

## ARTICLES

## Polarizability and optical rotation calculated from the approximate coupled cluster singles and doubles CC2 linear response theory using Cholesky decompositions

Thomas Bondo Pedersen<sup>a)</sup> and Alfredo M. J. Sánchez de Merás  
*Institute of Molecular Science, Department of Physical Chemistry, University of Valencia,  
E-46100 Burjassot, Valencia, Spain*

Henrik Koch  
*Department of Chemistry, Norwegian University of Science and Technology, N-7491 Trondheim, Norway*

(Received 30 January 2004; accepted 23 February 2004)

A new implementation of the approximate coupled cluster singles and doubles CC2 linear response model using Cholesky decomposition of the two-electron integrals is presented. Significantly reducing storage demands and computational effort without sacrificing accuracy compared to the conventional model, the algorithm is well suited for large-scale applications. Extensive basis set convergence studies are presented for the static and frequency-dependent electric dipole polarizability of benzene and C<sub>60</sub>, and for the optical rotation of CNOFH<sub>2</sub> and (–)-trans-cyclooctene (TCO). The origin-dependence of the optical rotation is calculated and shown to persist for CC2 even at basis set convergence. © 2004 American Institute of Physics.  
[DOI: 10.1063/1.1705575]

### I. INTRODUCTION

In a recent communication,<sup>1</sup> we have demonstrated that significant reductions in computational effort as well as storage requirements are attainable using Cholesky decompositions in electronic structure theory. Specifically, we have shown that the Cholesky representation of the two-electron integral matrix<sup>2</sup> can be immediately exploited for second order Møller–Plesset (MP2) perturbation theory. As mentioned in Ref. 1, Cholesky decomposition of the integral matrix is formally similar to the so-called resolution-of-identity (RI) (Refs. 3 and 4) or density-fitting (DF) (Ref. 5) method. One noticeable difference is that while the Cholesky decomposition yields a general-purpose integral representation of selected precision relative to the exact Hamiltonian (within the chosen basis set), RI methods usually employ a pre-optimized auxiliary basis set aimed at a specific subset of integrals needed for the calculation at hand. For example, the integrals needed for an MP2 energy calculation are  $(ai|bj)$ , where  $i, j$  refer to occupied and  $a, b$  to virtual orbitals in the zeroth order wave function determined from a previous self-consistent field (SCF) computation. The Cholesky integral representation is used for both steps, whereas different auxiliary basis sets (i.e., different Hamiltonians) are used for RI-SCF and RI-MP2.

The CC2 model was introduced by Christiansen, Koch, and Jørgensen<sup>6</sup> as a perturbation-based approximation to the full coupled cluster singles and doubles (CCSD) (Refs. 7 and

8) model with a computational cost similar to that of MP2. The foremost advantage of CC2 over MP2 is the possibility to calculate frequency-dependent molecular response properties, such as dynamic polarizabilities or optical rotations. This additional feature, however, is accompanied by a computational penalty arising from the iterative nature of CC2.

In analogy to canonical MP2, the CC2 doubles amplitudes may be calculated analytically from integrals and a given set of singles amplitudes, while the CC2 singles amplitudes are treated formally exact through the usual coupled cluster equations. Consequently, an algorithm may be formulated in which the CC2 doubles amplitudes need not be permanently stored, neither on disk nor in core, but rather calculated “on-the-fly” and immediately contracted with the necessary integrals to give three-index intermediates to be stored on disk as opposed to the four-index intermediates that must be stored using the integral-direct algorithm.<sup>9</sup> Clearly, in order to make feasible calculations on large systems and basis sets, the “doubles-direct” algorithm, developed for RI-CC2 by Hättig and co-workers,<sup>10–14</sup> is mandatory for a Cholesky-based CC2 implementation.

An obvious utilization of linear response theory is the calculation of electric dipole polarizabilities. Although the correlation level of the CC2 model has been shown to be insufficient for highly accurate calculations of polarizabilities,<sup>15,16</sup> CC2 linear response theory still represents an improvement over uncorrelated methods. Thus, we present in this paper calculations of electric dipole polarizabilities to demonstrate the large-scale applicability of the Cholesky-based algorithm.

Since absolute configuration of a chiral molecule may in

<sup>a)</sup>Present address: Department of Theoretical Chemistry, Chemical Center, University of Lund, P.O. Box 124 S-221 00 Lund, Sweden. Electronic mail: Thomas.Pedersen@teokem.lu.se

principle be determined from knowledge of its optical rotation, recent years have witnessed increased activity in *ab initio* calculations in this field, see, e.g., Ref. 17, and references therein. The major obstacle in *ab initio* calculations of optical rotation has been, and still is, the often intractable size of chiral molecules which, moreover, tend to have little or no point group symmetry to help reduce computational effort as well as storage demands. As electron correlation has been shown to be important for the accurate calculation of optical rotation,<sup>17</sup> we find it natural to test the doubles-direct Cholesky-based CC2 linear response model in this area.

This paper is organized as follows: Section II contains a brief survey of the Cholesky decomposition of two-electron integrals and presents the implementation of CC2 linear response theory on this basis. Sections III and IV present calculations of polarizabilities and optical rotations, respectively, and a summary and concluding remarks are given in Sec. V.

## II. IMPLEMENTATION

In this section we first give a brief outline of the Cholesky integral representation followed by a discussion of the implementation of CC2 linear response theory, mainly focusing on the differences between our Cholesky-based approach and the RI-CC2 method of Hättig and co-workers.<sup>10–14</sup> The code has been included in a development version of the DALTON program,<sup>18</sup> and the overall strategy (apart, of course, from the integral handling) closely follows the one described in Ref. 9.

### A. The Cholesky integral representation

From the positive (semi-)definite nature of the atomic orbital (AO) two-electron integral matrix it follows that a set of Cholesky vectors  $\{L_{\alpha\beta}^J, J=1,2,3,\dots,M\}$  may be calculated such that<sup>1,2</sup>

$$(\alpha\beta|\alpha\beta) - \sum_J (L_{\alpha\beta}^J)^2 \leq \Delta, \quad (1)$$

where  $\alpha$  and  $\beta$ , and later  $\gamma$  and  $\delta$ , denote AOs. The decomposition threshold  $\Delta \geq 0$  may in principle be chosen arbitrarily small, although floating-point arithmetic imposes a lower limit in practice. As demonstrated through SCF and MP2 sample calculations in Ref. 1, the Cholesky integral representation

$$(\alpha\beta|\gamma\delta) \approx \sum_J L_{\alpha\beta}^J L_{\gamma\delta}^J, \quad (2)$$

produces errors in total ground state energies on the order of  $\Delta$ , i.e., the Hamiltonian operator

$$H = \sum_{pq} h_{pq} E_{pq} + \sum_J \sum_{pqrs} L_{pq}^J L_{rs}^J e_{pqrs}, \quad (3)$$

is approximated with an accuracy  $\mathcal{O}(\Delta)$ . In Eq. (3), indices  $p, q, r, s$  refer to an orthonormal set of molecular orbitals (MOs),  $h_{pq}$  are the usual one-electron MO integrals,  $e_{pqrs} = E_{pq} E_{rs} - \delta_{qr} E_{ps}$ , and  $E_{pq} = \sum_{\sigma} a_{p\sigma}^\dagger a_{q\sigma}$  is expressed in terms of creation and annihilation operators (the summation being over spin projection). It should be noted that the inte-

gral accuracy is absolute and, consequently, the relative accuracy shows considerable variation with the numerical value of the exact integral.

In the RI approach, the integrals are approximated in a similar fashion according to

$$(\alpha\beta|\gamma\delta) \approx \sum_P B_{\alpha\beta}^P B_{\gamma\delta}^P, \quad (4)$$

$$B_{\alpha\beta}^P = \sum_Q (\alpha\beta|Q) [V^{-1/2}]_{QP}, \quad (5)$$

where  $P$  and  $Q$  denote auxiliary basis functions, usually preoptimized Gaussian functions centered on the atomic nuclei, and  $V_{QP} = (Q|P)$ . The accuracy of the RI approximation is determined by the ability of the auxiliary basis set to describe the subset of the integrals needed for the calculation at hand. Thus, different auxiliary basis sets are optimized for different wave function models, although we note that auxiliary basis sets optimized for MP2 have been shown to be applicable to the CC2 model as well.<sup>10–14</sup> In contrast, the Cholesky decomposition length  $M \leq N(N+1)/2$ , where  $N$  is the number of basis functions, is implicitly determined by the chosen threshold and as a rule of thumb,<sup>1</sup>  $M \sim 10N$  for  $\Delta = 10^{-8}$ . Furthermore, the Cholesky integrals may be employed in conjunction with any electronic structure model. While the storage requirements clearly depend on  $M$ , the dependence is not necessarily linear, as the sparsity of the Cholesky vectors may be utilized to further minimize storage as well as I/O. Moreover,  $M$  affects the computational cost in a twofold manner. First,  $M$  equals the minimum number of columns of the integral matrix that need to be calculated during the decomposition, although the number actually calculated usually is somewhat larger, as the integrals are calculated in shell-pair distributions.<sup>1</sup> Second, if sparsity is not employed,  $M$  directly determines the operation count for a subsequent electronic structure calculation.

An integral-direct Cholesky decomposition algorithm is presented in Ref. 1 along with applications to SCF and MP2 calculations.

### B. Cholesky-based CC2 linear response theory

The CC2 model was introduced by Christiansen, Koch, and Jørgensen<sup>6</sup> as an approximation to the full CCSD model based on perturbation theory and with a computational cost similar to MP2. An integral-direct implementation of CC2 linear response theory is described in Ref. 9 where the model is treated as a special case of the CCSD algorithm. Memory and disk requirements are consequently the same for integral-direct CC2 and CCSD, thus severely limiting the size of the systems and basis sets that can be treated.

Recently, Hättig and co-workers<sup>10–14</sup> have presented an alternative algorithm in which the doubles amplitudes are assembled on-the-fly from integrals and singles amplitudes, and their contributions to key intermediates are immediately calculated. While a doubles- and integral-direct algorithm formally is feasible, an efficient implementation would still require storage (in core or on disk) of half-transformed integrals roughly the same size as the doubles amplitudes them-

selves. Therefore, in order not to trade efficiency for storage demands, an alternative integral representation is mandatory. Comparing Eqs. (2) and (4), we see that the Cholesky vectors may straightforwardly be used in the doubles–direct algorithm of Hättig *et al.*<sup>10–14</sup> Thus, we refer to Refs. 10 and 12 for the partitioning of the CC2 equations and for formulas, and focus instead on some implementational aspects.

Apart from the prerequisite Cholesky decomposition and SCF optimization, a CC2 linear response calculation requires the solution of three sets of equations for the zeroth order amplitudes, for the zeroth order Lagrangian multipliers, and for the first order response of the amplitudes to an external (one-electron) perturbation of frequency  $\omega$ . Analytical expressions for the doubles parts of these amplitudes can be derived as

$$t_{ij}^{ab} = \text{it}_{ij}^{ab} = -\frac{(ai|\tilde{b}j)}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j}, \quad (6)$$

$$\begin{aligned} \bar{t}_{ij}^{ab} = \text{it}_{ij}^{\bar{a}b} + \text{nit}_{ij}^{\bar{a}b} = & -(2 - P_{ij}) \frac{(ia|\hat{j}b) + P_{ij}^{ab}[\bar{t}_{ai}\bar{F}_{jb}]}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j} \\ & - 2(2 - P_{ij}) \frac{(ia|jb)}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j}, \end{aligned} \quad (7)$$

$$\begin{aligned} {}^B t_{ij}^{ab}(\omega) = \text{it}_{ij}^{B,ab} + \text{nit}_{ij}^{B,ab} \\ = -\frac{(ai|\tilde{b}j)}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j - \omega} \\ - P_{ij}^{ab} \frac{\sum_c t_{ij}^{ac} \tilde{B}_{bc} - \sum_k t_{ik}^{ab} \tilde{B}_{kj}}{\epsilon_a - \epsilon_i + \epsilon_b - \epsilon_j - \omega}, \end{aligned} \quad (8)$$

for zeroth order amplitudes, multipliers, and first order amplitude response, respectively. Indices  $a, b, c$  and  $i, j, k$  are used to denote virtual and occupied MOs,  $\epsilon_p$  is the canonical SCF orbital energy of orbital  $p$ ,  $P_{ij} M_{ij}^{ab} = M_{ji}^{ab}$ , and  $P_{ij}^{ab} M_{ij}^{ab} = M_{ij}^{ab} + M_{ji}^{ba}$ . The various modified MO integrals are defined in terms of Cholesky vectors as

$$\tilde{F}_{ia} = \sum_{ck} (2(ia|kc) - (ka|ic)) t_{ck}, \quad (9)$$

$$(ia|jb) = (ia|\tilde{j}b) = \sum_J L_{ia}^J L_{jb}^J, \quad (10)$$

$$(ai|\tilde{b}j) = \sum_J \tilde{L}_{ai}^J \tilde{L}_{bj}^J, \quad (11)$$

$$(ia|\hat{j}b) = P_{ij}^{ab} \sum_J L_{ia}^J \hat{L}_{jb}^J, \quad (12)$$

$$(ai|\bar{b}j) = P_{ij}^{ab} \sum_J \tilde{L}_{ai}^J \tilde{L}_{bj}^J, \quad (13)$$

where

$$L_{ia}^J = \sum_{\alpha\beta} C_{\alpha i} C_{\beta a} L_{\alpha\beta}^J, \quad (14)$$

$$\tilde{L}_{ai}^J = \sum_{\alpha\beta} \Lambda_{\alpha a}^p \Lambda_{\beta i}^h L_{\alpha\beta}^J, \quad (15)$$

$$\hat{L}_{ia}^J = \sum_{\alpha\beta} (\hat{\Lambda}_{\alpha i}^p \Lambda_{\beta a}^h + \Lambda_{\alpha i}^p \hat{\Lambda}_{\beta a}^h) L_{\alpha\beta}^J, \quad (16)$$

$$\bar{L}_{ai}^J = \sum_{\alpha\beta} (\bar{\Lambda}_{\alpha a}^p \Lambda_{\beta i}^h + \Lambda_{\alpha a}^p \bar{\Lambda}_{\beta i}^h) L_{\alpha\beta}^J, \quad (17)$$

with similar definitions for the integrals of the one-electron operator  $B$  representing an external perturbation, such as electric or magnetic multipole operators. The “particle” and “hole” transformation matrices are given in terms of the SCF MO coefficient matrix,  $C$ , and the singles amplitudes and multipliers as

$$\Lambda^p = C(1 - \mathbf{t}_1^T), \quad (18)$$

$$\Lambda^h = C(1 + \mathbf{t}_1), \quad (19)$$

$$\hat{\Lambda}^p = \Lambda^p \bar{\mathbf{t}}_1, \quad (20)$$

$$\hat{\Lambda}^h = -\Lambda^h \bar{\mathbf{t}}_1^T, \quad (21)$$

$$\bar{\Lambda}^p = -\Lambda^p {}^B \mathbf{t}_1^T(\omega), \quad (22)$$

$$\bar{\Lambda}^h = \Lambda^h {}^B \mathbf{t}_1(\omega). \quad (23)$$

In Eqs. (6)–(8) the amplitudes and multipliers are written as a sum of iterative (“it”) and noniterative (“nit”) parts, the latter entering the right-hand side of the corresponding effective (partitioned) problem [for the general form of the effective linear equations for determining the multipliers, see Eqs. (31) and (32) of Ref. 12; similar expressions exist for the linear response equations for the amplitudes].

The vast majority of the CPU time of a Cholesky-based CC2 linear response calculation is spent calculating a particular class of intermediates, defined through an expression of the form

$$Y_{ai}^J = \sum_{bj} X_{ij}^{ab} Z_{bj}^J, \quad (24)$$

where  $X_{ij}^{ab}$ , in turn, is defined in terms of the relevant iterative doubles vector of Eqs. (6)–(8). For the calculation of the CC2 singles vector function,  $X_{ij}^{ab} = (2 - P_{ij}) \text{it}_{ij}^{ab}$  and  $Z_{bj}^J = L_{jb}^J$ , for left transformations with the effective Jacobian,  $X_{ij}^{ab} = \text{it}_{ij}^{\bar{a}b}$  and  $Z_{bj}^J = \tilde{L}_{bj}^J$ , and finally for right transformations with the effective Jacobian,  $X_{ij}^{ab} = (2 - P_{ij}) \text{it}_{ij}^{B,ab}(\omega)$  and  $Z_{bj}^J = L_{jb}^J$ . Assuming that the relevant MO Cholesky vectors are available on disk, the calculation of these  $Y$ -intermediates proceeds in two steps scaling as  $O^2 V^2 M \sim O^2 N^3$  in the following manner:

- Loop  $\mathcal{A}$  (subset of virtual orbitals).
  - Loop  $\mathcal{B}$  (subset of virtual orbitals).
    - \* Calculate  $X_{ij}^{ab}$  for  $a \in \mathcal{A}$  and  $b \in \mathcal{B}$ , contracting as many MO Cholesky vectors according to Eq. (6), (7), or (8) as can be stored in remaining memory.
  - \* Loop  $\mathcal{J}$  (subset of Cholesky vectors).
    - Read  $Y_{ai}^J$  for  $a \in \mathcal{A}$  and  $J \in \mathcal{J}$ .
    - Read  $Z_{bj}^J$  for  $b \in \mathcal{B}$  and  $J \in \mathcal{J}$ .

- Calculate contribution to  $Y_{ai}^J$  according to Eq. (24) for  $J \in \mathcal{J}$ ,  $b \in \mathcal{B}$ , and  $a \in \mathcal{A}$ .
- Store updated  $Y_{ai}^J$  for  $a \in \mathcal{A}$  and  $J \in \mathcal{J}$  on disk.
- \* End loop  $\mathcal{J}$ .
- End loop  $\mathcal{B}$ .
- End loop  $\mathcal{A}$ .

As indicated, main memory is split in two parts, one for the amplitudes and another one for MO Cholesky vectors in the first step and  $Y$ - and  $Z$ -vectors in the second step. The split is set up dynamically, depending on the relative sizes of amplitudes and vectors, thereby ensuring that even very large calculations may be carried out with minimum memory requirements. On the other hand, I/O operations on the various vectors in the two steps decrease as available memory is increased. In almost any case, more than 90% of the total time is spent calculating matrix–matrix multiplications.

The ground state amplitudes, Eq. (6), are constructed from the positive definite integrals ( $ai|bj$ ) and, consequently,  $t_{ij}^{ab}$  form a negative definite matrix. Hence, as has been demonstrated for the MP2 energy,<sup>1</sup> we may employ Cholesky decomposition of either the ( $ai|bj$ ) integral matrix or of the amplitudes themselves to reduce the operation count of the first step in the calculation of the ground state  $Y$ -intermediates. Specifically, using either

$$(ai|bj) \approx \sum_{J=1}^{M'} N_{ai}^J N_{bj}^J, \quad (25)$$

or

$$t_{ij}^{ab} \approx - \sum_{J=1}^{M''} N_{ai}^J N_{bj}^J, \quad (26)$$

the scaling of the first step in the  $Y$ -intermediate calculation is reduced, provided that the number of  $N_{ai}^J$  vectors is less than the number of original vectors  $M$ . As for MP2,<sup>1</sup>  $M' \ll \min[OV, M]$  and so Eq. (25) will certainly reduce the computational cost of calculating the amplitudes when the cost of the decomposition itself is negligible. The number of vectors  $M''$  needed to decompose the amplitudes is also bounded by  $OV$ , but may, on account of the orbital energy denominator, in principle be larger than  $M$ . In practice we have found that  $M'' > M'$ , making Eq. (25) preferable to Eq. (26). However, the advantage is far less pronounced than in the MP2 case, as the second step of the  $Y$ -intermediate calculation must be done for the full original set of Cholesky vectors in any case. Instead, once the equations have been solved for the ground state amplitudes, Eq. (26) may be employed once and for all to obtain a compact representation, useful for response calculations. By default, we use the same threshold for the amplitude decomposition as for the integral decomposition. For sufficiently small thresholds, this default value does not significantly reduce the accuracy of the final linear response function. As we shall see below, care must be exercised for larger thresholds, as the errors in the amplitude representation may in some cases lead to significant errors in the calculated response properties. The reason for this behavior is straightforward: the numerical value of the diagonal elements of the matrix formed by the ground state amplitudes is

often in the neighborhood of  $10^{-3}$  and smaller, implying that a too large amount of the amplitudes becomes identically zero in the Cholesky representation with a decomposition threshold of, say,  $\Delta = 10^{-4}$ .

In the integral-direct approach,<sup>9</sup> evaluation of the linear response function from known amplitudes and their  $F$ -matrix transformed counterparts is a straightforward series of dot-products. In the doubles-direct approach, however, the linear response function must be evaluated with doubles amplitudes assembled on-the-fly from the expression

$$\langle\langle A, B \rangle\rangle_{\omega} = \frac{1}{2} (\mathcal{F}(A, B, \omega) + \mathcal{F}(A^{\dagger}, B^{\dagger}, -\omega)^*), \quad (27)$$

$$\begin{aligned} \mathcal{F}(A, B, \omega) = & \sum_{pq} D_{pq}^{\eta} [{}^A \mathbf{t}(-\omega)] \tilde{B}_{pq} \\ & + \sum_{pq} D_{pq}^{\eta} [{}^B \mathbf{t}(\omega)] \tilde{A}_{pq} + \sum_{ai} {}^A t_{ai}(-\omega) \\ & \times \left( \frac{1}{2} \sum_{bj} F_{ai,bj} {}^B t_{bj}(\omega) \right. \\ & + \sum_{bjck} F_{ai,bjck} {}^B t_{jk}^{bc}(\omega) + \sum_{ai} {}^B t_{ai}(\omega) \\ & \times \left. \left( \frac{1}{2} \sum_{bj} F_{ai,bj} {}^A t_{bj}(-\omega) \right. \right. \\ & \left. \left. + \sum_{bjck} F_{ai,bjck} {}^A t_{jk}^{bc}(-\omega) \right) \right), \quad (28) \end{aligned}$$

for one-electron operators  $A$  and  $B$ . The modified MO integrals  $\tilde{A}_{pq}$  and  $\tilde{B}_{pq}$  are defined from the AO integrals of the  $A$  and  $B$  operators in analogy to the definition of modified MO Cholesky vectors in Eq. (15). Immediately after solving the equations for the first order amplitudes, the  $F$ -matrix transformations are carried out along lines similar to the Jacobian transformations,<sup>12</sup> thus involving calculation of  $Y$ -intermediates, Eq. (24), from full amplitudes (iterative as well as noniterative parts), and the resulting singles vectors are stored on disk. The partial first order density matrices are calculated from first order amplitudes according to

$$D_{ij}^{\eta} [{}^B \mathbf{t}(\omega)] = - \sum_a {}^B t_{bi}(\omega) D_{bj}^{\Lambda} - \sum_{ack} {}^B t_{ik}^{ac} \bar{t}_{jk}^{ac}, \quad (29)$$

$$\begin{aligned} D_{ia}^{\eta} [{}^B \mathbf{t}(\omega)] = & \sum_j {}^B t_{aj}(\omega) D_{ij}^{\Lambda} - \sum_b D_{ba}^{\Lambda} {}^B t_{bi}(\omega) \\ & + \sum_{bj} {}^B t_{ij}^{ab} D_{bj}^{\Lambda}, \quad (30) \end{aligned}$$

$$D_{ai}^{\eta} [{}^B \mathbf{t}(\omega)] = 0, \quad (31)$$

$$D_{ab}^{\eta} [{}^B \mathbf{t}(\omega)] = \sum_i D_{ai}^{\Lambda} {}^B t_{bi}(\omega) + \sum_{ijc} \bar{t}_{ij}^{ac} {}^B t_{ij}^{bc}, \quad (32)$$

where  $\mathbf{D}^{\Lambda}$  is the CC2 ground state density matrix calculated from multipliers and Cholesky decomposed zeroth order amplitudes. Obviously, the most expensive terms in the calculation of the partial first order densities are those that require

TABLE I. The  $zz$ -component of the static polarizability (in a.u.) calculated from frozen core CC2 linear response theory for benzene using conventional and Cholesky algorithms. For Cholesky, the errors  $(\alpha_{zz}[\text{Cholesky}] - \alpha_{zz}[\text{Conventional}])/\Delta$  are given. Total CPU timings (in minutes) are also given for the coupled cluster calculations, i.e., excluding integral evaluations/decompositions and SCF optimizations.  $N$  is the number of basis functions.

Basis set	$N$	Conventional		$\Delta = 10^{-4}$		$\Delta = 10^{-6}$		$\Delta = 10^{-8}$	
		$\alpha_{zz}$	Time	Error	Time	Error	Time	Error	Time
6-311++G(2d,2p)	222	43.978 998 82	52.0	-16.5	2.8	-19.8	4.3	-20.1	6.0
6-311++G(3df,3pd)	342	45.218 442 97	347.8	-25.0	8.6	-25.1	12.7	-14.0	17.8
aug-cc-pVDZ	192	46.014 923 45	26.5	-23.1	2.1	-6.8	3.0	20.0	4.0
aug-cc-pVTZ	414	45.652 888 91	604.9	-37.0	11.1	-2.7	16.4	-5.9	27.5

assembly of doubles according to Eqs. (7) and (8) using a loop structure similar to that of the  $Y$ -intermediates.

With the AO Cholesky vectors available on disk, the additional disk space requirements for a CC2 calculation based on Cholesky decomposed integrals is thus  $\sim OVM$ , which is substantially smaller than the  $\sim O^2V^2$  requirement of the integral-direct approach, especially for large molecules (large number of occupieds,  $O$ ). Whereas the memory requirements for Cholesky-based CC2 and RI-CC2 are identical, the latter requires less disk space, as the three-center integrals of Eq. (5) are never stored on disk but recalculated as needed.<sup>10</sup>

While crucial for formulating the doubles-direct algorithm, the partitioning of the CC2 Jacobian<sup>10-14</sup> leads to a frequency-dependent effective Jacobian (as well as right-hand sides) and, thus, has the side-effect that the response equations must be solved separately for each perturbation frequency. Therefore, the total number of linear transformations (effective Jacobian times trial vector) is usually larger for the doubles-direct implementation than for the conventional one. On the other hand, the partitioning also implies that the doubles parts of the equations are in fact explicitly solved, which tends to decrease the number of iterations needed for convergence and thus decrease the total number of linear transformations. The calculation of static properties (zero frequency) therefore generally requires fewer linear transformations for the doubles-direct implementation, whereas calculation of dynamic properties (nonzero frequencies) requires more. The reason for this somewhat peculiar behavior is that the Hermitian symmetrization of the linear response function in Eq. (27) forces us to solve the linear response equations not only at the perturbation frequencies of interest but also at minus those frequencies. Thus, in general, the doubles-direct algorithm requires solution of twice the number of perturbation frequencies distinct blocks of equations, whereas a single block in principle suffices in the conventional approach.

### III. POLARIZABILITY

The Cartesian  $jk$  component of the electric dipole polarizability tensor is related to the linear response function according to

$$\alpha_{jk}(\omega) = -\langle\langle \mu_j, \mu_k \rangle\rangle_{\omega}, \quad (33)$$

where  $\mu$  is the electric dipole operator. We report the polarizability in atomic units, noting that the conversion factor to

SI units is  $e^2 a_0^2 / E_h \approx 1.648\,778 \times 10^{-24} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$ , while that to volume polarizability is  $a_0^3 \approx 0.148\,184\,7 \text{ \AA}^3$ . To evaluate the CC2 polarizability at a given nonzero frequency, we must solve six equations for the first order amplitude responses: two for each Cartesian component of the electric dipole operator (one at  $\omega$  and one at  $-\omega$ ). In the static case ( $\omega = 0$ ), only three equations need to be solved. The worst-case scenario within the doubles-direct algorithm (i.e., solving the partitioned response problem) occur when the three components of the electric dipole operator transform according to different irreducible representations of the molecular point group. In this case, we are forced to solve each of the six (times the number of frequencies) amplitude response equations separately, thereby losing the advantages of spanning all response amplitudes in one common set of trial vectors.

Polarizability calculations are performed for the benzene and  $C_{60}$  molecules. For benzene we employ the following geometry:  $r_{CC} = 1.397 \text{ \AA}$ ,  $r_{CH} = 1.080 \text{ \AA}$ , and  $\angle(\text{CCC}) = \angle(\text{CCH}) = 120^\circ$ , placing the molecule in the  $xy$ -plane with the  $z$ -axis through the center of mass. For the fullerene, we use the same geometry as Ruud, Jonsson, and Taylor.<sup>19,20</sup>

Table I presents results obtained with rather modest basis sets for benzene using conventional as well as Cholesky decomposition-based doubles-direct algorithms. In order to ensure that discrepancies between the two algorithms are due solely to the Cholesky decomposition of the two-electron integrals, we have imposed very tight convergence thresholds (i.e.,  $10^{-10}$  for errors in MOs,  $10^{-10}$  for the CC2 energy, and  $10^{-8}$  for the residual norm for the CC2 response equations). All calculations were carried out on an IBM/RS6000 workstation equipped with a PowerIII processor. For the conventional calculations, the AO integrals were stored on disk and thus avoiding expensive recalculations. Nonetheless, the saving in computational time is substantial, up to a factor of 54 for the aug-cc-pVTZ basis set for  $\Delta = 10^{-4}$ , and even more pronounced savings can of course be anticipated for calculations requiring integral-direct techniques in the conventional scheme. We observe that the errors stemming from Cholesky decomposition vary systematically and, as expected, the error approaches zero for low decomposition thresholds. Note, however, that the error in general is one order of magnitude larger than  $\Delta$ , which can be traced back to the decomposition of the ground state amplitudes. As an example, decreasing the threshold for amplitude decomposition to  $10^{-10}$  while keeping  $\Delta = 10^{-4}$  with the aug-cc-pVTZ basis set reduces the error to  $-2.2 \times 10^{-4}$  a.u. We point out

TABLE II. The  $xx$ - and  $zz$ -components of the static polarizability (in a.u.) calculated from frozen core CC2 linear response theory for benzene using Cholesky algorithm.  $N$  is the number of basis functions.

Basis set	$N$	$\alpha_{xx}$	$\alpha_{zz}$
3-21G	66	67.95	14.70
4-31G	66	69.68	16.64
6-31G	66	70.87	18.40
6-31G*	96	71.89	20.33
6-31G**	114	72.40	20.76
6-31G(3df,3pd)	282	79.85	37.82
6-311G	96	74.04	25.19
6-311G(2df,2pd)	264	78.29	31.89
3-21++G	96	76.95	39.72
6-31++G	96	78.32	40.97
6-31++G*	126	79.19	41.29
6-31++G**	144	79.56	41.48
6-311++G(2d,2p)	222	82.34	43.98
6-311++G(3df,3pd)	342	83.79	45.22
aug-cc-pVDZ	192	84.46	46.01
aug-cc-pVTZ	414	84.51	45.65
aug-cc-pVQZ	756	84.36	45.37
aug-cc-pV5Z	1242	84.28	45.28
aug-cc-pV6Z	1896	84.24	45.22
aug-cc-pV7Z	2742	84.22	45.21

that for practical purposes, a decomposition threshold of  $\Delta = 10^{-6}$  should be sufficient to obtain essentially exact results.

A basis set convergence study of the static polarizability components for benzene is given in Table II using Pople- and Dunning-style basis sets and  $\Delta = 10^{-6}$ . These calculations were performed with default thresholds for convergence of SCF and CC2 wave function parameters (that is,  $10^{-8}$  for MOs in the SCF procedure,  $10^{-8}$  for the CC2 energy, and  $10^{-5}$  for the response equations). Both components of the polarizability show remarkable stability with respect to basis set, provided that diffuse functions are included. In particular, the aug-cc-pVXZ series is essentially converged already at the double-zeta level. The largest calculation, i.e., aug-cc-pV7Z, involves 100 million doubles amplitudes and would require more than 13 Gb of memory using the conventional integral-direct algorithm. The average of the polarizability tensor is 71.22 a.u. at the aug-cc-pV7Z level, in reasonable agreement with experiment (70.05 a.u. by Proutière and Camail,<sup>21</sup> 67.48 a.u. by Alms *et al.*,<sup>22</sup> and 68.7 a.u. by Okruss *et al.*<sup>23</sup>). Numerous theoretical calculations of the polarizability of benzene are available in the literature, see, e.g., Refs. 24–33. Most of these use the SCF method with results ranging from approximately 64 a.u. to 68 a.u. for the average polarizability. Among the correlated calculations, Christiansen, Hättig, and Jørgensen<sup>31</sup> obtained an average polarizability of 69.17 a.u. at the CCSD level with Sadlej's basis set, while Soscún *et al.*<sup>33</sup> reported 67.25 a.u. using DFT/B3LYP and a modified 6-31+G( $d,p$ ) basis set. Thus, the CC2 method seems to overestimate the correlation effects on the average polarizability of benzene. Recently, we have carried out a detailed study of the static and frequency-dependent polarizability of benzene and a few small  $[4n+2]$ -annulenes.<sup>34</sup>

TABLE III. The average polarizability (in a.u.) calculated from frozen core CC2 linear response theory for  $C_{60}$  using Cholesky algorithm.  $N$  is the number of basis functions.

Basis set	$N$	$\alpha_{ave}$	
		$\lambda = \infty$	$\lambda = 1064$ nm
6-31G	540	495.94	506.88
6-31++G	780	586.75	600.77
6-31++G*	1080	606.80	622.59
cc-pVDZ	840	541.46	554.46
aug-cc-pVDZ <sup>a</sup>	1380	623.70	640.15

<sup>a</sup>All electrons correlated.

Table III reports a large-scale application of the Cholesky CC2 linear response formalism to the calculation of static and dynamic polarizabilities for the  $C_{60}$  molecule using decomposition threshold  $\Delta = 10^{-6}$  and the following wave function convergence thresholds:  $5 \times 10^{-7}$  for MOs (SCF) and the CC2 energy, and  $5 \times 10^{-5}$  for the CC2 response equations. With the aug-cc-pVDZ basis set, the CC2 wave function contains 2.8 billion doubles amplitudes and an integral-direct calculation would require approximately 113 Gb of memory. As for benzene, we find that diffuse functions are essential for accurately calculating the polarizability. Note, however, that the percentwise changes going from 6-31++G or 6-31++G\* to aug-cc-pVDZ are in fact smaller for  $C_{60}$  than for benzene. Hence, the assumption used by Ruud *et al.*<sup>19</sup> (for SCF calculations of the polarizability) that 6-31++G constitutes a reasonable compromise between computational effort and quality of the results seems applicable to CC2 calculations as well. Comparing to the SCF results of Ruud *et al.*<sup>19</sup> and to the SCF and DFT results of Xie *et al.*,<sup>35</sup> we find that the CC2 polarizability is somewhat larger. This may to some extent be explained by the use of more diffuse basis sets in the present study. At the same time, the rather low level of correlation of the CC2 approach opens for possible reductions using full CCSD or better. Such reductions are to be expected, as the experimental polarizabilities are  $516 \pm 54$  a.u. ( $76.5 \pm 8 \text{ \AA}^3$ ) (Ref. 36) at  $\lambda = \infty$  and  $533 \pm 27$  a.u. ( $79 \pm 4 \text{ \AA}^3$ ) (Ref. 37) at  $\lambda = 1064$  nm. Note, however, that the dispersion is reasonably well described at the CC2 level, the polarizability increasing approximately 3% going from the static limit to 1064 nm.

## IV. OPTICAL ROTATION

### A. Theoretical background

The specific rotation, in units of  $\text{deg} [\text{dm g/cm}^3]^{-1}$ , of a chiral molecule is defined as<sup>38</sup>

$$[\alpha] = 28800 \pi^2 N_A a_0^4 \frac{\bar{\nu}^2 \beta'(\omega)}{M}, \quad (34)$$

where  $N_A$  is Avogadro's number,  $a_0$  is the bohr radius in cm,  $\bar{\nu}$  is the frequency of the light in  $\text{cm}^{-1}$  ( $\omega$  is the frequency in atomic units),  $M$  is the molecular weight in g/mol, and

$$\beta'(\omega) = -\text{Im} \frac{\text{Tr}(\langle \mu, \mathbf{m} \rangle)_\omega}{3\omega}, \quad (35)$$

TABLE IV. All-electron CC2 specific rotation ( $[\alpha]_D$ , in deg  $[\text{dm g/cm}^3]^{-1}$ ) at the sodium D line in length gauge (origin at the center of mass) for CNOFH<sub>2</sub> using conventional and Cholesky algorithms. For Cholesky, the errors,  $[\alpha]_D(\text{Cholesky}) - [\alpha]_D(\text{Conventional})$ , are given. Total CPU timings (in hours) are also given for the coupled cluster calculations, i.e., excluding integral evaluations/decompositions and SCF optimizations.  $N$  is the number of basis functions.

Basis set	$N$	Conventional		$\Delta = 10^{-4}$		$\Delta = 10^{-6}$		$\Delta = 10^{-8}$	
		$[\alpha]_D$	Time	Error	Time	Error	Time	Error	Time
cc-pVDZ	66	44.684 230	1.2	-0.093 165	0.1	0.000 816	0.2	0.000 005	0.3
cc-pVTZ	148	68.968 184	21.9	0.026 980	1.4	-0.000 357	2.3	-0.000 005	3.7
aug-cc-pVDZ	110	73.557 159	8.0	-0.156 513	0.5	-0.000 608	0.9	0.000 003	1.3
aug-cc-pVTZ	230	77.270 983	113.4	-0.304 409	5.6	-0.000 175	8.4	0.000 005	13.0

in atomic units ( $a_0^4$ ). Following tradition, we have employed the length gauge in Eq. (35), and the trace (Tr) is over Cartesian elements of the electric dipole-magnetic dipole polarizability tensor ( $\mu$  is the electric dipole operator,  $\mathbf{m}$  the magnetic dipole operator). Evaluation of the specific rotation thus requires exactly twice as much effort as for the frequency-dependent electric dipole polarizability, that is, twelve first order amplitude response equations for each requested frequency, making the calculation of optical rotation dispersion a rather costly affair. On the other hand, specific rotations are often reported at the sodium D line (i.e., at 589.3 nm, denoted  $[\alpha]_D$ ) so that calculations need be done for only one frequency.

Origin dependence is a recurring problem in coupled cluster calculations of response properties for perturbations that involve a time-dependent magnetic field. As discussed on several other occasions,<sup>17,39–41</sup> London (or gauge including) orbitals can not be used to ensure origin invariance in standard coupled cluster theory in these cases, although the problem may be solved by enforcing gauge invariance through optimizing orbitals alongside the double and higher order excitation cluster amplitudes (the NOCC model).<sup>42</sup> Under a translation of the origin along the vector  $\mathbf{a}$ , i.e.,  $\mathbf{r} \rightarrow \mathbf{r} - \mathbf{a}$ , the optical rotation transforms according to

$$[\alpha] \rightarrow [\alpha] + \mathbf{a} \cdot \Delta^{\text{OR}}. \quad (36)$$

The component  $\Delta_j^{\text{OR}}$ , i.e., the change in specific rotation per unit length along the  $j$ th Cartesian direction, can be derived as

$$\Delta_j^{\text{OR}} = -28800\pi^2 N_A a_0^4 \frac{\tilde{v}^2}{M} \text{Im} \frac{\sum_{k,l=1}^3 \epsilon_{jkl} \langle\langle r_k \cdot p_l \rangle\rangle_{\omega}}{6\omega}, \quad (37)$$

where  $\epsilon$  is the Levi-Civita tensor,  $\mathbf{r}$  is the position operator, and  $\mathbf{p}$  the momentum operator. While  $\Delta^{\text{OR}}$  vanishes for variational methods in the limit of basis set completeness, it is nonzero for truncated coupled cluster models, including CC2 and CCSD, even with a complete orbital basis. Thus, the length gauge specific rotation calculated from CC2 linear response theory depends on the chosen (*mathematical*) origin.

Ruud *et al.*<sup>17</sup> calculate specific rotations with the origin at the molecular center of mass. Although certainly a reasonable choice, inconsistencies would arise when calculating vibrational effects for different isotopomers, as  $\beta'(\omega)$  values would be different.

In the following sections, we report sample calculations of  $[\alpha]_D$  for two molecules, CNOFH<sub>2</sub> and (-)-trans-cyclooctene (TCO).

## B. Optical rotation of CNOFH<sub>2</sub>

Following Ruud *et al.*,<sup>17</sup> we employ CNOFH<sub>2</sub>, derived from 2-F-oxirane by substituting the ring CH<sub>2</sub> with NH, as test system for probing the basis set convergence of the specific rotation. We use the DFT/B3LYP/6-31G\* C<sub>1</sub> geometry of Ref. 17.

In Table IV we have collected some results for the optical rotation using conventional as well as Cholesky CC2 linear response theory with the same tight convergence thresholds as for the case of benzene above. All calculations were performed on the same computer as used for benzene. The basis sets are sufficiently small that the conventional calculations may be carried out with integrals on disk and, again, the Cholesky algorithm is nevertheless significantly faster than the conventional one. The errors in the specific rotation caused by the approximate Cholesky integral representation are clearly controllable, decreasing systematically with the decomposition threshold. Note that the errors measured in a.u. [i.e., the errors in  $\beta'(\omega)$ ] are on the order of  $\Delta$ , except for the  $\Delta = 10^{-4}$  results for the augmented basis sets where the error is roughly an order of magnitude larger. Again, this increased error can be attributed to the decomposition of the ground state amplitudes. For example, using a threshold of  $10^{-10}$  for the amplitude decomposition while keeping  $\Delta = 10^{-4}$ , the error in the specific rotation becomes  $-0.015 883 \text{ deg} [\text{dm g/cm}^3]^{-1}$  with the aug-cc-pVTZ basis set. In all, as for the benzene static polarizability, we thus find that  $\Delta = 10^{-6}$  suffices for essentially exact results.

Taking advantage of the open-ended nature of the doubles-direct algorithm, we perform an extensive basis set convergence study of the specific rotation and its origin-dependence at the CC2 level. The results are presented in Table V with all electrons correlated and in Table VI with frozen core approximation using  $\Delta = 10^{-6}$  and default convergence thresholds for the wave function parameters (cf. the benzene polarizability calculations presented above). The largest calculation (all-electron CC2 with the aug-cc-pV6Z basis set) incorporates nearly 126 million doubles amplitudes and would require more than 8.5 Gb of memory for the conventional algorithm.

As expected,<sup>17</sup> the all-electron results in Table V show

TABLE V. All-electron CC2 specific rotation at the sodium D line ( $[\alpha]_D$ , in deg  $[\text{dm g/cm}^3]^{-1}$ ) in length gauge (origin at the center of mass) for CNOFH<sub>2</sub> using Cholesky algorithm. The Cartesian origin-dependence vector (in deg  $[\text{dm g/cm}^3]^{-1} \text{ bohr}^{-1}$ ) is also reported.  $N$  is the number of basis functions.

Basis set	$N$	$[\alpha]_D$	$\Delta_x^{\text{OR}}$	$\Delta_y^{\text{OR}}$	$\Delta_z^{\text{OR}}$
STO-3G	22	163.69	10.345	-5.426	-15.012
3-21G	40	111.76	31.000	-4.095	-2.579
4-31G	40	113.37	21.063	-7.249	4.753
6-31G	40	115.82	17.061	-5.699	6.155
6-31G*	60	88.20	8.262	-7.512	6.475
6-31G**	66	88.59	7.101	-7.677	5.376
6-31G(3df,3pd)	156	104.56	2.334	-0.668	-4.658
6-311G	58	80.18	6.437	-6.239	9.640
6-311G*	78	58.54	1.522	-10.945	8.474
6-311G**	84	57.62	-3.264	-9.725	3.766
6-311G(2df,2pd)	148	70.90	-0.875	-2.094	1.119
3-21++G	58	63.06	1.320	9.172	6.275
6-31++G	58	111.67	-10.149	-0.303	6.216
6-31++G*	78	94.30	-8.697	0.781	6.788
6-31++G**	84	96.42	-10.737	0.340	4.399
6-311++G**	102	114.52	-10.321	-3.249	1.191
6-311++G(2d,2p)	128	101.11	-4.743	0.502	0.195
6-311++G(3df,3pd)	192	82.13	-2.109	1.712	-1.865
cc-pVDZ	66	44.69	-0.194	-3.105	5.265
cc-pVTZ	148	68.97	-3.586	1.182	2.839
cc-pVQZ	280	80.58	-2.858	2.257	0.759
cc-pV5Z	474	78.44	-3.176	2.501	-1.088
cc-pV6Z	742	81.51	-2.376	1.755	-1.594
aug-cc-pVDZ	110	73.56	-3.679	3.711	-4.323
aug-cc-pVTZ	230	77.27	-2.022	0.775	-3.246
aug-cc-pVQZ	412	79.23	-1.641	0.335	-3.025
aug-cc-pV5Z	668	80.60	-1.745	0.281	-3.051
aug-cc-pV6Z	1010	81.04	-1.729	0.286	-2.983
daug-cc-pVDZ	154	71.76	-1.158	-0.871	-4.414
daug-cc-pVTZ	312	76.69	-1.714	-0.186	-3.296
daug-cc-pVQZ	544	79.49	-1.734	0.105	-3.115
daug-cc-pV5Z	862	80.60	-1.745	0.210	-3.024
cc-pCVDZ	82	44.70	-0.009	-3.351	5.303
cc-pCVTZ	200	70.06	-3.088	0.488	3.477
cc-pCVQZ	396	80.72	-2.906	2.230	0.836
cc-pCV5Z	690	78.55	-3.123	2.482	-1.072
aug-cc-pCVDZ	126	72.69	-3.410	3.701	-4.105
aug-cc-pCVTZ	282	77.48	-1.781	0.948	-3.111
aug-cc-pCVQZ	528	79.44	-1.613	0.408	-2.954
aug-cc-pCV5Z	884	80.72	-1.733	0.306	-3.017

that diffuse functions are essential for describing optical rotation. It should be noted, however, that for the popular 6-31G-type basis sets, augmentation with diffuse functions seems to correct the specific rotation in the opposite direction of the basis set limit. In general, the Pople-style basis sets are rather ill-suited for the correlated calculation of specific rotation, as also reflected by the, in most cases, severe origin-dependence. Nevertheless, it should be stressed that the 6-31++G\* and 6-31++G\*\* basis sets might provide a reasonable compromise between computational effort and accuracy for large-scale applications. If possible, the 6-311++G(3df,3pd) basis set should also be considered for such applications. Dunning's basis sets, on the other hand, provide a well balanced hierarchy for CC2 calculations of specific rotation. Note also that the origin-dependence is

reduced to a minimum with the augmented series, although  $\Delta^{\text{OR}}$  definitely does not approach zero. Double augmentation of the basis set is seen to be less important than to increase the cardinal number of the basis set. Moreover, inclusion of core-correlating functions does not significantly improve the description, indicating validity of the frozen core approximation. This is confirmed by comparing to the frozen core results in Table VI, although we note that for the largest basis sets, the effect of freezing core orbitals is comparable to the effect of increasing the cardinal number of the basis set. However, for the smaller sets of interest for larger calculations, one may invoke the frozen core approximation with impunity. Selecting among Dunning's basis sets, the aug-cc-pVDZ or, preferably, the aug-cc-pVTZ basis set should be used for accurate calculations.



TABLE VI. As Table V but using the frozen core approximation.

Basis set	$N$	$[\alpha]_D$	$\Delta_x^{OR}$	$\Delta_y^{OR}$	$\Delta_z^{OR}$
6-31G	40	116.92	16.978	-6.049	5.912
6-31G*	60	88.87	8.199	-7.776	6.344
6-311G	58	82.05	6.068	-7.028	9.139
6-311G(2df,2pd)	148	71.33	-1.279	-2.645	0.906
6-31++G	58	112.60	-10.430	-0.604	5.951
6-31++G*	78	94.84	-8.994	0.533	6.596
6-311++G(2d,2p)	128	101.43	-5.374	0.037	-0.108
6-311++G(3df,3pd)	192	82.06	-2.704	1.272	-2.142
aug-cc-pVDZ	110	73.78	-3.918	3.550	-4.438
aug-cc-pVTZ	230	77.22	-2.501	0.428	-3.486
aug-cc-pVQZ	412	78.92	-2.246	-0.147	-3.367
aug-cc-pV5Z	668	80.10	-2.368	-0.240	-3.422
aug-cc-pV6Z	1010	80.45	-2.363	-0.250	-3.370

### C. Optical rotation of TCO

Whereas numerous *ab initio* studies of the electronic natural circular dichroism of TCO are available in the literature, see, e.g., Refs. 40, 41, 43, and 44 and references therein, none has appeared on the specific rotation. We use the DFT/B3LYP/cc-pVTZ  $C_2$  geometry of Ref. 41 for the present computations. As TCO represents a more realistic application of CC2 linear response theory to the calculation of specific rotation, we perform a limited basis set convergence study to test the main conclusions of the previous section.

Table VII contains a study of accuracy and timings. As for the similar calculations above, tight convergence criteria were employed for the wave function parameters. Once again, significantly reduced computational effort is gained from using the Cholesky-based algorithm and, measured in a.u., the errors are on the order of  $\Delta$ . In this case, too, the larger errors occurring for the diffuse basis sets at  $\Delta = 10^{-4}$  are due to the ground state amplitude decomposition: decreasing the threshold for amplitude decomposition to  $10^{-10}$  and keeping  $\Delta = 10^{-4}$  with the aug-cc-pVDZ basis set gives an error of  $-0.029892 \text{ deg} [\text{dm g/cm}^3]^{-1}$ . The sizes of the cc-pVDZ and 6-31++G\* basis sets differ by 4, which is reflected in the similar CPU times for these two basis sets. Note, however, that while the conventional algorithm is more time consuming for the largest basis set, the opposite is observed for the Cholesky algorithm. The reason for this peculiarity is that slightly more Cholesky vectors are needed for the decomposition of the cc-pVDZ integrals than for the

6-31++G\* integrals, implying that the CC2 work load is slightly larger for the former even though the number of basis functions is smaller. A decomposition threshold of  $\Delta = 10^{-6}$  is, again, sufficient for minimal Cholesky errors.

Using, therefore,  $\Delta = 10^{-6}$  and default convergence thresholds for wave function parameters (see above), specific rotations for TCO with various basis sets are given in Table VIII along with the origin-dependence vector. There are 207 million doubles amplitudes in the largest calculation (aug-cc-pVQZ basis set) which would require more than 11.6 Gb of memory with the conventional algorithm. Diffuse functions are once again observed to be important, although the cc-pVXZ series is in fact closer to the converged result than the smaller augmented Pople-style basis sets. The exceptions are 6-311++G(2d,2p) and 6-311++G(3df,3pd) which yield specific rotations in the proximity of the aug-cc-pVTZ and aug-cc-pVQZ results despite the substantially smaller number of basis functions. The basis set convergence of the origin-dependence towards a nonzero limit is even more pronounced for TCO than for CNOFH<sub>2</sub> above. Experimentally,<sup>45</sup> the specific rotation of TCO is  $-440 \text{ deg} [\text{dm g/cm}^3]^{-1}$  in CH<sub>2</sub>Cl<sub>2</sub> solution. Our best CC2 result is 38% above this value. While some of this discrepancy may be ascribed to insufficient correlation treatment of the CC2 model,<sup>17</sup> solvent effects can be expected to contribute significantly to the experimental value. In conclusion, the aug-cc-pVDZ and 6-31++G\* basis sets are found to be reasonable candidates for large-scale applications.

TABLE VII. Frozen core CC2 specific rotation ( $[\alpha]_D$ , in  $\text{deg} [\text{dm g/cm}^3]^{-1}$ ) at the sodium D line in length gauge (origin at the center of charge) for TCO using conventional and Cholesky algorithms. For Cholesky, the errors,  $[\alpha]_D(\text{Cholesky}) - [\alpha]_D(\text{Conventional})$ , are given. Total CPU timings (in hours) are also given for the coupled cluster calculations, i.e., excluding integral evaluations/decompositions and SCF optimizations.  $N$  is the number of basis functions.

Basis set	$N$	Conventional		$\Delta = 10^{-4}$		$\Delta = 10^{-6}$		$\Delta = 10^{-8}$	
		$[\alpha]_D$	Time	Error	Time	Error	Time	Error	Time
6-31G*	140	-168.945833	7.5	0.189215	0.6	-0.005446	1.0	-0.000012	1.5
6-31++G*	186	-247.278086	21.8	2.596908	1.2	-0.001086	2.0	-0.000018	3.3
cc-pVDZ	182	-259.507167	19.4	-0.045570	1.3	0.001378	2.2	0.000004	3.6
aug-cc-pVDZ	310	-286.227181	141.1	2.462699	5.2	-0.001901	9.1	-0.000019	15.5

TABLE VIII. Frozen core CC2 specific rotation ( $[\alpha]_D$ , in deg  $[\text{dm g/cm}^3]^{-1}$ ) at the sodium D line in length gauge (origin at the center of charge) for TCO using Cholesky algorithm. The only symmetry-allowed Cartesian element of the origin-dependence vector is the  $z$ -component, which is reported in deg  $[\text{dm g/cm}^3]^{-1} \text{ bohr}^{-1}$ .  $N$  is the number of basis functions.

Basis Set	$N$	$[\alpha]_D$	$\Delta_{\xi}^{\text{OR}}$
6-31G	100	-102.32	20.068
6-31G*	140	-168.95	-2.201
6-31G**	182	-167.19	-0.643
6-311G	146	-158.67	14.718
6-311G(2df,2pd)	436	-137.79	-10.697
6-31++G	146	-194.85	16.882
6-31++G*	186	-247.28	-3.533
6-31++G**	228	-243.22	-3.267
6-311++G(2d,2p)	356	-268.21	-9.650
6-311++G(3df,3pd)	564	-273.68	-8.439
cc-pVDZ	182	-259.51	-7.606
cc-pVTZ	436	-258.95	-9.538
cc-pVQZ	860	-268.46	-8.751
aug-cc-pVDZ	310	-286.23	-9.309
aug-cc-pVTZ	690	-276.35	-8.392
aug-cc-pVQZ	1284	-272.06	-8.374

## V. SUMMARY AND CONCLUDING REMARKS

We have presented an implementation of the linear response function at the CC2 level of theory based on Cholesky decompositions, notably of the two-electron integrals.<sup>1,2</sup> The algorithm is designed to avoid storage, in core as well as on disk, of four-index intermediates along the lines devised by Hättig and co-workers<sup>10–14</sup> for the RI–CC2 method. Through examples involving as many as 2.8 billion coupled cluster doubles amplitudes, we have demonstrated the large-scale applicability of the implementation. By comparing CPU times to those of the conventional integral-direct algorithm,<sup>9</sup> we have shown that significant reductions in computational effort are obtained using Cholesky decompositions. Equally important, we have explicitly demonstrated that the errors introduced by the Cholesky representation of the two-electron integrals are controllable through the decomposition threshold and, in particular, our results indicate that a decomposition threshold of  $\Delta = 10^{-6}$  guarantees essentially exact linear response properties as compared to those obtained from a conventional calculation.

In the present implementation, no use has been made of the symmetric nature of the amplitudes, i.e.,  $X_{ij}^{ab} = X_{ji}^{ba}$ . Employing this symmetry to restrict summations in the time-determining steps [i.e., calculation of  $Y$ -intermediates, Eq. (24)] will additionally decrease the computational effort involved in the Cholesky-based CC2 linear response model.

Basis set convergence studies of static and dynamic electric dipole polarizabilities and of optical rotations have been reported and discussed. As might have been expected, the frozen core approximation may be invoked and, as a general rule of thumb, one should employ Dunning's aug-cc-pVXZ basis set series for calculating these properties whenever possible. If more modest basis set sizes are called for, Pople's

6-31++G set with an additional set of polarization functions, is a good candidate.

Calculations of specific rotations in the length gauge formulation are somewhat obscured by the inherent origin-dependence. Our basis set study clearly demonstrates that this problem does not disappear in the limit of a complete basis set for the nonvariational CC2 model. Investigations are underway regarding inherently origin invariant calculations of specific rotations without resorting to London orbitals for variational as well as nonvariational models.<sup>46</sup>

## ACKNOWLEDGMENTS

The authors acknowledge support from the EU network MOLPROP, the Danish Center for Scientific Computing, and the Spanish "Plan Nacional I+D+I" of Ministerio de Ciencia y Tecnología with European FEDER funds (Project No. BQU2001-2935-C02-01). H.K. acknowledges a grant from "Programa Ramón y Cajal" of the Spanish Ministerio de Ciencia y Tecnología.

- H. Koch, A. Sánchez de Merás, and T. B. Pedersen, *J. Chem. Phys.* **118**, 9481 (2003).
- N. H. F. Beebe and J. Lindberg, *Int. J. Quantum Chem.* **12**, 683 (1977).
- M. W. Feyereisen, G. Fitzgerald, and A. Komornicki, *Chem. Phys. Lett.* **208**, 359 (1993).
- O. Vahtras, J. E. Almlöf, and M. W. Feyereisen, *Chem. Phys. Lett.* **213**, 514 (1993).
- H.-J. Werner, F. R. Manby, and P. J. Knowles, *J. Chem. Phys.* **118**, 8149 (2003).
- O. Christiansen, H. Koch, and P. Jørgensen, *Chem. Phys. Lett.* **243**, 409 (1995).
- G. D. Purvis and R. J. Bartlett, *J. Chem. Phys.* **76**, 1910 (1982).
- T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2000).
- O. Christiansen, A. Halkier, H. Koch, P. Jørgensen, and T. Helgaker, *J. Chem. Phys.* **108**, 2801 (1998).
- C. Hättig and F. Weigend, *J. Chem. Phys.* **113**, 5154 (2000).
- C. Hättig, A. Köhn, and K. Hald, *J. Chem. Phys.* **116**, 5401 (2002).
- C. Hättig and A. Köhn, *J. Chem. Phys.* **117**, 6939 (2002).
- C. Hättig, *J. Chem. Phys.* **118**, 7751 (2003).
- A. Köhn and C. Hättig, *J. Chem. Phys.* **119**, 5021 (2003).
- O. Christiansen, C. Hättig, and J. Gauss, *J. Chem. Phys.* **109**, 4745 (1998).
- H. Larsen, J. Olsen, C. Hättig, P. Jørgensen, O. Christiansen, and J. Gauss, *J. Chem. Phys.* **111**, 1917 (1999).
- K. Ruud, P. J. Stephens, F. J. Devlin, P. R. Taylor, J. R. Cheeseman, and M. J. Frisch, *Chem. Phys. Lett.* **373**, 606 (2003).
- T. Helgaker, H. J. Aa. Jensen, P. Jørgensen *et al.*, DALTON, a molecular electronic structure program, Release 1.2 (2001), available from <http://www.kjemi.uio.no/software/dalton/dalton.html>
- K. Ruud, D. Jonsson, and P. R. Taylor, *J. Chem. Phys.* **114**, 4331 (2001).
- K. Ruud (private communication).
- A. Proutière and M. Camail, *Mol. Phys.* **29**, 1473 (1975).
- R. J. Alms, A. K. Burham, and W. H. Flygare, *J. Chem. Phys.* **63**, 3321 (1975).
- M. Okrus, R. Müller, and A. Hesse, *J. Chem. Phys.* **110**, 10393 (1999).
- R. Mathies and A. C. Albrecht, *J. Chem. Phys.* **60**, 2500 (1974).
- O. Zamani-Khamiri, E. F. McIntyre, and H. F. Hameka, *J. Chem. Phys.* **72**, 1280 (1980).
- E. Perrin, P. N. Prasad, P. Mougnot, and M. Dupuis, *J. Chem. Phys.* **91**, 4728 (1989).
- J. D. Augsburger and C. E. Dykstra, *Mol. Phys.* **76**, 229 (1992).
- E. F. Archibong and A. J. Thakkar, *Mol. Phys.* **81**, 557 (1994).
- S. Nakawaga, *Chem. Phys. Lett.* **246**, 256 (1995).
- P. Norman, D. Jonsson, and H. Ågren, *Chem. Phys. Lett.* **268**, 337 (1997).
- O. Christiansen, C. Hättig, and P. Jørgensen, *Spectrochim. Acta, Part A* **55**, 509 (1999).
- L. Jensen, O. H. Schmidt, K. V. Mikkelsen, and P.-O. Åstrand, *J. Phys. Chem. B* **104**, 10462 (2000).

- <sup>33</sup>H. Soscuín, J. Hernández, R. Escobar, C. Toro-Mendoza, Y. Alvarado, and A. Hinchliffe, *Int. J. Quantum Chem.* **90**, 497 (2002).
- <sup>34</sup>I. García Cuesta, T. B. Pedersen, H. Koch, and A. M. J. Sánchez de Merás, *Chem. Phys. Lett.* (to be published).
- <sup>35</sup>R.-H. Xie, G. W. Bryant, L. Jensen, J. Zhao, and V. H. Smith, Jr., *J. Chem. Phys.* **118**, 8621 (2003).
- <sup>36</sup>R. Antoine, P. Dugourd, D. Rayane, E. Benichou, M. Broyer, F. Chandezon, and C. Guet, *J. Chem. Phys.* **110**, 9771 (1999).
- <sup>37</sup>A. Ballard, K. Bonin, and J. Louderback, *J. Chem. Phys.* **113**, 5732 (2000).
- <sup>38</sup>P. L. Polavarapu, *Mol. Phys.* **91**, 551 (1997).
- <sup>39</sup>T. B. Pedersen and H. Koch, *J. Chem. Phys.* **106**, 8059 (1997).
- <sup>40</sup>T. B. Pedersen, H. Koch, and K. Ruud, *J. Chem. Phys.* **110**, 2883 (1999).
- <sup>41</sup>T. B. Pedersen and H. Koch, *J. Chem. Phys.* **112**, 2139 (2000).
- <sup>42</sup>T. B. Pedersen, B. Fernández, and H. Koch, *J. Chem. Phys.* **114**, 6983 (2001).
- <sup>43</sup>Aa. E. Hansen and T. D. Bouman, *Adv. Chem. Phys.* **44**, 545 (1980).
- <sup>44</sup>K. L. Bak, Aa. E. Hansen, K. Ruud, T. Helgaker, J. Olsen, and P. Jørgensen, *Theor. Chim. Acta* **90**, 441 (1995).
- <sup>45</sup>A. C. Cope, C. R. Ganellin, H. W. Johnson, Jr., T. V. Van Auken, and H. J. S. Winkler, *J. Am. Chem. Soc.* **85**, 3276 (1963).
- <sup>46</sup>T. B. Pedersen, A. M. J. Sánchez de Merás, L. Boman, and H. Koch (unpublished).