives are easily derived for RI using prefitted auxiliary bases.[23] Although not apparent in the early discussion of gradients by O'neal and Simons,[24] the equivalence of RI and Cholesky decomposition implies that analogous gradient expressions exist for Cholesky decomposed integrals. Recently, this was exploited to formulate and implement analytic gradients by Aquilante, Lindh, and Pedersen.[25]

In this contribution, we propose a two-step algorithm where only the elements of the auxiliary basis are determined in the pivoting procedure (step I). Once the basis has been identified, the Cholesky vectors are constructed using the RI formulation of Cholesky decomposition (step II). As a consequence, both the columns and rows of the integral matrix may be screened in step I, giving a reduction in both memory usage and computational cost. The idea of a two-step algorithm, where the basis is determined in step I, was proposed in Ref. 12. Here, we present an efficient implementation of the first step, fully exploiting the available screening, and introduce the RI-approach to construct the Cholesky vectors in the second step.

By restricting the set of Cholesky basis elements and/or changing the screening criterion in the algorithm, a number of decomposition approaches are readily obtained. To illustrate this flexibility, we have, in addition to the regular decomposition of the electron repulsion integrals, implemented 1C-CD[17] and a Cholesky decomposition for multilevel calculations. We also present a Cholesky decomposition where, initially, the diagonal blocks of the matrix are decomposed separately in order to qualify elements for the Cholesky basis. This partitioned Cholesky decomposition (PCD) offers significant computational and memory savings.

II. THEORY

The electron repulsion integral matrix $\boldsymbol{M}$ is symmetric positive semidefinite and may therefore be Cholesky decomposed,

$$M_{\alpha\beta,\gamma\delta} = (\alpha\beta|\gamma\delta) = \sum_J L^J_{\alpha\beta} L^J_{\gamma\delta} = (\boldsymbol{L}\boldsymbol{L}^T)_{\alpha\beta,\gamma\delta}, \tag{1}$$

where α, β, ..., denote the real atomic orbitals (AOs) $\{\chi_\alpha(\boldsymbol{r})\}_\alpha$. Alternatively, $\boldsymbol{M}$ may be expressed as an inner projection,

$$M_{\alpha\beta,\gamma\delta} = \sum_{JK} (\alpha\beta|\rho_J)(\boldsymbol{S}^{-1})_{JK}(\rho_K|\gamma\delta), \tag{2}$$

where $S_{JK} = (\rho_J|\rho_K)$. The auxiliary functions $\{\rho_J(\boldsymbol{r})\}_J$ form a basis for the space spanned by $\{\chi_\gamma(\boldsymbol{r})\chi_\delta(\boldsymbol{r})\}_{\gamma\delta}$. Since $\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$, where $\boldsymbol{Q}$ is the Cholesky factor of $\boldsymbol{S}$, we may identify the Cholesky vectors as

$$L^J_{\alpha\beta} = \sum_K (\alpha\beta|\rho_K) Q^{-\mathrm{T}}_{KJ}. \tag{3}$$

That is, a Cholesky decomposition is equivalent to an RI approximation.[1,14]

The simplicity of the Cholesky decomposition approach is most easily seen from the full-pivoting algorithm, where one first selects the largest diagonal element M_{JJ} as the pivot. Then, the corresponding Cholesky vector

$$L^J_p = \frac{M_{pJ}}{\sqrt{M_{JJ}}} \tag{4}$$

is constructed. Finally, $\boldsymbol{M}$ is updated according to

$$M_{pq} \leftarrow M_{pq} - L^J_p L^J_q. \tag{5}$$

These steps are repeated until all diagonal elements of $\boldsymbol{M}$ are below a given threshold $\tau > 0$. From the Cauchy-Schwarz inequality,

$$M^2_{pq} \le M_{pp} M_{qq}, \tag{6}$$

all elements of $\boldsymbol{M}$ will then be smaller than τ in absolute value. We may thus conclude that

$$M_{pq} \approx \sum_J L^J_p L^J_q, \tag{7}$$

where the error in M_{pq} is less than τ. However, the full-pivoting algorithm is not well suited to decompose the integral matrix. As the electron repulsion integrals are computed in batches of shell quartets, and as the matrix is generally to large to store, specialized algorithms are necessary.

We propose an algorithm where only the pivot indices $\mathcal{B} = \{J\}_J$ are determined in the pivoting procedure. As contributions from new vectors are subtracted from $\boldsymbol{M}$, its diagonal elements decrease monotonously. Consequently, a diagonal $D_p = M_{pp}$ below τ will never be selected as a pivot element. Since we only determine the pivots, we may screen out elements M_{pq} for which at least one of the corresponding diagonals, D_p or D_q, is below τ, thus applying the same screening on the rows and columns of $\boldsymbol{M}$. In algorithms where the Cholesky vectors are constructed during the pivoting procedure, screening on the rows of $\boldsymbol{M}$ must instead be with respect to the Cauchy-Schwarz inequality.[5]

Below we outline the procedure to determine $\mathcal{B}$ (step I):

1. Define a set $\mathcal{D}_0$ of initial diagonals from which elements in $\mathcal{B}$ are to be selected.
2. Set $\mathcal{B} = \{\}$.
3. Determine the significant diagonal indices $\mathcal{D}$. The standard criterion is that $p \in \mathcal{D}$ if $p \in \mathcal{D}_0$ and $D_p \ge \tau$. For $J \in \mathcal{B}$, only keep L^J_p for $p \in \mathcal{D}$. See Fig. 1. If there are no elements in $\mathcal{D}$, stop. Otherwise, move on to step 4.
4. Find $D_{\max} = \max_{p\in\mathcal{D}} D_p$ and determine the set of qualified diagonal indices $\mathcal{Q}$,

$$\mathcal{Q} = \{p \in \mathcal{D} : D_p \ge \sigma D_{\max}\}, \tag{8}$$

Published under license by AIP Publishing

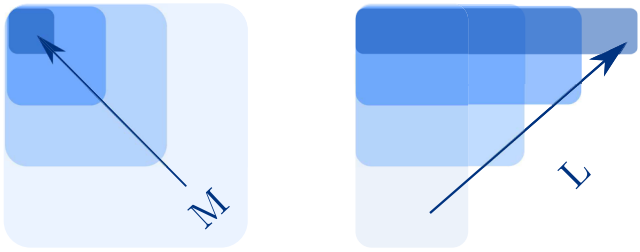


FIG. 1. While determining the basis $\mathcal{B}$, we screen out elements of $\boldsymbol{M}$ and $\boldsymbol{L}$ that are no longer needed. This means that we consider M_{pq} and L_p^J for $p, q \in \mathcal{D}$ and $J \in \mathcal{B}$. Only $\boldsymbol{L}$ is kept in memory throughout the decomposition procedure.

such that the number of elements in $\mathcal{Q}$ does not exceed a user-specified maximum. The parameter σ, which ensures that qualified diagonals are not too small, is called the span factor.

5. For each $q \in \mathcal{Q}$ compute M_{pq} for all $p \in \mathcal{D}$. If there are any previous Cholesky vectors, subtract their contributions to $\boldsymbol{M}$:

$$\tilde{M}_{pq} = M_{pq} - \sum_{J \in \mathcal{B}} L_p^J L_q^J, \quad p \in \mathcal{D}, \quad q \in \mathcal{Q}. \tag{9}$$

6. Let $\mathcal{C}$ be the set of qualified indices for which the associated Cholesky vector has been constructed. Initially, $\mathcal{C} = \{\}$. As long as there is a significant diagonal in $\mathcal{Q}$, select $q \in \mathcal{Q}$ such that $D_q = \max_{p \in \mathcal{Q}} D_p$, construct the Cholesky vector

$$L_p^q = \frac{\tilde{M}_{pq} - \sum_{J \in \mathcal{C}} L_p^J L_q^J}{\sqrt{\tilde{M}_{qq}}}, \quad p \in \mathcal{D}, \quad q \in \mathcal{Q}, \tag{10}$$

update $\mathcal{Q}$ and $\mathcal{C}$,

$$\mathcal{Q} = \mathcal{Q} \backslash \{q\}, \quad \mathcal{C} = \mathcal{C} \cup \{q\}, \tag{11}$$

and the diagonal elements,

$$D_p = D_p - (L_p^q)^2, \quad p \in \mathcal{D}. \tag{12}$$

7. Finally, update the pivots $\mathcal{B}$,

$$\mathcal{B} = \mathcal{B} \cup \mathcal{C}, \tag{13}$$

and return to step 3.

The memory needed for the Cholesky vectors reaches a maximum during the pivoting procedure and then drops off due to the reduction in the number of elements in $\mathcal{D}$; we only keep L_p^J for $p \in \mathcal{D}$. When $\mathcal{B}$ has been determined, $\mathcal{D}$ is empty, and the memory requirement has therefore dropped to zero.

Having determined $\mathcal{B}$ (step I), we need to construct the Cholesky vectors (step II). When $\boldsymbol{M}$ is the electron repulsion integral matrix, each pivot $J = \gamma\delta \in \mathcal{B}$ defines a Cholesky basis function $\rho_J(\boldsymbol{r}) = \chi_\gamma(\boldsymbol{r})\chi_\delta(\boldsymbol{r})$. The RI expression,

$$L_{\alpha\beta}^J = \sum_{K \in \mathcal{B}} (\alpha\beta|K) Q_{KJ}^{-T}, \quad J \in \mathcal{B}, \tag{14}$$

may then be used to construct the Cholesky vectors. To construct $L_{\alpha\beta}^J$, we use the Cauchy-Schwarz screening

$$(\alpha\beta|K)^2 \leq (\alpha\beta|\alpha\beta) \cdot \max_{\gamma\delta} D_{\gamma\delta} \leq \left(\min(\tau, 10^{-8})\right)^2, \tag{15}$$

meaning that we set $L_{\alpha\beta}^J = 0$ if $\alpha\beta$ satisfies the criterion in Eq. (15).

To evaluate Eq. (14), we decompose $\boldsymbol{S}$, $\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{Q}^T$, and then invert $\boldsymbol{Q}$. Note that $\boldsymbol{Q}^{-1}$ may alternatively be obtained by inverse factorization of $\boldsymbol{S}$.[26] We will not consider this alternative here since the inversion of $\boldsymbol{Q}$, which costs the same as decomposing $\boldsymbol{S}$ (about $N_J^3/3$), is not a bottleneck in our implementation.

Both Cholesky and RI based methods are available using the two-step Cholesky decomposition approach. The integrals may be constructed from the Cholesky vectors, computed as in Eq. (14), using Eq. (1). One may also use the RI expression given in Eq. (2) directly. For systems where $\boldsymbol{L}$ cannot be stored, we may instead store $\boldsymbol{Q}^{-1}$ or $\boldsymbol{S}^{-1}$, together with $\mathcal{B}$, thereby reducing the storage requirement to be proportional to N_{AO}^2 rather than N_{AO}^3.

Due to the cost of electron repulsion integral evaluation, special considerations are necessary in step I. We use the Libint integral package[27] in which $(\alpha\beta|\gamma\delta)$ is computed together with all the integrals in the shell quartet $(AB|CD)$, where $\alpha \in A$, $\beta \in B$, $\gamma \in C$, and $\delta \in D$. Therefore, we modify the screening and qualification steps such that shell pairs are treated instead of AO pairs. For instance, $\alpha\beta \in \mathcal{D}$ if at least one diagonal in AB exceeds τ. There is also a trade-off between numerical stability and efficiency: we want to both qualify diagonal indices (add AO pairs to $\mathcal{Q}$) in descending order and compute as few integrals as possible. Shell pairs AB are therefore ordered with respect to their maximal diagonal element

$$D_{\text{max}}^{\text{AB}} = \max_{\alpha\beta \in AB} D_{\alpha\beta}. \tag{16}$$

Diagonals are then qualified from the AB with the largest diagonal before the next shell pair in the ordered list is considered. To ensure that selected diagonals are not too small, which may result in numerical instabilities, we use $\sigma = 10^{-2}$ as proposed in Ref. 12. In this way, $\mathcal{Q}$ may involve relatively few shell pairs while also containing potential basis elements $J = \alpha\beta$ associated with large diagonals $D_{\alpha\beta}$.[5] Similarly, when the Cholesky vectors are constructed from Eq. (14), the screening in Eq. (15) is with respect to shell pairs, not individual AO pairs.

By using the standard criterion for screening the diagonals,

$$\mathcal{D} = \{p \in \mathcal{D}_0 : D_p \geq \tau\}, \tag{17}$$

the integral matrix is reproduced to within the decomposition threshold τ. To reduce the memory and computational requirements, we propose a partitioned Cholesky decomposition (PCD). In PCD, we first partition the initial diagonal, $\mathcal{D}_0 = \mathcal{D}_{0,1} \cup \mathcal{D}_{0,2} \cup \ldots \cup \mathcal{D}_{0,K}$, and then determine the Cholesky basis for each diagonal block separately, resulting in $\mathcal{B}_1, \mathcal{B}_2, \ldots$, and $\mathcal{B}_K$. A final decomposition is then performed using $\mathcal{D}_0 = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_K$. Note that the decomposition threshold τ is not an upper bound on the error in PCD, although we have found the error to be controlled by τ in practice.

The proposed algorithm may be used in method specific and one-center decomposition approaches. We have implemented the one-center approximation 1C-CD, for which only a minor modification relative to regular Cholesky decomposition is needed. The $J = \gamma\delta$ are selected such that $\chi_\gamma(\boldsymbol{r})$ and $\chi_\delta(\boldsymbol{r})$ are centered on the same

Published under license by AIP Publishing

atom.[17] We therefore exclude all $\gamma\delta$ from $\mathcal{D}_0$ that do not satisfy the one-center requirement.

Furthermore, we present a method specific Cholesky decomposition targeting the molecular orbital (MO) integrals in an active space. The active orbital space is generated by partial Cholesky decomposition of the occupied and virtual AO densities, $\boldsymbol{D}^{\mathrm{o}}$ and $\boldsymbol{D}^{\mathrm{v}} = \boldsymbol{S}^{-1} - \boldsymbol{D}^{\mathrm{o}}$, where $\boldsymbol{S}$ is the AO overlap matrix. A brief description follows. For more details, see Refs. 28 and 29. Both $\boldsymbol{D}^{\mathrm{o}}$ and $\boldsymbol{D}^{\mathrm{v}}$ are decomposed with the restriction that pivot elements are centered on the preselected atoms. This results in the active occupied density,

$$(\boldsymbol{D}^{\mathrm{o}}_{\mathrm{a}})_{\alpha\beta} = \sum_i C^{\mathrm{a}}_{\alpha i} C^{\mathrm{a}}_{\beta i}, \tag{18}$$

and the active virtual density,

$$(\boldsymbol{D}^{\mathrm{v}}_{\mathrm{a}})_{\alpha\beta} = \sum_a C^{\mathrm{a}}_{\alpha a} C^{\mathrm{a}}_{\beta a}. \tag{19}$$

The inactive densities are defined as $\boldsymbol{D}^{\mathrm{o}}_{\mathrm{i}} = \boldsymbol{D}^{\mathrm{o}} - \boldsymbol{D}^{\mathrm{o}}_{\mathrm{a}}$ and $\boldsymbol{D}^{\mathrm{v}}_{\mathrm{i}} = \boldsymbol{D}^{\mathrm{v}}_{\mathrm{a}} - \boldsymbol{D}^{\mathrm{v}}$. We have adopted the multilevel Hartree-Fock approach[22] where $\boldsymbol{D}^{\mathrm{o}}$ is a superposition of atomic densities[30] guess that has been made idempotent by a single Fock matrix diagonalization. We define the active space screening criterion to be

$$\mathcal{D} = \{\alpha\beta \in \mathcal{D}_0 : D_{\alpha\beta} v_\alpha v_\beta \geq \tau \text{ and } D_{\alpha\beta} \geq \tau\}, \tag{20}$$

where

$$v_\alpha = \max_p (C^{\mathrm{a}}_{\alpha p})^2. \tag{21}$$

The accuracy of the active MO integrals, rather than the AO integrals, is then controlled by the decomposition threshold τ. The reader is referred to the work of Boman *et al.*[19] for more details on the method specific decomposition approach.

III. RESULTS AND DISCUSSION

The algorithm was implemented in eT, a coupled cluster program currently under development by the authors and collaborators.[32] To demonstrate its performance, we report wall time comparisons to the OpenMolcas program[7] for regular Cholesky decomposition on the formaldehyde-water system in Fig. 2. We compare our implementation to both the default one-step algorithm and the two-step algorithm described in Ref. 12, also implemented in OpenMolcas. In these calculations, we use the Dunning basis sets aug-cc-pVXZ, X ∈ {D, T, Q, 5}.[33] The results are summarized in Table I. A tight decomposition threshold of $\tau = 10^{-8}$ is used to demonstrate the numerical stability of the algorithm. Compared to the default algorithm in OpenMolcas, the total decomposition time T is reduced by about an order of magnitude. Similarly, T is significantly reduced, in all calculations, compared to the two-step implementation in OpenMolcas. Consequently, as the number of AOs increase, T rapidly becomes negligible compared to the time spent converging the Hartree-Fock self-consistent field (SCF) equations using QChem, where the integrals are by default computed directly.

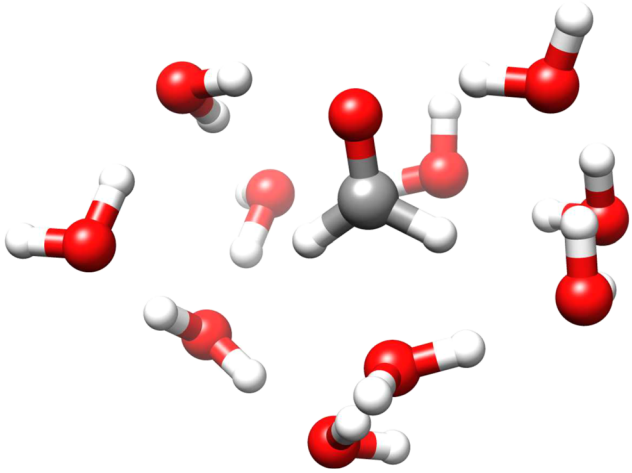


FIG. 2. Formaldehyde surrounded by ten water molecules.

For all decomposition carried out with eT in Table I, we computed the maximum error in the approximated integral matrix from

TABLE I. Wall time comparisons between eT and OpenMolcas[7] for formaldehyde surrounded by ten water molecules. The total decomposition time in eT is $T = T_1 + T_2$, where T_1 and T_2 are the times to determine $\mathcal{B}$ and $\boldsymbol{Q}^{-1}$ and to construct the Cholesky vectors, respectively. T_{default} and $T_{\text{two-step}}$ are the total decomposition times of the default and two-step algorithms in OpenMolcas. Also given is the time to converge the Hartree-Fock equations in QChem,[31] T_{SCF}. Time is in minutes unless other units are specified. In all calculations, $\tau = 10^{-8}$. Timings were made on an Intel Xeon CPU E5-2699 v4 with 1.5TB shared memory using 22 threads.

		OpenMolcas			eT				QChem[a]
N_{AO}		$\{N_J\}_{\text{default}}$	T_{default}	$T_{\text{two-step}}$	N_J	T	T_1	T_2	T_{SCF}
aug-cc-pVDZ	474	5 481	5	193*s*	5 374	63*s*	35*s*	28*s*	94*s*
aug-cc-pVTZ	1058	11 184	60	43	11 212	11	5	6	25
aug-cc-pVQZ	1972	19 336	553	534	19 297	79	34	45	249
aug-cc-pV5Z	3284	30 635	5277	3946	30 950	498	186	312	7985

[a]Version 5.0.2.

Published under license by AIP Publishing

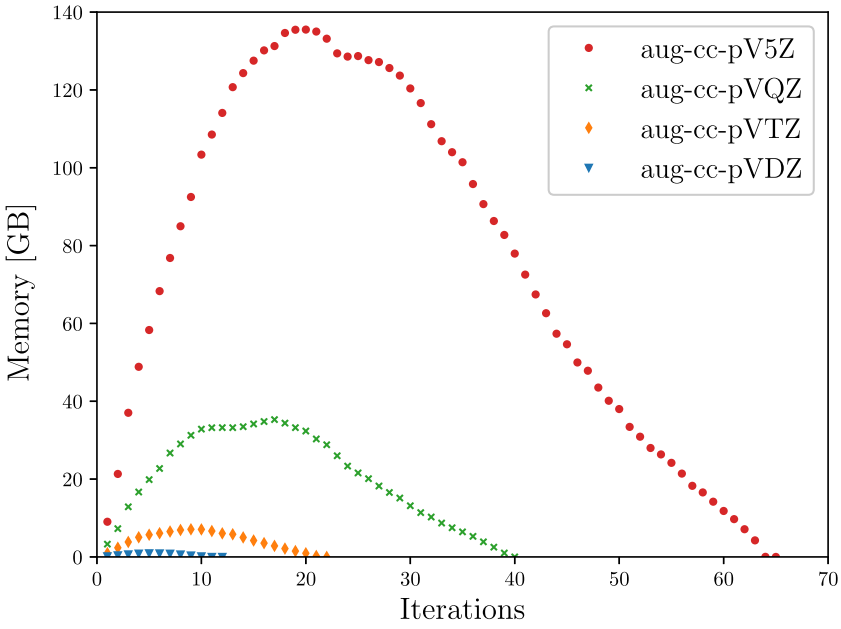


FIG. 3. Memory required to hold the Cholesky vectors $\boldsymbol{L}$ in each iteration of the pivoting procedure for formaldehyde surrounded by ten water molecules.

the diagonal of the matrix [see Eqs. (5) and (6)]. As expected, we found that the largest error in the matrix is less than τ in all cases, confirming that the integral matrix has been correctly decomposed.

The memory required to hold $\boldsymbol{L}$ varies as expected (see Fig. 3). It increases to a maximum and then drops off to zero during step I, giving a large reduction in memory usage compared to the previous algorithms.[5] To reduce the memory requirements further, one may use PCD. We present PCD calculations on the formaldehyde-water system using the aug-cc-pV5Z basis for a set of K values (see Table II). The peak memory usage is significantly reduced for all $K > 1$, and the time to determine $\mathcal{B}$ and $\boldsymbol{Q}^{-1}$ is reduced by up to a factor of two. However, the maximal error ϵ in the electron repulsion integral matrix, listed in Table II, is about an order of magnitude larger than τ. This error may be lowered by decreasing τ in all decomposition or only in the final decomposition.

TABLE II. Cholesky decomposition with K diagonal batches on formaldehyde surrounded by ten water molecules using the aug-cc-pV5Z basis. Here, N_J is the number of Cholesky vectors, T_1 is the time to determine $\mathcal{B}$ and $\boldsymbol{Q}^{-1}$, and ϵ is the maximal error in the matrix $\boldsymbol{M}$. Also given is the peak memory requirement to hold the Cholesky vectors. In all calculations, $\tau = 10^{-8}$.

K	N_J	T_1 (min)	Memory (GB)	ϵ
1	30 950	186	134	$<\tau$
2	30 313	158	56	15τ
4	30 374	123	22	17τ
6	30 486	106	22	15τ
8	30 450	90	24	19τ
10	30 459	102	25	13τ
12	30 407	103	28	16τ

To assess the reduction in computational cost, we estimate the PCD operation count and compare it to regular Cholesky decomposition (CD). The CD operation count, disregarding the screening on the rows in step I and assuming that all qualified diagonals are included in the basis, is

$$O_{\mathrm{CD}} = N_{\alpha\beta} N_J^2, \tag{22}$$

where $N_{\alpha\beta}$ is the number of elements in the initial $\mathcal{D}$ [see Eq. (9)]. Under the same assumptions, the operation count of PCD is

$$O_{\mathrm{PCD}} = \sum_{i=1}^{K} \left(\frac{N_{\alpha\beta}}{K} (N_J^i)^2 + N_J^i N_J^2 \right) \approx N_{\alpha\beta} (N_J^{\mathrm{b}})^2 + K N_J^{\mathrm{b}} N_J^2, \tag{23}$$

where N_J^i is the number of elements in $\mathcal{B}_i$ and we have assumed that N_J is independent of K. In the second equality, we assume that each diagonal block has the rank N_J^{b}. The ratio of operation counts is therefore

$$\frac{O_{\mathrm{PCD}}}{O_{\mathrm{CD}}} = \frac{(N_J^{\mathrm{b}})^2}{N_J^2} + \frac{N_J^{\mathrm{b}} K}{N_{\alpha\beta}}. \tag{24}$$

In the limiting cases where $K = 1$ and $K = N_{\alpha\beta}$, we have $N_J^{\mathrm{b}} = N_J$ and $N_J^{\mathrm{b}} = 1$, respectively, and there are no savings with PCD. Considering the two fractions, we see that K should be chosen such that $N_J^{\mathrm{b}} \ll N_J$ and $N_J^{\mathrm{b}} \ll N_{\alpha\beta}/K$. In other words, the block dimension $(N_{\alpha\beta}/K)$ should be significantly smaller than the full matrix dimension $(N_{\alpha\beta})$, yet also large enough for each block to be rank deficient.

Method specific Cholesky decomposition may also be used to treat large systems. Here, we apply a multilevel screening, where regions of the system are chosen to be active and the target quantities are the active space MO integrals. We consider an active formaldehyde molecule surrounded by 10–200 water molecules. In Fig. 4, we

Published under license by AIP Publishing

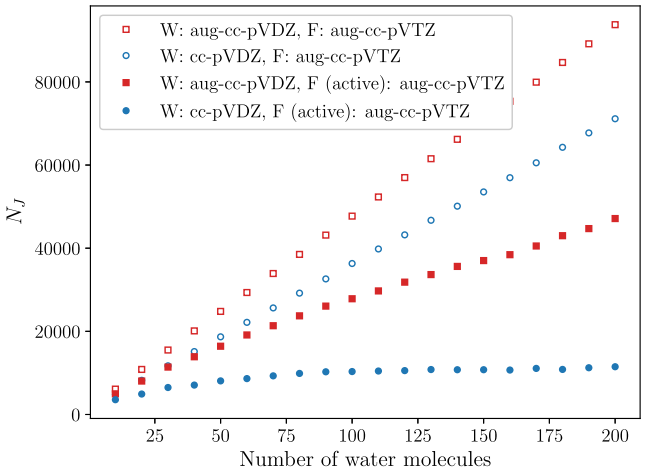


FIG. 4. The number of Cholesky vectors N_J, in full decomposition and active space decomposition, for formaldehyde (F) surrounded by 10–200 water (W) molecules.

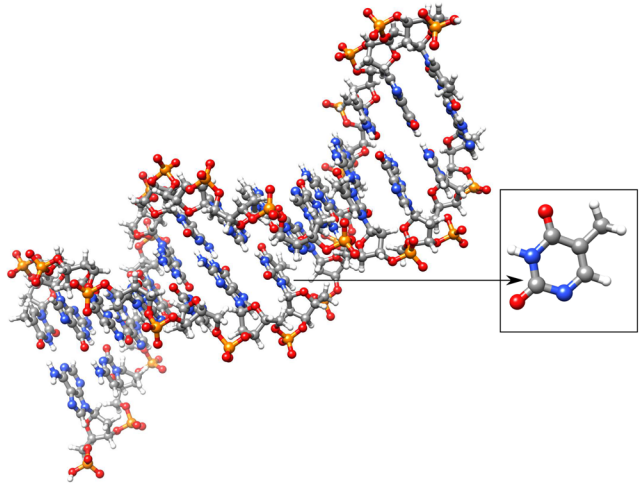


FIG. 5. DNA fragment with active thymine.

show the number of vectors obtained with the standard and active space screenings defined in Eqs. (17) and (20). With the standard screening, the number of Cholesky vectors increases linearly with system size, whereas it flattens out with the active space screening. We construct the active orbitals as follows. The active occupied orbitals are generated from $\boldsymbol{D}^{\mathrm{o}}$ by restricting the number of pivots to equal half the number of electrons on the active atoms. In the general case, one pivot is added if an active atom is bound to an inactive atom, effectively adding an orbital to the active occupied space. The number of pivots used to decompose $\boldsymbol{D}^{\mathrm{v}}$ is restricted such that one obtains the same fraction of virtual to occupied orbitals as in the full orbital space. Alternatively, a decomposition threshold may be used to determine the number of pivots in the decomposition of $\boldsymbol{D}^{\mathrm{o}}$ and $\boldsymbol{D}^{\mathrm{v}}$.[22]

With the standard screening in Eq. (17), the algorithm may be used to decompose the integral matrix of systems with more than 10 000 basis functions. With the method specific and one-center approaches, the applicability of the algorithm is further extended. To show that the algorithm can handle large systems, we determine $\mathcal{B}$ and $\boldsymbol{Q}^{-1}$ for the DNA fragment in Fig. 5. The time T_1 to determine $\mathcal{B}$ and $\boldsymbol{Q}^{-1}$, and N_J, are given in Table III. Decomposition using active space screening and the one-center approximation are also listed. For the active space calculations, a single thymine is active. Furthermore, we present full, active space, and one-center calculations on the retinal bound to rhodopsin (see Fig. 6). Retinal is active in the active space calculations. The number of Cholesky vectors is given in Table IV.

Finally, coupled cluster singles and doubles (CCSD)[36] calculations were performed on the formaldehyde-water system in Fig. 2 using approximated integrals [see Eq. (7)]. The Cholesky-based CCSD calculations were performed using our spin adapted singlet CCSD implementation in the eT program. In Table V, we list CCSD correlation energies for different τ using the CD and

TABLE III. Full, active space, and one-center Cholesky decomposition for the DNA fragment. Here, N_{AO} is the number of AOs, τ is the decomposition threshold, N_J is the number of Cholesky vectors, and T_1 is the wall time to determine $\mathcal{B}$ and $\boldsymbol{Q}^{-1}$.

Method	Basis	N_{AO}	τ	N_{J}	T_1 (min)
Full decomposition	aug-cc-pVDZ	15 064	10^{-2}	24 591	49[a]
			10^{-4}	53 719	330[a]
			10^{-6}	95 292	1065[a]
			10^{-8}	158 811	5506[b]
Active space decomposition	cc-pVDZ/aug-cc-pVTZ	9 447	10^{-8}	19 375	20[c]
	aug-cc-pVDZ/aug-cc-pVTZ	15 341	10^{-8}	90 551	1389[b]
One-center decomposition	aug-cc-pVDZ	15 064	10^{-4}	49 533	54[c]
			10^{-8}	89 489	802[c]

[a]Intel Xeon Gold 6152 and 1.5TB shared memory. Calculation on 44 threads.
[b]Intel Xeon Gold 6132 and 6TB shared memory. Calculation on 140 threads.
[c]Intel Xeon E5-2699 v4 and 1.5TB shared memory. Calculation on 22 threads.

Published under license by AIP Publishing

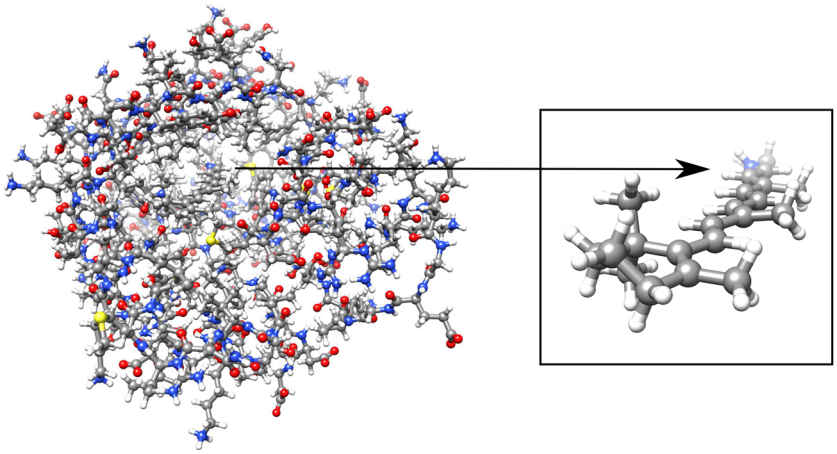


FIG. 6. Retinal bound to rhodopsin with active retinal.

TABLE IV. Full, active space, and one-center Cholesky decomposition for the retinal-rhodopsin system. Here, N_{AO} is the number of AOs, τ is the decomposition threshold, and N_J is the number of Cholesky vectors.

Method	Basis	N_{AO}	τ	N_J
Full decomposition	aug-cc-pVDZ	36 787	10^{-4}	124 632
Active space decomposition	cc-pVDZ/aug-cc-pVTZ	23 134	10^{-8}	77 719
One-center decomposition	cc-pVDZ	21 840	10^{-8}	119 357
	aug-cc-pVDZ	36 787	10^{-8}	202 935
	aug-cc-pVDZ	36 787	10^{-4}	112 592
	aug-cc-pVTZ	79 420	10^{-4}	257 198

TABLE V. CCSD/aug-cc-pVDZ correlation energies E_{corr} for formaldehyde surrounded by ten water molecules using CD and PCD bases as well as prefitted RI. Errors ΔE_{corr} are computed as the deviation from a CCSD calculation in DALTON[6] with no electron repulsion integral approximation. The RI-CCSD calculations were done in Psi4[34] using aug-cc-pVDZ with standard auxiliary bases (aug-cc-pVDZ-JKFIT for the Hartree-Fock part and aug-cc-pVDZ-RI for the coupled cluster part). The number of elements in the auxiliary basis is denoted N_J for CD, PCD, and RI.

Method	N_J	τ	$E_{corr}(E_h)$	$\Delta E_{corr}(E_h)$
CD	844	10^{-2}	−2.707 587 13	$1.0 \cdot 10^{-2}$
	1185	10^{-3}	−2.716 886 88	$1.1 \cdot 10^{-3}$
	1920	10^{-4}	−2.717 962 18	$-1.8 \cdot 10^{-5}$
	5374	10^{-8}	−2.717 944 46	$-5.3 \cdot 10^{-8}$
PCD	801	10^{-2}	−2.705 702 68	$1.2 \cdot 10^{-2}$
	1184	10^{-3}	−2.716 792 90	$1.2 \cdot 10^{-3}$
	1889	10^{-4}	−2.717 967 10	$-2.3 \cdot 10^{-5}$
	5333	10^{-8}	−2.717 944 42	$-9.3 \cdot 10^{-9}$
RI-CCSD	1370[a]	...	−2.719 302 45	$-1.4 \cdot 10^{-3}$
No approximation	...	...	−2.717 944 41	...

[a]aug-cc-pVDZ-RI.

PCD decomposition methods. The results are compared to an RI-CCSD calculation, using Psi4,[34] and a calculation with the exact electron repulsion integrals obtained with the DALTON software.[6] In Table VI, the lowest singlet excitation energy is given, for different τ, and compared to the calculation using the exact electron repulsion integrals.

TABLE VI. Lowest CCSD/aug-cc-pVDZ singlet excitation energy ω of formaldehyde surrounded by ten water molecules using CD bases obtained with eT. Errors $\Delta\omega$ are computed as the deviation from a CCSD calculation in DALTON[6] with no electron repulsion integral approximation. The used Hartree-to-eV conversion factor is the CODATA 2014 value 27.211 386 02.[35] The threshold τ is in atomic units.

Method	N_J	τ	ω (eV)	$\Delta\omega$ (eV)
CD	844	10^{-2}	4.745 280 86	$6.3 \cdot 10^{-3}$
	1185	10^{-3}	4.743 395 31	$4.4 \cdot 10^{-3}$
	1920	10^{-4}	4.739 269 87	$3.1 \cdot 10^{-4}$
	5374	10^{-8}	4.738 958 24	$4.2 \cdot 10^{-8}$
No approximation	...	...	4.738 958 20	...

Published under license by AIP Publishing

As expected, the error in the correlation energy and excitation energy is proportional to τ, reflecting the improved accuracy of the approximated integrals with decreasing τ. This trend is observed for both the CD and PCD methods. Considering the RI-CCSD calculation, we note that the error ($1.4 \cdot 10^{-3}$ E_h) is comparable to that of CD and PCD for $\tau = 10^{-3}$ ($1.1 \cdot 10^{-3}$ E_h and $1.2 \cdot 10^{-3}$ E_h, respectively). However, the size of the RI auxiliary basis is 16% larger than the CD and PCD bases obtained with $\tau = 10^{-3}$. As the cost in nonintegral-direct post-HF methods depends on the number of auxiliary functions, CD and PCD are in this case superior to prefitted RI. In the case of integral-direct implementations, however, one must take into account that integrals involving the two-center functions of the Cholesky bases are more expensive. We also mention that if an accuracy of <1 mE_h in the excitation energy is desired, CD with $\tau = 10^{-2}$ is sufficient in this case.

IV. CONCLUDING REMARKS

In recent decades, the Cholesky decomposition of the electron repulsion integral matrix has been implemented in popular quantum chemistry programs. While the technique allows for complete control of the error, a drawback has been its computational cost compared to prefitted RI. With this contribution, the application range of Cholesky decomposition is extended, and its competitiveness with other inner-projection methods is improved. We have already performed full decomposition for systems with tens of thousands of atomic orbitals, yet we expect that the partitioned decomposition approach may be applied to much larger systems. While useful in its own right, the Cholesky decomposition may also be used as an accurate starting point for the development of other integral approximations, such as the reduced-scaling tensor hypercontraction schemes.[37]

ACKNOWLEDGMENTS

We thank A. Sánchez de Merás for insightful discussions of the algorithm and its implementation in the early stages of the project. We also thank E. Valeev for assistance with the Libint integral package, A. Krylov for assistance with QChem, and T. B. Pedersen for assistance with the two-step algorithm in OpenMolcas.

We acknowledge computer resources from NOTUR through Project No. nn2962k and the high performance computer facilities of the SMART Laboratory. We also acknowledge funding from the Marie Skłodowska-Curie European Training Network through Grant Agreement No. 765739 (COSINE) and the Norwegian Research Council grants CCGPU (Grant No. 263110) and TheoLight (Grant No. 275506). H.K. acknowledges the Otto Mønsted Fond, and S.D.F. and E.F.K acknowledge Fondet til professor Leif Tronstads minne.

REFERENCES

[1] N. H. F. Beebe and J. Linderberg, Int. J. Quantum Chem. **12**, 683 (1977).
[2] P.-O. Löwdin, Phys. Rev. **139**, A357 (1965).
[3] P.-O. Löwdin, Int. J. Quantum Chem. **5**, 231 (2009).
[4] I. Røeggen and E. Wisløff-Nilssen, Chem. Phys. Lett. **132**, 154 (1986).
[5] H. Koch, A. Sánchez de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003).
[6] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thŕgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ĺgren, Wiley Interdiscip. Rev. Comput .Mol. Sci. **4**, 269 (2013).
[7] F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. F. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, J. Comput. Chem. **37**, 506 (2015).
[8] F. Aquilante, T. B. Pedersen, R. Lindh, B. O. Roos, A. Sánchez de Merás, and H. Koch, J. Chem. Phys. **129**, 024113 (2008).
[9] T. B. Pedersen, A. M. J. Sánchez de Merás, and H. Koch, J. Chem. Phys. **120**, 8887 (2004).
[10] T. D. Crawford, L. S. Owens, M. C. Tam, P. R. Schreiner, and H. Koch, J. Am. Chem. Soc. **127**, 1368 (2005).
[11] F. Aquilante, P.-Å. Malmqvist, T. B. Pedersen, A. Ghosh, and B. O. Roos, J. Chem. Theory Comput. **4**, 694 (2008).
[12] F. Aquilante, L. Boman, J. Boström, H. Koch, R. Lindh, A. S. de Merás, and T. B. Pedersen, "Cholesky decomposition techniques in electronic structure theory," in *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*, edited by R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski (Springer Netherlands, Dordrecht, 2011), pp. 301–343.
[13] P. Merlot, T. Kjærgaard, H. Trygve, R. Lindh, F. Aquilante, S. Reine, and T. B. Pedersen, J. Comput. Chem. **34**, 1486 (2013).
[14] O. Vahtras, J. Almlöf, and M. W. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).
[15] K. Eichkorn, O. Treutler, H. öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).
[16] K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Theor. Chem. Acc. **97**, 119 (1997).
[17] F. Aquilante, R. Lindh, and T. Bondo Pedersen, J. Chem. Phys. **127**, 114107 (2007).
[18] T. B. Pedersen, F. Aquilante, and R. Lindh, Theor. Chem. Acc. **124**, 1 (2009).
[19] L. Boman, H. Koch, and A. Sánchez de Merás, J. Chem. Phys. **129**, 134107 (2008).
[20] R. H. Myhre, A. M. J. Sánchez de Merás, and H. Koch, J. Chem. Phys. **141**, 224105 (2014).
[21] R. H. Myhre and H. Koch, J. Chem. Phys. **145**, 044111 (2016).
[22] S. Sæther, T. Kjærgaard, H. Koch, and I.-M. Høyvik, J. Chem. Theory Comput. **13**, 5282 (2017).
[23] C. Hättig, J. Chem. Phys. **118**, 7751 (2003).
[24] D. W. O'neal and J. Simons, Int. J. Quantum Chem. **36**, 673 (1989).
[25] F. Aquilante, R. Lindh, and T. B. Pedersen, J. Chem. Phys. **129**, 034106 (2008).
[26] M. Benzi, C. Meyer, and M. Tůma, SIAM J. Sci. Comput. **17**, 1135 (1996).
[27] E. Valeev, "Libint: A library for the evaluation of molecular integrals of many-body operators over gaussian functions" (2017).
[28] F. Aquilante, T. Bondo Pedersen, A. Sánchez de Merás, and H. Koch, J. Chem. Phys. **125**, 174101 (2006).

Published under license by AIP Publishing

[29]A. M. J. Sánchez de Merás, H. Koch, I. G. Cuesta, and L. Boman, J. Chem. Phys. **132**, 204105 (2010).
[30]J. H. Van Lenthe, R. Zwaans, H. J. J. Van Dam, and M. F. Guest, J. Comput. Chem. **27**, 926 (2006).
[31]Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng *et al.*, Mol. Phys. **113**, 184 (2015).
[32]The eT program will be described in a forthcoming publication.
[33]T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[34]J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams *et al.*, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 556 (2012).
[35]P. J. Mohr, D. B. Newell, and B. N. Taylor, J. Phys. Chem. Ref. Data **45**, 043102 (2016).
[36]G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).
[37]E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, J. Chem. Phys. **137**, 044103 (2012).

Published under license by AIP Publishing