

Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/tmph20>

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

Ronit Sarangi, Marta L. Vidal, Sonia Coriani & Anna I. Krylov

To cite this article: Ronit Sarangi, Marta L. Vidal, Sonia Coriani & Anna I. Krylov (2020) On the basis set selection for calculations of core-level states: different strategies to balance cost and accuracy, Molecular Physics, 118:19-20, e1769872, DOI: [10.1080/00268976.2020.1769872](https://doi.org/10.1080/00268976.2020.1769872)

To link to this article: <https://doi.org/10.1080/00268976.2020.1769872>



Published online: 28 May 2020.



Submit your article to this journal [↗](#)



Article views: 215



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 13 View citing articles [↗](#)

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

Ronit Sarangi ^a, Marta L. Vidal ^b, Sonia Coriani ^b and Anna I. Krylov ^a

^aDepartment of Chemistry, University of Southern California, Los Angeles, CA, USA; ^bDTU Chemistry – Department of Chemistry, Technical University of Denmark, Lyngby, Denmark

ABSTRACT

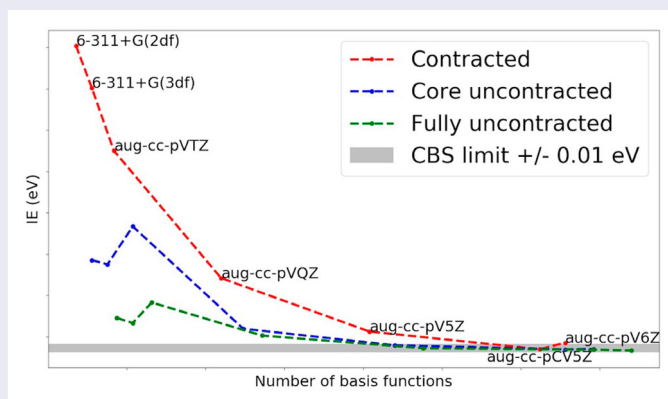
We present a study on basis set effects in correlated calculations of core-level states. While it is well recognised that the core-level states require using more extensive basis sets than their valence counterparts, the standard strategy has been to use large contracted basis sets, such as the cc-pVXZ or cc-pCVXZ series. Building upon the ideas of Besley *et al.* [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 numeric results. We also discuss other cost-saving strategies, such as virtual space truncation and mixed precision execution.

ARTICLE HISTORY

Received 3 April 2020
Accepted 8 May 2020

KEYWORDS

Benchmark; core ionized states; coupled cluster theory; basis set selection; core-valence separation; contracted basis sets



1. 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 the chemical environment. Compact and localised 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 modelling 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, modelling photoelectron spectra entails calculations 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-ionised and core-ionised 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 modelling. Challenges in correlated treatment of core-level states and possible solutions have been analyzed in recent reviews [7,8] and original research papers [16–34]. Particularly effective is the extension of the EOM-CC methods to core-level states via the core-valence separation (CVS) scheme [24,25,29,30,35,36]. Numerous benchmarks illustrated that CVS-EOM-IP/EE-CCSD [24,25,27,29,30,36,37] provides an effective and reliable description of core-ionised and core-excited states, including treatment of non-linear optical properties such as RIXS cross sections [31–33,38].

The special requirements to one-electron basis sets in calculations of core-level states have been recognised and documented in many papers [16,20,39–52]. 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 already in uncorrelated calculations, e.g. at the Hartree-Fock or Kohn-Sham DFT levels [16,46,48,50–52]. 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. Several studies pointed out that at the Hartree-Fock or DFT Δ SCF levels, the cc-pCVTZ basis [53], designed to describe core-valence correlation effects, delivers good performance for core-ionised states [16,39,48,49]. This is a considerably larger basis than typically used in uncorrelated calculations; hence, a number of strategies towards designing basis sets capable of providing a balanced treatment of the parent neutral and target core-ionised species have been explored. In particular, IGLO bases (individual gauge for localised orbitals), originally developed to improve the description of the electron density around the nuclei (which is needed for NMR spectroscopy), were found to provide good balance between accuracy and

cost [40,47,50]. A more general strategy was introduced by Besley and coworkers [50], who have shown that core relaxation effects can be effectively described by augmenting standard bases with functions for the next highest nuclear charge ($Z + 1$) than the element that is being ionised. By using $Z + 1$ augmenting functions, Δ SCF core-ionisation energies computed with double- ζ bases (6-31G* and cc-pVDZ) are within 0.5 eV from the cc-pCVQZ results [50]. This idea was further developed by Ambrose and Jensen, who proposed to use functions with interpolated exponents (between Z and $Z + 1$) within polarisation-consistent basis sets [51]. They observed a near-optimal balance of treating the neutral and core-ionised states with bases augmented by $Z + \frac{1}{2}$ functions.

In correlated calculations, the basis-set requirements are higher, as the basis should be sufficiently flexible to treat both orbital relaxation and electron correlation in the parent and target species. Thus, an optimal basis should afford a more flexible description of the core and a balanced treatment of the electron correlation. In the previous CVS-EOM-CC benchmark studies, series of standard contracted basis sets have been tested [36,41]. In Ref. [41], 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 ionisation 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 coworkers [41] exploited this smooth convergence of the results to extrapolate the computed excitation energies to the complete basis set (CBS) limit. They noted that for ionisation energies the aug-cc-pCV5Z basis (the largest used for ionisation energies in their study) appears to be close to the convergence limit, judging from the small differences between this and smaller bases. The authors also noted good performance of (aug)-cc-pCVTZ: the core IEs were within 0.1 eV from the aug-cc-pCV5Z results. While additional diffuse functions are required to properly describe core-excited states, they were found to be less important for core-ionised states. The relativistic effects were found to be less sensitive to the basis set [16,36]. These observations confirm that the main reason for extended basis sets in core-level calculations is orbital relaxation.

Here we systematically explore an alternative strategy, used by Gill and coworkers [16] and by us in recent applications [31,38]. Instead of following the hierarchy of Dunning's bases, optimised 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 [54,55] is much more effective than using Dunning's bases [53,56,57]. 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.

2. Study design, theoretical methods, and computational protocols

In this study, we focus on the calculations of ionised states using the fc-CVS-EOM-IP-CCSD method [29,58,59], with the goal of investigating the ability of various basis sets to describe orbital relaxation effects at a correlated level of theory. By focussing on core-ionised 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 rationale 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 is the rationale behind the $Z + 1$ approach of Besley [50,51]). 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- ζ (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 [60], 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 polarised-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. [16,31,38]. 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- ζ quality and above and their partially or fully decontracted variants. The full description of the basis sets is given below.

2.1. Computational details

All calculations were carried out using the *Q-Chem* electronic structure program [61,62]. We employ the fc-CVS-EOM-IP-CCSD method [29] in which the target ionised 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 ionising 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 Φ_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 [29]: the edge of interest and all lower edges are frozen at the CCSD step and active in the EOM calculation while all higher edges are treated normally. In this study we focus on 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 at the nitrogen edge, only oxygen and nitrogen $1s$ orbitals are frozen while carbon's $1s$ orbitals are active. Likewise, in calculations at the oxygen edge, only $1s_O$ 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_2), three

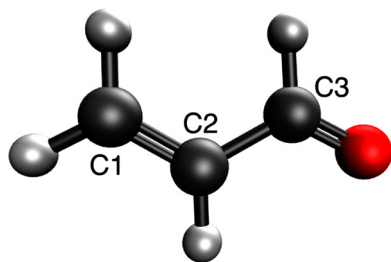


Figure 1. Acrolein structure with atom labels.

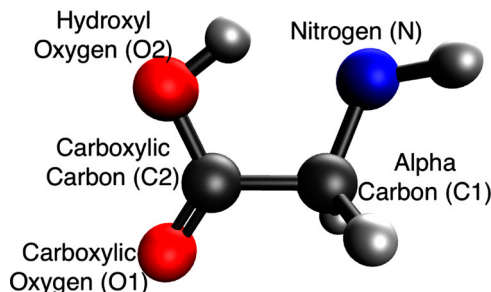


Figure 2. Glycine structure (canonical isomer) with atom labels.

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. [63] ($R_{\text{NN}} = 1.097685 \text{ \AA}$ and $R_{\text{CO}} = 1.128323 \text{ \AA}$). The hydrides' structures were taken from Ref. [26], where they were optimised with RI-MP2/cc-pVTZ. For acrolein (Figure 1), we used an MP2/cc-pVTZ optimised structure. Glycine calculations were performed for the canonical isomer (the main form of the gas-phase glycine) using the RI-MP2/cc-pVTZ optimised structure taken from Ref. [26] (Figure 2).

We used *Q-Chem*'s default convergence thresholds, except for the EOM amplitudes for 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 Section 3.5), 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- ζ basis into a 5- ζ one. The redundant basis

functions, which appear in decontracted Dunning's sets, were removed from the calculations. We note that in the segmented bases with optimised contractions (such as Pople's bases), there is a significant overlap between the exponents of the primitives in the contracted core and in the valence functions; thus, one may expect that further optimisation of fully uncontracted bases is possible. In this study we only removed exactly redundant functions and did not attempt to remove strongly overlapping ones.

Table 1 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 short-hand notations for the uncontracted bases. We used pure angular momentum functions (5d, 7f, 9g, 11h, ...) for all bases. For Dunning's bases, we used the versions with optimised contraction, as implemented in *Q-Chem*. The aug-cc-pCV5Z and aug-cc-pV6Z bases were taken from the Basis Set Exchange database [64], without optimising the general contractions (numeric tests indicated that using the variants of these bases with optimised general contractions leads to essentially the same results).

3. Results and discussion

3.1. Nitrogen molecule, N_2

The results for N_2 are collected in Table 2 and shown graphically in Figure 3. Table 2 shows the total CCSD energy of the neutral reference state and two core IEs, corresponding to ionisation 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 [41], the IEs decrease monotonously in the series of contracted basis sets of increasing size. Here we observe that the IEs also generally decrease upon uncontraction. This is a manifestation of core-relaxation effects, which lower the energy of the target ionised 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 polarised-core basis is used, because these bases already afford sufficient flexibility in describing core electrons.

Table 1. Basis sets, contraction schemes, and the number of functions per atom^a.

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	(11s6p3d2f)/[11s6p3d2f]	58
aug-cc-pVQZ	Original	(13s7p4d3f2g)/[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-pCVTZ	Original	(13s8p4d2f)/[7s6p4d2f]	59
uC-aug-cc-pCVTZ	Core-uncontracted	(13s8p4d2f)/[13s6p4d2f]	65
u-aug-cc-pCVTZ	Fully uncontracted	(13s8p4d2f)/[13s8p4d2f]	71
aug-cc-pCVQZ	Original	(16s10p6d4f2g)/[9s8p6d4f2g]	109
uC-aug-cc-pCVQZ	Core-uncontracted	(16s10p6d4f2g)/[16s8p6d4f2g]	116
u-aug-cc-pCVQZ	Fully uncontracted	(16s10p6d4f2g)/[16s10p6d4f2g]	122
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

^aFor a 2nd row element (C, N, O, etc).**Table 2.** Core IEs for N₂, fc-CVS-EOM-IP-CCSD with various basis sets.

Basis	CCSD energy ^a (a.u.)	IE1 ^b (eV)	IE2 (eV)	ΔIE (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
u-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-pCVTZ	-109.365969	410.0192	409.9146	0.1046
uC-aug-cc-pCVTZ	-109.366344	410.0033	409.8987	0.1046
u-aug-cc-pCVTZ	-109.369113	409.9582	409.8536	0.1046
aug-cc-pCVQZ	-109.388436	409.8997	409.7946	0.1051
uC-aug-cc-pCVQZ	-109.388576	409.8955	409.7905	0.1050
u-aug-cc-pCVQZ	-109.388920	409.8912	409.7861	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

^aTotal energy for the neutral reference state.^bExperimental IE_{σ_u} = 409.9 eV is taken from Ref. [26].

The results in Table 2 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 optimised for the core description (although they are optimised 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.

Figure 3 shows the lower core IE (σ_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 basis sets), the difference between contracted and core-uncontracted bases is remarkable. For example, uC-aug-cc-pVQZ gives better results 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

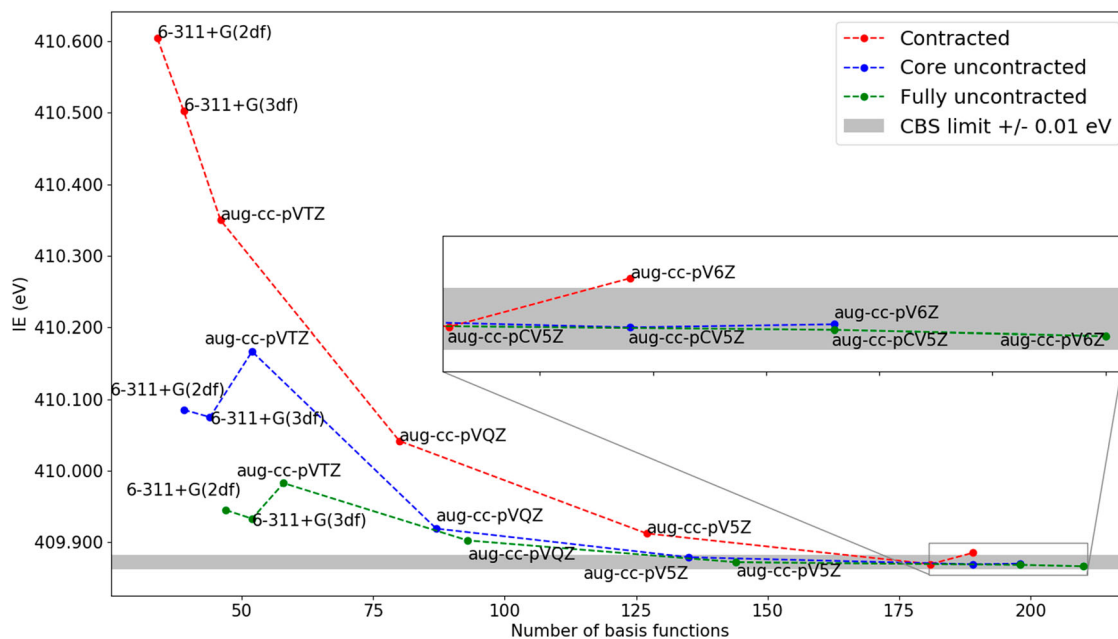


Figure 3. Core IEs for N_2 versus the number of basis functions per atom. Gray shaded area marks ± 0.01 eV interval around the basis-set limit (u-aug-cc-pV5Z).

(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 2). 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.

Dunning's bases with core-valence correlation (aug-cc-pCVTZ and aug-cc-pCVQZ, Table 2) perform better than their aug-cc-pVXZ counterparts, but are less effective than the uncontracted Pople's bases, e.g. the aug-cc-pCVTZ result is within 0.16 eV from the basis-set limit, to be compared with 0.06 eV for u-6-311+G(3df), despite the latter having fewer basis functions (52 versus 59, see Table 1).

3.2. Carbon monoxide, CO

The results for CO are collected in Tables 3 and 4 and shown graphically in Figures 4 and 5. The experimental values were taken from Ref. [65]. Overall, the trends for both edges follow very closely the trends observed for N_2 , 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). Here again we observe that while the aug-cc-pCVXZ bases deliver better results than the respective aug-cc-pVXZ

variants, they are less effective than the uncontracted Pople bases.

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.

3.3. 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 [41]. Table 5 shows the results computed with the Pople and Dunning basis sets on the heavy atom, combined with the matching contracted bases on hydrogen.

The results in Table 5 show that the differences between the smaller bases are similar to the results for N_2 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. [41].

Table 3. CO, carbon edge. Total and ionisation energies; fc-CVS-EOM-IP-CCSD with various basis sets.

Basis on C	Basis on O	CCSD energy ^a (a.u.)	IE ^b (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)	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)	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)	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)	u-6-311+G(3df)	-113.147177	296.2968
aug-cc-pVTZ	aug-cc-pVTZ	-113.144520	296.6638
uC-aug-cc-pVTZ	aug-cc-pVTZ	-113.144989	296.4907
uC-aug-cc-pVTZ	uC-aug-cc-pVTZ	-113.145565	296.4911
u-aug-cc-pVTZ	aug-cc-pVTZ	-113.146335	296.3600
u-aug-cc-pVTZ	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	uC-aug-cc-pVQZ	-113.172132	296.3164
u-aug-cc-pVQZ	aug-cc-pVQZ	-113.171948	296.3048
u-aug-cc-pVQZ	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	uC-aug-cc-pV5Z	-113.180128	296.2916
u-aug-cc-pV5Z	aug-cc-pV5Z	-113.180193	296.2868
u-aug-cc-pV5Z	u-aug-cc-pV5Z	-113.180538	296.2871
aug-cc-pV6Z	aug-cc-pV6Z	-113.183007	296.3027
uC-aug-cc-pV6Z	aug-cc-pV6Z	-113.183014	296.2871
uC-aug-cc-pV6Z	uC-aug-cc-pV6Z	-113.183029	296.2871
u-aug-cc-pV6Z	aug-cc-pV6Z	-113.183076	296.2846
u-aug-cc-pV6Z	u-aug-cc-pV6Z	-113.183257	296.2847
aug-cc-pCVTZ	aug-cc-pCVTZ	-113.149006	296.3988
uC-aug-cc-pCVTZ	aug-cc-pCVTZ	-113.149135	296.3807
uC-aug-cc-pCVTZ	uC-aug-cc-pCVTZ	-113.149289	296.3805
u-aug-cc-pCVTZ	aug-cc-pCVTZ	-113.150101	296.3442
u-aug-cc-pCVTZ	u-aug-cc-pCVTZ	-113.152347	296.3453
aug-cc-pCVQZ	aug-cc-pCVQZ	-113.173548	296.3036
uC-aug-cc-pCVQZ	aug-cc-pCVQZ	-113.173603	296.2988
uC-aug-cc-pCVQZ	uC-aug-cc-pCVQZ	-113.173707	296.2988
u-aug-cc-pCVQZ	aug-cc-pCVQZ	-113.173777	296.2918
u-aug-cc-pCVQZ	u-aug-cc-pCVQZ	-113.174502	296.2926
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	uC-aug-cc-pCV5Z	-113.180917	296.2853
u-aug-cc-pCV5Z	aug-cc-pCV5Z	-113.180928	296.2850
u-aug-cc-pCV5Z	u-aug-cc-pCV5Z	-113.181004	296.2851

^aTotal energy for the neutral reference state.^bExperimental IE for the C edge, 296.2 eV, is taken from Ref. [65].

In contrast to the observation in Ref. [41], that the effect of the basis set beyond triple- ζ is moderate (0.1 eV difference between the triple- ζ to quadruple- ζ results), we observe somewhat larger effects for this set (~ 0.4 eV), as well as for the N₂ and CO molecules discussed above. This difference is likely due to the different treatment of core electrons in the ground-state optimisation step in the CVS-EOM-CCSD and fc-CVS-EOM-CCSD (we also note that the structures used in Ref. [41] are slightly different). However, going from quadruple to quintuple- ζ , 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- ζ basis. As the largest basis in the present calculations, we

Table 4. CO, oxygen edge. Total and ionisation energies; fc-CVS-EOM-IP-CCSD with various basis sets.

Basis on O	Basis on C	CCSD energy ^a (a.u.)	IE ^b (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)	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)	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)	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)	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	uC-aug-cc-pVTZ	-113.166897	543.2388
u-aug-cc-pVTZ	aug-cc-pVTZ	-113.162900	542.9741
u-aug-cc-pVTZ	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	uC-aug-cc-pVQZ	-113.201736	542.8770
u-aug-cc-pVQZ	aug-cc-pVQZ	-113.200312	542.8504
u-aug-cc-pVQZ	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	uC-aug-cc-pV5Z	-113.214943	542.8152
u-aug-cc-pV5Z	aug-cc-pV5Z	-113.214077	542.8047
u-aug-cc-pV5Z	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	uC-aug-cc-pV6Z	-113.223047	542.8010
u-aug-cc-pV6Z	aug-cc-pV6Z	-113.222351	542.7953
u-aug-cc-pV6Z	u-aug-cc-pV6Z	-113.233567	542.7948
aug-cc-pCVTZ	aug-cc-pCVTZ	-113.197758	543.0234
uC-aug-cc-pCVTZ	aug-cc-pCVTZ	-113.197912	543.0079
uC-aug-cc-pCVTZ	uC-aug-cc-pCVTZ	-113.198510	543.0079
u-aug-cc-pCVTZ	aug-cc-pCVTZ	-113.200001	542.9422
u-aug-cc-pCVTZ	u-aug-cc-pCVTZ	-113.202259	542.9441
aug-cc-pCVQZ	aug-cc-pCVQZ	-113.226705	542.8494
uC-aug-cc-pCVQZ	aug-cc-pCVQZ	-113.226810	542.8449
uC-aug-cc-pCVQZ	uC-aug-cc-pCVQZ	-113.226909	542.8450
u-aug-cc-pCVQZ	aug-cc-pCVQZ	-113.227430	542.8274
u-aug-cc-pCVQZ	u-aug-cc-pCVQZ	-113.227817	542.8283
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	uC-aug-cc-pCV5Z	-113.235389	542.8026
u-aug-cc-pCV5Z	aug-cc-pCV5Z	-113.235461	542.8012
u-aug-cc-pCV5Z	u-aug-cc-pCV5Z	-113.235509	542.8013

^aTotal energy for the neutral reference state.^bExperimental IE for the O edge, 542.5 eV, is taken from [65].

use uC-aug-cc-pV5Z (the IEs drop by ~ 0.03 eV upon uncontraction); below, we refer to these results as the basis-set limit.

Table 6 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- ζ basis (or even smaller) on hydrogens in calculations of larger molecules as a viable cost-saving strategy.

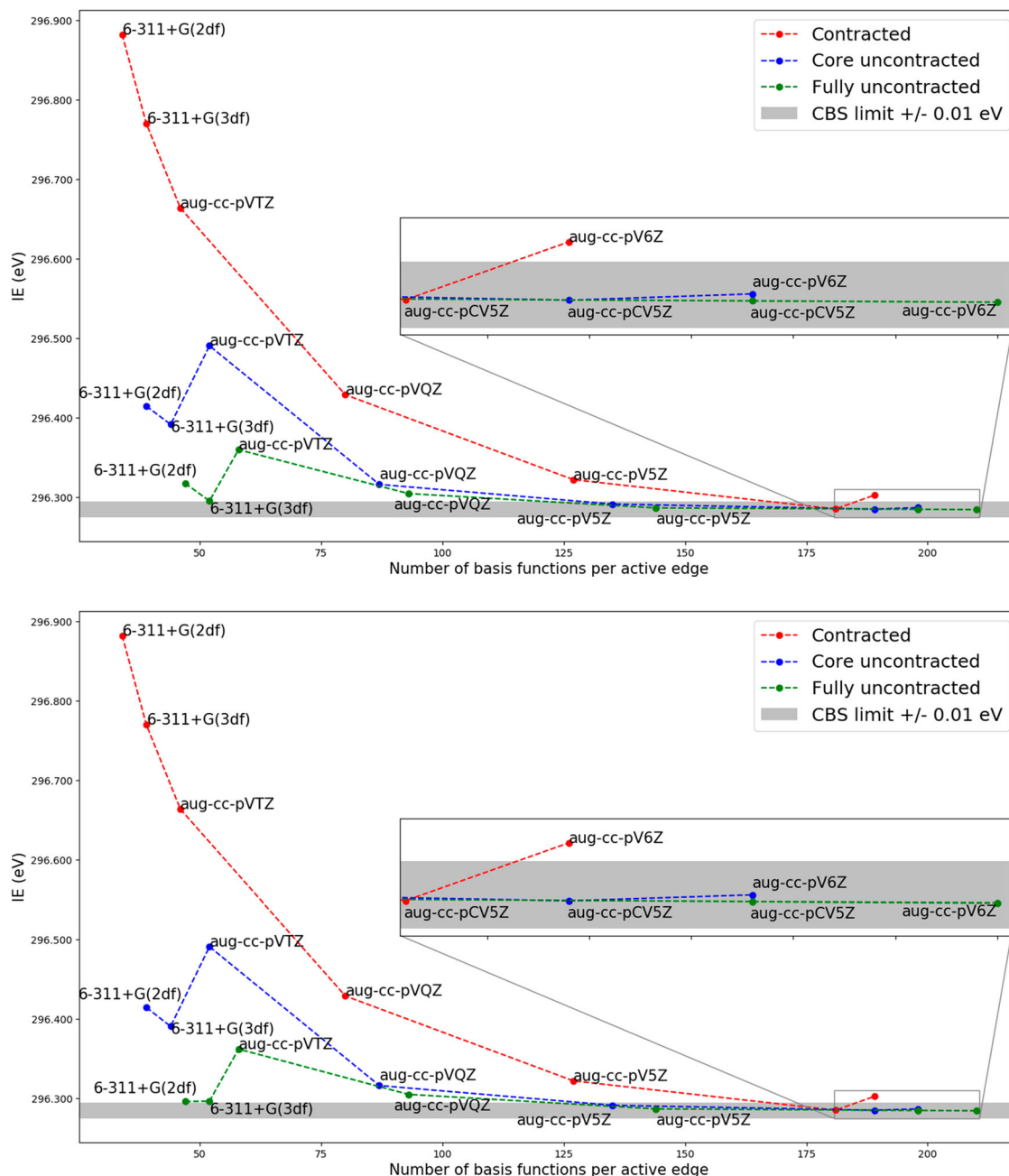


Figure 4. CO, carbon edge IEs. Top: Only the carbon basis is uncontracted, whereas the original matching basis is used for oxygen. Bottom: Both carbon and oxygen bases are uncontracted. Gray shaded area marks the ± 0.01 eV interval around the basis-set limit (u-aug-cc-pV5Z).

3.4. Using mixed basis sets for molecules with multiple edges

In this section we further investigate the idea of using mixed basis sets. This strategy has been exploited in numerous previous studies (see, for example, Refs. [46,50,52]), which have shown that the results are most sensitive to the basis set on the atom being ionised, while the rest of the atoms can be treated with standard bases. We use the CO molecule as an example and employ a larger basis (uC-aug-cc-pV5Z) on the active

edge, and a smaller basis on the inactive edge. Table 7 shows the results of these calculations. We observe that using a quadruple- ζ or even a triple- ζ 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.

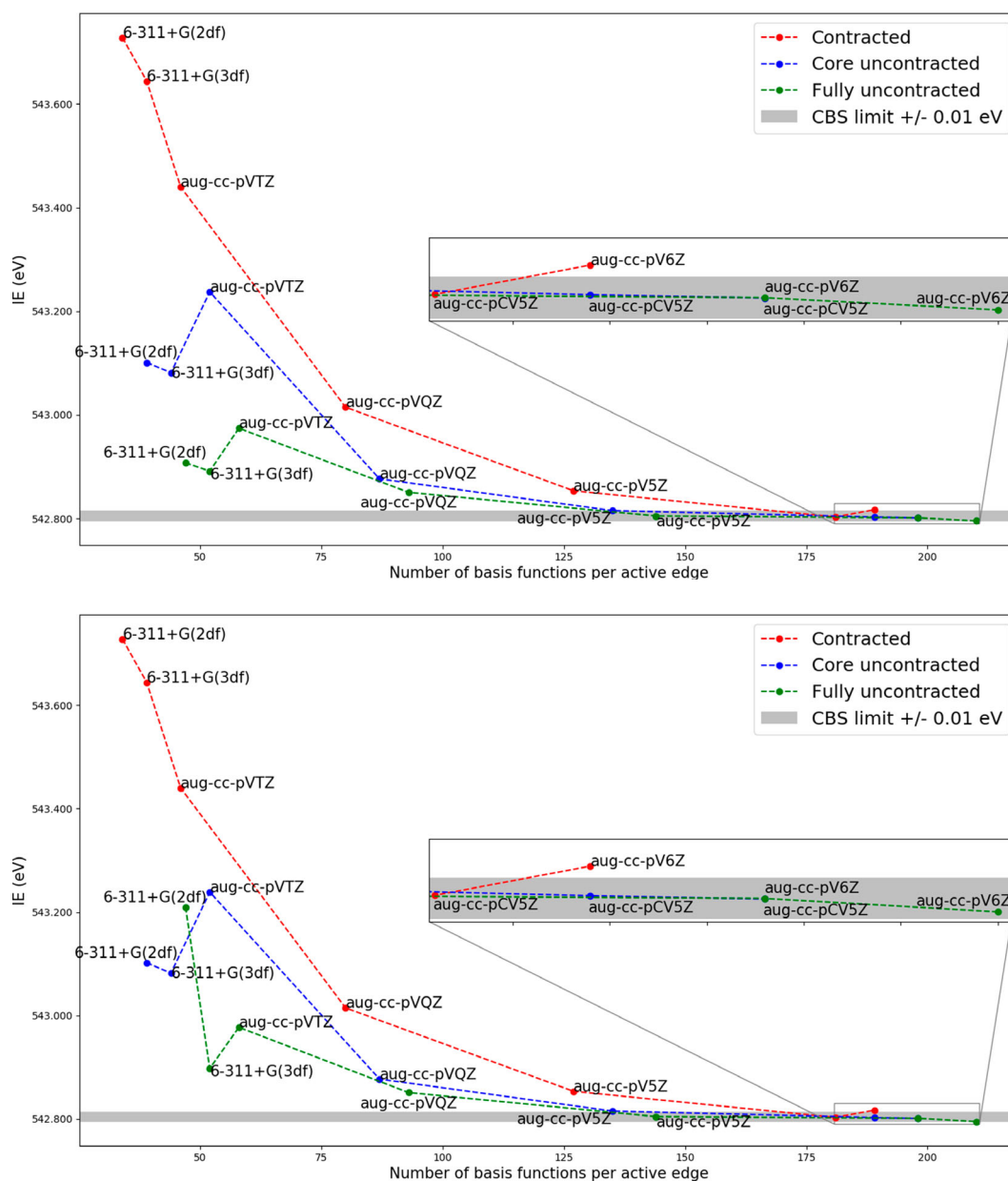


Figure 5. CO, oxygen edge IEs. Top: Only the oxygen basis is uncontracted, whereas the original matching basis is used for carbon. Bottom: Both carbon and oxygen bases are uncontracted. Gray shaded area marks the ± 0.01 eV interval around the basis-set limit (u-aug-cc-pV5Z).

Table 5. Core IEs for H₂O, NH₃, and CH₄ computed with mixed basis sets.

Basis on active edge	Basis on H	#b.f. (H)	Core IE (eV)		
			H ₂ O	NH ₃	CH ₄
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 ^a (eV)			539.9	405.6	290.76

^aRef. [66].

Table 6. Core IEs for H₂O, NH₃, and CH₄ computed with mixed basis sets.

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

Table 7. Core IEs in CO computed with mixed basis sets on carbon and oxygen edges^a.

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

^aThe uC-aug-cc-pV5Z basis is used on active edges.^bTotal energy for the neutral reference state.**Table 8.** Core IEs for H₂O, NH₃, and CH₄ computed with single and double precision CCSD^a.

Molecule	Precision	CCSD energy (a.u.) ^b	Core IE (eV)
H ₂ O	Double	−76.35986226	540.0151
	Single	−76.35986228	540.0151
NH ₃	Double	−56.49013066	405.6189
	Single	−56.49013076	405.6189
CH ₄	Double	−40.44666824	290.5843
	Single	−40.44666822	290.5843

^aActive edge basis: uC-aug-cc-pV5Z, H basis: uC-aug-cc-pVQZ.^bTotal 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.

3.5. 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

basis set, the cost of the calculations can be reduced by using single-precision execution [67] and truncation of virtual space using frozen natural orbitals (FNO) approach [68,69].

Using single-precision execution limits the convergence thresholds. Because in the benchmark study we

Table 9. Core IEs for H₂O, NH₃, and CH₄ computed with the FNO-based truncation of the virtual space.

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

^aActive edge basis: uC-aug-cc-pV5Z, H basis: uC-aug-cc-pVQZ.^bPopulation threshold: this fraction of total natural occupation is recovered by the active virtual orbitals.^cTotal energy for the neutral reference state.^dIE shift relative to the full orbital space values in Table 8.**Table 10.** Acrolein, core IEs for the oxygen edge.

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

^aTotal energy for the neutral reference state.

Table 11. Acrolein, core IEs for the carbon edge.

Basis on C	Basis on O	Basis on H	CCSD energy ^a (a.u)	IE (eV) ^b	Shift (eV) ^c
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

^aTotal energy for the neutral reference state.^bThe IEs are arranged in the order C2, C1, and C3, refer to Figure 1.^cExperimental shifts [70] in IEs are 0.0, 0.0, and 2.6 with respect to C1 (the experimental resolution is ± 0.17 eV).

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 8. In agreement with previous benchmarks [67], the loss of accuracy due to using single-precision at the CCSD step is negligible.

The FNO results are collected in Table 9. 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 ionisation.

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.

3.6. Acrolein

Acrolein (shown in Figure 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. 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. [70].

Tables 10 and 11 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 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- ζ to double- ζ 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 11, follow similar trends. We observe that the chemical shifts

Table 12. Glycine core IEs for all edges with mixed basis sets^a.

C Edge			
Basis on H	CCSD energy ^b (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 ^b (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 ^b (a.u.)	IE N (eV)	
aug-cc-pVDZ	−284.037027	406.5789	
aug-cc-pVTZ	−284.053208	406.5922	

^aActive edge basis: uC-aug-cc-pV5Z, inactive edge basis: aug-cc-pVTZ.^bTotal energy for the neutral reference state.

(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. [70]), we note excellent agreement for C3 (2.63 eV versus 2.6 eV), however, for the C2 shift we consistently obtain ~ 0.3 eV, versus near zero shift reported in Ref. [70]. We note that the experimental resolution in this study was 0.17 eV and that the calculations reported in the original paper also suggested a larger value of the shift.

Table 13. Glycine core IEs with Pople's basis sets.

C Edge ^a			
Basis on active edge	CCSD energy ^b (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 ^a			
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 ^a			
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	

^aSee Figure 2 for notations.^bTotal energy for the neutral reference state.^cInactive 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. [72]).

3.7. Glycine

Glycine ($C_2H_5NO_2$, canonical form shown in Figure 2) is a polyatomic molecule featuring multiple core IEs

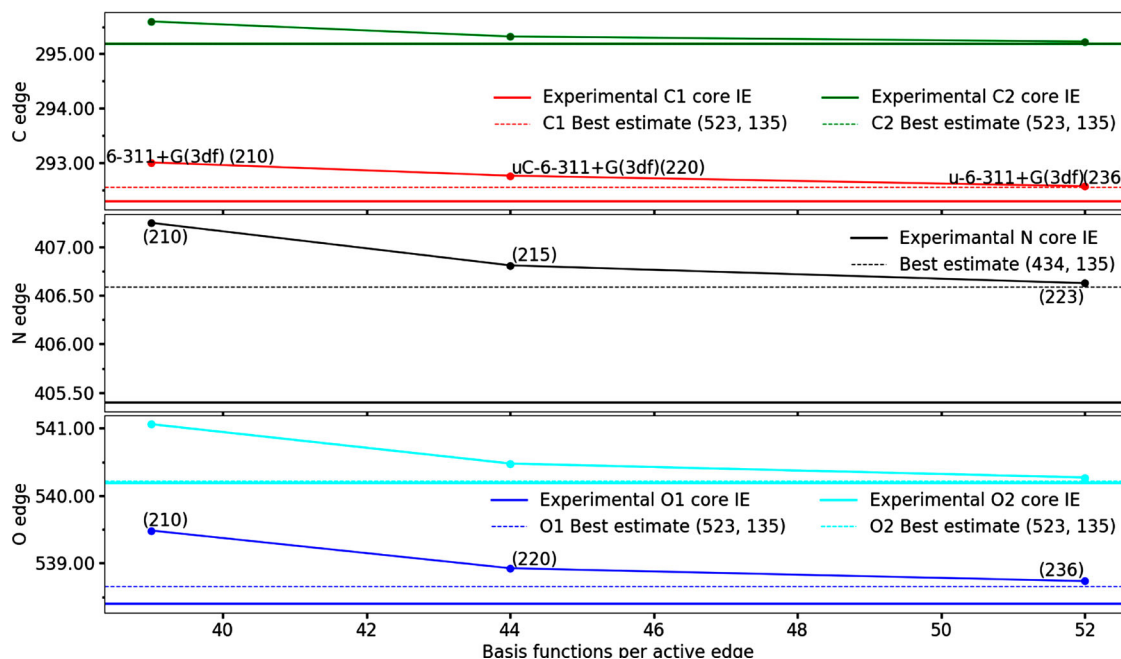


Figure 6. Glycine IEs for the carbon, nitrogen and oxygen edges versus the number of basis functions per active edge atom. The total number of basis functions in each calculation is shown in parentheses in the respective panel. The best estimate is obtained with the uC-aug-cc-pV5Z/aug-cc-pVTZ/aug-cc-pVTZ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

and three different edges [26,71,72]. Table 12 shows the results for the mixed basis sets in which we used uC-aug-cc-pV5Z for the active edge and aug-cc-pVTZ 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 13 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, Figure 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 experimental values. The shifts between C1/C2 and O1/O2 are also reproduced well.

4. Conclusion

We presented a computational study of basis-set effects in calculations of core-ionised 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 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% fewer 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. Future work entails investigation of basis-set effects for calculations of core-level states of heavier elements. While preliminary calculations confirm that uncontracting the standard bases improves the description of lower edges as well, detailed benchmarks including spin-orbit couplings are necessary for quantitative assessment of optimal basis set choices for heavier elements. This work is currently in progress.

Acknowledgments

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

Disclosure statement

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

Funding

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 - Department of Chemistry, Technical University of Denmark 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'.

ORCID

Ronit Sarangi  <http://orcid.org/0000-0002-5838-003X>

Marta L. Vidal  <http://orcid.org/0000-0003-0653-2078>

Sonia Coriani  <http://orcid.org/0000-0002-4487-897X>

Anna I. Krylov  <http://orcid.org/0000-0001-6788-5016>

References

- [1] S. Mobilio, F. Boscherini, and C. Meneghini, editors, *Synchrotron Radiation: Basics, Methods and Applications* (Springer, Berlin, Germany, 2014).
- [2] J.A. van Bokhoven and C. Lamberti, editors, *X-Ray Absorption and X-ray Emission Spectroscopy; Theory and Applications* (Wiley & Sons, Chichester, Germany, 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, *Chem. Rev.* **117**, 10760 (2017).

- [5] M. Ahmed and O. Kostko, *Phys. Chem. Chem. Phys.* **22**, 2713 (2020).
- [6] O. Kostko, B. Bandyopadhyay and M. Ahmed, *Annu. Rev. Phys. Chem.* **67**, 19 (2016).
- [7] P. Norman and A. Dreuw, *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, *Chem. Rev.* **116**, 7551 (2016).
- [9] J.A. Pople, 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 (Wiley, New York, 1973), pp. 51–61.
- [10] K. Emrich, *Nucl. Phys.* **A351**, 379 (1981).
- [11] J.F. Stanton and R.J. Bartlett, *J. Chem. Phys.* **98**, 7029 (1993).
- [12] A.I. Krylov, *Annu. Rev. Phys. Chem.* **59**, 433 (2008).
- [13] R.J. Bartlett, *Mol. Phys.* **108**, 2905 (2010).
- [14] K. Snegov and O. Christiansen, *WIREs: Comput. Mol. Sci.* **2**, 566 (2012).
- [15] R.J. Bartlett, *WIREs: Comput. Mol. Sci.* **2**, 126 (2012).
- [16] N.A. Besley, A.T.B. Gilbert and P.M.W. Gill, *J. Chem. Phys.* **130**, 124308 (2009).
- [17] M. Nooijen and R.J. Bartlett, *J. Chem. Phys.* **102**, 6735 (1995).
- [18] S. Coriani, O. Christiansen, T. Fransson and P. Norman, *Phys. Rev. A* **85**, 022507 (2012).
- [19] S. Coriani, T. Fransson, O. Christiansen and P. Norman, *J. Chem. Theory Comput.* **8**, 1616 (2012).
- [20] T.J. Watson and R.J. Bartlett, *Chem. Phys. Lett.* **555**, 235 (2013).
- [21] S. Sen, A. Shee and D. Mukherjee, *Mol. Phys.* **11**, 2625 (2013).
- [22] D. Zuev, E. Vecharynski, C. Yang, N. Orms and A.I. Krylov, *J. Comput. Chem.* **36**, 273 (2015).
- [23] B. Peng, P.J. LeStrange, J.J. Goings, M. Caricato and X. Li, *J. Chem. Theory Comput.* **11**, 4146 (2015).
- [24] S. Coriani and H. Koch, *J. Chem. Phys.* **143**, 181103 (2015).
- [25] S. Coriani and H. Koch, *J. Chem. Phys.* **145**, 149901 (2016).
- [26] A. Sadybekov and A.I. Krylov, *J. Chem. Phys.* **147**, 014107 (2017).
- [27] R.H. Myhre, S. Coriani and H. Koch, *J. Chem. Theory Comput.* **12**, 2633 (2016).
- [28] A.P. Bazante, A. Perera and R.J. Bartlett, *Chem. Phys. Lett.* **683**, 68 (2017).
- [29] M.L. Vidal, X. Feng, E. Epifanovsky, A.I. Krylov and S. Coriani, *J. Chem. Theory Comput.* **15**, 3117 (2019).
- [30] B.N.C. Tenorio, T. Moitra, M.A.C. Nascimento, A.B. Rocha and S. Coriani, *J. Chem. Phys.* **150**, 224104 (2019).
- [31] K. Nanda, M.L. Vidal, R. Faber, S. Coriani and A.I. Krylov, *Phys. Chem. Chem. Phys.* **22**, 2629 (2020).
- [32] R. Faber and S. Coriani, *J. Chem. Theory Comput.* **15**, 520 (2019).
- [33] R. Faber and S. Coriani, *Phys. Chem. Chem. Phys.* **22**, 2642 (2020).
- [34] Y.C. Park, A. Perera and R.J. Bartlett, *J. Chem. Phys.* **151**, 164117 (2019).
- [35] L.S. Cederbaum, W. Domcke and J. Schirmer, *Phys. Rev. A* **22**, 206 (1980).
- [36] J. Liu, D. Matthews, S. Coriani and L. Cheng, *J. Chem. Theory Comput.* **15**, 1642 (2019).
- [37] F. Frati, F. de Groot, J. Cerezo, F. Santoro, L. Cheng, R. Faber and S. Coriani, *J. Chem. Phys.* **151**, 064107 (2019).
- [38] L. Kjellsson, K. Nanda, J.-E. Rubensson, G. Doumy, S.H. Southworth, P.J. Ho, A.M. March, A. Al Hadad, 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, *Phys. Rev. Lett.* (2020), submitted; < <https://arxiv.org/abs/2003.03909> > .
- [39] G. Cavagliasso and D.P. Chong, *J. Chem. Phys.* **111**, 9485 (1999).
- [40] A. Mijovilovich, L.G.M. Pettersson, S. Mangold, M. Janousch, J. Susini, M. Salome, F.M.F. de Groot and B.M. Weckhuysen, *J. Phys. Chem. A* **113**, 2750 (2009).
- [41] J.P. Carbone, L. Cheng, R.H. Myhre, D. Matthews, H. Koch and S. Coriani, *Advances in Quantum Chemistry* (Elsevier Inc, 2019), Vol. 79, Chapt. 11, pp. 241–261.
- [42] S. Shirai, S. Yamamoto and S. Hyodo, *J. Chem. Phys.* **121**, 7586 (2004).
- [43] Y. Takahata and D.P. Chong, *J. Electron. Spectrosc. Relat. Phenom.* **133**, 69 (2003).
- [44] J. Wenzel, M. Wormit and A. Dreuw, *J. Comput. Chem.* **35**, 1900 (2014).
- [45] J. Wenzel, A. Holzer, M. Wormit and A. Dreuw, *J. Chem. Phys.* **142**, 214104 (2015).
- [46] B. Kovac, I. Ljubic, A. Kivimaki, M. Coreno and I. Novak, *Phys. Chem. Chem. Phys.* **16**, 10734–10742 (2014).
- [47] T. Fransson, I. Zhovtobriukh, S. Coriani, K.T. Wikfeldt, P. Norman and L.G.M. Pettersson, *Phys. Chem. Chem. Phys.* **18**, 566 (2016).
- [48] I. Tolbatov and D.M. Chipman, *Theor. Chem. Acc.* **136**, 82 (2017).
- [49] A.E.A. Fouda and N.A. Besley, *Theor. Chem. Acc.* **137**, 6 (2018).
- [50] M. Hanson-Heine, M. George and N. Besley, *Chem. Phys. Lett.* **699**, 279 (2018).
- [51] M.A. Ambrose and F. Jensen, *J. Chem. Theory Comput.* **15**, 325 (2019).
- [52] D. Hait and M. Head-Gordon, *J. Phys. Chem. Lett.* **11**, 775 (2020).
- [53] K.A. Peterson and T.H. Dunning, Jr., *J. Chem. Phys.* **117**, 10548 (2002).
- [54] R. Krishnan, J.S. Binkley, R. Seeger and J.A. Pople, *J. Chem. Phys.* **72**, 650 (1980).
- [55] M.J. Frisch, J.A. Pople and J.S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).
- [56] T.H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- [57] R.A. Kendall, T.H. Dunning, Jr. and R.J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- [58] M.L. Vidal, A.I. Krylov and S. Coriani, *Phys. Chem. Chem. Phys.* **22**, 2693 (2020).
- [59] M.L. Vidal, A.I. Krylov and S. Coriani, *Phys. Chem. Chem. Phys.* **22**, 3744 (2020).
- [60] J.C. Slater, *Phys. Rev.* **36**, 57 (1930).
- [61] 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, *Mol. Phys.* **113**, 184 (2015).
- [62] A.I. Krylov and P.M.W. Gill, *WIREs: Comput. Mol. Sci.* **3**, 317 (2013).
- [63] P.U. Manohar, J.F. Stanton and A.I. Krylov, *J. Chem. Phys.* **131**, 114112 (2009).
- [64] B.P. Pritchard, D. Altarawy, B. Didier, T.D. Gibson and T.L. Windus, *J. Chem. Inf. Model.* **59**, 4814 (2019).
- [65] W.L. Jolly, K.D. Bomben and C.J. Eyermann, *Atom Data Nucl Data Tables* **31**, 433 (1984).
- [66] J. Schirmer, A. Trofimov, K. Randall, J. Feldhaus, A.M. Bradshaw, Y. Ma, C.T. Chen and F. Sette, *Phys. Rev. A* **47**, 1136 (1993).
- [67] P. Pokhilko, E. Epifanovskii and A.I. Krylov, *J. Chem. Theory Comput.* **14**, 4088 (2018).
- [68] A. Landau, K. Khistyayev, S. Dolgikh and A.I. Krylov, *J. Chem. Phys.* **132**, 014109 (2010).
- [69] P. Pokhilko, D. Izmodenov and A.I. Krylov, *J. Chem. Phys.* **152**, 034105 (2020).
- [70] D. Duflot, J.-P. Flament, I.C. Walker, J. Heinesch and M.-J. Hubin-Franskin, *J. Chem. Phys.* **118**, 1137 (2003).
- [71] R.H. Myhre, S. Coriani and H. Koch, *J. Phys. Chem. A* **123**, 9701 (2019).
- [72] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K.C. Prince and V. Carravetta, *J. Phys. Chem. A* **111**, 10998 (2007).