Article

Equation-of-Motion MLCCSD and CCSD-in-HF Oscillator Strengths and Their Application to Core Excitations

Sarai Dery Folkestad and Henrik Koch*

Read Online

ACCESS | 📊 Metrics & More | 📄 Article Recommendations

ABSTRACT: We present an implementation of equation-of-motion oscillator strengths for the multilevel CCSD (MLCCSD) model where CCS is used as the lower level method (CCS/CCSD). In this model, the double excitations of the cluster operator are restricted to an active orbital space, whereas the single excitations are unrestricted. Calculated nitrogen K-edge spectra of adenosine, adenosine triphosphate (ATP), and an ATP-water system are used to demonstrate the performance of the model. Projected atomic orbitals (PAOs) are used to partition the virtual space into active and inactive orbital sets. Cholesky decomposition of the Hartree−Fock density is used to partition the occupied orbitals. This Cholesky-PAO partitioning is cheap, scaling as $O(N^3)$, and is suitable for the calculation of core excitations, which are localized in character. By restricting the single excitations of the cluster operator to the active space, as well as the double excitations, the CCSD-in-HF model is obtained. A comparison of the two models—MLCCSD and CCSD-in-HF—is presented for the core excitation spectra of the adenosine and ATP systems.
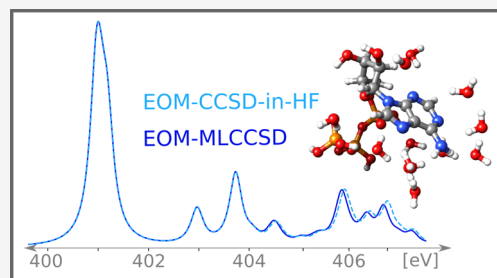
INTRODUCTION

The multilevel coupled cluster (MLCC) approach can be used to calculate excitation energies of molecular systems that are too large for the standard coupled cluster models. In MLCC, the higher order excitations that are included in the cluster operator are restricted to an active orbital space. One can view the approach as applying a higher level of coupled cluster theory to the active orbitals. The MLCC approach was introduced by Myhre et al.[1−3] but is similar to the active space approach, which has resulted from the multireference coupled cluster method of Oliphant and Adamowicz et al.[4−7]

In the multilevel coupled cluster singles and doubles (MLCCSD) model,[1,2] CCSD[8] is applied to the active orbital space. Coupled cluster singles (CCS) and/or singles and perturbative doubles[9] (CC2) are used for the inactive orbital space. With carefully selected active orbitals, excitation energies of CCSD quality are obtained. When the active orbital space is enlarged, the MLCCSD excitation energies converge smoothly toward the CCSD excitation energies. The MLCCSD oscillator strengths, within the coupled cluster response formalism,[10,11] were reported in ref 12. However, these proof-of-concept calculations did not exploit the computational reductions offered by the multilevel framework and were performed using a standard CCSD code.

Recently, we have reformulated and implemented the MLCCSD ground and excited state equations.[13] We have found it sufficient to use CCS as the lower level model to obtain accurate valence excitation energies. This CCS/CCSD model is cheaper and simpler than the CC2/CCSD and CCS/

CC2/CCSD models. Moreover, the CCS/CCSD model is compatible with properties derived within the equation-of-motion[14−17] (EOM) framework as well as with the coupled cluster response theory. This is because the only modification with respect to CCSD is a restriction of the double part of the cluster operator to the active orbital space. The CCS/CCSD model has been used to calculate valence excitation energies for a system with more than 50 second row atoms and the computational scaling approaches that of CCS for sufficiently large inactive spaces.[18] In this paper, we present an implementation of EOM oscillator strengths[19] for the CCS/CCSD model.

The success of an MLCC calculation relies heavily on the choice of the active orbital space. Two strategies are used to obtain the active orbitals: either information from a cheaper electronic structure model is used or localized (or semi-localized) orbitals in a subregion of the molecular system defines the active orbital space. The success of the first strategy relies on the accuracy of the cheaper electronic structure model. The use of correlated natural transition orbitals[13,20] (CNTOs) to determine the active space is an example of such an approach. The CNTOs are similar to natural transition
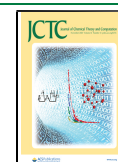
orbitals[21,22] (NTOs), which are extensively used for analysis[23,24] and in reduced cost methods for excited states,[25−29] but are defined by using excitation vectors that are parameterized with both single and double substitutions with respect to the reference determinant.

When an electronic excitation is localized in a region of the molecule, localized or semilocalized Hartree−Fock orbitals can be used to determine the active space. Cholesky orbitals have been used in MLCC calculations for both core and valence excitation energies.[2,3,12] Occupied Cholesky orbitals can be obtained through a partial, limited Cholesky decomposition of the idempotent Hartree−Fock density in the atomic orbital (AO) basis.[30,31] Virtual Cholesky orbitals can be obtained in the same way by considering the virtual Hartree−Fock density. This localization scheme is non-iterative and has cubic scaling with respect to the system size. Another option to determine the active virtual orbitals, which can be used in conjunction with occupied Cholesky orbitals, are the projected atomic orbitals (PAOs). PAOs have been used extensively in reduced cost electronic structure methods.[32−38] The construction of PAOs is also a non-iterative procedure with cubic scaling.

In MLCCSD, the double excitations of the cluster operator are restricted to the active orbital space. By also restricting the single excitations, we obtain a reduced space CCSD approach. There are several reduced space coupled cluster approaches, such as the frozen core approximation, the frozen natural orbital approaches,[39−44] and the LoFEx[28,29] and CorNFLEx[45] methods. The LoFEx and CorNFLEx methods are specialized for the calculation of accurate excitation energies. A truncated set of molecular orbitals (MOs) is determined by considering the dominant NTOs or CNTOs, obtained from a cheaper electronic structure method, and localized orbitals that overlap with these dominant NTOs/CNTOs. The reduced orbital space is increased until the excitation energy is converged to within a predefined threshold. As NTOs/CNTOs from a single excited state are used to determine the MOs that enter the coupled cluster calculation, LoFEx and CorNFLEx are state specific methods; the reduced space differs depending on the excited state.

In this work, we consider a reduced space CCSD approach (CCSD-in-HF) where Cholesky occupied orbitals and PAOs are used to obtain the active orbital space for a region of interest. Several excited states can be treated using the same truncated set of molecular orbitals, as long as the excitation processes are located in the region of interest. Preliminary studies using CC-in-HF to describe valence excitations have been reported.[18,31,46] An iterative procedure where the active space is increased and the excitation energies recomputed until convergence, as is done in LoFEx/CorNFLEx, is possible but has not yet been implemented.

In near edge X-ray absorption fine structure (NEXAFS) spectroscopy,[47] a core electron is excited. Since the binding energy of a core electron is unique to a given atomic number, specific energy ranges correspond to the K-edge NEXAFS spectrum for the different atoms. The excitation energies are sensitive to the environment of the core excited atom and NEXAFS spectra can be used to probe the local environment. Because of the strong interaction between the core hole and the excited electron, core excitation processes are generally localized in character.

With the development of the liquid microjet technique, studies of solutions and liquids with NEXAFS can be performed routinely. For reviews of the liquid microjet technique in soft X-ray spectroscopies, we refer the reader to refs 48 and 49. Proper interpretation of NEXAFS spectra relies on accurate theoretical modeling. While it can be challenging to accurately model the NEXAFS spectra of small molecules *in vacuo*, it is significantly more complicated for complex systems such as solutions and liquids.

The coupled cluster hierarchy of models can be used to accurately calculate core excitations, for instance by use of the core−valence separation[50] (CVS) approach of Coriani and Koch.[51,52] Typical errors of CCSD core excitation energies, obtained within the CVS approximation, are on the order of 1 eV. The errors can be significantly reduced by including triple excitations.[53−55] Intensities can be obtained from coupled cluster linear response theory or from EOM coupled cluster theory. Myhre et al.[12] calculated the MLCCSD NEXAFS spectra at the carbon and oxygen edge for ethanal, propenal, and butanal, demonstrating excellent agreement with the CCSD spectra. Their work showed that the multilevel coupled cluster models, using localized orbitals to determine the active space, is appropriate for the description of core excitation processes. While illustrating the usefulness of the MLCCSD model, this implementation was, as mentioned previously, not optimal and calculations on larger systems have not yet been performed.

In this paper, we consider the MLCCSD and CCSD-in-HF nitrogen K-edge spectra of adenine, adenosine, adenosine triphosphate (ATP), and an ATP−water system. As core excitation processes are spatially localized, the orbital space can be partitioned using occupied Cholesky orbitals and PAOs. For valence excitations, which are generally more delocalized in character, orbital selection can be more challenging and active spaces determined from CNTOs are often preferable. An exception is when the excitation of interest is localized in some known region of the system, e.g., in solvent−solute systems. The CCSD-in-HF approach relies on the selection of an active region and to treat delocalized valence excitations, a reduced space approach like LoFEx/CorNFLEx is more appropriate.

With the calculations presented in this paper, we outline a procedure to obtain accurate NEXAFS spectra for larger molecular systems, liquids, or solutions. First, a model system is used to determine the basis set and to ensure that the active space of the MLCCSD and CCSD-in-HF calculations is suitable for accurate treatment of the core excitations. Here, we use adenine and adenosine for this purpose. Afterward, the MLCCSD and CCSD-in-HF calculations are performed on the full system, i.e., ATP and the ATP−water system. The systems were selected because experimental spectra are available[56,57] and because ATP ($C_{10}H_{16}N_5O_{13}P_3$) is large enough that the full CCSD NEXAFS spectra is computationally expensive to generate. Another theoretical NEXAFS study on adenine and ATP, both in vacuum and in aqueous solution, has been performed at the DFT level of theory, using polarizable density embedding to describe the solvent.[58] In that study, a series of representative geometries for ATP solvated in water were considered; this is likely necessary in order to accurately describe the NEXAFS spectra of a solute. Our mission in the present study is not an accurate description of the experiment but rather to establish the performance of the EOM-MLCCSD and EOM-CCSD-in-HF implementations and their usefulness for modeling the NEXAFS spectra of complex systems.

■ THEORY

The coupled cluster wave function is given by

$$|CC\rangle = \exp(X)|HF\rangle, \quad X = \sum_{\mu} x_{\mu}\tau_{\mu} \tag{1}$$

where X is the cluster operator, $|HF\rangle$ is the Hartree–Fock reference, x_{μ} is the cluster amplitude, and τ_{μ} is the excitation operator. The standard models within the coupled cluster hierarchy are obtained by restricting X to include excitation operators up to a certain order. The cluster amplitudes are determined through the projected coupled cluster equations,

$$\Omega_{\mu} = \langle\mu|\bar{H}|HF\rangle = 0 \tag{2}$$

and the energy is obtained from

$$E_{CC} = \langle HF|\bar{H}|HF\rangle \tag{3}$$

where $\bar{H} = \exp(-T)H\exp(T)$ and

$$H = \sum_{pq} h_{pq}E_{pq} + \frac{1}{2}\sum_{pqrs} g_{pqrs}(E_{pq}E_{rs} - E_{ps}\delta_{qr}) + h_{nuc} \tag{4}$$

is the non-relativistic electronic Hamiltonian operator in terms of the singlet operators E_{pq}.[59] The $g_{pqrs} = (pq|rs)$ are the electron repulsion integrals in the Mulliken notation.

Equation-of-Motion Coupled Cluster Theory. In the equation-of-motion (EOM) coupled cluster framework,[14−17] a general state is expressed as

$$|k\rangle = e^{T}\left(|HF\rangle R_{0}^{k} + \sum_{\mu>0}|\mu\rangle R_{\mu}^{k}\right)$$
$$= \sum_{\mu\geq0} e^{T}R_{\mu}^{k}|\mu\rangle \tag{5}$$

where $|\mu\rangle = \tau_{\mu}|HF\rangle$, and R^{k} is obtained as the right eigenvectors of the similarity transformed Hamiltonian,

$$\bar{H} = \begin{pmatrix} 0 & \eta^{T} \\ 0 & A \end{pmatrix} + E_{CC}I \tag{6}$$

Here, A is the Jacobian matrix, with elements $A_{\mu\nu} = \langle\mu|[\bar{H}, \tau_{\nu}]|HF\rangle$ and $\eta_{\nu} = \langle HF|[\bar{H}, \tau_{\nu}]|HF\rangle$, and we have assumed that the ground state amplitudes have been determined from eq 2. As $\bar{H}$ is not Hermitian, the left eigenvectors differ from the right eigenvectors. We have the EOM coupled cluster left states

$$\langle k| = \sum_{\mu\geq0} L_{\mu}^{k}\langle\mu|\exp(-T)) \tag{7}$$

and we require that the left and right eigenvectors satisfy the biorthonormalization criterion

$$\langle k|l\rangle = \delta_{kl} \tag{8}$$

The ground state solutions, L^{0} and R^{0}, are given by

$$L^{0} = \begin{pmatrix} 1 \\ l_{0} \end{pmatrix}, \quad R^{0} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{9}$$

where l_{0} is the ground state multipliers determined from

$$A^{T}l_{0} = -\eta \tag{10}$$

The excited state solutions, L^{k} and R^{k} for $k > 0$, are given by

$$L^{k} = \begin{pmatrix} 0 \\ l_{k} \end{pmatrix}, \quad R^{k} = \begin{pmatrix} r_{0} \\ r_{k} \end{pmatrix} \tag{11}$$

where l_{k} and r_{k} are left and right eigenvectors of A, respectively, and $r_{0} = -l_{0} \cdot r_{k}$. The eigenvalues of A are the excitation energies, ω_{k}.

Oscillator strengths for transitions between the ground and the kth excited state are given by

$$f_{k} = \frac{2}{3}\omega_{k}\langle0|\mu|k\rangle\cdot\langle k|\mu|0\rangle \tag{12}$$

where $\mu^{\alpha} = \sum_{pq} \mu_{pq}^{\alpha}E_{pq}$ is the α component of dipole operator.[19,60]

Multilevel CCSD. In the multilevel coupled cluster theory, we restrict the higher order excitations of the cluster operator to an active orbital space. In the two-level CCS/CCSD approach, the cluster operator assumes the form

$$X^{MLCCSD} = X_{1} + T_{2} \tag{13}$$

The single excitation operator,

$$X_{1} = \sum_{\mu_{1}} x_{\mu_{1}}\tau_{\mu_{1}} = \sum_{AI} x_{I}^{A}E_{AI} \tag{14}$$

includes single excitations in the entire orbital space, that is, the summation indices A and I label general (active and inactive) virtual and occupied orbitals, respectively. The double excitation operator T_{2} is restricted to the active orbital space,

$$T_{2} = \sum_{\mu_{2}} t_{\mu_{2}}\tau_{\mu_{2}} = \frac{1}{2}\sum_{aibj} t_{ij}^{ab}E_{ai}E_{bj} \tag{15}$$

where the summation indices a, b and i, j label active virtual and occupied orbitals, respectively.

The MLCCSD ground state equations for the two-level CCS/CCSD model are

$$\Omega_{\mu_{1}} = \langle\mu_{1}|\hat{H} + [\hat{H}, T_{2}]|HF\rangle = 0 \tag{16}$$

$$\Omega_{\mu_{2}} = \langle\mu_{2}|\hat{H} + [\hat{H}, T_{2}] + \frac{1}{2}[[\hat{H}, T_{2}],T_{2}]|HF\rangle = 0 \tag{17}$$

where $\hat{H}$ is the X_{1}-transformed Hamiltonian and the doubles projection space is associated with T_{2}. These equations are equivalent to the standard CCSD ground state equations, except for the restriction of the T_{2} operator to the active space.

Properties of this MLCCSD model can be obtained within the EOM framework. The excited states ($|k\rangle$, $\langle k|$) are constructed by solving the eigenvalue equations of the MLCCSD (CCS/CCSD) Jacobian matrix,

$$A^{MLCCSD}r_{k} = \omega_{k}r_{k} \tag{18}$$

$$(A^{MLCCSD})^{T}l_{k} = \omega_{k}l_{k} \tag{19}$$

where

$$A^{MLCCSD} = \begin{pmatrix} \langle\mu_{1}|[\hat{H}, \tau_{\nu_{1}}] + [[\hat{H}, \tau_{\nu_{1}}],T_{2}]|HF\rangle & \langle\mu_{1}|[\hat{H}, \tau_{\nu_{2}}]|HF\rangle \\ \langle\mu_{2}|[\hat{H}, \tau_{\nu_{1}}] + [[\hat{H}, \tau_{\nu_{1}}],T_{2}]|HF\rangle & \langle\mu_{2}|[\hat{H}, \tau_{\nu_{2}}] + [[\hat{H}, \tau_{\nu_{2}}],T_{2}]|HF\rangle \end{pmatrix} \tag{20}$$

Note that A^{MLCCSD} assumes the same form as A^{CCSD} except for restriction of the operator T_{2} and the corresponding projection space. Core excited states can be obtained using the core−valence separation (CVS) approach of Coriani and

Koch.[51,52] In this approach, the non-zero elements of the excitation vectors have at least one occupied index belonging to the excited core orbital. It can be implemented as a projection[51,52] or by implementing the linear transformation by the CVS Jacobian matrix directly.[61,62] Once the left and right states are constructed and the multipliers are determined from eq (10), then the MLCCSD oscillator strengths can be calculated according to eq (12).

Partitioning the Orbital Space. The first step of any multilevel coupled cluster calculation is to partition the molecular orbitals into the active and inactive orbital sets. The canonical Hartree–Fock orbitals are not suitable to determine the active space. If the property of interest is spatially localized in the molecular system, such as core excitations or excitations in a target molecule in a solvent, localized orbitals can be used.

For the occupied space, there are many widely used iterative localization procedures, such as the Boys,[63] Pipek–Mezey,[64] and Edmiston–Ruedenberg[65] procedures. In this work, we use the semilocalized Cholesky orbitals described in refs 30, 31, which can be obtained in a non-iterative procedure. A set of active atoms are selected and the idempotent Hartree–Fock density,

$$D_{\alpha\beta} = \sum_I C_{\alpha I} C_{\beta I} \tag{21}$$

is Cholesky decomposed in a specialized procedure where the pivoting elements are restricted to correspond to AOs on the active atoms. The decomposition procedure ends when all "active" diagonals fall below a given threshold. After the decomposition, the Cholesky factors are the orbital coefficients of the active occupied orbitals, C^a:

$$D = C^a (C^a)^T + D^e \tag{22}$$

The density of the inactive space, D^e, can be fully Cholesky decomposed to yield the inactive occupied orbitals.

The iterative localization procedures, which are extensively used for the occupied space, can also be applied to the virtual space. However, convergence for the virtual space is more challenging and the use of sophisticated level-shift and trust-radius solvers are often necessary.[66] The projected atomic orbitals (PAOs), is an alternative to such iterative localization procedures for the virtual space. To construct PAOs in an active region of the molecular system, the occupied orbitals are projected out of the AOs centered on the active atoms. The orbital coefficient matrix for the active virtual PAOs is

$$C^{\mathrm{PAO}} = I - DS' \tag{23}$$

where S' is rectangular and contains the columns of the AO overlap matrix that correspond to AOs centered on the active atoms. These orbitals are non-orthogonal and linearly dependent. The Löwdin canonical orthonormalization procedure[67] can be used to obtain a set of orthonormal active virtual orbitals. Linear dependence in the full set of AOs should be removed before the PAO construction. The inactive virtual orbitals are obtained in a similar way. The active virtual orbitals, as well as the occupied orbitals, are projected out of the full set of AOs. The resulting orbitals are orthonormalized.

After the orbitals have been partitioned, we block diagonalize the occupied–occupied and virtual–virtual Fock matrices such that the active–active and inactive–inactive blocks become diagonal. This is achieved by rotating among the active orbitals and among the inactive orbitals separately.

This semicanonical basis is used throughout the MLCCSD calculation as this significantly improves convergence.

Reduced Space CCSD. In MLCCSD, the double excitations included in the cluster operator are restricted to an active space. In the reduced space CCSD-in-HF approach, the single excitations are also restricted to the active space. The active orbitals are determined from occupied Cholesky orbitals and PAOs. The inactive orbitals contribute through the Fock matrix,

$$
\begin{aligned}
F_{pq} &= h_{pq} + \sum_i \left(2g_{pqii} - g_{piiq}\right) + \sum_{\mathcal{I}} \left(2g_{pq\mathcal{I}\mathcal{I}} - g_{p\mathcal{I}\mathcal{I}q}\right) \\
&= h_{pq} + \sum_i \left(2g_{pqii} - g_{piiq}\right) + F_{pq}^{\mathrm{e}}
\end{aligned}
\tag{24}
$$

where p and q are general active space indices, i denotes an active occupied orbital, and the index $\mathcal{I}$ denotes an inactive occupied orbital. The CCSD-in-HF calculation is performed as a standard CCSD calculation but with a truncated MO basis consisting of the active orbitals ($N_{\mathrm{MO}} < N_{\mathrm{AO}}$) and with the effective Fock matrix of eq 23.

■ RESULTS AND DISCUSSION

As a test study for the EOM-MLCCSD and EOM-CCSD-in-HF implementations, we consider the nitrogen core excitations of adenine, adenosine, adenosine triphosphate (ATP) *in vacuo*, and ATP with 12 water molecules, see Figure 1. These systems
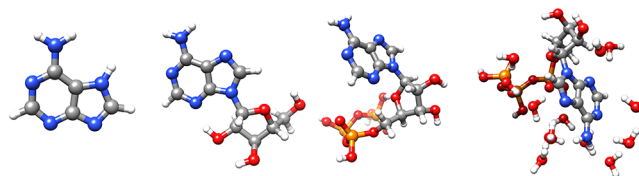

Figure 1. Adenine, adenosine, adenosine triphosphate (ATP), and ATP with 12 water molecules.

are chosen because of their biological importance and the availability of experimental studies.[56,57] In particular, experimental NEXAFS spectra at the nitrogen and carbon edge of adenosine triphosphate in aqueous solution has been reported.[57] Our goal in this paper is not to perform an accurate application study but rather to demonstrate the performance, in terms of accuracy and cost, of the MLCCSD and CCSD-in-HF methods. This has dictated our choice of basis sets and the number of computed states. Furthermore, in order to properly describe the effects of solvents, one should sample the spectra at several representative geometries, e.g., obtained from a molecular dynamics simulation. Bulk solvent should also be included in the system, for instance, by using the QM/MM framework,[68–70] polarizable continuum model,[71,72] or by treating all water molecules at the Hartree–Fock level of theory. In general, triple excitations (CC3[73]) are needed to obtain quantitative, unshifted NEXAFS spectra; this is demonstrated for adenine. However, shifted CCSD spectra can be useful for qualitative interpretation of experiments.

All geometries, except the ATP–water and methylamine–water geometries, are obtained at the B3LYP/aug-cc-pVDZ level using the NWChem[74] software. The ATP–water and methylamine–water geometries were built using the Avogadro
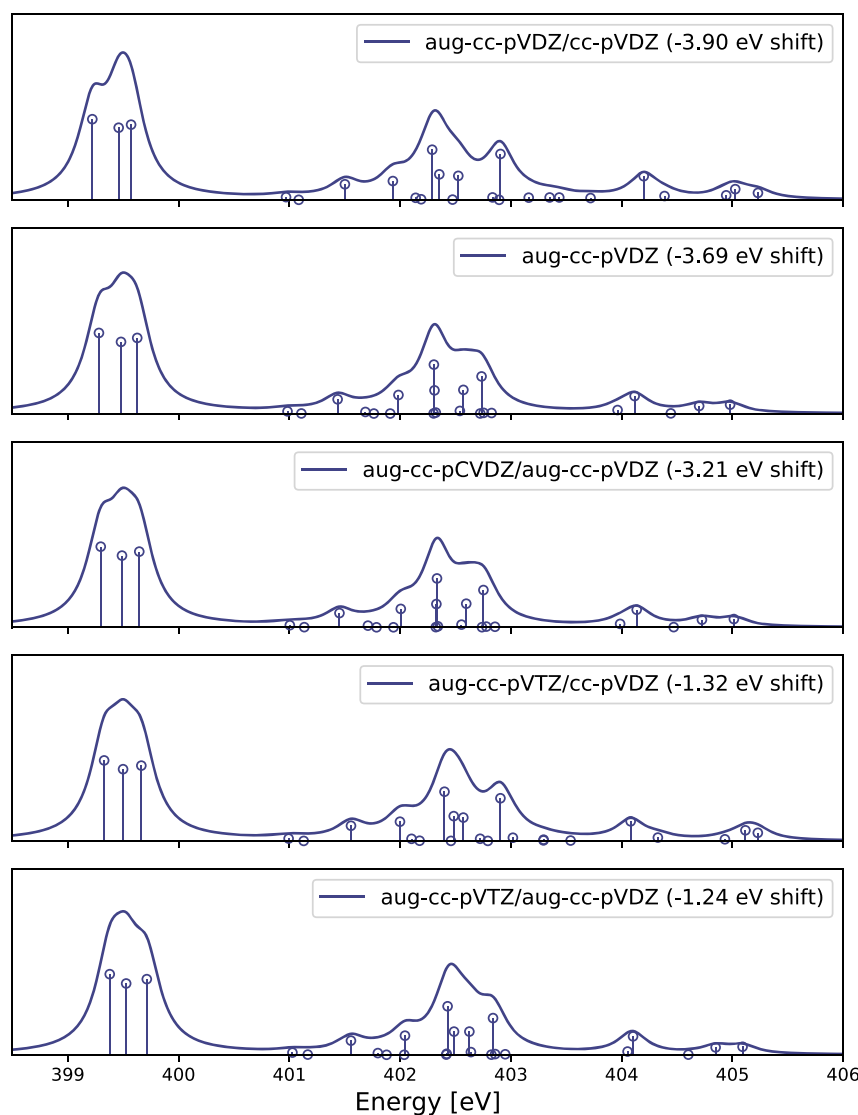
Figure 2. Adenine NEXAFS nitrogen edge calculated at the CCSD level of theory with combinations of the Dunning correlation consistent basis sets. When a larger basis set is used on the nitrogen atoms, we use the notation: *basis-on-nitrogen/basis-on-other-atoms*. Lorentzian broadening with 0.3 eV FWHM has been applied. The first peak has been shifted to the experimental value 399.5 eV, as reported in ref 56. Six roots have been calculated for each nitrogen atom.

software package.[75] All geometries are available from ref 76. Visualization of the molecular systems is done using the Chimera software package.[77]

The EOM-MLCCSD oscillator strengths were implemented in a development branch of the eT program,[46] and all calculations are performed with eT. The following thresholds have been used: For the ground state, we used a threshold of 10^{-6} on $|\Omega|$ and on the residual of the multiplier equations. For the excited states, a threshold of 10^{-4} was used for the residual and 10^{-6} on the change in the excitation energies. The electron repulsion integrals are Cholesky decomposed, and the decomposition threshold is 10^{-8} for adenine and 10^{-6} for the remaining calculations. Generally, a decomposition threshold of 10^{-3} is sufficient for accurate excitation energies. All timings were performed on two Intel Xeon Gold 6138 processors, using 40 threads, and 360 GB of memory was available in all calculations. In the EOM-MLCCSD and EOM-CCSD-HF calculations, we use Cholesky-PAOs to partition the orbital space and the "adenine part" of adenosine and ATP is considered active.

Adenine. In order to select the basis set for the larger systems, we start by considering the EOM-CCSD nitrogen K-edge spectrum of adenine for different combinations of Dunning correlation consistent basis sets.[79,80] The results are given in Figure 2. Generally, we use a larger basis set on the nitrogen atoms than on the carbon and hydrogen atoms. From the spectra with aug-cc-pVDZ and aug-cc-pCVDZ/aug-cc-pVDZ, we observe that additional core functions on the nitrogen atoms shifts the spectrum but that the overall shape is unchanged. From the two lower panels, we see that the shift, with respect to the experimental value of 399.5 eV for the first peak,[56] is significantly reduced by using triple-zeta rather than double-zeta on the nitrogen atoms. Although the finer details might differ, the main features of the spectrum are the same for all basis set combinations. As a compromise between accuracy and cost, we use the aug-cc-pVTZ/cc-pVDZ basis sets for the remaining calculations.

In Figure 3, the EOM-CCSD/aug-cc-pVTZ/cc-pVDZ and EOM-CC3/aug-cc-pVTZ/cc-pVDZ[62] are compared to the experimental spectrum from ref 56. The inclusion of triple
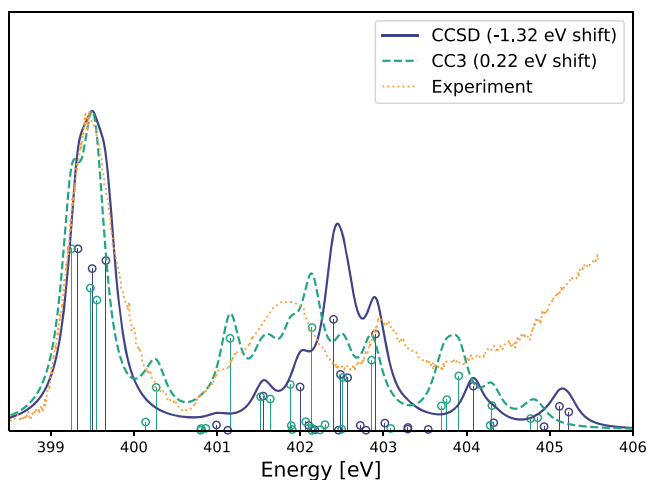
Figure 3. Adenine NEXAFS nitrogen edge calculated at the CCSD and CC3 level of theory with the aug-cc-pVTZ/cc-pVDZ basis set. The larger basis set is used on the nitrogen atoms. The first peak has been shifted to the experimental value 399.5 eV, as reported in ref 56. Six roots have been calculated for each nitrogen atom. Lorentzian broadening with 0.3 eV FWHM has been applied. Experimental data collected from ref 56 using WebPlotDigitizer.[78]

excitations significantly improves the computed spectrum for the second feature at approximately 402 eV. Furthermore, more than six roots per nitrogen atom are necessary in order to describe the spectrum from 403 eV. To accurately describe Rydberg states, additional diffuse basis functions are likely needed.

Adenosine. To illustrate the performance of EOM-MLCCSD and EOM-CCSD-in-HF compared to EOM-CCSD for core excitations and oscillator strengths, we consider the nitrogen edge NEXAFS spectrum of adenosine (conformer 1 in Figure 4) calculated using the aug-cc-pVTZ/cc-pVDZ
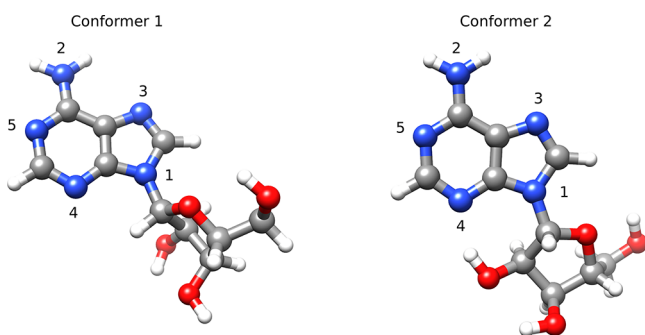

Figure 4. Adenosine conformers 1 and 2 with labels corresponding to Tables 1 and 2.

basis. The results are given in Figure 5, Table 1, and Table 2. In the MLCCSD and CCSD-in-HF calculations for conformer 1, there are 40 active occupied orbitals and 301 active virtual orbitals, and there are 30 inactive occupied orbitals and 120 inactive virtual orbitals.

For nitrogen atoms 2−5 (see Table 1), the calculated MLCCSD and CCSD-in-HF excitation energies for adenosine has an error with respect to CCSD of less than 0.1 eV. This is well within the expected error of CCSD for K-edge core excitations. The errors are generally larger for nitrogen atom 1 (N1). This is because not all nearest neighbours of N1 are defined as active. The errors for N1 can be reduced by

including the neighboring carbon on the ribose into the set of active atoms, as seen from Table 1. As seen from Figure 5 and Tables 1 and 2, the errors of MLCCSD and CCSD-in-HF, compared to full CCSD, are small: the spectra coincide for all but the low intensity peaks at around 407 eV. The MLCCSD and CCSD-in-HF spectra are almost indistinguishable, with small differences observed for the high energy excitations only.

Wall times are given in Table 3. We report average timings from the calculations of the nitrogen atoms for the cluster amplitudes (t^{gs}) and the average time to transform by the Jacobian matrix, A and A^T (t^A and t^{A^T}). Although the active space is quite large in these calculations, the computational savings are significant with approximately a factor of five for t^{gs}, t^A, and t^{A^T} compared to full CCSD. Specialized active spaces for each of the five MLCCSD and CCSD-in-HF calculations can be used where we include only the neighbors of the nitrogen atom being excited. This can significantly reduce the cost but will probably lead to increased errors with respect to the full CCSD calculation.

The time to converge the excited states varies greatly because the number of iterations required to reach convergence varies for the different nitrogen atoms. To give a perspective on the computational savings achieved in these MLCCSD calculations, we compare the calculation time of the cheapest CCSD calculation (N2) with the most expensive MLCCSD calculation (N4); the cheapest CCSD calculation used 9 days and 16 h, whereas the most expensive MLCCSD calculation used 1 day and 8 h. There are two contributing factors to the savings with MLCCSD: the savings due to the reduced cost of constructing the Ω vector and performing the linear transformation by A and A^T, and the reduction of IO in the MLCCSD calculations. The IO is reduced because the reduced space of the Davidson procedure can be stored in memory for MLCCSD, whereas this is not possible for CCSD with 360 GB of memory available.

In Figure 5, we also present the EOM-MLCCSD/aug-cc-pVTZ/cc-pVDZ and EOM-CCSD-in-HF/aug-cc-pVTZ/cc-pVDZ nitrogen K-edge spectra for conformer 2 of adenosine. For conformer 2, there are 42 active occupied orbitals and 303 active virtual orbitals, and there are 28 inactive occupied orbitals and 118 inactive virtual orbitals. We observe a change in the spectrum, compared to conformer 1, resulting from the change in the environment of the nitrogen atoms. Again, the MLCCSD and CCSD-in-HF models give very similar results, as seen for conformer 1.

ATP and ATP−Water. The EOM-MLCCSD and EOM-CCSD-in-HF methods can be used to treat systems for which full EOM-CCSD is too expensive. The EOM-MLCCSD/aug-cc-pVTZ/cc-pVDZ and EOM-CCSD-in-HF/aug-cc-pVTZ/cc-pVDZ nitrogen K-edge spectrum of ATP is given Figure 6. There are 44 active occupied orbitals and 303 active virtual orbitals and 86 inactive occupied orbitals and 253 inactive virtual orbitals. The most expensive of the five MLCCSD calculations for this system completed in 2 days and 23 h. As for adenosine, the MLCCSD and CCSD-in-HF spectra coincide for all but the high energy excitations around 407 eV.

In Figure 7, we present the nitrogen K-edge spectrum of the ATP−water system (see Figure 1). In this system, there are 58 active occupied and 310 active virtual orbitals and 132 inactive occupied and 474 inactive virtual orbitals. The MLCCSD and CCSD-in-HF results coincide well, but some differences are observed for higher excitation energies. This is likely because
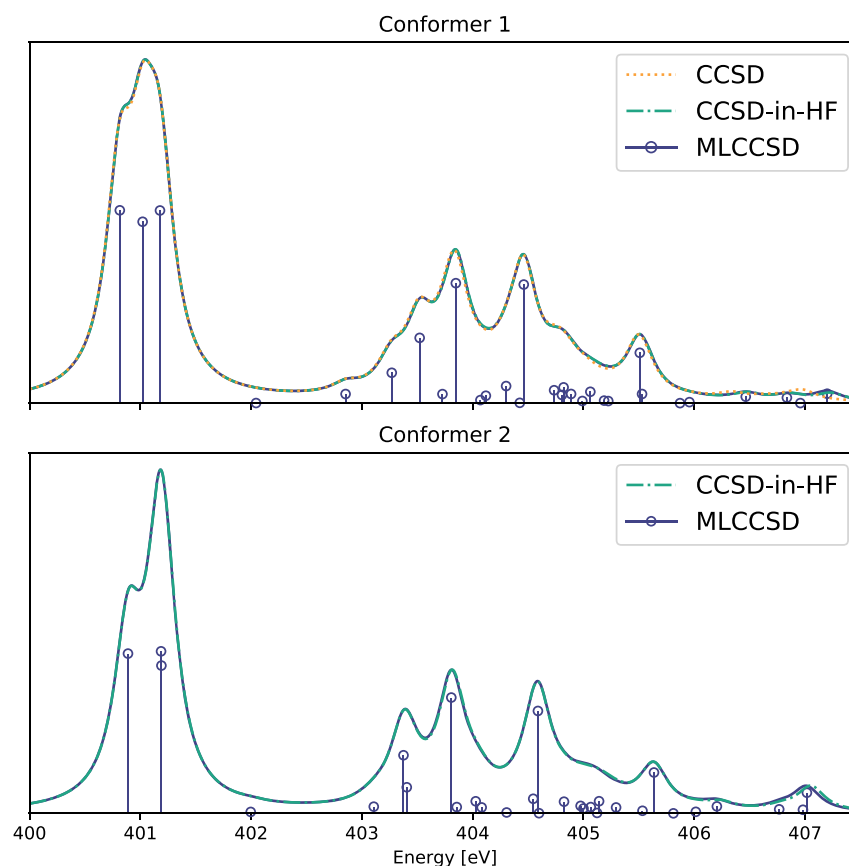
Figure 5. Nitrogen K-edge NEXAFS spectra of two conformers of Adenosine calculated at the EOM-MLCCSD/aug-cc-pVTZ/cc-pVDZ and EOM-CCSD-in-HF/aug-cc-pVTZ/cc-pVDZ levels of theory. Six roots have been calculated for each nitrogen atom. Lorentzian broadening with 0.3 eV FWHM has been applied.

Table 1. MLCCSD/aug-cc-pVTZ/cc-pVDZ Core Excitations out of the Nitrogen 1s Orbitals for Adenosine (Conformer 1)[a]

nitrogen	$\omega_1 \ (\Delta\omega_1)$	$\omega_2 \ (\Delta\omega_2)$	$\omega_3 \ (\Delta\omega_3)$	$\omega_4 \ (\Delta\omega_4)$	$\omega_5 \ (\Delta\omega_5)$	$\omega_6 \ (\Delta\omega_6)$
N1	403.848 (0.011)	405.529 (0.041)	406.465 (0.130)	406.836 (0.207)	406.958 (0.120)	407.200 (0.219)
N2	403.269 (0.004)	403.521 (0.007)	404.461 (0.004)	405.509 (0.006)	405.871 (0.012)	405.956 (0.009)
N3	400.815 (0.004)	404.119 (0.031)	404.302 (0.013)	404.821 (0.020)	404.990 (0.068)	405.186 (0.053)
N4	401.021 (<0.001)	402.853 (0.001)	404.069 (0.026)	404.806 (0.022)	405.062 (0.046)	405.224 (0.028)
N5	401.177 (<0.001)	402.044 (<0.001)	403.727 (0.004)	404.425 (0.006)	404.735 (0.013)	404.888 (0.007)
N1†	403.845 (0.008)	405.503 (0.015)	406.407 (0.072)	406.755 (0.126)	406.915 (0.077)	407.104 (0.123)

[a]Excitation energies, ω_i, and errors with respect to CCSD, $\Delta\omega_i$, are given in eV. † = nearest neighboring carbon on ribose included in active space.

Table 2. CCSD-in-HF/aug-cc-pVTZ/cc-pVDZ Core Excitations out of the Nitrogen 1s Orbitals for Adenosine (Conformer 1)[a]

nitrogen	$\omega_1 \ (\Delta\omega_1)$	$\omega_2 \ (\Delta\omega_2)$	$\omega_3 \ (\Delta\omega_3)$	$\omega_4 \ (\Delta\omega_4)$	$\omega_5 \ (\Delta\omega_5)$	$\omega_6 \ (\Delta\omega_6)$
N1	403.850 (0.013)	405.526 (0.038)	406.474 (0.140)	406.854 (0.224)	406.961 (0.123)	407.232 (0.251)
N2	403.269 (0.004)	403.522 (0.008)	404.461 (0.004)	405.508 (0.004)	405.874 (0.016)	405.957 (0.010)
N3	400.815 (0.004)	404.119 (0.031)	404.300 (0.011)	404.818 (0.018)	404.998 (0.076)	405.187 (0.054)
N4	401.021 (<0.001)	402.855 (0.002)	404.074 (0.032)	404.807 (0.023)	405.068 (0.052)	405.226 (0.030)
N5	401.178 (0.001)	402.045 (0.001)	403.726 (0.003)	404.426 (0.007)	404.736 (0.014)	404.888 (0.008)

[a]Excitation energies, ω_i, and errors with respect to CCSD, $\Delta\omega_i$, are given in eV.

Table 3. Average Wall Times to Solve for the Ground State Equations, t^{gs}, and Average Time for Transformations by A^T (t^{A^T}) and A (t^A) of Adenosine (Conformer 1)

model	t^{gs} [min]	t^A [min]	t^{A^T} [min]
CCSD-in-HF	19	2	3
MLCCSD	22	2	3
CCSD	128	11	14

MLCCSD, with CCS on the whole system, offers an improved description with more diffuse core excited states. Including additional Rydberg functions on the active atoms could be important for the description of these states in both methods. The MLCCSD calculations are more expensive than the CCSD-in-HF calculations; timings for transformations by A and A^T are given in Table 4.
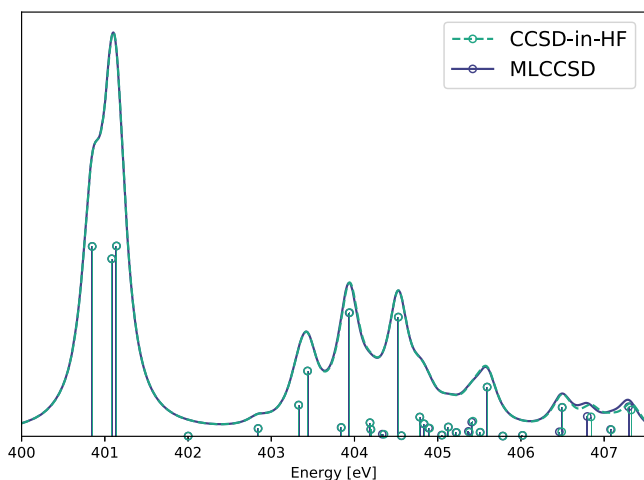
Figure 6. Nitrogen K-edge NEXAFS spectra calculated at the MLCCSD/aug-cc-pVTZ/cc-pVDZ and CCSD-in-HF/aug-cc-pVTZ/cc-pVDZ level for ATP. Six roots have been calculated for each nitrogen atom. Lorentzian broadening with 0.3 eV FWHM has been applied.
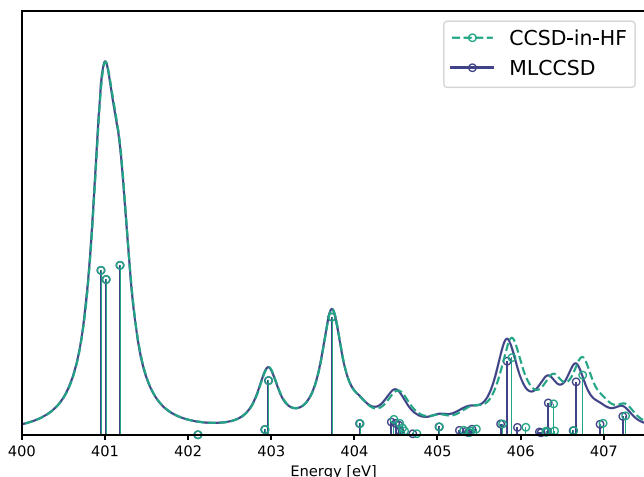

Figure 7. Nitrogen K-edge NEXAFS spectra calculated at the MLCCSD/aug-cc-pVTZ/cc-pVDZ and CCSD-in-HF/aug-cc-pVTZ/cc-pVDZ level for ATP and 12 water molecules. Six roots have been calculated for each nitrogen atom. Lorentzian broadening with 0.3 eV FWHM has been applied.

Table 4. MLCCSD and CCSD-in-HF Calculations on the ATP–Water System Using the aug-cc-pVTZ/cc-pVDZ Basis and Average Wall Times to Transformations by A^T (t^{A^T}) and A (t^A)

model	t^A[min]	t^{A^T}[min]
CCSD-in-HF	5	7
MLCCSD	33	37

With the calculation on ATP and the ATP–water system, we demonstrate that the MLCCSD and CCSD-in-HF approaches can be used to treat sizable molecules and to include solvent effects explicitly. To fully capture the effects of the solvent, the calculations must be performed on several representative geometries. Additionally, one should increase the number of water molecules included in the calculation, treating most water molecules at a lower level of theory. For such a study, the CCSD-in-HF approach is preferable since

accuracy is comparable to MLCCSD, but the cost is significantly lower.

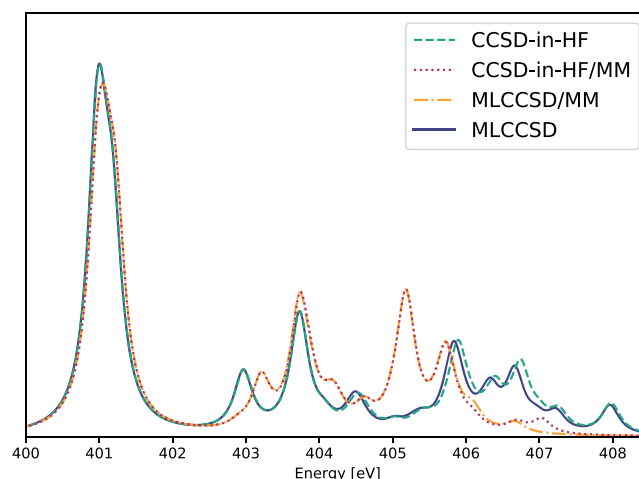The MLCCSD and CCSD-in-HF models can be used in a QM/MM framework. In Figure 8, we compare the MLCCSD

Figure 8. Nitrogen K-edge NEXAFS spectra calculated at the MLCCSD/aug-cc-pVTZ/cc-pVDZ, MLCCSD/MM/aug-cc-pVTZ/cc-pVDZ, CCSD-in-HF/aug-cc-pVTZ/cc-pVDZ, and CCSD-in-HF/MM/aug-cc-pVTZ/cc-pVDZ levels for ATP and 12 water molecules. Six roots have been calculated for each nitrogen atom. Lorentzian broadening with 0.3 eV FWHM has been applied.

and CCSD-in-HF spectra for the ATP–water system to the QM/MM approach where the QM region (ATP) is treated with either CCSD-in-HF or MLCCSD and the MM region (the 12 water molecules) is treated with electrostatic embedding.[81] There is good agreement between the pure QM and the QM/MM spectra for the peaks in the range of 401−405 eV. For the remaining peaks of higher energy, the differences are significant.

In order to assess the quality of the MLCCSD and CCSD-in-HF approaches compared to the QM/MM approach with electrostatic embedding, we consider the nitrogen K-edge spectrum of methylamine and eight water molecules (see Figure 9) The results are given in Figure 10. The active region, and the QM region of the QM/MM calculation, is methylamine. The MLCCSD, CCSD-in-HF and CCSD/MM spectra are shifted such that the most intense peak is aligned with the CCSD spectrum. Contrary to the CCSD/MM spectrum, the
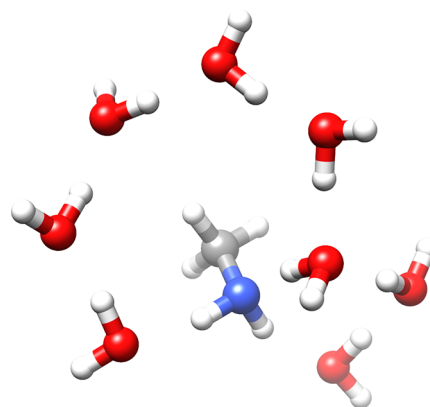

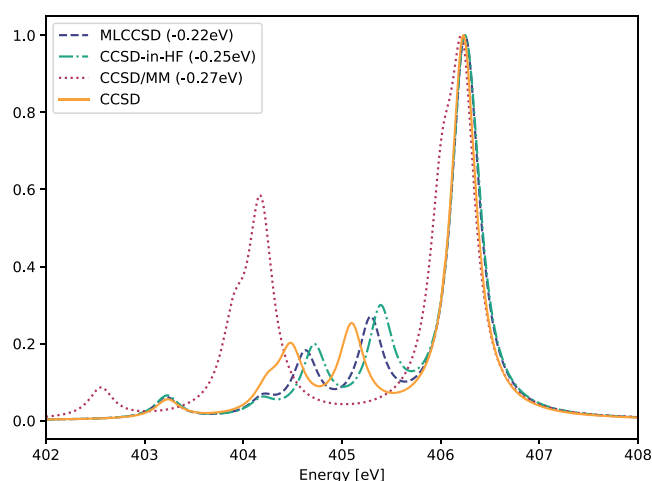**Figure 9.** Methylamine and eight water molecules.

Figure 10. Nitrogen K-edge of methylamine−water system. Comparison of CCSD, MLCCSD, CCSD-in-HF, and CCSD/MM using the aug-cc-pVTZ basis on the nitrogen atom and the cc-pVDZ basis on the remaining atoms. The MLCCSD, CCSD-in-HF, and CCSD/MM spectra are shifted so that the most intense peak is aligned with the corresponding CCSD peak. Lorentzian broadening with 0.3 eV FWHM has been applied.

MLCCSD and CCSD-in-HF spectra capture the features of the CCSD spectrum. MLCCSD performs better than CCSD-in-HF, but not significantly.

CONCLUDING REMARKS

We have presented an implementation of EOM-MLCCSD oscillator strengths for the two level CCS/CCSD model. The model can be used to simulate UV/visible and NEXAFS spectroscopies with CCSD quality at a significantly reduced cost, given that an adequate active orbital space is employed. In this paper, we have partitioned the orbital space by using occupied Cholesky orbitals and PAOs localized in a subregion of the molecular system. This orbital selection procedure is suitable for localized excitation processes such as core excitations. The CCS/CCSD model, and a reduced space CCSD model (CCSD-in-HF), has been applied to the nitrogen K-edge spectra of adenosine, adenosine triphosphate (ATP), and an ATP−water system. With these calculations, we have demonstrated that MLCCSD and CCSD-in-HF are useful for the accurate modeling of the NEXAFS spectra of complex molecular systems. Our results indicate that CCSD-in-HF may be the preferable approach to treat low lying core excitations, as the method has lower costs at no significant loss of accuracy. The MLCCSD approach, or a combined MLCCSD-in-HF approach, may be preferable for calculations with smaller active spaces and for more delocalized core excitation processes. Furthermore, MLCCSD can be used to assess the quality of CCSD-in-HF calculations. We believe that both models, together with a geometry sampling from, e.g., a molecular dynamics simulation, will provide a useful theoretical tool for the interpretation of experimental NEXAFS spectra of solvents and liquids.

AUTHOR INFORMATION

Corresponding Author

Henrik Koch − *Scuola Normale Superiore, Pisa 56126, Italy; Department of Chemistry, Norwegian University of Science and Technology, Trondheim N-7491, Norway;* ⬤ orcid.org/0000-0002-8367-8727; Email: henrik.koch@sns.it

Author

Sarai Dery Folkestad − *Department of Chemistry, Norwegian University of Science and Technology, Trondheim N-7491, Norway;* ⬤ orcid.org/0000-0002-8569-4886

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c00707

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Myhre, R. H.; Sánchez de Merás, A. M. J.; Koch, H. The extended CC2 model ECC2. *Mol. Phys.* **2013**, *111*, 1109−1118.

(2) Myhre, R. H.; Sánchez de Merás, A. M. J.; Koch, H. Multi-level coupled cluster theory. *J. Chem. Phys.* **2014**, *141*, 224105.

(3) Myhre, R. H.; Koch, H. The multilevel CC3 coupled cluster model. *J. Chem. Phys.* **2016**, *145*, 44111.

(4) Oliphant, N.; Adamowicz, L. Multireference coupled-cluster method using a single-reference formalism. *J. Chem. Phys.* **1991**, *94*, 1229−1235.

(5) Piecuch, P.; Oliphant, N.; Adamowicz, L. A state-selective multireference coupled-cluster theory employing the single-reference formalism. *J. Chem. Phys.* **1993**, *99*, 1875−1900.

(6) Köhn, A.; Olsen, J. Coupled-cluster with active space selected higher amplitudes: Performance of seminatural orbitals for ground and excited state calculations. *J. Chem. Phys.* **2006**, *125*, 174110.

(7) 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.

(8) Purvis, G. D., III; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910−1918.

(9) Christiansen, O.; Koch, H.; Jørgensen, P. The second-order approximate coupled cluster singles and doubles model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409−418.

(10) Koch, H.; Jo/rgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, *93*, 3333−3344.

(11) Pedersen, T. B.; Koch, H. Coupled cluster response functions revisited. *J. Chem. Phys.* **1997**, *106*, 8059−8072.

(12) Myhre, R. H.; Coriani, S.; Koch, H. Near-edge X-ray absorption fine structure within multilevel coupled cluster theory. *J. Chem. Theory Comput.* **2016**, *12*, 2633−2643.

(13) Folkestad, S. D.; Koch, H. Multilevel CC2 and CCSD Methods with Correlated Natural Transition Orbitals. *J. Chem. Theory Comput.* **2019**, *16*, 179.

(14) Sekino, H.; Bartlett, R. J. A linear response, coupled-cluster theory for excitation energy. *Int. J. Quantum Chem.* **1984**, *26*, 255−265.

(15) Geertsen, J.; Rittby, M.; Bartlett, R. J. The equation-of-motion coupled-cluster method: Excitation energies of Be and CO. *Chem. Phys. Lett.* **1989**, *164*, 57−62.

(16) Bartlett, R. J. Coupled-cluster theory and its equation-of-motion extensions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 126−138.

(17) Krylov, A. I. Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker's guide to Fock space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433.

(18) Folkestad, S. D.; Kjønstad, E. F.; Goletto, L.; Koch, H. Multilevel CC2 and CCSD in reduced orbital spaces: electronic excitations in large molecular systems. Submitted 2020.

(19) 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*, 7029−7039.

(20) Høyvik, I.-M.; Myhre, R. H.; Koch, H. Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models. *J. Chem. Phys.* **2017**, *146*, 144109.

(21) Luzanov, A. V.; Sukhorukov, A. A.; Umanskii, V. É. Application of transition density matrix for analysis of excited states. *Theor. Exp. Chem.* **1976**, *10*, 354−361.

(22) Martin, R. L. Natural transition orbitals. *J. Chem. Phys.* **2003**, *118*, 4775−4777.

(23) Dreuw, A.; Head-Gordon, M. Single-reference ab initio methods for the calculation of excited states of large molecules. *Chem. Rev.* **2005**, *105*, 4009−4037.

(24) Plasser, F.; Wormit, M.; Dreuw, A. New tools for the systematic analysis and visualization of electronic excitations. I. Formalism. *J. Chem. Phys.* **2014**, *141*, No. 024106.

(25) Mata, R. A.; Stoll, H. An incremental correlation approach to excited state energies based on natural transition/localized orbitals. *J. Chem. Phys.* **2011**, *134*, No. 034122.

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

(27) Helmich, B.; Hättig, C. A pair natural orbital implementation of the coupled cluster model CC2 for excitation energies. *J. Chem. Phys.* **2013**, *139*, No. 084114.

(28) 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.

(29) 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*, 2135−2144.

(30) Aquilante, F.; Pedersen, T. B.; Sánchez de Merás, A.; Koch, H. Fast noniterative orbital localization for large molecules. *J. Chem. Phys.* **2006**, *125*, 174101.

(31) Sánchez de Merás, A. M. J.; Koch, H.; Cuesta, I. G.; Boman, L. Cholesky decomposition-based definition of atomic subsystems in electronic structure calculations. *J. Chem. Phys.* **2010**, *132*, 204105.

(32) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151−154.

(33) Saebo, S.; Pulay, P. Local treatment of electron correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.

(34) Hampel, C.; Werner, H.-J. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.* **1996**, *104*, 6286−6297.

(35) Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661−681.

(36) Korona, T.; Werner, H.-J. Local treatment of electron excitations in the EOM-CCSD method. *J. Chem. Phys.* **2003**, *118*, 3006−3019.

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

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

(39) Taube, A. G.; Bartlett, R. J. Frozen natural orbital coupled-cluster theory: Forces and application to decomposition of nitro-ethane. *J. Chem. Phys.* **2008**, *128*, 164101.

(40) Landau, A.; Khistyaev, K.; Dolgikh, S.; Krylov, A. I. Frozen natural orbitals for ionized states within equation-of-motion coupled-cluster formalism. *J. Chem. Phys.* **2010**, *132*, No. 014109.

(41) DePrince, A. E., III; Sherrill, C. D. Accurate noncovalent interaction energies using truncated basis sets based on frozen natural orbitals. *J. Chem. Theory Comput.* **2013**, *9*, 293−299.

(42) DePrince, A. E., III; Sherrill, C. D. Accuracy and efficiency of coupled-cluster theory using density fitting/cholesky decomposition, frozen natural orbitals, and at 1- transformed hamiltonian. *J. Chem. Theory Comput.* **2013**, *9*, 2687−2696.

(43) Kumar, A.; Crawford, T. D. Frozen virtual natural orbitals for coupled-cluster linearresponse theory. *J. Phys. Chem. A* **2017**, *121*, 708−716.

(44) 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.

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

(46) Folkestad, S. D.; Kjønstad, E. F.; Myhre, R. H.; Andersen, J. H.; Balbi, A.; Coriani, S.; Giovannini, T.; Goletto, L.; Haugland, T. S.; Hutcheson, A.; et al. e^T 1.0: An open source electronic structure program with emphasis on coupled cluster and multilevel methods. *J. Chem. Phys.* **2020**, *152*, 184103.

(47) Stöhr, J. *NEXAFS spectroscopy*; Springer Science & Business Media: 2013; *25*.

(48) Lange, K. M.; Aziz, E. F. Electronic structure of ions and molecules in solution: a view from modern soft X-ray spectroscopies. *Chem. Soc. Rev.* **2013**, *42*, 6840−6859.

(49) Smith, J. W.; Saykally, R. J. Soft x-ray absorption spectroscopy of liquids and solutions. *Chem. Rev.* **2017**, *117*, 13909−13934.

(50) Cederbaum, L. S.; Domcke, W.; Schirmer, J. Many-body theory of core holes. *Phys. Rev. A* **1980**, *22*, 206.

(51) Coriani, S.; Koch, H. Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework. *J. Chem. Phys.* **2015**, *143*, 181103.

(52) Coriani, S.; Koch, H. Erratum:"Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework". *J. Chem. Phys.* **2015**, *143*, 181103.

(53) Myhre, R. H.; Wolf, T. J.; Cheng, L.; Nandi, S.; Coriani, S.; Gühr, M.; Koch, H. A theoretical and experimental benchmark study of core-excited states in nitrogen. *J. Chem. Phys.* **2018**, *148*, No. 064106.

(54) Myhre, R. H.; Coriani, S.; Koch, H. X-ray and UV Spectra of Glycine within Coupled Cluster Linear Response Theory. *J. Phys. Chem. A* **2019**, *123*, 9701−9711.

(55) Matthews, D. A. EOM-CC methods with approximate triple excitations for NEXAFS and XPS. *arXiv preprint* arXiv:2001.09218 2020

(56) Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; De Simone, M.; Prince, K. C.; Trofimov, A. B.; Gromov, E. V.; Zaytseva, I. L.; Schirmer, J. A theoretical and experimental study of the near edge X-ray absorption fine structure (NEXAFS) and X-ray photoelectron spectra (XPS) of nucleobases: Thymine and adenine. *Chem. Phys.* **2008**, *347*, 360−375.

(57) Kelly, D. N.; Schwartz, C. P.; Uejio, J. S.; Duffin, A. M.; England, A. H.; Saykally, R. J. Communication: Near edge x-ray absorption fine structure spectroscopy of aqueous adenosine triphosphate at the carbon and nitrogen K-edges. *J. Chem. Phys.* **2010**, 101103.

(58) Hršak, D.; Nørby, M. S.; Coriani, S.; Kongsted, J. One-photon absorption properties from a hybrid Polarizable Density Embedding/Complex Polarization Propagator approach for polarizable solutions. *J. Chem. Theory Comput.* **2018**, *14*, 2145−2154.

(59) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular electronic-structure theory*; John Wiley & Sons: 2014.

(60) Koch, H.; Kobayashi, R.; Sanchez de Merás, A.; Jo/rgensen, P. Calculation of sizeintensive transition moments from the coupled cluster singles and doubles linear response function. *J. Chem. Phys.* **1994**, *100*, 4393−4400.

(61) Vidal, M. L.; Feng, X.; Epifanovsky, E.; Krylov, A. I.; Coriani, S. New and efficient equation-of-motion coupled-cluster framework for core-excited and core-ionized states. *J. Chem. Theory Comput.* **2019**, *15*, 3117−3133.

(62) Paul, A. C.; Myhre, R. H.; Koch, H. *A new and efficient implementation of CC3*; 2020.

(63) Boys, S. F. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **1960**, *32*, 296.

(64) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916−4926.

(65) Edmiston, C.; Ruedenberg, K. Localized atomic and molecular orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457.

(66) Høyvik, I.-M.; Jørgensen, P. Characterization and generation of local occupied and virtual Hartree−Fock orbitals. *Chem. Rev.* **2016**, *116*, 3306−3327.

(67) Löwdin, P.-O. *Advances in quantum chemistry*; Elsevier: 1970; 5; 185−199.

(68) Warshel, A.; Karplus, M. Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. *J. Am. Chem. Soc.* **1972**, *94*, 5612−5625.

(69) Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694.

(70) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(71) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3094.

(72) Mennucci, B. Polarizable continuum model. WIREs. *Comput. Mol. Sci.* **2012**, *2*, 386−404.

(73) Koch, H.; Christiansen, O.; Jo/rgensen, P.; Sanchez de Merás, A. M.; Helgaker, T. The CC3 model: An iterative coupled cluster approach including connected triples. *J. Chem. Phys.* **1997**, *106*, 1808−1818.

(74) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; et al. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477−1489.

(75) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Aust. J. Chem.* **2012**, *4*, 17.

(76) Folkestad, S. D.; Koch, H. *Geometries for Equation-of-motion MLCCSD and CCSDin-HF oscillator strengths and their application to core excitations*; 2020 DOI: 10.5281/zenodo.3931992.

(77) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

(78) Rohatgi, A. *WebPlotDigitizer*; 2019, https://automeris.io/WebPlotDigitizer.

(79) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(80) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(81) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.