A new and efficient equation-of-motion coupled-cluster framework for core-excited and core-ionized states

Marta López Vidal,[†] Xintian Feng,[‡,¶] Evgeny Epifanovsky,[¶] Anna I. Krylov,[§] and Sonia Coriani[*,†]

[†]*DTU Chemistry - Department of Chemistry, Technical University of Denmark, DK-2800, Kongens Lyngby, Denmark*

[‡]*Department of Chemistry, University of California, Berkeley, California 94720, United States*

[¶]*Q-Chem Inc., 6601 Owens Drive, Suite 105 Pleasanton, CA 94588*

[§]*Department of Chemistry, University of Southern California, Los Angeles, California 90089-0482*

E-mail: soco@kemi.dtu.dk

January 13, 2019

Abstract

We present a fully analytical implementation of the core-valence separation (CVS) scheme for the equation-of-motion (EOM) coupled-cluster singles and doubles (CCSD) method for calculations of core-level states. In the spirit of the original CVS approximation proposed by Cederbaum, Domcke and Schirmer, pure valence excitations are excluded from the EOM target space and the frozen-core approximation is imposed on

the reference-state amplitudes and multipliers. This yields an efficient, robust, and accurate EOM-CCSD framework for calculations of excitation and ionization energies as well as state and transition properties (e.g., spectral intensities, natural transition and Dyson orbitals). The accuracy of the new scheme is improved relative to the results obtained applying the CVS only during the solution of the EOM eigenvalue equations. The errors in absolute excitation/ionization energies relative to the experimental reference data are of the order of 0.2–3.0 eV, depending on the K-edge considered and on the basis set used, and the shifts are systematic for each edge.

1 Introduction

By providing tunable high-energy radiation, advanced light sources such as X-ray free electron laser (X-FEL) and synchrotron installations enable a variety of X-ray based spectroscopies.[? ? ?] Recent advances in beam quality greatly expanded possible applications of X-rays, giving rise to a proliferation of techniques, including those operating in time-resolved and non-linear regimes.[? ? ? ? ?] Fundamentally, these spectroscopies exploit electronic transitions involving core orbitals. Since core-level binding energies are characteristic of a species, X-ray absorption and X-ray photo-electron spectroscopies (XAS and XPS, respectively) are powerful techniques for probing electronic structure of atoms and molecules.[?] The localized nature of core orbitals makes X-ray-based spectroscopies sensitive to local environment. However, just as in the case of VUV-based techniques,[?] theoretical modeling is required to unambiguously assign spectral features and to relate experimental measurements to molecular structures. The experimental advances over the past decade have been driving the interest in developing highly accurate theoretical methods for X-ray spectroscopy and, in particular, for time-resolved XAS, which is proving to be a powerful mean to investigate molecular dynamics.[? ? ? ?]

Owing to its low computational costs relative to *ab initio* methods, time-dependent (TD) density functional theory (DFT) has been among the most commonly used techniques for

modeling absorption spectra, including X-ray absorption.[?] However, TDDFT often fails to deliver an accurate description of spectroscopic properties. TDDFT core excitations computed with conventional exchange–correlation functionals, for instance, are grossly underestimated with MAE of the order of 20 eV.[?] Time-independent DFT-based procedures like ΔKS, transition-potential DFT (TP-DFT)[?] and orthogonality constrained DFT (OC-DFT)[? ?] perform better for regular X-ray absorption, but their extension to the simulation of, for instance, transient-state core spectra is, to the best of our knowledge, so far unexplored.

Thus, there is an evident need for reliable wave-function-based methods for calculations of XAS spectra of both ground and excited states. Although more expensive, these methods can be systematically improved, thereby yielding to results of controlled accuracy.[?] The focus of this contribution is on extending coupled-cluster (CC) approaches,[? ? ? ? ? ? ? ? ?] which are among the most successful *ab initio* methods for molecular properties and electronic spectra in the UV-vis region, to absorption and ionization phenomena in the X-ray domain.[? ? ? ? ?]

Although ionization or excitations of core electrons superficially appear to be similar to valence transitions, the numerical experiments have shown that direct application of standard approaches to core-level transitions leads to unsatisfactory results,[?] due to the following essential features of core-ionized and core-excited states. First, these states have open-shell character.[?] Second, the core-level states lie very high in energy (hundreds of electron volts, depending on the edge). Third, orbital relaxation effects are much more important for core states than for valence states because the outer orbitals are more delocalized and better shielded from the nuclear charge than the tight and localized core orbitals. Fourth, these transitions appear to be much more sensitive to the effect of the environment.[? ?] Fifth, these high-lying states are metastable with respect to electron ejection,[? ?] i.e., they are Feshbach resonances that can autoionize via two-electron transitions in which one valence electron fills the core hole and a second valence electron is ejected. Thus, they are embedded in the ionization continuum and their description within Hermitian quantum mechanics is

problematic.

Multistate methods, such as EOM-CC,[? ? ? ? ? ? ?] can effectively describe multiconfigurational wavefunctions, tackle open-shell character and orbital relaxation effects in one computational scheme. The effect of the environment can be included by using a variety of implicit or explicit solvent models. However, dealing with high-energy interior states, especially with those embedded in autodetaching continuum, is more problematic. Although the Davidson procedure can be modified to solve for the eigenstates dominated by the desired transition (MOM-like[?]) or lying within the desired energy range,[? ?] the convergence and numeric stability are strongly affected by the presence of the continuum. An attempt to compute such states often produces pseudo-continuum states in which one electron occupies the most diffuse orbital.[?] Because in Hermitian quantum mechanics, the resonances are not represented by a single state, but rather by an increased density of states in the continuum,[?] the representation of resonances in a discretized continuum is inherently prone to numeric instabilities. Although the EOM-CC methods can describe resonances by using complex-scaled and CAP-augmented approaches,[?] such calculations are much more expensive than regular bound-state calculations. Here we focus on alternative, less expensive approximate methods for modeling spectroscopic properties of the core-level states.

Our strategy for tackling issues due to the continuum nature of the core levels is based on employing the core-valence separation (CVS) approximation proposed by **?** in 1980.[?] By decoupling the core excitations from the rest of the configurational space, the CVS allows one to extend standard methods for excited and ionized states to the core-level states.[? ? ?] Effectively, CVS results in a decoupling of the highly excited core states from the continuum of valence excitations; it also leads to a significant reduction of the computational costs. The core states can also be decoupled from the continuum by excluding double excitations from the configurational space and treating them perturbatively.[?] Although numerically stable, this approach suffers from insufficient description of orbital relaxation.[?] Because the CVS approximation admits selected double excitations that are crucial for describing orbital

relaxation, CVS-EOM/LR-CCSDs[?] yields negligible decline in accuracy, as compared to full EOM-CCSD/LR-CCSD results.[? ? ?]

An implementation of the CVS within EOM-CC and Linear Response (CCLR) theories has been reported by **?**.[?] In this work, the CVS was deployed as an *a posteriori* projection applied at each iteration of the solution of the EOM/CCLR equations to eliminate excitations that do not involve at least one core electron, whereas the ground state amplitudes and Lagrangian multipliers retained all possible types of excitations. Here, a different strategy is presented: the zero-order wave-function parameters are computed within the frozen-core (fc) approximation and the subsequent EOM/CCLR equations are solved analytically imposing the core-valence separation. The respective state and transition properties such as oscillator strengths, natural transition orbitals (NTOs), Dyson orbitals, exciton descriptors, and so on are obtained from the appropriate densities between different target-state manifolds. Specifically, transition properties can be computed between the ground state and a core-excited state, as well as between a valence-excited and a core-excited state. This yields an efficient scheme for simulating, e.g., near-edge absorption fine structure (NEXAFS) spectra of both ground and excited states. The latter is required to simulate UV-pump/X-ray-probe experiments and to obtain core-ionization potentials and Dyson orbitals for XPS spectra of medium size molecules. We report illustrative results for NEXAFS and core IEs of all K-edges in neon, water, ammonia, ethylene, vinylfluoride, ozone, adenine, as well as the transient (time-resolved) NEXAFS spectrum of uracil.

2 Theory

2.1 EOM-CCSD

In the EOM-CC approach, the target-state wave functions are parameterized using the following ansatz:[? ? ? ?]

$$\langle \Psi_L| = \langle \Phi_0| L e^{-T} \quad ; \qquad |\Psi_R\rangle = e^T R |\Phi_0\rangle \tag{1}$$

where $|\Phi_0\rangle$ is a reference Slater determinant and T, R and $L^\dagger$ are excitation operators. The excitation operator T is the cluster operator:[?]

$$\begin{aligned} T &= \sum_\mu t_\mu \tau_\mu = T_1 + T_2 + \cdots + T_N \\ T_1 &= \sum_{ia} t_i^a a_a^\dagger a_i, \quad T_2 = \frac{1}{4} \sum_{ijab} t_{ij}^{ab} a_a^\dagger a_i a_b^\dagger a_j, \quad \ldots \end{aligned} \tag{2}$$

with τ_μ being the excitation operator and t_μ the corresponding cluster amplitudes determined by the CC equations for the reference state:

$$\langle \Phi_\mu | \bar{H} - E_{\mathrm{CC}} | \Phi_0 \rangle = 0 \quad ; \qquad E_{\mathrm{CC}} = \langle \Phi_0 | \bar{H} | \Phi_0 \rangle \tag{3}$$

where $\langle \Phi_\mu|$'s represent μ-tuple excited determinants and $\bar{H}$ is the similarity transformed Hamiltonian:

$$\bar{H} = e^{-T} H e^T. \tag{4}$$

The operator R is a general excitation operator:

$$R = R_0 + R_1 + \cdots + R_n \equiv \sum_\mu r_\mu \tau_\mu \tag{5}$$

and L is a de-excitation operator:

$$L = L_0 + L_1 + \cdots + L_n \equiv \sum_\mu l_\mu \tau_\mu^\dagger \tag{6}$$

The choice of the reference state and the exact form of R and L depend on the EOM method to be used.[??] Furthermore, in practical calculations the excitation and de-excitation operators must be truncated to some tractable level of excitation. In this work, we focus on the EOM-CCSD family of methods in which the cluster operator T is truncated to single (S) and double (D) excitations, and so are the excitation R and de-excitation L operators.

One of the most obvious applications of EOM theory is the calculation of electronically excited states.[??] In this approach, referred to as EOM-EE (EOM for excitation energies), the optimal reference state is usually the closed-shell ground-state Hartree-Fock determinant and the R and L operators conserve the number of electrons and their spin, taking the following forms:

$$R^{\mathrm{EE}} = r_0 + \sum_{ia} r_i^a a_a^\dagger a_i + \frac{1}{4} \sum_{ijab} r_{ij}^{ab} a_a^\dagger a_i a_b^\dagger a_j + \ldots \tag{7}$$

$$L^{\mathrm{EE}} = \sum_{ia} l_i^a a_i^\dagger a_a + \frac{1}{4} \sum_{ijab} l_{ij}^{ab} a_i^\dagger a_a a_j^\dagger a_b + \ldots \tag{8}$$

Another common use of EOM is calculations of ionization energies by the EOM-IP (EOM for ionization potentials) method.[???] In this case, the excitation operator changes the number of electrons in the system, assuming the following from:

$$R^{\mathrm{IP}} = \sum_i r_i a_i + \frac{1}{4} \sum_{ija} r_{ij}^a a_a^\dagger a_j a_i + \ldots \tag{9}$$

The EOM amplitudes r_μ and l_μ are found as stationary points of the EOM functional:

$$E = \frac{\langle \Psi_L | H | \Psi_R \rangle}{\langle \Psi_L | \Psi_R \rangle} \tag{10}$$

By applying the bi-variational principle,[??] one arrives to the non-symmetric eigenvalue

problem:

$$\langle \Phi_\mu | \bar{H} - E | R\Phi_0 \rangle = 0 \;\; ; \qquad\qquad \langle \Phi_0 L | \bar{H} - E | \Phi_\mu \rangle = 0, \tag{11}$$

where the eigenvectors of the Hamiltonian are chosen to form a biorthogonal set:[?]

$$\langle \Phi_0 L_i | R_j \Phi_0 \rangle = \delta_{ij} \tag{12}$$

Thus, the implementation of EOM-CCSD method boils down to the diagonalization of the effective Hamiltonian $\bar{H}$ in the basis of the reference and of singly and doubly excited determinants, which can be written in matrix form as:

$$\bar{H} = \begin{pmatrix} E_{CC} & \bar{H}_{OS} & \bar{H}_{OD} \\ 0 & \bar{H}_{SS} & \bar{H}_{SD} \\ 0 & \bar{H}_{DS} & \bar{H}_{DD} \end{pmatrix} \tag{13}$$

giving rise to the EOM-CCSD right and left eigenvalue equations:

$$\begin{pmatrix} \bar{H}_{SS} - E_{CC} & \bar{H}_{SD} \\ \bar{H}_{DS} & \bar{H}_{DD} - E_{CC} \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} = \omega \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \tag{14}$$

$$\begin{pmatrix} L_1 & L_2 \end{pmatrix} \begin{pmatrix} \bar{H}_{SS} - E_{CC} & \bar{H}_{SD} \\ \bar{H}_{DS} & \bar{H}_{DD} - E_{CC} \end{pmatrix} = \omega \begin{pmatrix} L_1 & L_2 \end{pmatrix} \tag{15}$$

where ω is the energy difference with respect to the reference state.

In practice, Eqs. (14) and (15) are solved iteratively using a generalized Davidson's iterative diagonalization procedure[? ? ?] which requires the computation of the right σ and left $\tilde{\sigma}$ trial vectors:

$$\sigma_1 = ([\bar{H}_{SS} - E_{CC}]R_1)_1 + (\bar{H}_{SD}R_2)_1 \tag{16}$$

$$\sigma_2 = (\bar{H}_{DS}R_1)_2 + ([\bar{H}_{DD} - E_{CC}]R_2)_2 \tag{17}$$

$$\tilde{\sigma}_1 = (L_1[\bar{H}_{SS} - E_{CC}])_1 + (L_2\bar{H}_{DS})_1 \tag{18}$$

$$\tilde{\sigma}_2 = (L_1\bar{H}_{SD})_2 + ([L_2\bar{H}_{DD} - E_{CC}])_2 \tag{19}$$

In EOM-CC, molecular properties are commonly calculated as expectation values, using unrelaxed one-particle density matrices like

$$\gamma_{pq} = \langle \Psi_L | a_p^\dagger a_q | \Psi_R \rangle \ . \tag{20}$$

In the present study, we focus on transition properties involving core states, specifically, the oscillator strengths f between the ground state and the core-excited states, and between valence-excited and core-excited states. These can be formulated via contractions of property integral matrices and transition density matrices between state i and j, for instance,

$$f(i \to j) = \frac{2}{3}(\omega_j - \omega_i) \sum_{\alpha=x,y,z} \mu_\alpha^{i\to j} \mu_\alpha^{j\to i} \tag{21}$$

where

$$\mu_\alpha^{i\to j} = \sum_{pq} \mu_{pq}^\alpha \gamma_{pq}^{i\to j} \ ; \qquad \mu_\alpha^{j\to i} = \sum_{rs} \mu_{rs}^\alpha \gamma_{rs}^{j\to i} \tag{22}$$

Above, μ^α refers to a specific Cartesian component of the electric dipole operator and the state i can be either the ground state and an excited state.

The transition density matrices $\gamma_{pq}^{i\to j}$ and $\gamma_{pq}^{j\to i}$ are generalizations of Eq. (20), and their specific form in the fc-CVS-EOM-CCSD framework is discussed in the next sections. We here used the so-called unrelaxed EOM transition densities, which differ from the density matrices derived within the coupled-cluster linear response (CCLR) formalism in that they do not contain terms originating from the amplitudes' response to the external perturbation. In the framework of EOM-CC, the amplitude-response terms can also be included by using, for example, a Lagrangian approach, giving rise to the so-called relaxed state and transition

density matrices, which are identical to those in the CCLR formalism.[? ? ? ? ? ?] Furthermore, orbital relaxation to the external perturbation is also omitted, as typically done within coupled-cluster response theory.[?]

Transition density matrices describe the changes in electron density upon excitation and can be interpreted as exciton wavefunction[? ? ? ?] when expressed in coordinate space:

$$\chi(r_e, r_h) = \sum_{pq} \gamma_{pq} \phi_p(r_e) \phi_q(r_h), \tag{23}$$

where r_e and r_h denote particle (electron) and hole coordinates, respectively (using $r_e = r_p = r$, χ is reduced to the transition density). Properties of the exciton can be quantified by various expectation values, i.e., exciton size, hole-particle separation and correlation, and so on. These exciton descriptors help to assign state characters.[? ? ?] Applying a singular value decomposition (SVD) to the transition matrices yields the so-called NTOs,[? ? ? ? ? ?] which can be used to describe the electronic excitations in terms of hole-particle excitons:

$$\chi(r_e, r_h) = \sum_{K} \sigma_K \tilde{\phi}^e_K(r_e) \tilde{\phi}^h_K(r_h), \tag{24}$$

where $\tilde{\phi}^e_K$ and $\tilde{\phi}^h_K$ are particle and hole orbitals obtained by SVD of γ and σ_K are the corresponding singular values. Usually, only a few σ_K are non-zero, giving rise to simple interpretation of excited-state characters in terms of one-electron excitations. Note that the sum of σ^2_K is equal to the squared norm of γ, which provides a simple metric quantifying the one-electron character of the transition, i.e., for pure single excitations, $||\gamma||$=1.

2.2 The fc-CVS-EOM-CCSD method

Different flavors of separation between core and valence excitations have been realized for a variety of approximate wavefunction methods as well as for TDDFT, and we refer the interested reader to Ref. **?** for a recent overview.

Within the CVS-CC framework of Ref. [?], an effective core-valence separation was introduced (as a projection) in the solution of the CC eigenvalue equations since, physically, the core-excited states are energetically far away from the valence excited states.

Already in 1980, [??] suggested that, since core orbitals are strongly localized in space and energetically are well separated from the valence orbitals, one can decouple core and valence orbitals by zeroing out the respective blocks of the model Hamiltonian, such that the solutions of the Schrödinger equation can be separated into the core and valence domains. In 1985, [??] implemented this idea within the second-order algebraic diagrammatic construction method ADC(2)[??] by observing that, in view of negligible coupling between core and valence orbitals, all the following two-electron Coulomb integrals are negligible

$$\begin{aligned} \langle Ip|qr\rangle = \langle pI|qr\rangle = \langle pq|Ir\rangle = \langle pq|rI\rangle \approx 0 \\ \langle IJ|pq\rangle = \langle pq|IJ\rangle \approx 0 \\ \langle IJ|Kp\rangle = \langle IJ|pK\rangle = \langle Ip|JK\rangle = \langle pI|JK\rangle \approx 0 \quad , \end{aligned} \tag{25}$$

where small letters indicate general valence orbitals, and capital letters core orbitals. As a consequence, one can omit the blocks of the ADC propagator matrix $\mathbf{M}$ involving these integrals and therefore decouple pure valence excitations from excitations involving core electrons. This idea was later extended to all other members of the ADC family of propagator methods.[????]

Since MP2 is an underlying wave-function for the ADC propagator,[?] omitting the integrals in Eq. (25) immediately corresponds to using the frozen-core approximation for the MP2 ground state energy.

Similarly, it is straightforward to show that, if the above criterion of negligible Coulomb integrals is used within the CCSD ansatz in the canonical basis,[?] the resulting CCSD energy and the amplitude equations for the "valence-only" amplitudes t_i^a and t_{ij}^{ab} are identical to those obtained within the frozen-core approximation—that is, assuming that all ground-state

amplitudes and multipliers where at least one occupied index refers to a core orbital are zero.

Thus, in the here-proposed fc-CVS-EOM-CCSD method, we use the frozen core approximation during the determination of the ground-state amplitudes t_μ and Lagrangian multipliers λ_μ while constraining the excitation/ionization operators R and L to involve at least one core orbital. This introduces restrictions on the left and right EOM equations, and on the respective density matrices.

The programmable expressions for obtaining the right and left excitation vectors within the fc-CVS-EOM-CCSD methods have been derived from Ref. **?** and for the density matrices from Ref. **?** and can be found in the Supplementary Information.

As an illustrative example, we show below how the expression of the right linearly transformed vector elements σ_i^a is modified for the fc-CVS case. The general expression for σ_i^a is[?]

$$\sigma_i^a = \sum_b F_{ab} r_i^b - \sum_j F_{ij} r_j^a - \sum_{jb} I^1_{ibja} r_j^b + \sum_{jb} F_{jb} r_{ij}^{ab} - \frac{1}{2}\sum_{jkb} I^6_{jkib} r_{jk}^{ab} - \frac{1}{2}\sum_{jbc} I^7_{jabc} r_{ij}^{bc} \quad (26)$$

with

$$\begin{aligned}
F_{ia} &= f_{ia} + \sum_{jb} t_j^b \langle ij||ab\rangle \\
F_{ij} &= f_{ij} + \sum_a t_i^a f_{ja} + \sum_{ka} t_k^a \langle jk||ia\rangle + \sum_{kab} t_i^a t_k^b \langle jk||ab\rangle + \frac{1}{2}\sum_{kbc} t_{jk}^{ac} \langle jk||bc\rangle \\
F_{ab} &= f_{ab} - \sum_i t_i^a f_{ja} - \sum_{ic} t_i^c \langle ia||bc\rangle + \sum_{ijc} t_i^c t_j^a \langle ij||bc\rangle - \frac{1}{2}\sum_{jkc} t_{jk}^{ac} \langle jk||bc\rangle \\
I^1_{iajb} &= \langle ia||jb\rangle - \sum_k t_k^b \langle jk||ia\rangle - \sum_c t_i^c \langle jb||ac\rangle + \sum_{kc} t_i^c t_k^b \langle jk||ac\rangle - \sum_{kc} t_{ik}^{bc} \langle jk||ac\rangle \\
I^6_{ijka} &= \langle ij||ka\rangle - \sum_c t_k^c \langle ij||ac\rangle \\
I^7_{iabc} &= \langle ia||bc\rangle - \sum_j t_j^a \langle ij||bc\rangle
\end{aligned} \quad (27)$$

It is convenient to further split the occupied orbitals $(i, j, k, l, \ldots)$ into the two sub-blocks:

occupied valence labeled with an additional v subindex, $(i_v, j_v, k_v, l_v, \ldots)$, and occupied core orbitals denoted by a capital letter $(I, J, K, L, \ldots)$. The CVS approximation then entails: (i) reducing the set of occupied orbitals to only the core ones in the excitation process and, (ii) freezing the core orbitals in optimization of the ground state CC wavefunction parameters (amplitudes and multipliers). Hence, by restricting the EOM excitations to core excitations only, all terms involving only valence excitations disappear in the equations above. The frozen-core approximation further simplifies the equations since the terms in Eqs. (27) involving ground-state amplitudes vanish for the core orbitals, or, in other words, only ground state amplitudes with valence occupied orbitals are retained in the fc-CVS case.

The fc-CVS expression for the linearly transformed trial vector in Eq. (26) thus reads as follows:

$$\begin{aligned}\sigma_I^a =& \sum_b F_{ab} r_I^b - \sum_J F_{IJ} r_J^a - \sum_{Jb} I^1_{IbJa} r_J^b + \sum_{Jb} F_{Jb} r_{IJ}^{ab} + \sum_{j_v b} F_{j_v b} r_{Ij_v}^{ab} \\ & - \sum_{Jk_v b} \langle Jk_v || Ib \rangle r_{Jk_v}^{ab} - \frac{1}{2} \sum_{JKb} \langle JK || Ib \rangle r_{JK}^{ab} - \frac{1}{2} \sum_{Jbc} I^7_{Jabc} r_{IJ}^{bc} - \frac{1}{2} \sum_{j_v bc} I^7_{j_v abc} r_{Ij_v}^{bc}\end{aligned} \tag{28}$$

with

$$\begin{aligned}F_{Ia} &= f_{Ia} + \sum_{j_v b} t_{j_v}^b \langle Ij || ab \rangle \\ F_{IJ} &= f_{IJ} + \sum_{k_v a} t_{k_v}^a \langle Jk_v || Ia \rangle \\ F_{i_v a} &= f_{i_v a} + \sum_{j_v b} t_{j_v}^b \langle i_v j_v || ab \rangle \\ I^1_{IaJb} &= \langle Ia || Jb \rangle - \sum_{k_v} t_{k_v}^b \langle Jk_v || Ia \rangle \\ I^7_{Iabc} &= \langle Ia || bc \rangle - \sum_{j_v} t_{j_v}^a \langle Ij_v || bc \rangle\end{aligned} \tag{29}$$

We followed the same strategy to derive the expressions for other linearly transformed vector blocks and for the (transition) density matrices; all programmable expressions are given

in the Supplementary Information. We emphasize that we only used the integral screening as per Eq. (25) to justify the use of the frozen-core approximation in the determination of the ground-state CC wavefunction parameters and that we retained the terms containing such integrals in the equations determining the target states energies and transition properties. Although they are expected to be small, these terms increase the flexibility of our ansatz to address, via correlation, the electron relaxation effects that follow core excitation.

We conclude this session by discussing size-extensivity of fc-CVS-EOM-CCSD. As explained, for instance, in Refs. **?** and **?** , size-extensivity of the total energy and size-intensivity of the excitation energies is guaranteed if the singles and doubles block of the first column of the matrix representation of the effective Hamiltonian $\bar{H}$ are zero, as per Eq. (13). The blocks in question contain the amplitude-constraint terms

$$\Omega_\mu = \langle \Phi_\mu | \exp(-T) H \exp(T) | \Phi_0 \rangle \ , \tag{30}$$

where, for CCSD, $\langle \Phi_\mu |$ is either a single or a double-excitation determinant with respect to the reference Slater determinant $|\Phi_0\rangle$, see Eq. (3).

Obviously, in fc-CVS-EOM-CCSD

$$\Omega^a_{i_v} = \langle \Phi^a_{i_v} | \exp(-T) H \exp(T) | \Phi_0 \rangle = 0 \tag{31}$$

$$\Omega^{ab}_{i_v j_v} = \langle \Phi^{ab}_{i_v j_v} | \exp(-T) H \exp(T) | \Phi_0 \rangle = 0 \tag{32}$$

since these terms are identical to the valence-only amplitude equations obtained by application of the negligible-integrals/frozen-core approximation. As for the remaining terms, in which the occupied indices are either a core I, a core J or both, let us consider, for instance,

elements Ω_I^a. In the frozen-core approximation, they become:

$$\begin{aligned}\Omega_I^a &= \langle \Phi_I^a | \exp(-T) H \exp(T) | \Phi_0 \rangle \\ &= f_{aI} - \sum_{k_v} f_{k_v I} t_{k_v}^a + \sum_{k_v c} \langle k_v a || cI \rangle t_{k_v}^c - \frac{1}{2} \sum_{k_v l_v c} \langle k_v l_v || cI \rangle t_{k_v l_v}^{ca} - \sum_{k_v l_v c} \langle k_v l_v || cI \rangle t_{k_v}^c t_{l_v}^a \end{aligned} \tag{33}$$

In the canonical basis, the respective Fock-matrix terms are zero. Then, invoking the negligible integrals conditions as per Eq. (25), these blocks become exactly zero. Similar arguments apply for the blocks Ω_{Ij}^{ab} and Ω_{IJ}^{ab}.

Alternatively, one can consider the Ω_I^a, Ω_{Ij}^{ab} and Ω_{IJ}^{ab} blocks when only the negligible integral conditions are applied. One can set these blocks equal to zero and solve for the amplitudes t_I^a, t_{IJ}^{ab} and $t_{Ij_v}^{ab}$. It is then straightforward to prove that the frozen-core conditions $t_I^a = 0$, $t_{IJ}^{ab} = 0$ and $t_{Ij_v}^{ab} = 0$ are trivial solutions to the equations $\Omega_I^a = 0$, $\Omega_{Ij}^{ab} = 0$ and $\Omega_{IJ}^{ab} = 0$.

Thus, in the fc-CVS-EOMCCSD method, all elements (except the first one) of the first column of $\bar{H}$ are zero by construction, hereby yielding a rigorous size-intensive approach for the core excitation energies. Numerical evidence of size-intensivity is shown in Table 1 below, where we report core-excitation energies computed for a cluster of Ne and H_2O separated by R=100 bohr, as well as for each system alone. Exactly the same core-excitation energies are obtained in both cases.

Table 1: Size-intensivity of the core-excitation energies (eV) of the fc-CVS-EOMCCSD method. Basis set 6-311++G** (plus supplementary Rydberg functions in the case of Neon). The two species are separated by 100 bohr in the cluster Ne + H_2O.

Ne K-edge		O K-edge	
Ne + H_2O	Ne	H_2O + Ne	H_2O
867.94	867.94	535.22	535.22
869.55	869.55	537.01	537.01
870.13	870.13	539.73	539.73
		539.93	539.93
		540.08	540.08
		540.55	540.55

3 Computational details

We implemented the fc-CVS-EOM-CC method in the Q-Chem electronic structure package[? ?] using the libtensor library.[?] The geometries of H_2O, NH_3, CO, C_2H_4, C_2H_3F and O_3 were optimized at the fc-CCSD(T)/cc-pVQZ level using CFOUR.[?] For 9H-adenine, we considered both planar and non-planar structures taken from the literature. The non-planar structure was optimized at the fc-RI-MP2/cc-pVTZ level,[?] whereas the planar structure was optimized at the B3LYP/cc-pVTZ level.[?] In the TR-NEXAFS simulations of uracil, we used different structures: an optimized MP2/cc-pVTZ ground-state structure, a ground-state structure and an S_1 minimum both optimized at the SF-BH&HLYP/6-31+G(d,p) level from Ref. **?** , and two stationary-point (i.e., zero-gradient) structures for the S_1 and S_2 excited states obtained at the EOM-CCSD/aug-cc-pVDZ level of theory. All structures were planar or almost planar.

In the (TR-)NEXAFS calculations, we considered several different basis sets, and results are here reported for Pople's 6-311++G** (pure d functions) and Dunning's aug-cc-pVTZ and aug-cc-pCVTZ sets. In selected cases, the basis sets were further augmented with uncontracted Rydberg-type functions whose exponents were computed according to the prescription of **?**[?] Using such system-specific Rydberg functions affords a more compact description of Rydberg states than an alternative brute-force strategy of adding additional diffuse sets with even-tempered exponents.[?]

To analyze the EOM states (i.e., the extent of Rydberg character), we considered, for a few selected cases, the spatial extent of the respective wavefunctions as well as the size of the particle NTO of a transition.[?] Both approaches deliver similar information, however, when using the former, one needs to consider the difference between the expectation value of the second moment of charges $\langle r^2 \rangle$ of the target EOM state and the CCSD reference[?] because the size of the electronic wave function depends on the system size.

Experimental data were taken from: Ref. **?** for adenine, Ref. **?** for Neon, Ref. **?** for H_2O and NH_3, Ref. **?** and **?** for CO, Ref. **?** for C_2H_4 and C_2H_3F and Ref. **?** for O_3. The

experimental NEXAFS spectrum of uracil is from Ref. **?** . All experimental spectra were digitized from the original references using WebPlotDigitizer.[?] The spectra were generated using a Python script and NTOs were visualized using MOLDEN.[?]

4 Results and discussion

4.1 Near-Edge Absorption Spectroscopy and core ionization energies

To test the performance of our method, we first considered the neon atom. Table S2 reports the computed excitation energies and IEs with three different basis sets, all supplemented with Rydberg-type functions (with n=2.5–4.5); the corresponding spectra are shown in Fig. 1. As the atomic NEXAFS spectra are due to transitions from the 1s orbital to Rydberg states,[?] the inclusion of Rydberg-type functions (or, alternative, a large set of diffuse functions) is mandatory in order to reproduce the Rydberg progression of peaks in the experimental spectrum.[?] The extent of Rydberg character in Neon is quantified in Table S3.

The NEXAFS spectra have been shifted, along with the IEs, to align with the first peak of the experimental NEXAFS spectrum (estimated to be at 867.10 eV). After the shift, the computed peaks match the experimental ones almost perfectly. Of the three sets, Dunning's aug-cc-pCVTZ (+Rydberg) yields the smallest absolute shift from experiment (+0.19 eV), followed by −0.43 eV of the aug-cc-pVTZ (+Rydberg) set, versus −0.84 eV of Pople's 6-311+G**(+Rydberg). The spectral profiles in the two Dunning sets overlap completely after the alignment.

With respect to previously reported CCSD results[?] obtained in the aug-cc-pCVTZ(+Rydberg) basis using the Lanczos algorithm with all electrons correlated (i.e., no CVS), the absolute shift from experiment in the fc-CVS-EOM-CCSD method is lower and has an opposite sign (+0.19 eV versus −1.07 eV). The shift is also smaller than that obtained in the CVS approach of Ref. **?** .

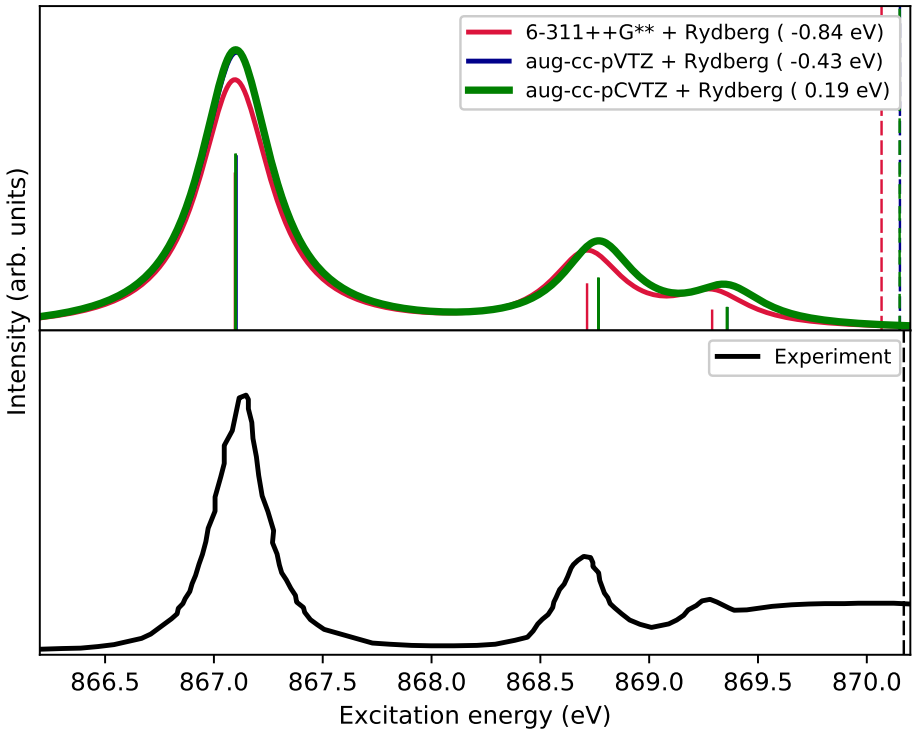


Figure 1: Neon. fc-CVS-EOMEE-CCSD X-ray absorption spectra obtained by convolution of the computed excitation energies and oscillator strengths with a Lorentzian function (FWHM = 0.4 eV). The experimental spectrum was digitized from Ref. **?** . The vertical dashed lines correspond to the core ionization energies. The experimental IE is 870.17 eV. The energy shifts required to align the NEXAFS profiles in each basis set with the experimental one are indicated in parenthesis. The computed IEs have been shifted by the same amount as used to align the NEXAFS profiles.

The NEXAFS and IE values of H_2O are reported in Table S4, with the corresponding spectra shown in Fig. 2. The upper panel of Fig. 2 shows the spectra for the chosen basis sets without Rydberg-type functions, whereas the middle panel shows those obtained including the Rydberg-type functions. Besides an overall shift (taking the value 534.0 eV as reference for the experimental first peak maximum, which varies slightly for the three bases, the separation between the two first peaks is practically the same, whereas huge differences are observed for the other bands known to have Rydberg character. Both relative intensity and position of the third band and the following ones are strongly overestimated in the bases

without Rydberg functions. Table S3 shows the sizes of particle NTOs and the $\langle r^2 \rangle$ values (i.e., the second moments of charge density), clearly revealing Rydberg character of certain transitions.

In H_2O, the energy shifts required to realign with the experimental spectrum are also smaller than those used in the LR-CVS calculation from Ref. **?** and in the full-space Lanczos calculation from Ref. **?** . Remarkably, in the aug-cc-pCVTZ basis the shift is smaller that in the aug-cc-pVTZ basis, whereas the reverse trend has been observed using the LR-CVS approach of Ref. **?** . Thus, the current approach shows a systematic improvement of the results with respect to the basis set increase.

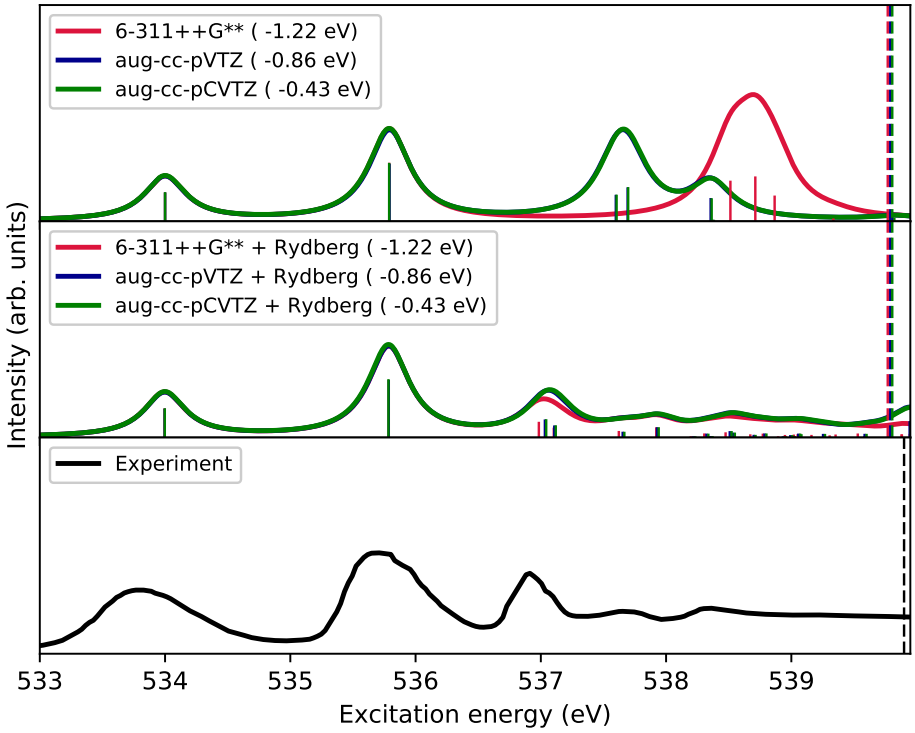


Figure 2: Water. fc-CVS-EOMEE-CCSD O-edge X-ray absorption spectra obtained by convolution of the spectral data in Table S4 with a Lorentzian function (FWHM = 0.4 eV). The experimental spectrum (0.12 eV resolution) was digitized from Ref. **?** . Dashed vertical lines correspond to the IEs. The energy shifts required to align the NEXAFS profiles in each basis set with the experimental one are indicated in parenthesis. The computed IEs have been shifted by the same amount as used to align the NEXAFS profiles.

Another system whose gas-phase NEXAFS is dominated by Rydberg states is NH_3. Fig. 3 shows the computed spectra; the raw data are in Table S5. The spectra were aligned with respect to the peak maximum of the first experimental band, estimated at 400.53 eV. As for the previous systems, the Dunning basis shows a smaller shift compared to the experimental peaks (−0.68 eV vs −1.04 eV). Neither Pople's 6-311++G** nor Dunning's aug-cc-pVTZ can correctly reproduce the third and higher bands without the inclusion of additional diffuse functions. Rydberg character of these bands is clearly revealed by the data in Table S3.

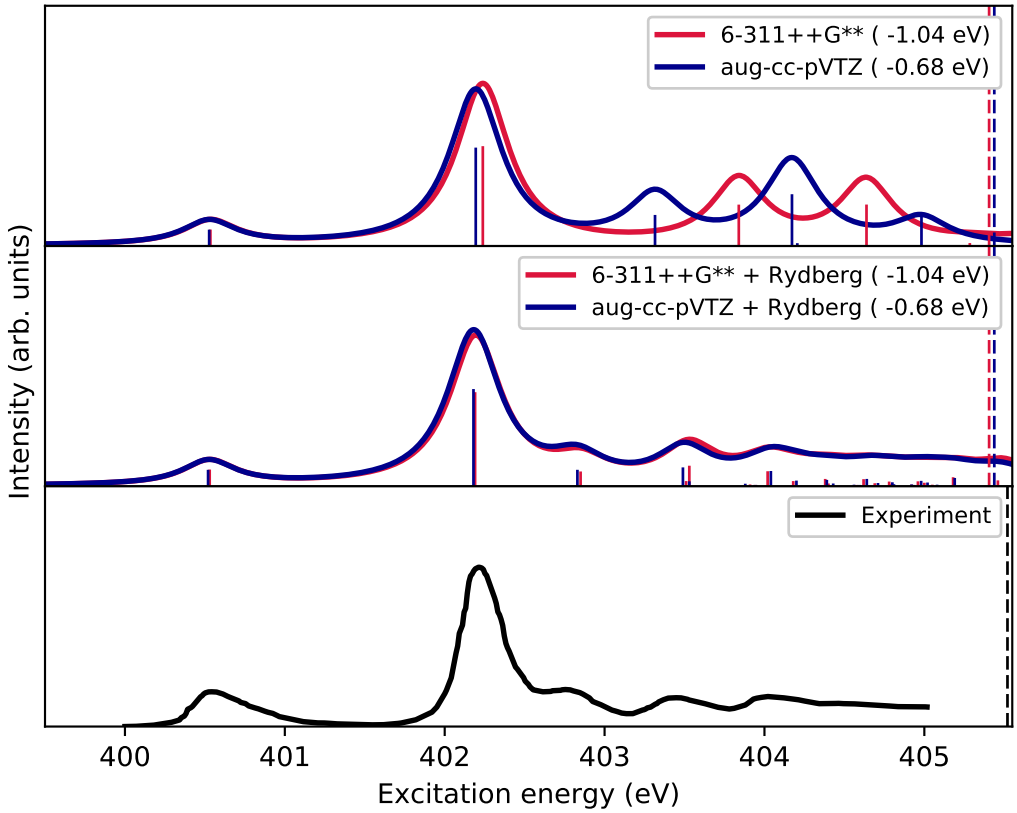
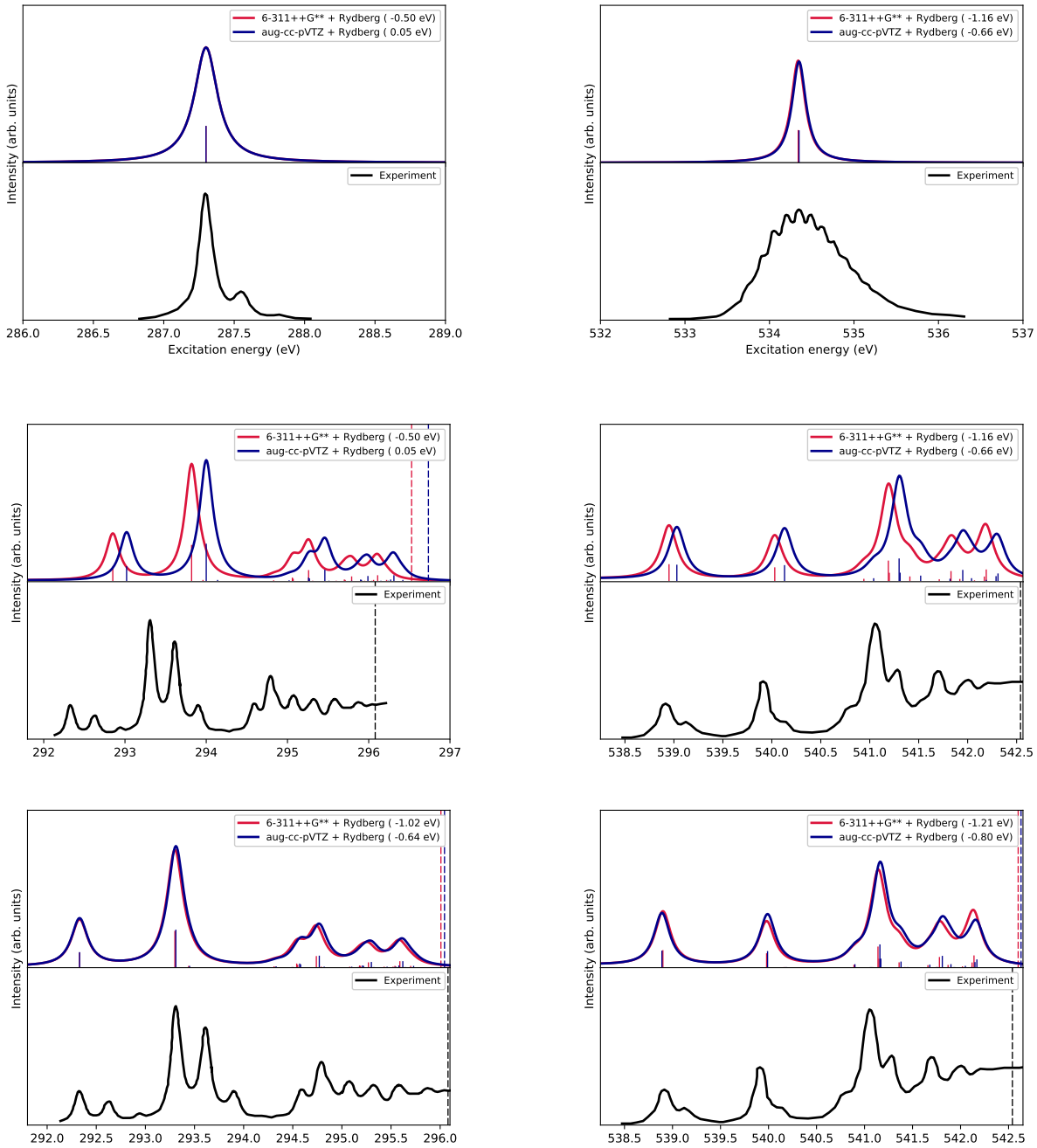


Figure 3: Ammonia. fc-CVS-EOMEE-CCSD N-edge X-ray absorption spectra obtained from convolution of the spectral data in Table S5 with a Lorentzian function (FWHM = 0.4 eV). The experimental spectrum (0.2 eV resolution) was digitized from Ref. **?** . Dashed vertical lines indicate the IEs. The energy shifts required to align the NEXAFS profiles in each basis set with the experimental one are indicated in parenthesis. The computed IEs have been shifted by the same amount as used to align the NEXAFS profiles.

Table S7 presents the spectral data for C and O edges of carbon monoxide and the corresponding spectra are shown in Fig. 4. The two upper panels in the figure show the main NEXAFS bands, experimentally observed in between 286.5 and 289.0 eV for carbon and in between 533 and 537 eV for oxygen. The middle and bottom panels of Fig. 4 show the (much weaker) peaks observed at higher frequencies below the ionization limit.

The position of the dominant C-edge $1s \rightarrow \pi^*$ band is overestimated by 0.50 eV in the 6-311++G** + Rydberg basis set, and underestimated by 0.05 eV in the aug-cc-pVTZ + Rydberg basis. The O-edge $1s \rightarrow \pi^*$ band is overestimated by about 1.16 eV in Pople's set, and by 0.66 eV in Dunning's basis. The additional features of the main experimental bands are due to the vibronic progression, which is not included in our calculations.

Upon alignment of the computed spectra with the main peak of the experimental ones, the Rydberg transitions are still slightly misaligned, see mid panels of Fig. 4. Nonetheless, all weaker $3s\sigma$, $3p\pi$, $3p\sigma$, $3d\pi$, $4s\sigma$, and $4s\pi$ transitions can be identified in the computed spectra of each edge, although once again without their finer vibronic progressions. The assignments can be verified by realignment of the first peak of the first progression, as shown in the bottom panels of Fig. 4.



C K-edge O K-edge

Figure 4: Carbon monoxide. fc-CVS-EOMEE-CCSD C K-edge (left) and O K-edge (right) X-ray absorption spectra, obtained by convolution of the computed excitation energies and oscillator strengths with a Lorentzian function (FWHM = 0.2 eV). The upper panels show the main $1s \rightarrow \pi^*$ band, the mid and bottom panels the band progressions of the weaker $1s \rightarrow 3s\sigma$, $3p\pi$, $3d\pi$, $4s\sigma$, and $4s\pi$ transitions. The experimental spectra (0.03 eV resolution at the C K-edge and 0.07 eV resolution at the O K-edge) were digitized from Ref. **?** . Vertical dashed lines correspond to the IEs.

Fig. 5 reports the computed spectra of ethylene obtained by convolution of the spectral data in Table S8. In this case, the Rydberg functions also improve the description of the higher energy region approaching the ionization limit (third experimental band [?]). The second band in the experimental spectrum corresponds to three excitations in the computed spectra. The overall shift is 0.44 eV in the aug-cc-pVTZ(+Rydberg) set and 0.90 eV for Pople's 6-311++G**(+Rydberg) set. Upon realignment with respect to the $1s \rightarrow \pi^*$ absorption energy, the IE obtained with Pople's set is slightly underestimated compared to the experimental IEs.

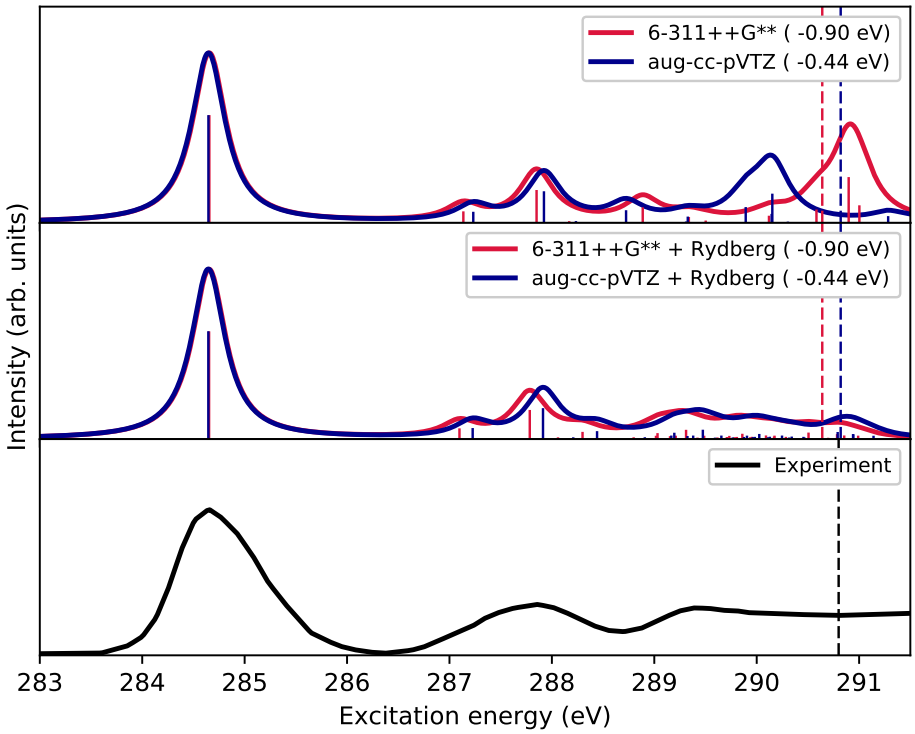


Figure 5: Ethylene. fc-CVS-EOMEE-CCSD X-ray absorption spectra by Lorentzian broadening (FWHM = 0.4 eV) of the computed excitation energies and oscillator strengths. The experimental spectrum (0.6 eV resolution) was digitized from Ref. [?] . The vertical dashed lines correspond to the IEs. The energy shifts required to align the NEXAFS profiles in each basis set with the experimental one are indicated in parenthesis. The computed IEs have been shifted by the same amount as used to align the NEXAFS profiles.

Fig. 6 shows the computed X-ray spectra at the C edge in vinylfluoride (CH_2CHF); the

raw data are given in Table S9. The computed spectra were shifted to align them to the first experimental peak,[?] whose position we estimated to be at 285 eV. The applied shift is −0.44 eV for Dunning's set, and −0.91 eV for Pople's set. Inclusion of Rydberg-type functions in the basis set has a more modest effect than in the case of ethylene. NTOs of the most intense core excitations obtained with the 6-311++G** basis set are shown in Table 2, allowing us to identify from which of the two C atoms they originate from and the character of the transition.

The X-ray absorption spectra obtained at the fluorine edge of CH_2CHF are shown in Fig. 7; the raw data are given in Table S10. In the experimental spectrum, digitized from Ref. **?** , only two peaks are clearly discernible, with absolute energies assigned at (689.2±2.0) eV and (690.6±2.0) eV ($1s_F \rightarrow \sigma^*$ (C-F)). In the experimental study, the first peak is assigned to a $1s_F \rightarrow \pi^*$(C=C) transition, and the second one to a $1s_F \rightarrow \sigma^*$ (C-F) transition. Inspection of the results in Table S10 and of the NTOs in Table 4 indicates that the first band results from two almost degenerate transitions, $1s_F \rightarrow \sigma^*$ (C–F) and $1s_F \rightarrow \pi^*$(C=C). The third excitation (second experimental band) also appears to be of $1s_F \rightarrow \sigma^*$ (C-F) character.

The experimental IE is at 693.26 eV.[?] The computed spectra in the Pople set (with and without Rydberg functions) are realigned by −1.98 eV, and those for the Dunning basis by −1.58 eV.

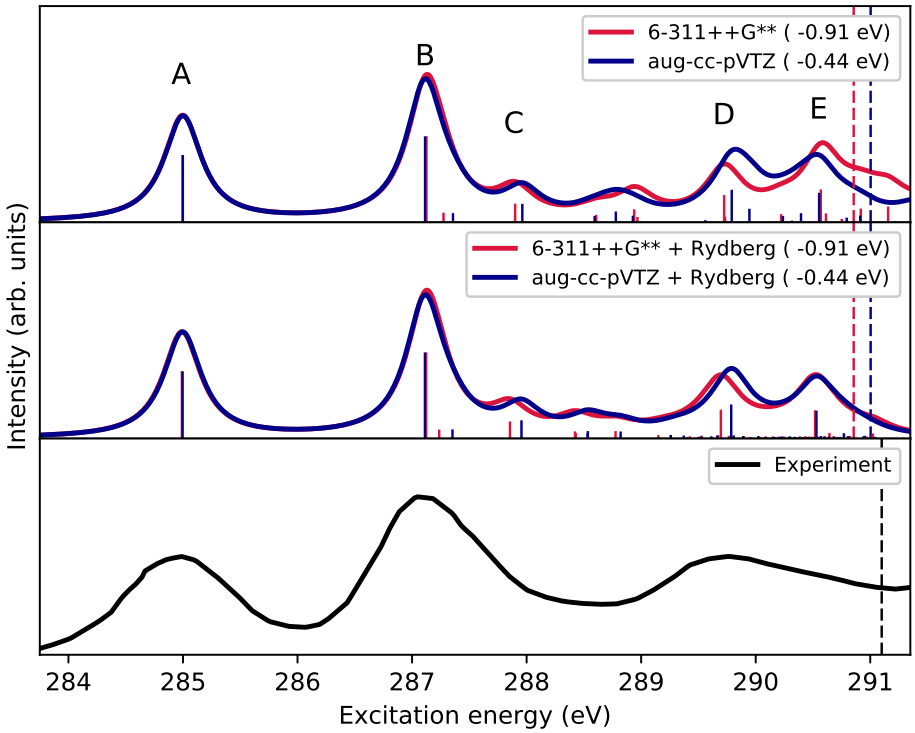


Figure 6: Vinyl fluoride. fc-CVS-EOMEE-CCSD X-ray absorption spectra at the C-edge obtained by convolution of the computed energies and oscillator strengths with a Lorentzian function (FWHM = 0.4 eV). The experimental spectrum (0.6 eV resolution) was digitized from Ref. **?** . The dashed vertical lines correspond to the IEs of the 1s electron on the carbon atom of the CH_2 group. The IEs of the 1s electron of C_{CHF} atom are outside the displayed frequency range (experimental IE 293.48 eV). The energy shifts required to align the NEXAFS profiles in each basis set with the experimental one are indicated in parenthesis. The computed IEs have been shifted by the same amount as used to align the NEXAFS profiles.

Table 2: Vinylfluoride. fc-CVS-EOM-CCSD/6-311++G** NTOs of 5 selected core-excited states at the C edge (NTO isosurface is 0.05).

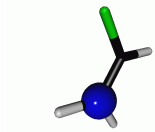
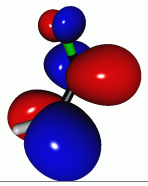
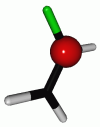
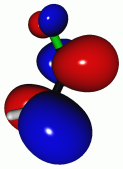
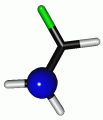
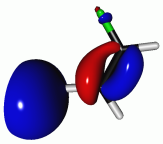
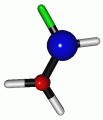
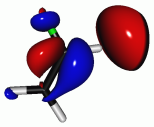
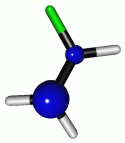
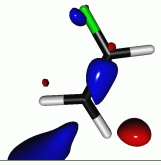
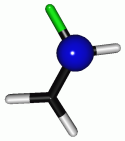
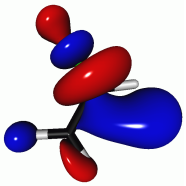
State	Hole	σ_K^2	Particle
A		0.78	
B		0.79	
C		0.82	
D		0.62	
		0.19	
E		0.82	

Table 3: Vinylfluoride. Changes in the second moments of charge density (in Å^2) and electron and hole size components (in Å) for selected core-excited states at the C K-edge.

State	$\Delta\langle x^2\rangle$	$\Delta\langle y^2\rangle$	$\Delta\langle z^2\rangle$	$\Delta\langle r^2\rangle$	x_e	y_e	z_e	r_e	x_h	y_h	z_h	r_h
A	-15.32	-4.51	-1.70	-21.23	1.92	0.68	0.81	1.40	0.10	0.10	0.10	0.17
B	-14.60	-4.63	-1.68	-20.60	0.92	0.67	0.81	1.39	0.10	0.10	0.10	0.17
C	-13.39	1.34	-0.95	12.69	1.42	2.39	1.22	3.03	0.10	0.10	0.10	0.17
D	-12.80	-0.60	-0.56	-13.66	1.81	1.78	1.36	2.88	0.51	0.29	0.10	0.59
E	-13.72	-2.03	-1.33	-16.78	1.50	1.58	1.03	2.40	0.10	0.10	0.10	0.17

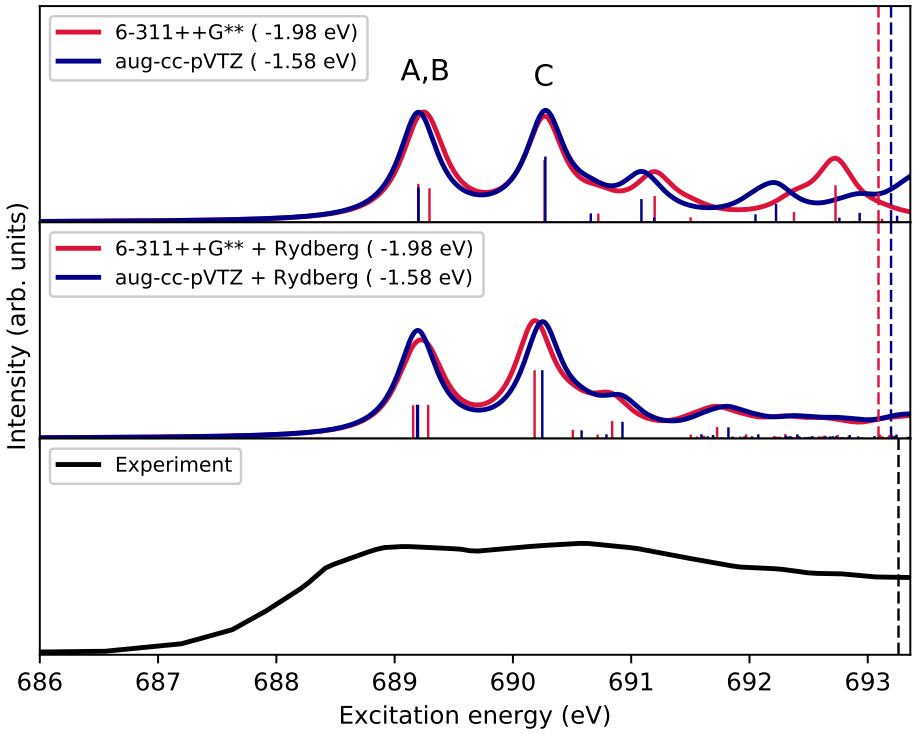


Figure 7: Vinylfluoride. fc-CVS-EOMEE-CCSD X-ray absorption spectra at the fluorine K edge, obtained by convolution of the computed energies and oscillator strengths with a Lorentzian function (FWHM = 0.4 eV). The experimental spectrum was digitized from Ref. **?** . The dashed vertical lines correspond to the IEs. The energy shifts required to align the NEXAFS profiles in each basis set with the experimental one are indicated in parenthesis. The computed IEs have been shifted by the same amount as used to align the NEXAFS profiles. The shift was computed based on the experimentally derived maximum at 689.2 eV.

Table 4: Vinylfluoride. fc-CVS-EOM-CCSD/6-311++G** NTOs of 3 selected core-excited states at the F edge. NTO isosurface is 0.05.

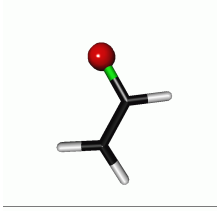
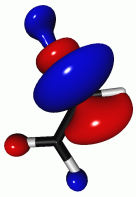
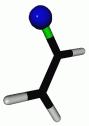
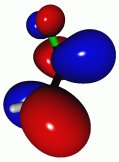
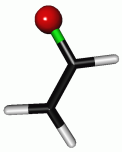
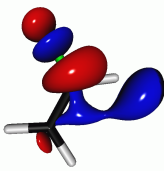
State	Hole	σ_K^2	Particle
A		0.84	
B		0.79	
C		0.83	

Fig. 8 shows the fc-CVS-EOM-CCSD NEXAFS spectra of O_3, based on the spectral data in Table S11. This molecule displays the largest overall shift relative to the experimental spectrum,[?] -2.35 eV in the Pople set, and -1.96 eV with Dunning's set. Apart from this, our calculations confirm the assignment in Ref. **?** : the first spectral feature is due to the terminal oxygens' $1s \rightarrow \pi^\star$, whereas the second (broad) band is due to both the central oxygens $1s \rightarrow \pi^\star$ and the terminal oxygens' $1s \rightarrow \sigma^\star$ excitations, see also the NTOs in Table 5. The shoulder at 530.7 eV in the experimental spectrum is known to be due to the O$1s \rightarrow \sigma^\star$ transition of a small amount of O_2 present in the sample.[?]

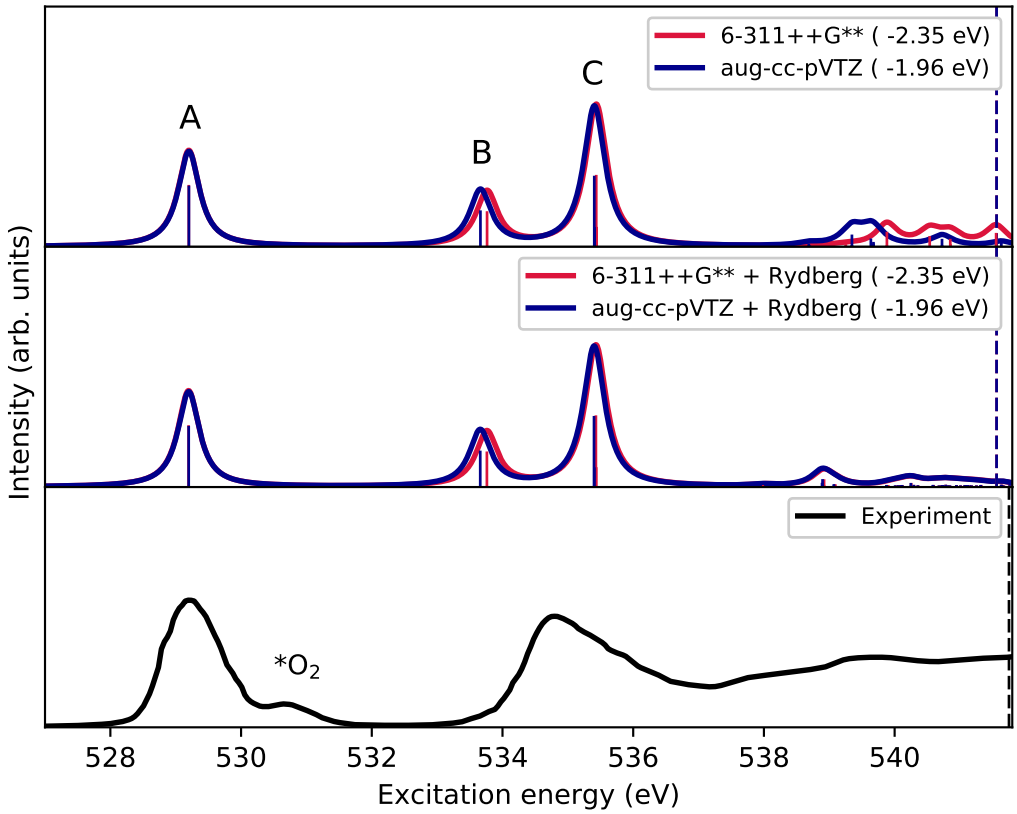
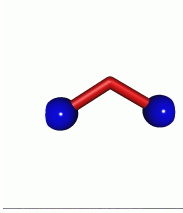
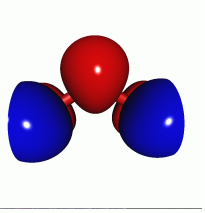
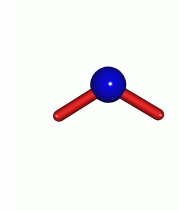
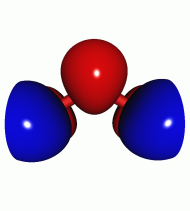
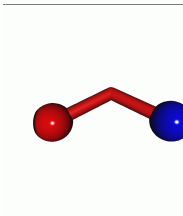
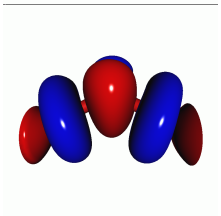
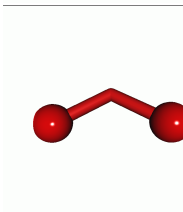
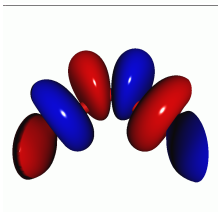


Figure 8: Ozone. fc-CVS-EOMEE-CCSD X-ray absorption spectra obtained from convolution with a Lorentzian function (FWHM = 0.4 eV) of the computed excitation energies and oscillator strengths. The experimental spectrum High-Resolution Inner-Shell (0.05 eV resolution) was digitized from Ref. **?** . The dashed vertical lines correspond to the ionization energy of the terminal O atom. The central oxygen's IE was omitted in figures as it lies above 545 eV.

Table 5: Ozone. fc-CVS-EOM-CCSD/6-311++G** NTOs of the first 3 core-excited states. NTO isosurface is 0.05

State	Hole	σ^2_K	Particle
A		0.70	
B		0.72	
C		0.63	
		0.19	

The final system considered here is adenine, whose NEXAFS and XPS spectra were experimentally recorded in gas-phase by **?**[?] We considered both carbon and nitrogen edges. Due to the relatively large size of this system, we only carried out calculations in the 6-311++G** basis set. The upper and lower panels of Figure 9 show the C and N K-edge spectra, respectively. The raw data are given in Table S12. The C K-edge spectra were shifted by −1.10(non-planar)/−1.14(planar) eV, and the N K-edge one by −1.43(non-planar)/−1.45(planar) eV, and one can expect an even smaller shift had the larger aug-cc-pVTZ basis set been used. The experimental features are, once again, quite well reproduced.

The agreement for the higher-energy peaks could probably be further improved by inclusion of Rydberg functions. Remarkably, the C-edge spectrum obtained from the planar geometry is more similar to the experimental spectrum, primarily due to the larger splitting between the 4th and 5th excitations in the non-planar structure. The spectral assignment for both structures is, nonetheless, identical. This is best appreciated looking at the NTOs for the first 5 excitations shown in Table 6. We also note that in this case, as in other examples, NTOs reveal that the electronic transitions have rather simple character and can be described by a single NTO pair. In contrast, the EOM wavefunctions often show multiple amplitudes with comparable weights, giving a misleading impression of the character of the transition.

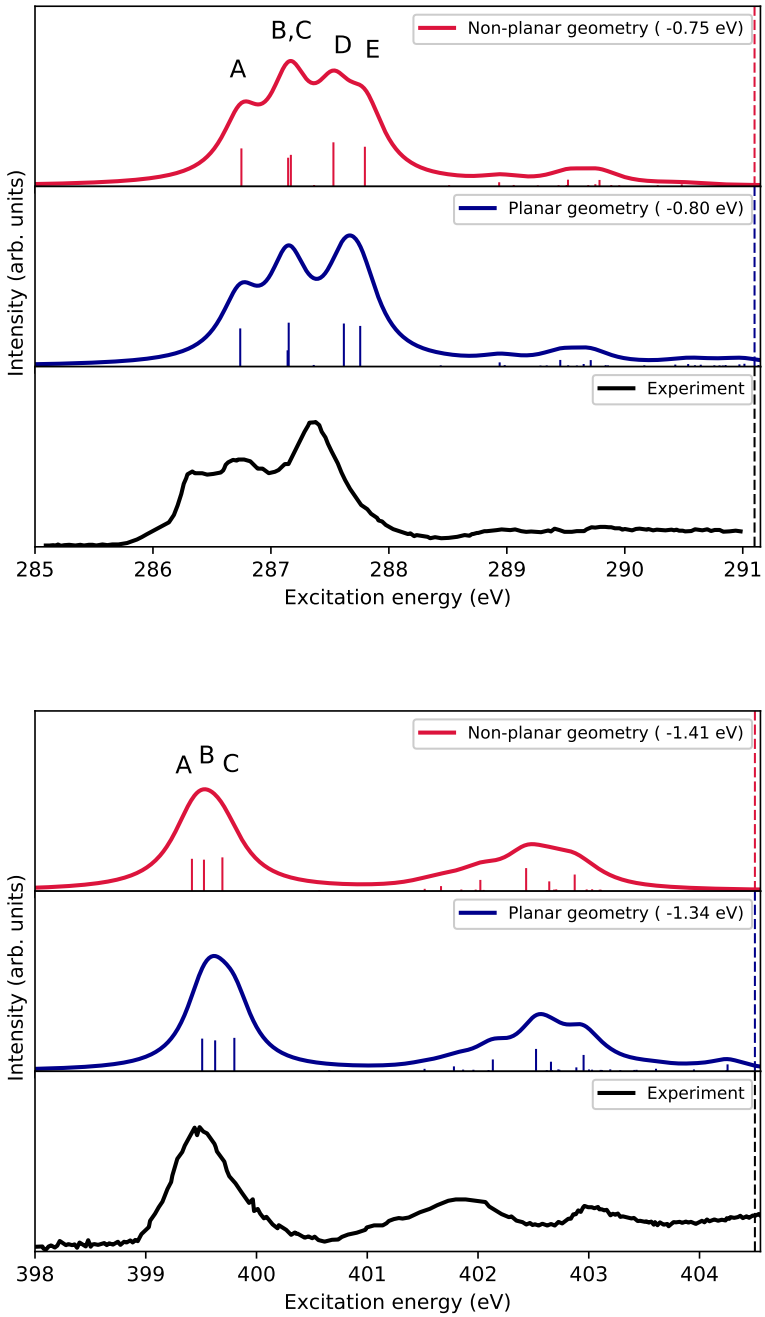


Figure 9: Adenine. C-edge (upper panels) and N-edge (lower panels) fc-CVS-EOMEE-CCSD/6-311++G** X-ray absorption spectra for two different molecular structures, obtained by convolution of the computed energies and oscillator strengths with a Lorentzian function (FWHM = 0.4 eV). The rigid shifts applied are indicated in parenthesis in the legends. They were determined with respect to the first experimental peak position in each spectrum, estimated to be at 286.4 eV for C and 399.4 eV for N. The vertical dashed line correspond to the first IE. The computed IEs have been shifted by the same amount as used to align the NEXAFS profiles. The experimental spectra (0.57 eV resolution at the C K-edge and 0.59 eV resolution at the N K-edge) were digitized from Ref. **?** .

Table 6: Adenine. fc-CVS-EOM-CCSD/6-311++G** NTOs of the first 5 core-excited states at the C K-edge at the non-planar RI-MP2/cc-pVTZ geometry (left) and planar B3LYP/cc-pVTZ geometry (right). NTO isosurface is 0.05.

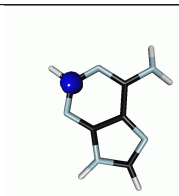
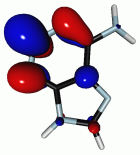
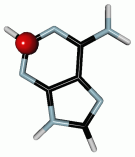
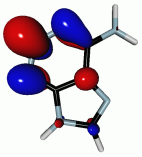
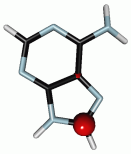
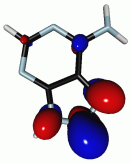
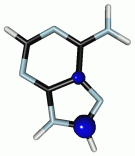
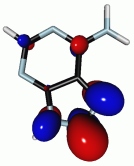
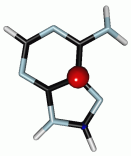
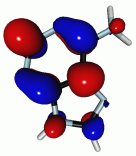
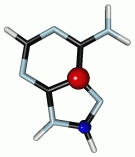
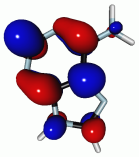
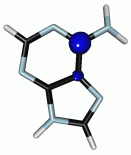
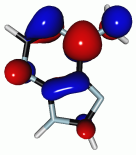
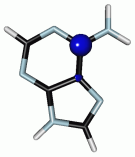
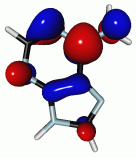
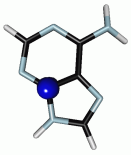
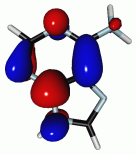
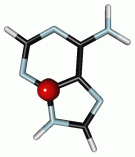
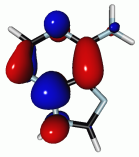
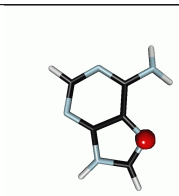
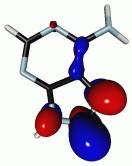
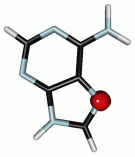
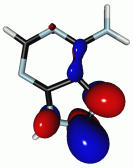
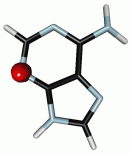
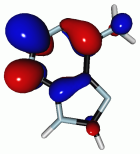
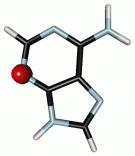
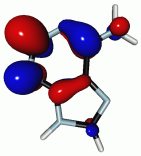
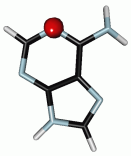
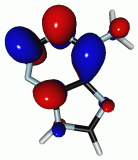
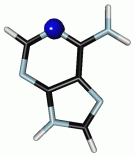
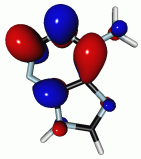
State	Non-planar Hole	σ_K^2	Particle	Planar Hole	σ_K^2	Particle
A		0.81			0.81	
B		0.79			0.71	
C		0.77			0.71	
D		0.75			0.82	
E		0.81			0.81	

Table 7: Adenine. fc-CVS-EOM-CCSD/6-311++G** NTOs of the first 3 core-excited states at the N K-edge at the non-planar RI-MP2/cc-pVTZ geometry (left) and planar B3LYP/cc-pVTZ geometry. NTO isosurface is 0.05.

	Non-planar			Planar		
State	Hole	σ_K^2	Particle	Hole	σ_K^2	Particle
A		0.78			0.78	
B		0.78			0.78	
C		0.79			0.79	

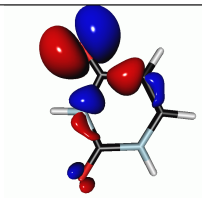
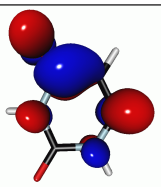
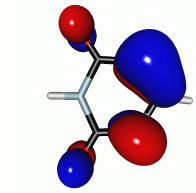
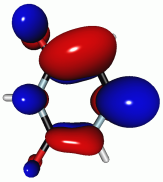
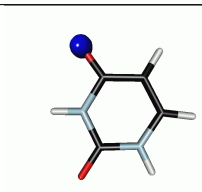
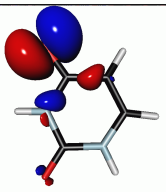
4.2 Core-level transient absorption spectroscopy

The advances in X-ray Free-Electron Lasers in the last decade have boosted the interest in computational methodologies to simulate of Time-Resolved X-ray Absorption (TR-XAS or TR-NEXAFS).[? ? ? ?] Typically, in TR-NEXAFS pump-probe experiments, the sample is first brought to a valence excited state by a UV pulse, and then probed, at different time delays, with X-ray radiation. To simulate these processes, methods to compute the intensity of valence-to-core transitions are needed. An EOM-CCSD/CC3 methodology, based on the CVS approach of Ref. **?** , has been devised and used, for instance, to simulate and interpret TR-NEXAFS experiments in thymine.[?] The study aimed at assessing the ability of K-edge resonant absorption spectroscopy to probe ultrafast $\pi\pi^\star$/n$\pi^\star$ internal conversion in organic

chromophores. Other methodologies have also been devised within the ADC framework by ?,[? ? ?] and at the TDDFT level by, e.g., [?]

We have extended the fc-CVS-EOM-CCSD formalism to the computation of the transition density matrices between two excited states, from which the transient X-ray absorption spectra can then be obtained. As illustrative example, we have considered the valence-to-core spectra of uracil at the O, C and N edges. TR-NEXAFS spectra of uracil have not been experimentally measured yet, but they are expected to bear strong similarities with those of thymine, whose O-edge TR-NEXAFS was measured in Ref. [?] . Two valence excited states were considered, the first bright $\pi\pi^*$ state (S_2 at FC geometry) and the first dark $n_O\pi^*$ (S_1 at FC geometry) state. The NTOs of these two states, obtained at the Franck-Condon geometry, are shown in Table 8.

Table 8: Uracil. EOM-CCSD/6-311++G** NTOs of the first 2 valence excited states and fc-CVS-EOM-CCSD/6-311++G** NTO of the core excitation from the S_1 valence excited state. NTO isosurface is 0.05.

Excitation	Hole	σ^2_K	Particle
$n_O\pi^*$		0.81	
$\pi\pi^*$		0.75	
$1s_On_O$		0.45	

Given the localized nature of the $n_O\pi^*$ (S_1) state on one of the two oxygen nuclei, and

similar to what has been observed for thymine,[?] one can expect that the TR-NEXAFS measurements at the O-edge are the best to probe the population of the $n_O\pi^*$ due to ultrafast internal conversion. Indeed, we show in Figs. 10 and 11 the X-ray absorption spectra obtained at the O edge for both the ground and the two excited states at different optimized geometries for the ground and the two valence excited states. In all cases, core excitation from the $n_O\pi^*$ state results in the emergence of a relatively strong and distinctive signal at around 526.0-526.5 eV, similar to what has been observed for thymine.[?] The NTO of this excitation, labeled $1s_On_O$, is also shown in Table 8, clearly illustrating that the core electron fills the vacancy in the $n_O\pi^*$ excited state.

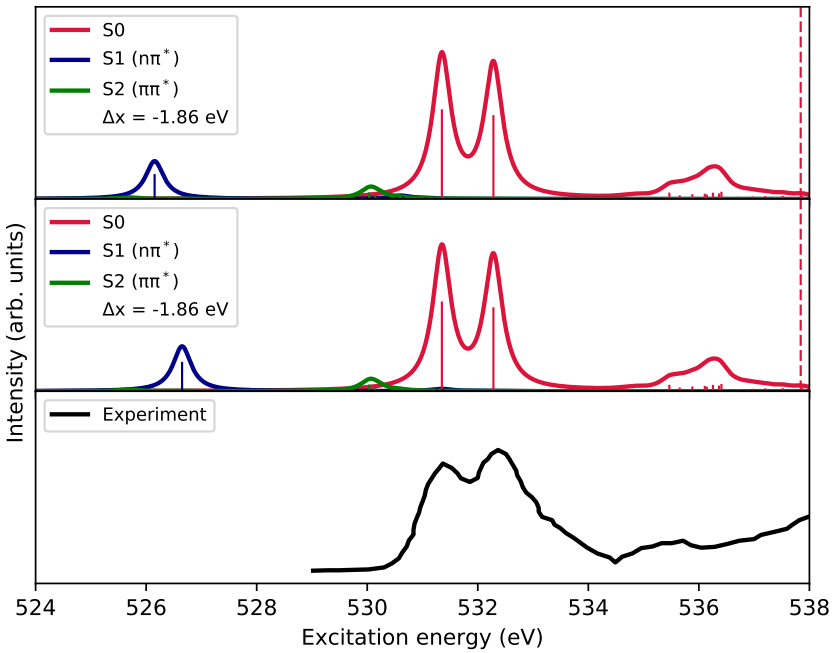


Figure 10: O-edge of uracil. Upper panel: fc-CVS-EOMEE-CCSD/6-311++G** ground and excited-state core absorption spectra at the DFT Franck-Condon geometry of Ref. **?** . Lower panel: fc-CVS-EOMEE-CCSD/6-311++G** ground and excited-state core-absorption spectra, at the Franck-Condon geometry for both the ground state (S_0) and the $\pi\pi^*$ (S_2) states, and at the TD-DFT optimized S_1 geometry of Ref. **?** . for S_1. In both cases, a Lorentzian convolution function (FWHM = 0.4 eV) was used.

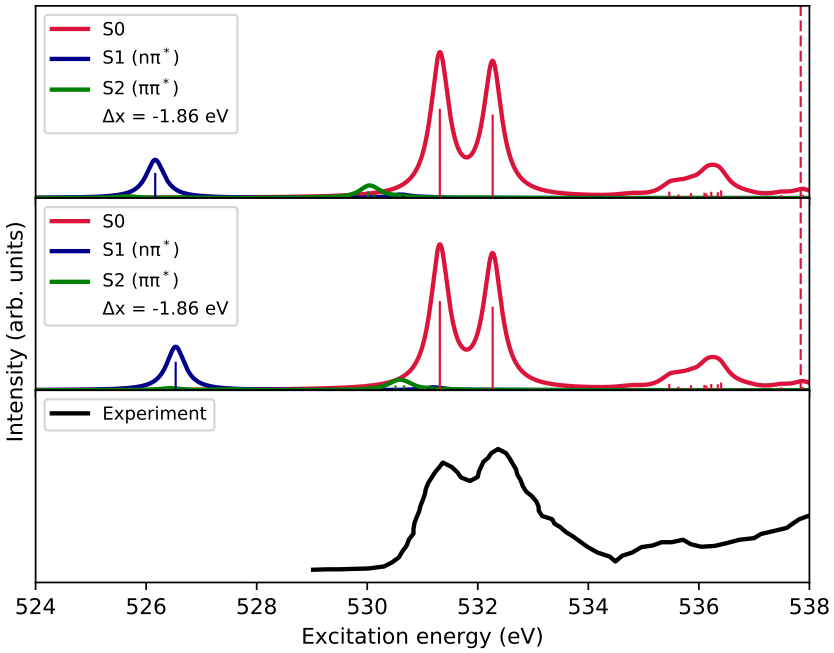


Figure 11: O-edge of uracil. Upper panel: fc-CVS-EOMEE-CCSD/6-311++G** ground and excited-state core-absorption spectra at the optimized MP2/cc-pVTZ Franck-Condon geometry. Lower panel: fc-CVS-EOMEE-CCSD/6-311++G** ground and excited-state core-absorption spectra at planar optimized geometries for each state, i.e., MP2/cc-pVTZ for the ground state, and EOM-CCSD/aug-cc-pVDZ for the two valence excited states. In both cases, a Lorentzian convolution function (FWHM = 0.4 eV) was used.

To conclude this section, we have also considered the transient state spectra that one could expect to observe if probing at the C and N edges after the initial pump, along with the computed ground state NEXAFS spectra and their experimental counterparts. Fig. 12 shows that at the C-edge the valence-to-core spectra are rather weak (signals have been enhanced by a factor 10 in the figures), and that, opposite to the O-edge case, the most intense features at this edge originate from the $\pi\pi^*$ excited state. At the N-edge (see Fig. 13) the intensities of the transient absorption spectra are higher than at the C edge (signals have been enhanced by a factor 5 in the figures) and, as in the C edge case, the dominant features are from the $\pi\pi^*$ excited state.

Figure 12: C-edge of uracil. fc-CVS-EOMEE-CCSD/6-311++G** ground and excited-state core-absorption spectra at different geometries. On the left, DFT geometries of Ref. **?** ; at the Franck-Condon DFT geometry for all states on the upper panel, and the Franck-Condon DFT geometry for S_0 and S_2 and the TD-DFT optimized S_1 geometry for S_1 on the middle panel. On the right, the planar MP2/cc-pVTZ Franck-Condon geometry for all states on the upper panel, and the MP2 optimized geometry for S_0 and planar optimized EOM-CCSD/aug-cc-pVDZ for both S_1 and S_2 on the middle panel.

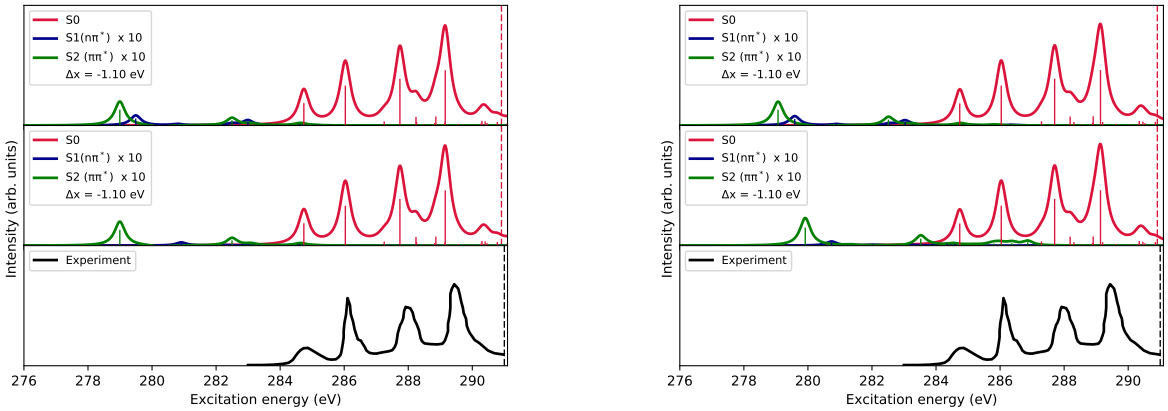
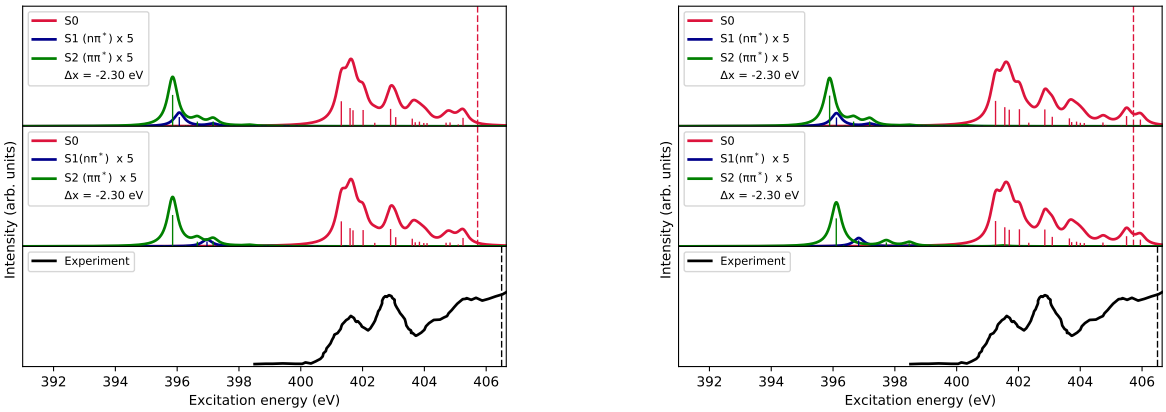


Figure 13: N-edge of uracil. fc-CVS-EOMEE-CCSD/6-311++G** ground and excited-state core-absorption spectra at different geometries: the Franck-Condon DFT geometry of Ref. **?** for all states on left upper panel; the Franck-Condon DFT geometry of Ref. **?** for S_0 and S_2 and the TD-DFT optimized S_1 geometry for S_1 on left lower panel; the planar MP2/cc-pVTZ Franck-Condon geometry for all states on right upper panel; the MP2 optimized geometry for S_0 and planar optimized EOM-CCSD/aug-cc-pVDZ geometry for both S_1 and S_2 on the bottom right panel.



5 Conclusions

We have presented a new, fully-analytic core-valence separated equation-of-motion approach, named fc-CVS-EOM-CCSD, for calculating spectral descriptors of X-ray absorption spectroscopies, specifically near-edge absorption fine structure, core-ionization energies and transient-state (time-resolved) X-ray absorption. The approach exploits the large energy separation of the core and valence orbitals both in the determination of the amplitudes of both coupled-cluster ground-state wavefunction and of the EOM target-state wavefunctions. By decoupling core-level states from the high-lying valence states that are embedded in the ionization continuum, CVS addresses the most challenging issue in modeling core states, that is, complications due to their auotoionizing nature. In contrast to full EOM-CCSD, CVS-EOM-CCSD features a robust numeric performance and reduced computational costs relative to the parent EOM-CCSD. Yet, it retains sufficient correlation, necessary to describe orbital relaxation effects. The approach also retains important properties of EOM-CCSD, such as size-extensivity and systematic convergence to the exact results upon inclusion of higher excitations.

We tested the methodology on a number of atomic and molecular systems. The shape of computed NEXAFS spectra agrees very well with the experimental one in terms of the relative heights of the individual peaks and the distance between them. However, the computed spectra are shifted with respect to the experiment. The magnitude of shifts required for the alignment varies between 0.2 and 3 eV, depending on the edge and basis set considered. The shifts are in all cases smaller than those obtained with a previously presented CVS-CCSD approach based on the energy separation between core and valence excited states,[?] whereas the spectral profiles are practically the same. Importantly, for all examples, we observed a systematic decrease of the shift upon the basis set increase. The fully analytical implementation also delivers reduced computational cost relative to the projection technique used in Ref. **?** while being strictly size-extensive. Simulations of the transient state NEXAFS spectra of uracil at all three edges supports the ability to probe the ultrafast internal conversion

of this DNA basis by TR-NEXAFS, as recently experimentally and computationally verified for the thymine molecule.[?]

Acknowledgement

M.L.V. and S.C. acknowledge support from DTU Chemistry (start-up Ph.D. grant). S.C. acknowledges support from the Independent Research Fund Denmark – DFF-Forskningsprojekt2 grant no. 7014-00258B and from the H2020-MSCA-ITN-2017 training network "COSINE – COmputational Spectroscopy In Natural sciences and Engineering". A.I.K. acknowledges support by the U.S. National Science Foundation (No. CHE-1566428) and by the Simons Foundation.

Note: The authors declare the following competing financial interest(s): A.I.K. is a part owner and a board member of Q-Chem, Inc.

Supporting Information Available: Tables with raw spectral data and implementation formulas. This information is available free of charge via the Internet at http://pubs.acs.org.

Graphical TOC Entry

Some journals require a graphical entry for the Table of Contents. This should be laid out "print ready" so that the sizing of the text is correct. Inside the `tocentry` environment, the font used is Helvetica 8 pt, as required by *Journal of the American Chemical Society*.
The surrounding frame is 9 cm by 3.5 cm, which is the maximum permitted for *Journal of the American Chemical Society* graphical table of content entries. The box will not resize if the content is too big: instead it will overflow the edge of the box.
This box and the associated title will always be printed on a separate page at the end of the document.