Correlated natural transition orbital framework for low-scaling excitation energy calculations (CorNFLEx)

Pablo Baudin and Kasper Kristensen[a)]
Department of Chemistry, qLEAP Center for Theoretical Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus C, Denmark

(Received 24 March 2017; accepted 18 May 2017; published online 7 June 2017)

We present a new framework for calculating coupled cluster (CC) excitation energies at a reduced computational cost. It relies on correlated natural transition orbitals (NTOs), denoted CIS(D′)-NTOs, which are obtained by diagonalizing generalized hole and particle density matrices determined from configuration interaction singles (CIS) information and additional terms that represent correlation effects. A transition-specific reduced orbital space is determined based on the eigenvalues of the CIS(D′)-NTOs, and a standard CC excitation energy calculation is then performed in that reduced orbital space. The new method is denoted CorNFLEx (Correlated Natural transition orbital Framework for Low-scaling Excitation energy calculations). We calculate second-order approximate CC singles and doubles (CC2) excitation energies for a test set of organic molecules and demonstrate that CorNFLEx yields excitation energies of CC2 quality at a significantly reduced computational cost, even for relatively small systems and delocalized electronic transitions. In order to illustrate the potential of the method for large molecules, we also apply CorNFLEx to calculate CC2 excitation energies for a series of solvated formamide clusters (up to 4836 basis functions). *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4984820]

I. INTRODUCTION

Nowadays, coupled cluster (CC) theory,[1,2] together with the response function[3–7] or the equation-of-motion[8–11] formalisms, is well-established as the method of choice for the calculation of electronic transition properties of molecules dominated by a single electronic configuration. However, the steep computational scaling with a system size of CC theory limits its application to molecules with a few tens of atoms. For the calculation of transition properties of large molecules, more affordable but less reliable methods, such as time-dependent density-functional-theory (TDDFT), are generally preferred.[12] The limitations of DFT methods are well-known,[13,14] and it is therefore important to provide more robust alternatives for large-scale calculations of transition properties.

The steep computational scaling of CC theory can be attributed to the fact that highly delocalized canonical Hartree-Fock (HF) orbitals are used to describe electron correlation effects that are of local nature.[15,16] For the calculation of CC ground state energies, many methods have been developed to take advantage of the locality of correlation effects to reduce the cost of CC calculations, and some efficient linear-scaling algorithms are now available.[17–22] We note that the ground state correlation energy is a size-extensive quantity, and, by definition, CC algorithms that target size-extensive properties have to scale at least linearly with the system size. The key to achieving a reduction in computational cost without affecting significantly the accuracy of the calculated quantities is to formulate the CC equations in a basis where the inherent locality of electron correlation can be efficiently exploited. While orthogonal localized occupied orbitals are universally used,[23–26] the virtual space has been described by local orthogonal virtual orbitals[26–28] or non-orthogonal alternatives, such as projected atomic orbitals (PAOs),[29] orbital-specific virtuals (OSVs),[30] or pair-natural orbitals (PNOs).[31–33]

More recently, some attention has also been given to CC algorithms with reduced scaling for the determination of excitation energies and other frequency-dependent properties. In this context, it is important to realize that electronic transition properties are size-intensive. For local electronic transitions, it should therefore be possible to devise a procedure where the computational cost depends only on the character of the electronic transition and not on the system size. However, unlike ground state correlation effects, some electronic transitions are not of local nature, and it is therefore not straightforward to use the same locality approximations for the calculation of transition properties as for the ground state energy.

In the design of approximated CC methods for excitation energy calculations, the size-intensivity of electronic transitions has often been ignored. Instead, the local approximations for ground state calculations have been extended to take into account the potential delocalized character of electronic transitions by relying on information from low-level calculations[34–43] [often from the configuration interaction singles (CIS) model]. Such a strategy allows for a more uniform description of size-intensive and size-extensive properties but limits significantly the potential computational savings. On the other hand, some approaches have been designed specifically for the calculation of size-intensive properties[44–49] and cannot be used to obtain, e.g., correlation energies. Finally,

a)kasperk@chem.au.dk

we note that computational savings can be obtained by simply truncating the virtual orbital space in the canonical orbital basis.[50,51]

In a previous publication, we have introduced a local framework for calculating CC excitation energies (LoFEx),[49] which provides a general approach for calculating excitation energies at a reduced computational cost. The orbital space in the LoFEx approach is a mixed space containing natural transition orbitals (NTOs) determined from a time-dependent Hartree-Fock (TDHF) calculation and localized molecular orbitals (LMOs). A subset of this mixed orbital space is then determined in a black-box manner to ensure the error control of the calculated quantities. In Ref. 49, it was shown that a compact description of electronic transitions could be obtained for transitions that are localized to a small part of the considered molecular system. However, for transitions that are delocalized over a large part of the molecule, no computational savings can be obtained with LoFEx.

In this paper, we investigate alternative orbital spaces that can be used to efficiently describe all types of transitions (including delocalized ones) and avoid the need for an optimization of the reduced orbital space. The new orbital spaces are made of generalized NTOs that include correlation effects such that an optimal reduced orbital space can be chosen solely based on the NTO eigenvalues as described in Secs. II and III. The new method entitled CorNFLEx (Correlated Natural transition orbital Framework for Low-scaling Excitation energy calculations) is tested on a set of spectroscopically interesting medium-sized molecules to calculate excitation energies at the second-order approximated CC singles and doubles (CC2) level (Sec. IV). We also calculate excitation energies of solvated formamide clusters of increasing size to illustrate the capabilities of the new method.

II. NATURAL TRANSITION ORBITALS

Our overall goal is to obtain computational savings by carrying out a standard CC excitation energy calculation within a reduced orbital space composed of transition-specific orthogonal molecular orbitals (MOs). In this section, we describe and compare different choices of NTOs that may be used for this purpose. Throughout the paper, we consider closed-shell molecules and spin-free orbitals using the following index convention (unless otherwise noted):

- i, j, k, l: occupied canonical MOs;
- a, b, c, d: virtual canonical MOs;
- I, J, K, L: occupied MOs in the CIS-NTO basis;
- A, B, C, D: virtual MOs in the CIS-NTO basis;
- I', J', K', L': occupied MOs in the *occupied correlation domain* (OCD);
- A', B', C', D': virtual MOs in the *virtual correlation domain* (VCD);
- p, q, r, s: MOs of unspecified nature and occupation;
- $\alpha, \beta, \gamma, \delta, \mu, \nu$: atomic orbitals;
- $\tilde{\mu}, \tilde{\nu}$: projected atomic orbitals.

In Sec. II A, we summarize how NTOs are obtained at the CIS level. In Secs. II B and II C, we generalize the concept of NTOs to include correlation effects at the CC2 and CIS(D)[52] levels of theory, while an approximation to CIS(D)-NTOs is introduced in Sec. II D.

A. CIS-NTOs

NTOs are usually obtained from a singular-value-decomposition (SVD) of one-particle excitation vectors.[53,54] At the CIS level, the excitation vector $\mathbf{R}^{\mathrm{CIS}}$ associated with a given electronic transition is obtained from the following eigenvalue problem:[55]

$$(\mathbf{H}^{\mathrm{CIS}} - E_{\mathrm{HF}}\mathbf{1})\mathbf{R}^{\mathrm{CIS}} = \omega^{\mathrm{CIS}}\mathbf{R}^{\mathrm{CIS}}, \tag{1}$$

where $\mathbf{H}^{\mathrm{CIS}}$ is the CIS Hamiltonian, E_{HF} is the HF ground-state energy, and the eigenvalue ω^{CIS} is the associated excitation energy. An SVD of $\mathbf{R}^{\mathrm{CIS}}$ is equivalent to the diagonalization of the following hole and particle density matrices:

$$M_{ij}^{\mathrm{CIS}} = \sum_a R_{ai}^{\mathrm{CIS}} R_{aj}^{\mathrm{CIS}}, \tag{2a}$$

$$N_{ab}^{\mathrm{CIS}} = \sum_i R_{ai}^{\mathrm{CIS}} R_{bi}^{\mathrm{CIS}}. \tag{2b}$$

CIS-NTOs are thus obtained by solving the eigenvalue equations,

$$\mathbf{M}^{\mathrm{CIS}}\mathbf{U}^{\mathrm{CIS}} = \boldsymbol{\lambda}^{\mathrm{CIS}}\mathbf{U}^{\mathrm{CIS}}, \tag{3a}$$

$$\mathbf{N}^{\mathrm{CIS}}\mathbf{V}^{\mathrm{CIS}} = \tilde{\boldsymbol{\lambda}}^{\mathrm{CIS}}\mathbf{V}^{\mathrm{CIS}}, \tag{3b}$$

where $\boldsymbol{\lambda}^{\mathrm{CIS}}$ and $\tilde{\boldsymbol{\lambda}}^{\mathrm{CIS}}$ are diagonal matrices with eigenvalues λ_p^{CIS} and $\tilde{\lambda}_p^{\mathrm{CIS}}$ on the diagonal. The $\mathbf{U}^{\mathrm{CIS}}/\mathbf{V}^{\mathrm{CIS}}$ matrices represent the transformation from the occupied/virtual canonical MO basis to the occupied/virtual CIS-NTO basis. In the following discussion, it is assumed that the eigenvalues in Eq. (3) are listed in order of decreasing magnitude and that the number of virtual orbitals is larger than the number of occupied orbitals ($V > O$), which is the case in all practical applications. In order to simplify the following discussion, we note that the rank of a matrix equals the number of non-zero singular values and that the following relations hold for a real matrix $\mathbf{B}$ of dimension (m, n):

$$\mathrm{rank}(\mathbf{B}^T\mathbf{B}) = \mathrm{rank}(\mathbf{B}\mathbf{B}^T) = \mathrm{rank}(\mathbf{B}) \leq \min(m, n). \tag{4}$$

Thus, from Eq. (2), it follows that

$$\mathrm{rank}(\mathbf{M}^{\mathrm{CIS}}) = \mathrm{rank}(\mathbf{R}^{\mathrm{CIS}}) \leq O, \tag{5a}$$

$$\mathrm{rank}(\mathbf{N}^{\mathrm{CIS}}) = \mathrm{rank}(\mathbf{R}^{\mathrm{CIS}}) \leq O, \tag{5b}$$

which implies that the last $(V - O)$ eigenvalues of $\mathbf{N}^{\mathrm{CIS}}$ are zero. Furthermore, the first O virtual eigenvalues $\tilde{\lambda}_p^{\mathrm{CIS}}$ equal the occupied eigenvalues λ_p^{CIS}, i.e.,[54]

$$\tilde{\lambda}_p^{\mathrm{CIS}} = \lambda_p^{\mathrm{CIS}} \qquad (1 \leq p \leq O), \tag{6a}$$

$$\tilde{\lambda}_p^{\mathrm{CIS}} = 0 \qquad (O < p \leq V). \tag{6b}$$

We also note that normalizing the CIS excitation vector is equivalent to requiring the sum of the eigenvalues to equal one,

$$\sum_p \lambda_p^{\mathrm{CIS}} = 1. \tag{7}$$

The magnitude of the eigenvalues λ_p^{CIS} can be related to the importance of a given pair of CIS-NTOs for describing the

transition of interest at the CIS level of theory. It would thus be convenient to choose a subset of CIS-NTOs based on their eigenvalues and then carry out a correlated CC calculation within this reduced orbital space in order to decrease the computational effort of the CC calculation. However, such a strategy is not useful in practice since it discards *at least* all the virtual CIS-NTOs with zero eigenvalues. This would result in unacceptable errors in the subsequent CC calculation and make it impossible to approach the full CC result in a systematic manner. In Secs. II B–II D, we generalize the concept of NTOs to include correlation effects with the objective of calculating accurate CC excitation energies within a reduced orbital space composed of correlated NTOs.

B. CC2-NTOs

In order to include the effects of CC2 doubles excitations in the NTOs, we consider the following generalized hole and particle density matrices at the CC2 level of theory:

$$M_{ij}^{\mathrm{CC2}} = \sum_a R_{ai}^{\mathrm{CC2}} R_{aj}^{\mathrm{CC2}} + \sum_{abk} \left(\frac{1+\delta_{ai,bk}}{2}\right)^{\frac{1}{2}} R_{aibk}^{\mathrm{CC2}} \left(\frac{1+\delta_{aj,bk}}{2}\right)^{\frac{1}{2}} R_{ajbk}^{\mathrm{CC2}}, \tag{8a}$$

$$N_{ab}^{\mathrm{CC2}} = \sum_i R_{ai}^{\mathrm{CC2}} R_{bi}^{\mathrm{CC2}} + \sum_{ijc} \left(\frac{1+\delta_{ai,cj}}{2}\right)^{\frac{1}{2}} R_{aicj}^{\mathrm{CC2}} \left(\frac{1+\delta_{bi,cj}}{2}\right)^{\frac{1}{2}} R_{bicj}^{\mathrm{CC2}}, \tag{8b}$$

where R_{ai}^{CC2} and R_{aibj}^{CC2} are the singles and doubles components of the solution vector of the right CC2 eigenvalue problem,[56]

$$\mathbf{A}\mathbf{R}^{\mathrm{CC2}} = \omega^{\mathrm{CC2}} \mathbf{R}^{\mathrm{CC2}}, \tag{9}$$

and $\mathbf{A}$ and ω^{CC2} are the CC2 Jacobian and CC2 excitation energies, respectively. We note that the density matrices in Eq. (8) (with a slightly different convention for the off-diagonal elements) were first introduced by Høyvik, Myhre, and Koch in connection with the multilevel CC scheme[57] and that $\mathrm{Tr}(\mathbf{M}^{\mathrm{CC2}}) = \mathrm{Tr}(\mathbf{N}^{\mathrm{CC2}}) = 1$ when the $\mathbf{R}^{\mathrm{CC2}}$ vector is normalized. In analogy with Eq. (3) for the CIS density matrices, the CC2-NTOs are obtained by diagonalizing $\mathbf{M}^{\mathrm{CC2}}$ and $\mathbf{N}^{\mathrm{CC2}}$,

$$\mathbf{M}^{\mathrm{CC2}}\mathbf{U}^{\mathrm{CC2}} = \boldsymbol{\lambda}^{\mathrm{CC2}}\mathbf{U}^{\mathrm{CC2}}, \tag{10a}$$

$$\mathbf{N}^{\mathrm{CC2}}\mathbf{V}^{\mathrm{CC2}} = \tilde{\boldsymbol{\lambda}}^{\mathrm{CC2}}\mathbf{V}^{\mathrm{CC2}}. \tag{10b}$$

For the following analysis, it is convenient to interpret the total $\mathbf{R}^{\mathrm{CC2}}$ eigenvector entering Eq. (8a) as a matrix $\mathbf{R}^M$ with a singles block of dimension (V,O) and elements $R_{a,i}^{\mathrm{CC2}}$ and a doubles block of dimension (V^2O, O) and elements $(\frac{1+\delta_{ai,bj}}{2})^{\frac{1}{2}} R_{bja,i}^{\mathrm{CC2}}$ (see Fig. 1, left). This enables us to write the $\mathbf{M}^{\mathrm{CC2}}$ matrix in Eq. (8a) as

$$\mathbf{M}^{\mathrm{CC2}} = (\mathbf{R}^M)^T \mathbf{R}^M. \tag{11}$$

Similarly, we consider a matrix $\mathbf{R}^N$ with a singles block of dimension (V,O) and elements $R_{a,i}^{\mathrm{CC2}}$ and a doubles block of dimension (V,O^2V) and elements $(\frac{1+\delta_{ai,bj}}{2})^{\frac{1}{2}} R_{a,ibj}^{\mathrm{CC2}}$ (see Fig. 1, right) and write the $\mathbf{N}^{\mathrm{CC2}}$ matrix in Eq. (8b) as

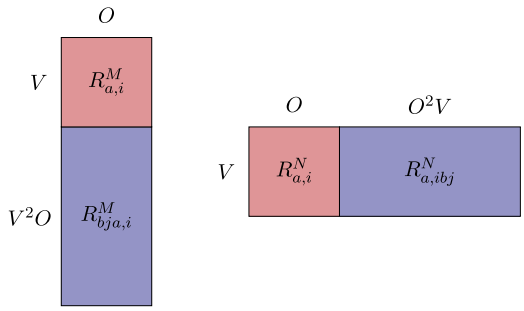

FIG. 1. Representation of the $\mathbf{R}^M$ (left) and $\mathbf{R}^N$ (right) matrices used to calculate generalized transition density matrices.

$$\mathbf{N}^{\mathrm{CC2}} = \mathbf{R}^N (\mathbf{R}^N)^T. \tag{12}$$

Using the general identity in Eq. (4) and the dimensions of the $\mathbf{R}^M$ and $\mathbf{R}^N$ matrices, it then follows that

$$\mathrm{rank}(\mathbf{M}^{\mathrm{CC2}}) \leq \min(V + V^2O, O) = O, \tag{13a}$$

$$\mathrm{rank}(\mathbf{N}^{\mathrm{CC2}}) \leq \min(V, O + O^2V) = V. \tag{13b}$$

This should be put in contrast with the ranks of the CIS density matrices in Eqs. (5). Comparing Eqs. (2) and (8), we thus see that the inclusion of the doubles excitation vector component enables the existence of nonzero eigenvalues for all of the CC2-NTOs generated by the diagonalization of $\mathbf{M}^{\mathrm{CC2}}$ and $\mathbf{N}^{\mathrm{CC2}}$. We also note that in contrast with the CIS case in Eq. (6), the occupied and virtual CC2-NTO eigenvalues are generally different, $\lambda_p^{\mathrm{CC2}} \neq \tilde{\lambda}_p^{\mathrm{CC2}}$.

The generation of the CC2-NTOs requires the determination of the CC2 excitation vector for the full molecular system, and the CC2-NTOs are therefore primarily of analytical interest if CC2 excitation energies are targeted. It is thus instructive to look at the explicit expression for the CC2 doubles excitation vector in order to consider approximations in Secs. II C and II D that could be invoked in practical calculations of CC2 excitation energies. From Ref. 58, the CC2 doubles excitation vector is given by

$$R_{aibj}^{\mathrm{CC2}} = \frac{1}{(1+\delta_{ai,bj})} \frac{(ai|\bar{b}j)^{\mathrm{CC2}}}{\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b + \omega^{\mathrm{CC2}}}, \tag{14}$$

where the transformed integrals are written as

$$(ai|\bar{b}j)^{\mathrm{CC2}} = P_{ij}^{ab} \sum_{\alpha\beta\gamma\delta} (\bar{\Lambda}_{\alpha a}^{\mathrm{CC2}} \Lambda_{\beta i}^{\mathrm{CC2}} + \Lambda_{\alpha a}^{\mathrm{CC2}} \bar{\Lambda}_{\beta i}^{\mathrm{CC2}}) \Lambda_{\gamma b}^{\mathrm{CC2}} \Lambda_{\delta j}^{\mathrm{CC2}} (\alpha\beta|\gamma\delta) \tag{15}$$

with

$$P_{ij}^{ab} f_{aibj} = f_{aibj} + f_{bjai} \tag{16}$$

and

$$\bar{\Lambda}_{\alpha a}^{\mathrm{CC2}} = -\sum_i C_{\alpha i} R_{ai}^{\mathrm{CC2}}, \tag{17a}$$

$$\bar{\Lambda}_{\alpha i}^{\mathrm{CC2}} = \sum_a C_{\alpha a} R_{ai}^{\mathrm{CC2}}, \tag{17b}$$

$$\Lambda_{\alpha a}^{\mathrm{CC2}} = C_{\alpha a} - \sum_i C_{\alpha i} t_{ai}, \tag{17c}$$

$$\Lambda_{\alpha i}^{\mathrm{CC2}} = C_{\alpha i} + \sum_a C_{\alpha a} t_{ai}, \tag{17d}$$

where $\mathbf{C}$ is the standard (canonical) MO coefficient matrix and t_{ai} are the ground state CC2 singles amplitudes, while ϵ_i and

ϵ_a represent the occupied and virtual orbital energies, respectively. The expensive part of the CC2-NTO procedure is the calculation of the doubles excitation vector components R^{CC2}_{aibj} in Eq. (14) and their contraction in Eq. (8), both of which scale as $\mathcal{O}(N^5)$ where N is a measure of the size of the molecular system. In particular, the determination of R^{CC2}_{aibj} requires the iterative solution of both the ground state CC2 equations (to determine t_{ai}) and the subsequent iterative solution of the CC2 eigenvalue problem in Eq. (9).

C. CIS(D)-NTOs

As a first approximation to the CC2-NTOs, we consider NTOs generated from the simpler CIS(D) model.[52] In Eq. (14), this effectively corresponds to setting the ground state singles amplitudes to zero ($t^a_i = 0$) and using the CIS excitation energy and excitation vector instead of the corresponding CC2 quantities. Thus, within the CIS(D) model, the $R^{\mathrm{CIS(D)}}_{aibj}$ amplitudes may be determined in a non-iterative $\mathcal{O}(N^5)$ process according to

$$R^{\mathrm{CIS(D)}}_{aibj} = \frac{1}{(1+\delta_{ai,bj})}\frac{(ai|bj)^{\mathrm{CIS}}}{\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b + \omega^{\mathrm{CIS}}}, \tag{18}$$

where the transformed integrals are given by

$$(ai|\bar{bj})^{\mathrm{CIS}} = P^{ab}_{ij}\sum_{\alpha\beta\gamma\delta}(\bar{\Lambda}^{\mathrm{CIS}}_{\alpha a}C_{\beta i} + C_{\alpha a}\bar{\Lambda}^{\mathrm{CIS}}_{\beta i})C_{\gamma b}C_{\delta j}(\alpha\beta|\gamma\delta), \tag{19}$$

$$\bar{\Lambda}^{\mathrm{CIS}}_{\alpha a} = -\sum_i C_{\alpha i}R^{\mathrm{CIS}}_{ai}, \tag{20a}$$

$$\bar{\Lambda}^{\mathrm{CIS}}_{\alpha i} = \sum_a C_{\alpha a}R^{\mathrm{CIS}}_{ai}. \tag{20b}$$

The CIS(D)-NTOs are then obtained by diagonalizing the CIS(D) hole and particle density matrices, which are constructed analogously to Eqs. (8),

$$M^{\mathrm{CIS(D)}}_{ij} = \sum_a R^{\mathrm{CIS}}_{ai}R^{\mathrm{CIS}}_{aj} + \sum_{abk}\left(\frac{1+\delta_{ai,bk}}{2}\right)^{\frac{1}{2}}R^{\mathrm{CIS(D)}}_{aibk}\left(\frac{1+\delta_{aj,bk}}{2}\right)^{\frac{1}{2}}R^{\mathrm{CIS(D)}}_{ajbk}, \tag{21a}$$

$$N^{\mathrm{CIS(D)}}_{ab} = \sum_i R^{\mathrm{CIS}}_{ai}R^{\mathrm{CIS}}_{bi} + \sum_{ijc}\left(\frac{1+\delta_{ai,cj}}{2}\right)^{\frac{1}{2}}R^{\mathrm{CIS(D)}}_{aicj}\left(\frac{1+\delta_{bi,cj}}{2}\right)^{\frac{1}{2}}R^{\mathrm{CIS(D)}}_{bicj}. \tag{21b}$$

D. CIS(D′)-NTOs

The generation of CIS(D)-NTOs is a non-iterative $\mathcal{O}(N^5)$ process, and it is therefore prohibitively expensive for large molecular systems. In this section, we describe a strategy for determining approximate CIS(D)-NTOs by using CIS information and locality arguments to approximate the doubles terms in Eq. (21).

The occupied and virtual CIS-NTO bases are defined by the $\mathbf{U}^{\mathrm{CIS}}$ and $\mathbf{V}^{\mathrm{CIS}}$ transformation matrices in Eq. (3). We now consider two mixed bases—one where the occupied orbitals are canonical and the virtual orbitals are expressed in the CIS-NTO basis, and vice versa—leading to the following approximate doubles vectors:

$$R^{\mathrm{CIS(D')}}_{AiBj} = \frac{1}{(1+\delta_{Ai,Bj})}\frac{(Ai|\bar{Bj})^{\mathrm{CIS}}}{\epsilon_i - F_{AA} + \epsilon_j - F_{BB} + \omega^{\mathrm{CIS}}}, \tag{22a}$$

$$R^{\mathrm{CIS(D')}}_{aIbJ} = \frac{1}{(1+\delta_{aI,bJ})}\frac{(aI|\bar{bJ})^{\mathrm{CIS}}}{F_{II} - \epsilon_a + F_{JJ} - \epsilon_b + \omega^{\mathrm{CIS}}}, \tag{22b}$$

where we have used capital letters to denote the CIS-NTO basis and made the diagonal Fock matrix approximation in the CIS-NTO basis. The use of Eq. (22) alone does not reduce the computational effort compared to Eq. (18). However, we now invoke a procedure where we freeze the most important CIS-NTOs and transform the remaining orbitals to a local basis in which the number of important doubles vector elements can be reduced. The philosophy behind this approach is that the frozen CIS-NTOs describe the main character of the transition at the CIS level of theory, while the remaining local orbitals allow for an efficient description of correlation effects. In the remainder of this section, we describe the technical details of this procedure which is also summarized in Figs. 2 and 3.

As a first step, we define a subset of Z occupied and Z virtual CIS-NTOs for which the eigenvalues add up to one minus some threshold τ_{CIS},

$$\min_Z\left(\sum_{p=1}^{Z}\lambda^{\mathrm{CIS}}_p\right) > 1 - \tau_{\mathrm{CIS}}, \tag{23}$$

where p refers to either occupied ($p = I$) or virtual ($p = A$) CIS-NTOs. In practice, we set $\tau_{\mathrm{CIS}} = 10^{-3}$. In order to account for correlation effects associated with the ($2Z$) important CIS-NTOs defined by Eq. (23), we define an *atomic correlation domain* (ACD), which is a list of atoms that represent the part of the molecular system where the important CIS-NTOs

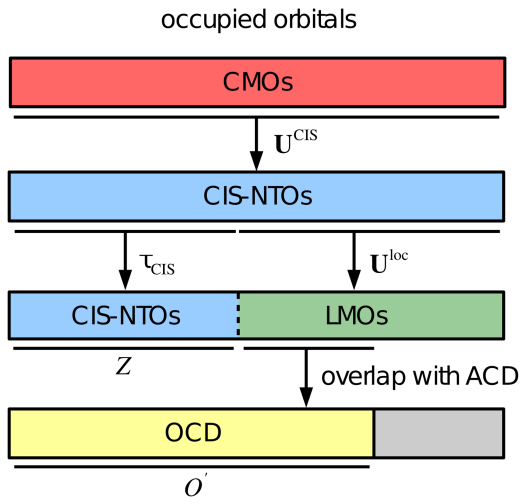

FIG. 2. Schematic representation of the generation of the occupied correlation domain (OCD). The $\mathbf{U}^{\mathrm{CIS}}$ matrix transforms from occupied canonical molecular orbitals (CMOs) to the CIS natural transition orbital (CIS-NTOs) basis. A subset of Z CIS-NTOs is selected using τ_{CIS} [see Eq. (23)], while the $\mathbf{U}^{\mathrm{loc}}$ matrix transforms the remaining CIS-NTOs to a local basis (LMOs). The atomic correlation domain (ACD) is constructed as a list of atoms located in the same part of space as the important CIS-NTOs. Finally, LMOs are included in the OCD if their atomic extents overlap with any of the atoms in the ACD, see text for details.

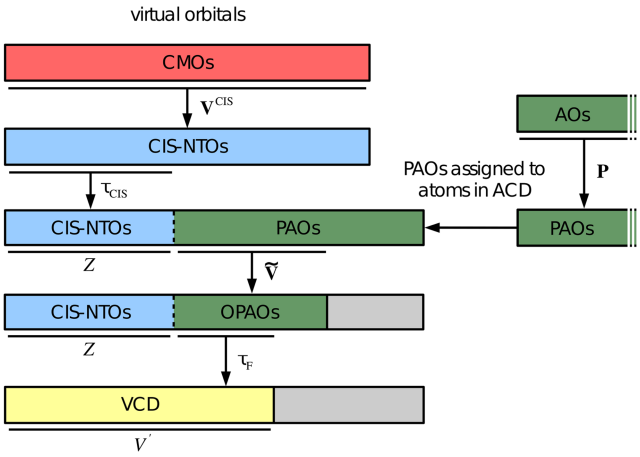

FIG. 3. Schematic representation of the generation of the virtual correlation domain (VCD). The $\mathbf{V}^{\mathrm{CIS}}$ matrix transforms from virtual canonical molecular orbitals (CMOs) to the CIS natural transition orbital (CIS-NTO) basis. A subset of Z CIS-NTOs is selected using τ_{CIS} [see Eq. (23)], while the $\mathbf{P}$ matrix is used to generate projected atomic orbitals (PAOs), see Eq. (25). The atomic correlation domain (ACD) is constructed as a list of atoms located in the same part of space as the important CIS-NTOs. A subset of PAOs assigned to atoms in the ACD is then orthonormalized using the $\tilde{\mathbf{V}}$ matrix [see Eq. (29)], and redundant PAOs are discarded to arrive at a set of non-redundant orthogonal PAOs (OPAOs). Finally, all OPAOs with a diagonal matrix element $F_{\tilde{\mu}\tilde{\mu}} > \tau_F$ are discarded. The remaining OPAOs and the Z important CIS-NTOs then constitute the VCD.

are located. Specifically, for each important CIS-NTO (p), we determine an orbital extent as the smallest possible reduced set of atomic orbitals (AOs) $\{\mu\}_p$ for which the accumulated Löwdin charge is above 0.95, i.e.,

$$\sum_{\mu\in\{\mu\}_p} Q_\mu^p > 0.95, \tag{24}$$

where Q_μ^p is the Löwdin charge of CIS-NTO p on AO μ. An atomic site is then included in the ACD if at least one of its AOs is contained in one of the orbital extents for the important CIS-NTOs.

Once the important CIS-NTOs have been identified and the ACD has been defined, we localize the remaining $(O - Z)$ occupied orbitals and determine the orbital extent for each of these as described above. We then define the *occupied correlation domain* (OCD), which contains the Z important occupied CIS-NTOs and those localized occupied orbitals for which the orbital extent contains at least one AO assigned to an atom in the ACD. The OCD is thus a mixed occupied CIS-NTO/local basis $\{I'\}$ of dimension O' with $Z \leq O' \leq O$. As a localization function, we use the second power of the second moment of the orbitals,[26] which provides a good compromise between the locality and computational cost.[28] The generation of the OCD is illustrated in Fig. 2.

A similar but slightly more involved procedure is used to obtain a reduced *virtual correlation domain* (VCD). We first define a set of projected atomic orbitals (PAOs) where all occupied orbitals *and* the Z important virtual CIS-NTOs have been projected out,

$$|\tilde{\mu}\rangle = \sum_\alpha P_{\alpha\mu}|\alpha\rangle, \tag{25}$$

where the projection matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \mathbf{1} - \mathbf{DS}, \tag{26}$$

$$D_{\mu\nu} = \sum_i C_{\mu i}C_{\nu i} + \sum_{A=1}^{Z} C_{\mu A}C_{\nu A}, \tag{27}$$

$$S_{\mu\nu} = \langle\mu|\nu\rangle. \tag{28}$$

We note that in Eq. (27), the i-summation runs over *all* occupied orbitals (both core and valence orbitals), while the A-summation runs only over the set of important virtual CIS-NTOs defined by Eq. (23). The PAOs in Eq. (25) are thus a redundant and nonorthogonal set of local orbitals that spans the part of the virtual space, which is orthogonal to the Z important virtual CIS-NTOs. Note that the PAOs defined by Eq. (25) are different from conventional PAOs[29] due to the second term in Eq. (27).

We now construct a reduced set of PAOs by considering only PAOs spatially local to the important CIS-NTOs, i.e., if atomic orbital μ does not belong to the ACD, then PAO $\tilde{\mu}$ is discarded. All the remaining PAOs are then orthonormalized and redunancies are removed, i.e., we diagonalize the PAO overlap matrix,

$$\tilde{\mathbf{V}}^T\tilde{\mathbf{S}}\tilde{\mathbf{V}} = \boldsymbol{\eta}, \tag{29}$$

$$\tilde{S}_{\tilde{\mu}\tilde{\nu}} = \langle\tilde{\mu}|\tilde{\nu}\rangle, \quad \eta_{\tilde{\mu}\tilde{\nu}} = \delta_{\tilde{\mu}\tilde{\nu}}\eta_{\tilde{\mu}\tilde{\mu}}, \tag{30}$$

and generate orthogonalized PAOs (OPAOs) using the $\tilde{\mathbf{V}}$ transformation matrix. OPAOs with small eigenvalues ($\tilde{\eta}_{\tilde{\mu}\tilde{\mu}} < 10^{-3}$) are removed from the set, and the remaining non-redundant OPAOs are normalized. Furthermore, from Eq. (22a), we note that large diagonal virtual Fock matrix elements will generally lead to small doubles excitation vector elements. We therefore propose to reduce the set of OPAOs further by considering only the OPAOs for which the diagonal Fock matrix elements are above a certain threshold τ_F, which we set to 50 eV. To summarize, the Z important CIS-NTOs and the reduced set of OPAOs define the VCD $\{A'\}$, which is an orthonormal set of virtual orbitals of dimension V' with $Z \leq V' \leq V$. The generation of the VCD is illustrated in Fig. 3.

Using the occupied $\{I'\}$ and virtual $\{A'\}$ correlation domains defined above, we may now approximate the doubles excitation vectors in Eq. (22) in the following manner:

$$M_{ij}^{\mathrm{CIS(D')}} = \sum_a R_{ai}^{\mathrm{CIS}}R_{aj}^{\mathrm{CIS}} + \sum_{A'B'k}\left(\frac{1+\delta_{A'i,B'k}}{2}\right)^{\frac{1}{2}} R_{A'iB'k}^{\mathrm{CIS(D')}}\left(\frac{1+\delta_{A'j,B'k}}{2}\right)^{\frac{1}{2}} R_{A'jB'k}^{\mathrm{CIS(D')}}, \tag{31a}$$

$$N_{ab}^{\mathrm{CIS(D')}} = \sum_i R_{ai}^{\mathrm{CIS}}R_{bi}^{\mathrm{CIS}} + \sum_{I'J'c}\left(\frac{1+\delta_{aI',cJ'}}{2}\right)^{\frac{1}{2}} R_{aI'cJ'}^{\mathrm{CIS(D')}}\left(\frac{1+\delta_{bI',cJ'}}{2}\right)^{\frac{1}{2}} R_{bI'cJ'}^{\mathrm{CIS(D')}}, \tag{31b}$$

where the prime denotes the occupied and virtual correlation domains of reduced dimensions O' and V', while the summations in the canonical basis run over the full set of orbitals. We refer to $\mathbf{M}^{\mathrm{CIS(D')}}$ and $\mathbf{N}^{\mathrm{CIS(D')}}$ as the CIS(D′) hole and particle density matrices. In analogy with Eq. (10), the CIS(D′)-NTOs may then be obtained by diagonalizing $\mathbf{M}^{\mathrm{CIS(D')}}$ and

$\mathbf{N}^{\mathrm{CIS(D')}}$. We note that while the primed indices are generally reduced compared to the full dimensions, $\mathbf{M}^{\mathrm{CIS(D')}}$ and $\mathbf{N}^{\mathrm{CIS(D')}}$ always have full molecular dimensions, and the diagonalization procedure therefore leads to O occupied and V virtual CIS(D′)-NTOs with *a priori* nonzero eigenvalues. This may be seen explicitly by considering the ranks of the CIS(D′) density matrices, which—in analogy with Eq. (13)—obey the following equations:

$$\mathrm{rank}(\mathbf{M}^{\mathrm{CIS(D')}}) \leq \min(O, V + O(V')^2) = O, \quad (32a)$$

$$\mathrm{rank}(\mathbf{N}^{\mathrm{CIS(D')}}) \leq \min(V, O + (O')^2 V) = V. \quad (32b)$$

It is seen that all the CIS(D′)-NTOs generally have nonzero eigenvalues, which can be used to quantify the physical significance of each orbital for describing the targeted transition.

The generation of CC2-NTOs and CIS(D)-NTOs is an iterative and a non-iterative $\mathcal{O}(N^5)$ process, respectively. An algorithm for generating CIS(D′)-NTOs is given in the Appendix where it is shown that the generation of CIS(D′)-NTOs has a formal $\mathcal{O}(N^3)$ scaling behaviour (assuming that the reduced dimensions O' and V' are independent of the system size for a given type of electronic transition). CIS(D′)-NTOs can thus be calculated for much larger molecular systems than CC2-NTOs or CIS(D)-NTOs.

We emphasize that even though the generation of CIS(D′)-NTOs described in this section and the Appendix is a bit technical, the end result is a set of prioritized orthogonal MOs that span the full molecular orbital space. A subset of CIS(D′)-NTOs can thus be chosen based on their eigenvalues. Furthermore, we may transform such a subset to a pseudo-canonical basis by diagonalizing the corresponding occupied-occupied and virtual-virtual blocks of the Fock matrix. These pseudo-canonical orbitals can be straightforwardly used in a standard CC code to calculate CC excitation energies. The main complexity compared to a conventional CC implementation thus occurs during the orbital generation rather than in the actual CC calculation.

E. Comparison of NTOs

In this section, we compare the different sets of NTOs introduced above. We consider dodecylbenzene and 11-*cis*-retinal protonated Schiff base (CRPSB), which are depicted in Figs. 4 and 5, respectively, together with the dominant pair of CIS-NTOs for the lowest electronic transition. All the calculations were performed using augmented basis sets of double-ζ quality as described in Sec. IV. For dodecylbenzene, it is clear that the transition is localized in a relatively small region of the molecule, while the transition in CRPSB involves almost the entire molecule. The dodecylbenzene and CRPSB cases are thus typical examples of a localized transition and a delocalized transition, respectively.

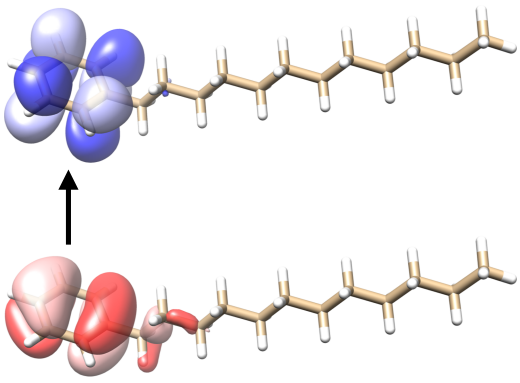

FIG. 4. Illustration of the lowest electronic transition in dodecylbenzene in terms of the dominant occupied (red, bottom) and virtual (blue, top) CIS-NTOs (aug-cc-pVDZ′ basis). The contour plot value was set to 0.02 a.u.[59]

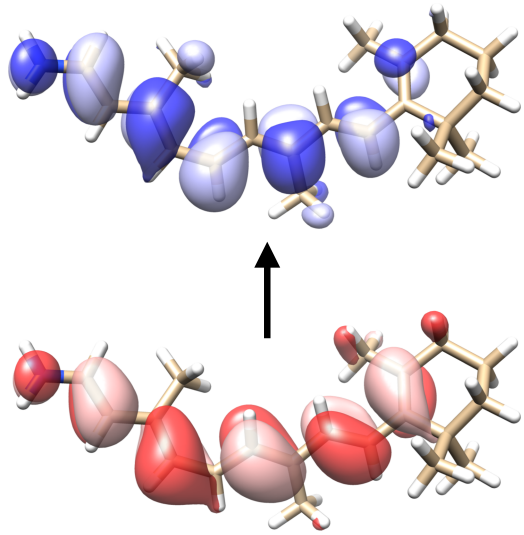

FIG. 5. Illustration of the lowest electronic transition in the 11-*cis*-retinal protonated Schiff base in terms of the dominant occupied (red, bottom) and virtual (blue, top) CIS-NTOs (aug-cc-pVDZ′ basis). The contour plot value was set to 0.02 a.u.[59]

For the CC2-, CIS(D)-, or CIS(D′)-NTOs (collectively referred to as correlated NTOs), we may list the orbitals in order of descending NTO eigenvalues. We then select a given number of orbitals from this list to generate an excitation orbital space (XOS), in which the CC2 ground state amplitude equation and CC2 eigenvalue equation are solved to yield an approximate CC2 excitation energy. In Figs. 6 and 7, we present the lowest CC2 excitation energies and associated errors for the dodecylbenzene and CRPSB molecules as a function of the number of MOs included in the XOS for the correlated NTOs. For comparison, we also give results for the mixed CIS-NTO/LMO space, which was defined in Refs. 49 and 60.

For a given XOS, errors compared to the full CC2 calculation arise since (i) the ground state amplitude equations and (ii) the CC2 eigenvalue problem in Eq. (9) are solved in the restricted XOS. We note that the errors of type (i) affect the errors of type (ii) indirectly, since the Jacobian matrix $\mathbf{A}$ in Eq. (9) is determined from ground state amplitudes.[56] All in all, the complex interplay of these two error sources ultimately leads to the errors observed in Figs. 6 and 7.

The CIS-NTO/LMO space behaves markedly different for the two molecules. For dodecylbenzene, the excitation energy error is below 0.1 eV when about 150 orbitals have been included in the XOS [see Fig. 6 (bottom)], while almost 500 orbitals are required to achieve the same accuracy for CRPSB [see Fig. 7 (bottom)]. This difference can be rationalized as follows. For the local transition in dodecylbenzene (Fig. 4), the use of LMOs leads to a significant reduction of the XOS, since a large region of the molecule is unaffected

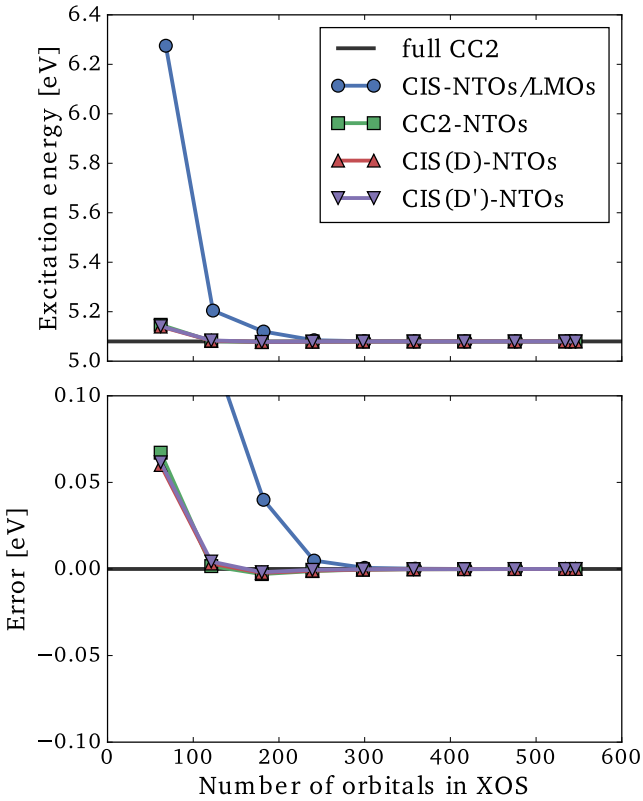

FIG. 6. Lowest CC2/aug-cc-pVDZ′ excitation energy (top) and error (bottom) as a function of the size of the excitation orbital space (XOS) for dodecylbenzene using different choices of orbital spaces.

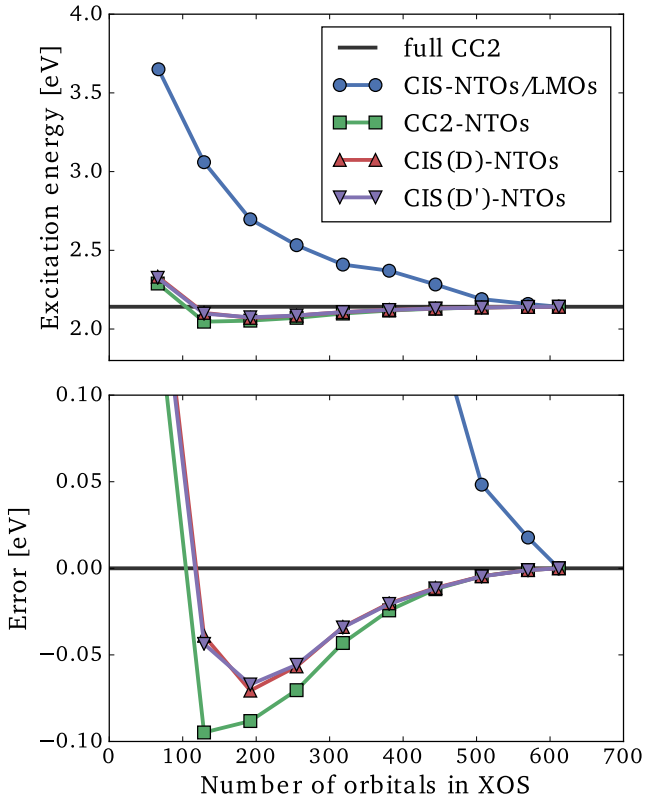

FIG. 7. Lowest CC2/aug-cc-pVDZ′ excitation energy (top) and error (bottom) as a function of the size of the excitation orbital space (XOS) for the 11-*cis*-retinal protonated Schiff base using different choices of orbital spaces.

by the transition. On the other hand, for the delocalized transition in CRPSB (Fig. 5), the transition induces changes in the electronic structure throughout the molecular system, and the use of LMOs offers very modest reductions of the XOS. These two examples clearly show the strength and weakness of the CIS-NTO/LMO approach introduced in Ref. 49, i.e., significant computational savings are possible for local transitions, while a delocalized transition essentially requires the full molecular CC calculation to be carried out.

For the local transition in Fig. 6, the use of any of the correlated NTOs leads to much smaller errors than the CIS-NTO/LMO mixed space for a given number of XOS orbitals. The correlated NTOs thus allow for a more compact representation of correlation effects than the LMOs. Even more striking is the behavior of the correlated NTOs for the delocalized transition in Fig. 7, where errors below 0.1 eV are observed with only around 100 orbitals, which is in stark contrast to the CIS-NTO/LMO curve. Thus, using correlated NTOs, it seems possible to obtain a compact representation of the orbital space for both localized and delocalized transitions.

The results in Figs. 6 and 7 obtained using the different correlated NTOs are very similar, indicating that all three sets of NTOs contain the basic information about correlation effects. This is a very important result, since the CIS(D′)-NTOs are much cheaper to generate than the CIS(D)- and CC2-NTOs (Sec. II D and the Appendix). The curves for CIS(D)- and CIS(D′)-NTOs are almost identical, and the approximations introduced in Sec. II D thus appear to be negligible for practical purposes. Both negative and positive errors occur for the correlated NTOs, and it is notable that the CIS(D′)- and CIS(D)-NTOs often yield smaller absolute errors than the CC2-NTOs even though the excitation energy is calculated at the CC2 level. This can happen due to error cancellations between the error sources (i) and (ii) mentioned above.

In summary, the dodecylbenzene and CRPSB results presented here indicate that the CIS(D′)-NTOs can be used to generate a reduced orbital space that may significantly lower the computational cost of CC2 calculations for both localized and delocalized transitions.

III. COUPLED CLUSTER EXCITATION ENERGIES USING CorNFLEx

In this section, we introduce a general framework for calculating CC excitation energies at a reduced computational cost using the concept of CIS(D′)-NTOs. We consider excitation energies calculated using the CC2 model introduced by Christiansen *et al.*,[56] where the resolution-of-the-identity (RI) approximation is used for the two-electron repulsion integrals, as introduced by Hättig and Weigend.[58] We refer to Refs. 56 and 58 for a detailed description of the CC2 model.

The main algorithm is presented in Sec. III A, while the treatment of multiple excitation energies is detailed in Sec. III B. In Sec. III C, we describe how the number of auxiliary functions in the XOS may be reduced.

A. The main CorNFLEx algorithm

The results in Sec. II E indicate that the CIS(D′)-NTOs provide a compact representation of electronic transitions at

the CC level. After solving the CIS eigenvalue problem for the full molecule, we thus propose the following procedure to reduce the computational cost of CC excitation energy calculations:

1. Perform an SVD of the CIS eigenvectors to determine the CIS-NTOs according to Eqs. (3).
2. Calculate CIS(D′) density matrices from Eqs. (31) as detailed in Sec. II D.
3. Diagonalize the CIS(D′) density matrices to determine the $\mathbf{U}^{\mathrm{CIS(D')}}$ and $\mathbf{V}^{\mathrm{CIS(D')}}$ transformation matrices that define the occupied and virtual CIS(D′)-NTOs,
$$\mathbf{M}^{\mathrm{CIS(D')}}\mathbf{U}^{\mathrm{CIS(D')}} = \boldsymbol{\lambda}^{\mathrm{CIS(D')}}\mathbf{U}^{\mathrm{CIS(D')}}, \tag{33a}$$
$$\mathbf{N}^{\mathrm{CIS(D')}}\mathbf{V}^{\mathrm{CIS(D')}} = \tilde{\boldsymbol{\lambda}}^{\mathrm{CIS(D')}}\mathbf{V}^{\mathrm{CIS(D')}}. \tag{33b}$$
4. List the CIS(D′)-NTO eigenvalues in descending order and scale them such that they add up to one (equivalent to normalizing the CIS(D′) excitation vectors),
$$\sum_p \lambda_p^{\mathrm{CIS(D')}} = \sum_p \tilde{\lambda}_p^{\mathrm{CIS(D')}} = 1. \tag{34}$$
5. Choose a subset of occupied and virtual CIS(D′)-NTOs based on the $\tau_{\mathrm{CIS(D')}}$ threshold,
$$\min_{Z_o}\left(\sum_{p=1}^{Z_o}\lambda_p^{\mathrm{CIS(D')}}\right) > 1 - \tau_{\mathrm{CIS(D')}}, \tag{35a}$$
$$\min_{Z_v}\left(\sum_{p=1}^{Z_v}\tilde{\lambda}_p^{\mathrm{CIS(D')}}\right) > 1 - \tau_{\mathrm{CIS(D')}}. \tag{35b}$$
This subset of orbitals containing Z_o occupied CIS(D′)-NTOs and Z_v virtual CIS(D′)-NTOs is denoted the excitation orbital space (XOS).
6. For the chosen CC model (CC2 in this work), solve the ground state amplitude equation and the Jacobian eigenvalue equation within the XOS to obtain the CC excitation energy corresponding to the CIS excitation vector used to generate the XOS.

We denote the approach defined by these six steps as the Correlated Natural transition orbital Framework for Low-scaling Excitation energy calculations (CorNFLEx). We stress that $\tau_{\mathrm{CIS(D')}}$ is our main threshold which controls the accuracy of the CorNFLEx calculation, e.g., in the limit where $\tau_{\mathrm{CIS(D')}} \to 0$ all orbitals are included and the full CC result is reproduced. The CorNFLEx procedure should be applied to each electronic transition of interest, and additional technical details related to the determination of multiple excitation energies are discussed in Sec. III B.

The use of CIS(D′)-NTOs implies that only single-replacement dominated electronic transitions that are described at the CIS level of theory can be addressed with CorNFLEx. However, many transitions of interest in typical organic and biological molecules are of this type. Furthermore, for the CC2 model, which is the target model in the present study, single-replacement dominated transitions are correct through second order in the fluctuation potential, while double-replacement dominated transitions are correct only to zeroth order.[61] The CC2 model thus provides accurate results only for transitions dominated by single-electron replacements. The fact that our CIS(D′)-NTOs can be generated only for such transitions is therefore not a limitation from a practical point of view.

B. Multiple excitation energies

When several excitation energies are requested, the CorNFLEx algorithm defined by the six steps in Sec. III A is applied to each CIS excitation vector of interest, starting with the lowest CIS excitation energy. For example, if the three lowest CIS vectors are determined $\{\mathbf{R}^{\mathrm{CIS}(k)}, k = 1,2,3\}$, we may calculate the corresponding three CC2 vectors, $\{\mathbf{R}^{\mathrm{CC2}(k)}, k = 1,2,3\}$, which, of course, do not necessarily correspond to the three lowest excitation energies at the CC2 level of theory. In CorNFLEx, it is thus implicitly assumed that there is a one-to-one correspondence between each CIS vector and the determined CC2 vector and care must be exercised if the CIS vectors mix considerably at the CC2 level of theory. For this reason, we consider the overlap, $S^{(k)}_{\mathrm{CIS/CC2}}$, between the CIS target vector, $\mathbf{R}^{\mathrm{CIS}(k)}$, and the final CC2 singles vector, $\mathbf{R}_1^{\mathrm{CC2}(k)}$, as a measure of the reliability of the results,
$$S^{(k)}_{\mathrm{CIS/CC2}} = |(\mathbf{R}^{\mathrm{CIS}(k)})^{\mathrm{T}}\mathbf{R}_1^{\mathrm{CC2}(k)}|^2. \tag{36}$$
If $\mathbf{R}^{\mathrm{CIS}(k)}$ and $\mathbf{R}_1^{\mathrm{CC2}(k)}$ are both normalized in the singles space, then $0 \leq S^{(k)}_{\mathrm{CIS/CC2}} \leq 1$. Thus, if $S^{(k)}_{\mathrm{CIS/CC2}}$ is much smaller than one, we recommend to reduce $\tau_{\mathrm{CIS(D')}}$ (increase the XOS) and repeat the calculation to check the validity of the calculated CC2 excitation energy. We return to this issue in Sec. IV and now turn our attention to some technical details regarding the determination of several excitation energies.

First, it is necessary to ensure that the solution of each CC2 eigenvalue equation (step 6 in Sec. III A) converges to the CC2 vector that has the largest overlap with the CIS target vector. A root homing procedure[62] is therefore applied to track the CIS vector of interest when the CC2 eigenvalue problem is solved using Davidson techniques.[63–65] Thus, in the Davidson procedure, we calculate the overlap ($S^{(k)}_{\mathrm{CIS/CC2}}$) between the CIS target vector and each of the optimal CC2 vectors generated in the current iteration. The optimal CC2 vector with the largest overlap is then used to calculate the residual vector and to extend the Davidson subspace with a new direction if the norm of the residual vector is above a convergence threshold.

Second, as an additional precaution, we project out the $k-1$ previously determined CC2 vectors from the CIS target vector k, i.e., we transform $\mathbf{R}^{\mathrm{CIS}(k)}$ in the following manner:
$$\mathbf{R}^{\mathrm{CIS}(k)} \leftarrow (\mathbf{1} - \mathbf{Q})\mathbf{R}^{\mathrm{CIS}(k)}, \tag{37a}$$
$$\mathbf{Q} = \sum_{j=1}^{k-1}\sum_{l=1}^{k-1}\mathbf{R}_1^{\mathrm{CC2}(j)}[\mathbf{Y}^{-1}]_{jl}(\mathbf{R}_1^{\mathrm{CC2}(l)})^{\mathrm{T}}, \tag{37b}$$
$$Y_{jl} = (\mathbf{R}_1^{\mathrm{CC2}(j)})^{\mathrm{T}}\mathbf{R}_1^{\mathrm{CC2}(l)}. \tag{37c}$$
This projection strategy is important when two or more CIS vectors mix considerably at the CC level of theory. For such cases, the use of Eq. (37) effectively ensures that any given CC2 vector is only determined once.

All in all, the root homing and projection strategies described in this section provide a way to calculate several excitation energies within independent XOSs.

C. Natural auxiliary functions

In Sec. III A, we have detailed a procedure to construct a reduced space of orthogonal occupied and virtual orbitals (the XOS). The MOs in the XOS are in principle expressed as a linear combination of *all* AOs which puts severe limitations on the performance of the method. However, as detailed in Appendix B of Ref. 66, the number of AOs in the XOS can be reduced by removing molecular orbital tails using a least squares fitting procedure. Similarly, for the RI approximation, we consider only the subset of auxiliary functions that are assigned to the subset of atoms defined by the AOs in the XOS (see Ref. 67 for details).

Once the XOS has been defined and the atomic and auxiliary functions have been reduced, a standard CC calculation is performed in that space. Our implementation of the CC2 model for excitation energies closely follows the RI specific algorithm described in Ref. 58 with the exception that the required tensors of dimension $N_\Gamma OV$ are kept in the core memory instead of a disk (where N_Γ denotes the number of auxiliary functions). It is thus very important to reduce the auxiliary dimension as much as possible. For that purpose, we invoke the use of natural auxiliary functions (NAFs) as introduced by Kállay in Ref. 68.

Using the RI approximation, the two-electron integrals involved in the CC2 algorithm can be written as follows:

$$(pq|rs) = \sum_{\Gamma\Upsilon}(pq|\Gamma)(\Gamma|\Upsilon)^{-1}(\Upsilon|rs) = \sum_{\Gamma\Upsilon} I^\Gamma_{pq} I^\Upsilon_{rs}, \tag{38}$$

$$I^\Gamma_{pq} = \sum_{\Upsilon} (\Gamma|\Upsilon)^{-1/2}(\Upsilon|rs), \tag{39}$$

where Γ and Υ denote the auxiliary basis functions. In our implementation, we calculate NAFs by diagonalizing the following symmetric matrix:

$$W_{\Gamma\Upsilon} = \sum_{pq} I^\Gamma_{pq} I^\Upsilon_{pq}, \tag{40}$$

where the summation runs over all possible pairs of MO indices (ai, ia, ab, ij). In Ref. 68, NAFs were obtained in a similar way but using I^Γ_{ai} or $I^\Gamma_{\alpha\beta}$ (depending on the targeted model) instead of I^Γ_{pq}. In a standard CC2 calculation, it might be more appropriate to use the $I^\Gamma_{\alpha\beta}$ tensor instead of I^Γ_{pq} since it avoids the contraction of the integrals with the MO coefficients. However, in the context of CorNFLEx, the number of MOs in the XOS is significantly smaller than the number of AOs, and using I^Γ_{pq} enables a larger reduction of the number of auxiliary functions. In our CC2 implementation, the integrals are thus stored in the form

$$I^{\tilde\Gamma}_{pq} = \sum_\Gamma (pq|\Gamma) X_{\Gamma\tilde\Gamma}, \tag{41}$$

where the coefficients $X_{\Gamma\tilde\Gamma}$—which transform a standard auxiliary function (Γ) to the corresponding NAF ($\tilde\Gamma$)—are obtained as described in Ref. 68. The number of NAFs is significantly reduced compared to the number of auxiliary functions by considering only those functions that correspond to an eigenvalue of $\mathbf{W}$ larger than a truncation threshold, τ_{NAF}. For the calculation of CC2 excitation energies, our test calculations indicate that a truncation threshold of $\tau_{\mathrm{NAF}} = 10^{-2}$ can be used without affecting the results (see Sec. IV A). This is particularly useful in the context of CorNFLEx where the number of NAFs and the memory requirements can be greatly reduced.

IV. RESULTS

In this section, we present numerical results and applications of the CorNFLEx procedure to real-life compounds and assess the performance and accuracy of the method when targeting CC2 excitation energies. As a test set, we consider the five lowest CIS transitions and the corresponding CC2 excitation energies of the following ten molecules, which are shown in Fig. 8:

- The molecules XLI, XLIV, XLIX, LI, LV, LVI, LVIII, and LX from the test set presented in Ref. 69.
- Dodecylbenzene optimized at the MP2/cc-pVDZ level using the ORCA program[70] (Cartesian coordinates are available in the supplementary material).
- 11-*cis*-retinal protonated Schiff base (CRPSB) using the geometry from Ref. 43.

The CorNFLEx CC2 excitation energies and timings for those systems are compared to full CC2 calculations in Sec. IV A. All calculations have been performed with a local version of the LSDalton program,[71,72] using the frozen core

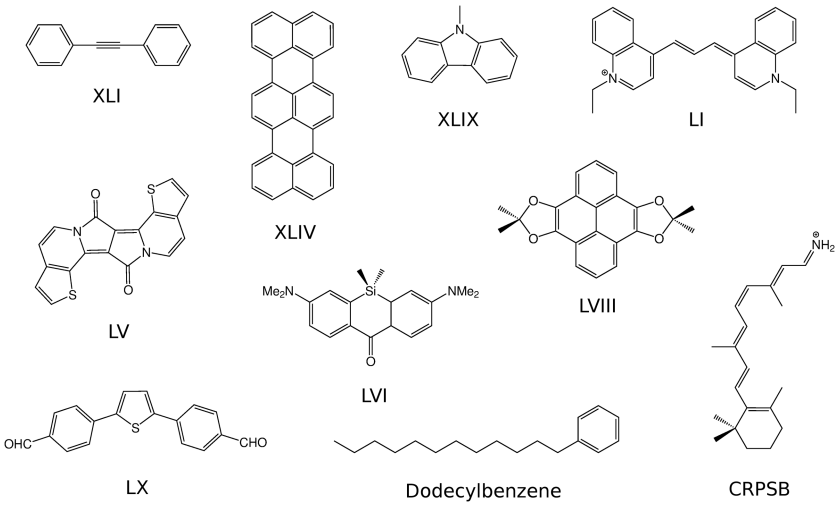

FIG. 8. Molecules contained in the test set used in Sec. IV A.

approximation. The correlation consistent aug-cc-pVDZ′ and aug-cc-pVTZ′ basis sets[73,74] were used with the corresponding auxiliary basis sets aug-cc-pVDZ-RI′ and aug-cc-pVTZ-RI′ for the RI approximation.[75] The prime in the basis set notation indicates that diffuse functions have been removed from the hydrogen atoms.

Finally, in order to demonstrate the potential of CorNFLEx, we also consider molecular clusters of solvated formamide containing up to 144 water molecules in Sec. IV B. The structures were obtained by the modification of a single snapshot taken from Ref. 76 and are available in the supplementary material.

A. Performance of CorNFLEx for the test set

In this section, we consider the five lowest excited states for the test set in Fig. 8 to compare the accuracy and computational cost of the CorNFLEx procedure described in Sec. III with full (canonical) CC2 calculations. In total, we thus consider 50 singlet electronic transitions using both augmented double- and triple-ζ quality basis sets.

In order to evaluate the convergence of the excitation energies with the size of the XOS, we have performed calculations using $\tau_{\mathrm{CIS(D')}} = 10^{-3}$, 10^{-4}, and 10^{-5}. In the supplementary material (Tables SI-SX), we report the individual errors in the excitation energies, $\Delta_i = \omega_i^{\mathrm{CorNFLEx}} - \omega_i^{\mathrm{ref}}$, while we present the mean absolute error,

$$\bar{\Delta}_{\mathrm{abs}} = \frac{1}{n}\sum_{i=1}^{n} |\Delta_i|, \tag{42}$$

the maximum absolute error,

$$\Delta_{\mathrm{max}} = \max_i |\Delta_i|, \tag{43}$$

the mean error,

$$\bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n} \Delta_i, \tag{44}$$

and the standard deviation,

$$\Delta_{\mathrm{std}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\Delta_i - \bar{\Delta})^2}, \tag{45}$$

for the aug-cc-pVDZ′ and aug-cc-pVTZ′ basis sets in Tables I and II, respectively. Here, $n = 50$ is the number of excitation energies calculated from the 10 molecules in the test set.

In general, we observe that the excitation energy errors decrease systematically with decreasing $\tau_{\mathrm{CIS(D')}}$ values for both basis sets. For example, the mean absolute errors are 0.11 eV, 0.05 eV, and 0.01 eV for $\tau_{\mathrm{CIS(D')}} = 10^{-3}$, 10^{-4}, and 10^{-5} when the aug-cc-pVTZ′ basis set is used. The mean absolute errors are small compared to typical CC2 errors, which are typically of the order of 0.2 eV.[58,77–79]

The maximum absolute errors reveal that there are some outliers. For example, using $\tau_{\mathrm{CIS(D')}} = 10^{-4}$ and the aug-cc-pVTZ′ basis set, the maximum absolute error is 0.40 eV. This error is associated with the third state of the XLIV molecule (Table SII in the supplementary material). As discussed in Sec. III B, the CorNFLEx procedure can only be expected to yield very accurate results when the CC2 singles vector is dominated by a single CIS vector. This is the case if the $S^{(k)}_{\mathrm{CIS/CC2}}$ measure in Eq. (36) is larger than 0.5. For the third state of XLIV molecule in the aug-cc-pVTZ′ basis set, we find that $S^{(3)}_{\mathrm{CIS/CC2}} = 0.42$. It is therefore not surprising that the associated error is rather large. However, by decreasing $\tau_{\mathrm{CIS(D')}}$ from 10^{-4} to 10^{-5}, the error for the considered state is effectively removed. In the lower parts of Tables I and II, we present the different error measures calculated from a subset of the 50 excitation energies where states with $S^{(k)}_{\mathrm{CIS/CC2}} \leq 0.5$ ($k = 1, 2, 3, 4, 5$) have been removed. This corresponds to removing between 2 and 4 states (out of the 50 states) and reduces most of the maximum errors quite dramatically, e.g., from 0.40 eV to 0.16 eV using $\tau_{\mathrm{CIS(D')}} = 10^{-4}$ and the aug-cc-pVTZ′ basis set.

All in all, we conclude that the CorNFLEx procedure allows us to calculate excitation energies of CC2 quality. One should, however, check the $S^{(k)}_{\mathrm{CIS/CC2}}$ diagnostics when analyzing the results. If $S^{(k)}_{\mathrm{CIS/CC2}} < 0.5$ for a given state k, it is recommended to decrease $\tau_{\mathrm{CIS(D')}}$ for that state to validate the calculated excitation energy. The exact CC2 result is obtained by reducing the $\tau_{\mathrm{CIS(D')}}$ threshold, and we will refer to the $\tau_{\mathrm{CIS(D')}}$ values of 10^{-3}, 10^{-4}, and 10^{-5} as *loose*, *standard*, and *tight*, respectively.

TABLE I. Statistical analysis of CC2 excitation energies for the test set using the aug-cc-pVDZ′ basis set. The mean absolute error ($\bar{\Delta}_{\mathrm{abs}}$), the maximum absolute error (Δ_{max}), the mean error ($\bar{\Delta}$), and the standard deviation (Δ_{std}) are reported. We also consider the same measures for the subset of CC2 states with $S^{(k)}_{\mathrm{CIS/CC2}} > 0.5$ (see text for details). All measures are given in eV.

All errors	$\tau_{\mathrm{CIS(D')}} = 10^{-3}$	$\tau_{\mathrm{CIS(D')}} = 10^{-4}$	$\tau_{\mathrm{CIS(D')}} = 10^{-5}$
$\bar{\Delta}_{\mathrm{abs}}$	0.07	0.02	0.00
Δ_{max}	0.37	0.22	0.02
$\bar{\Delta}$	−0.01	−0.01	0.00
Δ_{std}	0.10	0.04	0.01
Errors ($S^{(k)}_{\mathrm{CIS/CC2}} > 0.5$)	$\tau_{\mathrm{CIS(D')}} = 10^{-3}$	$\tau_{\mathrm{CIS(D')}} = 10^{-4}$	$\tau_{\mathrm{CIS(D')}} = 10^{-5}$
$\bar{\Delta}_{\mathrm{abs}}$	0.06	0.01	0.00
Δ_{max}	0.22	0.08	0.02
$\bar{\Delta}$	0.00	−0.01	0.00
Δ_{std}	0.08	0.02	0.01

TABLE II. Statistical analysis of CC2 excitation energies for the test set using the aug-cc-pVTZ′ basis set. The mean absolute error ($\bar{\Delta}_{\mathrm{abs}}$), the maximum absolute error (Δ_{max}), the mean error ($\bar{\Delta}$), and the standard deviation (Δ_{std}) are reported. We also consider the same measures for the subset of CC2 states with $S^{(k)}_{\mathrm{CIS/CC2}} > 0.5$ (see text for details). All measures are given in eV.

All errors	$\tau_{\mathrm{CIS(D')}} = 10^{-3}$	$\tau_{\mathrm{CIS(D')}} = 10^{-4}$	$\tau_{\mathrm{CIS(D')}} = 10^{-5}$
$\bar{\Delta}_{\mathrm{abs}}$	0.11	0.05	0.01
Δ_{max}	0.43	0.40	0.05
$\bar{\Delta}$	−0.04	−0.05	−0.01
Δ_{std}	0.16	0.08	0.01
Errors ($S^{(k)}_{\mathrm{CIS/CC2}} > 0.5$)	$\tau_{\mathrm{CIS(D')}} = 10^{-3}$	$\tau_{\mathrm{CIS(D')}} = 10^{-4}$	$\tau_{\mathrm{CIS(D')}} = 10^{-5}$
$\bar{\Delta}_{\mathrm{abs}}$	0.11	0.04	0.01
Δ_{max}	0.43	0.16	0.05
$\bar{\Delta}$	−0.04	−0.04	−0.01
Δ_{std}	0.15	0.05	0.01

In order to evaluate the computational cost of the CorNFLEx scheme, we consider the mean, maximum, and minimum speed-ups compared to a state-specific full calculation calculated as

$$\text{speed-up} = \frac{T_{\text{full}}}{T_{\text{CorNFLEx}}}. \tag{46}$$

For each of the 10 molecules, T_{full} is the time for the correlated part of the CC2 calculation, i.e., the determination of the CC2 ground state amplitudes and solution of the Jacobian eigenvalue problems for the five excited states in the complete canonical basis using the root homing procedure.[62] T_{CorNFLEx} includes both the time spent in the generation of the CIS(D′)-NTOs *and* the time spent in the CC2 calculation within the restricted XOS determined by $\tau_{\text{CIS(D′)}}$.

The timing results obtained using the aug-cc-pVDZ′ and aug-cc-pVTZ′ basis sets are given in Table III. It is seen that computational savings are obtained with all of the tested thresholds although the speed-ups of course decrease with decreasing $\tau_{\text{CIS(D′)}}$. The largest speed-ups are observed for dodecylbenzene where the XOS is very small compared to the full orbital space, since the transitions are localized on the benzene ring (see Fig. 4 for an illustration of the lowest transition). However, significant speed-ups are also obtained for the delocalized transitions. The CorNFLEx procedure thus cures the main flaw of our recently proposed LoFEx scheme where speed-ups were only obtained for localized transitions. In summary, CorNFLEx generally yields excitation energies of CC2 quality at a reduced computational cost for both localized and delocalized transitions.

Finally, we note that the CorNFLEx results in Tables I and II used the NAF approximation described in Sec. III C with $\tau_{\text{NAF}} = 10^{-2}$. In Tables SXI and SXII in the supplementary material, we present errors introduced by the NAF approximation in CorNFLEx calculations for CRPSB and dodecylbenzene. We note that the NAF errors are generally at least an order of magnitude smaller than the corresponding CorNFLEx errors (Tables SIX and SX in the supplementary material). We thus conclude that the main threshold $\tau_{\text{CIS(D′)}}$ controls the accuracy of a CorNFLEx calculation, while the NAF error is negligible. Finally, from Tables SXIII and SXIV in the supplementary material, we see that the use of NAFs reduces the number of auxiliary functions in the CorNFLEx CC2 calculations considerably compared to a full calculation.

TABLE III. Relative speed-ups of CorNFLEx calculations compared to canonical CC2 calculations for the test set using the aug-cc-pVDZ′ and aug-cc-pVTZ′ basis sets.

Basis	Measure	Loose	Standard	Tight
aug-cc-pVDZ′	Mean	5.9	3.3	1.8
	Max	19.0	11.3	5.0
	Min	1.9	1.5	1.1
aug-cc-pVTZ′	Mean	7.5	4.4	2.5
	Max	26.4	13.8	7.6
	Min	2.9	2.1	1.4

B. Solvated formamide

In order to investigate the potential of CorNFLEx for large molecular systems, we now consider the lowest excitation energy in a series of solvated formamide clusters with an increasing number of water molecules. All the calculations have been performed using the aug-cc-pVDZ′ basis set, and the CorNFLEx strategy has been applied with the loose, standard, and tight thresholds for clusters including up to 144 water molecules (4836 basis functions). Reference calculations using a conventional CC2 implementation have also been performed for clusters containing up to 63 water molecules (2163 basis functions).

In Fig. 9, we have plotted the convergence of the excitation energy as a function of the number of water molecules included in the cluster using the loose, standard, and tight CorNFLEx thresholds (explicit values are given in the supplementary material). The excitation energies obtained with conventional CIS and CC2 algorithms are also reported. Figure 9 shows that the solvation effects result in a blueshift of ≃0.8 eV at the CIS level, while the blueshift is lowered to around 0.5 eV at the CC2 level. Most of the solvation effects are already included with only 15 water molecules, and adding more waters results in small oscillations of around 0.03 eV. We note that the observed solvatochromatic shift is consistent with Ref. 76 but that a conformational sampling is required to obtain reliable results.

The excitation energies calculated using CorNFLEx are effectively of CC2 quality, even with the loose threshold. A closer look at the (absolute) errors for the different structures shows that they are at most 0.06 eV, 0.04 eV, and 0.02 eV for the loose, standard, and tight thresholds (see Table SXIII in the supplementary material). This is in accordance with the errors observed for the test set in Sec. IV A. The main correlation effect (i.e., a lowering of the blueshift compared to the CIS result) is thus well described with all of the tested thresholds.

In order to analyze the performance of the CorNFLEx strategy compared to a conventional algorithm, we report timings for the CIS calculation and for the correlated parts of the CC2 calculations in Fig. 10. For CorNFLEx, the time plotted includes the generation of the CIS(D′)-NTOs and the solution of the CC2 equations in the restricted XOS, while only the time

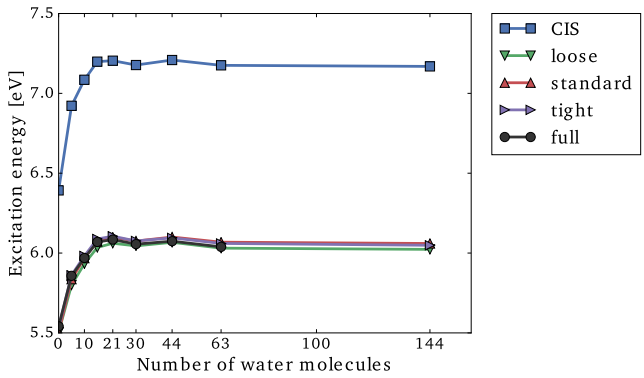

FIG. 9. Lowest excitation energy of a series of solvated formamide with an increasing number of water molecules using the aug-cc-pVDZ′ basis set. CC2-CorNFLEx results are reported using the **loose, standard,** and **tight** thresholds. For comparison, the CIS and full CC2 results are also reported.

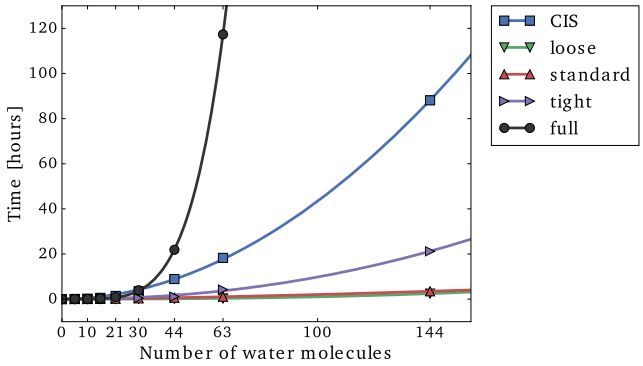

FIG. 10. Time spent in the calculation of the lowest excitation energy of a series of formamide-water clusters of increasing size using the aug-cc-pVDZ′ basis set. **CIS:** time spent in the HF ground state and CIS excitation energy calculations (prior to the CorNFLEx or full CC2 calculation). **Loose/standard/tight:** time spent in the CorNFLEx algorithm (generation of the CIS(D′)-NTOs and solution of the CC2 equations in the XOS). **Full:** solution of the CC2 equations in the full canonical basis. All calculations were performed on a single Lenovo nx360 M5 node with 28 cores @ 2.4 GHz and 256 GB memory.

for the solution of the CC2 equations is considered for the full calculation. Figure 10 shows that the CorNFLEx calculations (with all three thresholds) are systematically faster than the conventional CC2 calculations, and the speed-up grows rapidly as a function of the molecular cluster size. For example, for the cluster with 63 water molecules, the CorNFLEx calculation with the standard threshold is 111 times faster than the conventional CC2 calculation. The largest presented CorNFLEx calculations correspond to a molecular cluster containing 4836 basis functions, which is out of reach using our conventional CC2 implementation. It is also notable that the CorNFLEx calculation constitutes only a small fraction of the CIS time for the larger clusters.

V. CONCLUSION

In this work, we have introduced a new method for the calculation of CC excitation energies on large molecules denoted CorNFLEx. It relies on the generalization of the concept of NTOs to include correlation effects. The resulting CIS(D′)-NTOs are obtained based on an approximation to the CIS(D) model, and their generation formally scales cubically with the system size. Once the CIS(D′)-NTOs have been generated, a reduced orbital space can be determined by discarding a subset of occupied and virtual CIS(D′)-NTOs with low eigenvalues, and a conventional CC excitation energy calculation can then be performed in that reduced space. We have defined the reduced orbital space in terms of the loose, standard, and tight thresholds, which ultimately define the precision of a CorNFLEx calculation compared to a conventional CC2 calculation.

The CorNFLEx algorithm has been tested for the calculation of CC2 excitation energies on a set of 50 singlet electronic transitions resulting in mean absolute errors of 0.11 eV, 0.05 eV, and 0.01 eV (triple-ζ basis) with the loose, standard, and tight thresholds, respectively. Even though some of the considered molecules were of limited size and many of the transitions were nonlocal, significant computational savings were systematically obtained. We have also applied CorNFLEx to a series of solvated formamide clusters of increasing size (up to 4836 basis functions). The results indicate that our new scheme allows CC2 excitation energies to be determined at the cost of a CIS calculation and thus further extends the application range of the CC2 model.

The use of CIS(D′)-NTOs implies that only electronic transitions that are described at the CIS level of theory can be addressed (single-replacement dominated transitions). However, these include most transitions of interest in typical organic and biological molecules. We also note that the presented developments are not specific to the CC2 model as such, and more accurate CC models might also benefit from using the CorNFLEx scheme. Finally, we note that while our previously proposed LoFEx scheme only allows for computational reductions when the transitions of interest are local, the CorNFLEx procedure yields computational savings for both localized and delocalized transitions.

SUPPLEMENTARY MATERIAL

See supplementary material for the excitation energies of the test set and formamide clusters, the NAF results, and the Cartesian coordinates of dodecylbenzene and solvated formamide clusters.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (No. FP/2007-2013)/ERC Grant Agreement No. 291371. The numerical results presented in this work were performed at the Centre for Scientific Computing, Aarhus (http://phys.au.dk/forskning/cscaa/).

APPENDIX: ALGORITHM FOR GENERATING CIS(D′) DENSITY MATRICES

In this appendix, we present an algorithm for calculating the CIS(D′) density matrices in Eqs. (31). First, we note from Eq. (22) that the doubles excitation vectors occurring in Eq. (31) are given by

$$R^{\mathrm{CIS(D')}}_{A'iB'j} = \frac{1}{(1+\delta_{A'i,B'j})} \frac{(A'i|B'j)^{\mathrm{CIS}}}{\epsilon_i - F_{A'A'} + \epsilon_j - F_{B'B'} + \omega^{\mathrm{CIS}}}, \quad \text{(A1a)}$$

$$R^{\mathrm{CIS(D')}}_{aI'bJ'} = \frac{1}{(1+\delta_{aI',bJ'})} \frac{(aI'|bJ')^{\mathrm{CIS}}}{F_{I'I'} - \epsilon_a + F_{J'J'} - \epsilon_b + \omega^{\mathrm{CIS}}}, \quad \text{(A1b)}$$

where the two-electron integrals can be written as

$$(A'i|B'j)^{\mathrm{CIS}} = (\bar{A}'i|B'j) + (A'\bar{i}|B'j) + (A'i|\bar{B}'j) + (A'i|B'\bar{j}), \quad \text{(A2a)}$$

$$(aI'|bJ')^{\mathrm{CIS}} = (\bar{a}I'|bJ') + (a\bar{I}'|bJ') + (aI'|\bar{b}J') + (aI'|b\bar{J}'). \quad \text{(A2b)}$$

The lowercase indices refer to MOs in the full canonical basis, the primed capital indices denote the reduced CIS-NTO/local basis (OCD and VCD), and the barred indices are transformed

with the $\bar{\mathbf{\Lambda}}^{\mathrm{CIS}}$ matrices in Eqs. (20), e.g.,

$$(\bar{A}'i|B'j) = \sum_{\alpha\beta\gamma\delta} \bar{\Lambda}^{\mathrm{CIS}}_{\alpha A'} C_{\beta i} C_{\gamma B'} C_{\delta j} (\alpha\beta|\gamma\delta). \tag{A3}$$

As mentioned in Sec. III C, we employ the RI approximation to reduce the cost of the generation of the four-center integrals. For example, the integral in Eq. (A3) may be written as

$$(\bar{A}'i|B'j) = \sum_{\Gamma} I^{\Gamma}_{\bar{A}'i} I^{\Gamma}_{B'j}. \tag{A4}$$

We note that NAFs are not used in the generation of the CIS(D′)-NTOs, and the set of auxiliary basis functions is therefore not reduced.

Using Eqs. (A2) and (A4), the doubles excitation vectors in Eq. (A1) may be written as

$$R^{\mathrm{CIS(D')}}_{A'iB'j} = \frac{1}{(1+\delta_{A'i,B'j})} \frac{\sum_{\Gamma} P^{A'B'}_{ij} [I^{\Gamma}_{\bar{A}'i} I^{\Gamma}_{B'j} + I^{\Gamma}_{A'\bar{i}} I^{\Gamma}_{B'j}]}{\epsilon_i - F_{A'A'} + \epsilon_j - F_{B'B'} + \omega^{\mathrm{CIS}}}, \tag{A5a}$$

$$R^{\mathrm{CIS(D')}}_{aI'bJ'} = \frac{1}{(1+\delta_{aI',bJ'})} \frac{\sum_{\Gamma} P^{ab}_{I'J'} [I^{\Gamma}_{\bar{a}I'} I^{\Gamma}_{bJ'} + I^{\Gamma}_{a\bar{I}'} I^{\Gamma}_{bJ'}]}{F_{I'I'} - \epsilon_a + F_{J'J'} - \epsilon_b + \omega^{\mathrm{CIS}}}, \tag{A5b}$$

where the action of the permutation operator is defined in Eq. (16). From Eq. (A5), we see that the following integrals are required to evaluate the doubles contributions to the hole density matrix in Eq. (31a):

- $I^{\Gamma}_{A'i}$, $I^{\Gamma}_{\bar{A}'i}$, and $I^{\Gamma}_{A'\bar{i}}$,

while the particle density matrix in Eq. (31b) can be determined from the following integrals:

- $I^{\Gamma}_{aI'}$, $I^{\Gamma}_{\bar{a}I'}$, and $I^{\Gamma}_{a\bar{I}'}$.

We recall that the primed indices denote reduced sets of occupied and virtual orbitals of dimensions O' and V', respectively. The size-intensivity of excitation energies implies that O' and V' are independent of the system size for a given type of electronic transition. This will be assumed in the following.

In our current implementation, we first calculate the six integrals listed above and store them in memory. The construction of the integrals is detailed in Table IV. The first step is the construction of the $(\Gamma|\alpha\beta)$ integrals, which formally scales as $\mathcal{O}(N_\Gamma N_\alpha^2)$ where N_α and N_Γ denote the number of atomic and auxiliary basis functions, respectively. However, this scaling is effectively only quadratic with the system size if efficient integral screening techniques are employed. In step 2, one of the AO indices is transformed to a reduced MO index. These steps scale as $\mathcal{O}(N_\Gamma N_\alpha^2)$ with pre-factors O' or V'. In steps 3 and 4, the second AO index is transformed to an occupied and a virtual index, respectively, leading to formal scalings of $\mathcal{O}(N_\Gamma N_\alpha O)$ and $\mathcal{O}(N_\Gamma N_\alpha V)$. Finally, in steps 5 and 6, the auxiliary index is transformed with the $(\Gamma|\Upsilon)^{-1/2}$ matrix to provide the desired three-center integrals listed above. These last steps scale as $\mathcal{O}(N_\Gamma^2 O)$ or $\mathcal{O}(N_\Gamma^2 V)$ depending on the type of integrals. In summary, since the reduced dimensions O' or V' are independent of the system size for a given type of electronic transition, all steps involved in the generation of the integrals in Table IV scale cubically with the system size.

Once the integrals have been generated from Table IV, the doubles excitation vectors are calculated according to Eq. (A5). It is seen that the determinations of $R^{\mathrm{CIS(D')}}_{A'iB'j}$ and $R^{\mathrm{CIS(D')}}_{aI'bJ'}$ scale as $O^2 N_\Gamma$ and $V^2 N_\Gamma$ with pre-factors $(V')^2$ and $(O')^2$, respectively. Finally, the density matrices are evaluated according to Eq. (31) where the evaluation of $\mathbf{M}^{\mathrm{CIS(D')}}$ and $\mathbf{N}^{\mathrm{CIS(D')}}$ scales as O^3 and V^3 with pre-factors $(V')^2$ and $(O')^2$, respectively. In order to reduce the memory requirements, one may calculate the doubles excitation vectors in batches and contract them on the fly. Specifically, we batch over the k-index in Eq. (31a) and the c-index in Eq. (31b).

All in all, we conclude that the generation of the hole and particle density matrices scales cubically with the size of the molecular system for a given type of electronic transition. The pre-factor depends on the dimensions of the OCD and VCD (O' and V'), which depend on the type of electronic transition under consideration.

TABLE IV. Generation of two-electron RI integrals required for constructing CIS(D′) hole and particle density matrices. Batching of integrals and reordering of tensors have been omitted for clarity.

Step	Operation	Prefactor	Scaling
1	Calculate integrals $(\Upsilon\|\alpha\beta)$		$\mathcal{O}(N_\Gamma N_\alpha^2)$
2	$(\Upsilon\|A'\beta) \leftarrow \sum_\alpha C_{\alpha A'}(\Upsilon\|\alpha\beta)$ $(\Upsilon\|\bar{A}'\beta) \leftarrow \sum_\alpha \bar{\Lambda}^{\mathrm{CIS}}_{\alpha A'}(\Upsilon\|\alpha\beta)$ $(\Upsilon\|\alpha I') \leftarrow \sum_\beta C_{\beta I'}(\Upsilon\|\alpha\beta)$ $(\Upsilon\|\alpha\bar{I}') \leftarrow \sum_\beta \bar{\Lambda}^{\mathrm{CIS}}_{\beta I'}(\Upsilon\|\alpha\beta)$	V' V' O' O'	$\mathcal{O}(N_\Gamma N_\alpha^2)$
3	$(\Upsilon\|A'i) \leftarrow \sum_\beta C_{\beta i}(\Upsilon\|A'\beta)$ $(\Upsilon\|A'\bar{i}) \leftarrow \sum_\beta \bar{\Lambda}^{\mathrm{CIS}}_{\beta i}(\Upsilon\|A'\beta)$ $(\Upsilon\|\bar{A}'i) \leftarrow \sum_\beta C_{\beta i}(\Upsilon\|\bar{A}'\beta)$	V'	$\mathcal{O}(N_\Gamma N_\alpha O)$
4	$(\Upsilon\|aI') \leftarrow \sum_\alpha C_{\alpha a}(\Upsilon\|\alpha I')$ $(\Upsilon\|\bar{a}I') \leftarrow \sum_\alpha \bar{\Lambda}^{\mathrm{CIS}}_{\alpha a}(\Upsilon\|\alpha I')$ $(\Upsilon\|a\bar{I}') \leftarrow \sum_\alpha C_{\alpha a}(\Upsilon\|\alpha\bar{I}')$	O'	$\mathcal{O}(N_\Gamma N_\alpha V)$
5	$I^{\Gamma}_{A'i} \leftarrow \sum_\Upsilon (\Gamma\|\Upsilon)^{-1/2}(\Upsilon\|A'i)$ $I^{\Gamma}_{A'\bar{i}} \leftarrow \sum_\Upsilon (\Gamma\|\Upsilon)^{-1/2}(\Upsilon\|A'\bar{i})$ $I^{\Gamma}_{\bar{A}'i} \leftarrow \sum_\Upsilon (\Gamma\|\Upsilon)^{-1/2}(\Upsilon\|\bar{A}'i)$	V'	$\mathcal{O}(N_\Gamma^2 O)$
6	$I^{\Gamma}_{aI'} \leftarrow \sum_\Upsilon (\Gamma\|\Upsilon)^{-1/2}(\Upsilon\|aI')$ $I^{\Gamma}_{\bar{a}I'} \leftarrow \sum_\Upsilon (\Gamma\|\Upsilon)^{-1/2}(\Upsilon\|\bar{a}I')$ $I^{\Gamma}_{a\bar{I}'} \leftarrow \sum_\Upsilon (\Gamma\|\Upsilon)^{-1/2}(\Upsilon\|a\bar{I}')$	O'	$\mathcal{O}(N_\Gamma^2 V)$

[1] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).
[2] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: Many-Body Perturbation Theory and Coupled-Cluster Theory* (Cambridge University Press, Cambridge, UK, 2009).
[3] H. J. Monkhorst, Int. J. Quantum Chem. **12**, 421 (1977).
[4] E. Dalgaard and H. J. Monkhorst, Phys. Rev. A **28**, 1217 (1983).
[5] J. Olsen and P. Jørgensen, J. Chem. Phys. **82**, 3235 (1985).
[6] H. Koch and P. Jørgensen, J. Chem. Phys. **93**, 3333 (1990).
[7] A. E. Kondo, P. Piecuch, and J. Paldus, J. Chem. Phys. **104**, 8566 (1996).
[8] K. Emrich, Nucl. Phys. A **351**, 397 (1981).
[9] H. Sekino and R. J. Bartlett, Int. J. Quantum Chem. **26**, 255 (1984).
[10] J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).
[11] A. I. Krylov, Annu. Rev. Phys. Chem. **59**, 433 (2008).
[12] L. González, D. Escudero, and L. Serrano-Andrés, ChemPhysChem **13**, 28 (2012).
[13] A. J. Cohen, P. Mori-Sanchez, and W. Yang, Science **321**, 792 (2008).
[14] A. Dreuw and M. Head-Gordon, J. Am. Chem. Soc. **126**, 4007 (2004).

[15]R. K. Nesbet, Adv. Chem. Phys. **IX**, 321 (1965).
[16]S. Sæbø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[17]C. Riplinger, P. Pinski, U. Becker, E. Valeev, and F. Neese, J. Chem. Phys. **144**, 024109 (2016).
[18]H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, J. Chem. Theory Comput. **11**, 484 (2015).
[19]Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, J. Chem. Phys. **139**, 094105 (2013).
[20]P. R. Nagy, G. Samu, and M. Kállay, J. Chem. Theory Comput. **12**, 4897 (2016).
[21]W. Li, Z. Ni, and S. Li, Mol. Phys. **114**, 1447 (2016).
[22]T. Kjærgaard, P. Baudin, D. Bykov, J. J. Eriksen, P. Ettenhuber, K. Kristensen, J. Larkin, D. Liakh, F. Pawłowski, A. Vose, Y. M. Wang, and P. Jørgensen, Comput. Phys. Commun. **212**, 152 (2017).
[23]J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[24]S. F. Boys, Rev. Mod. Phys. **32**, 296 (1960).
[25]C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963).
[26]B. Jansík, S. Høst, K. Kristensen, and P. Jørgensen, J. Chem. Phys. **134**, 194104 (2011).
[27]I.-M. Høyvik, B. Jansík, and P. Jørgensen, J. Chem. Phys. **137**, 224114 (2012).
[28]I.-M. Høyvik and P. Jørgensen, Chem. Rev. **116**, 3306 (2016).
[29]P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[30]J. Yang, Y. Kurashige, F. R. Manby, and G. K.-L. Chan, J. Chem. Phys. **134**, 044123 (2011).
[31]P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955).
[32]C. Edmiston, J. Chem. Phys. **45**, 1833 (1966).
[33]F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).
[34]T. D. Crawford and R. A. King, Chem. Phys. Lett. **366**, 611 (2002).
[35]T. Korona and H.-J. Werner, J. Chem. Phys. **118**, 3006 (2003).
[36]D. Kats, T. Korona, and M. Schütz, J. Chem. Phys. **125**, 104106 (2006).
[37]D. Kats, T. Korona, and M. Schütz, J. Chem. Phys. **127**, 064107 (2007).
[38]D. Kats and M. Schütz, J. Chem. Phys. **131**, 124117 (2009).
[39]B. Helmich and C. Hättig, J. Chem. Phys. **139**, 084114 (2013).
[40]B. Helmich and C. Hättig, Comput. Theor. Chem. **1040-1041**, 35 (2014).
[41]T. D. Crawford, in *Recent Progress in Coupled Cluster Methods: Theory and Applications*, edited by P. Cársky, J. Paldus, and J. Pittner (Springer, Netherlands, Dordrecht, 2010), Vol. 53, Chap. 2, pp. 37–55.
[42]R. A. Mata and H. Stoll, J. Chem. Phys. **134**, 034122 (2011).
[43]A. K. Dutta, F. Neese, and R. Izsák, J. Chem. Phys. **145**, 034102 (2016).
[44]R. H. Myhre, A. M. J. Sánchez de Merás, and H. Koch, Mol. Phys. **111**, 1109 (2013).
[45]R. H. Myhre, A. M. J. Sánchez de Merás, and H. Koch, J. Chem. Phys. **141**, 224105 (2014).
[46]R. H. Myhre and H. Koch, J. Chem. Phys. **145**, 044111 (2016).
[47]M. Caricato, T. Vreven, G. W. Trucks, and M. J. Frisch, J. Chem. Phys. **133**, 054104 (2010).
[48]M. Caricato, T. Vreven, G. W. Trucks, and M. J. Frisch, J. Chem. Theory Comput. **7**, 180 (2011).
[49]P. Baudin and K. Kristensen, J. Chem. Phys. **144**, 224106 (2016).
[50]R. Send, V. R. I. Kaila, and D. Sundholm, J. Chem. Phys. **134**, 214114 (2011).
[51]A. Kumar and T. D. Crawford, J. Phys. Chem. A **121**, 708 (2017).
[52]M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee, Chem. Phys. Lett. **219**, 21 (1994).
[53]A. V. Luzanov, A. A. Sukhorukov, and V. E. Umanskii, Theor. Exp. Chem. **10**, 354 (1976).
[54]R. L. Martin, J. Chem. Phys. **118**, 4775 (2003).
[55]J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, J. Phys. Chem. **96**, 135 (1992).
[56]O. Christiansen, H. Koch, and P. Jørgensen, Chem. Phys. Lett. **243**, 409 (1995).
[57]I.-M. Høyvik, R. H. Myhre, and H. Koch, J. Chem. Phys. **146**, 144109 (2017).
[58]C. Hättig and F. Weigend, J. Chem. Phys. **113**, 5154 (2000).
[59]E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, J. Comput. Phys. **25**, 1605 (2004).
[60]In Ref. 49, we used TDHF-NTOs instead of CIS-NTOs, but this difference has a very minor effect on the presented results.
[61]K. Hald, P. Jørgensen, J. Olsen, and M. Jaszuński, J. Chem. Phys. **115**, 671 (2001).
[62]W. Butscher and W. E. Kammer, J. Comput. Phys. **20**, 313 (1976).
[63]E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[64]E. R. Davidson, Comput. Phys. Commun. **53**, 49 (1989).
[65]C. W. Murray, S. C. Racine, and E. R. Davidson, J. Comput. Phys. **103**, 382 (1992).
[66]P. Ettenhuber, P. Baudin, T. Kjærgaard, P. Jørgensen, and K. Kristensen, J. Chem. Phys. **144**, 164116 (2016).
[67]P. Baudin, P. Ettenhuber, S. Reine, K. Kristensen, and T. Kjærgaard, J. Chem. Phys. **144**, 054102 (2016).
[68]M. Kállay, J. Chem. Phys. **141**, 244113 (2014).
[69]D. Jacquemin, I. Duchemin, and X. Blase, J. Chem. Theory Comput. **11**, 5340 (2015).
[70]F. Neese, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 73 (2012).
[71]See http://daltonprogram.org for LSDalton, a linear-scaling molecular electronic structure program, Release Dalton 2016 (2015).
[72]K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenaes, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnaes, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawłowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. M. J. Sánchez de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziółkowski, and H. Ågren, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 269 (2014).
[73]T. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[74]R. Kendall, T. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6769 (1992).
[75]F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).
[76]K. Sneskov, T. Schwabe, J. Kongsted, and O. Christiansen, J. Chem. Phys. **134**, 104108 (2011).
[77]H. Koch, O. Christiansen, P. Jørgensen, and J. Olsen, Chem. Phys. Lett. **244**, 75 (1995).
[78]M. R. Silva-Junior, S. P. A. Sauer, M. Schreiber, and W. Thiel, Mol. Phys. **108**, 453 (2010).
[79]R. Send, V. R. I. Kaila, and D. Sundholm, J. Chem. Theory Comput. **7**, 2473 (2011).