

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/141368/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Halbert, Loïc, Vidal, Marta L., Shee, Avijit, Coriani, Sonia and Severo Pereira Gomes, André 2021. Relativistic EOM-CCSD for Core-Excited and Core-Ionized State Energies Based on the Four-Component Dirac–Coulomb(–Gaunt) Hamiltonian. *Journal of Chemical Theory and Computation* 17 (6) , pp. 3583-3598. 10.1021/acs.jctc.0c01203

Publishers page: <http://dx.doi.org/10.1021/acs.jctc.0c01203>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Relativistic EOM-CCSD for core-excited and core-ionized state energies based on the 4-component Dirac-Coulomb(-Gaunt) Hamiltonian

Loïc Halbert,<sup>1, a)</sup> Marta L. Vidal,<sup>2, b)</sup> Avijit Shee,<sup>3, c)</sup> Sonia Coriani,<sup>2, d)</sup> and André Severo Pereira Gomes<sup>1, e)</sup>

<sup>1)</sup>Université de Lille, CNRS, UMR 8523 – PhLAM – Physique des Lasers, Atomes et Molécules, F-59000 Lille, France

<sup>2)</sup>DTU Chemistry – Department of Chemistry, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

<sup>3)</sup>Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

We report an implementation of the core-valence separation approach to the 4-component relativistic Hamiltonian based equation-of-motion coupled-cluster with singles and doubles theory (CVS-EOM-CCSD), for the calculation of relativistic core-ionization potentials and core-excitation energies. With this implementation, which is capable of exploiting double group symmetry, we investigate the effects of the different CVS-EOM-CCSD variants, and the use of different Hamiltonians based on the exact 2-component (X2C) framework, on the energies of different core ionized and excited states in halogen (CH<sub>3</sub>I, HX and X<sup>-</sup>, X = Cl-At) and xenon containing (Xe, XeF<sub>2</sub>) species. Our results show that the X2C molecular mean-field approach [Sikkema et al. J. Chem. Phys. 2009, 131, 124116], based on 4-component Dirac-Coulomb mean-field calculations (<sup>2</sup>DC<sup>M</sup>) is capable of providing core excitations and ionization energies that are nearly indistinguishable from the reference 4-component energies for up to and including fifth-row elements. We observe that two-electron in-tegrals over the small-component basis sets yield non-negligible contributions to core binding energies for the K and L edges for atoms such as iodine or astatine, and that the approach based on Dirac-Coulomb-Gaunt mean-field calculations (<sup>2</sup>DCG<sup>M</sup>) are significantly more accurate than X2C calculations for which screened two-electron spin-orbit interactions are included via atomic mean-field integrals.

## I. INTRODUCTION

X-ray spectroscopies, which typically probe core electrons through electronic excitation or ionization, are particularly suitable techniques to study the local environment of atoms, molecules and materials, as the localized nature of the core orbitals makes them very selective and sensitive.<sup>1,2</sup>

Over the last few years, new X-ray free electron lasers (XFEL) and last generation synchrotrons have started operating. These advanced light sources have opened the door to a variety of new X-ray-based spectroscopies,<sup>3–5</sup> including those operating in time-resolved and non-linear regimes.<sup>2–6</sup> As stated for instance by Milne, Penfold and Chergui in their review:<sup>7</sup> “[...] The progress of experimental techniques for core level spectroscopies is unraveling subtle spectral features implying that high level theoretical approaches are required to interpret them [...]”. The experimental X-ray spectra need to be compared to highly accurate theoretical calculations to assign spectral features and to relate experimental measurements to the structure and dynamic properties of the probed molecular system.

Among the methods one can use to calculate core binding energies  $\Delta\text{SCF}$ <sup>8,9</sup> stands out due to the combination

of low computational cost and good results. By performing separate SCF calculations on the original (generally neutral, and closed-shell) and ionized (generally open-shell) species, it introduces the important orbital relaxation that accompanies the creation of the core hole, though at the expense of requiring several calculations for the edges of interest, and with the potential complication that the required open-shell calculations may be difficult to converge. The same approach can be applied in a straightforward manner to DFT,<sup>10</sup> with the additional complication that it will show a certain dependence on the chosen density functional. The same idea of performing separate calculations for the initial and the target states can be used to devise approaches based on correlated wavefunctions such as MP2<sup>11</sup> or coupled-cluster (CC).<sup>12,13</sup> Analogous methods for core excitations have also been formulated,<sup>14–22</sup> though this has proven somewhat more cumbersome than for ionizations, since one has to select orbital pairs that should represent the transitions, and converge such excited configurations. In a number of these approaches the orbitals for the ground and excited/ionized states are not orthogonal, making for instance the calculation of transition moments more involved.

The difficulties above are circumvented with approaches that diagonalize the same Hamiltonian for both the original and the ionized or excited states. As such, excited state approaches based on wavefunction theory, like, for instance, (multireference) configuration interaction (MRCI),<sup>23–26</sup> equation of motion CC (EOM-CC)<sup>27–35</sup> and multireference (EOM)CC,<sup>26,36</sup> and linear response complete active space SCF (LR-CASSCF),<sup>37</sup> on Green's functions such as the algebraic diagrammatic

a) Electronic mail: [loic.halbert@univ-lille.fr](mailto:loic.halbert@univ-lille.fr)

b) Electronic mail: [malop@kemi.dtu.dk](mailto:malop@kemi.dtu.dk)

c) Electronic mail: [ashee@umich.edu](mailto:ashee@umich.edu)

d) Electronic mail: [soco@kemi.dtu.dk](mailto:soco@kemi.dtu.dk)

e) Electronic mail: [andre.gomes@univ-lille.fr](mailto:andre.gomes@univ-lille.fr)

construction (ADC),<sup>38–41</sup> or on time-dependent density functional theory (TDDFT),<sup>10,42,43</sup> can all be used to target core excited or ionized states. One downside is, however, that now relaxation has to be accounted for by the correlated electronic structure approach. Among these, single-reference CC methods, and in particular those based on EOM and the closely related linear response (LRCC) formalisms,<sup>44–50</sup> are capable of treating electron correlation accurately and in a balanced way between ground and excited states, including a good deal of relaxation effects, and have an appealing “black box” nature (though they are less adept than MRCI/MRCC at treating cases for which the ground state shows a strong multireference nature).

Irrespective of the correlated method, a particular difficulty for determining core states by diagonalizing a many-body Hamiltonian is that such states appear at high energies, where spectra are potentially very dense (see for example figure 2). Thus, a naive application of procedures which target exterior roots (such as the Davidson algorithm), would require solving for a very large number of roots. One way to overcome this issue is through the use of methods capable of solving for interior roots without requiring that all of them are targeted, as those recently devised and applied to multiconfigurational,<sup>37</sup> EOMCC<sup>32,51</sup> and mean-field<sup>52,53</sup> wavefunctions.

A more widespread strategy,<sup>23,28,29,31,34,35,37–40,54–57</sup> which we follow in this paper, is to employ physically motivated approximations to the original Hamiltonian, in particular within the core valence separation (CVS) approximation.<sup>58</sup> The core-valence separation is justified by the large spatial and energetic separation between valence and core orbitals/electrons. On the one side, CVS allows to straightforwardly address the highly-energetic core states with minor modifications to preexisting eigenvalue solvers. On the other side, the CVS helps alleviating convergence issues due to the large number of valence ionized states in the frequency region of the core excitations, which is particularly afflict-ing methods that explicitly include double excitations in their parametrization.<sup>59,60</sup> One can also regard the core states as metastable Feshbach resonances embedded in the (valence) ionization continuum, and application of CVS helps stabilizing them. Further details on the CVS approach will be given in section II C.

A second issue to be addressed for core states is the importance of relativistic effects,<sup>61–63</sup> which significantly alter the energies of inner electrons, and consequently the core spectra of atoms and molecules: due to the high velocities of their electrons (a significant fraction of the speed of light), core s and p orbitals contract (with an associated lowering of their orbital energies) while, due to increased screening, d and f orbitals expand (with an associated increase in their orbital energies). At the same time, orbitals which are degenerate in a non-relativistic framework (p, d, f) are (strongly) split due to spin-orbit coupling, and particularly so for the innermost orbitals.

Relativistic effects are so pronounced in the core re-

gion that, even for molecules containing only elements of the first and second rows, they are very important for the accurate determination of K and L edge spectra. While rather approximate treatments of relativistic effect can yield accurate results for light elements,<sup>12,64–67</sup> a more consistent way of treating these effects is through an electronic Hamiltonian based on the 4-component Dirac operator, as it is accurate all the way down to the heaviest elements (5d and 6d transition metals, lanthanides and actinides). As such, 4-component based methods are ideally suited to treat the core spectra across the periodic table, and to probe K, L and M edges of heavy elements,<sup>11,68,69</sup> which are much more complex to interpret than those of lighter elements. Though the availability of efficient implementations<sup>70–72</sup> has in recent years enabled mean-field 4-component calculation for large-scale applications, the additional basis sets required to represent the small component part of the molecular spinors still places a considerable burden on such calculations compared to the non-relativistic case.

An important development to overcome such issues has been the introduction of the so-called eXact 2-Component (X2C) methods<sup>73–79</sup>, in which a transformation to decouple the positive and negative energy states of the Dirac Hamiltonian is available in matrix form and, unlike its more approximate counterparts, yields exactly the same positive energy spectrum as the original 4-component Hamiltonian. We refer the reader to a recent, comprehensive review on relativistic electronic structure for more details<sup>63</sup>.

We note that among the different X2C variants, those that construct the transformation matrix on the basis of converged 4-component atomic or molecular mean-field calculations<sup>75,77,79</sup> are particularly interesting to correlated calculations<sup>79–83</sup>, since they avoid the handling of two-electron integrals over small component basis functions in the transformation to molecular spinors, while largely avoiding errors with respect to the original 4-component Hamiltonian on the description of two-electron spin-orbit coupling contributions, compared to approaches in which the decoupling is performed based on the one-electron Dirac Hamiltonian prior to the mean-field step<sup>76</sup>.

In this study, we present an approach to study core ionization (and excitation) through the use of projectors, inspired by the work of Coriani and Koch,<sup>28,29</sup> but extended to the 4-component EOM-CCSD approach developed by Shee et al.<sup>84</sup> in the Dirac relativistic electronic structure package.<sup>85</sup> Because of the availability of different Hamiltonians in Dirac, we will also investigate the performance of 2-component approaches<sup>79</sup> with respect to the 4-component one. We refer to, e.g., Refs. 86–101 for other examples of relativistic EOM-CC implementations, and to Refs. 12,31,34,35,55,59,60,65,102,103 for other examples of CVS-EOM-CC implementations. We also wish to underline that the literature on the theoretical approaches for core spectroscopy is vast and rapidly increasing, and the works cited in the previous

paragraphs cannot be considered exhaustive. We refer the reader to a number of recent review papers for more thorough accounts.<sup>2,24,43,104</sup>

The paper is organized as follows: In Section II, we give a brief outline of EOM-CCSD (II A) and of CVS (II B), as well as the details of the current implementation (II C). In Section III, we provide the computational details of the calculations. In Section IV, we present the results obtained with the newly implemented method, where we discuss the accuracy of the CVS approximation (IV A), the performance of the different CVS variants (IV B), the influence of the Hamiltonian (IV C) and the comparison to experiment for ionization (IV D) and excitation (IV E) energies. Finally, Section V summarizes our conclusions.

## II. METHODS AND IMPLEMENTATION

In what follows, indices (i, j, k, l), (a, b, c, d) and (p, q, r, s) refer to occupied, virtual and general orbitals, respectively.

### A. EOM-CC

In EOM-CC, the ground-state is treated at the coupled-cluster level. Its eigenfunction is hence given by the exponential ansatz:

$$|\Psi_{CCi}\rangle = e^{\hat{T}} |\Phi_0\rangle; \quad \hat{T} = \sum_{\mu} t_{\mu} \hat{T}_{\mu} \quad (1)$$

where  $\Phi_0$  is the reference (typically Hartree-Fock) determinant and the operator  $T$  is the cluster operator. Truncating  $T$  to single (S) and double (D) excitations yields the coupled-cluster singles-and-doubles (CCSD) model:

$$\hat{T} = \hat{T}_1 + \hat{T}_2; \quad \hat{T}_1 = \sum_{ia} t_a a_i^\dagger; \quad \hat{T}_2 = \frac{1}{4} \sum_{ijab} t_{ijab} a_i^\dagger a_j^\dagger a_b a_j \quad (2)$$

The energy and the cluster amplitudes are found from the CC equations:

$$\langle \Phi_0 | \hat{H} | \Phi_0 \rangle = E \quad (3)$$

$$\langle \Phi_{\mu} | \hat{H} | \Phi_0 \rangle = 0; \quad |\Phi_{\mu}\rangle = \hat{T}_{\mu} |\Phi_0\rangle, \quad (4)$$

where the similarity-transformed Hamiltonian  $\hat{H}$  has

been defined as:

$$\hat{H} = e^{-\hat{T}} H e^{\hat{T}} \quad (5)$$

In the equation-of-motion coupled-cluster (EOM-CC) method<sup>44,47</sup> the target states are obtained by the diagonalization of the non-Hermitian similarity-transformed Hamiltonian. This non-Hermiticity gives rise to right (R) and left (L) eigenvectors that are not adjoints of each

other, obtained solving two different eigenvalue equations:

$$\hat{H} |R_{\mu}\rangle = E_{\mu} |R_{\mu}\rangle \quad (6)$$

$$\langle L_{\mu} | \hat{H} = E_{\mu} \langle L_{\mu} |, \quad (7)$$

for a given excited state  $\mu$  with energy  $E_{\mu}$ , where the eigenstates are chosen to satisfy the biorthogonality condition:

$$\langle L_{\mu} | R_{\nu} \rangle = \delta_{\mu\nu}. \quad (8)$$

Right and left wavefunctions of the target states have been thus obtained from a linear parametrization of the reference state through the R or L operators

$$|\Psi_{\mu}\rangle = e^{\hat{R}} |R_{\mu}\rangle |\Phi_0\rangle \quad (9)$$

and

$$\langle \Psi_{\mu} | = \langle \Phi_0 | e^{\hat{L}} |L_{\mu}\rangle \quad (10)$$

Therefore, the choice of the R and L operators defines which target states to study, yielding different EOM-CC models. In the present work we focus on the models for electronically excited (EE) states and ionization potentials (IP):

• EOM-CCSD-EE:

$$\hat{R}^{EE} = r_0 + \sum_{ia} r_i^a \{a_a^\dagger a_i\} + \sum_{i>j,a>b} r_{ij}^{ab} \{a_a^\dagger a_b^\dagger a_j a_i\} \quad (11)$$

$$\hat{L}^{EE} = l_0 + \sum_{ia} l_a^i \{a_i a_a^\dagger\} + \sum_{i>j,a>b} l_{ab}^{ij} \{a_i a_j a_b a_a^\dagger\} \quad (12)$$

• EOM-CCSD-IP:

$$\hat{R}^{IP} = r_x \{a_x\} + \sum_{i>j,a} r_{ij}^a \{a_a^\dagger a_j a_i\} \quad (13)$$

$$\hat{L}^{IP} = l_i \{a_i\} + \sum_{i>j,a} l_a^{ij} \{a_i a_j a_a^\dagger\} \quad (14)$$

where curly brackets refer to normal ordering with respect to the Fermi vacuum defined by the reference  $\Phi_0$ , and the sets  $\{r\}$ ,  $\{l\}$  to the amplitudes of the corresponding operators.

We have here truncated our R and L operators at the singles-doubles level since the same truncation is used for the T operators.

$$\hat{H}^{-1} \equiv e^{-\hat{T}} H e^{\hat{T}}.$$

## B. The core-valence separation approximation

The essence of the core-valence separation (CVS) approximation<sup>58</sup> is to decouple valence and core electrons based on their difference in energy and spatial extension. This allows to solve the regular, in this case, EOM-CC equations only in the space of the relevant orbitals. However, different flavours of CVS exist, which introduce different levels of approximation. Coriani and Koch first introduced the CVS approximation within coupled-cluster theory<sup>28,29</sup> by applying a projector that zeroes out the amplitudes of all excitations that do not involve at least one core electron. Recently, the frozen-core (fc) CVS-EOM-CCSD approach has been proposed that introduces a further approximation, in this case at the ground-state level: the core orbitals are frozen when solving the CC equations for the ground-state whereas they are the only active ones when solving the EOM equations. Both schemes retain the contribution from excitations simultaneously involving two core orbitals. However, these excitations are located in a much higher range of energy in the spectrum and might therefore also be decoupled. This is the strategy adopted for instance by Dreuw and co-workers in their implementation of the CVS within the ADC family of methods.<sup>39,40,105</sup>

In this work, the CVS implementation has been carried out following the recipe proposed by Coriani and Koch:<sup>28</sup> a projector  $P$  is applied at each iteration of the Davidson procedure during the resolution of the EOM-CC equations. This projector  $P$  selectively zeroes out the unwanted contributions to the EOM trial/solution vectors according to the approximation used. In the CVS approach, unwanted contributions generally correspond to excited determinants involving only occupied valence ( $v$ ) spinors. In the case of EOM-EE, the application of a projector  $P \equiv P^{CVS}$  to a target electronic state  $|\Psi_{\mu i}^{EE}\rangle$  corresponds to

$$P^{CVS}|\Psi_{\mu i}^{EE}\rangle \Rightarrow \mu_{ri}^a = \mu_{rij}^{ab} = 0 \text{ if } i, j \in v, \quad (15)$$

though one can also devise variants in which additional contributions are zeroed out, see Section II C for details. With this definition, the eigenvalue equation to be solved in order to get the energy of the target states becomes:

$$P^{CVS} H P^{CVS} |\mathbf{R}_{\mu i}\rangle = E_{\mu} P^{CVS} |\mathbf{R}_{\mu i}\rangle \quad (16)$$

It should be noted that the projection scheme does not yield any savings, in terms of memory usage or operations. Some of us<sup>31</sup> have recently gone beyond the use of projection operators and introduced an analytical formulation of CVS, in which only the sub-blocks of the similarity-transformed Hamiltonian that are relevant to the calculation of a given core level are actually formed, thus resulting in a more efficient implementation. This formulation is naturally combined with the frozen-core approximation separating valence and core spaces. However, one may argue it introduces at the same time an additional error from neglecting core correlation.<sup>31,103</sup>

## C. Implementation details

In this work we focus on exploring the definition of different CVS variants and on assessing their performance following the projection-based scheme, due to its ease of implementation, and with it in view of later implementing the analytical CVS<sup>31</sup> formulation in Dirac.

We have based our implementation on a flexible scheme to define projection operators.

- (a) A mapping is defined between the excited determinants (and the virtual and occupied spinors' energies associated with each) and the position of each excited determinant in the storage of the ground-state CC amplitudes and EOM-CC coefficients (in the RELCC module of Dirac these are stored in triangular form, and blocked by symmetry, for details see Refs. 106–108);
- (b) Information is gathered on how the excited determinant space  $\{v\}$  will be treated, i.e. whether retained ( $P^v = 1$ ) or projected out ( $P^v = 0$ ). We have made two options available to the user, based on spinor energies: (i) restricted excitation windows (REWs) for the occupied and virtual, which are defined by setting respective lower ( $L$ ) and upper ( $H$ ) bounds for the spinor's energies; and (ii) CVS, via a single energy that acts as the threshold ( $\epsilon$ ) for the separation between core and valence;
- (c) Using the information from the previous steps, suitable one-particle (singly ionized or excited) unit trial vectors for the core states are then generated, using the values of the diagonal of  $H$  within the subspaces defined in (b).

It is important to note that in our implementation, tracking the core excited or ionized states requires the use of a root homing procedure, in which new trial vectors are created by maximizing their overlap with the preceding ones.

In addition to CVS and REW for the excited states, we also used the projection setup to implement the frozen-core approximation, such that we can project out the amplitudes corresponding to core ( $c$ ) orbitals at each iteration during the resolution of the ground-state amplitude equations:

$$P^{core}|\Psi_{CCi}\rangle \Rightarrow t_i^a = t_{ij}^{ab} = 0 \text{ if } i, j \in c \quad (17)$$

As for the excited states, this projector is defined in terms of spinor energies, via a single threshold or upper and lower bounds defining a window. Finally, it should be noted that the thresholds defining the ground-state and excited state projectors are independent.

## III. COMPUTATIONAL DETAILS

All coupled-cluster calculations were carried out with the Dirac electronic structure code<sup>85</sup> (with the

DIRAC19<sup>109</sup> release and revisions dbbfa6a, 0757608, 323ab67, 2628039, 1e798e5, b9f45bd). The Dyall basis sets<sup>110–112</sup> of triple-zeta quality (dyall.acv3z) were employed for all species. For selected calculations, on heavy atoms only, we performed extrapolations to the complete basis set limit by also considering quadruple-zeta quality Dyall (dyall.acv4z) basis sets. In addition to the Dyall basis sets, the ANO-RCC basis<sup>113</sup> were employed for Xe and XeF<sub>2</sub>. The basis sets were kept uncontracted in all calculations.

Unless otherwise noted, all occupied and virtual spinors were considered in the correlation treatment.

Apart from the Dirac–Coulomb (<sup>4</sup>DC) Hamiltonian, for selected calculations we investigated: (a) the molecular mean-field<sup>79</sup> approximation to DC (<sup>2</sup>DC<sup>M</sup>) and the Dirac–Coulomb–Gaunt (<sup>2</sup>DCG<sup>M</sup>) Hamiltonians. For the latter, the Gaunt-type integrals are explicitly taken into account only during the 4-component SCF step, due to the fact that the transformation of these to MO basis is currently not implemented; (b) the DC Hamiltonian with projecting out all negative-energy solutions of the bare-nucleus one-electron Dirac Hamiltonian from the molecular spinor space, which corresponds to the Furry basis for QED and no-pair Hamiltonians<sup>114</sup> (<sup>4</sup>DC<sup>PF</sup>); and (c) the eXact 2-component (X2C)<sup>76</sup> Hamiltonian, in which we include two-electron spin-orbit contributions via the untransformed two-electron potential via atomic mean-field contributions calculated with the AMFI code<sup>115–117</sup>, namely spin-same-orbit (X2C-AMFI) and spin-same-orbit and spin-other-orbit (X2C-G-AMFI).

Unless otherwise noted, we employed the usual approximation of the energy contribution from (SS|SS)-type two-electron integrals by a point-charge model.<sup>118</sup> As it will be discussed below, this approximation introduces negligible errors (< 0.01 eV) for lower-energy edges. However, for the higher-energy edges of the heavier systems, its error becomes important.

The XeF<sub>2</sub> structure was taken from Ref. 119. It corresponds to an optimised structure obtained at the SFX2C-1e/coupled-cluster single double triple (CCSD(T))/unc-atomic natural orbital (ANO)-RCC level (r<sub>Xe-F</sub> = 1.9736 Å). The coordinates of CH<sub>3</sub>I come from experimental data in Ref. 120. Finally, the coordinates for HX were taken from Ref. 121.

The dataset can be retrieved at the Zenodo repository.<sup>122</sup>

#### A. Approximations used in CVS

Beside the original CC-CVS approach,<sup>28</sup> we have in addition investigated other approximations:

1. Further restricting the definition of the projectors to also drop excited configurations with two core indices (ND, for “No Doubly core hole determinants”); this corresponds to the method suggested,

e.g., in Ref. 40, making

$$\mu_{r_{ij}}^{ab} = 0 \text{ if } i, j \in c \quad (18)$$

2. Freezing the core in the ground-state calculation, retaining thus only the valence orbitals, typically defined as the orbitals with the highest principal quantum number  $n$  (FC-V); this essentially corresponds to using eq. 17 while setting  $c_V$  to a fairly high value (for example, in the atom of Xe,  $4d < c_V < 5s$ ).
3. Freezing all core orbitals except those that are to be targeted in the EOM step (FC-V-except); for example, this means treating a core spinor  $k$ , that should make up the most important core-excited configurations for state  $\mu$ , differently from other core spinors in the ground-state calculation. This corresponds to setting projectors for ground and excited states such that

$$t_i^a = t_{ij}^{ab} = 0 \text{ if } \begin{matrix} i, j \\ i, j \end{matrix} = \begin{matrix} k \\ c \end{matrix} \quad (19)$$

$$\mu_{r_i}^a = \mu_{r_{ij}}^{ab} = 0 \text{ if } i, j \in v \quad (20)$$

This approach is equivalent to the one of Sorensen et al.<sup>103</sup>

4. The definition of the core/valence spaces follows that of the CVS definition/threshold, that is, only the core orbitals with same or lower energy than the ones belonging to the edge under investigation are frozen (FC-f). This is the same approach adopted in Ref. 31, and it corresponds to setting projectors for ground and excited states such that

$$t_i^a = t_{ij}^{ab} = 0$$

$$\text{if } i, j \in v \left( \begin{matrix} \mu_{r_{ia}} \\ \mu_{r_{ijab}} \end{matrix} = 0 \right) \quad (21)$$

5. Further restricting the number of frozen core spinors to only those below the ones of the edge of interest (FC-fpMO frozen core - follow previous MO). For example, in the atom of Xe, if we are interested in the M<sub>5</sub> edge (3d<sub>5/2</sub>), this would correspond to a frozen core in the ground-state calculation including the 1s<sub>1/2</sub> to 3d<sub>3/2</sub> spinors, and the 4s<sub>1/2</sub> and higher spinors make up the valence space in the EOM calculations.

The variants ND and FC-V were also employed with REWs. Since the performance of REW variants was always found to be inferior to their CVS equivalents, they will not be discussed here. The results are however available as supplemental information.

## IV. RESULTS

As outlined above, apart from the different Hamiltonians at our disposal, the original CVS formulation itself

can be modified through a number of different approximations. In order to draw a clearer picture of their interactions, we proceed in a stepwise fashion: first, using the  ${}^2\text{DC}^{\text{M}}$  Hamiltonian and focusing on core ionizations, we compare (A) the performance of CVS-EOM-CCSD to the original EOM-CCSD; and (B) the impact of the different approximations that can be introduced in the CVS method itself (see Section III A). Then, having established the relative accuracy of CVS, we investigate (C) the effect of the Hamiltonians on the core ionization energies, and finally (D) proceed to a comparison to experiment of the most accurate setup, focusing mostly on core ionization but also discussing selected core excitations (E).

In what follows we will not discuss basis set effects. Nonetheless, in Figure 1, we show for  $\text{I}^-$  a typical behavior: first, there is very little difference between triple and quadruple zeta results (with differences between 0.05 and 0.2 eV), making the corrections due to basis set incompleteness so small that we consider these to be unnecessary. Second, if there are non-negligible differences between the convergence of s, p and d shells as the basis sets are improved, the double-zeta basis sets also yield results which are quite good, with differences in binding energies with respect to the triple zeta results no larger than around 0.4 eV, a behavior that may make these smaller basis sets interesting for calculations on larger molecules.

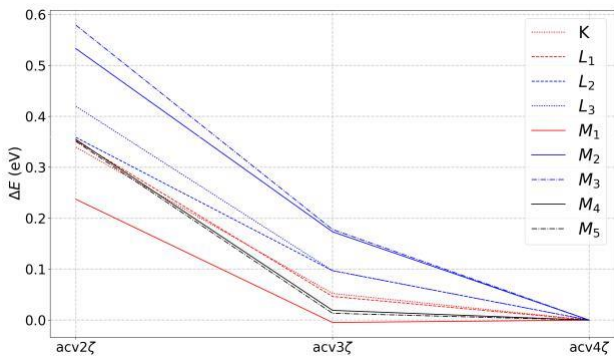


FIG. 1: Basis set influence (Dyall basis sets) on the core binding energies of  $\text{I}^-$  using the CVS approach. All calculations performed with the  ${}^2\text{DC}^{\text{M}}$  Hamiltonian, and basis sets are kept uncontracted.

These conclusions are in line with earlier basis set analyses on light elements.<sup>64,123</sup> In the basis set convergence study by Sarangi et al.,<sup>123</sup> in particular, it was shown that, for systems containing only light elements, uncontracted basis sets, even of relatively modest quality for the valence, can provide quite reliable core binding energies.

#### A. A comparison of CVS-EOM-IP-CCSD and EOM-IP-CCSD

As mentioned above, the CVS approach provides an efficient and robust way of targeting the low-lying core states. Quantifying the errors introduced by it is therefore an important issue, and we note proposals for methods to estimate it. Coriani and Koch<sup>28</sup> evaluated the errors of their CVS variant by comparing to full-space Lanczos results, finding deviations of at most a few hundredths of eV. The authors also proposed a perturbative correction obtained using a Löwdin partition of the Jacobian and the eigenvalue equation. A recent study by Herbst and Fransson<sup>124</sup> shows that the CVS error is small and stable across multiple systems within the algebraic diagrammatic method (ADC). Thus, we expect to find the same trend for CC. This article also reports the implementation of a post-processing step which removes the error to assess its significance.

Here, we can harness the ability of Dirac to exploit linear symmetry, and directly compare, for selected species, the CVS-EOM-IP-CCSD states with those obtained by the full diagonalization of the different symmetry blocks of the similarity-transformed Hamiltonian.

The difference between the two approaches is shown graphically for  $\text{I}^-$  in Figure 2. Due to the very dense spectra of H, in the top half of the figure we do not display individual states but rather the number of states per 10 eV intervals. We further discriminate between all states (black) and those containing contributions from singly ionized configurations contributing to at least 1% of the total wavefunction (red).

Upon having a closer look at the singly ionized states at the bottom half of the figure, we can more clearly see how the singly ionized states obtained with the CVS approximation (labelled “CVS”) closely match a subset of those obtained by the full diagonalization (labelled “full”) of H.

We note that the additional states presented for the “full” calculation correspond to states that contain significant singly-ionized character but also small but non-zero 2h1p contributions. The presence of such states hint at potential pitfalls when targeting only a subset of such highly excited states for some of the L and M edges— not so much in terms of convergence (we have not encountered particular difficulties in converging our calculations) but rather in terms of assignment and comparison to other theoretical or experimental results.

#### B. Performance of the CVS-EOM-CCSD variants

Next we turn our attention to the performance of the different approximations that can be employed on top of CVS (see Section III A for their description). Our results for the  $\text{X}^-$  systems are shown in Figure 3. As we see similar trends for Xe, the corresponding figure is shown in the supplemental information (Figure S1),

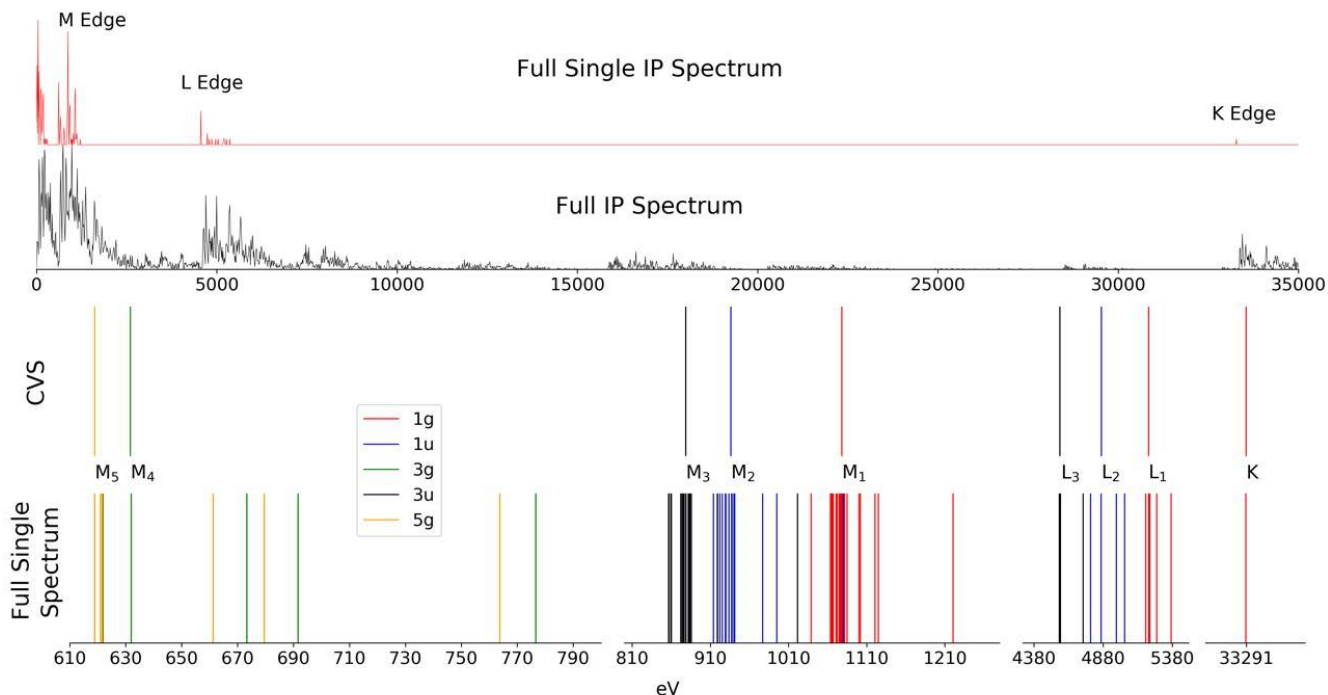


FIG. 2: Core ionization energies of iodide (I) from full diagonalization of H. Top: Representation of the eigenspectrum up to 35 keV, separated between all states (black) and those with non-negligible singly ionized contributions (red). Values in ordinate correspond to the number of states within a 10 eV interval, with the largest value re-scaled to one. Bottom: Singly ionized states obtained by full diagonalization of H and by the CVS approach. The labels  $ng/nu$  indicate the absolute value for the projection of the angular momentum ( $|m_j| = n/2$ ,  $n = 1, 3, 5$ ), for the components of spinors of gerade (g) or ungerade (u) symmetry. All calculations performed with the  ${}^2DC^M$  Hamiltonian.

along with results for the REW approach for the  $Cl^-$  to  $I^-$  (Figures S2, S3 and S4) and HCl to HI (Figures S5, S6 and S7).

### 1. Freezing core orbitals in the ground-state

Freezing the core orbitals in the ground-state CCSD calculation results in a significant lowering of the core ionization energies. This lowering is considerable if the frozen core is taken to represent all subvalence spinors (FC-V), and it remains non-negligible even if the same threshold is used to define both the frozen orbitals and the core-valence separation region (FC-f). However, if one does not freeze the aimed edge (FC-fpMO), the ionization energy obtained is much closer to the corresponding CVS one.

Freezing all subvalence spinors but the ones in the edge we are interested in (FC-V-ex), an approximation that could provide potentially lower computational cost for systems with large number of core electrons such as the elements in the fourth row and beyond, does not perform better than (FC-f)—in fact, the opposite appears to be the case for the inner core orbitals.

### 2. Projecting out double-core ionized configurations

Projecting out doubly-excited core excited determinants (ND) tends to result in an increase of the core ionization energies, which can be significant though always smaller than the underestimation produced by freezing all the core spinors.

As a result, the combination of these two approximations yields an approximate scheme that much better reproduces the CVS energies, though still with non-negligible discrepancies for the heavier elements. Particularly good results are shown by the (FC-ND-f) combination, for which error compensation yields results very similar to the regular CVS approximation. This can be rationalized because core correlation is not taken into account when using (FC-f), whereas the inclusion of the double occupied core orbitals in the EOM step does. Therefore, the combination of both approximations is in fact more consistent.



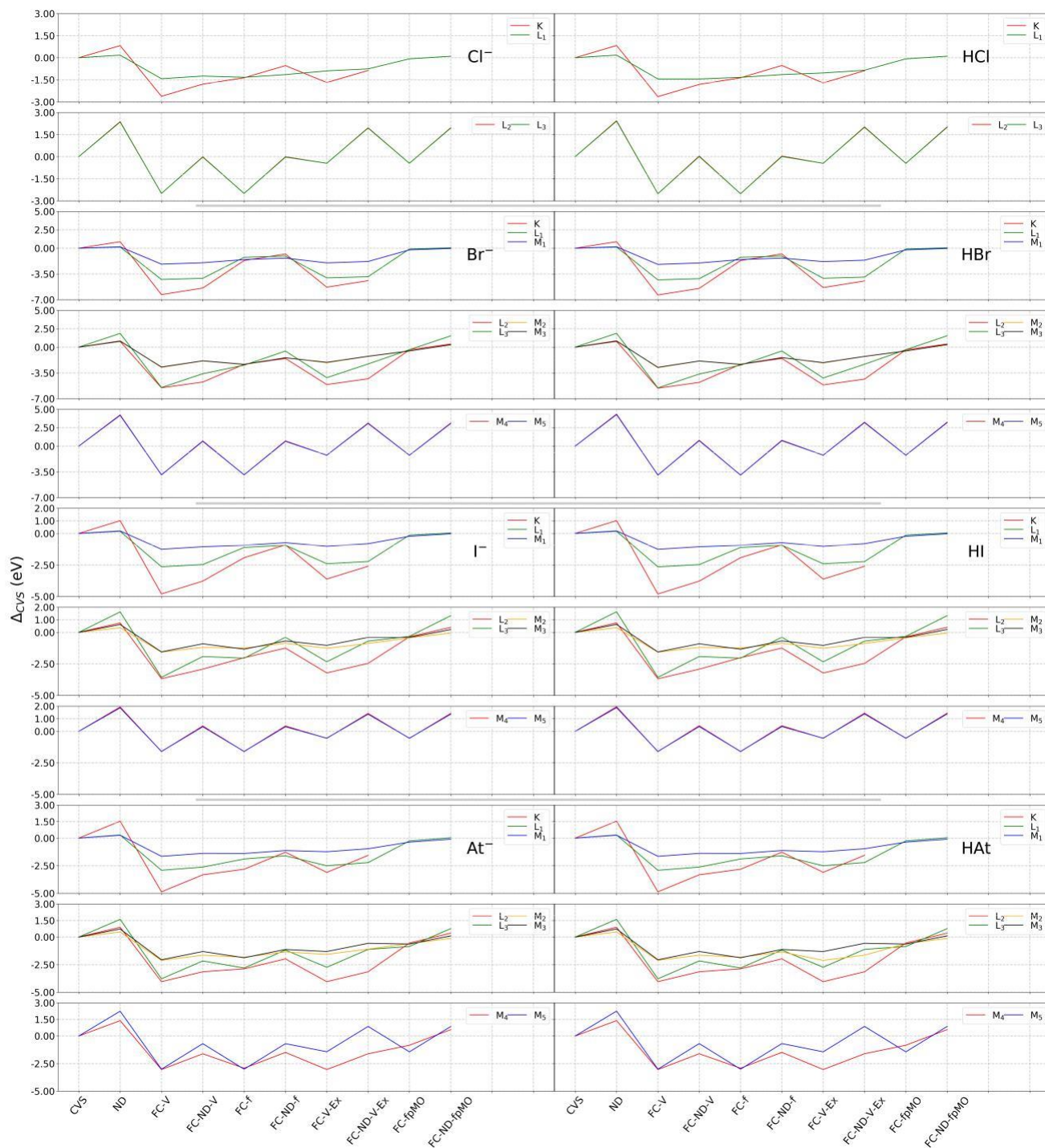


FIG. 3: Effect of the CVS variants on the binding energies of  $X^-$  and  $HX$  systems ( $X = \text{Cl}, \text{Br}, \text{I}$  and  $\text{At}$ ). Values (in eV) are relative to the original CVS approach. All calculations performed with the  $^2\text{DC}^M$  Hamiltonian.

### 3. Efficiency considerations

Since by using projection operators we do not save in memory or operation counts, the approximations dis-

cussed above are not of strong interest by themselves, and will not be used further since, even in the best combination, they will invariably degrade the performance of the original CVS scheme.

We can nevertheless use our findings to discuss the potential trade-offs between cost and accuracy, as a guide to efficient implementations such as those proposed in Ref. 31 for heavy elements. As discussed then, combining the elimination of doubly core excited determinants with the use of a frozen core (FC-ND-f, which is equivalent to the scheme in Ref. 31) seems to yield a computationally efficient approach with a good error cancellation balance, and this remains the case as one goes down the periodic table.

In the case of heavy elements, it may also be interesting to keep a large frozen core (CVS-FC-ND-V, or CVS-FC-V-Ex) since these introduce non-negligible, but seemingly systematic, errors, with the upside that a large core would translate in potentially large computational savings.

This is illustrated in Table S1, where we provide, for the representative systems  $\text{Cl}^-$  and  $\text{At}^-$ , the operation counts (without reductions due to point group symmetry) for the construction of EOM-EE and EOM-IP  $\sigma$  vectors (based on the expressions from Shee et al.<sup>84</sup>, see expressions in the supplemental information), in the case of an analytical implementation of the major approximations suggested in this work (FC-f and ND). As seen, the FC-f approximation yields the biggest reductions in operation counts, in particular for EOM-IP. It is also important to note that for a given edge, the heavier the element, the bigger these reductions are.

### C. Influence of the Hamiltonian on the core ionizations

We now turn our attention to the impact of the Hamiltonian on the ionization energies. Our results can be found in Figure 4, where we take the  $^4\text{DC}$  Hamiltonian as reference and we plot the difference in binding energies between it and the other Hamiltonians considered (see figure caption for details), as  $^4\text{DC}$  is the only 4-component Hamiltonian without approximations (apart from the treatment of the (SS)SS integrals which we address below) we can employ in correlated calculations with Dirac.

First, we observe a very good match between the  $^4\text{DC}$  and  $^2\text{DC}^{\text{M}}$  Hamiltonians across the halogen series—that is, differences in absolute values generally fall below 0.001 eV for the M edges, between around 0.01 and 0.1 eV for the L edges, and are of the order of 0.1 eV for the K edge of all species except astatine, for which the difference is 1.67 eV. This latter discrepancy is still small compared to the K edge binding energy of 96 keV.

To better understand how these discrepancies arise we should recall that in the  $^2\text{DC}^{\text{M}}$  ( $^2\text{DCG}^{\text{M}}$ ) approximation a calculation with the  $^4\text{DC}$  ( $^4\text{DCG}$ ) Hamiltonian is first carried out for whatever system we are interested in—in general a molecule, hence the denomination “molecular mean-field”, though in Figure 4 we are in effect dealing with atoms. Upon convergence, the transformation to 2-component is carried out on the Fock matrix itself (so

that the 2-component spinor energies correspond exactly to the 4-component ones), but the two-electron operator is left untransformed<sup>79,85</sup> and thus introduces a picture-change error in the electron-electron interaction. In valence only calculations, where core-core and core-valence electron correlation is not accounted for, this yields negligible errors for excitation, ionization and electron attachment processes<sup>84</sup>. Our results show that this is still the case for core states of light atoms such as chloride, and for the M and N edges down to the sixth row.

We can numerically assess the effect of using untransformed two-electron operator by comparing how the correlation energy ( $E_c$ ) differ between  $^2\text{DC}^{\text{M}}$  and  $^4\text{DC}$  ( $\Delta E_c$ ), as shown in table I. The  $E_c$  for both MP2 and CCSD yield the same trends, so we shall focus on the simpler MP2 model; as canonical orbitals are employed, and their energies the same for  $^2\text{DC}^{\text{M}}$  and  $^4\text{DC}$ , any differences between the two come from the picture change error in the two-electron integrals. We see that for chloride  $E_c$  is still relatively small, then it increases five fold going from chloride to bromide, doubles from bromide to iodide, and then increases nearly fourfold from iodide to astatide.

The effect of the two-electron picture change error can also be seen through a comparison of  $^4\text{DC}$  to a calculation in which negative-energy solutions of the bare-nucleus one-electron Dirac Hamiltonian are projected out from the molecular spinor space ( $^4\text{DC}^{\text{P F}}$ ), whereby rotations between negative and positive energy states are effectively eliminated. With this,  $^4\text{DC}^{\text{P F}}$  spinor energies are slightly different than  $^4\text{DC}$  ones (see dataset<sup>122</sup>), but the picture change error in the two-electron integrals is eliminated, and we would thus expect rather similar correlation energies. This picture is consistent with our numerical results, as correlation energies for  $^4\text{DC}^{\text{P F}}$  are closer to the  $^4\text{DC}$  ones by at least three orders of magnitude than the  $^2\text{DC}^{\text{M}}$  ones.

As it will be shown in the following, the sufficiently high accuracy of  $^2\text{DC}^{\text{M}}$  clearly makes it an asset for applications, due to its reduced computational cost in the index transformation step. However, our results indicate two-electron picture change errors are significant enough for deeper cores to require further attention for elements in the sixth row and beyond.

While we do not have  $^4\text{DCG}$  reference values to which compare the approximate Hamiltonians that include the Gaunt interaction such as  $^2\text{DCG}^{\text{M}}$ , we expect two-electron picture change errors to generally follow the same trends discussed above. We nevertheless compare  $^2\text{DCG}^{\text{M}}$  to  $^4\text{DC}$  in Figure 4 and Table I, as a way to underscore the importance of the Gaunt interaction for the inner edges. We clearly see that already starting with chloride, we have a non-negligible effect arising from the Gaunt interaction in  $^2\text{DCG}^{\text{M}}$  that lowers the core binding energy, and can amount to nearly 2 eV for the K edge with respect to the  $^2\text{DC}^{\text{M}}$  or  $^4\text{DC}$ .

We can also compare  $^2\text{DCG}^{\text{M}}$  to the approaches in which the transformation to the 2-component picture

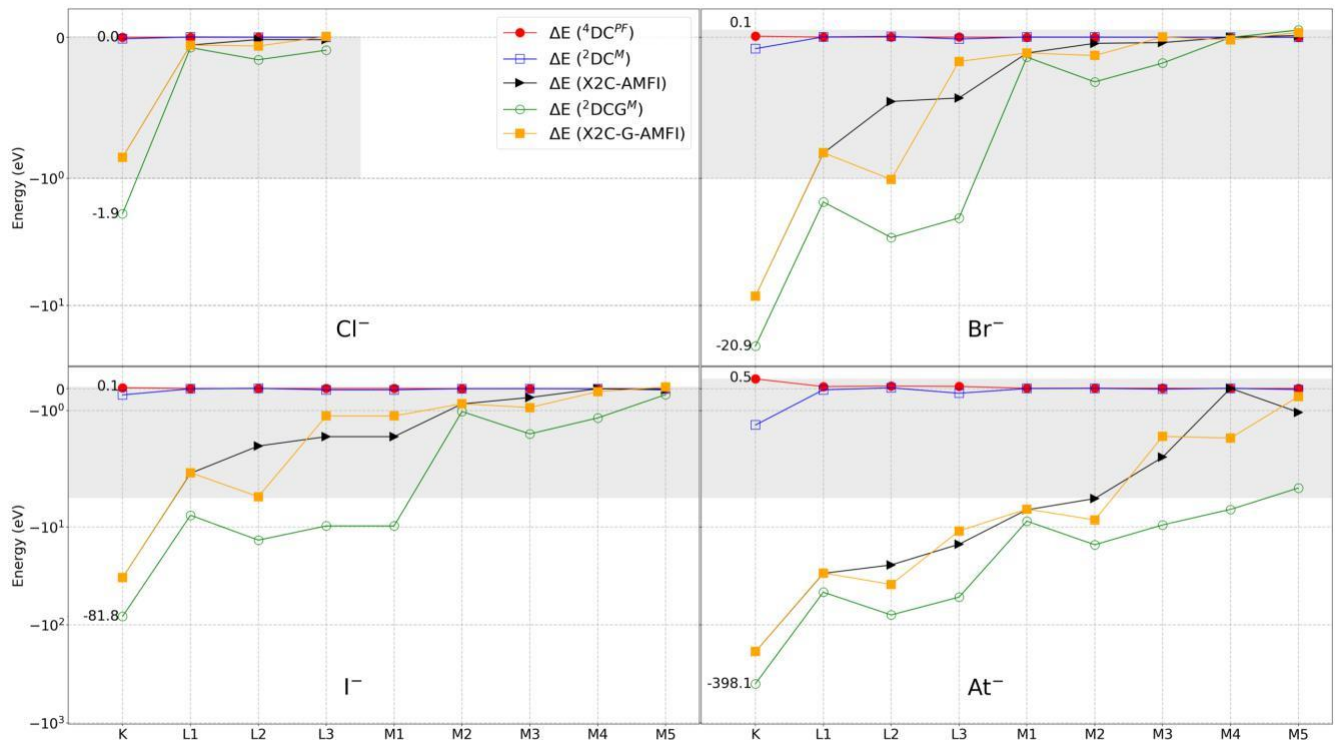


FIG. 4: Comparison of influence of the Hamiltonian to the core binding energies for different edges, for the halide ions ( $\text{Cl}^-$  to  $\text{At}^-$ ). Instead of binding energies themselves, we present the difference between the binding energies obtained for each Hamiltonian  $X$  and  ${}^4\text{DC}$  ( $\Delta E_m(X) = E_m(X) - E_m({}^4\text{DC})$ , in eV,  $m$  a particular edge). The scales are logarithmic except for the areas in grey for which the scales are linear.

TABLE I: MP2 and CCSD correlation energies as well as differences in correlation energy between  ${}^4\text{DC}$  ( $E_c$  and  $E_c$  respectively, in eV), for different Hamiltonians. X2C(a) refers to X2C-AMFI and X2C(b) to X2C-G-AMFI.

	${}^4\text{DC}^{\text{PF}}$	${}^2\text{DC}^{\text{M}}$	${}^2\text{DCG}^{\text{M}}$	X2C(a)	X2C(b)
$\text{Cl}^- E_c(\text{MP2})$	-16.12	-16.16	-16.16	-16.16	-16.16
$E_c < 2 \times 10^{-6}$	-0.04	-0.04	-0.04	-0.04	$E_c(\text{CCSD})$
	-16.43	-16.47	-16.47	-16.47	-16.47
$E_c$	$< 2 \times 10^{-6}$	-0.04	-0.04	-0.04	-0.04
$\text{Br}^- E_c(\text{MP2})$	-41.44	-41.65	-41.64	-41.65	-41.65
$E_c < 5 \times 10^{-5}$	-0.21	-0.20	-0.21	-0.21	$E_c(\text{CCSD})$
	-39.65	-39.86	-39.85	-39.86	-39.86
$E_c$	$< 5 \times 10^{-5}$	-0.21	-0.20	-0.21	-0.21
$\text{I}^- E_c(\text{MP2})$	-48.54	-49.07	-49.05	-49.06	-49.06
$E_c < 3 \times 10^{-4}$	-0.53	-0.51	-0.52	-0.52	$E_c(\text{CCSD})$
	-46.13	-46.65	-46.64	-46.65	-46.65
$E_c$	$< 3 \times 10^{-4}$	-0.52	-0.51	-0.52	-0.52
$\text{At}^- E_c(\text{MP2})$	-90.95	-92.90	-92.83	-92.86	-92.87
$E_c < 3 \times 10^{-3}$	-1.95	-1.88	-1.91	-1.92	$E_c(\text{CCSD})$
	-84.30	-86.24	-86.17	-86.20	-86.20
$E_c$	$< 3 \times 10^{-3}$	-1.94	-1.87	-1.90	-1.91

is done before the SCF step (X2C-AMFI and X2C-G-AMFI), as opposed to using the full atomic or molecular potential obtained the SCF step as in the  ${}^2\text{DCG}^{\text{M}}$  approach. We see that the latter qualitatively follow the changes in binding energy seen for  ${}^2\text{DCG}^{\text{M}}$  as we move across the rows, though with significant numerical differences for the K edge (with differences of over 40 eV for Iodine and Xenon, and over 200 eV for Astatine) as well as for the L and M edges.

Finally, we assess the effect of including the (SS|SS) integrals in  ${}^2\text{DC}^{\text{M}}$  and  ${}^2\text{DCG}^{\text{M}}$  calculations. Since these contributions are generally very small for elements before the fourth row, and they increase significantly the cost of the SCF step, we have only investigated the fourth and fifth row atoms (I, Xe and At).

For  $\text{I}^-$  and Xe, these contributions are relatively modest, with reductions of around 2.6 eV and 3 eV to their K edge binding energies, with other significant reductions for the L (0.4-0.7 eV) and  $\text{M}_{1,2}$  edges (0.1 eV). We note that there is little variation from these atomic values for the molecular systems (see dataset<sup>122</sup>), in line with the much more localized nature of the small component density compared to the large component one. Furthermore, we notice a very subtle difference (0.02 eV) between the  ${}^2\text{DC}^{\text{M}}$  and  ${}^2\text{DCG}^{\text{M}}$  Hamiltonians with (SS|SS) integrals.

For  $\text{At}^-$  the reductions in binding energies are much

TABLE II: Contributions (in eV) to the core binding energies of  $I^-$ , Xe and  $At^-$  from: (a) the Gaunt interaction and (b) the (SS|SS) integral. These contributions are calculated as energy difference between CVS-EOM-CCSD calculations employing (a) the  ${}^2DCG^M$  and  ${}^2DC^M$  Hamiltonians; and (b) the  ${}^2DCG^M$  without and with inclusion of the (SS|SS) integrals at the SCF step.

Edge	$I^-$		Xe		$At^-$	
	E(a)	E(b)	E(a)	E(b)	E(a)	E(b)
K	-81.57	-2.58	-86.75	-2.87	-396.04	-40.50
L <sub>1</sub>	-7.59	-0.39	-8.15	-0.44	-46.24	-7.28
L <sub>2</sub>	-13.64	-0.58	-14.62	-0.65	-78.99	-10.46
L <sub>3</sub>	-9.70	-0.21	-10.39	-0.23	-51.77	-3.56
M <sub>1</sub>	-1.05	-0.08	-1.14	-0.09	-8.72	-1.79
M <sub>2</sub>	-2.08	-0.21	-2.25	-0.12	-15.19	-2.35
M <sub>3</sub>	-1.34	-0.08	-1.46	-0.04	-9.51	-0.89
M <sub>4</sub>	-0.59	-0.06	-0.66	-0.03	-6.62	-0.82
M <sub>5</sub>	-0.26	-0.04	-0.31	0.01	-4.52	-0.09
N <sub>1</sub>					-1.81	-0.44
N <sub>2</sub>					-3.29	-0.55
N <sub>3</sub>					-1.81	-0.19
N <sub>4</sub>					-0.91	-0.10
N <sub>5</sub>					-0.42	0.02
N <sub>6</sub>					6.74	0.06
N <sub>7</sub>					0.48	0.12

more significant: around 40 eV for the K edge, 10 eV for the L<sub>2</sub> edge, and well above 1 eV for the L<sub>3</sub> to M<sub>2</sub> edges. Interestingly, these contributions remain around 0.5 eV for the N edges relating to ionizations from s and p spinors, which are comparable to the energies of the M<sub>1</sub> and M<sub>2</sub> edges for  $I^-$  and Xe, which fall between 0.7 and 1 keV.

These results underline the need for explicitly accounting for these integrals (or correcting the energies for their contribution), as soon as we are interested in edges arising from ionizations of s and p spinors for fifth-row elements and beyond.

#### D. Binding energies: comparison to experiment and prior theoretical works

We now compare our results for the  ${}^2DC^M$  and  ${}^2DCG^M$  Hamiltonians, including the (SS|SS) integrals, to experiment and other theoretical results for the K and L edges of Xe, XeF<sub>2</sub> and CH<sub>3</sub>I, for which recent high energy, gas-phase XPS experiments have been performed (for Xe and XeF<sub>2</sub>, results are also available for less energetic edges). Our results are found in Table III.

As all our calculations lack contributions from QED effects, and that the  ${}^2DCG^M$  does not include the gauge term that is necessary to recover the full (zero-frequency) Breit interaction, we correct our energies with the results from Koziol and Aucar,<sup>125</sup> who provide the values for the Breit and leading QED contributions (self interaction and vacuum polarization) to the atomic spinors of

selected closed-shell atoms (among which Xe). The same procedure was followed by Southworth et al.<sup>126</sup> to correct their calculations, which are based on a combination of 1-component CVS-EOM-CCSD (and CCSDT) calculations with energy estimates for the different effects (1- and 2-electron scalar relativity, spin-orbit coupling, nuclear size effects).

Taking the Xe atom, the QED effects are most important for the K and L<sub>1</sub> edges, lowering the associated binding energies by 42.7 and 5.5 eV, respectively. They are also non-negligible for the L<sub>3</sub> and M<sub>1</sub> edges, which are lowered respectively by 0.5 and 1.1 eV. The Breit interaction also lowers the binding energies and, apart from the L<sub>1</sub> edge (for which both are of the same magnitude) is in general much larger than the QED effects (80.7, 13.1 and 8.8 eV for the K, L<sub>2</sub> and L<sub>3</sub> edges, respectively) and still important for the M<sub>3</sub> edge (1.2 eV).

Since Koziol and Aucar only provide the full Breit term, we have estimated the magnitude of the gauge correction from the difference between our  ${}^2DCG^M$  results and their Breit values. From that, we have that the gauge term increases the binding energies (6.04 eV for the K edge, 0.51 eV for the L<sub>1</sub> edge, and around 1.57 and 1.58 eV for the L<sub>2</sub> and L<sub>3</sub> edges, respectively).

We observe our corrected 2-component results differ from experiment by around 2.1 eV, and if we employ the estimate for higher-order correlation contributions to EOM-CCSD from Southworth et al.,<sup>126</sup> which decreases the binding energies by 3.8 eV, the difference to experiment is now of -1.71 eV.

We have also investigated the use of uncontracted ANO-RCC basis, used by Southworth et al.<sup>127</sup>, which is slightly smaller for s and p primitive sets than the Dyall sets. We obtain differences with respect to experiment of -1.69 eV, which is consistent with our results using the Dyall basis sets. We attribute most of the small differences between ours and prior results to the shortcomings in the treatment of the two-electron interactions in  ${}^2DC^M$  or  ${}^2DCG^M$  discussed previously, as calculations with  ${}^4DC$  yield results in very good agreement with those of Southworth et al.

Beyond the K edge, our corrected calculations compare rather well to the recent experimental results of Oura et al.<sup>128</sup> (5452.7 eV, 5106.7 eV and 4786.7 eV; measurements carried out at BL29XU of SPring-8 in May 2016) with the exception of the L<sub>1</sub> edge, for which a larger discrepancy to experiment (7.79 eV) is observed. For the M edges, Southworth et al.<sup>127</sup> present experimental results for the M<sub>4,5</sub> edges, and our corrected calculations differ from experiment by -0.94 eV and -0.85 eV respectively. For the M<sub>4</sub> edge these are quite comparable to the theoretical calculations of Southworth and coworkers,<sup>127</sup> now also corrected for QED effects (-0.92 eV to experiment), while for the M<sub>5</sub> edge both theoretical results differ by around 0.24 eV.

Given the values for the higher-order correlation effects for the K and M<sub>4,5</sub> edges, and from the breakdown of relativistic, correlation and QED effects from

atomic many-body calculations on Xe at the K, L and M edges<sup>131</sup>—which indicate non-negligible differential correlation, relaxation and other effects (for example Auger shifts) for the different L edges, and to a lesser extent for the M edges—we consider future attempts to investigate higher-order electron correlation corrections for the L edges to be of significant interest.

For XeF<sub>2</sub>, applying the same corrections as above to our <sup>2</sup>DC<sup>M</sup> and <sup>2</sup>DCG<sup>M</sup> calculations in the Dyall basis sets, we arrive at K and M<sub>4,5</sub> edges edge binding energies differing by roughly −0.4 eV from experiment. These differences are smaller but consistent with those obtained for Xe, underscoring the largely atomic nature of these deep core energies. As was the case for the atom, a comparison to <sup>4</sup>DC results for the K edge indicate that part of the small differences between the results of Southworth et al. and our 2-component ones come from the shortcomings in the treatment of two-electron interactions.

For CH<sub>3</sub>I, we have only performed <sup>2</sup>DCG<sup>M</sup> calculations, first because, as illustrated above, there are no significant differences between <sup>2</sup>DCG<sup>M</sup> and <sup>2</sup>DC<sup>M</sup> based results once we account for QED and Breit or gauge contributions. Second, in the C<sub>s</sub> point group used, the coupled-cluster wavefunctions are complex-valued, making calculations computationally more expensive. We obtain 33213.79 eV and 5208.55 eV for the K and L<sub>1</sub> edge binding energies, values that overestimate the experimental values by 38.59 eV and 11.08 eV. For the K edge, this difference is rather close to the one found for the Xe and XeF<sub>2</sub> species.

As QED and Breit corrections are not provided by Koziol and Aucar<sup>125</sup> for iodine, we have used instead those of Boudjemia et al.,<sup>130</sup> which amount, for the K edge, to −77.0 eV for Breit, +6.10 eV for the gauge term and −39.0 eV for QED, and to −7.20 eV, +0.50 eV and −5.10 eV for the L<sub>1</sub> edge. With these corrections, we now overestimate the K and L<sub>1</sub> binding energies by 5.69 eV and 1.95 eV. The missing effect would be that of higher-order correlation corrections. While we cannot estimate this here, we speculate that if it follows roughly what is found for the Xe species, that would decrease the binding energies, and likely take the K edge to a few eV.

#### E. Excitation energies: comparisons to experiment and prior theoretical works

As anticipated, the current implementation allows calculation not only of ionization energies, but also of excitation energies, which will be briefly discussed hereafter. In this case, since (SS|SS) integrals do not significantly affect the energies, only results for the <sup>2</sup>DCG<sup>M</sup> and <sup>2</sup>DC<sup>M</sup> Hamiltonian are presented, without explicit inclusion of the aforementioned integrals.

TABLE III: Comparison between calculated and experimental gas-phase binding energies (in eV) for Xe, XeF<sub>2</sub> and CH<sub>3</sub>I. Calculations are broken down into values obtained with (a) Koopmans theorem, (b) CVS-EOM-CCSD method; (c) CVS-EOM-CCSD with atomic corrections for QED and Breit (in the case of <sup>2</sup>DCG<sup>M</sup> results, corrections for the gauge term) interactions; and (d) higher-order correlation corrections by Southworth and coworkers on (c). All of our results include contributions from the (SS|SS) integrals. Basis sets A: ANO-RCC, D: Dyall

Edge model	E(a)	E(b)	E(c)	E(d)	
Xe					
K	D <sup>4</sup> DC	34755.91	34690.82	34567.44	34563.64
	D <sup>2</sup> DC <sup>M</sup>	34755.91	34690.60	34567.22	34563.42
	D <sup>2</sup> DCG <sup>M</sup>	34669.08	34603.85	34567.22	34563.42
	A <sup>4</sup> DC	34755.89	34690.63	34567.48	34563.68
	A <sup>2</sup> DC <sup>M</sup>	34755.89	34690.63	34567.24	34563.44
	A/ref. <sup>127</sup> exp. <sup>127</sup>	34752.00	34690.90	34567.51	34563.71 34565.13
L <sub>1</sub>	D <sup>2</sup> DC <sup>M</sup>	5509.35	5473.66	5460.49	
	D <sup>2</sup> DCG <sup>M</sup> exp. <sup>128</sup>	5501.17	5465.51	5460.49	5452.7
L <sub>2</sub>	D <sup>2</sup> DC <sup>M</sup>	5161.45	5122.38	5109.30	
	D <sup>2</sup> DCG <sup>M</sup> exp. <sup>128</sup>	5146.77	5107.77	5109.30	5106.7
L <sub>3</sub>	D <sup>2</sup> DC <sup>M</sup>	4835.59	4796.85	4787.51	
	D <sup>2</sup> DCG <sup>M</sup> exp. <sup>128</sup>	4825.15	4786.46	4784.51	4786.7
M <sub>4</sub>	D <sup>2</sup> DC <sup>M</sup>	708.13	689.63	689.01	688.31
	D <sup>2</sup> DCG <sup>M</sup>	707.46	688.98	689.01	688.31
	A/ref. <sup>127</sup> exp. <sup>127</sup>	707.5	689.6	688.98	688.28 689.23
M <sub>5</sub>	D <sup>2</sup> DC <sup>M</sup>	694.90	676.71	676.32	675.62
	D <sup>2</sup> DCG <sup>M</sup>	694.58	676.30	676.32	675.62
	A/ref. <sup>127</sup> exp. <sup>127</sup>	694.6	676.7	676.08	675.38 676.44
XeF <sub>2</sub>					
K	D <sup>4</sup> DC	34759.79	34694.48	34567.08	34566.18
	D <sup>2</sup> DC <sup>M</sup>	34759.79	34693.30	34569.90	34566.00
	A <sup>2</sup> DC <sup>M</sup>	34759.76	34694.28	34570.89	34566.99
	A/ref. <sup>127</sup> exp. <sup>127</sup>	34755.80	34694.50	34571.10	34567.20 34567.4
M <sub>4</sub>	D <sup>2</sup> DC <sup>M</sup>	711.99	693.23	692.61	691.61
	A <sup>2</sup> DC <sup>M</sup>	711.74	693.32	692.70	691.70
	A/ref. <sup>127</sup> exp. <sup>129</sup>	711.9	693.9	693.28	692.28 692.09
M <sub>5</sub>	D <sup>2</sup> DC <sup>M</sup>	698.49	680.22	679.93	678.93
	A <sup>2</sup> DC <sup>M</sup>	698.46	680.30	680.01	679.01
	A/ref. <sup>127</sup> exp. <sup>129</sup>	698.6	680.6	680.31	679.31 679.31
CH <sub>3</sub> I					
K	D <sup>2</sup> DCG <sup>M</sup> exp. <sup>130</sup>	33278.46	33213.79	33180.89	33175.20
L <sub>1</sub>	D <sup>2</sup> DCG <sup>M</sup> exp. <sup>130</sup>	5244.04	5208.55	5203.95	5197.47

The results are shown in Table IV, which displays selected excitation energies from different edges to the LUMO. We chose these particular transitions since both experimental and computational data are available in the literature,<sup>126,127</sup> which can be used as reference values. In particular, the experimental values are tabulated with the theoretical energies obtained in this study.

As already observed for the ionization processes, inclusion of the Gaunt interaction, accounting for the magnetic interaction between the electrons, lowers the excitation energy. This was expected as, in general, the inclusion of this term shifts the orbital energies of the inner core orbitals up (or lowers them, in term of absolute energy) and, at the same time, it reduces the spin-orbit coupling,<sup>62</sup> which is also reflected by the results reported herein. Indeed, the difference in the excitation energy from 3d<sub>3/2</sub> and 3d<sub>5/2</sub> to LUMO is slightly higher when the <sup>2</sup>DC<sup>M</sup> Hamiltonian is employed.

Upon inclusion of the Gaunt term in the Hamiltonian, the excitation energies differ by at most 0.4% from the corresponding experimental value, specifically for 1s(F). This mismatch is higher than the others, which can be due to the fact that the CVS space was reduced so that this was the highest orbital, necessary in order to get this excitation.

Thus, this space excludes the 3d orbitals and in particular 3d<sub>3/2</sub>. The next higher disagreement between theory and experiment is found when exciting the 1s(Xe), although by only 0.1%. Finally, the best agreement is found for both 3d orbitals, with an error of only 0.04% (3d<sub>3/2</sub>) and 0.02% (3d<sub>5/2</sub>) from experiment. The computed energies are thus within the experimental error. In order to obtain the excitation from the (3d<sub>3/2</sub>) orbital, the same strategy of excluding all higher orbitals from the CVS space, hence excluding (3d<sub>5/2</sub>), might be the reason of the subtly larger disagreement.

Nonetheless, these errors are actually quite small. Therefore, we can conclude that the CVS-EOM-CC method implemented in this work gives satisfactory core excitation energies.

## V. CONCLUSIONS

We have presented an implementation of the core-valence separation for the equation-of-motion coupled-cluster method in the Dirac program. This implementation, which is based on a flexible framework for defining projection operators, enables the calculation ionization and excitation energies for all 4-component based Hamiltonians available in Dirac, and consequently for non-relativistic Hamiltonians as well. We have applied our implementation to the calculation of core electron binding energies for halogen (CH<sub>3</sub>I, X<sup>-</sup>, and HX, X = Cl–At) and Xe species (Xe, XeF<sub>2</sub>). For the latter, we also briefly explored the calculation of core excitations. With these systems, we have investigated the performance of different approximations to the original CVS approach, the

TABLE IV: XeF<sub>2</sub>. Comparison between experimental gas-phase excitation energies and calculations (in eV). The calculated excitation energies have been obtained at the CVS-EOM-CCSD level of theory using the <sup>2</sup>DC<sup>M</sup> and the <sup>2</sup>DCG<sup>M</sup> Hamiltonians without contributions from the (SS|SS) integrals and the Dyall (D) basis set

Edge transition	model	E
K 1s (Xe) → LUMO	D/ <sup>2</sup> DC <sup>M</sup>	34686.70
	D/ DCG	34599.95
	exp. <sup>126</sup>	34557.4
1s (F) → LUMO	D/ <sup>2</sup> DC <sup>M</sup>	685.91
	D/ DCG	685.70
	exp. <sup>127</sup>	682.8 ± 0.3
M <sub>4</sub> 3d <sub>3/2</sub> → LUMO	D/ <sup>2</sup> DC <sup>M</sup>	683.71
	D/ DCG	683.06
	exp. <sup>127</sup>	682.8 ± 0.3
M <sub>5</sub> 3d <sub>5/2</sub> → LUMO	D/ <sup>2</sup> DC <sup>M</sup>	670.34
	D/ DCG	670.04
	exp. <sup>127</sup>	669.9 ± 0.3

basis set effects, and the performance of different classes of 2-component approximations.

For highly symmetric species such as the lighter halides, for which we are able to exactly diagonalize the original similarity-transformed Hamiltonian, we show that the CVS energies closely match those from exact diagonalization at all edges. For the overall test set, our assessment of the different approximations indicates the one which more closely matches the performance of the original CVS scheme employs both frozen core and the removal of doubly excited determinants containing only core occupied spinors. Taken individually, these approximations yield sizeable overestimations and underestimations to the core ionization energies, respectively.

With respect to the Hamiltonians, we observe first that calculations in which the transformation to 2-component is performed after the SCF step (<sup>2</sup>DC<sup>M</sup>) are nearly indistinguishable from the equivalent 4-component ones (<sup>4</sup>DC), though non-negligible discrepancies appear at the K edge of fifth and sixth-row elements. We have traced these discrepancies to the use of the uncorrected 2-electron operator in the 2-component molecular mean-field scheme (<sup>2</sup>DC<sup>M</sup> and <sup>2</sup>DCG<sup>M</sup>). More approximate approaches in which the transformation to 2-component is carried out before the SCF step (X2C-AMFI and X2C-G-AMFI), on the other hand, offer at best qualitative accuracy. Second, our results underscore the importance of explicitly considering SSSS-type integrals, in particular from the fifth row onwards.

A comparison to experimental results for Xe, XeF<sub>2</sub> and CH<sub>3</sub>I underscores the importance of the QED, Breit and higher order correlation effects to approach experimental results. While our calculations including the Gaunt interaction recover a significant fraction of the Breit interaction, the gauge term remains quite significant for the K and L edges, and it must be accounted for.

In view of these findings, we consider the 4-component

based CVS-EOM-CCSD as a reliable approach for investigating core properties throughout the periodic table. Apart from its intrinsic interest, it may serve as a basis for further investigations of the reliability of more ap-proximate schemes for atoms beyond the fifth row, as well as to verify whether such approximations result in significant changes for properties such as transition mo-ments, that require the determination of the excited-state wavefunctions.

## VI. ACKNOWLEDGEMENTS

ASPG and LH acknowledge support from PIA ANR project CaPPA (ANR-11-LABX-0005-01), the Franco-German project CompRIXS (Agence nationale de la recherche ANR-19-CE29-0019, Deutsche Forschungsge-meinschaft JA 2329/6-1), I-SITE ULNE projects OVER-SEE and MESONM International Associated Labo-ratory (LAI) (ANR-16-IDEX-0004), the French Min-istry of Higher Education and Research, region Hauts de France council and European Regional Develop-ment Fund (ERDF) project CPER CLIMIBIO, and the French national supercomputing facilities (grants DARI A0050801859, A0070801859), as well as discussions with Trond Saue at the 2020 Dirac Developers' meeting and following the appearance of the manuscript as a preprint. MLV and SC acknowledge support from DTU Chemistry (start-up Ph.D. grant), from the Independent Research Fund Denmark–Natural Sciences, Research Project 2, grant no. 7014-00258B, and from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie European Training Network Computational Spectroscopy In Natural sciences and En-gineering (COSINE), grant agreement no. 765739. Sup-port from the COST action CM1405 MOLIM: Molecules in Motion (short-term mobility grant to MLV to visit ASPG) is also acknowledged.

- 1 J. Stöhr, NEXAFS Spectroscopy (Springer, Berlin, 1992).
- 2 P. Norman and A. Dreuw, *Chem. Rev.* **118**, 7208 (2018).
- 3 J. van Bokhoven and C. Lamberti, eds., *X-Ray Absorption and X-ray Emission Spectroscopy; Theory and Applications* (Wiley & Sons, 2016).
- 4 S. Mobilio, F. Boscherini, and C. Meneghini, eds., *Synchrotron Radiation: Basics, Methods and Applications* (Springer, 2014).
- 5 U. Bergmann, V. Yachandra, and J. Yano, eds., *X-Ray Free Electron Lasers: Applications in Materials, Chemistry and Bi-ology, Energy and Environment Series No. 18* (Royal Society of Chemistry, 2017).
- 6 M. Nisoli, P. Decleva, F. Calegari, A. Palacios, and F. Martín, *Chem. Rev.* **117**, 10760 (2017).
- 7 C. J. Milne, T. J. Penfold, and M. Chergui, *Coord. Chem. Rev.* **277-278**, 44 (2014).
- 8 P. S. Bagus, *Phys. Rev.* **139**, A619 (1965).
- 9 P. S. Bagus and H. F. Schaefer, *J. Chem. Phys.* **55**, 1474 (1971), <https://doi.org/10.1063/1.1676248>.
- 10 N. A. Besley, *WIREs Comput. Mol. Sci.* n/a, e1527 (2021), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1527>.
- 11 C. South, A. Shee, D. Mukherjee, A. K. Wilson, and T. Saue, *Phys. Chem. Chem. Phys.* **18**, 21010 (2016).
- 12 X. Zheng and L. Cheng, *J. Chem. Theory Comput.* **15**, 4945 (2019), <https://doi.org/10.1021/acs.jctc.9b00568>.
- 13 J. Lee, D. W. Small, and M. Head-Gordon, *J. Chem. Phys.* **151**, 214103 (2019), <https://doi.org/10.1063/1.5128795>.
- 14 N. A. Besley, A. T. B. Gilbert, and P. M. W. Gill, *J. Chem. Phys.* **130**, 124308 (2009), <https://doi.org/10.1063/1.3092928>.
- 15 Y. Liang, J. Vinson, S. Pemmaraju, W. S. Drisdell, E. L. Shirley, and D. Prendergast, *Phys. Rev. Lett.* **118**, 096402 (2017).
- 16 D. Hait and M. Head-Gordon, *J. Phys. Chem. Lett.* **11**, 775 (2020), <https://doi.org/10.1021/acs.jpclett.9b03661>.
- 17 C. Ehlert and T. Klamroth, *J. Comput. Chem.* **41**, 1781 (2020), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.26219>.
- 18 W. D. Derricotte and F. A. Evangelista, *Phys. Chem. Chem. Phys.* **17**, 14360 (2015).
- 19 P. Verma, W. D. Derricotte, and F. A. Evangelista, *J. Chem. Theory Comput.* **12**, 144 (2016), pMID: 26584082, <https://doi.org/10.1021/acs.jctc.5b00817>.
- 20 W. J. Hunt and W. A. Goddard, *Chemical Physics Letters* **3**, 414 (1969).
- 21 H. Agren, V. Carravetta, O. Vahtras, and L. G. M. Pettersson, *Chem. Phys. Lett.* **222**, 75 (1994).
- 22 N. Ferr' and X. Assfeld, *J. Chem. Phys.* **117**, 4119 (2002), <https://doi.org/10.1063/1.1496462>.
- 23 M. Lundberg and M. G. Delcey, "Multiconfigurational approach to x-ray spectroscopy of transition metal complexes," in *Transition Metals in Coordination Environments: Computational Chemistry and Catalysis Viewpoints*, edited by E. Broclawik, T. Borowski, and M. Radoń (Springer International Publishing, Cham, 2019) pp. 185–217.
- 24 S. I. Bokarev and O. Kühn, *WIREs Comput. Mol. Sci.* **10**, e1433 (2020), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1433>.
- 25 R. Klooster, R. Broer, and M. Filatov, *Chem. Phys.* **395**, 122 (2012).
- 26 D. Maganas, J. K. Kowalska, M. Nooijen, S. DeBeer, and F. Neese, *J. Chem. Phys.* **150**, 104106 (2019), <https://doi.org/10.1063/1.5051613>.
- 27 S. Coriani, O. Christiansen, T. Fransson, and P. Norman, *Phys. Rev. A* **85**, 022507 (2012).
- 28 S. Coriani and H. Koch, *J. Chem. Phys.* **143**, 181103 (2015).
- 29 S. Coriani and H. Koch, *J. Chem. Phys.* **145**, 149901 (2016), <https://doi.org/10.1063/1.4964714>.
- 30 A. Sadybekov and A. I. Krylov, *J. Chem. Phys.* **147**, 014107 (2017).
- 31 M. L. Vidal, X. Feng, E. Epifanovsky, A. I. Krylov, and S. Coriani, *J. Chem. Theory Comput.* **15**, 3117 (2019), <https://doi.org/10.1021/acs.jctc.9b00039>.
- 32 B. Peng, P. J. Lestrangle, J. J. Goings, M. Caricato, and X. Li, *J. Chem. Theory Comput.* **11**, 4146 (2015).
- 33 A. C. Paul, R. H. Myhre, and H. Koch, *Journal of Chemical Theory and Computation* **17**, 117 (2021), pMID: 33263255, <https://doi.org/10.1021/acs.jctc.0c00686>.
- 34 Y. C. Park, A. Perera, and R. J. Bartlett, *J. Chem. Phys.* **151**, 164117 (2019), <https://doi.org/10.1063/1.5117841>.
- 35 D. A. Matthews, *Mol. Phys.* **118**, e1771448 (2020), <https://doi.org/10.1080/00268976.2020.1771448>.
- 36 J. Brabec, K. Bhaskaran-Nair, N. Govind, J. Pittner, and K. Kowalski, *J. Chem. Phys.* **137**, 171101 (2012).
- 37 B. Helmich-Paris, *Int. J. Quantum Chem.* **121**, e26559 (2021), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.26559>.
- 38 J. Schirmer and A. Thiel, *J. Chem. Phys.* **115**, 10621 (2001).
- 39 J. Wenzel, M. Wormit, and A. Dreuw, *J. Comput. Chem.* **35**, 1900 (2014).
- 40 A. Dreuw and M. Wormit, *WIREs Comput. Mol. Sci.* **5**, 82 (2015).
- 41 S. P. Neville, V. Averbukh, S. Patchkovskii, M. Ruberti, R. Yun, M. Chergui, A. Stolow, and M. S. Schuurman, *Faraday Discuss.* **194**, 117 (2016).
- 42 U. Ekström, P. Norman, V. Carravetta, and H. Agren, *Phys. Rev. Lett.* **97**, 143001 (2006).
- 43 N. A. Besley, *Acc. Chem. Res.* **53**, 1306 (2020), <https://doi.org/10.1021/acs.accounts.0c00171>.

- 44 J. Stanton and R. Bartlett, *J. Chem. Phys.* **98**, 7029 (1993).
- 45 K. Sneskov and O. Christiansen, *WIREs Comput. Mol. Sci.* **2**, 566 (2012).
- 46 R. J. Bartlett, *WIREs Comput Mol Sci* **2**, 126 (2012).
- 47 A. I. Krylov, *Ann. Rev. Phys. Chem.* **59**, 433 (2008).
- 48 O. Christiansen, P. Jørgensen, and C. Hättig, *Int. J. Quantum Chem.* **98**, 1 (1998).
- 49 H. Koch and P. Jørgensen, *J. Chem. Phys.* **93**, 3333 (1990).
- 50 S. Coriani, F. Pawłowski, J. Olsen, and P. Jørgensen, *J. Chem. Phys.* **144**, 024102 (2016).
- 51 D. Zuev, E. Vecharynski, C. Yang, N. Orms, and A. I. Krylov, *J. Comp. Chem.* **36**, 273 (2015).
- 52 C. Huang, W. Liu, Y. Xiao, and M. R. Hoffmann, *J. Comput. Chem.* **38**, 2481 (2017), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.24907>.
- 53 C. Huang and W. Liu, *J. Comput. Chem.* **40**, 1023 (2019), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.25569>.
- 54 M. Stener, G. Fronzoni, and M. de Simone, *Chem. Phys. Lett.* **373**, 115 (2003).
- 55 B. N. C. Tenorio, T. Moitra, M. A. C. Nascimento, A. B. Rocha, and S. Coriani, *J. Chem. Phys.* **150**, 224104 (2019), <https://doi.org/10.1063/1.5096777>.
- 56 I. Seidu, S. P. Neville, M. Kleinschmidt, A. Heil, C. M. Marian, and M. S. Schuurman, *J. Chem. Phys.* **151**, 144104 (2019), <https://doi.org/10.1063/1.5110418>.
- 57 M. G. Delcey, L. K. Sørensen, M. Vacher, R. C. Couto, and M. Lundberg, *J. Comp. Chem.* **40**, 1789 (2019), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.25832>.
- 58 L. S. Cederbaum, W. Domcke, and J. Schirmer, *Phys. Rev. A: At. Mol. Opt. Phys.* **22**, 206 (1980).
- 59 R. Faber and S. Coriani, *Phys. Chem. Chem. Phys.* **22**, 2642 (2020).
- 60 K. D. Nanda, M. L. Vidal, R. Faber, S. Coriani, and A. I. Krylov, *Phys. Chem. Chem. Phys.* **22**, 2629 (2020).
- 61 K. G. Dyall and K. Faegri Jr, *Introduction to Relativistic Quantum Chemistry* (Oxford University Press, 2007).
- 62 T. Saue, *ChemPhysChem* **12**, 3077 (2011).
- 63 W. Liu, *J. Chem. Phys.* **152**, 180901 (2020), <https://doi.org/10.1063/5.0008432>.
- 64 J. P. Carbone, L. Cheng, R. H. Myhre, D. Matthews, H. Koch, and S. Coriani, in *State of The Art of Molecular Electronic Structure Computations: Correlation Methods, Basis Sets and More*, Advances in Quantum Chemistry, Vol. 79, edited by L. U. Ancarani and P. E. Hoggan (Academic Press, 2019) pp. 241–261.
- 65 J. Liu, D. Matthews, S. Coriani, and L. Cheng, *J. Chem. Theory Comput.* **15**, 1642 (2019), <https://doi.org/10.1021/acs.jctc.8b01160>.
- 66 M. L. Vidal, P. Pokhilko, A. I. Krylov, and S. Coriani, *J. Phys. Chem. Lett.* **11**, 8314 (2020), <https://doi.org/10.1021/acs.jpcclett.0c02027>.
- 67 L. Keller, V. Blum, P. Rinke, and D. Golze, *J. Chem. Phys.* **153**, 114110 (2020), <https://doi.org/10.1063/5.0018231>.
- 68 C. Fillaux, C. Den Auwer, D. Guillaumont, D. K. Shuh, and T. Tyliszczak, *J. Alloys Compd.* **444–445**, 443 (2007).
- 69 P. S. Bagus, E. S. Ilton, and C. J. Nelin, *Surf. Sci. Rep.* **68**, 273 (2013).
- 70 M. Repisky, S. Komorovsky, M. Kadek, L. Konecny, U. Ekström, E. Malkin, M. Kaupp, K. Ruud, O. L. Malkina, and V. G. Malkin, *J. Chem. Phys.* **152**, 184101 (2020).
- 71 L. Belpassi, M. De Santis, H. M. Quiney, F. Tarantelli, and L. Storchi, *J. Chem. Phys.* **152**, 164118 (2020), <https://doi.org/10.1063/5.0002831>.
- 72 Y. Zhang, B. Suo, Z. Wang, N. Zhang, Z. Li, Y. Lei, W. Zou, J. Gao, D. Peng, Z. Pu, Y. Xiao, Q. Sun, F. Wang, Y. Ma, X. Wang, Y. Guo, and W. Liu, *J. Chem. Phys.* **152**, 064113 (2020), <https://doi.org/10.1063/1.5143173>.
- 73 W. Kutzelnigg and W. Liu, *J. Chem. Phys.* **123**, 241102 (2005), <https://doi.org/10.1063/1.2137315>.
- 74 H. J. Aa. Jensen, (2005), Douglas–Kroll the Easy Way, Talk at Conference on Relativistic Effects in Heavy Elements - REHE, Mülheim, Germany, April, 2005. Available at <https://doi.org/10.6084/m9.figshare.12046158>.
- 75 W. Liu and D. Peng, *J. Chem. Phys.* **125**, 044102 (2006), <https://doi.org/10.1063/1.2222365>.
- 76 M. Iliaş and T. Saue, *J. Chem. Phys.* **126**, 064102 (2007).
- 77 D. Peng, W. Liu, Y. Xiao, and L. Cheng, *J. Chem. Phys.* **127**, 104106 (2007), <https://doi.org/10.1063/1.2772856>.
- 78 W. Liu and D. Peng, *J. Chem. Phys.* **131**, 031104 (2009), <https://doi.org/10.1063/1.3159445>.
- 79 J. Sikkema, L. Visscher, T. Saue, and M. Iliaş, *J. Chem. Phys.* **131**, 124116 (2009), <https://doi.org/10.1063/1.3239505>.
- 80 J. Liu, Y. Shen, A. Asthana, and L. Cheng, *J. Chem. Phys.* **148**, 034106 (2018), <https://doi.org/10.1063/1.5009177>.
- 81 J. Liu and L. Cheng, *J. Chem. Phys.* **148**, 144108 (2018), <https://doi.org/10.1063/1.5023750>.
- 82 L. Cheng, *J. Chem. Phys.* **151**, 104103 (2019), <https://doi.org/10.1063/1.5113796>.
- 83 A. Asthana, J. Liu, and L. Cheng, *J. Chem. Phys.* **150**, 074102 (2019), <https://doi.org/10.1063/1.5081715>.
- 84 A. Shee, T. Saue, L. Visscher, and A. S. P. Gomes, *J. Chem. Phys.* **149**, 174113 (2018).
- 85 T. Saue, R. Bast, A. S. P. Gomes, H. J. A. Jensen, L. Visscher, I. A. Aucar, R. Di Remigio, K. G. Dyall, E. Eliav, E. Fasshauer, T. Fleig, L. Halbert, E. D. Hedegård, B. Helmich-Paris, M. Iliaş, C. R. Jacob, S. Knecht, J. K. Laerdahl, M. L. Vidal, M. K. Nayak, M. Olejniczak, J. M. H. Olsen, M. Pernpointner, B. Sen-jean, A. Shee, A. Sunaga, and J. N. P. van Stralen, *J. Chem. Phys.* **152**, 204104 (2020), <https://doi.org/10.1063/5.0004844>.
- 86 H. Pathak, S. Sasmal, M. K. Nayak, N. Vaval, and S. Pal, *Comput. Theor. Chem.* **1076**, 94 (2016).
- 87 H. Pathak, B. K. Sahoo, T. Sengupta, B. P. Das, N. Vaval, and S. Pal, *J. Phys. B: Atom. Mol. Opt. Phys.* **48**, 115009 (2015).
- 88 H. Pathak, S. Sasmal, M. K. Nayak, N. Vaval, and S. Pal, *Phys. Rev. A* **90**, 062501 (2014).
- 89 H. Pathak, B. K. Sahoo, B. P. Das, N. Vaval, and S. Pal, *Phys. Rev. A* **89**, 042510 (2014).
- 90 H. Pathak, S. Sasmal, M. K. Nayak, N. Vaval, and S. Pal, *J. Chem. Phys.* **145**, 074110 (2016).
- 91 K. Klein and J. Gauss, *J. Chem. Phys.* **129**, 194106 (2008).
- 92 D.-D. Yang, F. Wang, and J. Guo, *Chem. Phys. Lett.* **531**, 236 (2012).
- 93 Z. Wang, S. Hu, F. Wang, and J. Guo, *J. Chem. Phys.* **142**, 144109 (2015).
- 94 E. Epifanovsky, K. Klein, S. Stopkowicz, J. Gauss, and A. I. Krylov, *J. Chem. Phys.* **143**, 064102 (2015).
- 95 Z. Cao, F. Wang, and M. Yang, *J. Chem. Phys.* **145**, 154110 (2016).
- 96 Z. Cao, Z. Li, F. Wang, and W. Liu, *Phys. Chem. Chem. Phys.* **19**, 3713 (2017).
- 97 S. Zhang and F. Wang, *J. Phys. Chem. A* **121**, 3966 (2017).
- 98 Y. Akinaga and T. Nakajima, *J. Phys. Chem. A* **121**, 827 (2017).
- 99 F. Wang, in *Handbook of Relativistic Quantum Chemistry*, edited by W. Liu (Springer Berlin Heidelberg, Berlin, Heidelberg, 2016) pp. 1–27.
- 100 M. Guo, Z. Wang, and F. Wang, *J. Chem. Phys.* **153**, 214118 (2020), <https://doi.org/10.1063/5.0032716>.
- 101 H. Pathak, S. Sasmal, K. Talukdar, M. K. Nayak, N. Vaval, and S. Pal, *J. Chem. Phys.* **152**, 104302 (2020), <https://doi.org/10.1063/1.5140988>.
- 102 S. D. Folkestad, E. F. Kjørstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, T. Moitra, A. C. Paul, M. Scavino, A. S. Skeidsvoll, A. H. Tveten, and H. Koch, *J. Chem. Phys.* **152**, 184103 (2020), <https://doi.org/10.1063/5.0004713>.
- 103 S. L. Sorensen, X. Zheng, S. H. Southworth, M. Patanen, E. Kokkonen, B. Oostenrijk, O. Travnikova, T. Marchenko, M. Simon, C. Bostedt, G. Doumy, L. Cheng, and L. Young, *Journal of Physics B: Atomic, Molecular and Optical Physics* **53**, 244011 (2020).



- <sup>104</sup> C. D. Rankine and T. J. Penfold, *The Journal of Physical Chemistry A* **0**, null (0), PMID: 33729774, <https://doi.org/10.1021/acs.jpca.0c11267>.
- <sup>105</sup> J. Wenzel, A. Holzer, M. Wormit, and A. Dreuw, *J. Chem. Phys.* **142**, 214104 (2015).
- <sup>106</sup> L. Visscher, T. J. Lee, and K. G. Dyall, *J. Chem. Phys.* **105**, 8769 (1996).
- <sup>107</sup> M. Pernpointner and L. Visscher, *J. Comp. Chem.* **24**, 754 (2003).
- <sup>108</sup> A. Shee, L. Visscher, and T. Saue, *J. Chem. Phys.* **145**, 184107 (2016).
- <sup>109</sup> DIRAC, a relativistic ab initio electronic structure program, Release DIRAC19 (2019), written by A. S. P. Gomes, T. Saue, L. Visscher, H. J. Aa. Jensen, and R. Bast, with contributions from I. A. Aucar, V. Bakken, K. G. Dyall, S. Dubillard, U. Ekström, E. Eliav, T. Enevoldsen, E. Faßhauer, T. Fleig, O. Fossgaard, L. Halbert, E. D. Hedegård, B. Heimlich-Paris, T. Helgaker, J. Henriksson, M. Iliaš, Ch. R. Jacob, S. Knecht, S. Komorovský, O. Kullie, J. K. Lærdahl, C. V. Larsen, Y. S. Lee, H. S. Nataraj, M. K. Nayak, P. Norman, G. Olejniczak, J. Olsen, J. M. H. Olsen, Y. C. Park, J. K. Pedersen, M. Pernpointner, R. di Remigio, K. Ruud, P. Salek, B. Schimmelpfennig, B. Senjean, A. Shee, J. Sikkema, A. J. Thorvaldsen, J. Thyssen, J. van Stralen, M. L. Vidal, S. Villaume, O. Visser, T. Winther, and S. Yamamoto (available at <http://dx.doi.org/10.5281/zenodo.3572669>, see also <http://www.diracprogram.org>).
- <sup>110</sup> K. G. Dyall, *Theor. Chem. Acc.* **108**, 335 (2002).
- <sup>111</sup> K. G. Dyall, *Theor. Chem. Acc.* **109**, 284 (2003).
- <sup>112</sup> K. G. Dyall, *Theor. Chem. Acc.* **135**, 128 (2016).
- <sup>113</sup> B. O. Roos, R. Lindh, P.-A. Malmqvist, V. Veryazov, and P.-O. Widmark, *J. Phys. Chem. A* **108**, 2851 (2004), <https://doi.org/10.1021/jp031064+>.
- <sup>114</sup> W. H. Furry, *Phys. Rev.* **81**, 115 (1951).
- <sup>115</sup> B. A. Hess, C. M. Marian, U. Wahlgren, and O. Gropen, *Chem. Phys. Lett.* **251**, 365 (1996).
- <sup>116</sup> C. M. Marian (Wiley-VCH, 2001) p. 99.
- <sup>117</sup> B. Schimmelpfennig, "AMFI, an atomic mean-field spin-orbit integral program," University of Stockholm, Stockholm, Sweden; 1999.
- <sup>118</sup> L. Visscher, *Theor. Chem. Acc.* **98**, 68 (1997).
- <sup>119</sup> L. Cheng, J. Gauss, and J. F. Stanton, *J. Chem. Phys.* **139**, 054105 (2013), <https://doi.org/10.1063/1.4816130>.
- <sup>120</sup> K. Kuchitsu, ed., *Structure of Free Polyatomic Molecules* (Springer Berlin Heidelberg, 1998).
- <sup>121</sup> L. Visscher, J. Styszyński, and W. C. Nieuwpoort, *J. Chem. Phys.* **105**, 1987 (1996).
- <sup>122</sup> L. Halbert, M. L. Vidal, A. Shee, S. Coriani, and A. S. P. Gomes, "Dataset: Relativistic eom-ccsd for core-excited and core-ionized state energies based on the 4-component dirac-coulomb(-gaunt) hamiltonian," (2020).
- <sup>123</sup> R. Sarangi, M. L. Vidal, S. Coriani, and A. I. Krylov, *Mol. Phys.* **118**, e1769872 (2020), <https://doi.org/10.1080/00268976.2020.1769872>.
- <sup>124</sup> M. F. Herbst and T. Fransson, *J. Chem. Phys.* **153**, 054114 (2020), <https://doi.org/10.1063/5.0013538>.
- <sup>125</sup> K. Koziol and G. A. Aucar, *J. Chem. Phys.* **148**, 134101 (2018).
- <sup>126</sup> S. H. Southworth, R. W. Dunford, D. Ray, E. P. Kanter, G. Doumy, A. M. March, P. J. Ho, B. Krässig, Y. Gao, C. S. Lehmann, A. Picón, L. Young, D. A. Walko, and L. Cheng, *Phys. Rev. A* **100**, 022507 (2019).
- <sup>127</sup> S. H. Southworth, R. Wehlitz, A. Picón, C. S. Lehmann, L. Cheng, and J. F. Stanton, *J. Chem. Phys.* **142**, 224302 (2015).
- <sup>128</sup> M. Oura, T. Gejo, K. Nagaya, Y. Kohmura, K. Tamasaku, M. N. Piancastelli, and M. Simon, *New J. Phys.* **21**, 043015 (2019).
- <sup>129</sup> T. X. Carroll, R. W. Shaw Jr, T. D. Thomas, C. Kindle, and N. Bartlett, *J. Am. Chem. Soc.* **96**, 1989 (1974).
- <sup>130</sup> N. Boudjemia, K. Jank'ala, T. Gejo, K. Nagaya, K. Tamasaku, M. Huttula, M. N. Piancastelli, M. Simon, and M. Oura, *Phys. Chem. Chem. Phys.* **21**, 5448 (2019).
- <sup>131</sup> T. Mooney, E. Lindroth, P. Indelicato, E. G. Kessler, and R. D. Deslattes, *Phys. Rev. A* **45**, 1531 (1992).

Supplemental Information: Relativistic EOM-CCSD for core-excited and core-ionized state energies based on the 4-component Dirac-Coulomb(-Gaunt) Hamiltonian.

Loïc Halbert,<sup>1, a)</sup> Marta L. Vidal,<sup>2, b)</sup> Avijit Shee,<sup>3, c)</sup> Sonia Coriani,<sup>2, d)</sup> and André Severo Pereira Gomes<sup>1, e)</sup>

<sup>1)</sup>Université de Lille, CNRS, UMR 8523 – PhLAM – Physique des Lasers, Atomes et Molécules, F-59000 Lille, France

<sup>2)</sup>Department of Chemistry, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

<sup>3)</sup>Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

---

a) Electronic mail: [loic.halbert@univ-lille.fr](mailto:loic.halbert@univ-lille.fr)

b) Electronic mail: [malop@kemi.dtu.dk](mailto:malop@kemi.dtu.dk)

c) Electronic mail: [ashee@umich.edu](mailto:ashee@umich.edu)

d) Electronic mail: [soco@kemi.dtu.dk](mailto:soco@kemi.dtu.dk)

e) Electronic mail: [andre.gomes@univ-lille.fr](mailto:andre.gomes@univ-lille.fr)

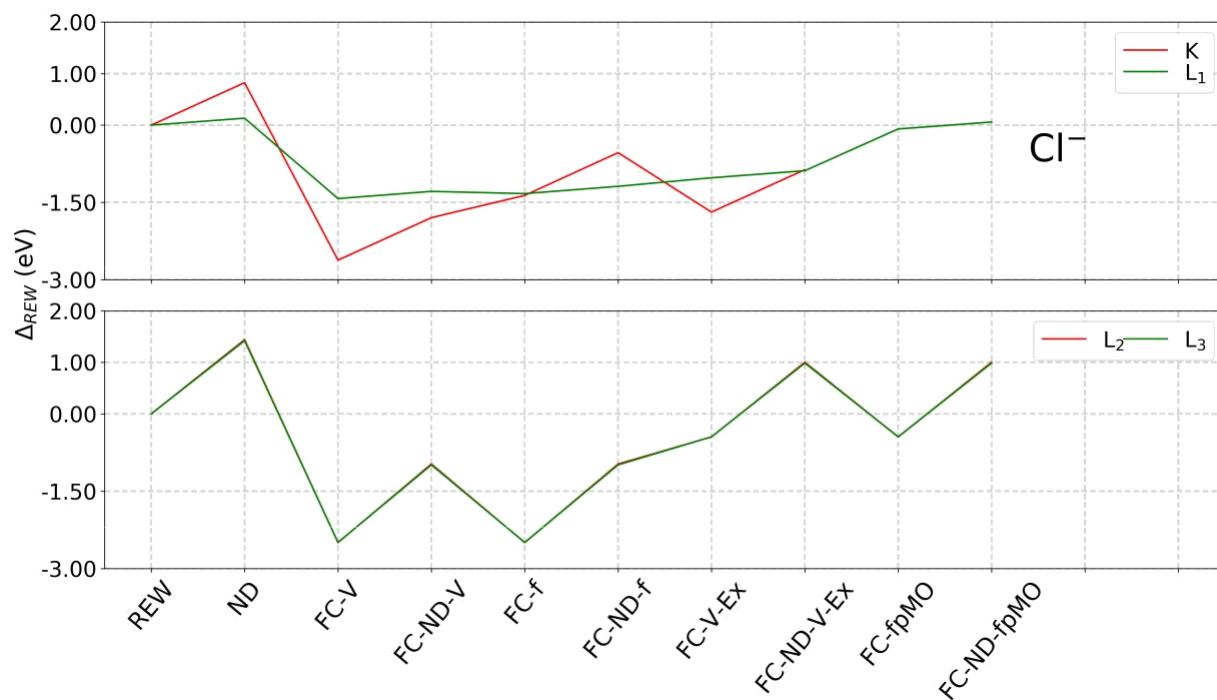


FIG. S2: Effect of the REW variants on the different binding energies of Cl<sup>-</sup>. Values (in eV) are relative to the original REW approach.

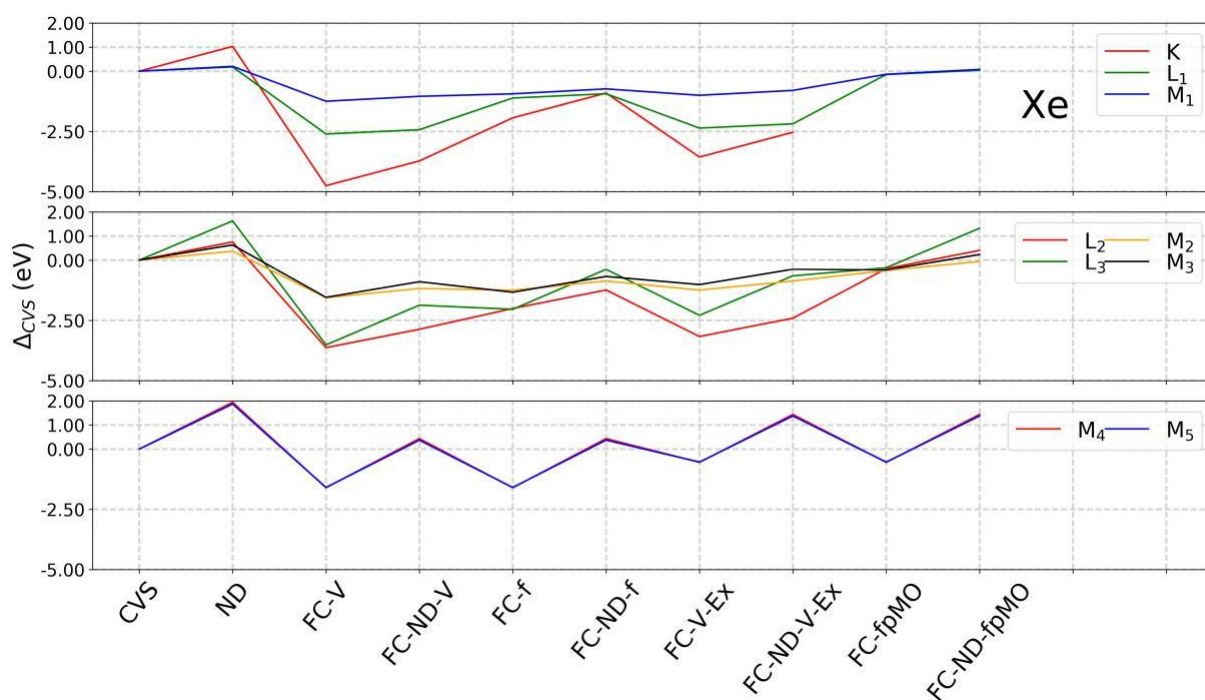


FIG. S1: Effect of the CVS variants on the different binding energies of Xe. Values (in eV) are relative to the original CVS approach.

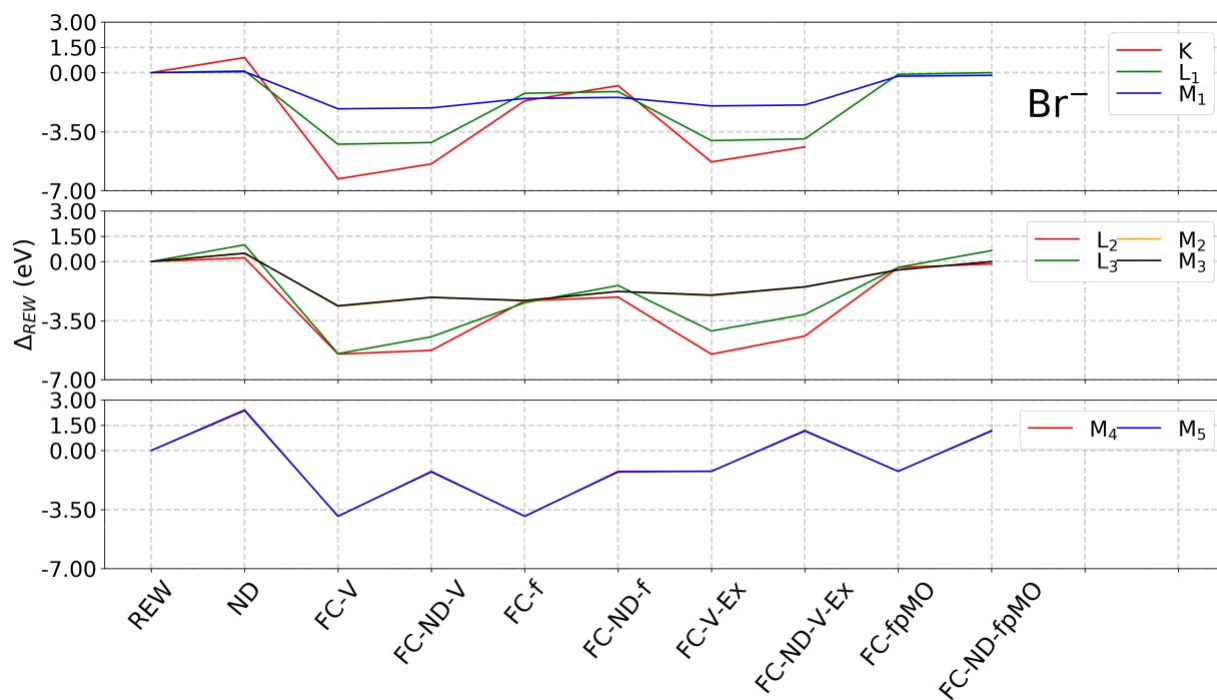


FIG. S3: Effect of the REW variants on the different binding energies of  $\text{Br}^-$ . Values (in eV) are relative to the original REW approach.

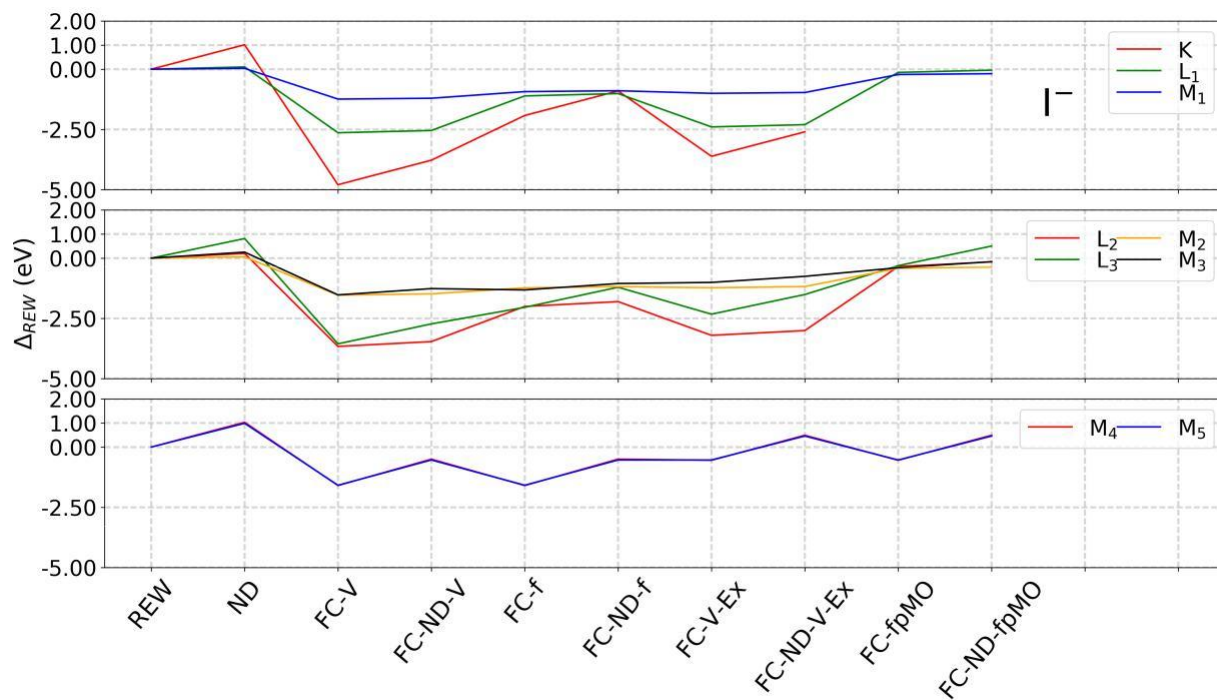


FIG. S4: Effect of the REW variants on the different binding energies of  $\text{I}^-$ . Values (in eV) are relative to the original REW approach.

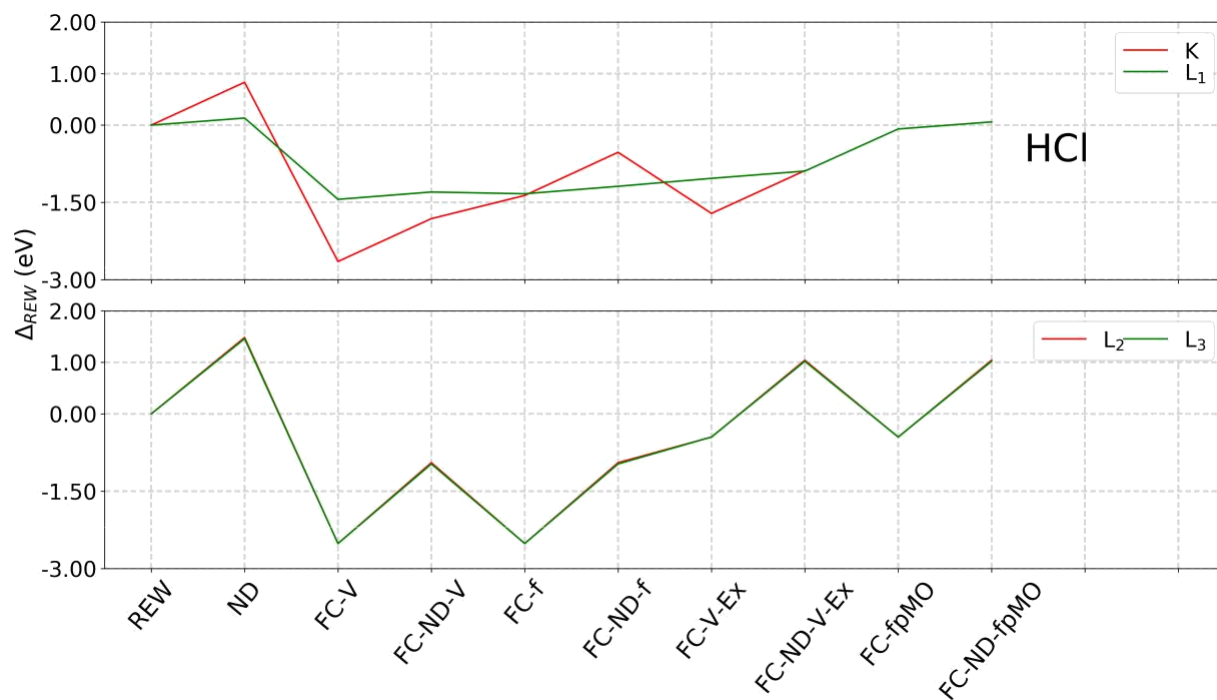


FIG. S5: Effect of the REW variants on the different binding energies of HCl. Values (in eV) are relative to the original REW approach.

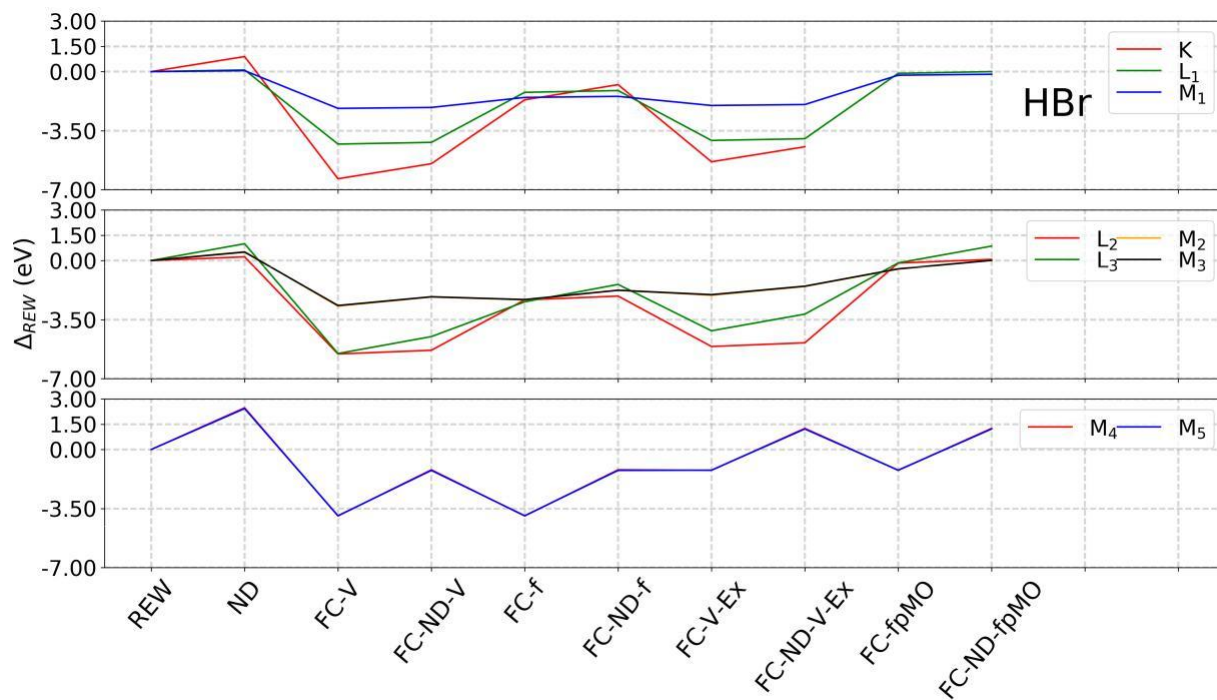


FIG. S6: Effect of the REW variants on the different binding energies of HBr. Values (in eV) are relative to the original REW approach.

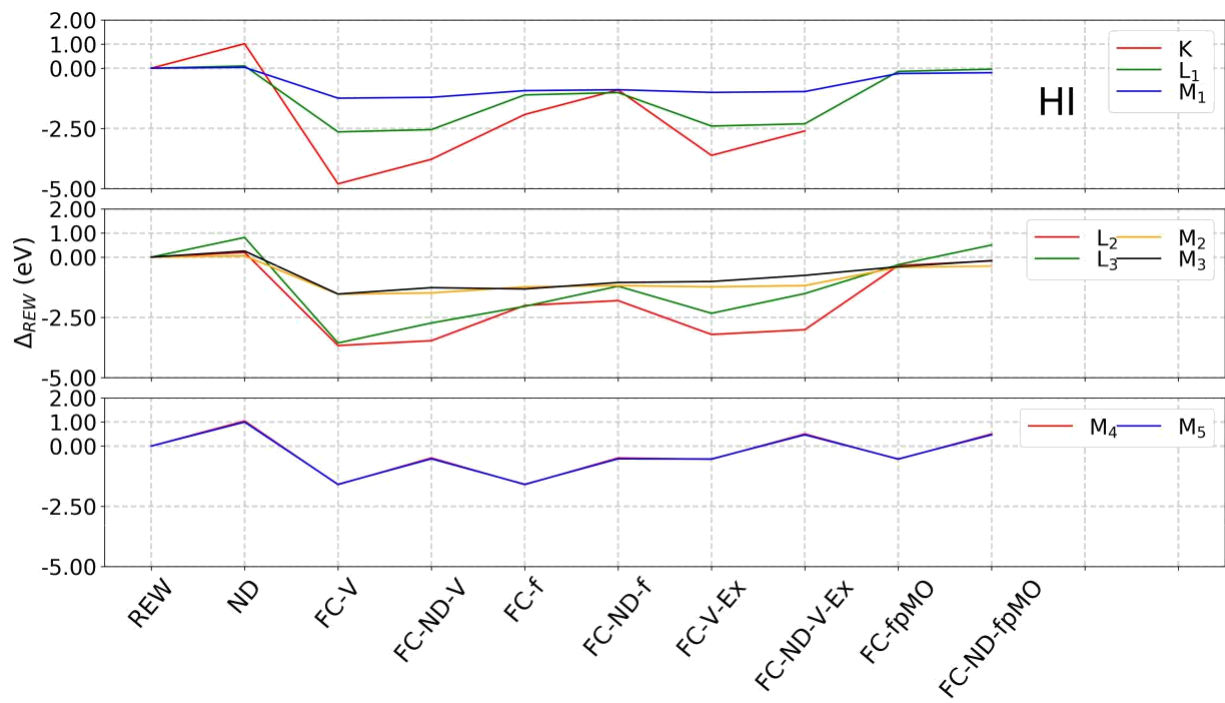


FIG. S7: Effect of the REW variants on the different binding energies of HI. Values (in eV) are relative to the original REW approach.

TABLE S1: Operation counts for the construction of a single  $\sigma$  vector for EOM-EE and EOM-IP wavefunctions for different variants (P: CVS-EOM using projectors; A: analytical CVS-EOM; A(FC-f) : analytical CVS-EOM with the frozen core approximation; A(ND): analytical CVS-EOM with the ND approximation; A(FC-f + ND): analytical CVS-EOM combining FC-f and ND; see manuscript for additional details). The operation counts are calculated employing the number of virtual ( $\text{Cl}^-$ : 208;  $\text{At}^-$ : 504 with dyall.acv3z basis sets) valence occupied and core spinors for each edge (with 2, 8, 18 and 60 core spinors for the K, L<sub>3</sub>, M<sub>5</sub> and N<sub>7</sub> edges, respectively) from the working equations from Shee et al. J. Chem. Phys. (2018) 149, 174113. For convenience these expressions are reproduced below, with different colors identifying the terms that relate to the A(FC-f), A(ND) and A(FC-f + ND) variants.

		CVS variants				
System	Edge	P	A	A(FC-f)	A(ND)	A(FC-f + ND)
EOM-EE						
$\text{Cl}^-$	K	2.0E+14	1.4E+13	8.6E+12	1.1E+13	7.7E+12
	L <sub>3</sub>	2.0E+14	1.1E+14	1.1E+13	3.2E+13	4.9E+12
$\text{At}^-$	K	3.6E+18	4.4E+16	3.9E+16	4.1E+16	3.8E+16
	L <sub>3</sub>	3.6E+18	2.7E+17	1.6E+17	2.0E+17	1.4E+17
	M <sub>5</sub>	3.6E+18	1.0E+18	2.6E+17	4.8E+17	1.8E+17
	N <sub>7</sub>	3.6E+18	2.6E+18	1.1E+17	4.6E+17	3.4E+16
EOM-IP						
$\text{Cl}^-$	K	5.2E+12	1.1E+10	3.3E+07	9.2E+09	2.7E+07
	L <sub>3</sub>	5.2E+12	2.0E+11	1.9E+08	5.8E+10	6.8E+07
$\text{At}^-$	K	2.0E+16	3.7E+12	3.9E+09	3.6E+12	3.7E+09
	L <sub>3</sub>	2.0E+16	8.9E+13	2.1E+10	7.4E+13	1.7E+10
	M <sub>5</sub>	2.0E+16	6.2E+14	6.7E+10	3.4E+14	3.7E+10
	N <sub>7</sub>	2.0E+16	2.2E+15	1.5E+11	3.1E+14	3.6E+10

Below are the equations for EOM-IP and EOM-EE right-hand side  $\sigma$ -vector equations (we employ I, J, . . . to denote occupied core orbitals and  $i_v, j_v, . . .$  occupied valence orbitals), color-coded as to indicate which terms would be present or absent for different approximations:

- Black: always calculated
- Blue: calculated if ND is not invoked
- Green: calculated if FC-f is not invoked
- Red: calculated if neither FC-f nor ND are invoked

EE:

$$R_{\sigma I}^a = \begin{aligned} & \overset{X}{F_e^a} r_{Ie}^e - \overset{X_M}{F_I^M} r_{M^a} + \overset{X_v}{F_e^{m_e}} r_{m^{ea}} + \overset{X}{F_e^M} r_{M^{ea}} + \overset{X}{W_{eI}^{Ma}} r_{M^e} \\ & + \overset{X}{W_{ef}^{amv}} r_{Imv}^{ef} + \overset{X}{W_{ef}^{aM}} r_{IM}^{ef} - \overset{X}{W_{le}^{Mn_v}} r_{Mn_v}^{ae} + \overset{X}{W_{le}^{MN}} r_{MN}^{ae} \end{aligned} \quad (S1)$$

$$R_{\sigma I_j}^{ab_v} = \begin{aligned} & -P(ab) \overset{X}{W_{Ij}^{Mb}} r_{M^a} + \overset{X}{W_{ej}^{ab}} r_{I^e} + P(ab) \overset{X}{(W_{fe}^{bM} r_{M^e})} t_{Ij}^{af_v} \\ & + \overset{X}{(W_{jv}^{n_v} r_{M^e})} t_{In}^{ab_v} - \overset{X}{(W_{le}^{n_vM} r_{M^e})} t_{jv}^{ab} + \overset{X}{(W_{jv}^{NM} r_{M^e})} t_{IN}^{ab} - \overset{X}{(W_{le}^{NM} r_{M^e})} t_{jv}^{ab} \\ & + P(ab) \overset{X}{F_e^b} r_{Ij}^{ae} - \overset{X}{F_{jv}^{m_v}} r_{Imv}^{ab} - \overset{X}{F_{jv}^M} r_{IM}^{ab} + \overset{X}{F_I^M} r_{jvM}^{ab} + \overset{X}{W_{Ij}^{Mn_v}} r_{Mn_v}^{ab} \\ & + \overset{X}{W_{Ijv}^{MN}} r_{ab} + P(ab) \overset{X}{W_{m_e}^{m_e}} r_{ejv}^{ae} + P(ab) \overset{X}{W_{ejv}^{Mbrae}} r_{IM} \\ & - P(ab) \overset{X}{W_{eI}^{Mb}} r_{jvM}^{ae} - P(ab) \overset{X}{(V_{fe}^{n_vM} r_{Mn_v}^{ea})} t_{Ij}^{fb_v} - P(ab) \overset{X}{(V_{fe}^{NM} r_{MN}^{ea})} t_{Ij}^{fb_v} \\ & - P(Ij_v) \overset{X}{(V_{fe}^{n_vM} r_{Imv}^{fe})} t_{jv}^{ba} - P(Ij_v) \overset{X}{(V_{fe}^{n_vM} r_{IM}^{fe})} t_{jv}^{ba} \\ & + \overset{X}{(V_{fe}^{n_vM} r_{jvM}^{fe})} t_{In}^{ba} - \overset{X}{(V_{fe}^{NM} r_{IM}^{fe})} t_{jv}^{ba} + \overset{X}{(V_{fe}^{NM} r_{jvM}^{fe})} t_{IN}^{ba} \\ & - \overset{X}{(V_{fe}^{Nmv} r_{Imv}^{fe})} t_{jvN}^{ba} + \overset{X}{W_{ef}^{ab}} r_{Ijv}^{ef} \end{aligned} \quad (S2)$$



$$\begin{aligned}
R_{\sigma_{IJ}^{ab}} = & -P(ab) \overset{X}{M} W_{IJ}^{Mb} r_{IM}^a + P(IJ) \overset{X}{e} W_{eJ}^{ab} r_e^e + P(ab) \overset{X}{efM} (W_{fe}^{bM} r_M^e) t_{IJ}^{af} \\
& + P(IJ) \overset{X}{Mn} (W_{nvM}^{re}) t_{JeM}^{ab} + P(IJ) \overset{X}{MNe} (W_{JeM}^{NIM} r_e^e) t_{MN}^{ab} + P(ab) \overset{X}{e} F_{eIJ}^{brae} \\
& - P(IJ) \overset{X}{m} \bar{F}_J^{mv} r_{Im}^{ab} - P(IJ) \overset{X}{M} \bar{F}_J^M r_{IM}^{ab} + \overset{X}{Mn} W_{IJ}^{Mnv} r_{Mn}^{ab} + \overset{X}{M>N} W_{IJ}^{MN} r_{MN}^{ab} \\
& + P(ab)P(IJ) \overset{X}{me} W_{eJ}^{mvb} r_{Im}^{ae} + P(ab)P(IJ) \overset{X}{Me} W_{eJ}^{Mb} r_{IM}^{ae} - P(ab) \overset{X}{ef,Mn} (V_{fe}^{nvM} r_{Mn}^{ea}) t_{IJ}^{fb} \\
& - P(ab) \overset{X}{ef,M>N} (V_{fe}^{NM} r_{MN}^{ea}) t_{IJ}^{fb} - P(IJ) \overset{X}{e>f,mn} (V_{fe}^{nmv} r_{Imv}^{fe}) t_{Jnv}^{ba} - P(IJ) \overset{X}{e>f,Mn} (V_{fe}^{nvM} r_{IM}^{fe}) t_{Jnv}^{ba} \\
& - P(IJ) \overset{X}{e>f,mN} (V_{fe}^{Nmv} r_{Imv}^{fe}) t_{JN}^{ba} - P(IJ) \overset{X}{e>f,MN} (V_{fe}^{NM} r_{IM}^{fe}) t_{JN}^{ba} + \overset{X}{e>f} W_{efIJ}^{ab}
\end{aligned} \tag{S3}$$

IP:

$$R_{\sigma_{I=-}} = \overset{X}{M} \bar{F}_I^M r_{IM} + \overset{X}{me} \bar{F}_e^{mv} r_{Im}^e + \overset{X}{Me} \bar{F}_e^M r_{IM}^e + \overset{X}{Mn,e} W_{Ie}^{Mnv} r_{Mn}^e + \overset{X}{M>N,e} W_{Ie}^{MN} r_{MN}^e \tag{S4}$$

$$\begin{aligned}
R_{\sigma_{Ij}^{av}} = & - \overset{X}{M} W_{IJ}^{Ma} r_{Mj}^v - \overset{X}{M} \bar{F}_I^{mv} r_{Mj}^{av} + \overset{X}{mnv} \bar{F}_j^{mv} r_{mv}^{av} + \overset{X}{M} \bar{F}_j^M r_{Mj}^a \\
& + \overset{X}{Mnv} W_{Mnv}^{ra} r_{Ijv}^{Mnv} + \overset{X}{M>N} W_{Mn}^{ra} r_{Ijv}^{MN} \\
& - \overset{X}{Me} W_{Ie}^{Ma} r_{Mj}^e + \overset{X}{me} W_j^{mve} r_{me}^{av} + \overset{X}{Me} W_j^{Ma} r_{Me}^e \\
& + \overset{X}{e} F_{eIjv}^{ae} + \overset{X}{Mn} (V_{ef}^{Mnv} r_{Mnv}^f) t_{Ijv}^{ea} + \overset{X}{M>N,ef} (V_{ef}^{MN} r_{MN}^f) t_{Ijv}^{ea}
\end{aligned} \tag{S5}$$

$$\begin{aligned}
R_{\sigma_{IJ}^a} = & - \overset{X}{m^{X,N}} W_{IJ}^{Ma} r_{mN}^a - P(IJ) \overset{X}{mnv} \bar{F}_I^{mv} r_{mN}^{av} - P(IJ) \overset{X}{Xv} \bar{F}_I^M r_{Mj}^a \\
& + \overset{X}{Mn} W_{IJ}^{mN} r_{mN}^a + \overset{X}{M>N} W_{IJ}^{MN} r_{MN}^a - P(IJ) \overset{X}{me} W_{Ie}^{mva} r_{me}^e \\
& - P(IJ) \overset{X}{Me} W_{Ie}^{Ma} r_{Mj}^e + \overset{X}{e} \bar{F}_e^a r_{IJ}^e + \overset{X}{Mn} (V_{ef}^{mn} r_{Mnv}^f) t_{IJ}^{ea} + \overset{X}{M>N,ef} (V_{ef}^{MN} r_{MN}^f) t_{IJ}^{ea}
\end{aligned} \tag{S6}$$

The  $\bar{F}$ ,  $W$  and  $G$  intermediates are defined as follows:

$$\bar{F}_{iv} = f_{iv} + f_{iv} t_e + V_{ivnv} t_e + V_{ivN} t_e + V_{ivnv} T_{ef} + V_{ivN} T_{ef}^{ef} \quad (S7)$$

$$\bar{F}_{iv} = f_{iv} + f_{iv} t_e + V_{ivnv} t_e + V_{ivN} t_e + V_{ivnv} T_{ef} + V_{ivN} T_{ef} \quad (S8)$$

$$\bar{F}_{iv} = f_{iv} + f_{iv} t_e + V_{ivnv} t_e + V_{ivN} t_e + V_{ivnv} T_{ef} + V_{ivN} T_{ef} \quad (S9)$$

$$\bar{F}_{iv} = f_{iv} + f_{iv} t_e + V_{ivnv} t_e + V_{ivN} t_e + V_{ivnv} T_{ef} + V_{ivN} T_{ef} \quad (S10)$$

$$\bar{F}_{iv} = f_{iv} + f_{iv} t_e + V_{ivnv} t_e + V_{ivN} t_e + V_{ivnv} T_{ef} + V_{ivN} T_{ef} \quad (S11)$$

$$\bar{F}_{iv} = f_{iv} + f_{iv} t_e + V_{ivnv} t_e + V_{ivN} t_e + V_{ivnv} T_{ef} + V_{ivN} T_{ef} \quad (S12)$$

$$\bar{F}_{iv} = f_{iv} + f_{iv} t_e + V_{ivnv} t_e + V_{ivN} t_e + V_{ivnv} T_{ef} + V_{ivN} T_{ef} \quad (S13)$$

$$W_{MN}^{IJ} = V_{MN}^{IJ} + P(MN) + V_{eNM}^{IJ} t_e + V_{ef}^{IJ} T_{ef}^{MN} \quad (S14)$$

$$W_{Mn}^{IJ_v} = V_{Mn}^{IJ_v} + e V_{en}^{IJ_v} t_e - e V_{en}^{IJ} t_{nv} + e V_{ef}^{IJ} T_{MN}^{ef_v} \quad (S15)$$

$$l_{jv} = l_{jv} + l_{jv} e - l_{jv} e - l_{jv} e \quad (S16)$$

X

S9 X

X

$$W_{MN}^{I_{jv}} = V_{MN}^{I_{jv}} + P(MN) \sum_e V_{eN}^{I_{jv}} t_e^e + \sum_{e>f} V_{ef}^{I_{jv}} T_{ef}^{MN} \quad (S17)$$

$$W_{ej}^{m_v b} = V_{ej}^{m_v b} + \sum_{nf} V_{ef}^{m_v b} t_{jv}^{nf} - \sum_{nv} V_{ej}^{m_v n_v} t_{nv}^b - \sum_N V_{ej}^{m_v N} t_N^b + \sum_{ef} V_{ef}^{m_v n_v} (t_{jv n_v}^{fb} - t_{jv n_v}^{fb}) + \sum_{Nf} V_{ef}^{m_v N} (t_{jv N}^{fb} - t_{jv N}^{fb}) \quad (S18)$$

$$W_{eJ}^{m_v b} = V_{eJ}^{m_v b} + \sum_{nf} V_{ef}^{m_v b} t_{jv}^{nf} - \sum_{nv} V_{eJ}^{m_v n_v} t_{nv}^b - \sum_N V_{eJ}^{m_v N} t_N^b + \sum_{ef} V_{ef}^{m_v n_v} (t_{Jn_v}^{fb} - t_{Jn_v}^{fb}) + \sum_{Nf} V_{ef}^{m_v N} (t_{JN}^{fb} - t_{JN}^{fb}) \quad (S19)$$

$$W_{ej}^{M_b v} = V_{ej}^{M_b v} + \sum_{nf} V_{ef}^{M_b v} t_{jv}^{nf} - \sum_{nv} V_{ej}^{M_b n_v} t_{nv}^b - \sum_N V_{ej}^{M_b N} t_N^b + \sum_{ef} V_{ef}^{M_b n_v} (t_{jv n_v}^{fb} - t_{jv n_v}^{fb}) + \sum_{Nf} V_{ef}^{M_b N} (t_{jv N}^{fb} - t_{jv N}^{fb}) \quad (S20)$$

$$W_{eJ}^{M_b v} = V_{eJ}^{M_b v} + \sum_{nf} V_{ef}^{M_b v} t_{jv}^{nf} - \sum_{nv} V_{eJ}^{M_b n_v} t_{nv}^b - \sum_N V_{eJ}^{M_b N} t_N^b + \sum_{ef} V_{ef}^{M_b n_v} (t_{Jn_v}^{fb} - t_{Jn_v}^{fb}) + \sum_{Nf} V_{ef}^{M_b N} (t_{JN}^{fb} - t_{JN}^{fb}) \quad (S21)$$

$$(S22)$$

$$W_{ef}^{a m_v} = V_{ef}^{a m_v} - \sum_{nv} V_{ef}^{n_v m_v} t_{nv}^a - \sum_N V_{ef}^{N m_v} t_N^a \quad (S23)$$

$$W_{ef}^{a M} = V_{ef}^{a M} - \sum_{nv} V_{ef}^{n_v M} t_{nv}^a - \sum_N V_{ef}^{N M} t_N^a \quad (S24)$$

$$W_{Ie}^{M n_v} = V_{Ie}^{M n_v} + \sum_f t_{Ij}^f V_{fe}^{M n_v} \quad (S25)$$

$$W_{Ie}^{MN} = V_{Ie}^{MN} + \sum_f t_{Ij}^f V_{fe}^{MN} \quad (S26)$$

$$W_{Ie}^{m_v N} = V_{Ie}^{m_v N} + \sum_f t_{Ij}^f V_{fe}^{m_v N}$$

$$W_{i ve}^{m v N} = V_{i ve}^{m v N} + X_{f t_{i v}} V_{f e}^{m v N} \quad (S27)$$

$$W_{Mn_v}^{le} = V_{Mn_v}^{le} + X_{f t_{Mn_v}} V_{f e}^{le} - X_{ovf} V_{Mn_v}^{lo v e} + X_{ovf} V_{Mn_v}^{IO e} + V_{fg}^{T_{Mn_v}} + V_{fn_v}^{le} + V_{M}^{fg} - V_{le f}^{le f} \quad (S28)$$

$$- X_{ovf} V_{Mn_v}^{lo v e} + X_{ovf} V_{Mn_v}^{IO e} + V_{fg}^{T_{Mn_v}} + V_{fn_v}^{le} + V_{M}^{fg} - V_{le f}^{le f} \quad (S29)$$

$$W_{am_v}^{ef} = V_{am_v}^{ef} + P(ef) X_{n g} V_{ag}^{env T_{m_v n_v}} + P(ef) X_{n g} V_{ag}^{e N T_{g f}} + X_{n g} W_{ag m_v}^{ef t g} + X_{n g} F_{a m_v n_v}^{n v t e f} + X_{n g} F_{a m_v N}^{N t e f} + X_{n > o} V_{am_v}^{n v o v T e f} + X_{n > o} V_{am_v}^{n v o T e f} + X_{n > o} V_{am_v}^{N O T e f} + X_{n > o} W_{am_v}^{ef t g} \quad (S30)$$

$$W_{ab}^{ef} = V_{ab}^{ef} - P(ef) X_{m v N} V_{ab}^{m v N T_{m_v N}^{ef}} + X_{m v N} V_{ab}^{m v N T_{m_v N}^{ef}} + X_{m v N} V_{ab}^{M N T_{M N}^{ef}} + X_{m v N} V_{ab}^{M N T_{M N}^{ef}} + X_{m v N} V_{ab}^{m v N T_{m_v N}^{ef}} + X_{m v N} V_{ab}^{m v N T_{m_v N}^{ef}} + X_{m v N} V_{ab}^{m v N T_{m_v N}^{ef}} \quad (S31)$$

M>N

$$\overline{W}_{ejv}^{m v b} = V_{ejv}^{m v b} - X_{n v f} V_{ef}^{m v n v} t_{n v j v}^{b f} - X_{n v f} V_{ef}^{m v N} t_{n v j v}^{b f} \quad (S32)$$

$$\overline{W}_{ejv}^{M b} = V_{ejv}^{M b} - X_{n v f} V_{ef}^{M n v} t_{n v j v}^{b f} - X_{n v f} V_{ef}^{M N} t_{n v j v}^{b f} \quad (S33)$$

$$\overline{W}_{ej}^{M b} = V_{ej}^{M b} - X_{n f} V_{ef}^{M n v} t_{n v j}^{b f} - X_{n f} V_{ef}^{M N} t_{n v j}^{b f} \quad (S34)$$