

# On the basis set selection for calculations of core-level states: Different strategies to balance cost and accuracy

Ronit Sarangi<sup>a</sup>, Marta L. Vidal<sup>b</sup>, Sonia Coriani<sup>b</sup>, and Anna I. Krylov<sup>a</sup>

<sup>a</sup> Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA

<sup>b</sup> DTU Chemistry - Department of Chemistry, Technical

University of Denmark, DK-2800, Kongens Lyngby, Denmark

We present a study on basis set effects in correlated calculations of core-level states. While it is well recognized that the core-level states require using more extensive basis sets than their valence counterparts, the standard strategy has been to use large contracted basis sets, such as the cc-pVXZ or cc-pCVXZ series. Building upon the ideas of Besley, Gilbert, and Gill [J. Chem. Phys. 130, 124308 (2009)], we show that a much more effective strategy is to use uncontracted bases, such as core or fully uncontracted Pople’s basis. The physical grounds behind this approach are explained and illustrated by numerical results. We also discuss other cost-saving strategies, such as virtual space truncation and mixed precision execution.

## I. INTRODUCTION

Owing to their unique capabilities, spectroscopies exploiting core-level transitions are gaining popularity[1–5]. The transitions involving core electrons are element-specific because of large energy gaps (hundreds of electron-volts) between different edges, yet they are sensitive to their chemical environment. Compact and localized shapes of core orbitals result in local sensitivity, which is particularly important for designing spectroscopic probes of the electronic structure and dynamics in complex environments. As in the case of valence spectroscopies[6], theoretical modeling is crucially important for the interpretation of the experimental spectra. Thus, the progress in experimental techniques, ranging from advanced light sources to table-top X-ray instruments, has stimulated vigorous theoretical developments[7, 8].

At first glance, electronic structure behind valence (UV-VIS) and core-level (X-ray) spectroscopies appear to be similar. For example, modeling photoelectron spectra entails calcula-

tions of  $N-1$ -electron states of a neutral molecule. Since the molecular Schrödinger equation contains the solutions for all states, regardless of their energy, one may expect that the same quantum-chemistry method could be used to describe both valence-ionized and core-ionized states. This is, however, not the case: although the Schrödinger equation is the same, the approximations to it, which are used to construct a practical quantum chemistry method (i.e., theoretical model chemistry, in John Pople’s terms[9]) may lead to manifestly different outcomes in the valence and core domains.

Theoretical model chemistry[9] is defined by the pair of approximations: one to the many-body problem (correlation treatment) and one to a one-electron basis set used to represent molecular orbitals and construct Slater determinants.

Equation-of-motion coupled-cluster (EOM-CC) theory[10–15] provides an effective and robust treatment of electron correlation and is capable of treating multiple electronic states on the same footing. Its accuracy can be systematically improved up to the exact limit. These properties make EOM-CC the method of choice for spectroscopy modeling. Challenges in correlated treatment of core-level states and possible solutions have been analyzed in recent reviews[7, 8] and original research papers[16–31]. Particularly effective approach is the extension of the EOM-CC methods to core-level states via the core-valence separation (CVS) scheme[24, 27, 32, 33]. The benchmarks illustrated that CVS-EOM-IP/EE-CCSD [24, 26–28, 33, 34] provides an effective and reliable description of core-ionized and core-excited states, including treatment of non-linear optical properties such as RIXS cross sections[29–31].

The special requirements to one-electron basis sets in calculations of core-level states have been recognized and documented in many papers[16, 20, 35–39]. In a nutshell, obtaining converged and accurate results for core-level states requires considerably larger bases than needed for their valence counterparts. This high sensitivity of the results to the one-electron basis is observed even in uncorrelated calculations, e.g., at the Hartree-Fock level[16]. Its physical origin is a strong perturbation caused by the creation of a core hole as a result of removing or exciting core electrons. To describe the ensuing changes in electronic structure, traditionally referred to as orbital relaxation, sufficiently flexible basis sets are needed.

In the previous CVS-EOM-CC benchmark studies, series of standard contracted basis sets have been tested[35, 40]. In Ref. [35], Coriani and coworkers investigated the convergence with respect to the basis-set size with coupled-cluster methods of increasing complexity

(CC2, CCSD, CC3, and CCSDT). The largest bases tested in this study were aug-cc-pCV5Z and d-aug-cc-pCV5Z. The authors observed monotonous decrease of the computed ionization and excitation energies towards the experimental values upon increasing the basis-set cardinal number, which illustrates that the target core-level states are more sensitive to the basis set than the ground-state reference. Coriani and co-workers[35] exploited this smooth convergence of the results to extrapolate the computed excitation energies to the complete basis set (CBS) limit. They noted that for ionization energies the aug-cc-pCV5Z basis (the largest used for IEs in their study) appears to be close to the convergence limit, judging from the small differences between this and smaller bases. While additional diffuse functions are required to properly describe core-excited states, they were found to be less important for core-ionized states. The relativistic effects were found to be less sensitive to the basis set[16, 40]. These observations confirm that the main reason for extended basis sets in core-level calculations is due to orbital relaxation.

Here we systematically explore an alternative strategy, used by Gill and coworkers[16] and by us in recent applications[29]. Instead of following the hierarchy of Dunning’s bases, optimized to describe electron correlation in ground-state molecules, we build series of basis sets by uncontracting the core and valence functions. We consider Pople’s and Dunning’s sets and show that using uncontracted Pople’s bases is much more effective than using Dunning’s bases. Our results provide a simple guideline for choosing basis sets for calculations of core-level states. In addition to effective basis-set choices, we also briefly explore other cost-saving strategies.

## II. STUDY DESIGN, THEORETICAL METHODS, AND COMPUTATIONAL PROTOCOLS

In this study, we focus on the calculations of ionized states using the fc-CVS-EOM-IP-CCSD method[27, 41, 42], with the goal of investigating the ability of various basis sets to describe orbital relaxation effects at a correlated level of theory. By focusing on core-ionized states, we can investigate perturbation of the electronic structure due to creation of the core hole. Because excitation of core electrons also creates a core hole, the results should be largely transferable to core-excited states, with the caveat that the calculations of the XAS transitions require additional diffuse functions to describe states of Rydberg

character. While we provide experimental results when available, our main emphasis is not on the differences relative to the experiment, but rather on the convergence of the theoretical values to the basis-set limit.

To explain the rational behind the design of our study, let us briefly discuss the effects caused by the removal of a core electron. Because core orbitals are compact, they screen the nuclear charge much more effectively than the valence orbitals do. Thus, removing a core electron from an atom is roughly equivalent to increasing the nuclear charge by one, in terms of the Coulomb field experienced by the remaining electrons. This increased Coulomb attraction causes the valence atomic orbitals to collapse toward the nucleus. To describe such collapse, the basis set must have significant radial flexibility; angular flexibility is less important. For this reason, one should use at least a triple- $\zeta$  (or better) basis. The collapse of the core orbitals has even a larger energetic effect because of the large contribution of core electrons to the total electronic energy. According to Slater’s rules [43], the shielding effect of one  $1s$  electron on the other  $1s$  electron is roughly 0.3 protons. This core collapse has huge energetic consequences; thus, it is essential to describe it well to obtain accurate results. The basis, therefore, should include a sufficient number of the core functions. This can be achieved by choosing polarized-core Dunning’s sets (cc-pCVXZ) or by decontracting the core functions, such as the “6-” core function in the split-valence Pople bases, as was done in Refs. [16, 29, 44]. Core-correlation effects are considerably smaller (in energy) than these “radial collapse” effects and, for that reason, one may expect core-correlated basis sets to be less effective than core-decontracted ones [16]. To verify whether this expectation holds when using a high-level correlated method, we consider series of the original contracted Dunning and Pople basis sets of a triple- $\zeta$  quality and above and their partially or fully decontracted variants. The full description of the basis sets is given below.

### A. Computational Details

All calculations were carried out using the *Q-Chem* electronic structure program[45, 46]. We employ the fc-CVS-EOM-IP-CCSD method[27] in which the target ionized states are described by the following ansatz:

$$\Psi^{N-1} = (R_1 + R_2) e^{T_1+T_2} \Phi_0(N), \quad (1)$$

where  $\Phi_0(N)$  denotes the reference determinant of an  $N$ -electron system, the singles and doubles excitation operators  $T_1$  and  $T_2$  contain the amplitudes for the reference state obtained by solving CCSD equations. The excitation operators  $R_1$  and  $R_2$  contain the EOM amplitudes obtained solving an EOM eigenproblem. While  $T_1$  and  $T_2$  are particle- and spin-conserving operators, the EOM-IP operators are of an ionizing type:

$$R_1 = \sum_i r_i i; \quad R_2 = \frac{1}{2} \sum_{ija} r_{ij}^a a^\dagger j i. \quad (2)$$

Following the standard notation, indices  $i, j, k, \dots$  denote occupied orbitals and  $a, b, c, \dots$  denote virtual orbitals, as defined by the choice of the reference determinant  $\Phi_0$ . In fc-CVS-EOM-IP-CCSD, the core electrons are frozen at the CCSD step (i.e., respective amplitudes in  $T_1$  and  $T_2$  are zero) and the EOM amplitudes should involve at least one core orbital, as prescribed by the CVS scheme.

The definition of the core in our CVS scheme depends on the edge[27]: the edge of interest and all lower edges are frozen while all higher edges are active. In this study we focus on the molecules containing first- and second-row elements (C,N,O, and H). Thus, in calculations of the carbon edge, the standard definition of the frozen core is used: all 1s orbitals of the second-row atoms are frozen. In calculations of the nitrogen edge, only oxygen and nitrogen 1s orbitals are frozen while carbon’s 1s orbitals are active. Likewise, in calculations of the oxygen edge, only 1s<sub>O</sub> orbitals are frozen and the rest of the core orbitals are active.

Our benchmark set comprises two simple diatomics, carbon monoxide (CO) and dinitrogen (N<sub>2</sub>), three hydrides (water, ammonia, methane), and two polyatomic molecules, acrolein and glycine. This set allows us to investigate basis set effects for carbon, nitrogen, and oxygen edges, including molecules with several atoms of the same type and molecules with more than one edge.

The calculations for dinitrogen and carbon monoxide were carried out at the experimental geometries taken from Ref. [47] ( $R_{NN} = 1.097685$  Å and  $R_{CO} = 1.128323$  Å). The hydrides’ structures were taken from Ref. [25], where they were optimized with RI-MP2/cc-pVTZ. For acrolein (Fig. 1), we used an MP2/cc-pVTZ optimized structure. Glycine calculations were performed for the canonical isomer (the main form of the gas-phase glycine) using the RI-MP2/cc-pVTZ optimized structure taken from Ref. [25] (Fig. 2).

We used *Q-Chem*’s default convergence thresholds, except for the EOM amplitudes for

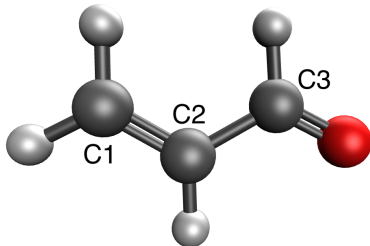


FIG. 1: Acrolein structure with atom labels.

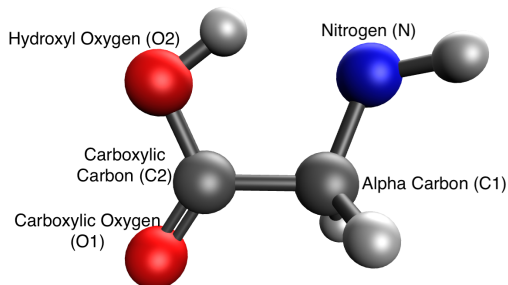


FIG. 2: Glycine structure (canonical isomer) with atom labels.

which a tighter threshold was used. SCF convergence was  $10^{-8}$ , CCSD convergence was  $10^{-6}$ , and the Davidson convergence was  $10^{-7}$ . In single-precision calculations (cf Sec. III E), CCSD convergence thresholds were  $10^{-4}$  for amplitudes and  $10^{-5}$  for energies.

The basis sets were decontracted manually and inputted as user-specified bases. For each basis, we considered two decontracted versions: one in which only the core orbitals were decontracted (this converts one core function from the 6-311+G(3df) basis set into six variationally independent functions) and one in which all functions were uncontracted. Using 6-311+G(3df) as an example, the latter procedure amounts to converting a triple- $\zeta$  basis into a 5- $\zeta$  one. The redundant basis functions, which appear in decontracted Dunning’s sets, were removed from the calculations.

Table I collects the basis sets used in this study, their contraction schemes, and the number of basis functions per atom for the second row elements. It also introduces shorthand notations for the uncontracted bases. We used pure angular momentum functions ( $5d$ ,  $7f$ ,  $9h$ ,  $11i$ , ...) for all bases. For Dunning’s bases, we used the versions with optimized contraction, as implemented in *Q-Chem*. The aug-cc-pCV5Z and aug-cc-pV6Z bases were taken from the Basis Set Exchange database[48], without optimizing the general contractions (numeric tests indicated that using the variants of these bases with optimized general

contractions leads to essentially the same results).

TABLE I: Basis sets, contraction schemes, and the number of functions per atom<sup>a</sup>.

Basis	Contraction level	Contraction scheme	#b.f.
6-311+G(2df)	Original	(12s6p2d1f)/[5s4p2d1f]	34
uC-6-311+G(2df)	Core-uncontracted	(12s6p2d1f)/[10s4p2d1f]	39
u-6-311+G(2df)	Fully uncontracted	(12s6p2d1f)/[12s6p2d1f]	47
6-311+G(3df)	Original	(12s6p3d1f)/[5s4p3d1f]	39
uC-6-311+G(3df)	Core-uncontracted	(12s6p3d1f)/[10s4p3d1f]	44
u-6-311+G(3df)	Fully uncontracted	(12s6p3d1f)/[12s6p3d1f]	52
aug-cc-pVTZ	Original	(11s6p3d2f)/[5s4p3d2f]	46
uC-aug-cc-pVTZ	Core-uncontracted	(11s6p3d2f)/[11s4p3d2f]	52
u-aug-cc-pVTZ	Fully uncontracted	(11s 6p 3d 2f)/[11s 6p 3d 2f]	58
aug-cc-pVQZ	Original	(13s7p4d3f)/[6s5p4d3f2g]	80
uC-aug-cc-pVQZ	Core-uncontracted	(13s7p4d3f2g)/[13s5p4d3f2g]	87
u-aug-cc-pVQZ	Fully uncontracted	(13s7p4d3f2g)/[13s7p4d3f2g]	93
aug-cc-pV5Z	Original	(15s9p5d4f3g2h)/[7s6p5d4f3g2h]	127
uC-aug-cc-pV5Z	Core-uncontracted	(15s9p5d4f3g2h)/[15s6p5d4f3g2h]	135
u-aug-cc-pV5Z	Fully uncontracted	(15s9p5d4f3g2h)/[15s9p5d4f3g2h]	144
aug-cc-pV6Z	Original	(17s11p6d5f4g3h2i)/[8s7p6d5f4g3h2i]	189
uC-aug-cc-pV6Z	Core-uncontracted	(17s11p6d5f4g3h2i)/[17s7p6d5f4g3h2i]	198
u-aug-cc-pV6Z	Fully uncontracted	(17s11p6d5f4g3h2i)/[17s11p6d5f4g3h2i]	210
aug-cc-pCV5Z	Original	(19s13p8d6f4g2h)/[11s10p8d6f4g2h]	181
uC-aug-cc-pCV5Z	Core-uncontracted	(19s13p8d6f4g2h)/[19s10p8d6f4g2h]	189
u-aug-cc-pCV5Z	Fully uncontracted	(19s13p8d6f4g2h)/[19s13p8d6f4g2h]	198

<sup>a</sup> For a 2nd row element (C,N,O, etc).

### III. RESULTS AND DISCUSSION

#### A. Nitrogen molecule, $N_2$

The results for  $N_2$  are collected in Table II and shown graphically in Fig. 3. Table II shows total CCSD energy of the neutral reference state and two core IEs, corresponding to ionization from  $\sigma_u(1s)$  (lower value, IE1) and  $\sigma_g(1s)$  (higher value, IE2) orbitals. The total energies show anticipated trends: they decrease upon uncontraction and the magnitude of the decrease is larger when the valence orbitals are uncontracted than when only the core orbitals are uncontracted. The magnitude of this decrease is larger for Pople's bases than for Dunning's bases, which is also expected because the relative increase in the number of basis functions is larger for Pople's bases.

As noted in the previous EOM-CC benchmark study[35], the IEs decrease monotonously in the series of contracted basis sets of increasing size. The IEs also generally decrease upon uncontraction. This is a manifestation of core-relaxation effects, which lower the energy of the target ionized state. In contrast to the total energies, the drop in IE is always larger when the core orbitals are uncontracted than when the valence orbitals are uncontracted. The magnitude of the change is larger for smaller bases than for larger bases because the increase in the basis size is larger for the smaller bases. We also observe that the changes are rather small when polarized-core basis is used, because these bases already afford sufficient flexibility in describing core electrons.

The results in Table II show that the IEs reach the basis-set limit within 0.01 eV for aug-cc-pCV5Z (and the respective uncontracted variants), uC-aug-cc-pV6Z/u-aug-cc-pV6Z, and u-aug-cc-pV5Z. The smallest among these bases is aug-cc-pCV5Z, which is not surprising, as this basis has more functions optimized for the core description (although they are optimized for describing the correlation of the core electrons in the ground state).

The energy gap between the two core states converges much faster with respect to the basis set than the absolute values of IEs, owing to error cancellation. For example, the difference between the smallest basis (6-311+G(2df)) and the basis-set limit is less than 0.001 eV.

Fig. 3 shows the lower core IE ( $\sigma_u$ ) for all basis sets as a function of the number of basis functions, which clearly indicates the effectiveness of different bases in describing core IEs.



While it is not surprising that larger bases perform better than smaller bases, as illustrated by the smooth trend of the red curve (original contracted bases sets), the difference between contracted and core-uncontracted bases is remarkable. For example, uC-aug-cc-pVQZ gives better result than aug-cc-pV5Z, despite being 1.5 times more compact. The performance of core and fully uncontracted Pople's bases is even more impressive—uncontracted u-6-311+G(3df) (with 52 functions per atom) delivers the same result as uC-aug-cc-pVQZ (87 functions per atom). Overall, the values with fully uncontracted u-6-311+G(3df) are within 0.06 eV from the basis set limit (u-aug-cc-pV5Z). Adding a second set of diffuse functions to uC-6-311+G(3df) lowers the IE by 0.002 eV (see Table II). Somewhat unexpectedly, the uncontracted aug-cc-pVTZ bases yield noticeably larger errors relative to the basis set limit than more compact uncontracted u-/uC-6-311+G(3df) bases.

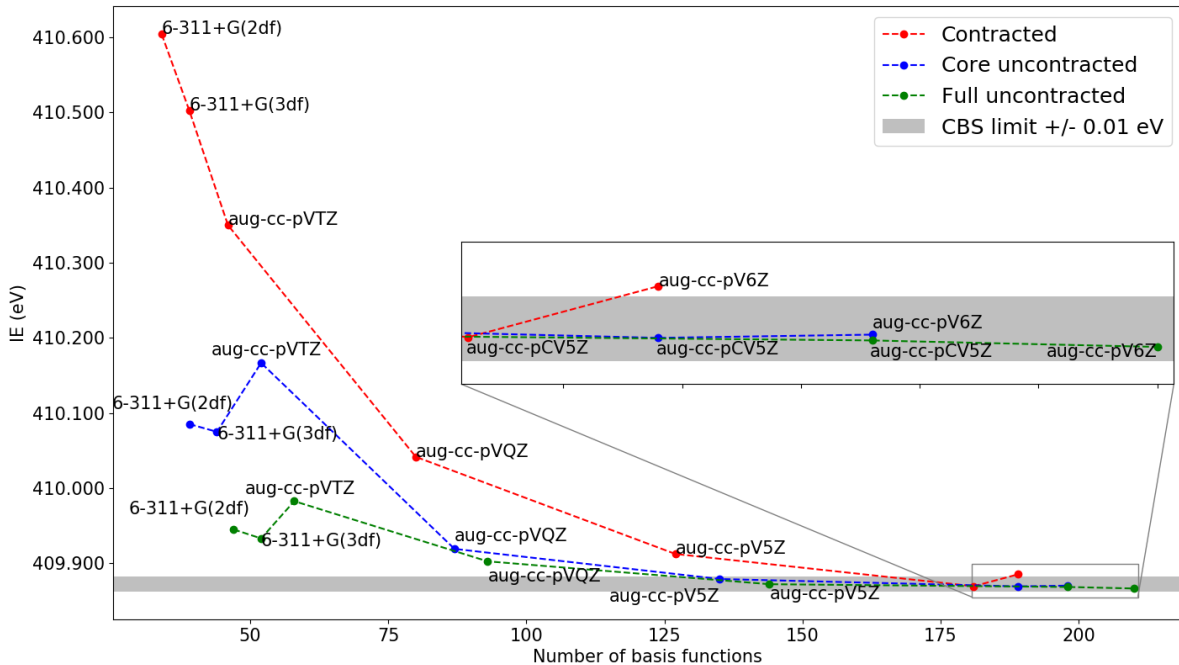


FIG. 3: Core IEs for  $N_2$  versus the number of basis functions per atom. Gray shaded area marks  $\pm 0.01$  eV interval around the basis-set limit (u-aug-cc-pV5Z).

## B. Carbon monoxide, CO

The results for CO are collected in Tables III and IV and shown graphically in Figs. 4 and 5. The experimental values were taken from Ref. [49]. Overall, the trends for both

TABLE II: Core IEs for N<sub>2</sub>, fc-CVS-EOM-IP-CCSD with various basis sets.

Basis	CCSD energy <sup>a</sup> (a.u.)	IE1 <sup>b</sup> (eV)	IE2 (eV)	$\Delta$ IE (eV)
6-311+G(2df)	-109.350207	410.6041	410.4996	0.1045
uC-6-311+G(2df)	-109.353623	410.0853	409.9802	0.1051
u-6-311+G(2df)	-109.357797	409.9451	409.8402	0.1049
6-311+G(3df)	-109.354688	410.5026	410.3981	0.1045
uC-6-311+G(3df)	-109.357388	410.0750	409.9701	0.1049
u-6-311+G(3df)	-109.361379	409.9331	409.8282	0.1049
6-311(2+)G(3df)	-109.354828	410.4998	410.3953	0.1045
uC-6-311(2+)G(3df)	-109.357517	410.0725	409.9676	0.1049
u-6-311(2+)G(3df)	-109.361498	409.9315	409.8266	0.1049
aug-cc-pVTZ	-109.361574	410.3500	410.2454	0.1046
uC-aug-cc-pVTZ	-109.363030	410.1667	410.0620	0.1047
u-aug-cc-pVTZ	-109.366680	409.9827	409.8781	0.1046
aug-cc-pVQZ	-109.386793	410.0417	409.9370	0.1047
uC-aug-cc-pVQZ	-109.387264	409.9192	409.8143	0.1049
u-aug-cc-pVQZ	-109.387920	409.4026	409.7976	0.1050
aug-cc-pV5Z	-109.394586	409.9124	409.8073	0.1051
uC-aug-cc-pV5Z	-109.394648	409.8791	409.7740	0.1051
u-aug-cc-pV5Z	-109.395016	409.8722	409.7671	0.1051
aug-cc-pV6Z	-109.397296	409.8853	409.7802	0.1051
uC-aug-cc-pV6Z	-109.397313	409.8703	409.7652	0.1051
u-aug-cc-pV6Z	-109.397514	409.8665	409.7614	0.1051
aug-cc-pCV5Z	-109.395330	409.8695	409.7644	0.1051
uC-aug-cc-pCV5Z	-109.395322	409.8693	409.7642	0.1051
u-aug-cc-pCV5Z	-109.395413	409.8685	409.7634	0.1051

<sup>a</sup> Total energy for the neutral reference state. <sup>b</sup> Experimental IE <sub>$\sigma_u$</sub> =409.9 eV is taken from Ref. [25].

edges follow very closely the trends observed for N<sub>2</sub>, reinforcing the main finding—impressive effectiveness of uncontracted Pople’s bases in describing the core IEs. For the carbon edge, u-6-311+G(3df) is within 0.01 eV from the basis-set limit (u-aug-cc-pV5Z), whereas for the oxygen edge the difference is slightly larger (0.1 eV).

Using this molecule with two edges as an example, we tested the protocol of using different bases for active and inactive edges, e.g., using uncontracted bases for both C and O or using an uncontracted basis for the active edge and an original, contracted basis for the inactive edge. The results show that the difference between the two schemes is small (except for

the smallest basis, 6-311+G(2df)), suggesting an effective compromise for calculations of polyatomic heteronuclear molecules.

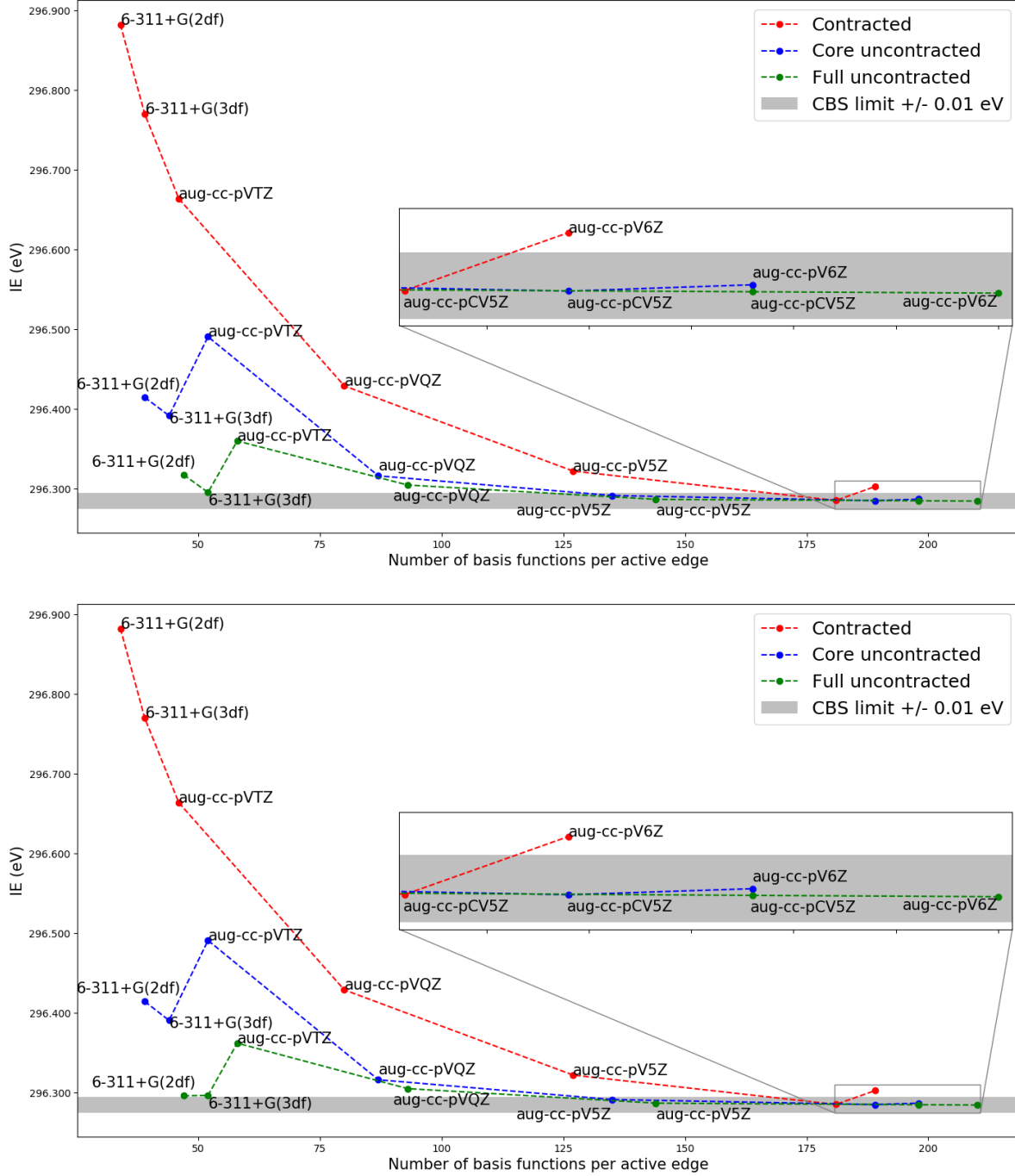


FIG. 4: CO, carbon edge IEs. Top: Only carbon basis is uncontracted, original matching basis is used for oxygen. Bottom: Both carbon and oxygen bases are uncontracted. Gray shaded area marks the  $\pm 0.01$  eV interval around the basis-set limit (u-aug-ccp-pV5Z).

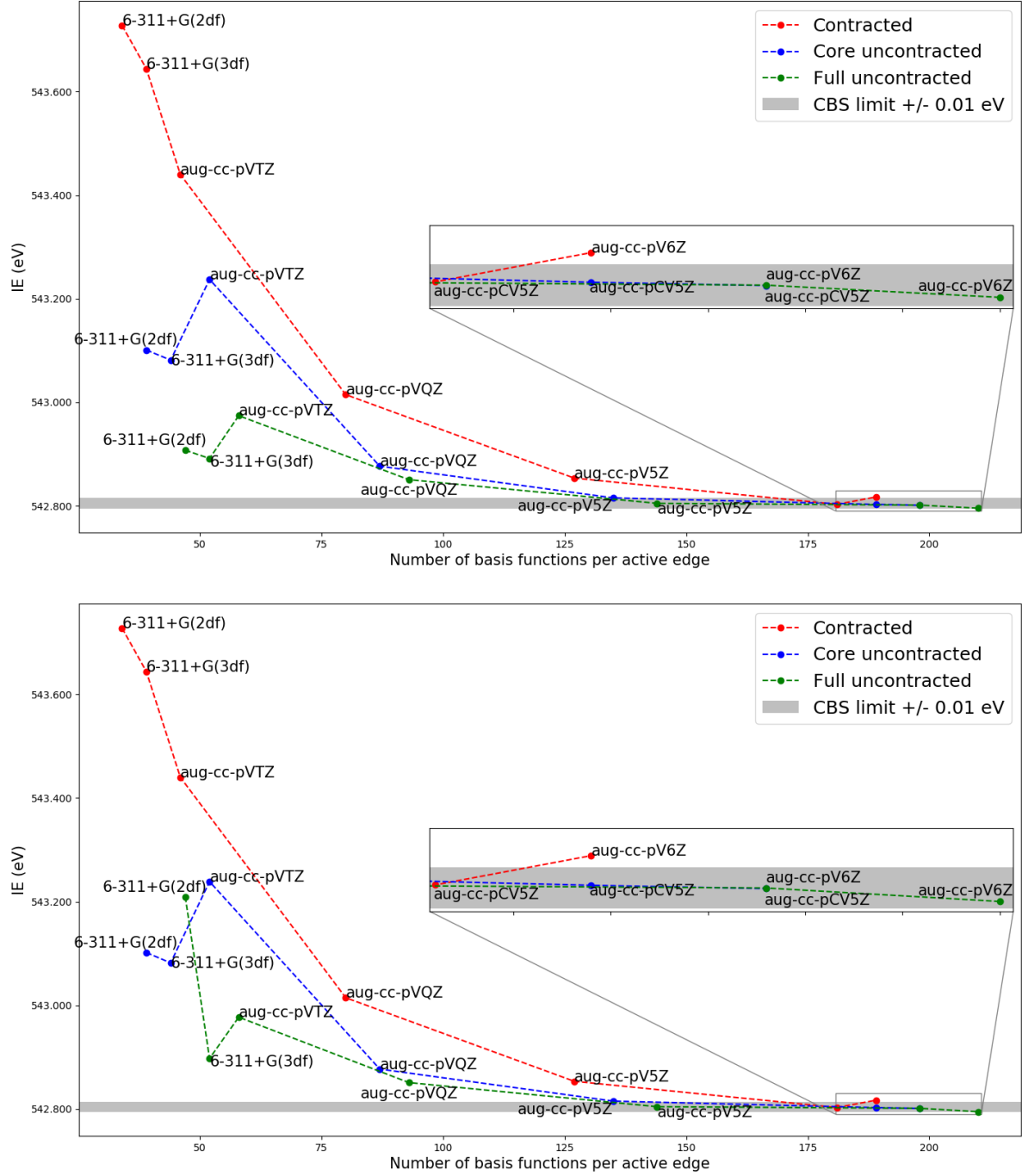


FIG. 5: CO, oxygen edge IEs. Top: Only oxygen basis is uncontracted, original matching basis is used for carbon. Bottom: Both carbon and oxygen bases are uncontracted. Gray shaded area marks the  $\pm 0.01$  eV interval around the basis-set limit (u-aug-ccp-pV5Z).

TABLE III: CO, carbon edge. Total and ionization energies;  
fc-CVS-EOM-IP-CCSD with various basis sets.

Basis on C	Basis on O	CCSD energy <sup>a</sup> (a.u.)	IE <sup>b</sup> (eV)
6-311+G(2df)	6-311+G(2df)	-113.133212	296.8822
uC-6-311+G(2df)	6-311+G(2df)	-113.134520	296.4147
	uC-6-311+G(2df)	-113.142902	296.3908
u-6-311+G(2df)	6-311+G(2df)	-113.136044	296.3176
	u-6-311+G(2df)	-113.122722	296.2963
6-311+G(3df)	6-311+G(3df)	-113.139488	296.7699
uC-6-311+G(3df)	6-311+G(3df)	-113.140761	296.3917
	uC-6-311+G(3df)	-113.142902	296.3908
u-6-311+G(3df)	6-311+G(3df)	-113.142200	296.2955
	u-6-311+G(3df)	-113.147177	296.2968
aug-cc-pVTZ	aug-cc-pVTZ	-113.1445197	296.6638
uC-aug-cc-pVTZ	aug-cc-pVTZ	-113.144989	296.4907
	uC-aug-cc-pVTZ	-113.145565	296.4911
u-aug-cc-pVTZ	aug-cc-pVTZ	-113.146335	296.3600
	u-aug-cc-pVTZ	-113.149699	296.3623
aug-cc-pVQZ	aug-cc-pVQZ	-113.171609	296.4291
uC-aug-cc-pVQZ	aug-cc-pVQZ	-113.171748	296.3163
	uC-aug-cc-pVQZ	-113.172132	296.3164
u-aug-cc-pVQZ	aug-cc-pVQZ	-113.171948	296.3048
	u-aug-cc-pVQZ	-113.172888	296.3053
aug-cc-pV5Z	aug-cc-pV5Z	-113.180060	296.3223
uC-aug-cc-pV5Z	aug-cc-pV5Z	-113.180078	296.2916

	uC-aug-cc-pV5Z	-113.180128	296.2916
u-aug-cc-pV5Z	aug-cc-pV5Z	-113.180193	296.2868
	u-aug-cc-pV5Z	-113.180538	296.2871
aug-cc-pV6Z	aug-cc-pV6Z	-113.183007	296.3027
uC-aug-cc-pV6Z	aug-cc-pV5Z	-113.183014	296.2871
	uC-aug-cc-pV6Z	-113.183029	296.2871
u-aug-cc-pV6Z	aug-cc-pV6Z	-113.183076	296.2846
	u-aug-cc-pV6Z	-113.183257	296.2847
aug-cc-pCV5Z	aug-cc-pCV5Z	-113.180920	296.2854
uC-aug-cc-pCV5Z	aug-cc-pCV5Z	-113.180916	296.2852
	uC-aug-cc-pCV5Z	-113.180917	296.2853
u-aug-cc-pCV5Z	aug-cc-pCV5Z	-113.180928	296.2850
	u-aug-cc-pCV5Z	-113.181004	296.2851

<sup>a</sup> Total energy for the neutral reference state. <sup>b</sup> Experimental IE for the C edge 296.2 eV is taken from Ref. [49]

TABLE IV: CO, oxygen edge. Total and ionization energies;  
fc-CVS-EOM-IP-CCSD with various basis sets.

Basis on O	Basis on C	CCSD energy <sup>a</sup> (a.u.)	IE <sup>b</sup> (eV)
6-311+G(2df)	6-311+G(2df)	-113.153403	543.7282
uC-6-311+G(2df)	6-311+G(2df)	-113.156050	543.1010
	uC-6-311+G(2df)	-113.159353	543.1019
u-6-311+G(2df)	6-311+G(2df)	-113.159201	542.9073
	u-6-311+G(2df)	-113.167966	543.2097
6-311+G(3df)	6-311+G(3df)	-113.160756	543.6443
uC-6-311+G(3df)	6-311+G(3df)	-113.163070	543.0814
	uC-6-311+G(3df)	-113.166218	543.0820
u-6-311+G(3df)	6-311+G(3df)	-113.166134	542.8913
	u-6-311+G(3df)	-113.193741	542.8975
aug-cc-pVTZ	aug-cc-pVTZ	-113.159233	543.4400
uC-aug-cc-pVTZ	aug-cc-pVTZ	-113.159820	543.2373
	uC-aug-cc-pVTZ	-113.166897	543.2388
u-aug-cc-pVTZ	aug-cc-pVTZ	-113.162900	542.9741
	u-aug-cc-pVTZ	-113.194170	542.9772
aug-cc-pVQZ	aug-cc-pVQZ	-113.199349	543.0147
uC-aug-cc-pVQZ	aug-cc-pVQZ	-113.199742	542.8764
	uC-aug-cc-pVQZ	-113.201736	542.8770
u-aug-cc-pVQZ	aug-cc-pVQZ	-113.200312	542.8504
	u-aug-cc-pVQZ	-113.220245	542.8509
aug-cc-pV5Z	aug-cc-pV5Z	-113.213669	542.8534
uC-aug-cc-pV5Z	aug-cc-pV5Z	-113.213736	542.8152
	uC-aug-cc-pV5Z	-113.214943	542.8152
u-aug-cc-pV5Z	aug-cc-pV5Z	-113.214077	542.8047

	u-aug-cc-pV5Z	-113.229862	542.8044
aug-cc-pV6Z	aug-cc-pV6Z	-113.222164	542.8169
uC-aug-cc-pV6Z	aug-cc-pV5Z	-113.222180	542.8010
	uC-aug-cc-pV6Z	-113.223047	542.8010
u-aug-cc-pV6Z	aug-cc-pV6Z	-113.222351	542.7953
	u-aug-cc-pV6Z	-113.233567	542.7948
aug-cc-pCV5Z	aug-cc-pCV5Z	-113.235384	542.8028
uC-aug-cc-pCV5Z	aug-cc-pCV5Z	-113.235385	542.8026
	uC-aug-cc-pCV5Z	-113.235389	542.8026
u-aug-cc-pCV5Z	aug-cc-pCV5Z	-113.235461	542.8012
	u-aug-cc-pCV5Z	-113.235509	542.8013

<sup>a</sup> Total energy for the neutral reference state. <sup>b</sup> Experimental IE for the O edge IE 542.5 eV is taken from [49]



### C. Simple hydrides and the effect of basis on hydrogen

Small hydrides, water, ammonia, and methane, represent 3 different edges in molecules with hydrogen atoms. We use this set to investigate the effect that the basis set on the H atoms has on the heavy atoms' core IEs. We compare our findings with those of the previous study[35]. Table V shows the results computed with the Pople (3df) and Dunning basis sets on the heavy atom, combined with the matching contracted bases on hydrogen.

The results in Table V show that the differences between the smaller bases are similar to the results for N<sub>2</sub> and CO and that the IEs converge from above to the basis-set limit. The basis-set limit results are slightly above the experimental values; the largest difference from the experiment is observed for carbon (0.18 eV). This is similar to the findings in Ref. [35].

In contrast to the observation in Ref. [35], that the effect of the basis set beyond triple- $\zeta$  is moderate (0.1 eV difference between the triple- $\zeta$  to quadruple- $\zeta$  results), we observe somewhat larger effects for this set ( $\sim 0.4$  eV), as well as for the N<sub>2</sub> and CO molecules discussed above. This difference is likely due to the slightly different treatment of core electrons in the ground-state optimization step in the CVS-EOM-CCSD and fc-CVS-EOM-CCSD (we also note that the structures used in Ref. [35] are slightly different). However, going from quadruple to quintuple- $\zeta$ , we observe a similar change of  $< 0.1$  eV. Thus, the results of both studies indicate near-convergence to the basis-set limit at the quintuple- $\zeta$  basis. As the largest basis in the present calculations, we use uC-aug-cc-pV5Z (the IEs drop by  $\sim 0.03$  eV upon uncontraction); below, we refer to these results as the basis-set limit.

Table VI collects the IEs computed with uC-aug-cc-pV5Z on the heavy atom and smaller bases on hydrogens. As expected, the effect of the hydrogen basis on the core IEs is not large. For example, using aug-cc-pVQZ or even aug-cc-pVTZ instead of aug-cc-pV5Z changes the IEs by less than 0.001 and 0.005 eV, respectively, while significantly reducing the number of basis functions. Thus, one may consider using a contracted triple- $\zeta$  basis (or even smaller) on hydrogens in calculations of larger molecules as a viable cost-saving strategy.

### D. Using mixed basis sets for molecules with multiple edges

In this section we further investigate the idea of using mixed bases sets. We use the CO molecule and employ a larger basis (uC-aug-cc-pV5Z) on the active edge, and a smaller basis

TABLE V: Core IEs for H<sub>2</sub>O, NH<sub>3</sub>, and CH<sub>4</sub>.

Basis on active edge	Basis on H	#b.f. (H)	Core IE (eV)		
			H <sub>2</sub> O	NH <sub>3</sub>	CH <sub>4</sub>
6-311+G(3df)	6-311G	3	540.9000	406.3149	291.0717
uC-6-311+G(3df)	6-311G	3	540.2500	405.8179	290.7117
aug-cc-pVTZ	aug-cc-pVTZ	23	540.6570	406.0944	290.9248
uC-aug-cc-pVTZ	aug-cc-pVTZ	23	540.4573	405.9537	290.8383
aug-cc-pVQZ	aug-cc-pVQZ	46	540.2110	405.7655	290.6862
uC-aug-cc-pVQZ	aug-cc-pVQZ	46	540.0857	405.6690	290.6209
aug-cc-pV5Z	aug-cc-pV5Z	80	540.0530	405.6505	290.6103
uC-aug-cc-pV5Z	aug-cc-pV5Z	80	540.0162	405.6197	290.5845
Experimental core IEs <sup>a</sup> (eV)			539.9	405.6	290.76

<sup>a</sup> Ref. [50].TABLE VI: Core IEs for H<sub>2</sub>O, NH<sub>3</sub>, and CH<sub>4</sub> computed with mixed basis sets.

Basis on active edge	Basis on H	#b.f. (H)	Core IE (eV)		
			H <sub>2</sub> O	NH <sub>3</sub>	CH <sub>4</sub>
uC-aug-cc-pV5Z	aug-cc-pV5Z	80	540.0162	405.6197	290.5845
	u-aug-cc-pVQZ	48	540.0151	405.6189	290.5843
	aug-cc-pVQZ	46	540.0148	405.6185	290.5839
	u-aug-cc-pVTZ	25	540.0107	405.6144	290.5805
	aug-cc-pVTZ	23	540.0103	405.6139	290.5795
	aug-cc-pVDZ	9	539.9974	405.5971	290.5617
	cc-pVDZ	5	539.9965	405.5958	290.5621

on the inactive edge. Table VII shows the results of these calculations. We observe that using a quadruple- $\zeta$  or even a triple- $\zeta$  basis on inactive edges leads to relatively small differences in IEs (less than 0.05 eV). However, the IEs no longer approach the basis-set limit from above. For example, the calculation with aug-cc-pVDZ on the inactive edge yields smaller IE than the calculation with aug-cc-pVQZ on the inactive edge, which indicates potential imbalance of such approach.

### E. Other cost-saving strategies

In larger molecules, using uC-aug-cc-pV5Z on all heavy atoms quickly becomes prohibitively expensive. For example, for molecules with just 4 second row atoms, the total number of basis functions in uC-aug-cc-pV5Z exceeds 500. Aside the obvious choice of using a smaller set, the cost of the calculations can be reduced by using single-precision execution[51] and truncation of virtual space using frozen natural orbitals approach[52, 53].

TABLE VII: Core IEs in CO computed with mixed basis sets on carbon and oxygen edges<sup>a</sup>.

Basis on C	Basis on O	CCSD energy <sup>b</sup> (a.u.)	IE <sup>a</sup> (eV)
	aug-cc-pVDZ	-113.118361	296.2136
uC-aug-cc-pV5Z	aug-cc-pVTZ	-113.159656	296.2589
	aug-cc-pVQZ	-113.174465	296.2832
	aug-cc-pV5Z	-113.180916	296.2852
aug-cc-pVDZ		-113.172430	542.7505
aug-cc-pVTZ	uC-aug-cc-pV5Z	-113.193331	542.7850
aug-cc-pVQZ		-113.206596	542.8045
aug-cc-pV5Z		-113.235385	542.8026

<sup>a</sup> The uC-aug-cc-pV5Z basis is used on active edges. <sup>b</sup> Total energy for the neutral reference state.

Using single-precision execution limits the convergence thresholds. Thus, because in the benchmark study we desire tight convergence for the EOM energies, here we use single precision for the CCSD step only. Because CCSD is the scaling-determining step in the EOM-IP-CCSD calculations, using single precision leads to noticeable speedup. The results are shown in Table VIII. In agreement with previous benchmark[51], the effect of using single-precision at the CCSD step is negligible.

The FNO results are collected in Table IX. We use an occupation criterion to control the truncation of the virtual space (for example, the FNO threshold of 99.99 % means that this much of the total natural occupation is recovered within the truncated orbital space). We observe that the errors due to virtual space truncation for a particular value of FNO threshold are similar for all three hydrides. The errors with FNO threshold of 99.99%, which amounts to freezing 27-28% of the virtual orbitals, are around 0.06 eV. This relatively large value illustrates the importance of virtual orbitals in describing the relaxation effects due to core ionization.

Combining the FNO approximation with single precision at the CCSD level leads to noticeable reduction of computational time in methane and ammonia (about 7-fold speedup), while the effect in water was much smaller.

## F. Acrolein

Acrolein (shown in Fig. 1) is an interesting model system with 3 chemically distinct carbon atoms: C1 is connected to 2 hydrogens and one carbon, C2 is connected to two carbons and one hydrogen, and C3 is connected to one hydrogen, one carbon, and one oxygen.

TABLE VIII: Core IEs for H<sub>2</sub>O, NH<sub>3</sub>, and CH<sub>4</sub> computed with single and double precision CCSD<sup>a</sup>.

Molecule	Precision	CCSD energy (a.u.) <sup>b</sup>	Core IE (eV)
H <sub>2</sub> O	Double	-76.35986226	540.0151
	Single	-76.35986228	540.0151
NH <sub>3</sub>	Double	-56.49013066	405.6189
	Single	-56.49013076	405.6189
CH <sub>4</sub>	Double	-40.44666824	290.5843
	Single	-40.44666822	290.5843

<sup>a</sup> Active edge basis: uC-aug-cc-pV5Z, H basis: uC-aug-cc-pVQZ. <sup>b</sup> Total energy for the neutral reference state. 8 decimal places are shown in order to demonstrate that the difference is only in the 8th decimal place. CCSD convergence thresholds in single-precision calculation:  $10^{-4}$  for amplitudes and  $10^{-5}$  for energies.

TABLE IX: Core IEs for H<sub>2</sub>O, NH<sub>3</sub>, and CH<sub>4</sub> computed with the FNO-based truncation of the virtual space.

Molecule <sup>a</sup>	FNO thresh <sup>b</sup>	Act. virt.	Frzn. virt.	CCSD energy (a.u.) <sup>c</sup>	$\Delta$ IE <sup>d</sup> (eV)
H <sub>2</sub> O	99.00	48	178	-76.348978	-0.6967
NH <sub>3</sub>		55	219	-56.481735	-0.5238
CH <sub>4</sub>		65	257	-40.440410	-0.5280
H <sub>2</sub> O	99.90	115	111	-76.358738	-0.1541
NH <sub>3</sub>		135	139	-56.489228	-0.1781
CH <sub>4</sub>		159	163	-40.446014	-0.1632
H <sub>2</sub> O	99.99	165	61	-76.359781	-0.0656
NH <sub>3</sub>		200	74	-56.490056	-0.0670
CH <sub>4</sub>		233	89	-40.446609	-0.0614

<sup>a</sup> Active edge basis: uC-aug-cc-pV5Z, H basis: uC-aug-cc-pVQZ. <sup>b</sup> Population threshold: this fraction of total natural occupation is recovered by the active virtual orbitals. <sup>c</sup> Total energy for the neutral reference state. <sup>d</sup> IE shift relative to the full orbital space values in Table VIII.

We use this molecule to test how multiple IEs corresponding to the same edge are described with different bases and test whether our observations based on CO are transferable to a larger molecule. The available experimental results for the carbon edge, reported as shifts relative to C1, are from Ref. [54].

Table X and XI collect the results obtained using aug-cc-pV5Z and uC-aug-cc-pV5Z for the active edge. We observe that for both edges uncontracting the core in this basis leads

TABLE X: Acrolein, core IEs for the oxygen edge.

Basis on C	Basis on O	Basis on H	CCSD energy <sup>a</sup> (a.u)	IE (eV)
	6-311+G(3df)		-191.605984	540.0669
6-311+G(3df)	uC-6-311+G(3df)	6-311G	-191.608320	539.5005
	u-6-311+G(3df)		-191.611347	539.3078
aug-cc-pVQZ		aug-cc-pVTZ	-191.712598	539.2886
aug-cc-pVTZ	aug-cc-pV5Z	aug-cc-pVTZ	-191.661197	539.2722
aug-cc-pVTZ		aug-cc-pVDZ	-191.643593	539.2678
aug-cc-pVQZ		aug-cc-pVTZ	-191.712657	539.2502
aug-cc-pVTZ	uC-aug-cc-pV5Z	aug-cc-pVTZ	-191.661307	539.2325
aug-cc-pVTZ		aug-cc-pVDZ	-191.643713	539.2278

<sup>a</sup> Total energy for the neutral reference state.

to 0.03-0.04 eV drop in IE. Let us first discuss the results for the oxygen edge. In these calculations, our largest basis for the inactive edge was aug-cc-pVQZ. Further reducing this basis to aug-cc-pVTZ leads to a change of 0.02 eV. The effect of the basis on the hydrogen is even smaller — for example, reducing the basis on hydrogens from triple- $\zeta$  to double- $\zeta$  changes the IEs by 0.005 eV only. The trend in IEs computed with Pople’s bases is similar to the previous cases. The results for u-6-311+G(3df)/6-311+g(3df)/6-311G are within 0.06 eV from the uC-aug-cc-pV5Z/aug-cc-pVQZ/aug-cc-pVTZ (our largest basis in this calculation). The total number of basis functions in these two calculations are 181 and 467, respectively.

The results for the carbon edge, shown in Table XI, follow similar trends. We observe that the chemical shifts (the difference between  $1s_C$  IEs relative to C1) converge with respect to the basis much faster than the absolute values. As said, the largest basis used in this calculation is uC-aug-cc-pV5Z/aug-cc-pVTZ/aug-cc-pVTZ. The results for u-6-311+G(3df)/6-311+g(3df)/6-311G are within 0.02 eV from that value. Comparing our best estimates to the experiment (only the shifts were reported in Ref. [54]), we note excellent agreement for C3 (2.63 eV versus 2.6 eV), however, for the C2 shift we consistently obtain  $\sim 0.3$  eV, versus zero shift reported in Ref. [54].

## G. Glycine

Glycine ( $C_2H_5NO_2$ , canonical form shown in Fig. 2) is a polyatomic molecule featuring multiple core IEs and three different edges[25, 55, 56]. Table XII shows the results for the mixed basis sets in which we used uC-aug-cc-pV5Z for the active edge and aug-cc-pVTZ

TABLE XI: Acrolein, core IEs for the carbon edge.

Basis on C	Basis on O	Basis on H	CCSD energy <sup>a</sup> (a.u)	IE (eV) <sup>b</sup>	Shift (eV) <sup>c</sup>
6-311+G(3df)	6-311+G(3df)	6-311G	-191.538887	291.9013	0.0
				292.1980	0.2967
				294.5433	2.6420
uC-6-311+G(3df)	6-311+G(3df)	6-311G	-191.541545	291.6330	0.0
				291.9216	0.2886
				294.2588	2.6258
u-6-311+G(3df)	6-311+G(3df)	6-311G	-191.545560	291.5079	0.0
				291.7872	0.2793
				294.1488	2.6409
aug-cc-pVQZ	aug-cc-pVTZ	aug-cc-pVTZ	-191.601682	291.6321	0.00
				291.9026	0.2705
				294.2649	2.6328
uC-aug-cc-pVQZ	aug-cc-pVTZ	aug-cc-pVTZ	-191.602010	291.5453	0.00
				291.8155	0.2702
				294.1730	2.6277
aug-cc-pV5Z	aug-cc-pVTZ	aug-cc-pVTZ	-191.611970	291.5364	0.00
				291.8071	0.2707
				294.1650	2.6286
aug-cc-pV5Z	aug-cc-pVTZ	aug-cc-pVDZ	-191.606131	291.5298	0.00
				291.7991	0.2693
				294.1585	2.6287
uC-aug-cc-pV5Z	aug-cc-pVTZ	aug-cc-pVTZ	-191.612021	291.5086	0.00
				291.7794	0.2708
				294.1359	2.6273
uC-aug-cc-pV5Z	aug-cc-pVTZ	aug-cc-pVDZ	-191.606239	291.5012	0.00
				291.7702	0.2690
				294.1286	2.6274

<sup>a</sup> Total energy for the neutral reference state. <sup>b</sup> The IEs are arranged in the order C2, C1, and C3, refer to Fig. 1. <sup>c</sup> Experimental shifts in IEs are 0.0, 0.0, and 2.6 with respect to C1.

basis for other heavy atoms. For the hydrogens, we used aug-cc-pVDZ and aug-cc-pVTZ. Similarly to the acrolein example, the difference in IEs between these calculations is 0.01 eV.

Table XIII shows the results with Pople's bases. As in other cases, we see that the u-6-311+G(3df)/6-311+G(3df)/6-311G results are within 0.07 eV from our best estimates. We also performed calculations with the fully uncontracted Pople bases on all atoms and, as in previous cases did not observe much difference (results not shown). Finally, Fig. 6 compares the selected results against the available experimental values and our best estimate. The IEs computed with u-6-311+G(3df)/6-311+G(3df)/6-311G are within 0.2 eV for oxygen and carbon edges, and 1 eV for nitrogen edge from the experiments. The shifts between C1/C2 and O1/O2 are also reproduced well.

TABLE XII: Glycine core IEs for all edges with mixed basis sets<sup>a</sup>.

C Edge			
Basis on H	CCSD energy <sup>b</sup> (a.u.)	IE C1 (eV)	IE C2 (eV)
aug-cc-pVDZ	-284.013614	292.5430	295.1946
aug-cc-pVTZ	-284.027506	292.5531	295.1988
O Edge			
Basis on H	CCSD energy <sup>b</sup> (a.u.)	IE O1 (eV)	IE O2 (eV)
aug-cc-pVDZ	-284.089201	538.6577	540.2119
aug-cc-pVTZ	-284.110647	538.6632	540.2216
N Edge			
Basis on H	CCSD Energy <sup>b</sup> (a.u.)	IE N (eV)	
aug-cc-pVDZ	-284.037027	406.5789	
aug-cc-pVTZ	-284.053208	406.5922	

<sup>a</sup> Active edge basis: uC-aug-cc-pV5Z, inactive edge basis: aug-cc-pVTZ. <sup>b</sup> Total energy for the neutral reference state.

TABLE XIII: Glycine core IEs with Pople basis sets.

C Edge <sup>a</sup>			
Basis on active edge	CCSD Energy <sup>b</sup> (a.u.)	IE C1 (eV)	IE C2 (eV)
6-311+G(2df)	-283.922367	293.0310	295.6453
uC-6-311+G(2df)	-283.924455	292.7187	295.3349
u-6-311+G(2df)	-283.927506	292.5870	295.2341
6-311+G(3df)	-283.941538	293.0077	295.5965
uC-6-311+G(3df)	-283.943366	292.7032	295.3202
u-6-311+G(3df)	-283.946117	292.5704	295.2269
O Edge <sup>a</sup>			
Basis	CCSD Energy (a.u.)	IE O1 (eV)	IE O2 (eV)
6-311+G(2df)	-283.986182	539.5400	541.0882
uC-6-311+G(2df)	-283.991309	538.9286	540.4764
u-6-311+G(2df)	-283.997698	538.7314	540.2612
6-311+G(3df)	-284.009612	539.4839	541.0594
uC-6-311+G(3df)	-284.014362	538.9247	540.4761
u-6-311+G(3df)	-284.020468	538.7336	540.2695
N Edge <sup>a</sup>			
Basis	CCSD Energy (a.u.)	IE N (eV)	
6-311+G(2df)	-283.964956	407.3069	
uC-6-311+G(2df)	-283.966768	406.8375	
u-6-311+G(2df)	-283.969069	406.6510	
6-311+G(3df)	-283.986780	407.2477	
uC-6-311+G(3df)	-283.988335	406.8096	
u-6-311+G(3df)	-283.990529	406.6262	

<sup>a</sup> See Fig. 2 for notations. <sup>b</sup> Total energy for the neutral reference state. <sup>c</sup> Inactive edge and H basis is the contracted version of the basis on active edge Experimental IEs: C1: 292.3 eV, C2: 295.2 eV, O1: 538.4 eV, O2: 540.2 eV, N: 405.4 eV(from Ref. [56]).

#### IV. CONCLUSION

We presented a computational study of basis-set effects in calculations of core-ionized states using a correlated method, fc-CVS-EOM-IP-CCSD. In agreement with previous studies, we observed that core-level states require higher-quality basis sets than valence states because of the large perturbation on the electronic structure due to removal of a core electron. Although the converged results can be obtained by using very large Dunning's bases, such as aug-cc-pCV5Z, we investigated a different strategy, that is, using core- and fully uncontracted basis sets. Our results indicate that this approach is much more effective. We



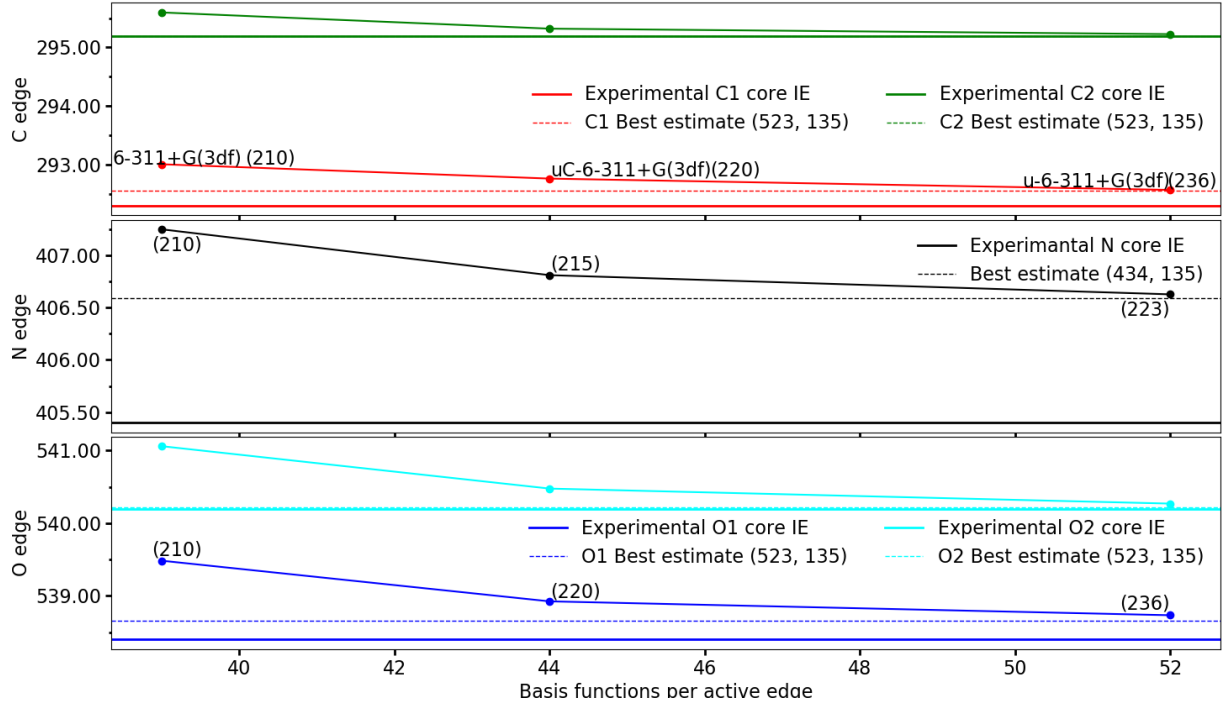


FIG. 6: Glycine IEs for the carbon, nitrogen and oxygen edges versus the number of basis functions per active edge atom. The total number of basis functions in each calculation is shown in parentheses in the respective panel. The best estimate is obtained with the uC-aug-cc-pV5Z/aug-cc-pVTZ/aug-cc-pVTZ basis; the respective total number of basis functions and the number of basis functions per active edge atom are shown in parentheses.

observe especially good performance for uncontracted Pople’s bases. For example, the results with u-6-311G+(3df) are of nearly the same quality as with aug-cc-pV5Z, despite having 60% less basis functions. For the systems we studied, the results with uC-6-311+G(3df) and u-6-311+G(3df) are within 0.07 eV from the basis-set limit. These errors are smaller than the anticipated errors due to an incomplete treatment of electron correlation. Slightly smaller bases, uC-6-311+G(2df) and u-6-311+G(2df) also perform very well. Thus, our recommended approach to core-level calculations is to use the uncontracted variants of the standard bases. The largest gain is achieved by uncontracting the core. The results show that it is sufficient to uncontract only the basis used for the active edge, while treating the rest of the atoms with matching contracted bases. Smaller bases can be used on hydrogens, without significant effect on the core IEs.

We also investigated more aggressive cost-saving strategies: using mixed bases on active and inactive edges, using single-precision at the CCSD step, and using the FNO-based truncation of the virtual space. The results pave a way towards cost-effective and accurate

calculations of core-level states.

### **Acknowledgments**

We are grateful to Prof. Peter Gill from the University of Sydney for his insightful remarks on the physics of core-ionized states and anticipated consequences for basis set selection, which motivated the present study.

This work was supported by the U.S. National Science Foundation (No. CHE-1856342). A.I.K. is a grateful recipient of the Simons Fellowship in Theoretical Physics and Mildred Dresselhaus Award from CFEL/DESY, which supported her sabbatical stay in Germany. M.L.V. and S.C. acknowledge financial support from DTU Chemistry and from the Independent Research Fund Denmark–Natural Sciences, DFF-RP2 grant no. 7014-00258B. S.C. also acknowledges the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 765739, “COSINE- European Training Network on COmputational Spectroscopy In Natural sciences and Engineering.”

### **Conflicts of interest**

A.I.K. is the President and a part-owner of Q-Chem, Inc.

- 
- [1] S. Mobilio, F. Boscherini, and C. Meneghini, editors, *Synchrotron Radiation: Basics, Methods and Applications*. Springer, 2014.
- [2] J.A. van Bokhoven and C. Lamberti, editors, *X-Ray Absorption and X-ray Emission Spectroscopy; Theory and Applications*. Wiley & Sons, 2016.
- [3] U. Bergmann, V. K. Yachandra, and J. Yano, editors, *X-Ray Free Electron Lasers: Applications in Materials, Chemistry and Biology*, Number 18 in Energy and Environment Series. Royal Society of Chemistry, 2017.
- [4] M. Nisoli, P. Decleva, F. Calegari, A. Palacios, and F. Martín, Attosecond electron dynamics in molecules, *Chem. Rev.* **117**, 10760 (2017).
- [5] M. Ahmed and O. Kostko, From atoms to aerosols: probing clusters and nanoparticles with synchrotron based mass spectrometry and X-ray spectroscopy, *Phys. Chem. Chem. Phys.* **22**, 2713 (2020).
- [6] O. Kostko, B. Bandyopadhyay, and M. Ahmed, Vacuum ultraviolet photoionization of complex chemical systems, *Annu. Rev. Phys. Chem.* **67**, 19 (2016).
- [7] P. Norman and A. Dreuw, Simulating X-ray spectroscopies and calculating core-excited states of molecules, *Chem. Rev.* **118**, 7208 (2018).
- [8] T. Fransson, Y. Harada, N. Kosugi, N. A. Besley, B. Winter, J. J. Rehr, L. G. M. Pettersson, and A. Nilsson, X-ray and electron spectroscopy of water, *Chem. Rev.* **116**, 7551 (2016).
- [9] J. A. Pople, Theoretical models for chemistry, in *Energy, Structure and Reactivity: Proceedings of the 1972 Boulder Summer Research Conference on Theoretical Chemistry*, edited by D.W. Smith and W.B. McRae, pages 51–61. Wiley, New York, 1973.
- [10] K. Emrich, An extension of the coupled-cluster formalism to excited states (I), *Nucl. Phys.* **A351**, 379 (1981).
- [11] J. F. Stanton and R. J. Bartlett, The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties, *J. Chem. Phys.* **98**, 7029 (1993).
- [12] A. I. Krylov, Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker’s guide to Fock space, *Annu. Rev. Phys. Chem.* **59**, 433 (2008).
- [13] R. J. Bartlett, The coupled-cluster revolution, *Mol. Phys.* **108**, 2905 (2010).

- [14] K. Sneskov and O. Christiansen, Excited state coupled cluster methods, *WIREs: Comput. Mol. Sci.* **2**, 566 (2012).
- [15] R. J. Bartlett, Coupled-cluster theory and its equation-of-motion extensions, *WIREs: Comput. Mol. Sci.* **2**, 126 (2012).
- [16] N.A. Besley, A.T.B. Gilbert, and P.M.W. Gill, Self-consistent-field calculations of core excited states, *J. Chem. Phys.* **130**, 124308 (2009).
- [17] M. Nooijen and R. J. Bartlett, Description of core-excitation spectra by the open-shell electron-attachment equation-of-motion coupled cluster method, *J. Chem. Phys.* **102**, 6735 (1995).
- [18] S. Coriani, O. Christiansen, T. Fransson, and P. Norman, Coupled-cluster response theory for near-edge x-ray-absorption fine structure of atoms and molecules, *Phys. Rev. A* **85**, 022507 (2012).
- [19] S. Coriani, T. Fransson, O. Christiansen, and P. Norman, Asymmetric-Lanczos-chain-driven implementation of electronic resonance convergent coupled-cluster linear response theory, *J. Chem. Theory Comput.* **8**, 1616 (2012).
- [20] T. J. Watson and R. J. Bartlett, Infinite order relaxation effects for core ionization energies with a variational coupled cluster ansatz, *Chem. Phys. Lett.* **555**, 235 (2013).
- [21] S. Sen, A. Shee, and D. Mukherjee, A study of the ionisation and excitation energies of core electrons using a unitary group adapted state universal approach, *Mol. Phys.* **11**, 2625 (2013).
- [22] D. Zuev, E. Vecharynski, C. Yang, N. Orms, and A. I. Krylov, New algorithms for iterative matrix-free eigensolvers in quantum chemistry, *J. Comput. Chem.* **36**, 273 (2015).
- [23] B. Peng, P.J. LeStrange, J.J. Goings, M. Caricato, and X. Li, Energy-specific equation-of-motion coupled-cluster methods for high-energy excited states: Application to K-edge X-ray absorption spectroscopy, *J. Chem. Theory Comput.* **11**, 4146 (2015).
- [24] S. Coriani and H. Koch, Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework, *J. Chem. Phys.* **143**, 181103 (2015).
- [25] A. Sadybekov and A. I. Krylov, Coupled-cluster based approach for core-ionized and core-excited states in condensed phase: Theory and application to different protonated forms of aqueous glycine, *J. Chem. Phys.* **147**, 014107 (2017).
- [26] R. H. Myhre, S. Coriani, and H. Koch, Near-edge X-ray absorption fine structure within

- multilevel coupled cluster theory, *J. Chem. Theory Comput.* **12**, 2633 (2016).
- [27] M. L. Vidal, X. Feng, E. Epifanovski, A. I. Krylov, and S. Coriani, A new and efficient equation-of-motion coupled-cluster framework for core-excited and core-ionized states, *J. Chem. Theory Comput.* **15**, 3117 (2019).
- [28] B. N. C. Tenorio, T. Moitra, M. A. C. Nascimento, A. B. Rocha, and S. Coriani, Molecular inner-shell photoabsorption/photoionization cross sections at core-valence-separated coupled cluster level: Theory and examples, *J. Chem. Phys.* **150**, 224104 (2019).
- [29] K. Nanda, M. L. Vidal, R. Faber, S. Coriani, and A. I. Krylov, How to stay out of trouble in RIXS calculations within the equation-of-motion coupled-cluster damped response theory framework? Safe hitchhiking in the excitation manifold by means of core-valence separation, *Phys. Chem. Chem. Phys.* **22**, 2629 (2020).
- [30] R. Faber and S. Coriani, Resonant inelastic X-ray scattering and nonresonant X-ray emission spectra from coupled-cluster (damped) response theory, *J. Chem. Theory Comput.* **15**, 520 (2019).
- [31] R. Faber and S. Coriani, Core-valence-separated coupled-cluster-singles-and-doubles complex-polarization-propagator approach to X-ray spectroscopies, *Phys. Chem. Chem. Phys.* **22**, 2642 (2020).
- [32] L. S. Cederbaum, W. Domcke, and J. Schirmer, Many-body theory of core holes, *Phys. Rev. A* **22**, 206 (1980).
- [33] S. Coriani and H. Koch, Erratum: "Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework" [*J. Chem. Phys.* 143, 181103 (2015)], *J. Chem. Phys.* **145**, 149901 (2016).
- [34] F. Frati, F. de Groot, J. Cerezo, F. Santoro, L. Cheng, R. Faber, and S. Coriani, Coupled cluster study of the x-ray absorption spectra of formaldehyde derivatives at the oxygen, carbon, and fluorine K-edges, *J. Chem. Phys.* **151**, 064107 (2019).
- [35] J. P. Carbonea, L. Cheng, R. H. Myhre, D. Matthews, H. Koch, and S. Coriani, *Advances in Quantum Chemistry*, volume 79, chapter An analysis of the performance of coupled cluster methods for K-edge core excitations and ionizations using standard basis sets, pages 241–261. Elsevier Inc, 2019.
- [36] J. Wenzel, M. Wormit, and A. Dreuw, Calculating core-level excitations and x-ray absorption spectra of medium-sized closed-shell molecules with the algebraic-diagrammatic construction

- scheme for the polarization propagator, *J. Comput. Chem.* **35**, 1900 (2014).
- [37] S. Shirai, S. Yamamoto, and S. Hyodo, Accurate calculation of core-electron binding energies: Multireference perturbation treatment, *J. Chem. Phys.* **121**, 7586 (2004).
- [38] Y. Takahata and D. P. Chong, DFT calculation of core-electron binding energies, *J. Electron. Spectrosc. Relat. Phenom.* **133**, 69 (2003).
- [39] J. Wenzel, A. Holzer, M. Wormit, and A. Dreuw, Analysis and comparison of CVS-ADC approaches up to third order for the calculation of core-excited states, *J. Chem. Phys.* **142**, 214104 (2015).
- [40] J. Liu, D. Matthews, S. Coriani, and L. Cheng, Benchmark calculations of K-edge ionization energies for first-row elements using scalar-relativistic corevalence-separated equation-of-motion coupled-cluster methods, *J. Chem. Theory Comput.* **15**, 1642 (2019).
- [41] M. L. Vidal, A. I. Krylov, and S. Coriani, Dyson orbitals within the fc-CVS-EOM-CCSD framework: theory and application to X-ray photoelectron spectroscopy of ground and excited states, *Phys. Chem. Chem. Phys.* **22**, 2693 (2020).
- [42] M. L. Vidal, A. I. Krylov, and S. Coriani, Correction to: "Dyson orbitals within the fc-CVS-EOM-CCSD framework: theory and application to X-ray photoelectron spectroscopy of ground and excited states", *Phys. Chem. Chem. Phys.* **22**, 3744 (2020).
- [43] J. C. Slater, Atomic shielding constants, *Phys. Rev.* **36**, 57 (1930).
- [44] L. Kjellsson, K. Nanda, J.-E. Rubensson, G. Doumy, S. H. Southworth, P. J. Ho, A. M. March, A. Al Haddad, Y. Kumagai, M.-F. Tu, T. Debnath, M. S. Bin Mohd Yusof, C. Arnold, W. F. Schlotter, S. Moeller, G. Coslovich, J. D. Koralek, M. P. Minitti, M. L. Vidal, M. Simon, R. Santra, Z.-H. Loh, S. Coriani, A. I. Krylov, and L. Young, RIXS reveals hidden local transitions of the aqueous OH radical, *Phys. Rev. Lett.* (2020), submitted; <https://arxiv.org/abs/2003.03909>.
- [45] Y. Shao, Z. Gan, E. Epifanovsky, A.T.B. Gilbert, M. Wormit, J. Kussmann, A.W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P.R. Horn, L.D. Jacobson, I. Kaliman, R.Z. Khaliullin, T. Kus, A. Landau, J. Liu, E.I. Proynov, Y.M. Rhee, R.M. Richard, M.A. Rohrdanz, R.P. Steele, E.J. Sundstrom, H.L. Woodcock III, P.M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G.J.O. Beran, Y.A. Bernard, E. Berquist, K. Brandhorst, K.B. Bravaya, S.T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S.H. Chien, K.D. Closser, D.L. Crittenden, M. Diedenhofen, R.A. DiStasio Jr., H. Do, A.D. Dutoi, R.G. Edgar, S.

- Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M.W.D. Hanson-Heine, P.H.P. Harbach, A.W. Hauser, E.G. Hohenstein, Z.C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyayev, J. Kim, J. Kim, R.A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C.M. Krauter, K.U. Laog, A. Laurent, K.V. Lawler, S.V. Levchenko, C.Y. Lin, F. Liu, E. Livshits, R.C. Lochan, A. Luenser, P. Manohar, S.F. Manzer, S.-P. Mao, N. Mardirossian, A.V. Marenich, S.A. Maurer, N.J. Mayhall, C.M. Oana, R. Olivares-Amaya, D.P. O'Neill, J.A. Parkhill, T.M. Perrine, R. Peverati, P.A. Pieniazek, A. Prociuk, D.R. Rehn, E. Rosta, N.J. Russ, N. Sergueev, S.M. Sharada, S. Sharma, D.W. Small, A. Sodt, T. Stein, D. Stuck, Y.-C. Su, A.J.W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M.A. Watson, J. Wenzel, A. White, C.F. Williams, V. Vanovschi, S. Yeganeh, S.R. Yost, Z.-Q. You, I.Y. Zhang, X. Zhang, Y. Zhou, B.R. Brooks, G.K.L. Chan, D.M. Chipman, C.J. Cramer, W.A. Goddard III, M.S. Gordon, W.J. Hehre, A. Klamt, H.F. Schaefer III, M.W. Schmidt, C.D. Sherrill, D.G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A.T. Bell, N.A. Besley, J.-D. Chai, A. Dreuw, B.D. Dunietz, T.R. Furlani, S.R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D.S. Lambrecht, W.Z. Liang, C. Ochsenfeld, V.A. Rassolov, L.V. Slipchenko, J.E. Subotnik, T. Van Voorhis, J.M. Herbert, A.I. Krylov, P.M.W. Gill, and M. Head-Gordon, Advances in molecular quantum chemistry contained in the Q-Chem 4 program package, *Mol. Phys.* **113**, 184 (2015).
- [46] A. I. Krylov and P. M. W. Gill, Q-Chem: An engine for innovation, *WIREs: Comput. Mol. Sci.* **3**, 317 (2013).
- [47] P. U. Manohar, J. F. Stanton, and A. I. Krylov, Perturbative triples correction for the equation-of-motion coupled-cluster wave functions with single and double substitutions for ionized states: Theory, implementation, and examples, *J. Chem. Phys.* **131**, 114112 (2009).
- [48] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus, New basis set exchange: An open, up-to-date resource for the molecular sciences community, *J. Chem. Inf. Model.* **59**, 4814 (2019).
- [49] W.L. Jolly, K.D. Bomben, and C.J. Eyermann, Core-electron binding energies for gaseous atoms and molecules, *Atomic Data and Nuclear Data Tables* **31**, 433 (1984).
- [50] J. Schirmer, A. Trofimov, K. Randall, J. Feldhaus, A. M. Bradshaw, Y. Ma, C.T. Chen, and F. Sette, K-shell excitation of the water, ammonia, and methane molecules using high-resolution photoabsorption spectroscopy, *Phys. Rev. A* **47**, 1136 (1993).

- [51] P. Pokhilko, E. Epifanovskii, and A. I. Krylov, Double precision is not needed for many-body calculations: Emergent conventional wisdom, *J. Chem. Theory Comput.* **14**, 4088 (2018).
- [52] A. Landau, K. Khistyayev, S. Dolgikh, and A. I. Krylov, Frozen natural orbitals for ionized states within equation-of-motion coupled-cluster formalism, *J. Chem. Phys.* **132**, 014109 (2010).
- [53] P. Pokhilko, D. Izmodenov, and A. I. Krylov, Extension of frozen natural orbital approximation to open-shell references: Theory, implementation, and application to single-molecule magnets, *J. Chem. Phys.* **152**, 034105 (2020).
- [54] D. Duflot, J.-P. Flament, I. C. Walker, J. Heinesch, and M.-J. Hubin-Franskin, Core shell excitation of 2-propenal (acrolein) at the O 1s and C 1s edges: An experimental and ab initio study, *J. Chem. Phys.* **118**, 1137 (2003).
- [55] R. H. Myhre, S. Coriani, and H. Koch, X-ray and UV spectra of glycine within coupled cluster linear response theory, *J. Phys. Chem. A* **123**, 9701 (2019).
- [56] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, and V. Carravetta, Investigation of the amino acids glycine, proline, and methionine by photoemission spectroscopy, *J. Phys. Chem. A* **111**, 10998 (2007).