Physical Chemistry Chemical Physics



# PCCP

# Efficient Basis Sets for Core-Excited States Motivated by Slater's Rules

Journal:	Physical Chemistry Chemical Physics
Manuscript ID	CP-ART-08-2021-003931.R2
Article Type:	Paper
Date Submitted by the Author:	25-Dec-2021
Complete List of Authors:	Qian, Jin; Lawrence Berkeley National Laboratory, Crumlin, Ethan; Lawrence Berkeley National Laboratory, Advanced Light Srouce Prendergast, David; Lawrence Berkeley National Laboratory, Molecular Foundry

SCHOLARONE<sup>™</sup> Manuscripts

# ARTICLE

Received 26th August 2021, Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

# Efficient Basis Sets for Core-Excited States Motivated by Slater's Rules

Jin Qian<sup>a,b,c</sup>, Ethan J. Crumlin<sup>a,b\*</sup>, David Prendergast<sup>c\*</sup>

X-ray photoemission spectroscopy is a commonly applied characterization technique that probes the local chemistry of atoms in molecules and materials via the photoexcitation of electrons from atomic core orbitals. These measurements can be interpreted by comparison with previous literature or through the calculation of core-electron binding energies (CEBEs) for model systems. However, physically and numerically accurate description of the core-excited electronic structures demand specializations beyond routine ground state setups. Inspired by Slater's rules, we focus on developing computationally efficient and physically motivated contractions to reproduce the core-excited atomic orbitals which led to improved numerical accuracy of calculated CEBEs. When applied to carbon 1*s* excitations in a wide range of molecules, these core-excited basis sets produce total energy differences ( $\Delta$ SCF) using a hybrid exact-exchange density functional (B3LYP) that can reproduce core-excitation energies within experimental accuracy (~0.1 eV). Due to missing relativistic effects, heavier elements (N, O) exhibit slightly larger systematic absolute errors, but still maintain a satisfactory 0.2 eV mean average error for relative CEBEs. We also connect the known variability in the core level binding energy with local atomic charge to demonstrate how the transferability of a given model should be measured against a diverse test set. We conclude by exploring one outlier, CO, and the outlook for extending this approach to other elements

#### Introduction

XPS is a very common characterization technique that provides estimates of core-level binding energies of the elements in their local chemical or materials environment context with an accuracy typically of ~0.1 eV or less.<sup>1–3</sup> This technique is commonly applied in a surface science<sup>4</sup> or vacuum/low pressure context with multiple standards published for reference.<sup>3</sup> More recently this technique has been applied to *in-situ* or *operando* experiments,<sup>5–9</sup> where the context may lack an experimental reference and subtle shifts in binding energies may reveal details of the local conditions within which this atom, molecule, surface, or interface finds itself. The need for theoretical predictions is therefore increasing as more nuanced experiments require atomic and molecular scale interpretation.

Concomitantly, there are multiple theoretical approaches for estimating the core-level binding energy with a lack of clear consensus on the best approach that gives the most reliable results in a computationally efficient manner. These can be divided into four broad categories: (1) Effective single-particle orbital energy estimates, following Koopman's theorem applied to the ground state<sup>10</sup>; Slater's transition-state theory, employing orbitals from a modified self-consistent field with a half-occupied core orbital<sup>11</sup>, or generalizations of the transition-state theory<sup>12,13</sup>; or restricted open-shell Kohn Sham (ROKS) using orbitals from a mixed energy estimate<sup>14</sup>. (2) Simple self-consistent field estimates employing total energy differences ( $\Delta$ SCF) between ground and core-excited states using Hartree Fock (HF) or Density Functional Theory (DFT), in various approximations<sup>15</sup> (LDA<sup>16</sup>, GGA<sup>17</sup>, meta-GGA<sup>18</sup>, hybrid<sup>19,20</sup>, range-separated hybrid<sup>21</sup>). (3) Post-HF methods, such as configuration interaction<sup>22,23</sup>, coupled cluster singles and doubles (CCSD)<sup>24,25</sup> or equation of motion coupled cluster (EOM-CCSD)<sup>26</sup>. (4) Green's function GW methods<sup>27</sup>, or by using hybrid functionals with a finite amount of exact exchange as starting point in the  $G_0W_0$  approximation<sup>28</sup>, geared towards the prediction in solid systems.

Within specific theoretical approaches, the basis sets' contributions to the sensitivity of calculated core-electron binding energy (CEBE) estimates to numerical representation have been explored. Krylov. et al.<sup>29</sup> extended their EOM-CCSD calculations to the (so-called) complete basis set limit, Illas. et al.<sup>30</sup> applied TPSS functional with fully uncontracted basis set, while Besley et al. considered utilizing published basis sets from the next element in the periodic table to better describe the core-excited atom, inspired by the Z+1 approximation<sup>31</sup>. Chong. et al.<sup>32</sup> considered using exponent-scaled Gaussian basis functions, according to Clementi's reformulation<sup>33</sup> of Slater's rules<sup>34</sup>, within generalized transition state theory. There are several other basis sets that purportedly have the variational freedom to describe the inner-shell regions around atomic nuclei (IGLO-III, PCX<sup>35</sup>, etc.). However, in the end, the results for some of these studies still fail to meet the accuracy requirements for predictive estimates of XPS binding energies (~0.1 eV)<sup>31,32</sup> or they avoided the problem by comparisons with other theoretical results<sup>29,35</sup> that require dramatic increases in the basis set size and computational expense. Furthermore, the strategy of increasing basis set size to the so-called complete basis set limit, when utilizing contracted Gaussian type orbitals (GTOs) is perhaps more appropriate to describing valence (i.e., more diffuse) electronic

<sup>&</sup>lt;sup>a.</sup> Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

<sup>&</sup>lt;sup>b.</sup> Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

<sup>&</sup>lt;sup>c.</sup> Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See

# Page 2 of 8

#### Journal Name

### ARTICLE

structure rather than subtle variations in the vicinity of the atomic nucleus.

Coming from a physics motif, we noticed that there is a fine balance between the numerical aspect for improved CEBEs accuracy and the interesting physical aspect of accurate electronic structure description. Dated from the birth of basis sets, they were meant to describe the atomic orbitals. It is therefore worth asking in what degree does the orbital change from ground state to core-excited states (see **Figure SI1-4**) and hence, would reproducing this key piece of physical information leads to more satisfactory numerical results? While we acknowledge that the adoption of basis close to basis set limits may circumvent these questions, it is, nevertheless a "black box" like optimization approach with exploded computational expense.

# Methodology

Here, we adopt a robust and chemically intuitive approach to derive a new suite of specialized basis set contractions reminiscent of the tactics employed by Dunning *et al.*<sup>36</sup> in the 1970s for the development of ground-state basis sets employing contracted Gaussian orbitals (i.e., multiple Gaussian exponents summed to define a single orbital basis function). Dunning aimed to describe the inner region of atoms near the atomic nuclei using Gaussian contractions that best reproduced atomic natural orbitals (ANOs)<sup>37</sup>, provided by Huzinaga *et al.*<sup>38</sup>, leaving free Gaussian orbitals with smaller exponents to describe the interatomic regions, bonding, diffuse states, or polarization response. This standard for contractions is adopted in most basis sets that we are familiar with (e.g., 6-31 series<sup>39</sup>, def2 series<sup>40</sup>, cc series<sup>36</sup>, etc.), which are, therefore, by construction, variationally inflexible to deviations in the inner regions near atomic nuclei.

There are multiple atomic electronic structure codes capable of producing numerically accurate results for Hartree-Fock or DFT approximations by employing log-radial grid-based finite difference representations<sup>41–44</sup> and, more recently, based on finite elements<sup>45,46</sup>. These calculations often provide the inputs for pseudopotential or effective core potential generation in valence only calculations. Using Vanderbilt's atomic code<sup>43</sup>, we provide (**Figure 1a**) a comparison of the radial component of the three

atomic orbitals (1s, 2s, 2p) of the carbon atom, calculated using DFT in its electronic ground state, with those of a constrained occupancy (and spin-averaged) approximation to the core-excited atom (based on a  $1s^{1}2s^{2}2p^{2}$  charged configuration). What we observe are distinct differences in the inner region near the nucleus, as we might imagine for the 1s orbital, and even longer-range differences in the valence orbitals, especially for 2p.

Our aim is to describe the orbitals of the core-excited state using an efficient Gaussian basis set. In the development of the correlationconsistent basis sets, Dunning<sup>36</sup> and co-workers provided contractions to accurately reproduce the atomic natural orbitals (ANOs)<sup>37,38</sup>. The goal was to best reproduce the inner-shell regions near the nuclei, which we would expect to remain unaffected by the molecular context, while leaving the outer Gaussians uncontracted and freely varying to describe the interatomic regions where bonding occurs. There is a close analogy here to the spirit of the pseudopotential or effective core potential approximation, in which core-valence separation is deemed appropriate. Such contractions can effectively describe radicals or bonding involving valence orbitals. However, these ground state contractions (either for the element itself or its next neighbour in the periodic table) are not flexible enough to describe orbital variation due to core excitations (Figure 1b). The use of the next element in the periodic table, nitrogen in this case, to model a core-excited state - referred to as the Z+1 approximation - is also insufficient. And since the level of agreement is mixed, depending on the orbital, it also seems clear that a simple linear combination of either Z or Z+1 orbital will be incapable of accurately describing the core-excited atom.

Therefore, we set about understanding the origins of changes in the atomic orbitals in response to the core-excitation. This led us to Slater's rules<sup>34</sup>, which were developed ca. 1930 to estimate the change in the screening of the nuclear charge experienced by each orbital shell (defined by its principal quantum number, *n*). These effective (i.e., screened) nuclear charges,  $Z_{eff} = Z - \sigma$  were then employed as the exponents in Slater-type orbitals. Coincidentally, in Slater's 1930 paper<sup>34</sup> he chooses a specific example of the 1s core-excited iron atom to demonstrate how its electronic structure is modified by the loss of a 1s electron. By analogy we can consider the change in the effective screening, between the ground and core-excited states to rescale the exponents within a given Gaussian



Figure 1 Comparison of carbon atomic orbitals: (a) The effect of 1s<sup>1</sup> core excitation illustrated by comparing ground state and core-excited orbitals computed on a log-radial grid via finite differences<sup>41-44</sup>. (b) Neither a ground state carbon (cc-pVTZ) basis nor ground state nitrogen (Z+1 approximation) basis accurately describes the inner-region of the core-excited state orbitals. Closer detail can be found in Supplementary Information.

ARTICLE

#### Journal Name

basis set, such that each  $a_n \rightarrow a'_n = a_n \left(\frac{Z'_{\text{eff}}}{Z_{\text{eff}}}\right)^2$ , where the primes indicate the core-excited state (details provided in Supplementary Information). As already stated, based on Dunning's convention, there is a contracted GTO in the basis set corresponding to each atomic orbital of the ground state. For certain orbitals, this rescaling of the exponents within that GTO is sufficient to transform the orbital to a very close approximation to the core-excited state. In other cases, we must go further and update the contraction terms themselves, which is a simple matter of either (1) linear regression if we have decided to use the rescaled exponents and have reference core-excited orbitals to fit or (2) a new calculation for the core-excited atom using a fully uncontracted basis set with these rescaled exponents (see **Supplementary Information** for details).

Interestingly, Slater's example for core-excited Fe illustrates that a different rescaling is associated with the n=1 and n=2 shells in the core-excited state. This, in part, explains why previous efforts employing the Z+1 approximation or a linear combination of Z and Z+1 failed to generate consistent improvements in CEBEs<sup>31,47</sup>, since what is really needed is different Gaussian exponents per orbital. In our chosen application to all-electron GTOs, we noticed that Slater's rules need some modification to simultaneously describe both the inner and outer regions of orbitals. Slater-type orbitals used in the original work ignore the inner nodal structure of the true allelectron orbital, focusing instead on describing the outer, bonding region, analogous to pseudopotential approaches which define pseudo-orbitals with similar properties. We find that the C 2s orbital is actually better described in the core-excited state by the rescaling corresponding to the 1s orbital. Our convention was to apply the same rescaling to all orbitals of the same angular momentum, defined by that orbital with the smallest principal quantum number. However, despite these nuances, the ultimate goal remains the same: to reproduce the core-excited atomic basis, based on the cc-pVTZ basis set, can be found in the **Supplementary Information**, and they can directly adopt and/or converted by the readers for future quantum chemistry calculations.

#### **Results and Discussion**

It is worth mentioning that there are additional complications when attempting to predict the core-excitation energies of condensed phases or interfaces<sup>48</sup> within extended models employing periodic boundary conditions, such as surfaces with adsorbates, due to the requirement to model a charged excited state within an inhomogeneous or anisotropic environment. Furthermore, benchmarking computational results with experiments for extended systems<sup>49</sup> also presents unique challenge due to controversial alignment schemes<sup>3</sup>. Therefore, in this work, we will focus entirely on isolated systems, limiting the possible sources of errors. We will model electronic structure using localized orbitals defined by Gaussian basis sets and leave extended systems for future work. We benchmark and compare our results with gas phase molecular experiments<sup>50</sup>, which can be unambiguously aligned to the vacuum level<sup>3</sup>, yielding absolute binding energies for comparison with our theoretical estimates.



rescaled exponents and optimal contractions of our excited-state

Figure 2 Comparison of calculated and experimental C 1s CEBEs computed with B3LYP for a broad set of C-containing molecules. (a) Consistent improvement in terms of mean average error (MAE) and maximum error (ME) are observed using the rescaled core-excited basis set for the excited carbon atom. (b) CEBE estimates using the carbon and nitrogen (Z+1 approximation) ground-state cc-pVTZ basis sets lead to larger systematic errors.

Journal Name



Figure 3 (a) CEBE is loosely proportional to the local charge on the carbon atom. (b) The quality of testing suites can be judged in terms of the width of the distribution of charges sampled in the test suite, based on the values provided in the literature (see also **Table 1**). All charges are calculated using the Mulliken population analysis, at the present B3LYP functional level. All structures were self-consistently obtained from geometry optimization at the B3LYP level.

Using this approach to construct the core-excited basis set, we applied it to study CEBEs of carbon-containing molecules, because: (i) carbon is chemically interesting due to its different hybridization possibilities and (ii) there are significantly more high-quality, gasphase XPS experimental data<sup>50</sup> for carbon-containing species (i.e. ~300 times more than boron, ~5 times more than oxygen and nitrogen) for us to benchmark with. We chose a broad set of molecules, for specific reasons highlighted below. Consistent improvement in terms of mean average error (MAE) and maximum error (ME) with respect to experiment are observed using the new core-excited basis set (Figure 2a). Across a wide range of molecular contexts, almost all of our calculations using the hybrid density functional, B3LYP<sup>19</sup>, and our core-excited contracted Gaussian basis, consistent with the cc-pVTZ family, lead to CEBE estimates within the desired accuracy of ~0.1 eV – an MAE of 0.1 eV and an ME of 0.4 eV (for CO) when compared with experiment. This is very encouraging by contrast to CEBE estimates using the carbon and nitrogen (Z+1 approximation) ground-state cc-pVTZ basis sets, which generate larger errors: MAE of 0.3 eV/ME of 0.7 eV (for CO), MAE of 1.4 eV/ME of 2.0 eV (for CO), respectively. The singular exception, where our new basis appears to be at odds with XPS experiments, is the CO molecule, whose peculiarities we discuss in detail later.

To explore the transferability of our approach to the broadest range of variability in CEBE estimates, we turn to the simple charge model for XPS binding energies proposed by Siegbahn<sup>51</sup>, which indicates approximate proportionality with respect to local atomic charge or oxidation state. This is validated by the approximate diagonal distribution of our calculated CEBEs with respect to the Mulliken charge on the core-excited carbon (**Figure 3a**). In this study, we aimed to describe the broadest range of possible oxidation states for a given element before reporting statistics on accuracy, such as mean absolute errors -from the most oxidized <u>CF4</u> to the most reduced <u>CH3</u>SiH3 and <u>CH4</u>. Since the relation between local charge and binding energy is not exactly linear, one can see that agreement within a narrow range of oxidation states might not be transferable to a broader context, see illustration in Figure 3b as well as results in Table 1.

CEBE estimates in the literature do not converge to experimental values with increasing basis set size, unlike studies that focus on properties defined by valence electronic structure, such as bonding. Increases in basis set size or complexity are often biased towards the description of more diffuse aspects of electronic structure, rather than details close to the atomic nuclei. Also, even though they might employ more Gaussian exponents, the contracted GTOs often appearing within these basis sets continue to reinforce the reproduction of the ground state atomic orbitals. Therefore, although diffusion and polarization are commonly thought to improve the performance for charged systems for the outer shell, it should be expected that such basis sets are ill-suited to describe core excitations more accurately for the inner shell without being fully uncontracted. For example, Ågren *et al.*<sup>11</sup> showed that the larger cc-pV5Z basis set is unable to produce satisfactory CEBE estimates. Our validation of accuracy with respect to theory (comparing CCSD with B3LYP, see Table S5) also highlights that higher levels of theory do not necessarily offer solutions, particularly when we realize that we may have large systematic errors due to our choice of basis set.

We return to the singular outlier in our C 1s data set. The CO molecule exhibited a CEBE error of 0.4 eV. It has already been highlighted that the local atomic charge state and polarizability of the CO molecule is very sensitive to the internuclear distance near the equilibrium bond length.<sup>52</sup> Correlated with these electronic properties, our B3LYP calculations predict that when first core-excited, the gradient of the excited-state potential energy surface would drive the bond length to shorten (by 0.06 Å), which has been experimentally observed (by 0.048 Å).<sup>53</sup> **Figure 4** reveals how dramatically the initial and final state energetics in the CO molecule change significantly with bond length near equilibrium, consistent with the theoretical findings of Dunning *et al.*<sup>52</sup> By contrast, CF<sub>4</sub> has a relatively flat potential energy surface. We propose that CO is a



Figure 4. Potential energy surfaces of (a) CO and (b) CF4 in ground state (bottom row) and C 1s core-excited states (top row) with respect to bond length. Equilibrium bond lengths in both ground and excited states are indicated using vertical dashed lines. For visualization purposes, the energy axes span the same range (0.54 eV or 0.02 Ha), referenced to the equilibrium ground state bond length.

special case, where (i) the equilibrium geometry samples a broad bond length distribution due to activation of vibrational modes and (ii) this distribution is sampled in an unbalanced manner in the core excited state due to the impulse to shorten the bond length. This tends to lower the core excitation energy for the departing electron, depending on its time scale. Potentially this could be studied with varying photon energy or time-resolved IR pump/X-ray probe. A higher level of theory may not be needed to explain the 0.4 eV error in our  $\Delta$ SCF/Maximum Overlap Method (MOM) calculations using the B3LYP functional, because we can successfully reduce the error to within 0.1 eV if we include the effect of CO bond shrinkage during the time of flight of the photoemitted electron. These results, together with Table S4, indicate that an appropriate basis set provides a consistent estimate of accurate CEBEs for each geometry. More detailed investigation using either ultrafast quantum dynamics or vibrational wave packet analysis is warranted for future work.

#### Extension to other elements

We also constructed Slater's Rule motivated basis sets for describing the core-excited states of boron, nitrogen, and oxygen.

The degree to which the core-excited electron alters the 1s, 2s, 2p orbitals in each case is visualized in the SI. We note that for lighter elements such as B and C, rescaling factors for exponents from Slater's Rule (see Table S1) generates satisfactory exited state basis sets, whereas heavier elements such as N and O would require linear regression of contraction terms (see  $\ensuremath{\text{SI}}$  for details) to match with the respective core-excited orbitals computed on a log-radial grid via finite differences<sup>41-44</sup>. We observe that the variational principle is consistently followed using our new Slater's Rule inspired basis sets (see Table S3 and S6). The core-excited basis sets perform well in providing B3LYP/MOM CEBE estimates for B, C, and N. However, we noticed larger systematic errors for O. One possible source of error is our lack of inclusion of relativistic effects, which amount to ~ 0.4 eV for  $O^{11,54}$ , by contrast with corrections that are negligible for B, and 0.1-0.2 eV for C and N. This may in part explain the observed systematic error of 0.5-0.7 eV for O 1s CEBEs of Ocontaining molecules (Table 1) but will be the focus of our future work.

Molecules	Charge	Experi ment <sup>50</sup>	Slater's Rule Motivat ed Basis	Present Work Error	Reference a <sup>55</sup> Error (ROKS/SCAN functional)	Reference b <sup>31</sup> Error (Z+1/B3LYP functional)	Reference c <sup>11</sup> Error (Δ SCF/PD86 functional)	Reference c <sup>11</sup> Error (UGTS/PD86 functional)
<u>B</u> H₃CO	-0.40	195.2	195.20	0.0	N/A	0.8*	N/A	N/A
<u>B</u> (OCH <sub>3</sub> ) <sub>3</sub>	0.21	197.8	197.95	0.1	N/A	1.2*	N/A	N/A
<u>B</u> F <sub>3</sub>	0.52	202.8	202.77	0.0	N/A	1.2	N/A	N/A
<u>C</u> H <sub>4</sub>	-0.41	290.8	290.92	0.1	0.0	1.1	0.31	-0.21
<u>C</u> H₃SiH₃	-0.39	290.3	290.40	0.1	N/A	1.3*	N/A	N/A
<u>C</u> H₃SH	-0.28	291.4	291.55	0.1	N/A	1.3*	N/A	N/A
H <u>C</u> N	-0.09	293.5	293.59	0.1	0.1	1.3	N/A	N/A

J. Name., 2013, 00, 1-3 | 5

#### ARTICLE

ļ	ο	u	r	r	١	а	L	ſ	V	а	ľ	γ	۱	e
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

<u>C</u> H₃OH	-0.06	292.4	292.48	0.1	0.2	1.3*	N/A	N/A
H <sub>2</sub> N <u>C</u> N	-0.05	293.5	293.47	0.0	N/A	1.4*	N/A	N/A
<u>C</u> O	-0.03	296.2	296.56	0.4	0.4	2.0*	0.34	-0.15
<u>C</u> Cl <sub>4</sub>	0.06	296.4	296.45	0.1	N/A	1.8*	N/A	N/A
H <sub>2</sub> <u>C</u> O	0.16	294.5	294.63	0.1	0.1	1.3	0.29	-0.23
NH₂ <u>C</u> HO	0.17	294.5	294.46	0.0	N/A	1.2*	N/A	N/A
HN <u>C</u> O	0.26	295.9	296.03	0.1	N/A	1.5*	N/A	N/A
$\underline{C}H_2F_2$	0.34	296.4	296.33	-0.1	N/A	1.2*	N/A	N/A
<u>C</u> O <sub>2</sub>	0.37	297.7	297.84	0.1	0.5	1.6*	N/A	N/A
<u>C</u> OF <sub>2</sub>	0.46	299.6	299.64	0.0	N/A	1.6*	N/A	N/A
<u>C</u> HF <sub>3</sub>	0.48	299.1	299.10	0.0	N/A	1.4*	N/A	N/A
<u>C</u> F <sub>4</sub>	0.54	301.9	301.76	-0.1	N/A	1.4*	N/A	N/A
<u>N</u> H <sub>3</sub>	-0.46	405.5	405.59	0.1	0.5	1.5*	N/A	N/A
CH <sub>3</sub> NH <sub>2</sub>	-0.33	405.1	405.10	0.0	N/A	1.4*	N/A	N/A
<u>N</u> H <sub>2</sub> CHO	-0.20	406.3	406.47	0.2	N/A	1.5*	N/A	N/A
HC <u>N</u>	-0.06	406.8	406.82	0.0	0.2	1.3	N/A	N/A
(CH <sub>3</sub> )C <u>N</u>	-0.04	405.6	405.48	-0.1	N/A	1.1*	0.23	-0.79
H <sub>2</sub> O	-0.44	539.92	539.35	-0.6	0.4	1.4*	0.07	-0.41
H <sub>2</sub> CO	-0.24	539.48	538.90	-0.6	0.2	1.4	0.35	-0.30
COF <sub>2</sub>	-0.21	540.77	540.07	-0.7	0.1	1.2*	N/A	N/A
C <u>O</u>	0.03	542.57	542.13	-0.4	0.3	1.5	0.53	-0.16
MAE				0.2	0.3	1.4	0.30	0.30
ME				0.7	0.5	2.0	0.53	-0.79

Table 1 Performance comparison of our Slater's rule motivated core-excited basis sets employed with  $\Delta$ SCF MOM-B3LYP calculations to state-of-the-art CEBE computationalstrategies in the literature. Molecules with more than 0.1 eV (outside of 1.3-1.5 eV for Z+1) error are highlighted in bold. We have completed the table of Z+1 entry with our owncalculationsindicatedby\*. Allstructureswereself-consistentlyobtainedfromgeometryoptimizationattheB3LYPlevel.

## Conclusions

In summary, we have derived compact contracted Gaussian basis sets, inspired by Slater's Rules, to efficiently retain an accurate local description of the atomic orbitals in the coreexcited states of molecules. Focusing initially on carbon, we motivated the need for rescaled and optimized contractions to describe each orbital in the core-excited state and provided a general procedure for their generation. The core electron binding energy (CEBE) estimates with these basis sets are computed with the B3LYP hybrid functional and the maximum overlap method for constraining the core-excited state across a large set of C-containing molecules, with an accuracy of ~0.1 eV with respect to experiment, barring a single outlier (CO: 0.4 eV) which is proposed to exhibit some ultrafast excited state dynamics. The breadth of our data set is characterized by the range of local electronic charge on the excited C atom. Extensions to other elements (B, N, O) demonstrate the generality of our approach, albeit with some systematic errors that relativistic effect can largely correct back. Overall, these core-excited contracted basis sets provide an efficient means to improve our ability to interpret XPS measurements.

### **Conflicts of interest**

There are no conflicts to declare.

## Acknowledgements

This work was supported by a User Project at The Molecular Foundry and its compute cluster (Lawrencium), managed by the High-Performance Computing Services Group, at Lawrence Berkeley National Laboratory (LBNL). Portions of this work used the computing resources of the National Energy Research Scientific Computing Center (NERSC), LBNL, supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. J.Q., and E.J.C. were supported by an Early Career Award in the Condensed Phase and Interfacial Molecular Science Program, in the Chemical Sciences Geosciences and Biosciences Division of the Office of Basic Energy Sciences of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

# Notes and references

#### Journal Name

- 1 J. J. Pireaux, S. Svensson, E. Basilier, P.-Å. Malmqvist, U. Gelius, R. Caudano and K. Siegbahn, Core-electron relaxation energies and valence-band formation of linear alkanes studied in the gas phase by means of electron spectroscopy, Phys. Rev. A, 1976, 14, 2133–2145.
- 2 V. Carravetta, G. lucci, A. Ferri, M. V. Russo, S. Stranges, M. de Simone and G. Polzonetti, Synchrotron radiation photoemission study of some  $\pi$ -conjugated alkynes in the gas phase: Experiment and theory, Chemical Physics, 2001, 264, 175–186.
- 3 G. Greczynski and L. Hultman, X-ray photoelectron spectroscopy: Towards reliable binding energy referencing, Progress in Materials Science, 2020, 107, 100591.
- 4 C. J. Powell and A. Jablonski, Surface sensitivity of X-ray photoelectron spectroscopy, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2009, 601, 54–65.
- 5 M. Favaro, H. Xiao, T. Cheng, W. A. Goddard, J. Yano and E. J. Crumlin, Subsurface oxide plays a critical role in CO2 activation by Cu(111) surfaces to form chemisorbed CO2, the first step in reduction of CO2, PNAS, 2017, 114, 6706–6711.
- 6 M. Favaro, W. S. Drisdell, M. A. Marcus, J. M. Gregoire, E. J. Crumlin, J. A. Haber and J. Yano, An Operando Investigation of (Ni–Fe–Co–Ce)Ox System as Highly Efficient Electrocatalyst for Oxygen Evolution Reaction, ACS Catal., 2017, 7, 1248–1258.
- 7 M. Favaro, J. Yang, S. Nappini, E. Magnano, F. M. Toma, E. J. Crumlin, J. Yano and I. D. Sharp, Understanding the Oxygen Evolution Reaction Mechanism on CoOx using Operando Ambient-Pressure X-ray Photoelectron Spectroscopy, J. Am. Chem. Soc., 2017, 139, 8960–8970.
- 8 K. A. Stoerzinger, M. Favaro, P. N. Ross, J. Yano, Z. Liu, Z. Hussain and E. J. Crumlin, Probing the Surface of Platinum during the Hydrogen Evolution Reaction in Alkaline Electrolyte, J. Phys. Chem. B, 2018, 122, 864–870.
- 9 C. J. Corcoran, H. Tavassol, M. A. Rigsby, P. S. Bagus and A. Wieckowski, Application of XPS to study electrocatalysts for fuel cells, Journal of Power Sources, 2010, 195, 7856–7879.
- 10 N. P. Bellafont, F. Illas and P. S. Bagus, Validation of Koopmans' theorem for density functional theory binding energies, Physical Chemistry Chemical Physics, 2015, 17, 4015–4019.
- 11 L. Triguero, O. Plashkevych, L. G. M. Pettersson and H. Ågren, Separate state vs. transition state Kohn-Sham calculations of X-ray photoelectron binding energies and chemical shifts, Journal of Electron Spectroscopy and Related Phenomena, 1999, 104, 195–207.
- 12 D. P. Chong, Density functional calculation of core electron binding energies of C, N, O, and F, J. Chem. Phys., 1995, 103, 1842–1845.
- 13 D. P. Chong, Accurate calculation of core-electron binding energies by the density-functional method, Chemical Physics Letters, 1995, 232, 486–490.
- 14 T. Kowalczyk, T. Tsuchimochi, P.-T. Chen, L. Top and T. Van Voorhis, Excitation energies and Stokes shifts from a restricted open-shell Kohn-Sham approach, J. Chem. Phys., 2013, 138, 164101.
- 15 J. P. Perdew and K. Schmidt, Jacob's ladder of density functional approximations for the exchange-correlation energy, AIP Conference Proceedings, 2001, 577, 1–20.

- 16 P. a. M. Dirac, Note on Exchange Phenomena in the Thomas Atom, Mathematical Proceedings of the Cambridge Philosophical Society, 1930, 26, 376–385.
- 17 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett., 1996, 77, 3865–3868.
- 18 Y. Zhao and D. G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, Theor Chem Account, 2008, 120, 215– 241.
- 19 A. D. Becke, Density functional thermochemistry. III. The role of exact exchange, J. Chem. Phys., 1993, 98, 5648–5652.
- 20 C. Adamo and V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, J. Chem. Phys., 1999, 110, 6158–6170.
- 21 J. Heyd and G. E. Scuseria, Efficient hybrid density functional calculations in solids: Assessment of the Heyd– Scuseria–Ernzerhof screened Coulomb hybrid functional, J. Chem. Phys., 2004, 121, 1187–1192.
- 22 A. J. Nelson, J. G. Reynolds and J. W. Roos, Core-level satellites and outer core-level multiplet splitting in Mn model compounds, Journal of Vacuum Science & Technology A, 2000, 18, 1072–1076.
- 23 M. W. D. Hanson-Heine, M. W. George and N. A. Besley, A scaled CIS(D) based method for the calculation of valence and core electron ionization energies, J. Chem. Phys., 2019, 151, 034104.
- 24 R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, Rev. Mod. Phys., 2007, 79, 291–352.
- 25 G. D. Purvis and R. J. Bartlett, A full coupled cluster singles and doubles model: The inclusion of disconnected triples, J. Chem. Phys., 1982, 76, 1910–1918.
- 26 M. Musiał, S. A. Kucharski and R. J. Bartlett, Equation-ofmotion coupled cluster method with full inclusion of the connected triple excitations for ionized states: IP-EOM-CCSDT, J. Chem. Phys., 2003, 118, 1128–1136.
- 27 T. Aoki and K. Ohno, Accurate quasiparticle calculation of xray photoelectron spectra of solids, J Phys Condens Matter, 2018, 30, 21LT01.
- 28 D. Golze, L. Keller and P. Rinke, Accurate Absolute and Relative Core-Level Binding Energies from GW, J. Phys. Chem. Lett., 2020, 11, 1840–1847.
- 29 R. Sarangi, M. L. Vidal, S. Coriani and A. I. Krylov, On the basis set selection for calculations of core-level states: different strategies to balance cost and accuracy, Molecular Physics, 2020, 0, e1769872.
- 30 N. Pueyo Bellafont, F. Viñes and F. Illas, Performance of the TPSS Functional on Predicting Core Level Binding Energies of Main Group Elements Containing Molecules: A Good Choice for Molecules Adsorbed on Metal Surfaces, J. Chem. Theory Comput., 2016, 12, 324–331.
- 31 M. W. D. Hanson-Heine, M. W. George and N. A. Besley, Basis sets for the calculation of core-electron binding energies, Chemical Physics Letters, 2018, 699, 279–285.
- 32 G. Cavigliasso and D. P. Chong, Accurate density-functional calculation of core-electron binding energies by a totalenergy difference approach, J. Chem. Phys., 1999, 111, 9485–9492.

- 33 E. Clementi and D. L. Raimondi, Atomic Screening Constants from SCF Functions, J. Chem. Phys., 1963, 38, 2686–2689.
- 34 J. C. Slater, Atomic Shielding Constants, Phys. Rev., 1930, 36, 57–64.
- 35 M. A. Ambroise and F. Jensen, Probing Basis Set Requirements for Calculating Core Ionization and Core Excitation Spectroscopy by the Δ Self-Consistent-Field Approach, J. Chem. Theory Comput., 2019, 15, 325–337.
- 36 T. H. Dunning, Gaussian Basis Functions for Use in Molecular Calculations. I. Contraction of (9s5p) Atomic Basis Sets for the First - Row Atoms, J. Chem. Phys., 1970, 53, 2823–2833.
- 37 J. Almlöf and P. R. Taylor, General contraction of Gaussian basis sets. I. Atomic natural orbitals for first and second row atoms, J. Chem. Phys., 1987, 86, 4070–4077.
- 38 S. Huzinaga, Basis sets for molecular calculations, Computer Physics Reports, 1985, 2, 281–339.
- 39 R. Ditchfield, W. J. Hehre and J. A. Pople, Self Consistent Molecular - Orbital Methods. IX. An Extended Gaussian -Type Basis for Molecular - Orbital Studies of Organic Molecules, J. Chem. Phys., 1971, 54, 724–728.
- 40 F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, Phys. Chem. Chem. Phys., 2005, 7, 3297–3305.
- 41 M. Fuchs and M. Scheffler, Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems using density-functional theory, Computer Physics Communications, 1999, 119, 67–98.
- 42 D. R. Hamann, Generalized norm-conserving pseudopotentials, Phys. Rev. B, 1989, 40, 2980–2987.
- 43 D. Vanderbilt, Soft self-consistent pseudopotentials in a generalized eigenvalue formalism, Phys. Rev. B, 1990, 41, 7892–7895.
- 44 N. Troullier and J. L. Martins, Efficient pseudopotentials for plane-wave calculations, Phys. Rev. B, 1991, 43, 1993– 2006.
- 45 S. Lehtola, Fully numerical Hartree-Fock and density functional calculations. II. Diatomic molecules, International Journal of Quantum Chemistry, 2019, 119, e25944.
- 46 S. Lehtola, Fully numerical Hartree-Fock and density functional calculations. I. Atoms, , DOI:10.1002/qua.25945.
- 47 A. E. A. Fouda and N. A. Besley, Assessment of basis sets for density functional theory-based calculations of coreelectron spectroscopies, Theor Chem Acc, 2017, 137, 6.
- 48 F. Viñes, C. Sousa and F. Illas, On the prediction of core level binding energies in molecules, surfaces and solids, Phys. Chem. Chem. Phys., 2018, 20, 8403–8410.
- 49 J. M. Kahk and J. Lischner, Accurate absolute core-electron binding energies of molecules, solids, and surfaces from first-principles calculations, Phys. Rev. Materials, 2019, 3, 100801.
- 50 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, 1984, 31, 433–493.
- 51 K. M. G. Siegbahn, W. C. Price and D. W. Turner, A Discussion on photoelectron spectroscopy - Electron spectroscopy for chemical analysis (e.s.c.a.), Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1970, 268, 33–57.

- 52 K. A. Peterson and T. H. Dunning, The CO molecule: the role of basis set and correlation treatment in the calculation of molecular properties, Journal of Molecular Structure: THEOCHEM, 1997, 400, 93–117.
- 53 T. X. Carroll, K. J. Børve, L. J. Sæthre, J. D. Bozek, E. Kukk, J. A. Hahne and T. D. Thomas, Carbon 1s photoelectron spectroscopy of CF4 and CO: Search for chemical effects on the carbon 1s hole-state lifetime, J. Chem. Phys., 2002, 116, 10221–10228.
- 54 C. L. Pekeris, Ground State of Two-Electron Atoms, Phys. Rev., 1958, 112, 1649–1658.
- 55 D. Hait and M. Head-Gordon, Highly Accurate Prediction of Core Spectra of Molecules at Density Functional Theory Cost: Attaining Sub-electronvolt Error from a Restricted Open-Shell Kohn–Sham Approach, J. Phys. Chem. Lett., 2020, 11, 775–786.