JCTC Journal of Chemical Theory and Computation

Quantum Electronic Structure

Reduced-scaling correlation methods for the excited states of large molecules: implementation and benchmarks for the second-order algebraic-diagrammatic construction approach

David Mester, Péter R. Nagy, and Mihaly Kallay

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.9b00735 • Publication Date (Web): 25 Sep 2019

Reduced-scaling correlation methods for the excited states of large molecules: Implementation and benchmarks for the second-order algebraic-diagrammatic construction approach

Dávid Mester,* Péter R. Nagy, and Mihály Kállay*

Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, H-1521 Budapest, P.O.Box 91, Hungary

E-mail: mester.david@mail.bme.hu; kallay@mail.bme.hu

Abstract

A framework for the reduced-scaling implementation of excited-state correlation methods is presented. An algorithm is introduced to construct excitation-specific local domains, which include all the important molecular orbitals for the excitation as well as for the electron correlation. The orbital space dimensions of the resulting compact domains are further decreased utilizing our reduced-cost techniques developed previously [*J. Chem. Phys.* **148**, 094111 (2018)] based on the natural auxiliary function and local natural orbital approaches. Additional methodological improvements for the evaluation of density matrices are also discussed. Benchmark calculations are presented at the second-order algebraic-diagrammatic construction level. Compared to our reduced-cost algorithm significant, up to 3–9-fold speedups are achieved even for systems of

*To whom correspondence should be addressed

smaller than 100 atoms. At the same time the additional errors introduced by the domain approximations are highly acceptable being about 2–4 meV on the average. The presented reduced-scaling algorithm allows us to carry out correlated excited-state calculations using triple-ζ basis sets with diffuse functions for systems of up to 400 atoms or 13000 atomic orbitals in a matter of days using an 8-core processor.

1 Introduction

The interest in excited-state calculations of molecular systems has been significantly increasing over the past decades, which encourages the development of more and more efficient quantum chemical approaches for excited states. Accordingly, many theories have become accessible over the time to investigate excited-state and transition properties, such as the widely-used time-dependent density functional theory (TD-DFT),[1,2] as well as the simple semi-empirical[3–6] and the more complex electron correlation[7–29] wave function-based methods. Nowadays, the most popular approaches are the various TD-DFT methods since their computational costs are relatively low, while the results obtained are adequate in most cases. Such methods can be routinely applied to molecular systems of greater than 100 atoms, while excitation energies can be recovered, in average, within 0.3 eV using the best functionals. However, the drawbacks of TD-DFT are well-known for some challenging cases,[30,31] such as Rydberg and charge transfer (CT) states, or $\pi \rightarrow \pi^*$ excitations of conjugated systems.

In general, more consistent and accurate results can be expected from the propagator-based[13,14] and coupled-cluster (CC) approaches utilizing the equation-of-motion[7,8] (EOM) or the linear-response[9–11] theories. Using such methods arbitrary accuracy can be attained, nevertheless, the computational demand of the high-accuracy methods imposes serious limitations in practice. Appropriate compromises between calculation time and accuracy are the simplest methods arising from the aforementioned theories, the second-order algebraic-diagrammatic construction [ADC(2)][25,32] approach and the approximate second-order CC singles and doubles (CC2) method,[16–18] respectively. While the first effective implementa-

tion and extensive applications of the latter were presented by Hättig and co-workers,[19–21] the former was mostly studied in detail by Dreuw *et al.*[33–37] Both methods are iterative and scale as the fifth power of the systems size, however, it should be noted that, despite their similar numerical accuracy, the ADC(2) method is more efficient. Since the secular matrix for ADC(2) is Hermitian, one should solve only one system of equations to compute excitation energies and transition properties.

Taking into account their scaling, the upper limit of the applicability of the ADC(2) and CC2 approaches is around 50 atoms or 2000 basis functions. To make them competitive with the less robust and accurate TD-DFT methods[38–40] for more extended systems, effective approximations are needed to reduce computation times and storage requirements. The most severe bottlenecks in the calculations are the evaluation and the storage of the four-center electron repulsion integrals (ERIs) and the calculation of the intermediates including them. Several approaches, such as the density fitting (DF) or the Cholesky-decomposition (CD) techniques, avoid this problem by forming ERIs or intermediates in an alternative way. In the DF scheme developed by Shavitt,[41] Whitten,[42] and Dunlap[43] the four-center ERIs are written as the products of two- and three-center integrals using an auxiliary (fitting) basis set, which greatly facilitates the storage of integrals. Two further related methods worth mentioning are the tensor hypercontraction[44–46] and the natural auxiliary function (NAF)[47] schemes, which go beyond the DF approximation. Both of them have been successfully extended to excited-state calculations.[48–50] In the CD approach introduced by Koch *et al.*[51,52] the four-center ERI tensor is decomposed, and the Cholesky-vectors giving negligible contributions are disregarded. The benefits of the CD approach were also demonstrated for CC2.[53–55] Another simple technique for reducing the costs of correlated excited-state methods is the restricted virtual space approach, where the high-lying canonical virtual molecular orbitals (MOs) are neglected. This approach was also tested at the ADC(2) and CC2 levels.[56–60]

The approximations discussed above do not use or contain any information about the

excited state. Significantly more efficient methods can be developed by determining the MOs that play an important role in the excitation. One of the most popular schemes is the natural orbital (NO) approximation,[61–63] with which the MO space where the equations are solved can be effectively reduced. In the NO approach, a one-particle density matrix, which is formed using a lower-level wave function, is diagonalized, and the orbitals with significant importance are selected from the resulting NOs. The approach is widely utilized for ground-state calculations,[64–67] and after a few early attempts, its importance for excited-state theories started to increase recently.[49,50,68–71] The developed approaches are not only suitable for relatively cheap methods, such as ADC(2) and CC2 but could also extend the applicability of higher-order *ab initio* methods to medium-sized molecules.

Further computational savings can be achieved if one takes advantage of the locality of the MOs.[72,73] In this case, not only the time required for the calculations is decreased, but at the same time the scaling of the methods is also reduced. The first excited-state local approaches were presented by Korona and Werner[74] and Crawford *et al.*,[75] who generalized the ground-state local CC singles and doubles (CCSD) method of Werner and co-workers[76] to EOM-CCSD. In the local EOM-CCSD method developed by the former authors, the doubles amplitudes were restricted using the information by inspecting the configuration interaction singles (CIS) wave function,[74] which idea has been taken over in several subsequent studies. Thereafter, Korona, Schütz, Kats, and their co-workers developed various excited-state CC methods using local approaches.[77–82] In later publications the development of local CC2 and ADC(2) methods was reported,[77,83] which were also extended to the calculation of molecular properties[78] and improved with Laplace transform techniques.[79–82] Parallel to those efforts, further papers were published by Russ and Crawford about the calculation of excited-state properties.[84,85] Promising results were also obtained by Hättig *et al.* extending the pair natural orbital (PNO) approach to excited-state theories.[86–89] The chain of spheres exchange[90] and the back transformed PNO based[91–93] approaches developed by Izsák *et al.*, and the state-averaged PNO-EOM-CCSD scheme[94] of Valeev *et al.* represent further

cost-reducing ideas in this category. The recent local framework for calculating excitation energies (LoFEx)[95–97] and the correlated natural transition orbital framework (CorNFLEx)[71] approaches of Baudin and Kristensen introduce somewhat different strategies. The latter is an encouraging combination of the NO and the local approaches, where the reduced domains of the MOs are constructed by analyzing an approximate second-order density matrix and considering distance criteria for the orbitals. A comprehensive study was recently published on the topic of reduced-cost approximations by Crawford, Kumar, and co-workers.[98]

In this paper we introduce a new approach to reduce the computation time for correlated excited-state calculations. Our recent reduced-cost technique[50] is combined with a state-dependent local domain construction scheme, which also reduces the scaling of the algorithm. It is worth emphasizing that the domain construction is completely automatic and free from any distance-based parameters. The proposed algorithm adopts to the complexity of both the ground- and the excited-state wave function. We discuss the most important considerations about the domain assembly, and further improvements are presented for the density matrix construction with respect to our latest work. The errors introduced by the approximations are assessed in detail in various benchmark comparisons at the ADC(2) level. Finally, we carry out calculations for extended molecular systems, which have not been possible so far.

2 Theory and implementation

2.1 Domain construction

The basic assumption of the theory presented here is that only a subset of molecular orbitals contribute dominantly to an excited-state wave function, and the number of these orbitals does not increase with the size of the system. Consequently, it is a good approximation to evaluate the corresponding matrix elements within domains compiled from such orbitals. In order to find an ideal domain construction algorithm, one important consideration should be kept in mind. Since it is highly advantageous to solve the ground- and excited-state

equations in the same basis, the domain should contain all the MOs which are required for the adequate description of both states. Accordingly, for the excited state the domain has to contain all the occupied and virtual MOs involved in the excitation. However, as these orbitals can be far apart, it is desirable to augment the MO list with other orbitals that are spatially close to the former ones. While the first step is essential for the excited state, the latter is required for the accurate calculation of the ground-state wave function. To facilitate the following discussion the most frequently used indices are collected in Table 1.

Table 1: Summary of index notations

$\mu, \nu, \ldots$	atomic orbitals (AOs)
$p, q, \ldots$	general (quasi-)canonical orbitals
$i, j, \ldots$	(quasi-)canonical occupied orbitals
$a, b, \ldots$	(quasi-)canonical virtual orbitals
$i', j', \ldots$	occupied localized molecular orbitals (LMOs)
$a', b', \ldots$	projected atomic orbitals (PAOs)
$P, Q, \ldots$	auxiliary functions

Following the early ideas of Korona and co-workers,[74,77] to select the important MOs involved in the excitation, we first solve the following CIS eigenvalue problem:

$$\sigma^{\mathrm{CIS}} = \mathbf{A}^{\mathrm{CIS}}\mathbf{c} = \omega_{\mathrm{CIS}}\mathbf{c}, \tag{1}$$

where $\mathbf{A}^{\mathrm{CIS}}$ is the CIS Jacobian, ω_{CIS} stands for the CIS excitation energy, $\mathbf{c}$ contains the corresponding singles coefficients, and σ^{CIS} is the so-called CIS sigma vector. It is important to mention that this step does not require any additional calculation because the CIS or another lower-level solution is anyway required for the demanding correlation calculations as a starting guess. Since the CIS wave function is only necessary as an initial guess and to determine the dominant orbitals, approximations can be introduced that speed up the CIS calculations, such as the recently developed local fitting approach.[99] After the CIS calculation the canonical occupied MOs are localized using the Boys algorithm, while, to

span the virtual space, projected atomic orbitals (PAOs)[72] are constructed as

$$|a'\rangle = \left(\mathbf{1} - \sum_{i'} |i'\rangle\langle i'|\right)|\mu\rangle. \tag{2}$$

The CIS coefficients are transformed to the bases obtained as $c_{i'}^{a'} = \sum_{ia} C_{ii'} C_{aa'} c_i^a$, where $C_{ii'}$ ($C_{aa'}$) stands for the MO coefficient of the occupied LMO (PAO) basis. Subsequently, we determine the order of the orbitals characterizing the importance of their contribution to the wave function. To that end, motivated by Refs. 74 and 77, the norm of each column and row of matrix $c_{i'}^{a'}$ is evaluated for all LMOs (PAOs), and these values are sorted into ascending order. Starting from the largest one the squares of the norms are summed up until the sum becomes larger than a predefined threshold T_{LMO} (T_{PAO}). The selected LMOs (PAOs) form the $\mathcal{P}_1(i')$ [$\mathcal{P}_2(a')$] domain. Obviously, the size of the domains can be arbitrarily controlled with the threshold, furthermore, if the corresponding threshold is set to 1.0, all the orbitals will be selected.

It is important to keep in mind that this procedure selects the important orbitals only for the CIS wave function. If the overlap of the CIS and the single excitation part of the final wave function is relatively small, it may be required to augment the domains with additional orbitals. This problem was also realized by Kats and Schütz in the their local CC2 approach,[79] and it was resolved by the on-the-fly extension of the orbital domains in the course of the diagonalization. Here, we propose a simpler, *a priori* extension scheme. It can be assumed that the occupied orbitals involved in the excitations are close to each other for a particular state, for example, they can be found on a chromophore group. Analogously, the similar can be supposed for the virtual orbitals. Thus, if the domains are supplemented with the environment of the selected orbitals, presumably all the important orbitals will be chosen for the excitation. The orbital list extension procedure is illustrated in Fig. 1 and performed as follows. First, a loose Boughton–Pulay (BP) atom list[100] is defined for all the LMOs using the T_{BPol} completeness criteria. If this value is sufficiently chosen, only a few

atoms are selected on which the corresponding LMO is practically localized. Thereafter, we inspect for each LMO that is not included in the domain $\mathcal{P}_1(i')$ whether its BP domain has a common atom with the BP atom list of the LMOs already included in the $\mathcal{P}_1(i')$ list. This descriptor can be considered as a coarse-grained overlap measure of the two LMOs, hence it provides a system specific, wave function-based tool and allows us to avoid the use of distance-based criteria. If their overlap is significant with the LMOs of $\mathcal{P}_1(i')$ in the form of a common atom, the domain is augmented with the corresponding LMO. The $\mathcal{P}_1(i')$ extended with the surrounding LMOs will be denoted by $\mathcal{P}_3(i')$. A similar process is executed for the PAO domain. In this case, the atom domain assigned to a PAO consists of just one atom, the atom on which AO μ of Eq. (2) is located. For simplicity, this atom will be referred to as the central atom of the PAO. For each PAO included in the $\mathcal{P}_2(a')$ list, all the other PAOs are added to the domain which have central atoms common with PAOs included previously. The extended domains will be denoted by $\mathcal{P}_4(a')$. It is easy to see that $\mathcal{P}_1(i')$ $[\mathcal{P}_2(a')]$ is a subset of the $\mathcal{P}_3(i')$ $[\mathcal{P}_4(a')]$ domain, and the latter contains information about the environment of the orbitals involved in the excitation.

The excitations can also take place between two distant parts of the system. In this case, the $\mathcal{P}_3(i')$ and $\mathcal{P}_4(a')$ domains can be very far from each other. If only the selected orbitals were used in the calculations, the ground state correlation energy and amplitudes would be close to zero, and of course, the excitation energy would also be highly inaccurate. Accordingly, further supplementation of the domains is carried out, which is an important advancement in our scheme compared to the related models.[71,95] For that purpose we select those PAOs which are close to the LMOs already included in the domain, and *vice versa*. At this point we could use both the corresponding concise $[\mathcal{P}_1(i'), \mathcal{P}_2(a')]$ and extended $[\mathcal{P}_3(i'), \mathcal{P}_4(a')]$ lists. Our numerical experience shows that the smaller domains are sufficient for this purpose. Thus, in domain $\mathcal{P}_5(a')$ all the PAOs are collected whose central atom can be found in the BP atom list of the LMOs of the $\mathcal{P}_1(i')$ list. Analogously, we inspect the central atoms of the PAOs from the $\mathcal{P}_2(a')$ list. If at least one of these atoms can be found in

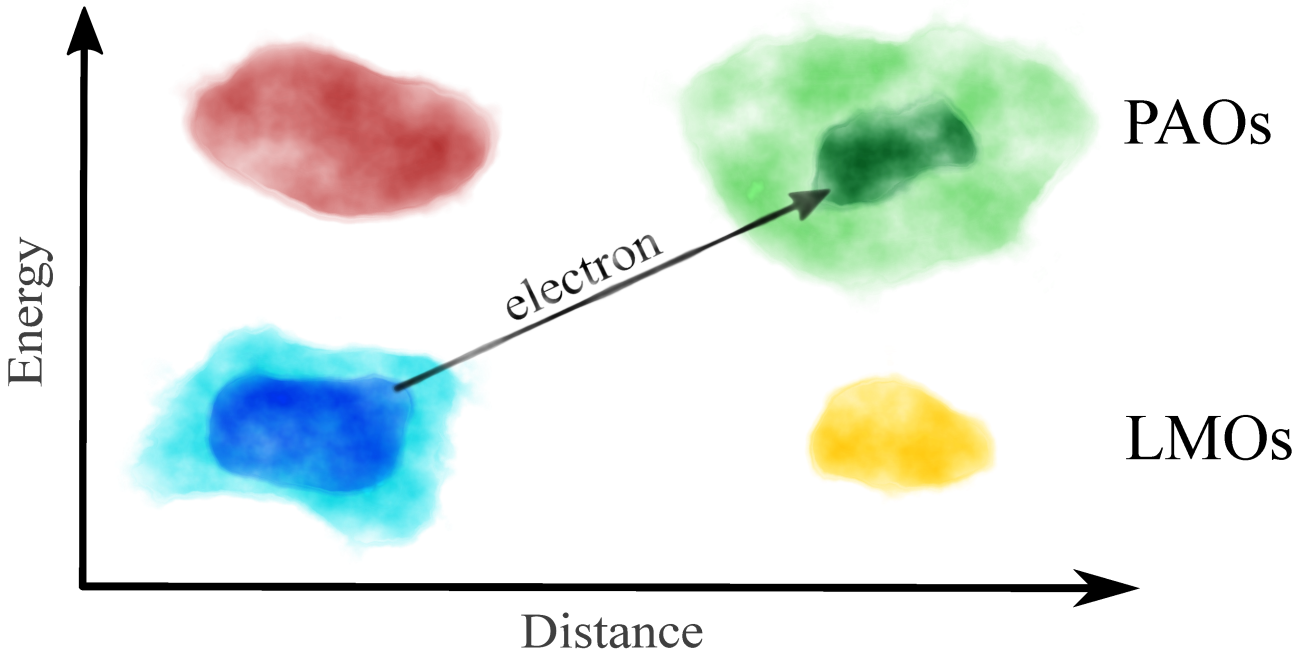

Figure 1: Illustration of the domain construction scheme. The dark blue, dark green, light blue, light green, red, and yellow colors refers to the $\mathcal{P}_1(i'), \mathcal{P}_2(a'), \mathcal{P}_3(i'), \mathcal{P}_4(a'), \mathcal{P}_5(a')$, and $\mathcal{P}_6(i')$ domains, respectively.

the BP domain of any LMO, the latter is added to the $\mathcal{P}_6(i')$ list. The final domain, which presumably includes all the important orbitals for the excitation and the electron correlation, is formed as the union of the compiled lists: $\mathcal{P}_{\mathrm{f}}(i', a') = \mathcal{P}_3(i') \cup \mathcal{P}_4(a') \cup \mathcal{P}_5(a') \cup \mathcal{P}_6(i')$. It is important to note that, in practice, there may be a significant overlap among the $\mathcal{P}_1(i'), \mathcal{P}_2(a'), \mathcal{P}_3(i'), \mathcal{P}_4(a'), \mathcal{P}_5(a')$, and $\mathcal{P}_6(i')$ domains.

The AO and auxiliary bases are also restricted to a smaller part of the molecular system in order to achieve reduced scaling. These restrictions are again made system specifically to ensure the accurate representation of the MOs and integrals required in the domain. To exploit the locality of the LMOs in the integral transformation steps, each of the LMOs of $\mathcal{P}_3(i') \cup \mathcal{P}_6(i')$ is projected onto a BP domain constructed with a significantly tighter BP criterion (T_{BPot}). For more details we refer to the documentation of our ground state local correlation approaches.[101,102] The union of such extended atom lists defines the initial atom and correspondingly the AO list of the domain. In our experience the complete AO list of

the domain is also sufficient to accurately expand the PAOs of the domain, in spite of the fact that the PAOs are significantly more delocalized. In rare cases, PAOs located at the edge of the domain are truncated too severely to be useful in the forthcoming correlation energy calculations. In our ground-state schemes those handful of PAOs are simply discarded from the higher level treatment, which can be afforded because their correlation energy contribution would be negligible. We adopt this approach for the PAOs of $\mathcal{P}_4(a')$ and $\mathcal{P}_5(a')$. In order to retain the most important PAOs of $\mathcal{P}_2(a')$ in the domain, if necessary at all, the atom list of the domain is extended with the most important BP atoms of the $\mathcal{P}_2(a')$ PAOs (controlled by T_{BPp}), yielding the final domain atom and AO lists. Analogously, for the accurate and efficient density fitting of each LMO-PAO pair density of the domain, the auxiliary functions are used that are placed on the atoms of the union of the T_{BPol} atom lists of the LMOs included in $\mathcal{P}_3(i') \cup \mathcal{P}_6(i')$.

The quasi-canonical MO basis construction of the domain follows closely the scheme of our ground-state methods.[101,103] In brief, the truncated LMOs are re-orthogonalized in the metric of the domain's AO basis, and then they are canonicalized utilizing the projection of the Fock matrix onto the AO basis of the domain. The PAOs are also projected onto the entire AO basis of the domain. The resulting functions are orthogonalized, the possible quasi-linear-dependency of this basis is removed, and then all the PAOs are canonicalized within the domain. This procedure yields the occupied (i) and virtual (a) quasi-canonical MO bases of the domain. The required occupied-virtual and occupied-occupied three-center two-electron integrals are then constructed relying on our highly-optimized, integral-direct, low-scaling integral transformation implementation as discussed in Refs. 101 and 102.

We note that, as mentioned above, we use Boys LMOs, which do not preserve the separation of the σ- and π-orbitals. Of course, the domain construction algorithm outlined in this subsection can be applied with other types of LMOs, such as Pipek–Mezey[104] or intrinsic bond orbitals,[105] which are free of this issue. We prefer Boys orbitals because our experience shows that they are somewhat more localized than the aforementioned alternatives resulting

in smaller domains and shorter computation times.[67,101–103] On the other hand, the mixing of σ- and π-orbitals does not worsen the accuracy of the computed spectral properties or increase the computation time as the surrounding σ- and π-orbitals are anyway selected by our scheme.

2.2 Utilizing the NAF and NO approximations in the domain

Since the procedure performed in the domain is very similar to the algorithm described in detail in our previous publications,[49,50] only the most important formulas and modifications will be discussed in this subsection. In the DF approach the four-center ERIs can be approximated with the

$$(pq|rs) = \sum_{Q} J^{Q}_{pq} J^{Q}_{rs}\,, \tag{3}$$

expression, and the **J** quantities are constructed from two- and three-center two-electron integrals, $(P|Q)$ and $(pq|P)$, respectively, in the form of

$$J^{Q}_{pq} = \sum_{P} (pq|P)(P|Q)^{-1/2}, \tag{4}$$

where $(P|Q)^{-1/2}$ is the corresponding element of the inverse square root of the two-center integral matrix. The NAF approach[47] is a very efficient tool to reduce the size of the auxiliary basis for excited-state calculations.[49,50,71] Although the approximation is not based on any physical or chemical consideration, it is very similar to the well-known natural orbital approach. It utilizes the singular value decomposition of **J** to construct a reduced auxiliary basis. In practice, it is more favorable to compute the singular vectors as the eigenvectors of matrix **W** with elements

$$W_{PQ} = \sum_{pq} J^{P}_{pq} J^{Q}_{pq}\,. \tag{5}$$

The eigenvectors of this matrix, the so-called NAFs, are the right singular vectors of matrix **J**, and the eigenvalues are the squares of the singular values. Using a predefined threshold,

$\varepsilon_{\mathrm{NAF}}$, the less important NAFs can be dropped to get a more compact representation of **J**. It is easy to see that the **J** matrix transformed to the compressed basis is the best approximation to the original **J**.

In the NO approximation the virtual-virtual block of the one-particle density matrix,

$$D_{ab} = \langle \Psi | a^{+} b^{-} | \Psi \rangle \,, \tag{6}$$

is constructed and diagonalized, where a^{+} and b^{-} are creation and annihilation operators, as well as Ψ denotes a lower-level wave function. The eigenvectors of this matrix are the virtual natural orbitals (VNOs), while its eigenvalues are interpreted as the corresponding occupation numbers of the NOs. The VNOs with smaller occupation numbers usually give a smaller contribution to the wave function. Accordingly, using a predefined truncation threshold denoted by $\varepsilon_{\mathrm{VNO}}$, the less important VNOs can be selected and disregarded. Our previous studies[49,50] have shown that VNOs derived from state-averaged density matrices are highly suitable for the calculation of excitation energies and transition properties. This density matrix is formed as $\mathbf{D} = (\mathbf{D}^{\mathrm{MP2}} + \mathbf{D}^{\mathrm{CIS(D)}})/2$, where $\mathbf{D}^{\mathrm{MP2}}$ and $\mathbf{D}^{\mathrm{CIS(D)}}$ denote the density matrices obtained from the second-oder Møller–Plesset (MP2) and the CIS with perturbative second-order correction [CIS(D)] wave functions. The virtual-virtual block of the approximate one-particle MP2 density matrix in a spatial orbital basis can be written in the

$$D_{ab}^{\mathrm{MP2}} = 2 \sum_{ijc} t_{ij}^{ca} t_{ij}^{cb} \tag{7}$$

form, where the above first-order amplitudes, t_{ij}^{ab}, are given as

$$t_{ij}^{ab} = \frac{(ia|jb)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} = \frac{(ia|jb)}{D_{ij}^{ab}} = \frac{\sum_Q J_{ia}^Q J_{jb}^Q}{D_{ij}^{ab}} \,, \tag{8}$$

with ε_i (ε_a) as the occupied (virtual) orbital energy. One of the bottlenecks of the MP2 density calculation is the assembly of the $(ai|bj)$ integral list, which scales as $n_{\mathrm{occ}}^2 n_{\mathrm{virt}}^2 n_{\mathrm{aux}}$,

where n_{occ}, n_{virt}, and n_{aux} stand for the number of occupied orbitals, virtual orbitals, and auxiliary functions, respectively. It can be seen that the computational requirements of the above operation can be linearly reduced by decreasing the number of auxiliary functions. If the three-center two-electron integrals represented in the NAF basis are employed in the expression, the density can be calculated with arbitrary precision by changing a threshold, which is different from $\varepsilon_{\mathrm{NAF}}$ and will be denoted by $\varepsilon_{\mathrm{NAFd}}$. As we will see it in Sect. 3.2, our numerical experience shows that about half of the NAFs can be dropped without any significant inaccuracy in the final results. Accordingly, the computation times required for the density construction can be about halved.

The CIS(D) density matrix is formed as the sum of the density matrices obtained from the CIS wave function and its second-order perturbative correction. The latter, the approximate $\mathbf{D}^{(\mathrm{D})}$ contribution, analogously to the MP2 density, is defined as

$$D_{ab}^{(\mathrm{D})} = 2 \sum_{ijc} c_{ij}^{ca} c_{ij}^{cb}. \tag{9}$$

Here, the CIS(D) doubles coefficient, c_{ij}^{ab}, is given as

$$c_{ij}^{ab} = \frac{\sum_c [(ac|bj) c_i^c + (ai|bc) c_j^c] - \sum_k [(kj|ai) c_k^b + (ki|bj) c_k^a]}{D_{ij}^{ab} + \omega_{\mathrm{CIS}}} = \frac{V_{ab}^{ij} + V_{ba}^{ji}}{D_{ij}^{ab} + \omega_{\mathrm{CIS}}}, \tag{10}$$

where we have introduced a shorthand notation

$$V_{ij}^{ab} = \sum_Q J_{bj}^Q Y_{ai}^Q = \sum_Q J_{bj}^Q \left(\sum_c J_{ac}^Q c_i^c - \sum_j J_{ij}^Q c_j^a \right). \tag{11}$$

In general, it can be stated that the computation and storage of the occupied-virtual and occupied-occupied blocks of the $\mathbf{J}$ integrals are feasible in the main memory. However, it is not true for the more demanding virtual-virtual block. In order to avoid the expensive index transformation from the AO to the MO basis and the unfavorable disk input/output operations, an integral-direct route is followed for the calculation of intermediate Y_{ai}^Q. To

that end, the first term in the definition of the intermediate is recast in the

$$Y_{ai}^{Q} \leftarrow \sum_{c} J_{ac}^{Q} c_{i}^{c} = \sum_{c\mu\nu} C_{\mu a} C_{\nu c} J_{\mu\nu}^{Q} c_{i}^{c} \tag{12}$$

form. First, we perform the half-transformation of the CIS coefficient by $C_{\nu c}$ and then the contraction of the AO integral list with the resulting half-transformed coefficient. At this operation, the sparsity of the half-transformed coefficient matrix can be exploited utilizing that Eq. (9) is invariant to the unitary transformation of the occupied indices. If the occupied index is transformed to the LMO basis, a restricted domain can be constructed for each occupied orbital inspecting the $c_{i'}^{\nu} = \sum_i C_{ii'} c_i^{\nu}$ coefficients, and the transformation in Eq. (12) can only be carried out for the AOs of the domain. This restricted domain for a given LMO contains solely the AOs of atoms for which at least one AO has a large element in the coefficient matrix and gives significant contribution to intermediate $\mathbf{Y}$. To select the corresponding atom list for LMO i', the square of the matrix elements belonging to AOs on the given atom are summed. If this value is larger than a predefined threshold, $\varepsilon_{\mathrm{MOd}}$, the AOs of the selected atom are added to the domain. It is easy to see that the size of the domain can be arbitrarily controlled with the threshold. In addition, the NAF approach can also be utilized for the calculation of intermediate $\mathbf{V}$. The auxiliary index of three-index quantities, that is, integrals J_{bj}^{Q} and intermediates Y_{ai}^{Q}, can be replaced by NAFs. Unlike in the previous case, at this step we have found it advantageous to construct NAFs that are optimal for both the $\mathbf{J}$ and the $\mathbf{Y}$ matrices. Therefore matrix $\mathbf{W}$ is constructed with elements $W_{PQ} = \sum_{ai}(J_{ai}^{P} J_{ai}^{Q} + Y_{ai}^{P} Y_{ai}^{Q})$ and the $2\varepsilon_{\mathrm{NAFd}}$ truncation threshold is applied. Nevertheless, similar to the MP2 density, half of the NAFs can be safely neglected also in this case.

Having the state-averaged density matrix for the corresponding excited state at hand, the matrix is diagonalized and the NOs are canonicalized. Subsequently, the integrals are transformed to the NO basis, which can be performed much more economically as the size

of the NO basis is significantly smaller than that of the original MO basis. The final NAFs for the excited-state calculations are formed at this point using all the occupied-occupied, occupied-virtual, and virtual-virtual blocks of the NO integral list. The three-center quantities expressed in the compact NO and NAF bases can be easily stored in the main memory, and the calculation can be performed without any modification in the canonical code.

2.3 General algorithm

To conclude this section we overview our general algorithm for the present reduced-scaling approach.

1. Solve Hartree–Fock equations
2. Localize orbitals using Boys algorithm, construct the Boughton–Pulay atom lists
3. Solve CIS equations for all the excited states using our integral-direct local-fitting algorithm,[99] transform the CIS wave function to the LMO/PAO basis
4. Loop over excited states
    - 4.a. Analyze the CIS wave function, select LMOs and PAOs important for the excited state to construct domains $\mathcal{P}_1(i')$ and $\mathcal{P}_2(a')$
    - 4.b. Augment the domains with LMOs and PAOs important for the correlation utilizing the BP atoms lists to construct domains $\mathcal{P}_3(i')$, $\mathcal{P}_4(a')$, $\mathcal{P}_5(a')$, and $\mathcal{P}_6(i')$, and their union, the excited-state-dependent local domain $\mathcal{P}_\mathrm{f}(i', a')$
    - 4.c. Diagonalize the Fock matrix within the $\mathcal{P}_\mathrm{f}(i', a')$ domain to get the excited-state-specific quasi-canonical MOs
    - 4.d. Integral transformation to compute the occupied-virtual and occupied-occupied three-center two-electron integrals in the quasi-canonical MO basis

4.e. Calculate the MP2 and CIS(D) density with the aid of the NAF approach, diagonalize the density matrices to construct the NO basis of the domain, truncate the NO basis

4.f. Diagonalize the Fock matrix within the truncated NO basis to construct the state-specific canonicalized NO basis

4.g. Transform the MO indices of the three-center integrals to the canonicalized NO basis

4.h. Calculate NAFs in the canonicalized NO basis and transform the auxiliary function index of the three-center integrals to the final NAF basis

4.i. Solve the excited-state problem within the state-specific canonicalized NO/NAF basis

End loop

The algorithm presented is similar but contains essential changes compared to our previous reduced-cost scheme.[49,50] First, the CIS problem is solved utilizing our effective and almost error-free local-fitting approximation.[99] Second, a state-dependent local domain is constructed. The operations performed in the domain are very similar to the steps of the reduced-cost approach, but some further approximations have been introduced for the density matrix calculation. That is, the NAF approach is exploited at the calculation of the densities, and, in addition, an LMO-based AO domain is constructed for the intermediate calculation utilizing the sparsity of the CIS coefficients in the domain. The virtual-virtual block of the three-center integrals is not constructed explicitly for the densities, it is first calculated in the small NO basis.

Notice that, in our scheme, each excited state is computed independently with a state-specific MO and auxiliary function basis. Consequently, the resulting excited-state wave functions will not be orthogonal to each other, which also means that the computation of transition moments between two excited states would require further considerations.

3 Results

3.1 Computational details

The new approach has been implemented in the MRCC suite of quantum chemical programs and will be available in the next release of the package.[106]

Since a CT excitation between two distant systems would probably be the biggest challenge for domain construction, we chose a well-known example from the literature and customized it to inspect the errors introduced by the MO space reduction. The basic idea originates from the paper of Dreuw *et al.*,[31] where a CT transition of an ethylene – tetrafluoroethylene system was studied. If the separation is large enough, more than around 10 Å, the matrix elements for a ground-state correlation calculation between the occupied orbitals on one of the subunits and the virtual orbitals on the other subunit are practically zero. From this point of view, the chosen example is fortunate, however, the final domain would contain the entire molecules as they are rather small. Therefore, to be able to study the effects of the truncation of the MO space and to determine the truncation parameters, an undecane skeleton was connected to both subunits. The resulting tridec-1-ene – 1,1,2-trifluorotridec-1-ene system, hereafter referred to as the CT system, is satisfactory in all aspects. The coordinates are available in the supplementary material. For this test system Dunning's correlation consistent triple-ζ basis set (cc-pVTZ) were used,[107,108] and the corresponding auxiliary bases developed by Weigend *et al.* were employed.[109–111]

To benchmark the cutoff parameters for the density calculation with the selected MO space truncation thresholds, the phenothiazine-isoalloxazine dyad (dyad for short)[77] was chosen. This molecule is one of the smallest ones from our benchmark set but large enough to test the effects. In addition, its four lowest excited states include several important types of excitations. In these calculations, the triple-ζ AO and auxiliary basis sets augmented with diffuse functions (aug-cc-pVTZ)[109,112] were applied.

With the selected truncation parameters, further benchmark calculations were performed

to assess the errors and speedups for real-life compounds. For this purpose, a set of realistic systems containing 51-127 atoms was assembled. These widely-studied molecules were taken from the literature,[49,50,77–81,95,96,99,113] and most of them are real challenges for local excited-state methods since they have Rydberg or CT excitations, as well as conjugated or delocalized electronic structures. Our test set includes the two borondipyrromethene-flavin dyads [Flv(a) and Flv(b)],[78] the above dyad molecule,[77] a perylene bisimide derivative (bisimide derivative),[113] leupeptin,[95] met-enkephalin,[95] D21L6,[80] and the pyrene-phenothiazine-isoalloxazine triad (triad).[77] For these calculations, the aug-cc-pVTZ basis set was applied as well. To demonstrate the applicability of the method, additional calculations were carried out for even larger systems including up to 400 atoms and 13000 AOs. Two of the selected compounds (C60Im–ZnP–BDP and WW-6 dye) play important roles in photovoltaics,[114–116] bivalirudin is a notable synthetic polypeptide in biochemistry,[96] and the hydrated formamide (FA) model is an excellent system to study the effects of the explicit solvation.[71] The structure of the molecules is depicted in the supplementary material. For the Zn atom in the solar cell dyes, the auxiliary functions for the def2-QZVPPD[117] basis were used. The sizes of the investigated systems are collected in Table 2.

In the excited-state calculations the core orbitals were kept frozen. The oscillator strengths (f) were evaluated in the dipole length approximation. The reported computation times are wall-clock times determined on a machine with 128 GB of main memory and an 8-core 1.7 GHz Intel Xeon E5-2609 v4 processor.

3.2 Convergence with the truncation thresholds

First, using the constructed CT system, convergence tests were performed at the ADC(2) level to determine the threshold parameters used for the domain construction. The structure of the model system is presented in Fig. 2. Since the primary aim of these studies was to identify the errors introduced by the domain approximations, the NO and NAF approximations were not utilized in these calculations. The BP parameters for the LMO (T_{BPot}) and

Table 2: The size of the systems studied and the number of the basis functions.

Molecule	Number of atoms	Number of AOs	Number of auxiliary functions
Flv(a)	51	2001	4506
Dyad	53	2051	4639
Bisimide derivative	60	2346	5280
Leupeptin	68	2254	4928
Met-enkephalin	75	2649	5891
CT system	78	1556	3819
Flv(b)	78	2829	6288
D21L6	98	3412	7590
Triad	127	4650	10383
C60Im–ZnP–BDP	202	8097	18431
Bivalirudin	293	10304	22778
WW-6 dye	311	10604	23445
FA@144 H_2O	438	13455	28968

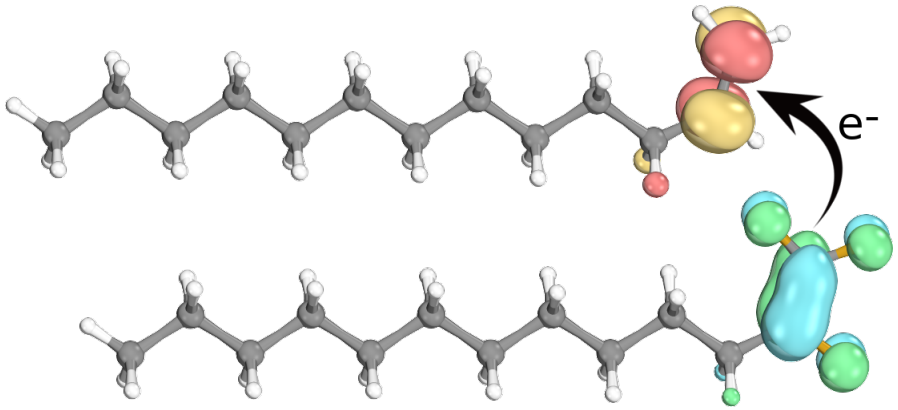

Figure 2: The CT model system and the orbitals involved in the CT excitation studied.

PAO (T_{BPp}) atom lists were carefully selected for our ground-state local correlation methods,[101,102] and no circumstances require their modification for the excited-state approach. Accordingly, we employ the values determined in our previous studies, $T_{\mathrm{BPot}} = 0.9999$ and $T_{\mathrm{BPp}} = 0.98$. In addition, the loose LMO atom list threshold, T_{BPol}, is only necessary to select the most important atoms, so the $T_{\mathrm{BPol}} = 0.95$ value is expected to be appropriate. Consequently, systematic benchmarks are presented only for two of the so far introduced

thresholds, for the most important ones, namely T_{LMO} and T_{PAO}.

Unfortunately, the effects of these two parameters are difficult to examine separately in practice. The primary purpose of the T_{LMO} (T_{PAO}) threshold is to control the number of important LMOs (PAOs), which are required to properly describe the excitation (see the scheme of Sect. 2.1). However, it is obvious that the number of the LMOs (PAOs) included in domain $\mathcal{P}_6(i')$ [$\mathcal{P}_5(a')$] depends on the value of T_{PAO} (T_{LMO}) and the size of domain $\mathcal{P}_4(a')$ [$\mathcal{P}_3(i')$]. In other words, if all PAOs are selected to domain $\mathcal{P}_4(a')$, regardless of the value of T_{LMO}, all LMOs would be added to the final domain. This is required to accurately evaluate the correlation contribution of the orbitals. The same statements can be made about the relationship of domain $\mathcal{P}_3(i')$ and the final number of the PAOs. Nevertheless, the role of the two thresholds are much less coupled for two distant systems, and their effects can be discussed separately for such cases. For the selected model system, the excitation occurs from the trifluorotridecene unit to the tridecene unit. If one sets T_{LMO} to 1.0, all the LMOs and PAOs are selected on the trifluorotridecene unit, however, the number of orbitals on the other unit can be controlled arbitrarily. In this case, the size of $\mathcal{P}_2(a')$ is directly influenced by threshold T_{PAO}, but, of course, this also affects the additional selected orbitals in domains $\mathcal{P}_4(a')$ and $\mathcal{P}_6(i')$. On the other hand, if T_{PAO} is set to 1.0, all the LMOs and PAOs are selected on tridecene, but the number of the selected LMOs, PAOs, as well as the sizes of the domains $\mathcal{P}_1(i')$, $\mathcal{P}_3(i')$, and $\mathcal{P}_5(a')$ can be varied with threshold T_{LMO} on the trifluorotridecene unit. The errors with respect to the canonical ADC(2) calculation and the sizes of the domains are visualized in Fig. 3.

Inspecting the plots in both cases we can observe that the decrease of the errors, apart from a short interval, are monotonic in the entire range. Considering that the error should be less than an order of magnitude smaller than the intrinsic error of the ADC(2) method, which is 0.2 to 0.3 eV, the T_{PAO} and T_{LMO} thresholds must necessarily be equal to or tighter than 0.92 and 0.999, respectively. Relying on the results of further numerical experiments and threshold combinations, we recommend $T_{\mathrm{PAO}} = 0.94$ and $T_{\mathrm{LMO}} = 0.999$ as the default

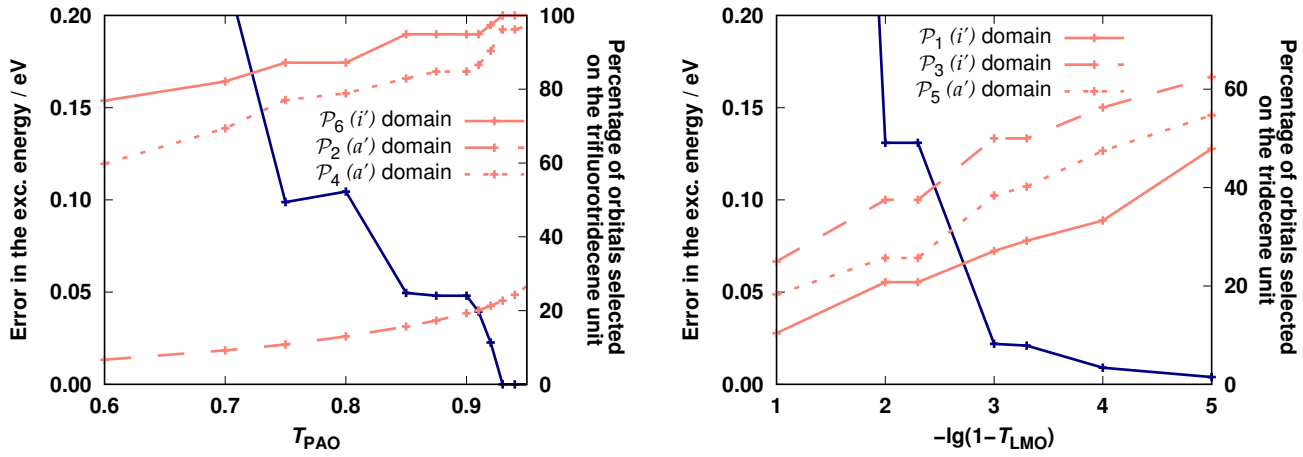

Figure 3: Error of the ADC(2) excitation energy (blue) and the size of the domains (red) as a function of the corresponding truncation threshold. A charge transfer state with the canonical ADC(2) excitation energy of 12.03 eV was selected as reference. See the text for further details.

truncation parameters, as well as $T_{\mathrm{LMO}} = 0.9999$ as a tight threshold for occupied orbitals. This choice ensures that the error in the excitation energies introduced by the LMO and PAO truncation is sufficiently small with the default thresholds, while the tight value for the LMO truncation allows the calculation of more accurate properties. For our CT system, with $T_{\mathrm{PAO}} = 0.94$, only 24% of the PAOs are selected based on the CIS excitation vector, but they are taken from many atoms. As a result, unfortunately, almost all the PAO orbitals are included in domain $\mathcal{P}_4(a')$, and all the LMOs are selected to the domain $\mathcal{P}_6(i')$. Hence, the PAO truncation on trifluorotridecene unit is practically error-free. In contrast, significant gains come from neglecting LMOs on the tridecene unit. At $T_{\mathrm{LMO}} = 0.999$ (0.9999), 27% (33%) of the LMOs are included in domain $\mathcal{P}_1(i')$. This is supplemented to 50% (56%) based on the BP domains, and 38% (47%) of the PAOs are selected for the correlation. The truncation of the AO and auxiliary bases seems to be very conservative because it does not have any significant effect on the error introduced. With the default (tight) threshold, if both the PAOs and LMOs are restricted, the error in the excitation energy is 0.02 (0.01) eV.

The selection of the appropriate $\varepsilon_{\mathrm{MOd}}$ and $\varepsilon_{\mathrm{NAFd}}$ parameters for the density calculation requires further numerical analysis. For this study, the dyad molecule was chosen, and the

NAF and VNO approaches within the domain were utilized with the default $\varepsilon_{\mathrm{NAF}} = 0.1$ a.u. and $\varepsilon_{\mathrm{VNO}} = 7.5 \times 10^{-5}$ cutoff thresholds for the NAF and VNO selection, respectively. The results are visualized in Fig. 4. Inspecting the plots we can observe that the errors decrease

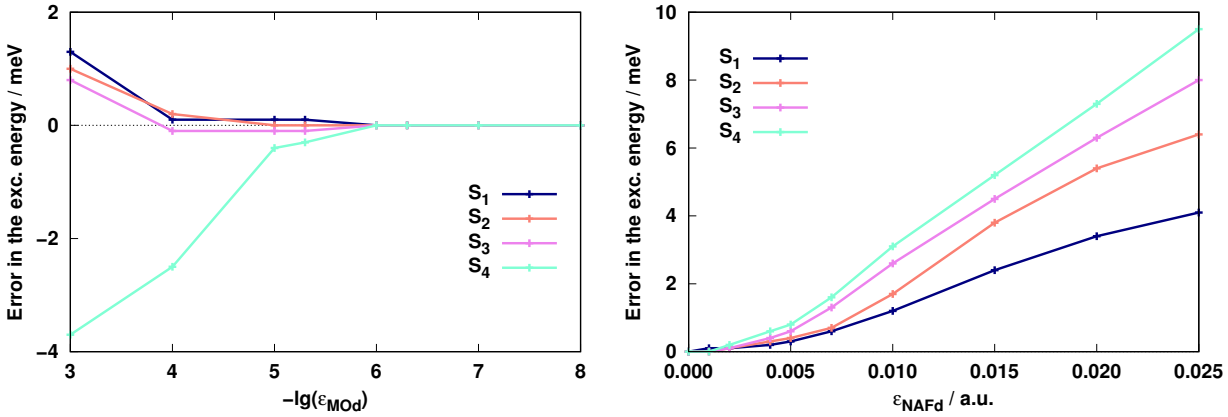

Figure 4: Error of the ADC(2) excitation energy as a function of the $\varepsilon_{\mathrm{MOd}}$ (left panel) and $\varepsilon_{\mathrm{NAFd}}$ (right panel) truncation thresholds for the four lowest excited states of the dyad molecule. The reference is the case where no further approximations are used for the density construction.

continuously with tightening the parameters. The use of the LMO-based AO domains instead of the full AO list in Eq. (12) becomes practically error-free at the $\varepsilon_{\mathrm{MOd}} = 10^{-6}$ value for all the excitations of this test system. Note also that the evaluation of intermediate $\mathbf{Y}$ does significantly not contribute to the total wall-clock time since the rate-determining step in Eq. (12) scales as $n_{\mathrm{AO}}^2 n_{\mathrm{occ}} n_{\mathrm{aux}}$, where n_{AO} stands for the number of AOs. For these reasons the conservative $\varepsilon_{\mathrm{MOd}} = 10^{-7}$ threshold is suggested. With this $\varepsilon_{\mathrm{MOd}}$ choice, the AO basis can be compressed by about 20% for the system investigated. Presumably, this gain can be more favorable for larger systems. Although the errors are somewhat larger in the case of the NAF approximation, they are still under 10 meV in the inspected range. In order to minimize the error introduced by the NAF approach, the $\varepsilon_{\mathrm{NAFd}} = 0.005$ a.u. cutoff parameter is proposed. With the selected threshold, the errors do not exceed 1 meV, while the percentage of the dropped NAFs are around 55 and 65% for the MP2 and CIS(D) density, respectively. We expect that the number of operations required for Eqs. (8) and (11) can be halved with negligible errors in general since the NAF approximation is fairly system-independent.

3.3 Benchmark calculations

Further benchmark calculations were performed for the selected realistic molecular systems at the ADC(2) level. These systems of 51 to 127 atoms are primarily important in the field of photochemistry. We attempted to inspect all important types of excitations, such as valence excited states ($n \rightarrow \pi^*$, $\sigma \rightarrow \pi^*$, and $\pi \rightarrow \pi^*$), Rydberg, as well as charge transfer excitations. In order to compare the errors introduced by the reduced-cost and the reduced-scaling algorithms, extensive benchmark calculations were carried out using different truncation thresholds. For the reduced-cost algorithm, $\varepsilon_{\mathrm{VNO}} = 7.5 \times 10^{-5}$ and $\varepsilon_{\mathrm{NAF}} = 0.1$ a.u. thresholds were used as default values, while $\varepsilon_{\mathrm{VNO}} = 1.5 \times 10^{-5}$ and $\varepsilon_{\mathrm{NAF}} = 0.1$ a.u. were applied in the tight case. In line with our previous results, we have found that the NAF approximation is practically error-free, so further tightening of the NAF threshold is not necessary. In the case of the reduced-scaling calculations, the $\varepsilon_{\mathrm{VNO}} = 7.5 \times 10^{-5}$ and $\varepsilon_{\mathrm{NAF}} = 0.1$ a.u. thresholds were used to obtain the VNOs and final NAFs for both the default and tight calculations. For the MP2 and CIS(D) density construction the $\varepsilon_{\mathrm{NAFd}} = 0.005$ a.u. parameter was used in every case. The default and tight calculations only differ in the domain construction. For the former, the $T_{\mathrm{LMO}} = 0.999$ and $T_{\mathrm{PAO}} = 0.94$ cutoff parameters were applied, while $T_{\mathrm{LMO}} = 0.9999$ and $T_{\mathrm{PAO}} = 0.94$ were used for the latter.

The errors of the excitation energies and the corresponding oscillator strengths with respect to the best estimates are presented in Table 3. The statistical error measures given in the table are the mean deviation (MD), the mean absolute deviation (MAD), and the maximum absolute deviation (MAX). First, we discuss the results obtained with the reduced-cost algorithm, because the relevant performance analysis was not part of our previous study.[50] In this case, all the errors are highly acceptable. The MD (MAD) is 5 (12) meV for the excitation energies using the default thresholds, while the MAX does not exceed 40 meV. The oscillator strengths, except for the S_4 state of the D21L6 molecule, are practically error-free: their MD is zero, while their MAD (MAX) is 0.002 (0.022). The errors obviously decrease by tightening the VNO threshold. The MAD and MAX are halved for the excitation

Table 3: Reference ADC(2) excitation energies (ω_{ref}, in eV), oscillator strengths (f_{ref}), the error of excitation energies ($\delta\omega$, in eV) and oscillator strengths (δf) with various approaches using the aug-cc-pVTZ basis set.

				Reduced-cost algorithm				Reduced-scaling algorithm			
				Tight		Default		Tight		Default	
Molecule	Character	ω_{ref} [a]	f_{ref} [a]	$\delta\omega$	δf	$\delta\omega$	δf	$\delta\omega$	δf	$\delta\omega$	δf
Flv(a)	$\pi \rightarrow \pi^*$	2.593	0.294	0.000	0.002	0.008	0.008	0.003	0.015	0.006	0.029
	$\pi \rightarrow \pi^*$	2.863	0.185	-0.003	0.000	0.005	0.002	-0.003	0.014	0.001	0.020
	$n, \sigma \rightarrow \pi^*$	3.207	0.000	-0.004	0.000	0.012	0.000	0.016	0.000	0.017	0.000
	$n, \sigma \rightarrow \pi^*$	3.319	0.001	-0.008	0.000	-0.009	0.000	-0.008	0.000	0.001	0.000
Dyad	$\pi \rightarrow \pi^*$	2.939	0.170	-0.003	0.002	0.005	-0.001	-0.002	0.002	-0.001	0.003
	CT	3.150	0.006	-0.002	0.000	0.017	0.000	0.016	0.000	0.016	0.000
	$n, \sigma \rightarrow \pi^*$	3.312	0.001	-0.006	0.000	0.002	0.000	-0.010	0.000	-0.006	0.000
	$n, \sigma \rightarrow \pi^*$	3.376	0.001	-0.006	0.000	0.012	0.000	0.016	0.000	0.025	0.000
Bisimide der.	$\pi \rightarrow \pi^*$	2.464	0.686	0.002	-0.001	0.016	-0.003	-0.004	0.005	-0.004	0.006
	$\pi \rightarrow \pi^*$	3.415	0.000	0.003	0.000	0.018	0.000	0.028	0.000	0.028	0.000
	$\pi \rightarrow \pi^*$	3.636	0.000	0.000	0.000	0.015	0.000	0.006	0.000	0.006	0.000
	$\pi \rightarrow \pi^*$	3.670	0.019	0.002	0.000	0.020	0.002	0.027	0.007	0.042	0.025
Leupeptin	Rydberg	4.065	0.001	-0.013	0.000	-0.025	0.000	-0.019	0.000	-0.011	0.000
	Rydberg	5.123	0.001	-0.006	0.000	-0.005	0.000	0.007	0.000	0.025	0.000
	Rydberg	5.137	0.002	-0.009	-0.001	-0.006	-0.001	0.003	-0.001	0.033	0.000
	Rydberg	5.289	0.003	-0.008	0.000	-0.007	0.000	-0.005	0.000	0.011	0.000
Met-enkephalin	Rydberg	4.749	0.018	0.002	0.000	0.013	0.001	0.012	0.001	0.017	0.001
	Rydberg	4.998	0.000	0.004	0.000	0.020	0.000	0.019	0.000	0.025	0.000
	Rydberg	5.303	0.000	-0.003	0.000	-0.040	0.000	-0.036	0.000	-0.026	0.000
	Rydberg	5.351	0.000	-0.014	0.000	-0.010	0.000	-0.003	0.000	0.007	0.000
Flv(b)	$\pi \rightarrow \pi^*$	2.423	0.362	0.000	0.001	0.008	0.006	-0.006	0.018	-0.005	0.037
	CT	2.746	0.000	-0.006	0.000	0.000	0.000	-0.009	0.000	0.025	0.000
	$\pi \rightarrow \pi^*$	2.843	0.168	-0.003	0.000	0.005	0.001	-0.002	0.020	-0.004	-0.044
	$n, \sigma \rightarrow \pi^*$	3.126	0.000	-0.005	0.000	0.011	0.000	0.018	0.000	0.024	0.000
D21L6	CT	2.588	1.051	–	–	0.011	0.005	-0.005	0.012	0.011	0.022
	$\pi \rightarrow \pi^*$	3.353	0.107	–	–	0.015	0.002	-0.006	0.000	0.009	-0.010
	Rydberg	3.449	0.057	–	–	0.014	0.003	0.002	-0.004	0.002	0.042
	Rydberg	4.036	0.040	–	–	0.008	-0.022	0.010	-0.019	0.009	-0.024
Triad	$\pi \rightarrow \pi^*$	2.811	0.111	–	–	–	–	-0.002	-0.010	-0.029	0.040
	$n, \sigma \rightarrow \pi^*$	3.188	0.001	–	–	–	–	0.012	0.000	0.036	0.000
	$\pi \rightarrow \pi^*$	3.584	0.001	–	–	–	–	-0.001	0.000	0.002	0.000
	$\pi \rightarrow \pi^*$	3.758	0.354	–	–	–	–	-0.005	0.028	0.000	-0.027
MD				-0.004	0.000	0.005	0.000	0.002	0.003	0.009	0.004
MAD				0.005	0.000	0.012	0.002	0.010	0.005	0.015	0.010
MAX				0.014	0.002	0.040	0.022	0.036	0.028	0.042	0.044

[a] The reference value for the properties is taken from the best available calculation, which is the canonical ADC(2) value for the Flv(a), dyad, bisimide derivative, leupeptin, met-enkephalin, and Flv(b) molecules. For the D21L6 and the triad molecules, because of their size, only reduced-cost results are available as a reference.

energies, while the MAD is also zero for the oscillator strengths. Significant difference among the various types of excitations cannot be observed, which suggests that the approximations can be used in a black-box manner for arbitrary type of excited state. Concerning the reduced-scaling algorithm, the results are also very encouraging. With the default thresholds, the MD (MAD) of the excitation energies is 9 (15) meV, while the MAX is still 42 meV.

In other words, with the domain construction, the error measures increase by 2-4 meV with respect to the reduced-cost algorithm. However, if one takes a closer look at the errors, a small amount of error compensation occurs between the domain and the reduced-cost approximations. Despite the excellent results for the excitation energies, the errors are more notable for the oscillator strengths. Their MD (MAD) is 0.004 (0.010), while their MAX is 0.044, which are significantly higher compared to the reduced-cost algorithm. However, as it can be seen, the relative error in the most intense transitions is still acceptable being at most about 10%. Accordingly, these errors have no effect on the assignation of absorption spectra. Of course, in particular cases, it may be important to carry out more accurate calculations, and that is why we have introduced a tighter cutoff parameter for the domain construction. With the tight settings the error measures can be reduced by 5-7 meV for the excitation energies and can be halved for the oscillator strengths. In this case, all the MD, MAD, and MAX errors of the excitation energies are slightly lower compared to the default reduced-cost algorithm, which can be explained by error compensation. The errors in the oscillator strengths are somewhat still higher, however, they are firmly more moderate.

The above results show the accuracy of the reduced-cost and reduced-scaling algorithms, but the computational resources required by them are also important. To characterize this, the sizes of the bases in which the time-consuming operations were performed, the total wall-clock times, and the overall speedups with respect to the canonical calculations are collected. The results for the reduced-cost algorithm are presented in Table 4. Inspecting the results we can observe that 59.7% of the VNOs and 83.3% of the NAFs can be dropped on the average using the default thresholds. These averages are fairly representative, as the differences between the maximum and minimum values are around 5% and 2% for the VNOs and NAFs, respectively. Accordingly, both the operation counts in the rate-determining steps and the memory requirement can be reduced by about a factor of 35. As the calculation of the one-particle densities and integral transformations requires extra operations, it is important to determine the overall speedups. The total wall-clock time in the canonical

Table 4: The percentage of VNOs and NAFs dropped, total wall-clock times (in hours), and overall speedups with the reduced-cost algorithm using various thresholds.

	Tight				Default			
Molecule	Dropped VNOs	Dropped NAFs	Total wall time	Speedup	Dropped VNOs	Dropped NAFs	Total wall time	Speedup
Flv(a)	35.8	71.8	23.7	5.2	58.1	82.2	10.8	11.5
	35.7	71.8			58.1	82.1		
	35.9	71.9			58.4	82.2		
	36.0	71.9			58.4	82.2		
Dyad	36.2	73.1	32.8	5.0	58.3	82.8	14.5	11.3
	35.6	72.9			58.0	82.7		
	36.4	73.1			58.6	82.9		
	35.7	72.9			58.0	82.7		
Bisimide der.	37.6	72.7	41.7	5.7	59.7	82.7	23.2	10.3
	37.5	72.7			59.8	82.8		
	35.2	72.4			57.4	82.5		
	37.6	72.7			59.8	82.8		
Leupeptin	39.5	73.5	27.1	5.0	61.6	84.2	13.4	10.1
	38.2	73.3			60.3	83.9		
	37.7	73.2			59.8	83.9		
	38.4	73.3			60.5	83.9		
Met-enkephalin	36.2	72.8	90.7	4.7	58.6	83.3	45.0	9.5
	38.0	73.1			60.5	83.5		
	36.9	72.9			59.2	83.3		
	34.0	72.6			56.3	82.9		
Flv(b)	38.7	73.0	114.6	5.9	60.7	83.3	54.5	12.3
	39.1	73.0			61.3	83.4		
	39.3	73.1			61.5	83.5		
	39.4	73.1			61.6	83.5		
D21L6	39.8	74.6	288.8	–	61.9	84.4	115.8	–
	38.5	74.4			60.6	84.2		
	39.3	74.5			61.5	84.3		
	39.3	74.5			61.5	84.3		
Triad	–	–	–	–	61.4	83.7	549.8	–
	–	–			59.8	83.5		
	–	–			58.9	83.4		
	–	–			61.7	83.8		
Average	37.4	73.0		5.3	59.7	83.3		10.8
Maximum	39.8	74.6		5.9	61.9	84.4		12.3
Minimum	34.0	71.8		4.7	56.3	82.1		9.5

case contains the time required for the integral-direct CIS, MP2, and ADC(2) calculations, which algorithms are also well-optimized. For the reduced-cost algorithm, it includes the integral transformation from the AO to the MO basis, the computation of the complete MO space NAFs, the time required for the CIS solution, the one-particle density calculations, the integral transformation to the VNO basis, the final NAF construction and transformation, as well as the time spent in the MP2 and ADC(2) calculations.[50] Taking into account all of

these steps, the overall speedup gained with the default thresholds is 10.8 in average, while the minimum (maximum) speedup is 9.5 (12.3). The balanced speedups can be attributed to the systematic behavior of the VNO and NAF space truncations: the percentage of both the VNOs and NAFs retained fluctuates within a narrow range. Similar conclusions can be drawn for the tight parameters. Then 37.4% and 73.0% of the VNOs and NAFs can be neglected on the average, respectively, which means a reduction of about 10-times in the size of the integral list and in the operation count of the most expensive steps. The average speedup is 5.3, and the minimum and maximum values of the space reductions and the speedups are also well-balanced.

The above values were also collected for the reduced-scaling algorithm together with the corresponding parameters which are only relevant in the state-dependent local domain. The percentage of various orbitals and functions dropped as well as the speedups with respect to the reduced-cost algorithm are collected in Table 5.

Table 5: Percentage of AOs, auxiliary functions, LMOs, and PAOs dropped at the domain construction; percentage of VNOs and NAFs dropped inside the local domain with respect to the canonical bases of the domain; as well as the speedups with respect to the reduced-cost algorithm using various thresholds.

	Tight								Default							
	Dropped							Overall	Dropped							Overall
Molecule	AOs	Aux.	LMOs	PAOs	VNOs	NAFs	Speedup[a]	speedup[b]	AOs	Aux.	LMOs	PAOs	VNOs	NAFs	Speedup[a]	speedup[b]
Flv(a)	0.0	10.5	8.1	12.8	61.4	83.5	0.8	0.9	0.0	20.6	20.7	24.6	66.7	85.9	1.0	1.2
	0.0	11.9	9.2	12.3	61.2	83.4	0.8		0.0	11.9	16.1	25.3	65.0	85.2	1.0	
	0.0	11.9	9.2	17.1	61.7	83.5	0.9		0.0	11.9	16.1	26.5	65.4	85.3	1.1	
	0.0	16.6	25.3	28.4	68.7	86.7	1.3		9.2	40.1	40.2	44.4	74.9	89.2	1.7	
Dyad	0.0	1.0	0.0	5.4	58.3	82.5	0.7	0.8	0.0	3.0	4.7	11.8	60.5	83.4	0.7	0.8
	0.0	1.0	0.0	0.7	57.8	82.4	0.7		0.0	1.0	0.0	0.7	57.8	82.4	0.7	
	10.1	31.7	32.6	33.9	72.1	88.2	1.8		23.7	41.8	39.5	45.8	75.5	89.6	2.1	
	0.0	1.0	0.0	0.0	57.8	82.4	0.7		0.0	1.0	0.0	2.3	58.0	82.4	0.7	
Bisimide der.	0.0	3.5	0.0	6.7	59.7	82.4	0.8	1.0	0.0	3.5	0.0	8.8	59.7	82.4	0.8	1.1
	0.0	28.8	28.6	32.2	71.4	87.4	1.7		0.0	28.8	28.6	32.2	71.4	87.4	1.7	
	0.0	5.2	2.0	12.7	58.8	82.7	0.8		0.0	5.2	2.0	12.7	58.8	82.7	0.8	
	0.0	9.2	12.2	20.5	62.2	84.5	1.1		0.0	28.8	28.6	32.2	68.9	87.1	1.6	
Leupeptin	3.1	50.1	52.3	63.0	82.5	92.7	2.4	1.7	14.3	64.6	66.3	73.6	87.6	94.8	2.8	2.6
	0.0	18.6	15.1	29.2	67.5	86.7	1.2		0.0	35.0	39.5	51.4	77.1	90.7	2.1	
	0.0	25.4	22.1	41.9	70.7	88.0	2.0		16.3	64.3	65.1	72.5	86.7	94.6	4.2	
	0.0	25.7	23.3	32.4	70.5	88.1	1.5		3.1	36.8	38.4	48.1	76.4	90.4	2.1	
Met-enkephalin	53.1	63.6	64.8	67.0	85.1	94.0	5.0	5.1	54.9	67.0	69.4	71.4	87.0	94.8	5.1	5.3
	53.1	63.6	64.8	64.5	85.5	94.1	5.9		54.9	67.0	69.4	68.7	87.2	94.9	6.1	
	33.9	43.9	49.1	48.2	78.5	91.4	4.3		37.3	51.1	56.5	54.9	81.1	92.5	4.6	
	37.3	47.5	51.9	50.9	78.1	91.5	5.2		38.2	51.8	58.3	56.8	80.6	92.6	5.5	
Flv(b)	0.0	7.8	8.8	12.6	64.0	84.6	1.1	1.3	1.6	14.3	18.4	20.9	67.7	86.3	1.4	1.9
	0.0	2.9	0.0	2.4	61.2	83.2	0.9		0.0	7.8	9.7	11.7	64.6	84.8	1.1	
	0.0	8.5	9.7	14.6	64.9	84.9	1.5		0.0	26.7	29.0	39.8	72.7	88.1	2.6	
	0.0	21.4	22.8	33.0	70.3	87.0	2.1		26.0	52.0	50.9	57.2	81.0	91.7	4.0	
D21L6	12.1	24.5	22.9	27.9	70.6	87.7	2.0	2.2	14.8	30.3	28.2	40.3	73.1	88.6	2.6	2.7
	12.1	24.5	22.9	27.9	69.6	87.5	2.3		14.8	29.7	27.5	35.6	71.5	88.2	2.7	
	6.7	18.7	17.6	22.7	68.2	86.8	1.9		6.7	24.1	25.2	29.9	70.8	88.1	2.6	
	12.1	24.5	22.9	27.9	70.3	87.6	2.5		13.5	27.1	26.7	31.6	71.5	88.2	2.8	
Triad	4.0	17.6	15.7	23.5	66.5	85.9	1.2	2.5	9.4	24.4	23.8	29.9	69.4	87.2	1.8	3.9
	7.4	14.5	11.9	21.5	64.9	85.2	1.3		7.9	25.7	29.2	36.9	72.0	88.3	2.7	
	43.6	68.1	68.1	69.7	86.8	94.6	8.1		44.1	69.2	71.4	71.8	87.9	95.1	8.1	
	25.7	58.6	59.5	64.5	84.8	93.4	7.9		42.5	69.5	68.7	74.0	88.3	94.9	8.7	
Average	9.8	23.8	23.5	29.0	69.1	87.0	2.3	1.9	13.5	32.4	33.4	38.9	73.0	88.7	2.7	2.4
Maximum	53.1	68.1	68.1	69.7	86.8	94.6	8.1	5.1	54.9	69.5	71.4	74.0	88.3	95.1	8.7	5.3
Minimum	0.0	1.0	0.0	0.0	57.8	82.4	0.7	0.8	0.0	1.0	0.0	0.7	57.8	82.4	0.7	0.8

[a] The time required for the reduced-cost CIS calculations is distributed equally among the excited states. [b] Overall speedup for the entire calculation for all the investigated states.

The results are in line with the expectations. The sizes of the bases are heavily influenced by several factors, thus, their comparison is rather difficult. Perhaps one of the most relevant factors is the size of the molecule, but, in addition, the properties of the electronic structure, the character of the excited state, and the shape of the molecular orbitals involved in the excitation are also important. For the first three molecules no significant improvement can be achieved. The Flv(a) and dyad molecules are relatively small (51 and 53 atoms, respectively), while the bisimide derivative has an extended delocalized electronic structure. However, the effective size of the system was successfully reduced for some excitations via the domain approximation [the S_4 state of Flv(a), the S_3 state of the dyad, as well as the S_2 and S_4 transitions of the bisimide derivative]. In these cases the time required for the rate-determining steps are decreased but the overall speedups are moderate. This can be explained by the fact that the local-fitting CIS algorithm is more expensive than the in-core reduced-cost CIS algorithm for these systems, and the costly integral transformation to the NO basis must be performed for each state separately. For larger systems, significant speedups can be gained: for the molecules above 78 atoms we observe up to 3–9-fold improvement with respect to the reduced-cost algorithm. Comparing the default and tight thresholds, we can observe that about 10% more LMOs are retained with the latter, however, the number of the VNOs and NAFs does not differ significantly. That is, the accuracy of the calculation is primarily influenced by the number occupied orbitals.

3.4 Extended systems

To demonstrate the efficiency of the algorithm presented, further calculations were performed for more extended molecular systems. Such extensive calculations with more than 200 atoms using the aug-cc-pVTZ basis set are not feasible on our hardware even with our reduced-cost algorithm. The calculated excitation energies and oscillator strengths, the sizes of the various bases, and the wall-clock times measured are collected in Table 6.

Table 6: ADC(2) excitation energies (ω, in eV) and oscillator strengths (f) computed with the present approach, the percentage of basis set reduction (see the caption of Table 5), as well as the wall-clock times (in hours) required for the various rate-determining steps of the calculations using the aug-cc-pVTZ basis set.

				Dropped						Wall-clock time				
Molecule	Character	ω	f	AOs	Aux.	LMOs	PAOs	VNOs	NAFs	CIS	Domain[a]	Density[b]	ADC(2)	Total
C60Im–ZnP–BDP	$\pi \rightarrow \pi^*$	1.826	0.000	48.0	59.3	59.5	62.5	83.6	92.9	71.1	1.9	14.7	22.9	110.6
	$\pi \rightarrow \pi^*$	1.895	0.000	48.0	59.9	60.1	63.3	83.8	93.0	71.3	1.9	13.8	21.8	108.8
	$\pi \rightarrow \pi^*$	2.235	0.048	46.0	61.3	62.8	64.8	84.8	93.7	58.8	1.9	11.8	5.6	78.1
	$\pi \rightarrow \pi^*$	2.250	0.046	46.0	61.3	62.8	64.8	84.8	93.7	56.2	1.9	11.8	5.8	75.7
Bivalirudin	$\pi \rightarrow \pi^*$	4.808	0.022	80.4	90.9	91.8	91.7	96.4	98.6	19.6	0.7	0.2	0.0	20.5
	$\pi \rightarrow \pi^*$	5.164	0.001	70.3	85.1	88.4	87.8	94.8	98.0	25.3	0.7	0.6	0.1	26.6
	$\pi \rightarrow \pi^*$	5.371	0.001	76.8	84.7	87.0	86.6	94.7	97.9	22.0	0.7	0.6	0.2	23.5
	$\pi \rightarrow \pi^*$	5.769	0.103	77.2	86.1	87.0	87.4	94.6	97.8	31.8	0.7	0.6	0.3	33.3
WW-6 dye	$\pi \rightarrow \pi^*$	1.990	0.693	33.2	56.6	53.9	65.0	84.0	92.9	114.9	2.4	53.5	38.9	209.7
FA@144 H_2O	$n \rightarrow \pi^*$	5.790	0.000	46.2	93.9	90.8	93.8	97.0	98.6	90.6	2.2	1.4	0.1	94.3

[a] Time required for the domain construction including the transformation of CIS coefficients, calculation of PAOs, construction of BP domains, compilation of atom and MO lists. [b] Time required for the calculation of NOs and NAFs including the integral transformations.

Inspecting the results similar statements can be made as for the smaller examples. The sizes of the bases and the wall-clock times highly depend on the factors mentioned in the previous subsection. However, for these large molecules, all the basis set sizes can be almost halved for any type of excitation by employing the domain approximation. Consequently, the calculation of the VNOs and NAFs, as well as the demanding excited-state correlation calculation can be performed with significant savings. As it can be seen, the cubic-scaling local density fitting CIS calculation is the rate-determining step of the whole procedure in all the cases. For the lowest excited state of the C60Im–ZnP–BDP system, which is the least suitable for sizeable cost-reduction, the occupied, virtual, and auxiliary indices can be cut by about 60% with the domain construction. On the basis of operation count estimates the corresponding speedup factor for the VNO and NAF construction is about 100. The final LMO, VNO, and NAF bases are compressed by about 63, 84, and 93% compared to the corresponding canonical bases, thus, the speedup in the ADC(2) part is about 2000-times. For this system the reduced-scaling CIS calculation takes 70 hours, which is approximately 65% of the total wall-clock time. This ratio and the basis set reduction are even better with increasing system sizes.

The effects of the explicit solvation was tested on the lowest excited state of a solvated formamide molecule with an increasing number of water molecules as introduced by Baudin and Kristensen.[71] The change of the excitation energy as a function of the size of the solvation shell is shown in Fig. 5. The tendencies are greatly in line with those of Ref. 71 in spite of the different basis sets (aug-cc-pVTZ here, aug'-cc-pVDZ in Fig. 9 of Ref. 71). As we can see, the canonical calculations were performed for up to 30 water molecules, whereas the reduced-cost approximation allowed us to carry out calculations with up to 44 solvent molecules. Since the errors of the reduced-cost and reduced-scaling CIS results are negligible (below 5 meV), they are not presented. The maximum error of the ADC(2) excitation energy in the considered range using the reduced-scaling algorithm is 30 meV with respect to either the canonical or the reduced-scaling calculation. The times required for the canonical ADC(2),

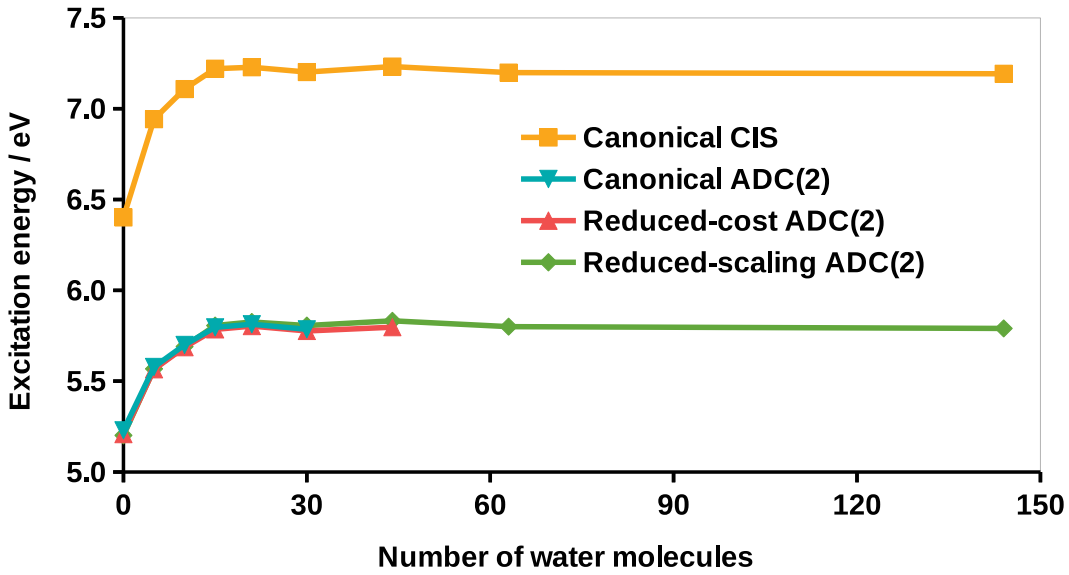

Figure 5: CIS and ADC(2) excitation energies (in eV) for solvated formamide with various approximations using the aug-cc-pVTZ basis set.

the reduced-scaling CIS, and the reduced-scaling ADC(2) calculations including the domain construction as well as the evaluation of VNOs and NAFs are presented in Fig. 6. The

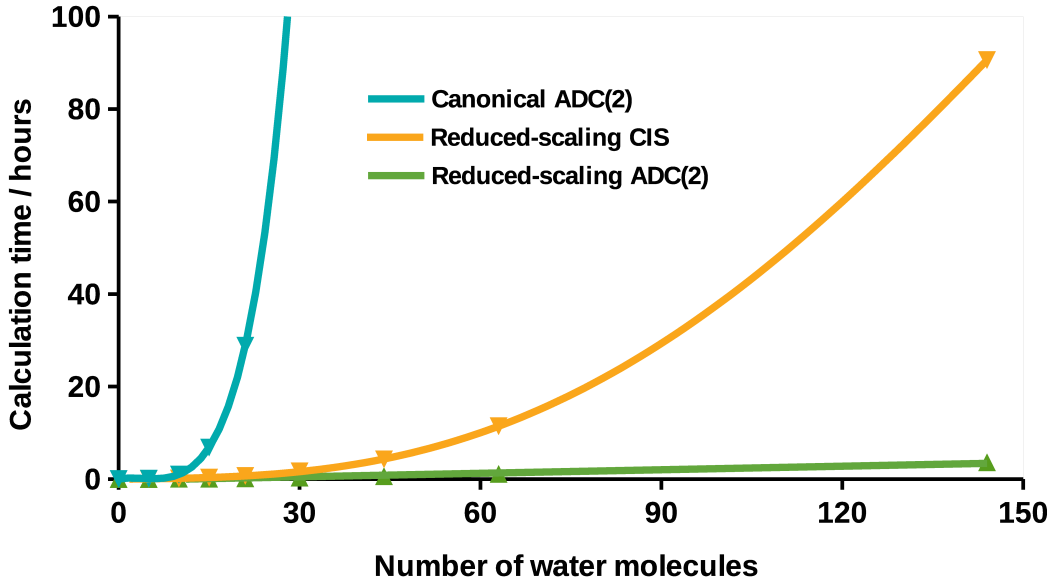

Figure 6: Wall-clock times (in hours) required for the canonical ADC(2), reduced-scaling CIS, and reduced-scaling ADC(2) calculations including the domain construction as well as the VNO and NAF evaluation for the solvation of formamide.

plots verify the effectiveness of the reduced-scaling method. As it can be seen, the ADC(2) calculation within the domain takes the least time for all solvent region sizes using the present reduced-scaling algorithm. The improvement in efficiency is significant even for 15 water molecules, while the time required for the ADC(2) step of the domain calculation hardly changes with increasing the number of water molecules. Here we can effectively exploit that the investigated excitation is localized on the formamide molecule and its first few solvent shells. The rate-determining step of the procedure is the cubic-scaling CIS calculation in all the cases. This highlights the importance of employing the local fitting approximation in our reduced-scaling CIS implementation. For instance, for the largest cluster with 144 water molecules the reduced-scaling CIS algorithm is almost two orders of magnitude faster than the conventional one. Hence the use of quartic-scaling CIS approaches, even with significantly smaller basis sets can be problematic for large molecules.[71] Accordingly, the local fitting approach is an effective tool to speed up the CIS, the rate-determining step of these calculations.

4 Conclusions

An algorithm was presented for the construction of excitation-dependent local orbital domains for the reduced-scaling calculation of excitation energies and transition properties. The various aspects of the selection of the basis functions, auxiliary functions, and molecular orbitals were discussed in detail. The approximation was successfully combined with our previously developed reduced-cost techniques,[49,50] further compressing the size of the orbital domains using virtual natural orbitals and natural auxiliary functions. Additional methodological improvements were also discussed for the calculation of MP2 and CIS(D) density matrices, such as the utilization of the NAF approach and the sparsity of the CIS coefficient vector.

To determine the cutoff parameters a suitable model system was constructed. Inspecting

a charge transfer excitation between two distant systems allowed us to separately study the effects of the thresholds and to define default and tight values for them. With the selected thresholds, additional benchmark calculations were performed for realistic molecular systems at ADC(2) level. We compared the errors of the reduced-cost and the reduced-scaling algorithms in detail, furthermore the performance of the approximations was also assessed. Our results show that the errors introduced by the domain construction were highly acceptable, being about 2-4 meV. In turn, on top of the about an order of magnitude performance improvement of the reduced-cost approach, an additional, up to 3–9-fold speedup can be achieved even for systems of smaller than 100 atoms. The conservative cutoff parameters and the reduced-scaling algorithm enable us to perform accurate, reduced-scaling ADC(2) calculations for excitation energies and transition properties using triple-ζ basis sets with diffuse functions for systems of up to 400 atoms in a matter of days using a moderate, 8-core processor.

Supporting Information

The structure of the molecules studied and the Cartesian coordinates for the CT system. This information is available free of charge via the Internet at http://pubs.acs.org

Acknowledgments

The authors are grateful for the financial support from the National Research, Development, and Innovation Office (NKFIH, Grant No. KKP126451). This work was also supported by the BME-Nanotechnology FIKP grant of EMMI (BME FIKP-NANO). The computing time granted on the Hungarian HPC Infrastructure at NIIF Institute, Hungary, is gratefully acknowledged.

References

(1) Casida, M. E. In *Computational Chemistry: Reviews of Current Trends*; Chong, D. P., Ed.; World Scientific: Singapore, 1999; Vol. 1.

(2) Casida, M. E.; Huix-Rotllant, M. Progress in Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.* **2012**, *63*, 287.

(3) Ridley, J.; Zerner, M. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theor. Chim. Acta* **1973**, *32*, 111.

(4) Zerner, M. C.; Loew, G. H.; Kirchner, R. F.; Mueller-Westerhof, U. T. An Intermediate Neglect of Differential Overlap Technique for Spectroscopy of Transition-Metal Complexes. Ferrocene. *J. Am. Chem. Soc.* **1980**, *102*, 589.

(5) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495.

(6) Koslowski, A.; Beck, M. E.; Thiel, W. Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach. *J. Comput. Chem.* **2003**, *24*, 714.

(7) Stanton, J. F.; Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **1993**, *98*, 7029.

(8) Watts, J. D.; Bartlett, R. J. The inclusion of connected triple excitations in the equation-of-motion coupled-cluster method. *J. Chem. Phys.* **1994**, *101*, 3073.

(9) Koch, H.; Jørgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, *93*, 3333.

(10) Koch, H.; Jensen, H. J. A.; Jørgensen, P.; Helgaker, T. Excitation energies from the coupled cluster singles and doubles linear response function (CCSDLR). Applications to Be, CH^+, CO, and H_2O. *J. Chem. Phys.* **1990**, *93*, 3345.

(11) Rico, R. J.; Head-Gordon, M. Single-reference theories of molecular excited states with single and double substitutions. *Chem. Phys. Lett.* **1993**, *213*, 224.

(12) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256.

(13) Oddershede, J. Polarization Propagator Calculations. *Adv. Quantum Chem.* **1978**, *11*, 275.

(14) Oddershede, J. Propagator methods. *Adv. Chem. Phys.* **1987**, *69*, 201.

(15) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **1980**, *48*, 157.

(16) Christiansen, O.; Koch, H.; Jørgensen, P. The second-order approximate coupled cluster singles and doubles model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409.

(17) Christiansen, O.; Koch, H.; Jørgensen, P.; Helgaker, T. Integral direct calculation of CC2 excitation energies: singlet excited states of benzene. *Chem. Phys. Lett.* **1996**, *263*, 530.

(18) Hald, K.; Hättig, C.; Yeager, D. L.; Jørgensen, P. Linear response CC2 triplet excitation energies. *Chem. Phys. Lett.* **2000**, *328*, 291.

(19) Hättig, C.; Weigend, F. CC2 excitation energy calculations on large molecules using the resolution of the identity approximation. *J. Chem. Phys.* **2000**, *113*, 5154.

(20) Hättig, C. Geometry optimizations with the coupled-cluster model CC2 using the resolution-of-the-identity approximation. *J. Chem. Phys.* **2003**, *118*, 7751.

(21) Köhn, A.; Hättig, C. Analytic gradients for excited states in the coupled-cluster model CC2 employing the resolution-of-the-identity approximation. *J. Chem. Phys.* **2003**, *119*, 5021.

(22) Oddershede, J.; Jørgensen, P. An order analysis of the particle-hole propagator. *J. Chem. Phys.* **1977**, *66*, 1541.

(23) Nielsen, E. S.; Jørgensen, P. Transition moments and dynamic polarizabilities in a second order polarization propagator approach. *J. Chem. Phys.* **1980**, *73*, 6238.

(24) Oddershede, J.; Jørgensen, P.; Yeager, D. L. Polarization propagator methods in atomic and molecular calculations. *Comp. Phys. Rep.* **1984**, *2*, 33.

(25) Schirmer, J. Beyond the random-phase approximation: A new approximation scheme for the polarization propagator. *Phys. Rev. A* **1982**, *26*, 2395.

(26) Trofimov, A. B.; Stelter, G.; Schirmer, J. A consistent third-order propagator method for electronic excitation. *J. Chem. Phys.* **1999**, *111*, 9982.

(27) Trofimov, A. B.; Stelter, G.; Schirmer, J. Electron excitation energies using a consistent third-order propagator approach: Comparison with full configuration interaction and coupled cluster results. *J. Chem. Phys.* **2002**, *117*, 6402.

(28) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; Wiley: Chichester, 2000.

(29) Buenker, R. J.; Peyerimhoff, S. D. Individualized configuration selection in CI calculations with subsequent energy extrapolation. *Theor. Chim. Acta* **1974**, *35*, 33.

(30) Grimme, S.; Parac, M. Substantial Errors from Time-Dependent Density Functional Theory for the Calculation of Excited States of Large Π Systems. *ChemPhysChem* **2003**, *4*, 292.

(31) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange. *J. Chem. Phys.* **2003**, *119*, 2943.

(32) Mertins, F.; Schirmer, J. Algebraic propagator approaches and intermediate-state representations. I. The biorthogonal and unitary coupled-cluster methods. *Phys. Rev. A* **1996**, *53*, 2140.

(33) Starcke, J. H.; Wormit, M.; Dreuw, A. Unrestricted algebraic diagrammatic construction scheme of second order for the calculation of excited states of medium-sized and large molecules. *J. Chem. Phys.* **2009**, *130*, 024104.

(34) Harbach, P. H. P.; Wormit, M.; Dreuw, A. The third-order algebraic diagrammatic construction method (ADC(3)) for the polarization propagator for closed-shell molecules: Efficient implementation and benchmarking. *J. Chem. Phys.* **2014**, *141*, 064113.

(35) Wormit, M.; Rehn, D. R.; Harbach, P. H. P.; Wenzel, J.; Krauter, C. M.; Epifanovsky, E.; Dreuw, A. Investigating Excited Electronic States using the Algebraic Diagrammatic Construction (ADC) Approach of the Polarisation Propagator. *Mol. Phys.* **2014**, *112*, 774.

(36) Wormit, M.; Dreuw, A. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *WIREs Comput. Mol. Sci.* **2015**, *5*, 82.

(37) Krauter, C. M.; Pernpointner, M.; Dreuw, A. Application of the scaled-opposite-spin approximation to algebraic diagrammatic construction schemes of second order. *J. Chem. Phys.* **2013**, *138*, 044107.

(38) Goerigk, L.; Grimme, S. Assessment of TD-DFT methods and of various spin scaled CIS(D) and CC2 versions for the treatment of low-lying valence excitations of large organic dyes. *J. Chem. Phys.* **2010**, *132*, 184103.

(39) Winter, N. O. C.; Graf, N. K.; Leutwyler, S.; Hättig, C. Benchmarks for 0-0 transitions of aromatic organic molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 compared to high-resolution gas-phase data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6623.

(40) Jacquemin, D.; Duchemin, I.; Blase, X. 0-0 Energies Using Hybrid Schemes: Benchmarks of TD-DFT, CIS(D), ADC(2), CC2, and BSE/GW formalisms for 80 Real-Life Compounds. *J. Chem. Theory Comput.* **2015**, *11*, 5340.

(41) Boys, S. F.; Cook, G. B.; Reeves, C. M.; Shavitt, I. Automatic Fundamental Calculations of Molecular Structure. *Nature* **1956**, *178*, 1207.

(42) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496.

(43) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On some approximations in applications of $X\alpha$ theory. *J. Chem. Phys.* **1979**, *71*, 3396.

(44) Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller–Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, 044103.

(45) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.

(46) Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Martínez, T. J. Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions. *J. Chem. Phys.* **2012**, *137*, 221101.

(47) Kállay, M. A systematic way for the cost reduction of density fitting methods. *J. Chem. Phys.* **2014**, *141*, 244113.

(48) Hohenstein, E. G.; Kokkila, S. I. L.; Parrish, R. M.; Martínez, T. J. Tensor Hypercontraction Equation-of-Motion Second-Order Approximate Coupled Cluster: Electronic Excitation Energies in $O(N^4)$ Time. *J. Phys. Chem. B* **2013**, *117*, 12972.

(49) Mester, D.; Nagy, P. R.; Kállay, M. Reduced-cost linear-response CC2 method based on natural orbitals and natural auxiliary functions. *J. Chem. Phys.* **2017**, *146*, 194102.

(50) Mester, D.; Nagy, P. R.; Kállay, M. Reduced-cost second-order algebraic-diagrammatic construction method for excitation energies and transition moments. *J. Chem. Phys.* **2018**, *148*, 094111.

(51) Koch, H.; Sánchez de Merás, A. M. Size-intensive decomposition of orbital energy denominators. *J. Chem. Phys.* **2000**, *113*, 508.

(52) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481.

(53) Pedersen, T. B.; Sánchez de Merás, A. M.; Koch, H. Polarizability and optical rotation calculated from the approximate coupled cluster singles and doubles CC2 linear response theory using Cholesky decompositions. *J. Chem. Phys.* **2004**, *120*, 8887.

(54) Baudin, P.; Sánchez de Marín, J.; García Cuesta, I.; Sánchez de Merás, A. M. Calculation of excitation energies from the CC2 linear response theory using Cholesky decomposition. *J. Chem. Phys.* **2014**, *140*, 104111.

(55) Nascimento, D. R.; DePrince III, A. E. Linear Absorption Spectra from Explicitly Time-Dependent Equation-of-Motion Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2016**, *12*, 5834.

(56) Send, R.; Kaila, V. R. I.; Sundholm, D. Reduction of the virtual space for coupled-cluster excitation energies of large molecules and embedded systems. *J. Chem. Phys.* **2011**, *134*, 214114.

(57) Send, R.; Suomivuori, C.-M.; Kaila, V. R. I.; Sundholm, D. Coupled-Cluster Studies of Extensive Green Fluorescent Protein Models Using the Reduced Virtual Space Approach. *J. Phys. Chem. B* **2015**, *119*, 2933.

(58) Gamiz-Hernandez, A. P.; Angelova, I. N.; Send, R.; Sundholm, D.; Kaila, V. R. I. Protein-Induced Color Shift of Carotenoids in β-Crustacyanin. *Angew. Chem. Int. Ed.* **2015**, *54*, 11564.

(59) Suomivuori, C.-M.; Winter, N. O. C.; Hättig, C.; Sundholm, D.; Kaila, V. R. I. Exploring the Light-Capturing Properties of Photosynthetic Chlorophyll Clusters Using Large-Scale Correlated Calculations. *J. Chem. Theory Comput.* **2016**, *12*, 2644.

(60) Yang, C.; Dreuw, A. Evaluation of the Restricted Virtual Space Approximation in the Algebraic-Diagrammatic Construction Scheme for the Polarization Propagator to Speed-up Excited-State Calculations. *J. Comput. Chem.* **2017**, *38*, 1528.

(61) Meyer, W. PNO–CI Studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane. *J. Chem. Phys.* **1973**, *58*, 1017.

(62) Hay, P. J. On the calculation of natural orbitals by perturbation theory. *J. Chem. Phys.* **1973**, *59*, 2468.

(63) Ahlrichs, R.; Lischka, H.; Staemmler, V.; Kutzelnigg, W. PNO–CI (pair natural orbital configuration interaction) and CEPA–PNO (coupled electron pair approximation with pair natural orbitals) calculations of molecular systems. I. Outline of the method for closed-shell states. *J. Chem. Phys.* **1975**, *62*, 1225.

(64) Taube, A. G.; Bartlett, R. J. Fozen Natural Orbital Coupled-Cluster Theory: Forces and Application to Decomposition of Nitroethane. *J. Chem. Phys.* **2008**, *128*, 164101.

(65) DePrince, A. E.; Sherrill, C. D. Accurate Noncovalent Interaction Energies Using Truncated Basis Sets Based on Frozen Natural Orbitals. *J. Chem. Theory Comput.* **2013**, *9*, 293.

(66) Rolik, Z.; Kállay, M. Cost-reduction of high-order coupled-cluster methods via active-space and orbital transformation techniques. *J. Chem. Phys.* **2011**, *134*, 124111.

(67) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.

(68) Landau, A.; Khistyaev, K.; Dolgikh, S.; Krylov, A. I. Frozen Natural Orbitals for Ionized States Whithin Equation-of-Motion Coupled-Cluster Formalism. *J. Chem. Phys.* **2010**, *132*, 014109.

(69) Kumar, A.; Crawford, T. D. Frozen Virtual Natural Orbitals for Coupled Cluster Linear-Response Theory. *J. Phys. Chem. A* **2017**, *121*, 708.

(70) Höfener, S.; Klopper, W. Natural transition orbitals for the calculation of correlation and excitation energies. *Chem. Phys. Lett.* **2017**, *679*, 52.

(71) Baudin, P.; Kristensen, K. Correlated natural transition orbital framework for low-scaling excitation energy calculations (CorNFLEx). *J. Chem. Phys.* **2017**, *146*, 214114.

(72) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151.

(73) Pulay, P.; Saebø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller–Plesset perturbation theory. *Theor. Chim. Acta* **1986**, *69*, 357.

(74) Korona, T.; Werner, H.-J. Local treatment of electron excitations in the EOM-CCSD method. *J. Chem. Phys.* **2003**, *118*, 3006.

(75) Crawford, T. D.; King, R. A. Locally correlated equation-of-motion coupled cluster theory for the excited states of large molecules. *Chem. Phys. Lett.* **2002**, *366*, 611.

(76) Hampel, C.; Werner, H.-J. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.* **1996**, *104*, 6286.

(77) Kats, D.; Korona, T.; Schütz, M. Local CC2 electronic excitation energies for large molecules with density fitting. *J. Chem. Phys.* **2006**, *125*, 104106.

(78) Kats, D.; Korona, T.; Schütz, M. Transition strengths and first-order properties of excited states from local coupled cluster CC2 response theory with density fitting. *J. Chem. Phys.* **2007**, *127*, 064107.

(79) Kats, D.; Schütz, M. A multistate local coupled cluster CC2 response method based on the Laplace transform. *J. Chem. Phys.* **2009**, *131*, 124117.

(80) Freundorfer, K.; Kats, D.; Korona, T.; Schütz, M. Local CC2 response method for triplet states based on Laplace transform: Excitation energies and first-order properties. *J. Chem. Phys.* **2010**, *133*, 244110.

(81) Ledermüller, K.; Kats, D.; Schütz, M. Local CC2 response method based on the Laplace transform: Orbital-relaxed first-order properties for excited states. *J. Chem. Phys.* **2013**, *139*, 084111.

(82) Ledermüller, K.; Schütz, M. Local CC2 response method based on the Laplace transform: Analytic energy gradients for ground and excited states. *J. Chem. Phys.* **2014**, *140*, 164113.

(83) Schütz, M. Oscillator strengths, first-order properties, and nuclear gradients for local ADC(2). *J. Chem. Phys.* **2015**, *142*, 214103.

(84) Russ, N. J.; Crawford, T. D. Local correlation in coupled cluster calculations of molecular response properties. *Chem. Phys. Lett.* **2004**, *400*, 104.

(85) Russ, N. J.; Crawford, T. D. Local correlation domains for coupled cluster theory: Optical rotation and magnetic field perturbations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3345.

(86) Helmich, B.; Hättig, C. Local pair natural orbitals for excited states. *J. Chem. Phys.* **2011**, *135*, 214106.

(87) Helmich, B.; Hättig, C. A pair natural orbital implementation of the coupled cluster model CC2 for excitation energies. *J. Chem. Phys.* **2012**, *139*, 084114.

(88) Helmich, B.; Hättig, C. A pair natural orbital based implementation of ADC(2)-x: Perspectives and challenges for response methods for singly and doubly excited states in large molecules. *Comput. Theoret. Chem.* **2014**, *1040-1041*, 35.

(89) Frank, M. S.; Hättig, C. A pair natural orbital based implementation of CCSD excitation energies within the framework of linear response theory. *J. Chem. Phys.* **2018**, *148*, 134102.

(90) Dutta, A. K.; Neese, F.; Izsák, R. Speeding up equation of motion coupled cluster theory with the chain of spheres approximation. *J. Chem. Phys.* **2016**, *144*, 034102.

(91) Dutta, A. K.; Neese, F.; Izsák, R. Towards a pair natural orbital coupled cluster method for excited states. *J. Chem. Phys.* **2016**, *145*, 034102.

(92) Dutta, A. K.; Nooijen, M.; Neese, F.; Izsák, R. Automatic active space selection for the similarity transformed equations of motion coupled cluster method. *J. Chem. Phys.* **2017**, *146*, 074103.

(93) Dutta, A. K.; Nooijen, M.; Neese, F.; Izsák, R. Exploring the accuracy of a low scaling

similarity transformed equation of motion method for vertical excitation energies. *J. Chem. Theory Comput.* **2018**, *14*, 72.

(94) Peng, C.; Clement, M. C.; Valeev, E. F. State-Averaged Pair Natural Orbitals for Excited States: A Route toward Efficient Equation of Motion Coupled-Cluster. *J. Chem. Theory Comput.* **2018**, *14*, 5597.

(95) Baudin, P.; Kristensen, K. LoFEx — A local framework for calculating excitation energies: Illustrations using RI-CC2 linear response theory. *J. Chem. Phys.* **2016**, *144*, 224106.

(96) Baudin, P.; Kjærgaard, T.; Kristensen, K. CC2 oscillator strengths within the local framework for calculating excitation energies (LoFEx). *J. Chem. Phys.* **2017**, *146*, 144107.

(97) Baudin, P.; Bykov, D.; Liakh, D.; Ettenhuber, P.; Kristensen, K. A local framework for calculating coupled cluster singles and doubles excitation energies (LoFEx-CCSD). *Mol. Phys.* **2017**, *115*, 2135.

(98) Crawford, T. D.; Kumar, A.; Bazanté, A. P.; Remigio, R. D. Reduced-scaling coupled cluster response theory: Challenges and opportunities. *WIREs Comput. Mol. Sci.* **2019**, e1406.

(99) Mester, D.; Kállay, M. Reduced-Scaling Approach for Configuration Interaction Singles and Time-Dependent Density Functional Theory Calculations Using Hybrid Functionals. *J. Chem. Theory Comput.* **2019**, *15*, 1690.

(100) Boughton, J. W.; Pulay, P. Comparison of the Boys and Pipek–Mezey Localizations in the Local Correlation Approach and Automatic Virtual Basis Selection. *J. Comput. Chem.* **1993**, *14*, 736.

(101) Nagy, P. R.; Samu, G.; Kállay, M. An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **2016**, *12*, 4897.

(102) Nagy, P. R.; Samu, G.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications. *J. Chem. Theory Comput.* **2018**, *14*, 4193.

(103) Nagy, P. R.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform. *J. Chem. Phys.* **2017**, *146*, 214106.

(104) Pipek, J.; Mezey, P. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916.

(105) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834.

(106) MRCC, a quantum chemical program suite written by M. Kállay, P. R. Nagy, Z. Rolik, D. Mester, G. Samu, J. Csontos, J. Csóka, B. P. Szabó, L. Gyevi-Nagy, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and B. Hégely. See http://www.mrcc.hu/ (accessed January 13, 2019).

(107) Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

(108) Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358.

(109) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175.

(110) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143.

(111) Weigend, F. Hartree–Fock Exchange Fitting Basis Sets for H to Rn. *J. Comput. Chem.* **2008**, *29*, 167.

(112) Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796.

(113) Goerigk, L.; Moellmann, J.; Grimme, S. Computation of accurate excitation energies for large organic molecules with double-hybrid density functionals. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4611.

(114) D'Souza, F.; Smith, P. M.; Zandler, M. E.; McCarty, A. L.; Itou, M.; Araki, Y.; Ito, O. Energy Transfer Followed by Electron Transfer in a Supramolecular Triad Composed of Boron Dipyrrin, Zinc Porphyrin, and Fullerene: A Model for the Photosynthetic Antenna-Reaction Center Complex. *J. Am. Chem. Soc.* **2004**, *126*, 7898.

(115) Furche, F.; Krull, B. T.; Nguyen, B. D.; Kwon, J. Accelerating molecular property calculations with nonorthonormal Krylov space methods. *J. Chem. Phys.* **2016**, *144*, 174105.

(116) Luo, J.; Xu, M.; Li, R.; Huang, K.-W.; Jiang, C.; Qi, Q.; Zeng, W.; Zhang, J.; Chi, C.; Wang, P.; Wu, J. *N*-Annulated Perylene as An Efficient Electron Donor for Porphyrin-Based Dyes: Enhanced Light-Harvesting Ability and High-Efficiency Co(II/III)-Based Dye-Sensitized Solar Cells. *J. Am. Chem. Soc.* **2014**, *136*, 265.

(117) Hellweg, A.; Rappoport, D. Development of new auxiliary basis functions of the Karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC calculations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 1010.

Graphical TOC Entry