

# Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models

Ida-Marie Høyvik, Rolf Heilemann Myhre, and Henrik Koch

Citation: *The Journal of Chemical Physics* **146**, 144109 (2017); doi: 10.1063/1.4979908

View online: <https://doi.org/10.1063/1.4979908>

View Table of Contents: <http://aip.scitation.org/toc/jcp/146/14>

Published by the [American Institute of Physics](#)

---

## Articles you may be interested in

[Excited states from modified coupled cluster methods: Are they any better than EOM CCSD?](#)

*The Journal of Chemical Physics* **146**, 144104 (2017); 10.1063/1.4979078

[Incremental full configuration interaction](#)

*The Journal of Chemical Physics* **146**, 104102 (2017); 10.1063/1.4977727

[A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory](#)

*The Journal of Chemical Physics* **146**, 164105 (2017); 10.1063/1.4981521

[Natural transition orbitals](#)

*The Journal of Chemical Physics* **118**, 4775 (2003); 10.1063/1.1558471

[Automatic active space selection for the similarity transformed equations of motion coupled cluster method](#)

*The Journal of Chemical Physics* **146**, 074103 (2017); 10.1063/1.4976130

[CC2 oscillator strengths within the local framework for calculating excitation energies \(LoFEx\)](#)

*The Journal of Chemical Physics* **146**, 144107 (2017); 10.1063/1.4979713

---

PHYSICS TODAY

WHITEPAPERS

### ADVANCED LIGHT CURE ADHESIVES

Take a closer look at what these environmentally friendly adhesive systems can do

READ NOW

PRESENTED BY  
 MASTERBOND  
ADHESIVES | SEALANTS | COATINGS

# Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models

Ida-Marie Høyvik,<sup>1</sup> Rolf Heilemann Myhre,<sup>1,2</sup> and Henrik Koch<sup>1,a)</sup>

<sup>1</sup>*Department of Chemistry, Faculty of Science, NTNU - Norwegian University of Science and Technology, Trondheim, Norway*

<sup>2</sup>*CTCC, Department of Chemistry, University of Oslo, 0315 Oslo, Norway*

(Received 13 January 2017; accepted 28 March 2017; published online 13 April 2017)

In this article, we present a black-box approach for the selection of orbital spaces when computing core excitation energies in the multilevel coupled cluster (MLCC) framework. Information available from the lower level of theory is used to generate correlated natural transition orbitals (CNTOs) for the high-level calculation by including both singles and doubles information in the construction of the transition orbitals. The inclusion of the doubles excitation information is essential to obtain a set of orbitals that all contain physical information, in contrast to the natural transition orbitals where only a small subset of the virtual orbitals contains physical information. The CNTOs may be included in an active space based on a cutoff threshold for the eigenvalues corresponding to the orbitals. We present MLCC results for core excitation energies calculated using coupled cluster singles and doubles (CCSD) in the inactive space and CCSD with perturbative triples (CC3) in the active space. The use of CNTOs results in small errors compared to full CC3. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4979908>]

## I. INTRODUCTION

The coupled cluster (CC) hierarchy of electronic wave function models has been successful for the accurate description of molecular systems and provides a systematic route for calculating accurate molecular properties such as excitation energies. However, a disadvantage of the CC hierarchy is the computational scaling with system size and much effort has been focused on formulating CC models with reduced computational scaling. Approaches exploiting the locality of electron correlation were proposed more than 30 years ago<sup>1-7</sup> and extensive steps in this direction have been taken<sup>8-20</sup> where each approach has numerous further developments and different flavours. Many molecular properties are size-intensive and only depend on a particular region of the molecular system. This is the foundation for the multilevel CC method<sup>21,22</sup> in which the region of interest is treated at high levels of theory and the remainder using lower level of theory. This is achieved by dividing the orbital space into two (or more) subspaces and associating an excitation manifold to each set.

Previously, the active space in multilevel CC theory<sup>21</sup> has been determined by classifying atoms and their associated atomic orbitals (AOs) as active or inactive. The active space is then determined by a Cholesky decomposition of diagonal elements of the Hartree-Fock (HF) density matrix corresponding to active AOs, and similarly for inactive AOs. This procedure is detailed by Myhre *et al.*<sup>21</sup> The Cholesky decomposition generates a set of orthogonal orbitals<sup>23</sup> that are less local than explicitly localized orbitals, but which are very convenient for partitioning the orbital space. Using explicitly

localized orbitals gives more compact orbital spaces than the Cholesky orbitals. However, if the excitation is nonlocal, the active space must be large to obtain sufficient accuracy for both types of orbitals. Furthermore, if the HF density matrix is nonlocal, the Cholesky orbitals are nonlocal and inappropriate for the description of an excitation in a reduced orbital space. Using explicitly localized HF orbitals one may generate local orbitals also for electronically nonlocal systems, see Ref. 24 and references therein for an overview of such methods. The local HF orbitals may remedy some of the problems of the Cholesky orbitals, but two important drawbacks remain: (1) local HF orbitals are not well-suited to express nonlocal excitation processes and (2) some user input to define a region in space is still required. For nonlocal excitations, we need a particular nonlocal basis that can give a good description in a reduced orbital space.

Recently, CC models that use natural transition orbitals (NTOs) have been developed for the efficient computation of electronic transitions. NTOs were introduced as a compact orbital representation for the transition density matrix by Luzanov *et al.*<sup>25</sup> and are generated by a diagonalization of matrices constructed from the single excitation vector of the transition. NTOs have been used in some mixed orbital representation schemes. Mata and Stoll<sup>26</sup> use a pair of NTOs and localized molecular orbitals for the occupied space, while the virtual space is unchanged. Helmich and Hättig<sup>27</sup> have discussed the use of pair natural orbitals (PNOs) in combination with NTOs and conclude that using an incomplete set of NTOs corrupts the localization and leads to an extended PNO space. Rather they use an implementation of CC2 excitation energies<sup>28</sup> with orbital-specific virtuals (OSVs) and PNOs to reduce the dimension of the virtual space. Baudin and Kristensen<sup>29</sup> use NTOs in combination with local occupied orbitals

<sup>a)</sup>Electronic mail: henrik.koch@ntnu.no

and local virtual orbitals to generate a reduced orbital space specific to a given transition in which a CC calculation is carried out. The orbital space is optimized to ensure that the obtained excitation energies are determined to a predefined precision. The need to use mixed orbital spaces with only one occupied-virtual pair of NTOs is a consequence of the way the NTOs are constructed. Since only single excitation information is used to construct the matrices, the diagonalization of the virtual matrix yields only the number of occupied ( $n_{occ}$ ) nonzero eigenvalues, and consequently only  $n_{occ}$  virtual orbitals of any significance to the given transition. This is not sufficient for an accurate description of the excitation.

In this article, we present a black-box procedure to obtain the active space in multilevel CC calculations by including information from the double excitation vector. The single and double excitation information is available from the lower level of theory and is used to tailor an active orbital space for the high-level calculation. We refer to the resulting orbitals as correlated natural transition orbitals (CNTOs). The CNTO eigenvalues provide a simple measure to determine which orbitals should be included in the orbital space. In this way, rather than selecting an active region of the molecule we use a threshold for which eigenvalues with corresponding orbitals to include. In the limit of a zero threshold, the full orbital space is included in the active space and the full molecular system is treated at the high level of theory.

## II. THEORY

The coupled cluster (CC) wave function can be written as

$$|\text{CC}\rangle = \exp(T)|\text{HF}\rangle, \quad (1)$$

where  $|\text{HF}\rangle$  is the HF reference determinant and  $T$  is the cluster operator. For an  $N$ -electron system, the cluster operator is given by  $T = T_1 + T_2 + \dots + T_N$ , where

$$T_1 = \sum_{ai} t_i^a E_{ai}, \quad (2)$$

$$T_2 = \sum_{(ai) \geq (bj)} t_{ij}^{ab} E_{ai} E_{bj}, \quad (3)$$

and so on for higher excitations.  $t_i^a$ ,  $t_{ij}^{ab}$ , etc., are the CC amplitudes and  $E_{pq}$  are singlet excitation operators in the second quantization formalism. We employ the usual notation for the orbital indices, where  $i, j, k, \dots$  denote orbitals which are occupied and  $a, b, c, \dots$  denote unoccupied (virtual) orbitals in the HF reference determinant. For convenience, we introduce the notation  $T = \sum_{\mu} t_{\mu} \tau_{\mu}$  for the cluster operator, where  $t_{\mu}$  are the cluster amplitudes and  $\tau_{\mu}$  the excitation operators. The energy and amplitude equations may be obtained from projecting the CC Schrödinger equation with  $\langle \text{HF} |$  and  $\langle \mu | \equiv \langle \text{HF} | \tau_{\mu}^{\dagger}$ , and can be written as

$$E = \langle \text{HF} | H | \text{CC} \rangle, \quad (4)$$

$$\Omega_{\mu} = \langle \mu | \exp(-T) H | \text{CC} \rangle = 0. \quad (5)$$

In CC response theory,<sup>30,31</sup> the excitation energies are usually determined as the eigenvalues of the non-symmetric CC Jacobian, e.g., for excitation  $\omega$ , we have

$$\mathbf{A}\mathbf{R} = \omega\mathbf{R}, \quad (6)$$

and the CC Jacobian is defined as

$$A_{\mu\nu} = \frac{\partial \Omega_{\mu}}{\partial t_{\nu}}. \quad (7)$$

The right excitation vector  $\mathbf{R}$  of Eq. (6) contains components from excitation levels, i.e., single and double excitation components  $\mathbf{R}_1$  and  $\mathbf{R}_2$  in the case of coupled cluster singles and doubles (CCSD).

### A. Multilevel CC3 theory

In this article, we present results from MLCC3 using CNTOs. The MLCC3 model is the multilevel coupled cluster model where CCSD<sup>32</sup> is used in the inactive orbital space and CCSD with perturbative triples<sup>33</sup> (CC3) is employed for the smaller active space. The theory and implementation for MLCC3 theory is presented by Myhre and Koch,<sup>22</sup> and here we reiterate the most important points. Note that we have used a notation for cluster operators opposite of what is used by Myhre *et al.* In MLCC3, we write the cluster operator introduced in Section II as

$$T = T_1 + T_2 + X_3, \quad (8)$$

where  $T_1$  and  $T_2$  are the singles and doubles cluster operators, and  $X_3$  is the triples cluster operator.  $T_1$  and  $T_2$  are treated the same way in CCSD and CC3, and no special notations are needed for these. All  $X_3$  amplitudes referencing one or more orbitals in the inactive orbital space are set to zero. Similar to the cluster operator, the projection manifold consist of all singles and doubles, but only the reduced set of triples

$$\{\langle \mu | \} = \{\langle \mu_1 | \} \oplus \{\langle \mu_2 | \} \oplus \{\langle \mu_3^X | \}. \quad (9)$$

For working equations for the amplitudes and an expression for the MLCC3 Jacobian, we refer the reader to the original MLCC3 article.<sup>22</sup>

### B. Construction of correlated natural transition orbitals

Exploiting the information from the low-level coupled cluster calculation in a multilevel coupled cluster approach (e.g., CC2<sup>28</sup> or CCSD), we may generate an orbital space which is efficient for carrying out higher level theory (e.g., CC3) in a reduced space. Using the excitation vector of Eq. (6), we may generate orthogonal transformation matrices for the occupied and virtual orbitals which provide the CNTOs. These are obtained by constructing and diagonalizing an occupied-occupied matrix,  $\mathbf{M}$ , and a virtual-virtual matrix,  $\mathbf{N}$ , from the excitation vectors.

$$M_{ij} = \sum_a R_{ai} R_{aj} + \frac{1}{2} \sum_{abk} (1 + \delta_{ai,bk} \delta_{ij}) R_{aibk} R_{ajbk}, \quad (10)$$

$$N_{ab} = \sum_i R_{ai} R_{bi} + \frac{1}{2} \sum_{ijc} (1 + \delta_{ai,cj} \delta_{ab}) R_{aicj} R_{bicj}. \quad (11)$$

We diagonalize the  $\mathbf{M}$  and  $\mathbf{N}$  matrices as

$$\mathbf{M}\mathbf{u}_i = \lambda_i^o \mathbf{u}_i, \quad i = 1, \dots, n_{occ}, \quad (12)$$

$$\mathbf{N}\mathbf{v}_a = \lambda_a^v \mathbf{v}_a, \quad a = 1, \dots, n_{vir}. \quad (13)$$

Due to the normalization of the excitation vectors we have

$$\sum_{ai} R_{ai} R_{ai} + \frac{1}{2} \sum_{aibj} (1 + \delta_{ai,bj}) R_{aibj} R_{aibj} = 1 \quad (14)$$

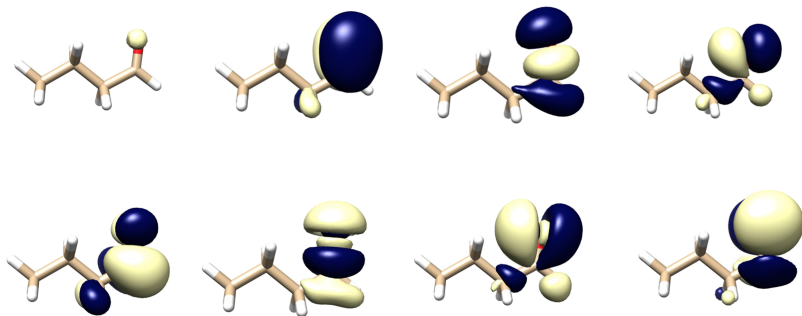


FIG. 1. CNTOs for butanal in aug-cc-pV(CT)DZ basis. The four occupied CNTOs with largest  $\lambda^o$  (top) and the four virtual CNTOs with largest  $\lambda^v$  (bottom) are plotted using contour value 0.03.

such that  $\text{Tr}(\mathbf{M}) = \text{Tr}(\mathbf{N}) = 1.0$  and  $\sum_i \lambda_i^o = 1.0$  and  $\sum_a \lambda_a^v = 1.0$ . The eigenvectors of  $\mathbf{M}$  and  $\mathbf{N}$  form the transformation matrices for the occupied and virtual space that generate the CNTOs. To determine which CNTOs to include in the active orbital space of the multilevel CC calculation, we use the sum of eigenvalues. Ordering the orbitals with descending  $\lambda_i^o$  and  $\lambda_a^v$ , we include orbitals in the active space until

$$\begin{aligned} 1 - \sum_{i \in \text{act}} \lambda_i^o &< \delta_o, \\ 1 - \sum_{a \in \text{act}} \lambda_a^v &< \delta_v. \end{aligned} \quad (15)$$

The inclusion of the doubles component of the excitation vector is essential for this scheme, because there are only  $n_{occ}$  nonzero eigenvalues with standard NTOs, as discussed previously. Including only  $n_{occ}$  virtual orbitals is not sufficient to describe the virtual space adequately. Thus, the described approach is not possible for the NTO approach of Luzanov.<sup>25</sup>

We treat excitations that are closer than approximately 1 eV in CCSD in the same basis, and a combined basis based on a sum of the excitation vectors is used. The sum is normalized so that Eq. (15) is still valid for selecting the active space.

### C. Example of correlated natural transition orbitals

Here we present illustrations of CNTOs for an oxygen core excitation in butanal in the MLCC3 framework.<sup>22</sup> The CNTOs are generated from a full space CCSD calculation, and in Figure 1 we have visualized the four most important occupied and virtual CNTOs corresponding to the four largest  $\lambda^o$  and  $\lambda^v$  values, respectively. Dunning's aug-cc-pCVTZ<sup>34</sup> basis set is used on the active oxygen and aug-cc-pVDZ<sup>35,36</sup> on the other atoms. We denote this basis as aug-cc-pV(CT)DZ. As can be seen from Figure 1, the most important occupied CNTOs for the core excitation resembles an oxygen 1s-orbital, whereas the other three CNTOs are localized around the oxygen atom. The remaining occupied CNTOs have also been investigated and orbitals with lower  $\lambda^o$  values are centered further away from the oxygen atom. The same trend is observed for the virtual orbitals.

### D. Storing double excitation vectors

For molecules of a moderate size, storing  $\mathbf{R}_2$  will not pose a problem. However, we eventually want to carry out these calculations on molecules where storing doubles information in the full space is unfeasible. In such cases, we may consider using more than two spaces and carry out the CC2 or CCSD calculation in a subspace. The lower level spaces may be treated

with CCS or even HF. Alternatively, if storing double information is to be avoided in general, one may consider CC2 and recognize the fact that CC2 doubles need not be stored, as they may be calculated on the fly. The doubles-doubles block,  $\mathbf{A}_{22}$ , of the CC2 Jacobian is diagonal if the Fock matrix is diagonal in the CC2 orbital space. If this is the case, one may construct the double excitation vector from the single excitation vector

$$\mathbf{R}_2 = -(\mathbf{A}_{22} - \mathbf{I}\omega)^{-1} \mathbf{A}_{21} \mathbf{R}_1, \quad (16)$$

where we have solved for  $\omega$ . In the current implementation, the doubles information has been stored on disk.

## III. ILLUSTRATIVE RESULTS

In this section, we first investigate the effect of  $\delta_v$  and  $\delta_o$  on the accuracy of core excitation energies, before we employ the experience gained to a more complicated example by performing a core excitation energy calculation on thymine. At the end, we discuss speedups of the MLCC3 calculations compared to full CC3 calculations.

### A. The effect of $\delta_v$ and $\delta_o$

We present oxygen core excitation energies obtained using CNTOs and the MLCC3 model. The core excitations were calculated using the core-valence separation approximation.<sup>37,38</sup> All calculations employ aug-cc-pV(CT)DZ (see Section II C). Using aug-cc-pCVTZ on the atom of interest can provide excitation energies comparable to experiment.<sup>39</sup> Excitations which are seen to be degenerate or nearly degenerate at the lower level of theory (CCSD in this case) are treated in a combined basis,

TABLE I. Two lowest MLCC3 oxygen core excitation energies for glycine using aug-cc-pV(CT)DZ. Since excitations are nearly degenerate in CCSD, they are computed in a combined CNTO basis. All excitation energies and errors are in eV.

$\log(\delta_o)$	$\log(\delta_v)$	Exc. 1	Exc. 2	Err. 1	Err. 2	nocc	nvirt
-4	-4	535.41	535.70	0.25	0.48	12	53
-4	-5	535.26	535.27	0.10	0.05	12	102
-4	-6	535.21	535.26	0.05	0.04	12	141
-4	-7	535.19	535.25	0.03	0.03	12	168
-5	-4	535.39	535.69	0.23	0.47	15	53
-5	-5	535.24	535.25	0.08	0.03	15	102
-5	-6	535.19	535.23	0.03	0.01	15	141
-5	-7	535.17	535.23	0.01	0.01	15	168
	CC3	535.16	535.22	...	...	20	180
	CCSD	537.34	537.88	...	...	20	180

TABLE II. Three lowest MLCC3 nitrogen core excitation energies for glycine using aug-cc-pV(CT)DZ. Excitations two and three are nearly degenerate in CCSD, and therefore computed in a combined CNTO basis. All excitation energies and errors are in eV.

$\log(\delta_o)$	$\log(\delta_v)$	Exc. 1	Err. 1	nocc	nvirt	Exc. 2	Exc. 3	Err. 2	Err. 3	nocc	nvirt
-4	-4	401.49	0.15	11	60	402.54	402.73	0.19	0.19	12	68
-4	-5	401.39	0.05	11	113	402.43	402.62	0.08	0.08	12	123
-4	-6	401.38	0.04	11	151	402.42	402.60	0.06	0.06	12	157
-4	-7	401.37	0.03	11	171	402.41	402.60	0.05	0.06	12	170
-5	-4	401.47	0.13	15	60	402.53	402.71	0.17	0.17	15	68
-5	-5	401.37	0.02	15	113	402.40	402.58	0.04	0.05	15	123
-5	-6	401.36	0.01	15	151	402.38	402.57	0.02	0.03	15	157
-5	-7	401.34	0.00	15	171	402.37	402.56	0.01	0.02	15	170
	CC3	401.34	...	20	176	402.36	402.54	...	...	20	176
	CCSD	402.90	...	20	176	403.99	404.52	...	...	20	176

as described at the end of Section II B. The thresholds considered range from  $10^{-4}$  to  $10^{-5}$  for  $\delta_o$  and from  $10^{-4}$  to  $10^{-7}$  for  $\delta_v$ .

Tables I–IV contain the results for glycine, butanal, and pentanal with geometries optimized using CCSD(T) and aug-cc-pVDZ in CFOUR<sup>40</sup> (see [supplementary material](#)). For glycine, we have computed the two lowest core excitation energies for the oxygen highlighted in Figure 2 and the three lowest core excitation energies of nitrogen, whereas for butanal and pentanal, we have computed the four lowest core excitation energies.

Considering the results for glycine (Tables I and II), we see that all combinations of thresholds except the combination

with  $\delta_v = 10^{-4}$  give errors on the order of 0.1 eV or less compared to CC3. Excluding the results for  $\delta_v = 10^{-4}$ , where the nitrogen errors are smaller than those for oxygen, the size of the orbital spaces and errors are similar for the oxygen and nitrogen core excitation energies. The difference between the CC3 and CCSD excitation energies range from 1.56 eV (lowest nitrogen excitation) to 2.66 (second oxygen excitation). Considering the errors are on the order of 0.1 eV, the effects of CC3 on the core excitation energies are captured by MLCC3.

For butanal and pentanal (Tables III and IV), the same trend as in glycine is observed for the first two excitations. All combinations of thresholds except the combination with  $\delta_v = 10^{-4}$  give errors on the order of 0.1 eV or less compared

TABLE III. Four lowest MLCC3 oxygen core excitation energies for butanal using aug-cc-pV(CT)DZ. Excitations 3 and 4 are determined in the same CNTO basis since they are nearly degenerate in CCSD. All excitation energies and errors are in eV.

$\log(\delta_o)$	$\log(\delta_v)$	Exc. 1	Err. 1	nocc	nvirt	Exc. 2	Err. 2	nocc	nvirt	Exc. 3	Exc. 4	Err. 3	Err. 4	nocc	nvirt
-4	-4	531.23	0.10	10	56	534.48	0.31	12	63	535.63	536.67	0.72	1.64	12	62
-4	-5	531.16	0.03	10	101	534.26	0.09	12	123	535.19	536.03	0.28	1.00	12	121
-4	-6	531.15	0.02	10	140	534.22	0.05	12	165	535.06	535.27	0.15	0.24	12	164
-4	-7	531.15	0.02	10	172	534.21	0.04	12	190	534.99	535.13	0.08	0.10	12	191
-5	-4	531.23	0.10	14	56	534.46	0.29	15	63	535.60	536.65	0.69	1.62	15	62
-5	-5	531.15	0.02	14	101	534.24	0.07	15	123	535.15	536.00	0.24	0.97	15	121
-5	-6	531.14	0.01	14	140	534.20	0.03	15	165	535.02	535.25	0.11	0.22	15	164
-5	-7	531.14	0.01	14	172	534.19	0.02	15	190	534.95	535.10	0.04	0.07	15	191
	CC3	531.13	...	20	203	534.17	...	20	203	534.91	535.03	...	...	20	203
	CCSD	532.80	...	20	203	537.49	...	20	203	538.21	538.44	...	...	20	203

TABLE IV. Four lowest MLCC3 oxygen core excitation energies for pentanal using aug-cc-pV(CT)DZ. Excitations 3 and 4 are determined in the same CNTO basis since they are nearly degenerate in CCSD. All excitation energies and errors are in eV.

$\log(\delta_o)$	$\log(\delta_v)$	Exc. 1	Err. 1	nocc	nvirt	Exc. 2	Err. 2	nocc	nvirt	Exc. 3	Exc. 4	Err. 3	Err. 4	nocc	nvirt
-4	-4	531.23	0.09	10	57	534.45	0.29	12	66	535.64	536.57	0.72	1.57	13	66
-4	-5	531.16	0.03	10	104	534.24	0.08	12	133	535.19	535.94	0.26	0.93	13	133
-4	-6	531.15	0.02	10	149	534.22	0.05	12	183	535.06	535.19	0.14	0.18	13	187
-4	-7	531.15	0.01	10	189	534.20	0.04	12	219	534.99	535.11	0.06	0.11	13	224
	CC3	531.14	...	24	240	534.16	...	24	240	534.92	535.00	...	...	24	240
	CCSD	532.80	...	24	240	537.49	...	24	240	538.21	538.43	...	...	24	240

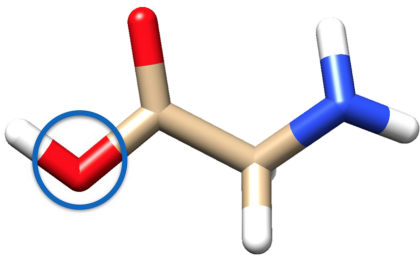


FIG. 2. The glycine molecule where we have circled the oxygen atom for which we have computed the core excitation energies.

to CC3. However, for excitations three and four, the errors are larger. These are more diffuse than excitations one and two and the errors are more dependent on  $\delta_v$ . In particular,  $\delta_v = 10^{-4}$  yield errors from 0.69 eV for excitation three in butanal to 1.57 eV for excitation four in pentanal. A threshold of  $\delta_v = 10^{-6}$  is necessary to obtain acceptable errors for these excitations. However, for the occupied orbitals,  $\delta_o = 10^{-4}$  is sufficient for excitations three and four as was the case for excitations one and two. The thresholds  $\delta_o = 10^{-4}$  and  $\delta_v = 10^{-6}$  appear to be sufficient to obtain MLCC3 core excitation energies with small errors relative to CC3. However, the error is larger for higher excitations, and one could consider running with a tighter threshold for the virtual space for higher excitations.

### B. A challenging example

To test the  $\delta_o = 10^{-4}$  and  $\delta_v = 10^{-6}$  combination of thresholds on a system that will have a nonlocal excitation and thus will be difficult to describe using reduced orbital space methods, we consider a calculation of the core excitation energy of an oxygen in thymine. This is a challenging case, due to the conjugated ring. Using Cholesky orbitals<sup>21</sup> to generate the active space for an MLCC3 calculation on thymine yields large errors in the excitation energies. In Table V, we have listed the results of a calculation using CNTOs and a calculation using Cholesky orbitals for comparison. The calculations are performed using the aug-cc-pCVDZ basis set on the active atom and aug-cc-pVDZ on the rest, denoted aug-cc-p(C)VDZ. From Table V, we see that the error in the thymine core excitation with  $\delta_o = 10^{-4}$  and  $\delta_v = 10^{-6}$  is 0.12 eV, which is larger than the errors presented in Tables I–IV. However, it is a good result considering the nonlocal nature of the virtual space needed for the calculation. Furthermore, the

TABLE V. The lowest MLCC3 oxygen core excitation energy for thymine using aug-cc-p(C)VDZ, computed using CNTOs ( $\delta_o = 10^{-4}$  and  $\delta_v = 10^{-6}$ ) and Cholesky orbitals (previous approach). All excitation energies and errors are in eV.

Approach	Exc.	Err.	nocc	nvirt
CNTOs	533.08	0.12	15	121
Cholesky	534.11	1.15	26	159
CC3	532.96	...	33	236
CCSD	535.46	...	33	236

difference between CCSD and CC3 core excitation energies is 2.5 eV, so the CC3 description of the excitation is captured with less than half of the orbital space. In Figure 3, we have plotted the four most important occupied and virtual CNTOs demonstrating their adaption to the nature of the excitation. With Cholesky orbitals, the error is approximately ten times larger, even though the active space included 11 more occupied and 38 more virtual orbitals.

### C. Computational savings

Computational savings of MLCC3 compared to full CC3 were extensively discussed in Ref. 22 and will only be briefly discussed here. The most expensive terms in full CC3 scale as  $V^4 O^3$ , where  $V$  is the number of virtual and  $O$  is the number of occupied orbitals. This is reduced to  $V V_A^3 O_A^3$  in MLCC3, where  $V_A$  and  $O_A$  are the numbers of active virtual and occupied orbitals. A theoretical speedup factor is then given by

$$\eta_{theo} = \left( \frac{V \times O}{V_A \times O_A} \right)^3. \quad (17)$$

In practice, a CCSD/CC3 calculation will rarely be close to the theoretical speedup factor because the CCSD calculation scales as  $V^3 O^3$  and will usually dominate the calculation. In addition, an integral transformation scaling as  $O_A N^4$ , where  $N$  is the number of basis functions, is required. This is demonstrated in Table VI where observed speedup factors are compared to the theoretical one. The active spaces are from the first excitation in Table IV. The total observed speedup factor changes relatively little due to the CCSD calculation while the speedup in the CC3 part is smaller than the theoretical due to the integral transformation. Looking only at the  $\Omega$  contribution, the speedup is greater than the

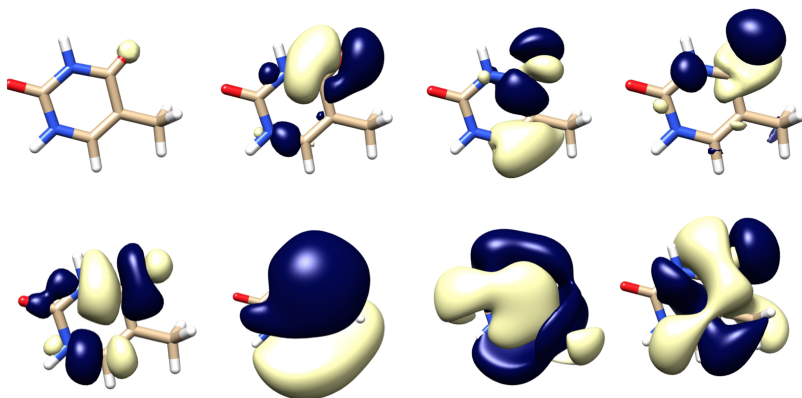


FIG. 3. CNTOs for thymine in aug-cc-p(C)VDZ basis. The four occupied CNTOs with largest  $\lambda^O$  (top) and the four virtual CNTOs with largest  $\lambda^V$  (bottom) are plotted using contour value 0.03.

TABLE VI. Observed and theoretical speedup factors for one energy iteration of pentanal using the active space for the first core excitation.  $\eta_{tot}$  includes the CCSD part and  $\eta_{CC3}$  is the speedup in the CC3 part.  $\eta_{\Omega}$  is the speedup in the CC3 contribution to the  $\Omega$  vectors without the integral transforms and is the most appropriate to compare to the theoretical speedup, as described in the text.

$\log(\delta_o)$	$\log(\delta_v)$	$\eta_{tot}$	$\eta_{CC3}$	$\eta_{\Omega}$	$\eta_{theo}$
-4	-4	14	44	1166	1032
-4	-5	11	37	187	170
-4	-6	11	25	60	58
-4	-7	9	17	28	28

theoretical, probably due to better cache utilization when reordering  $t_3$  amplitudes. The reported speedups in Table VI are for energy calculations. A Jacobi transformation normally takes two and a half times longer, but the reduced scaling due to the active space is the same. For core excitations, the Jacobi transformation takes about the same time as an energy iteration due to the effective number of occupied orbitals being much smaller.

#### IV. SUMMARY AND CONCLUDING REMARKS

In this article, we have presented an approach for the black-box selection of orbital spaces when computing excitation energies in the multilevel coupled cluster framework. Information available from the lower level of theory is used to generate correlated natural transition orbitals for the high-level calculation by including both singles and doubles information in the construction of the transition orbitals. The inclusion of the doubles excitation information is essential to obtain a set of orbitals that all contain physical information, in contrast to the natural transition orbitals where only a small subset of the virtual orbitals contains physical information. The correlated natural transition orbitals may be included in an active space, using a threshold for the eigenvalues corresponding to the orbitals. Calculated core excitation energies from the MLCC3 method are presented and the resulting errors compared to CC3 are small, even for nonlocal excitations. Clearly, the correlated natural transition orbitals are appropriately adapted for the electron transition. Using a threshold of  $10^{-4}$  for occupied orbitals and  $10^{-6}$  for virtual orbitals are seen to give good results for the molecules used in this article. However, a tighter threshold for the virtual space may be necessary when considering higher excitations. In the future, we will explore the approach for valence excitation energies.

#### SUPPLEMENTARY MATERIAL

See [supplementary material](#) for xyz-structures of the molecules used as examples in this article.

#### ACKNOWLEDGMENTS

We acknowledge computer time from NOTUR through Project No. nn2962k. H.K. acknowledges financial support from the FP7-PEOPLE-2013-IOF funding scheme (Project No. 625321). R.H.M. acknowledges financial support from the COST action ‘‘XUV/X-ray light and fast ions for ultrafast chemistry (XLIC).’’

- <sup>1</sup>J. M. Cullen and M. C. Zerner, *J. Chem. Phys.* **77**, 4088 (1982).
- <sup>2</sup>J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, *Int. J. Quantum Chem.* **14**, 545 (1978).
- <sup>3</sup>W. D. Laidig, G. D. Purvis, and R. J. Bartlett, *Int. J. Quantum Chem.* **22**, 561 (1982).
- <sup>4</sup>W. D. Laidig, G. D. Purvis, and R. J. Bartlett, *Chem. Phys. Lett.* **97**, 209 (1983).
- <sup>5</sup>W. Förner, J. Ladik, P. Otto, and J. Cizek, *Chem. Phys.* **97**, 251 (1985).
- <sup>6</sup>M. Takahashi and J. Paldus, *Phys. Rev. B* **31**, 5121 (1985).
- <sup>7</sup>S. Seibø and P. Pulay, *Annu. Rev. Phys. Chem.* **44**, 213 (1993).
- <sup>8</sup>M. Schütz, G. Hetzer, and H.-J. Werner, *J. Chem. Phys.* **111**, 5691 (1999).
- <sup>9</sup>G. E. Scuseria and P. Y. Ayala, *J. Chem. Phys.* **111**, 8330 (1999).
- <sup>10</sup>S. Li, J. Ma, and Y. Jiang, *J. Comput. Chem.* **23**, 237 (2002).
- <sup>11</sup>N. Flocke and R. J. Bartlett, *J. Chem. Phys.* **121**, 10935 (2004).
- <sup>12</sup>D. G. Fedorov and K. Kitaura, *J. Chem. Phys.* **123**, 134103 (2005).
- <sup>13</sup>J. E. Subotnik, A. Sodt, and M. Head-Gordon, *J. Chem. Phys.* **125**, 074116 (2006).
- <sup>14</sup>O. Christiansen, P. Manninen, P. Jørgensen, and J. Olsen, *J. Chem. Phys.* **124**, 084103 (2006).
- <sup>15</sup>J. Friedrich, M. Hanrath, and M. Dolg, *J. Chem. Phys.* **126**, 154110 (2007).
- <sup>16</sup>M. Kobayashi and H. Nakai, *J. Chem. Phys.* **129**, 044103 (2008).
- <sup>17</sup>F. Neese, F. Wennmohs, and A. Hansen, *J. Chem. Phys.* **130**, 114108 (2009).
- <sup>18</sup>M. Ziolkowski, B. Jansik, T. Kjærgaard, and P. Jørgensen, *J. Chem. Phys.* **133**, 014107 (2010).
- <sup>19</sup>Z. Rolik and M. Kallay, *J. Chem. Phys.* **135**, 104111 (2011).
- <sup>20</sup>J. Yang, G. K. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, *J. Chem. Phys.* **136**, 144105 (2012).
- <sup>21</sup>R. H. Myhre, A. Sánchez de Merás, and H. Koch, *J. Chem. Phys.* **141**, 224105 (2014).
- <sup>22</sup>R. H. Myhre and H. Koch, *J. Chem. Phys.* **145**, 044111 (2016).
- <sup>23</sup>F. Aquilante, T. Pedersen, A. Sánchez de Merás, and H. Koch, *J. Chem. Phys.* **125**, 174101 (2006).
- <sup>24</sup>I. Høyvik and P. Jørgensen, *Chem. Rev.* **116**, 3306 (2016).
- <sup>25</sup>A. V. Luzanov, A. A. Sukhorukov, and V. E. Umanskii, *Theor. Exp. Chem.* **10**, 354 (1976).
- <sup>26</sup>R. A. Mata and H. Stoll, *J. Chem. Phys.* **134**, 034122 (2011).
- <sup>27</sup>B. Helmich and C. Hättig, *J. Chem. Phys.* **139**, 084114 (2013).
- <sup>28</sup>O. Christiansen, H. Koch, and P. Jørgensen, *Chem. Phys. Lett.* **243**, 409 (1995).
- <sup>29</sup>P. Baudin and K. Kristensen, *J. Chem. Phys.* **144**, 224106 (2016).
- <sup>30</sup>H. Koch and P. Jørgensen, *J. Chem. Phys.* **93**, 3333 (1990).
- <sup>31</sup>T. Pedersen and H. Koch, *J. Chem. Phys.* **106**, 8059 (1997).
- <sup>32</sup>G. Purvis and R. J. Bartlett, *J. Chem. Phys.* **76**, 1910 (1982).
- <sup>33</sup>H. Koch, O. Christiansen, P. Jørgensen, A. Sánchez de Merás, and T. Helgaker, *J. Chem. Phys.* **106**, 1808 (1997).
- <sup>34</sup>D. E. Woon and T. Dunning, *J. Chem. Phys.* **103**, 4572 (1995).
- <sup>35</sup>T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>36</sup>R. Kendall, T. Dunning, and R. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- <sup>37</sup>L. Cederbaum, W. Domcke, and J. Schirmer, *Phys. Rev. A: At., Mol., Opt. Phys.* **22**, 206 (1980).
- <sup>38</sup>S. Coriani and H. Koch, *J. Chem. Phys.* **143**, 181103 (2015).
- <sup>39</sup>T. J. A. Wolf *et al.*, e-print [arXiv:1610.08498](https://arxiv.org/abs/1610.08498) [physics.chem-ph] (2016).
- <sup>40</sup>J. F. Stanton *et al.*, CFOUR, coupled-cluster techniques for computational chemistry, 2015, see <http://www.cfour.de>.