

## An efficient algorithm for Cholesky decomposition of electron repulsion integrals

Sarai D. Folkestad,<sup>1,2, a)</sup> Eirik F. Kjørstad,<sup>1,2, a)</sup> and Henrik Koch<sup>2, 1, b)</sup>

<sup>1)</sup>*Department of Chemistry, Norwegian University of Science and Technology, N-7491 Trondheim, Norway*

<sup>2)</sup>*Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy*

(Dated: 27 October 2019)

Approximating the electron repulsion integrals using an inner projection method, such as Cholesky decomposition, 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 80000 atomic orbitals.

---

<sup>a)</sup>Contributed equally to this work

<sup>b)</sup>Electronic mail: henrik.koch@sns.it

## I. INTRODUCTION

The Beebe-Linderberg<sup>1</sup> algorithm for the Cholesky decomposition 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.<sup>2,3</sup> The algorithm was later modified to include screening by Røeggen and Wisløff-Nilssen,<sup>4</sup> who also demonstrated that the numerical rank is proportional to the number of atomic orbitals, as had already been suggested.<sup>1</sup> An algorithm suited for large-scale applications was first proposed in 2003 by Koch, Sánchez de Merás and Pedersen.<sup>5</sup> This algorithm was implemented in Dalton<sup>6</sup> and subsequently included in the Molcas program.<sup>7,8</sup> A number of applications based on the Cholesky decomposition of the integrals have since been published.<sup>9-12</sup>

The resolution of identity (RI) method, also called density fitting,<sup>13</sup> is an inner projection technique introduced by Vahtras, Almlöf and Feyereisen.<sup>14</sup> 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.<sup>15,16</sup> 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 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 coworkers have advocated the atomic (aCD) and one-center (1C-CD) decomposition methods, where the Cholesky basis is restricted to one-center functions.<sup>17</sup> 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.<sup>17,18</sup>

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.<sup>19</sup> The approach is well suited for multilevel methods, where only subsets of integrals are needed in the correlated treatments.<sup>20–22</sup>

To be generally applicable, an integral approximation scheme must have analytic geometric derivatives. Such derivatives are easily derived for RI using prefitted auxiliary bases.<sup>23</sup> Although not apparent in the early discussion of gradients by O’neal and Simons,<sup>24</sup> 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.<sup>25</sup>

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<sup>17</sup> and a Cholesky decomposition for multilevel calculations. We also present a Cholesky decomposition where, initially, 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  $M$  is symmetric positive semidefinite and may therefore be Cholesky decomposed,

$$M_{\alpha\beta,\gamma\delta} = (\alpha\beta|\gamma\delta) = \sum_J L_{\alpha\beta}^J L_{\gamma\delta}^J = (LL^T)_{\alpha\beta,\gamma\delta}, \quad (1)$$



FIG. 1. While determining the basis  $\mathcal{B}$ , we screen out elements of  $M$  and  $L$  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  $L$  is kept in memory throughout the decomposition procedure.

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

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

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

$$L_{\alpha\beta}^J = \sum_K (\alpha\beta|\rho_K)Q_{KJ}^{-T}. \quad (3)$$

That is, a Cholesky decomposition is equivalent to an RI approximation.<sup>1,14</sup>

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

$$L_p^J = \frac{M_{pJ}}{\sqrt{M_{JJ}}} \quad (4)$$

is constructed. Finally,  $M$  is updated according to

$$M_{pq} \leftarrow M_{pq} - L_p^J L_q^J. \quad (5)$$

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

$$M_{pq}^2 \leq M_{pp}M_{qq}, \quad (6)$$

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

$$M_{pq} \approx \sum_J L_p^J L_q^J, \quad (7)$$

where the error in  $M_{pq}$  is less than  $\tau$ . 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 too 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  $M$ , its diagonal elements decrease monotonously. Consequently, a diagonal  $D_p = M_{pp}$  below  $\tau$  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  $\tau$ , thus applying the same screening on the rows and columns of  $M$ . In algorithms where the Cholesky vectors are constructed during the pivoting procedure, screening on the rows of  $M$  must instead be with respect to the Cauchy-Schwarz inequality.<sup>5</sup>

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  $D_p \geq \tau$ . For  $J \in \mathcal{B}$ , only keep  $L_p^J$  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  $\mathcal{Q}$ ,

$$\mathcal{Q} = \{p \in \mathcal{D} : D_p \geq \sigma D_{\max}\}, \quad (8)$$

such that the number of elements in  $\mathcal{Q}$  does not exceed a user-specified maximum. The parameter  $\sigma$ , 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  $M$ :

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

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

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

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

$$\mathcal{Q} = \mathcal{Q} \setminus \{q\}, \quad \mathcal{C} = \mathcal{C} \cup \{q\}, \quad (11)$$

and the diagonal elements,

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

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

$$\mathcal{B} = \mathcal{B} \cup \mathcal{C}, \quad (13)$$

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  $M$  is the electron repulsion integral matrix, each pivot  $J = \gamma\delta \in \mathcal{B}$  defines a Cholesky basis function  $\rho_J(r) = \chi_\gamma(r)\chi_\delta(r)$ . The RI expression,

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

may then be used to construct the Cholesky vectors. We decompose  $\mathbf{S}$  and then invert  $\mathbf{Q}$ . Note, however, that  $\mathbf{Q}$  may be inverted while  $\mathbf{S}$  is decomposed.<sup>26</sup> **To construct  $L_{\alpha\beta}^J$ , we use the Cauchy-Schwarz screening**

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

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

From the RI formulation, an integral-direct approach is available. By storing  $\mathbf{Q}^{-1}$  and  $\mathcal{B}$ , the Cholesky vector  $L_{\alpha\beta}^J$  may be constructed on-the-fly from Eq. (14). This may be useful for systems where  $\mathbf{L}$  cannot be stored—the memory required would be proportional to  $N_{\text{AO}}^2$  rather than  $N_{\text{AO}}^3$ .

We use the Libint integral package,<sup>27</sup> in which  $(\alpha\beta|\gamma\delta)$  is computed together with all the integrals in the shell quadruple  $(AB|CD)$ , where  $\alpha \in A$ ,  $\beta \in B$ ,  $\gamma \in C$ , and  $\delta \in D$ . Therefore, the screening and qualification steps are modified 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  $\tau$ . 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

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

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, we use  $\sigma = 10^{-2}$ . In this way,  $\mathcal{Q}$  may involve relatively few shell pairs while also containing potential basis elements  $J = \alpha\beta$  associated with large diagonals  $D_{\alpha\beta}$ .<sup>5</sup> **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.**

The procedure described thus far reproduces the integral matrix to within the decomposition threshold  $\tau$ . However, the framework easily allows for method specific approximations that further reduce the number of elements in  $\mathcal{B}$ . We have implemented an active space screening where the target quantities are the molecular orbital (MO) integrals in a selected active space. First, the occupied and virtual AO densities,  $\mathbf{D}^{\text{o}}$  and  $\mathbf{D}^{\text{v}} = \mathbf{S}^{-1} - \mathbf{D}^{\text{o}}$ , are Cholesky decomposed with the restriction that pivot elements are centered on active atoms. This results in the active occupied density,

$$(\mathbf{D}_{\text{a}}^{\text{o}})_{\alpha\beta} = \sum_i C_{\alpha i}^{\text{a}} C_{\beta i}^{\text{a}}, \quad (17)$$

and the active virtual density,

$$(\mathbf{D}_{\text{a}}^{\text{v}})_{\alpha\beta} = \sum_a C_{\alpha a}^{\text{a}} C_{\beta a}^{\text{a}}. \quad (18)$$

The inactive densities are defined as  $\mathbf{D}_i^o = \mathbf{D}^o - \mathbf{D}_a^o$  and  $\mathbf{D}_i^v = \mathbf{D}_a^v - \mathbf{D}^v$ .<sup>28,29</sup> To generate the active orbital space, we have adopted the multilevel Hartree-Fock approach of Sæther and coworkers;<sup>22</sup> they use, as  $\mathbf{D}^o$ , a superposition of atomic densities<sup>30</sup> guess that has been made idempotent by a single Fock matrix diagonalization. We define the active space screening by replacing the requirements on the diagonals (in steps 2, 3, and 5) with

$$\mathcal{D} = \{\alpha\beta : (\alpha\beta|\alpha\beta)v_\alpha v_\beta \geq \tau\}, \quad (19)$$

where

$$v_\alpha = \max_p (C_{\alpha p}^a)^2. \quad (20)$$

The accuracy of the active MO integrals, rather than the AO integrals, is then controlled by the decomposition threshold  $\tau$ . The reader is referred to Boman et al.<sup>19</sup> for more details on the method specific decomposition approach.

Similarly, only a minor modification of the algorithm is needed to obtain the one-center approximation 1C-CD. In 1C-CD, the  $J = \gamma\delta$  are selected such that  $\chi_\gamma(\mathbf{r})$  and  $\chi_\delta(\mathbf{r})$  are centered on the same atom.<sup>17</sup> To implement 1C-CD, we altered the initial screening to exclude all  $\gamma\delta$  from  $\mathcal{D}$  that do not satisfy the one-center requirement.

### III. RESULTS AND DISCUSSION

The algorithm was implemented in eT, a coupled cluster program currently under development by the authors and collaborators. To demonstrate its performance, we report wall time comparisons to the OpenMolcas program<sup>7</sup> on the formaldehyde-water system in Fig. 2. In these calculations, we use the Dunning basis sets aug-cc-pVXZ,  $X \in \{D, T, Q, 5\}$ .<sup>34</sup> The results are summarized in Table I. Compared to OpenMolcas, the total decomposition time  $T$  is reduced by about an order of magnitude. Consequently, as the number of AOs increase,  $T$  rapidly becomes negligible compared to the time spent converging the Hartree-Fock equations. **For all decompositions in Table I using eT, we computed the maximum error on the diagonal of the approximated integral matrix  $\mathbf{LL}^T$ . As  $\mathbf{M} - \mathbf{LL}^T$  is positive semidefinite, it follows from the Cauchy-Schwarz inequality that the element with the largest error is on the diagonal. For all basis sets, we found that the largest error in  $\mathbf{LL}^T$  was less than  $\tau$ .**

The memory required to hold  $\mathbf{L}$  varies as expected, see Fig. 3. It increases to a maximum during the decomposition and then drops off to zero, giving a large reduction in memory

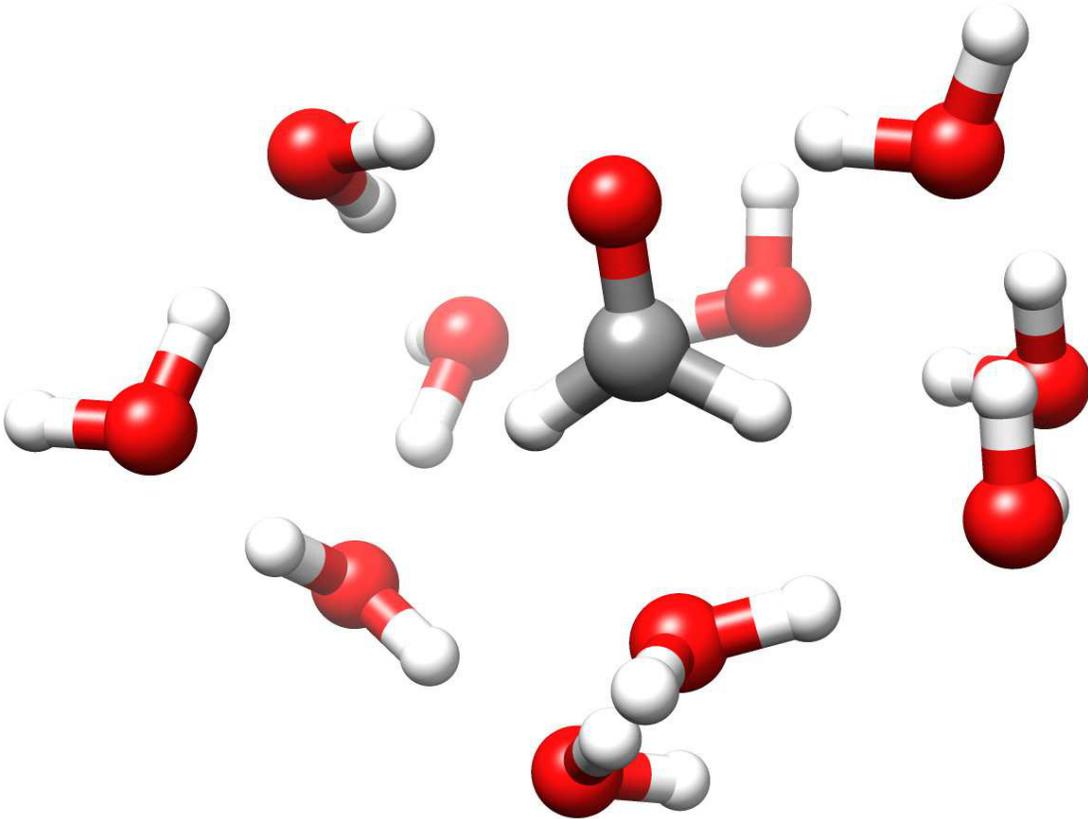


FIG. 2. Formaldehyde surrounded by ten water molecules.

usage compared to the previous algorithms.<sup>5</sup> However, to reduce the memory requirements further, the following partitioned matrix algorithm may be used. First, the significant diagonal is partitioned,  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \dots \cup \mathcal{D}_K$ , and each diagonal batch decomposed separately, resulting in  $\mathcal{B}_1, \mathcal{B}_2, \dots$ , and  $\mathcal{B}_K$ . A final decomposition is then performed using  $\mathcal{D} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_K$ . With this approach, the decomposition threshold  $\tau$  is not an upper bound on the error. However, we have found that the error is controlled by  $\tau$  in practice. The error may be lowered by decreasing  $\tau$  in all decompositions or only in the final decomposition. We present 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  $K$  considered, and the time to determine  $\mathcal{B}$  and  $\mathcal{Q}^{-1}$  is reduced by up to a factor of two.

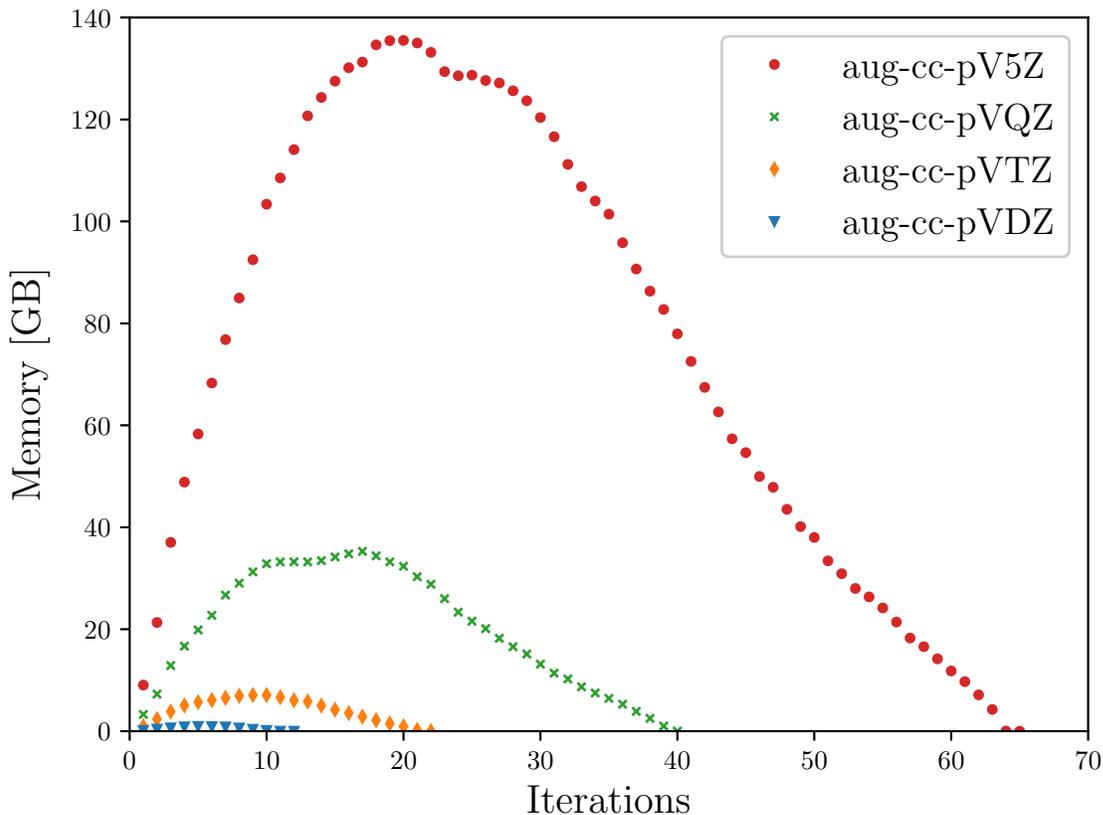


FIG. 3. Memory required to hold the Cholesky vectors  $\mathbf{L}$  in each iteration of the decomposition for formaldehyde surrounded by ten water molecules.

Method specific screenings 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 show the number of vectors obtained with the standard and active space screenings defined in Eqs. (??) and (19). 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  $\mathbf{D}^o$  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. Similarly, the number of pivots used to decompose  $\mathbf{D}^v$

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

	OpenMolcas			eT				QChem <sup>a</sup>
	$N_{\text{AO}}$	$N_J$	$T$	$N_J$	$T$	$T_1$	$T_2$	$T_{\text{SCF}}$
aug-cc-pVDZ	474	5481	7	5374	63 s	35s	28s	94 s
aug-cc-pVTZ	1058	11184	70	11212	11	5	6	25
aug-cc-pVQZ	1972	19336	589	19297	79	34	45	249
aug-cc-pV5Z	3284	30635	5534	30950	498	186	312	7985

<sup>a</sup> Version 5.0.2.

is restricted such that one obtains the same fraction of virtual to occupied orbitals as in the entire set of orbitals. Alternatively, a decomposition threshold may be used to determine the number of pivots in the decomposition of  $\mathbf{D}^o$  and  $\mathbf{D}^v$ .<sup>22</sup>

The algorithm may be used to decompose the integral matrix of systems with more than ten thousand basis functions. With the method specific and one-center approaches, the applicability of the algorithm is further extended. To show that the algorithm can tackle large systems, we determine  $\mathcal{B}$  and  $\mathbf{Q}^{-1}$  for the DNA fragment in Fig. 5. The time  $T_1$  to determine  $\mathcal{B}$  and  $\mathbf{Q}^{-1}$ , and  $N_J$ , are given in Table III. Decompositions 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 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)<sup>35</sup> calculations were performed on the formaldehyde-water system in Figure 2 using approximated integrals  $(pq|rs) = \sum_{J=1}^{N_J} L_{pq}^J L_{rs}^J$ . 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

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,  $T_1$  the time to determine  $\mathcal{B}$  and  $\mathbf{Q}^{-1}$ , and  $\epsilon$  is the maximal error in the matrix  $\mathbf{M}$ . 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]	$\epsilon$
1	30950	186	134	$< \tau$
2	30313	158	56	$15\tau$
4	30374	123	22	$17\tau$
6	30486	106	22	$15\tau$
8	30450	90	24	$19\tau$
10	30459	102	25	$13\tau$
12	30407	103	28	$16\tau$

$\tau$  using the CD and PCD decomposition methods. The results are compared to an RI-CCSD calculation, using Psi4,<sup>32</sup> and a calculation with exact electron repulsion integrals obtained with the DALTON software.<sup>6</sup> In Table VI, the lowest singlet excitation energy for different  $\tau$  is given and compared to the calculation using exact electron repulsion integrals.

As expected, the error in the correlation energy and excitation energy is proportional to  $\tau$ , reflecting the improved accuracy of the approximated integrals with decreasing  $\tau$ . 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 non-integral-direct post-HF methods depend on the number of auxiliary functions, CD and PCD is 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 should mention that if an accuracy of  $< 1 \text{ mE}_h$  in  $\omega$  is desired, CD with  $\tau = 10^{-2}$  is sufficient in this case.

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

Method	Basis	$N_{\text{AO}}$	$\tau$	$N_J$	$T_1$ [min]
Full decomposition	aug-cc-pVDZ	15064	$10^{-2}$	24591	49 <sup>a</sup>
			$10^{-4}$	53742	532 <sup>b</sup>
			$10^{-6}$	95403	1854 <sup>b</sup>
			$10^{-8}$	158811	5506 <sup>b</sup>
Active space decomposition	cc-pVDZ/aug-cc-pVTZ	9447	$10^{-8}$	19375	20 <sup>c</sup>
	aug-cc-pVDZ/aug-cc-pVTZ	15341	$10^{-8}$	90551	1389 <sup>b</sup>
One-center decomposition	aug-cc-pVDZ	15064	$10^{-4}$	49533	54 <sup>c</sup>
			$10^{-8}$	89489	802 <sup>c</sup>

<sup>a</sup> Intel Xeon Gold 6152 and 1.5TB shared memory. Calculation on 44 threads.

<sup>b</sup> Intel Xeon Gold 6132 and 6TB shared memory. Calculation on 140 threads.

<sup>c</sup> Intel Xeon E5-2699 v4 and 1.5TB shared memory. Calculation on 22 threads.

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

Method	Basis	$N_{\text{AO}}$	$\tau$	$N_J$
Full decomposition	aug-cc-pVDZ	36787	$10^{-4}$	124632
Active space decomposition	cc-pVDZ/aug-cc-pVTZ	23134	$10^{-8}$	77719
One-center decomposition	cc-pVDZ	21840	$10^{-8}$	119357
	aug-cc-pVDZ	36787	$10^{-8}$	202935
	aug-cc-pVDZ	36787	$10^{-4}$	112592
	aug-cc-pVTZ	79420	$10^{-4}$	257198

#### IV. CONCLUDING REMARKS

In recent decades, the Cholesky decomposition of the electron repulsion integrals has been implemented in popular quantum chemistry programs. While the technique allows

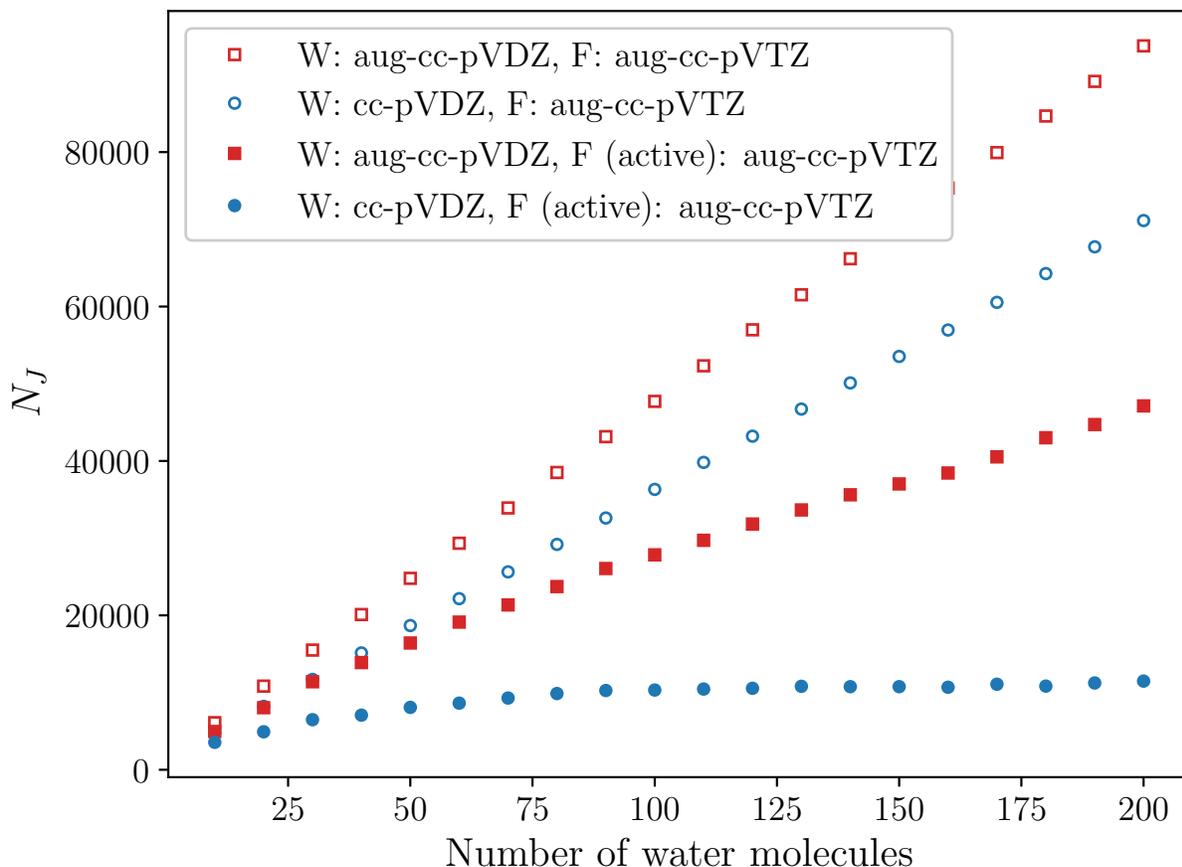


FIG. 4. The number of Cholesky vectors  $N_J$ , in full decomposition and active space decomposition, for formaldehyde (F) surrounded by 10–200 water (W) molecules.

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 improved. We have already performed full decompositions for systems with tens of thousands of atomic orbitals, yet we expect that the partitioned diagonal 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.<sup>36</sup>

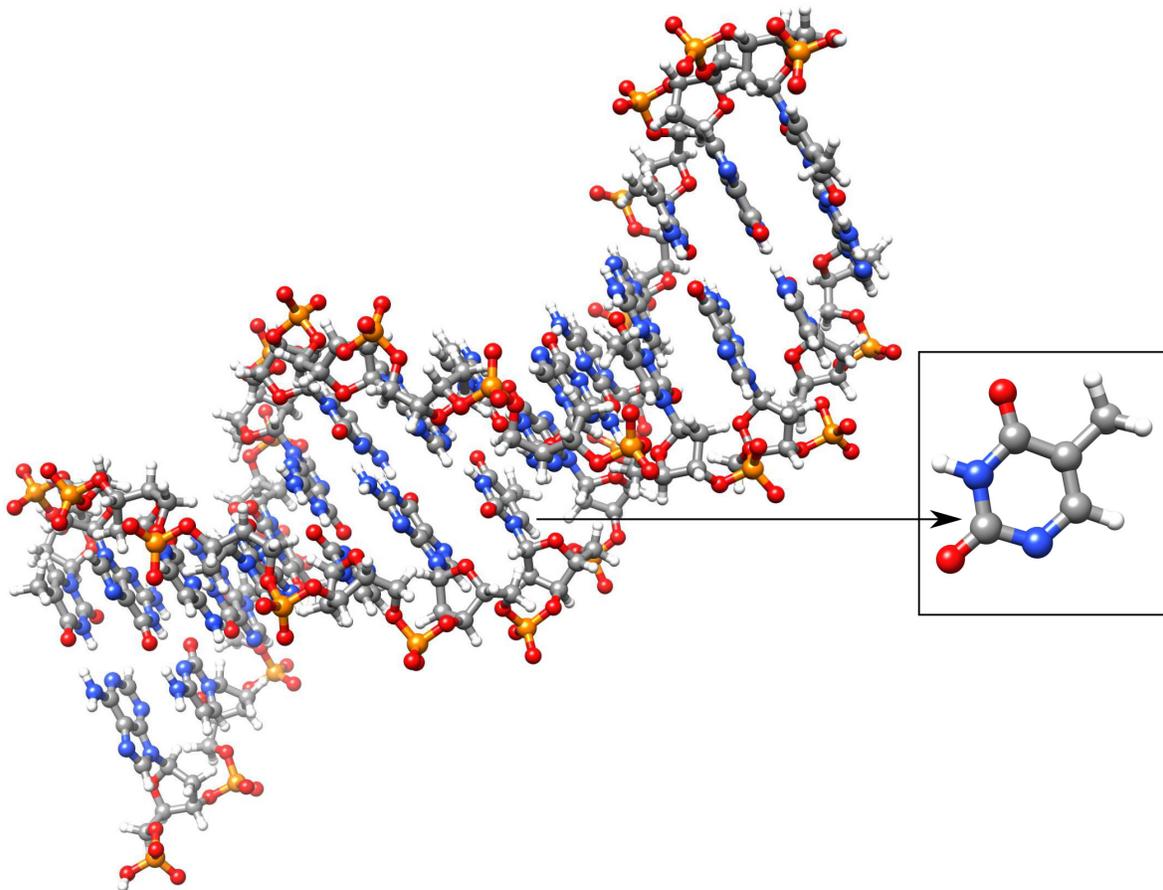


FIG. 5. DNA fragment with active thymine.

## V. ACKNOWLEDGEMENTS

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 and A. Krylov for assistance with QChem.

We acknowledge computer resources from NOTUR through project no. nn2962k and the high performance computer facilities of the SMART Laboratory. Furthermore, we acknowledge funding from the Marie Skłodowska-Curie European Training Network through grant agreement no. 765739 (COSINE) and the Norwegian Research Council through FRINATEK project no. 263110/F20. H. K. acknowledges the Otto Mønsted Fond and S. D. F. acknowledges Fondet til professor Leif Tronstads minne.

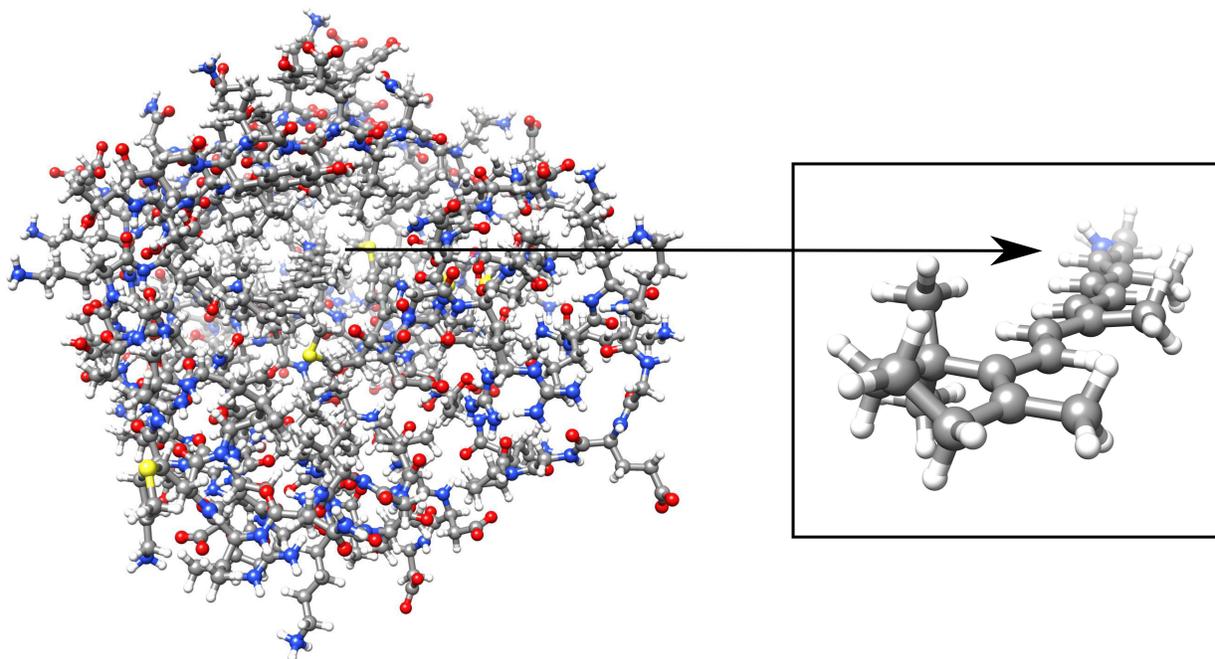


FIG. 6. Retinal bound to rhodopsin with active retinal.

## REFERENCES

- <sup>1</sup>N. H. F. Beebe and J. Linderberg, *Int. J. Quantum Chem.* **12**, 683 (1977).
- <sup>2</sup>P.-O. Löwdin, *Phys. Rev.* **139**, A357 (1965).
- <sup>3</sup>P.-O. Löwdin, *Int. J. Quantum Chem.* **5**, 231 (1971).
- <sup>4</sup>I. Røeggen and E. Wisløff-Nilssen, *Chem. Phys. Lett.* **132**, 154 (1986).
- <sup>5</sup>H. Koch, A. Sánchez de Merás, and T. B. Pedersen, *J. Chem. Phys.* **118**, 9481 (2003).
- <sup>6</sup>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. Pawłowski, 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,

TABLE V. CCSD/aug-cc-pVDZ correlation energies  $E_{\text{corr}}$  for formaldehyde surrounded by ten water molecules using CD and PCD bases as well as prefitted RI. Errors  $\Delta E_{\text{corr}}$  are computed as the deviation from a CCSD calculation in DALTON<sup>6</sup> with no electron repulsion integral approximation. The RI-CCSD calculations were done in Psi4<sup>32</sup> 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	$N_J$	$\tau$	$E_{\text{corr}}[E_h]$	$\Delta E_{\text{corr}}[E_h]$
CD	844	$10^{-2}$	-2.70758713	$1.0 \cdot 10^{-2}$
	1185	$10^{-3}$	-2.71688688	$1.1 \cdot 10^{-3}$
	1920	$10^{-4}$	-2.71796218	$-1.8 \cdot 10^{-5}$
	5374	$10^{-8}$	-2.71794446	$-5.3 \cdot 10^{-8}$
PCD	801	$10^{-2}$	-2.70570268	$1.2 \cdot 10^{-2}$
	1184	$10^{-3}$	-2.71679290	$1.2 \cdot 10^{-3}$
	1889	$10^{-4}$	-2.71796710	$-2.3 \cdot 10^{-5}$
	5333	$10^{-8}$	-2.71794442	$-9.3 \cdot 10^{-9}$
RI-CCSD	1370	–	-2.71930245	$-1.4 \cdot 10^{-3}$
No approximation	–	–	-2.71794441	–

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. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren, Wiley Interdiscip. Rev. Comput. Mol. Sci. **4**, 269 (2013).

<sup>7</sup>F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. 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. Vancollie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, J. Comput. Chem. **37**, 506 (2015).

<sup>8</sup>F. Aquilante, T. B. Pedersen, R. Lindh, B. O. Roos, A. Sánchez de Merás, and H. Koch,

TABLE VI. Lowest CCSD/aug-cc-pVDZ singlet excitation energy  $\omega$  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<sup>6</sup> with no electron repulsion integral approximation. The used Hartree-to-eV conversion factor is the CODATA 2014 value 27.21138602.<sup>33</sup> The threshold  $\tau$  is in atomic units.

Method	$N_J$	$\tau$	$\omega$ [eV]	$\Delta\omega$ [eV]
CD	844	$10^{-2}$	4.74528086	$6.3 \cdot 10^{-3}$
	1185	$10^{-3}$	4.74339531	$4.4 \cdot 10^{-3}$
	1920	$10^{-4}$	4.73926987	$3.1 \cdot 10^{-4}$
	5374	$10^{-8}$	4.73895824	$4.2 \cdot 10^{-8}$
No approximation	–	–	4.73895820	–

J. Chem. Phys. **129**, 024113 (2008).

<sup>9</sup>T. B. Pedersen, A. M. J. Sánchez de Merás, and H. Koch, J. Chem. Phys. **120**, 8887 (2004).

<sup>10</sup>T. D. Crawford, L. S. Owens, M. C. Tam, P. R. Schreiner, and H. Koch, J. Am. Chem. Soc. **127**, 1368 (2005).

<sup>11</sup>F. Aquilante, P.-Å. Malmqvist, T. B. Pedersen, A. Ghosh, and B. O. Roos, J. Chem. Theory Comput. **4**, 694 (2008).

<sup>12</sup>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.

<sup>13</sup>P. Merlot, T. Kjærgaard, H. Trygve, R. Lindh, F. Aquilante, S. Reine, and T. B. Pedersen, J. Comput. Chem. **34**, 1486 (2013).

<sup>14</sup>O. Vahtras, J. Almlöf, and M. W. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).

<sup>15</sup>K. Eichkorn, O. Treutler, H. öhm, M. Häser, and R. Ahlrichs, Chemical Physics Letters **240**, 283 (1995).

<sup>16</sup>K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Theoretical Chemistry Accounts **97**, 119 (1997).

- <sup>17</sup>F. Aquilante, R. Lindh, and T. Bondo Pedersen, *J. Chem. Phys.* **127**, 114107 (2007).
- <sup>18</sup>T. B. Pedersen, F. Aquilante, and R. Lindh, *Theor. Chem. Acc.* **124**, 1 (2009).
- <sup>19</sup>L. Boman, H. Koch, and A. Sánchez de Merás, *J. Chem. Phys.* **129**, 134107 (2008).
- <sup>20</sup>R. H. Myhre, A. M. J. Sánchez de Merás, and H. Koch, *J. Chem. Phys.* **141**, 224105 (2014).
- <sup>21</sup>R. H. Myhre and H. Koch, *J. Chem. Phys.* **145**, 044111 (2016).
- <sup>22</sup>S. Sæther, T. Kjærgaard, H. Koch, and I.-M. Høyvik, *J. Chem. Theory Comput.* **13**, 5282 (2017).
- <sup>23</sup>C. Hättig, *J. Chem. Phys.* **118**, 7751 (2003).
- <sup>24</sup>D. W. O’neal and J. Simons, *Int. J. Quantum Chem.* **36**, 673 (1989).
- <sup>25</sup>F. Aquilante, R. Lindh, and T. B. Pedersen, *J. Chem. Phys.* **129**, 034106 (2008).
- <sup>26</sup>M. Benzi, C. Meyer, and M. Tũma, *SIAM J. Sci. Comput.* **17**, 1135 (1996).
- <sup>27</sup>E. Valeev, “Libint: A library for the evaluation of molecular integrals of many-body operators over gaussian functions,” (2017).
- <sup>28</sup>F. Aquilante, T. Bondo Pedersen, A. Sánchez de Merás, and H. Koch, *J. Chem. Phys.* **125**, 174101 (2006).
- <sup>29</sup>A. M. J. Sánchez de Merás, H. Koch, I. G. Cuesta, and L. Boman, *J. Chem. Phys.* **132**, 204105 (2010).
- <sup>30</sup>J. H. Van Lenthe, R. Zwaans, H. J. J. Van Dam, and M. F. Guest, *J. Comput. Chem.* **27**, 926 (2006).
- <sup>31</sup>Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, *et al.*, *Mol. Phys.* **113**, 184 (2015).
- <sup>32</sup>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 Interdisciplinary Reviews: Computational Molecular Science* **2**, 556 (2012).
- <sup>33</sup>P. J. Mohr, D. B. Newell, and B. N. Taylor, *Journal of Physical and Chemical Reference Data* **45**, 043102 (2016), <https://doi.org/10.1063/1.4954402>.
- <sup>34</sup>T. H. Dunning Jr, *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>35</sup>G. D. Purvis and R. J. Bartlett, *The Journal of Chemical Physics* **76**, 1910 (1982), <https://doi.org/10.1063/1.443164>.
- <sup>36</sup>E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, *J. Chem. Phys.* **137**, 044103 (2012).