

On the basis set selection for molecular core-level GW calculations

Daniel Mejia-Rodriguez,^{*,†} Alexander Kunitsa,^{*,‡} Edoardo Aprà,^{*,†} and
Niranjan Govind^{*,¶}

[†]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory,
Richland, WA 99352, USA*

[‡]*Zapata Computing, Inc., 100 Federal Street, Boston, MA 02110, USA*

[¶]*Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory,
Richland, WA 99352, USA*

E-mail: daniel.mejia@pnnl.gov; aakunitsa@gmail.com; edoardo.apra@pnnl.gov;
niri.govind@pnnl.gov

Abstract

The *GW* approximation has been recently gaining popularity among the method for simulating molecular core-level X-ray photoemission spectra. Traditionally, *GW* core-level binding energies have been computed using either the cc-pVnZ or def2-nZVP basis set families, extrapolating the obtained results to the complete basis set limit, followed by an element-specific relativistic correction. Despite of achieving good accuracy, these binding energies are chronically underestimated. By using first-row elements and standard techniques known to offer good cost-accuracy ratio in other theories, we show that the cc-pVnZ and def2-nZVP families show large contraction errors and lead to unreliable complete basis set extrapolations. On the other hand, we demonstrate that uncontracted versions of these basis sets offer vastly improved convergence. Even faster convergence can be obtained using core-rich, property-optimized, basis sets families like pcSseg-*n*, pcJ-*n* and ccX-*n*Z. Finally, we also show that the improvement over the core properties does not degrade the calculation of the valence excitations, and thus offer a balanced description of both core and valence regions.

1 Introduction

X-ray-based spectroscopies are well suited to study the local environment of atoms in a molecule or material. Typically, a core electron is probed through electronic excitation or ionization, leading to a very sensitive and selective spectrum.¹ The use of state-of-the-art X-ray light sources has expanded the applicability and power of such techniques. Moreover, it has been noted that the interpretation of the X-ray spectra requires high-level theoretical approaches.^{2,3} Theoretical simulations of X-ray-based spectroscopies are thus increasingly relevant.

X-ray photoelectron spectroscopy (XPS)⁴ is one of the most widely used X-ray-based approaches. The key quantities in simulations of XPS spectra are core-level binding energies (CLBEs), which correspond to the ionization energies in core orbitals. The most popular methods for computing CLBEs can be classified into two main categories: Δ self-consistent-field-like (Δ SCF) methods, and response methods (including linear response, equation-of-motion, and Green's func-

tion approaches).

The GW approximation (GWA)⁵ to the self-energy is a Green’s-function-based method that can be used to obtain accurate molecular CLBES at a reasonable cost.⁶⁻⁹ The CLBES obtained with this approach include some level of orbital relaxation effects, however, the quality of the results might depend strongly on the starting point. For example, it has been known that a large fraction of exact exchange is needed at the one-shot G_0W_0 level, mainly due to a large self-interaction error (SIE) present when standard “pure” exchange-correlation density functional approximations (DFAs) are adopted.^{6,7,9} A possible workaround when starting from a pure Kohn-Sham result is to use the more demanding $evGW$ partial self-consistent approach,⁷⁻¹⁰ although some starting-point dependency will remain.

Besides these unique requirements, CLBES obtained with the GWA will also depend on the quality of the basis set. The accuracy of GWA-based CLBES has been studied recently by means of two benchmark datasets.^{6,8,10} Both studies extrapolated the CLBES to the complete basis-set (CBS) limit in order to eliminate the dependency on the basis set. In Ref. 6, the authors used the def2-TZVP and def2-QZVP¹¹ basis sets and omitted relativistic corrections; while in Refs. 7 and 8, the authors used Dunning’s correlation-consistent basis sets cc-pVnZ,¹² from triple- to sextuple-zeta quality, and included relativistic corrections. Both studies showed that GW can be used to predict CLBES with a mean absolute error of about 0.3 eV for first row elements, but that this error can only be “inconsistently” reduced by adding relativistic corrections.¹⁰

Both benchmark studies also relied on the extrapolation of the CLBES, obtained with standard basis sets, to the CBS limit. Standard CBS extrapolation techniques were developed to extrapolate the correlation energy of small molecules.¹³⁻¹⁵ This is usually a small-energy regime with well-known basis set convergence behavior,^{16,17} and asymptotic behavior $-E_c/Z \sim -0.02073 \ln Z + 0.0372$ ^{18,19}—for atoms. In contrast, the energy involved for ionizing the core is rather large— $\varepsilon_{1s} \sim Z^2/2$ —and does not follow the same convergence trend. This may explain why Golze and co-workers used a linear regression fit with respect to the inverse number of basis functions⁸ compared with Bruneval and co-workers, who used the more common exponential convergence

formula for GW valence binding energies.^{20,21}

Even when the correlation energy of molecules and the core ionization energy have similar asymptotic behaviors, the basis set requirements for a response method, like the GWA, might be very different compared with a single-point energy calculation. It has been shown that basis sets optimized for energies often fail to describe other molecular properties due to a poor description of the orbitals in regions important for the given property but otherwise energetically unimportant.^{22,23} This numerical issue prevents the application of systematic improvements and of reliable extrapolation approaches. A natural consequence is that extrapolation using non-optimal basis sets might result in values far away from the real CBS limit.

The importance on the selection of the basis set have already been addressed for other response methods, like the equation-of-motion coupled-cluster^{23,24} or the linear-response time-dependent density functional theory formalism.²⁵ Basis set convergence has also been studied within the GWA applied to molecules, mostly focused to valence ionization potentials,^{20,21,26} usually showing smooth, albeit slow, convergence. However, a systematic study comparing CLBES obtained with basis sets other than the cc-pVnZ and def2-nZVP families is not available in the literature.

In this study, we use the CORE65 dataset, comprised of 65 CLBES of first-row elements, in order to test the effect of the basis set core-level GW calculations. We assess two types of basis sets that are known to offer very good results in other theories: uncontracted versions of traditional energy-optimized basis sets, and core-rich property-optimized basis sets.

In turn, we show that both cc-pVnZ and def2 basis set families are not well suited to describe first-row 1s CLBES within the GWA, often resulting in very slow convergence and unreliable CBS extrapolations. Findings of this work are consistent with earlier studies dealing with other linear-response methods, by showing that while uncontracted versions of these basis sets offer vastly improved results, the CLBES seem to converge to a different value as compared to the CLBES obtained with other core-rich basis sets like pcJ- n ,²⁷ pcSseg- n ,²² and the very recent ccX- n Z²³ families. The difference is small (around 100 meV), however, this value is close to the intrinsic accuracy of the GWA for CLBES and, as such, it is worth noticing.

2 Theory

2.1 Overview of the *GW* Approximation

The central object of the GWA is the one-particle Green's function G describing particle and hole scattering in the interacting many-body system. In order to obtain the electron addition and removal energies from such a Green's function, a non-local and dynamic effective potential, the self-energy Σ , is often introduced. The self-energy Σ substitutes the mean-field exchange-correlation operator, and the *GW* quasiparticle (QP) energies ε_n^{GW} can be obtained as corrections to the mean-field energies ε_n :

$$\varepsilon_{n\sigma}^{GW} = \varepsilon_{n\sigma} + \Re(\Sigma_{n\sigma}(\varepsilon_{n\sigma}^{GW})) - V_{n\sigma}^{xc} \quad (1)$$

Here, σ denotes the spin index, and $V_{n\sigma}^{xc}$ and $\Sigma_{n\sigma}$ denote the n th diagonal element of the corresponding matrix representation in the molecular orbital basis. Equation 1 is non-linear and must be solved iteratively.

The self-energy operator Σ is given in terms of the Green's function G^σ and the screened Coulomb interaction W :

$$\Sigma_\sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\xi G^\sigma(\mathbf{r}, \mathbf{r}', \omega + \xi) W(\mathbf{r}, \mathbf{r}', \xi) e^{i\xi\eta} \quad (2)$$

with η being a positive infinitesimal. In practice, the GWA is often performed as a one-shot perturbative approach known as G_0W_0 . In this case, G_0^σ is the non-interacting mean-field Green's function,

$$G_0^\sigma(\mathbf{r}, \mathbf{r}', \omega) = \sum_m \frac{\phi_{m\sigma}(\mathbf{r})\phi_{m\sigma}(\mathbf{r}')}{\omega - \varepsilon_{m\sigma} - i\eta \text{sign}(\mu - \varepsilon_{m\sigma})}, \quad (3)$$

and W_0 is obtained using the random phase approximation (RPA) as

$$W_0(\mathbf{r}, \mathbf{r}', \omega) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}'', \omega) v(\mathbf{r}'', \mathbf{r}') \quad (4)$$

In the preceding equations, μ is the Fermi-level of the system, $v(\mathbf{r}, \mathbf{r}')$ is the bare Coulomb inter-

action, and $\epsilon(\mathbf{r}, \mathbf{r}'', \omega)$ is the RPA dynamical dielectric function:

$$\epsilon(\mathbf{r}, \mathbf{r}'', \omega) = \delta(\mathbf{r}, \mathbf{r}') - \int d\mathbf{r}'' v(\mathbf{r}, \mathbf{r}'') \chi_0(\mathbf{r}'', \mathbf{r}', \omega) \quad (5)$$

In the RPA, the irreducible polarizability χ_0 has a simple sum-over-states representation^{28,29}

$$\chi_0(\mathbf{r}, \mathbf{r}', \omega) = \sum_{\sigma} \sum_{i,a} \left[\frac{\phi_{i\sigma}(\mathbf{r}) \phi_{a\sigma}(\mathbf{r}) \phi_{i\sigma}(\mathbf{r}') \phi_{a\sigma}(\mathbf{r}')}{\omega - \varepsilon_{a\sigma} + \varepsilon_{i\sigma} + i\eta} + \frac{\phi_{i\sigma}(\mathbf{r}) \phi_{a\sigma}(\mathbf{r}) \phi_{i\sigma}(\mathbf{r}') \phi_{a\sigma}(\mathbf{r}')}{-\omega - \varepsilon_{a\sigma} + \varepsilon_{i\sigma} + i\eta} \right] \quad (6)$$

where the index i runs over the occupied orbitals while the index a runs over the virtual ones.

The contour deformation (CD) technique is used in this work for the accurate integration of Σ , as it has been shown to have a good cost-effective profile for the evaluation of CLBES. Further details about CD-*GW* and its implementation using local orbitals can be obtained from References 7, 30, and 9 for example.

3 Computational Details

The adequacy of the cc-pVnZ and def2-nZVP basis set families for core-level *GW* calculations was evaluated by computing the 65 CLBES from the CORE65 benchmark dataset.⁸ These energies were then compared with those obtained with the respective completely uncontracted versions, hereafter denoted as un-cc-pVnZ and un-def2-nZVP. The CD-*GW* approach, recently implemented in the open-source computational chemistry package NWChem,^{9,31} was used for this task. The frequency integral over the imaginary axis was evaluated with a modified Gauss-Legendre grid with 200 points. The QP equation were always solved iteratively (i.e. no linearized approximation was used). All calculations used the SIMINT library^{32,33} and a 10^{-14} Schwarz screening threshold for the evaluation of the electron repulsion integrals. All CLBES were obtained using the PBEh³⁴ functional with 45 % of exact exchange as starting point for the G_0W_0 calculations.⁷

In order to find other appropriate basis sets families, the 65 CLBES were also obtained with the pcJ-*n*, pcSseg-*n*, and ccX-*n*Z families, as well as their uncontracted versions.

Since our CD-*GW* implementation fits the four-center ERIs, an adequate fitting basis is needed. Here, we do not directly assess the impact of various fitting basis sets on the quality of CLBES but we do use automatically generated fitting bases using the “AutoAux” procedure³⁵ for each orbital basis set.

4 Results and Discussion

4.1 The cc-pVnZ and def2-nZVP families

We first begin studying the two families used in the CLBES benchmark studies in the literature, the cc-pVnZ and def2-nZVP families, respectively. Both basis set families were developed with total energies in mind. In energy-optimized basis sets, contraction of the core basis functions is usually performed in order to achieve high efficiency at the price of a small loss in accuracy.³⁶ The removal of a core electron, however, induces a significant relaxation of the core orbital incompatible with a fixed shape due to the contraction scheme. The contraction effect is clearly visible when comparing the results of a given basis set and its corresponding uncontracted version. Such an effect can be seen in Figures 1 and 2 for the def2-nZVP and cc-pVnZ families, respectively, where CLBES deviations with respect to the experiment (see Supplemental Material of reference 8) are shown. The contracted series are presented in a blue multi-hue color scheme, while the uncontracted series in a red multi-hue color scheme. A two-point CBS extrapolation is also shown in gold with dots.

There are several important remarks about the plots shown in Figures 1 and 2: first, contracted double- ζ basis sets lead to very large CLBES regardless of the element; second, CLBES obtained with uncontracted basis sets are almost always larger than those obtained with contracted basis sets, the only exception being the unusually large CLBES obtained with def2-SVP and cc-pVDZ; third, since def2-QZVP CLBES are smaller than def2-TZVP CLBES, a CBS extrapolation shifts the CLBES prediction further away from experimental values; fourth, CLBES obtained with uncontracted basis sets show a rather fast convergence, and a double- ζ basis set suffices for C, N, and O in most cases.

These results provide evidence that the contraction schemes used to generate the def2- n ZVP and cc-p Vn Z families are not well suited to describe CLBES within the GWA. As a consequence, the CBS extrapolations used in previous studies do not convey absolute CLBES converged with respect to the basis set size, often being 0.5 eV too small (when comparing the results from uncontracted basis with the ones from CBS extrapolation).

An arguably better description can be obtained by using the uncontracted versions of both basis set families. Figure 3 shows that the CLBES predicted by both uncontracted basis set families are rather similar. This supports the notion that this description might be indeed close to the basis set limit. It is important to mention that the CLBES obtained with the uncontracted bases are systematically larger than the CBS extrapolation used in Reference 8 by about 0.25 eV for carbon and nitrogen, 0.40 eV for oxygen, and 0.50 eV for fluorine.

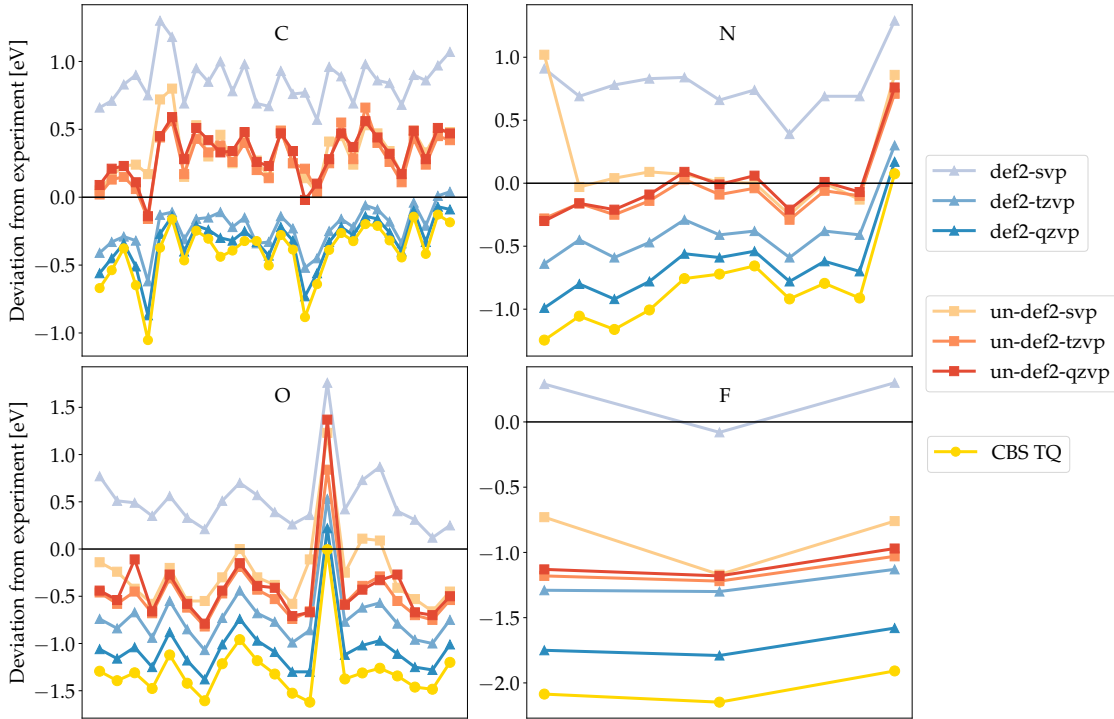


Figure 1: G_0W_0 1s binding energy error due to the contraction of the def2 basis set family. The plots show shifts with respect to the experimental results listed in Reference 8. No relativistic corrections were included.

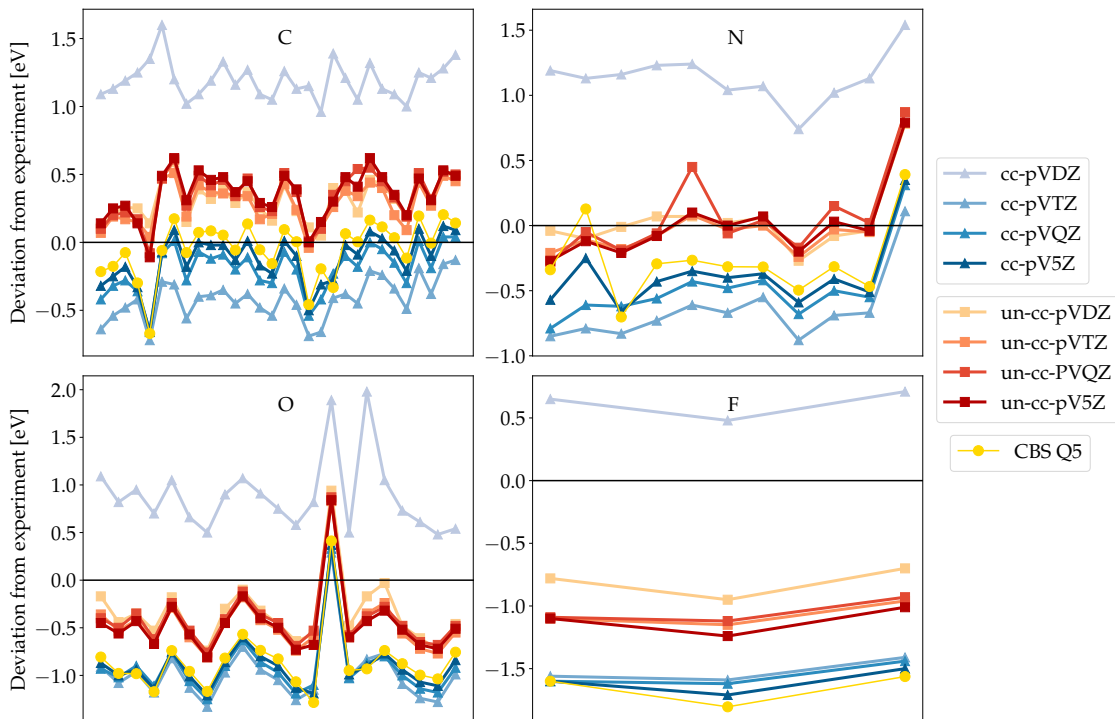


Figure 2: G_0W_0 1s binding energy error due to the contraction of the cc-pV n Z basis set family. The plots show shifts with respect to the experimental results listed in Reference 8. No relativistic corrections were included.

4.2 Comparison with core-rich basis sets

4.2.1 Absolute core-level binding energies

Given that the flexibility around the core seems to play a very important role in determining the quality of the GW CLBEs, the use of core-rich basis sets might be a better choice than the use of either un-cc-pV n Z or un-def2- n ZVP.²³ As mentioned earlier, we could not find a systematic study of GW CLBEs using these basis sets. The only data available was found in the Supplemental Material of reference,¹⁰ where the CLBEs obtained with the core-augmented cc-pCV n Z basis set was compared with CLBEs obtained with the standard cc-pV n Z basis sets. In that study, the authors noted that the errors of the cc-pCV n Z were not systematic. A similar conclusion was reached in Reference 23, where the authors noted that uncontracting the cc-pCV n Z family leads to larger deviations with respect to the CBS limit, a result which suggests an uncontrolled error cancellation in the contraction scheme.

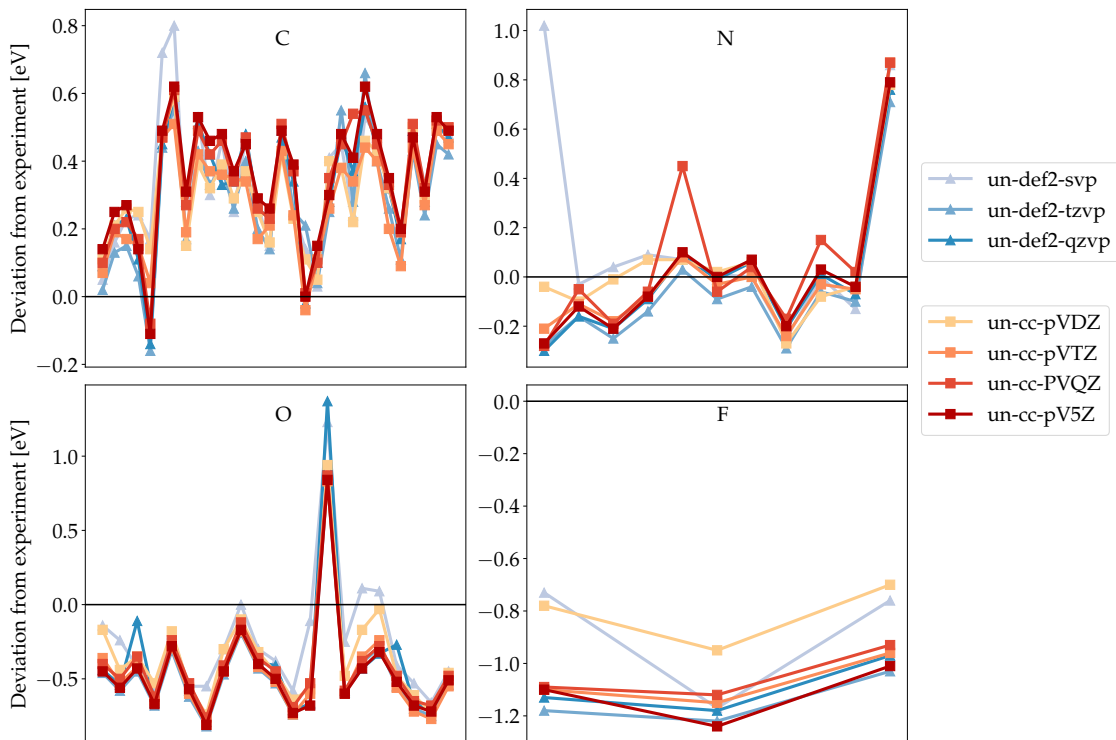


Figure 3: Comparison of G_0W_0 1s binding energies obtained with two families of uncontracted basis sets. No relativistic corrections were included.

Here, we have extended the results presented in Reference 10 and have included the pcJ- n , pcSseg- n , and ccX- n Z core-rich basis set families. Figure 4 shows CLBES deviations with respect to experiment for all basis set families tested. The same color code as in the previous sections is being used: blue hues for contracted sets, and red hues for uncontracted sets. Note that relativistic corrections have not been included.

The three core-rich basis set families predict converged CLBES in very good agreement to the ones obtained using the uncontracted def2- n ZVP or cc-pV n Z basis sets. Moreover, there is no longer a mismatch between the CLBES obtained with contracted and uncontracted versions of these basis sets. The fact that the same limit is achieved by all uncontracted basis sets, as well as with the contracted pcJ- n , pcSseg- n , and ccX- n Z, strongly supports the idea that this limit is in fact closer to the CBS than the extrapolations given in Reference 8.

Another aspect worth noting is the manner in which the CLBES converge for each basis set. As noted before, the def2- n ZVP basis sets show the wrong shift going from triple- to quadruple- ζ , the

cc-pVnZ family converges very slowly from below, the pcSseg- n family converges from above, the pcJ- n family converges from below, and the ccX- n Z family is very well converged even with double- ζ quality.

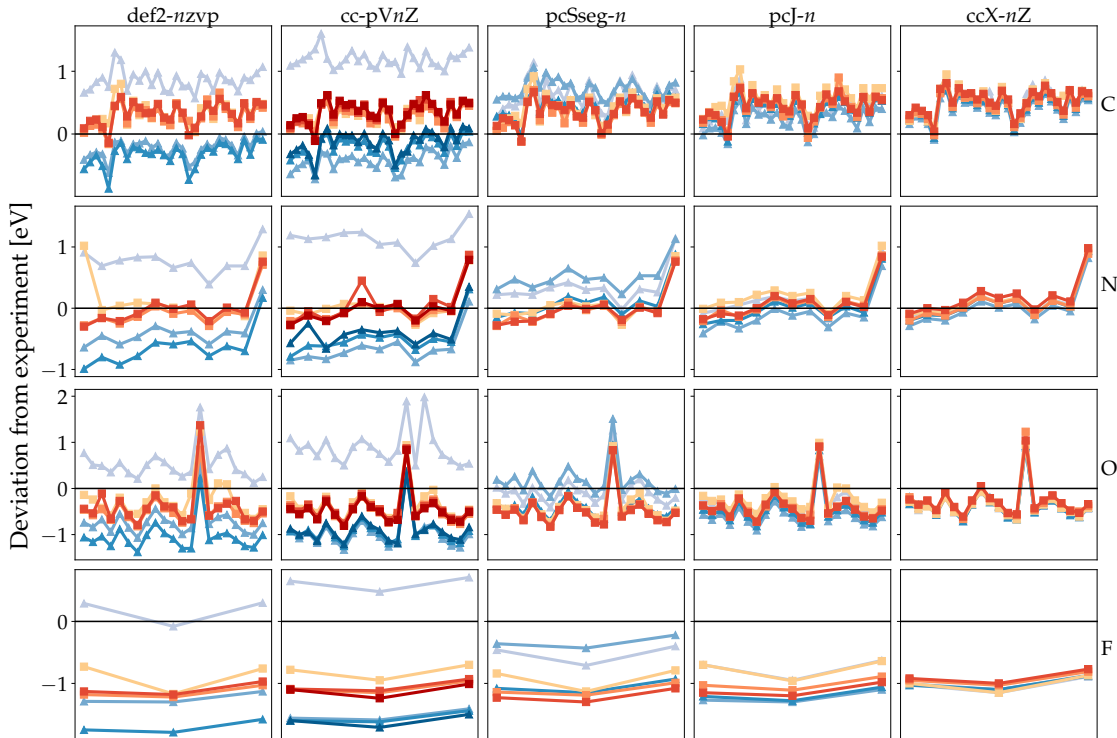


Figure 4: G_0W_0 1s binding energy deviations with respect to experiment. No relativistic corrections were included. All families include double-, triple-, and quadruple- ζ bases, and the cc-p n VZ family includes cc-pV5Z. The color code is analog to Figures 1–3, blue hues for contracted bases and red hues for uncontracted ones. The larger the n , the darker the color.

As a consequence of the overall larger CLBES predicted with core-rich basis sets, the statistics presented by Golze et al.⁸ also shift. All individual CLBES, as well as the mean errors and mean absolute errors with respect to experiment, are tabulated in the supplemental material. Taking the results obtained with the uncontracted ccX-QZ basis set as our approximation to the CBS limit, we find that $G_0W_0@PBEh(\alpha = 0.45)$ yields mean absolute errors of 0.65, 0.41, 0.21, and 0.19 eV for C, N, O, and F, respectively. In comparison, the values reported by Golze et al.⁸ are 0.24, 0.16, 0.48, and 0.83 eV, respectively (see Tables S11 and S12 of the Supplementary material for more details).

4.2.2 Relative core-level binding energies

Chemical shifts show faster convergence with any of the basis set used in this study, as shown in Figure 5. The contracted pcJ- n and ccX- n Z tend to show smaller dispersions overall, however, this effect is rather small and barely noticeable. A special case is obtained with cc-pV5Z for Nitrogen 1s chemical shifts, as Figure 5 shows that these chemical shifts are too small as compared to the consensus of the remaining bases. We have not been able to identify the cause for this discrepancy, but note that the chemical shifts obtained with the uncontracted version are in very good agreement with the rest of the results.

Since the relative binding energies are less sensitive to the basis set chosen, the statistics presented by Golze et al.⁸ does not change by more than 0.05 eV when compared to our ccX-QZ results.

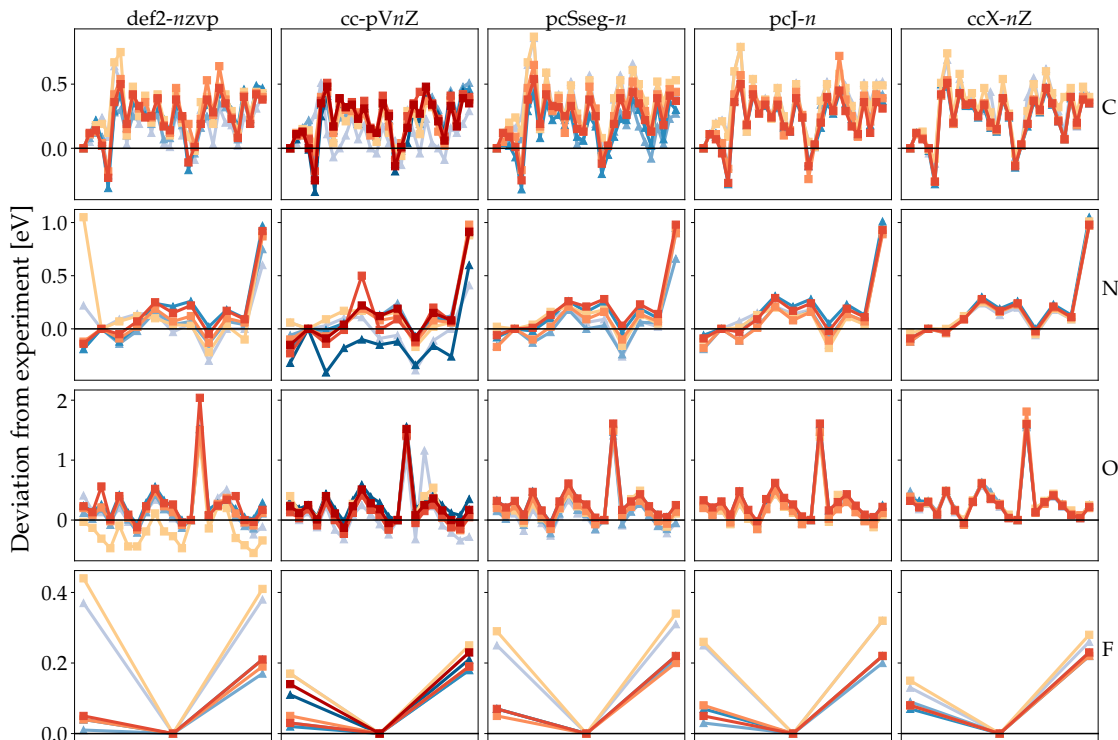


Figure 5: G_0W_0 1s chemical shift deviations with respect to experiment. No relativistic corrections were included. Same coloring scheme as in Figure 4.

4.2.3 Valence binding energies

Improvement in convergence of core-level binding energies should not, in principle, come at the expense of a poor description of the valence region. We evaluated the accuracy achieved with each one of the contracted basis set families by means of a small subset of the GW100 dataset.^{37,38} The subset is made exclusively of molecules containing B, C, N, O, and/or F (the ccX family has only been generated for these elements). The result is a test set with 10 vertical ionization potentials (N_2 , F_2 , C_4 , CF_4 , BF , BN , CO_2 , CO , O_3 , C_6F_6) and 6 vertical electron affinities corresponding to the experimentally-bound anions (F_2 , C_4 , BN , CO , O_3 , C_6F_6). In order to directly compare to existing benchmark data,^{38,39} the results presented in this section used the “pure” PBE functional at the G_0W_0 level without basis set extrapolation.

Table 1 shows the vertical ionization potentials (VIP), in eV, obtained with the different basis sets tested. It is immediately apparent that the core-rich basis sets converge as fast as the traditional cc-pVnZ and def2-nZVP families. In contrast with the CLBES case, the VIPs computed with all basis set families tend to the same value.

Slightly larger variations in the rate of convergence for each basis set family can be seen in the vertical electron affinities (VEA) series shown in Table 2. Here, it is evident that the core-rich basis sets offer better VEAs at the double- and triple- ζ level than the corresponding cc-pVnZ and def2-nzvp sets, especially for molecules containing fluorine atoms. The good performance of the core-rich basis sets for VEAs is partly explained by the presence of some more diffuse basis functions. Table 3 shows the smallest exponent of each angular momentum for all double- ζ basis sets. To facilitate, all exponents are given as the fraction of the corresponding def2-SVP ones. It is evident that the ccX-nZ family is consistently more diffuse than their counterparts, hence its good performance for GW VEAs calculations, but the story is not as evident for the other core-rich families. The remaining difference is explained by the additional variation flexibility core-rich basis set have in comparison to standard ones. For example, the uncontracted def2-SVP VEAs for F_2 and C_6F_6 (-1.40 eV and -1.45 eV, respectively) show a slight improvement toward the basis set limit with respect to the standard def2-SVP.

Table 1: Vertical ionization potentials (in eV) obtained at the $G_0W_0@PBE$ level of theory. Experimental results were taken from References 40–47.

	N ₂	F ₂	C ₄	CF ₄	BF	BN	CO ₂	CO	O ₃	C ₆ F ₆
Exptl.	15.58	15.70	12.54	16.20	11.00		13.77	14.01	12.73	10.20
def2-svp	14.48	14.39	10.40	14.67	10.19	10.67	12.61	13.15	11.50	8.77
def2-tzvp	14.72	14.80	10.66	15.13	10.40	10.91	13.04	13.43	11.84	9.25
def2-qzvp	14.89	14.96	10.78	15.37	10.56	11.01	13.25	13.57	11.96	9.49
cc-pVDZ	14.32	14.30	10.31	14.45	10.41	10.71	12.46	13.18	11.32	8.57
cc-pVTZ	14.81	14.84	10.70	15.15	10.58	10.94	13.06	13.53	11.75	9.19
cc-pVQZ	14.94	15.05	10.81	15.40	10.60	11.02	13.28	13.63	11.94	9.47
cc-pV5Z	14.98	15.10	10.86	15.50	10.61	11.06	13.37	13.65	12.03	9.59
pcSseg-1	14.56	14.58	10.46	14.89	10.45	10.82	12.80	13.37	10.90	9.04
pcSseg-2	14.73	14.79	10.67	15.18	10.42	10.92	13.06	13.45	11.90	9.36
pcSseg-3	14.92	14.94	10.81	15.38	10.55	11.02	13.27	13.59	11.98	9.54
pcSseg-4	14.96	15.03	10.86	15.48	10.56	11.05	13.36	13.64	12.03	9.61
pcJ-1	14.56	14.65	10.49	14.97	10.51	10.85	12.86	13.42	11.69	9.09
pcJ-2	14.84	14.93	10.78	15.32	10.56	10.97	13.17	13.55	11.93	9.44
pcJ-3	14.93	14.99	10.82	15.42	10.53	11.03	13.30	13.60	12.00	9.56
pcJ-4	14.95	15.06	10.87	15.51	10.55	11.06	13.38	13.64	12.04	9.63
ccX-DZ	14.52	14.63	10.46	14.97	10.39	10.76	12.80	13.27	11.11	9.18
ccX-TZ	14.78	14.84	10.71	15.25	10.48	10.94	13.13	13.49	11.92	9.42
ccX-QZ	14.91	14.99	10.82	15.43	10.52	11.03	13.30	13.60	12.00	9.56
ccX-5Z	14.96	15.07	10.87	15.52	10.54	11.06	13.38	13.64	12.04	9.63

It is reassuring to see that the core-rich basis sets are very well balanced to describe both core- and valence-excitations at the GW level.

5 Summary

Convergence of the core-level binding energies obtained in GW calculations have often been deemed hard to achieve with respect to the basis set quality. Here, we have shown that the def2- $nzvp$ and cc-p VnZ basis sets families are shown to have substantial contraction errors when used to compute GW core-level binding energies of molecules containing first row atoms. Furthermore, we have shown that basis set extrapolations using such bases might be unreliable, as in the case of

Table 2: Vertical electron affinities (in eV) obtained at the $G_0W_0@PBE$ level of theory. Experimental values taken from References 48- 51.

	F ₂	C ₄	BN	CO	O ₃	C ₆ F ₆
Exptl.	1.24	3.88	3.16	1.33	2.10	0.70
def2-svp	-1.61	2.14	3.49	-2.00	0.74	-1.54
def2-tzvp	0.17	2.72	3.81	-0.97	1.89	-0.96
def2-qzvp	0.70	2.94	3.95	-0.67	2.30	-0.66
cc-pVDZ	-1.63	1.95	3.27	-2.23	0.51	-1.73
cc-pVTZ	-0.23	2.63	3.72	-1.19	1.62	-1.32
cc-pVQZ	0.43	2.90	3.92	-0.78	2.13	-0.78
cc-pV5Z	0.87	3.03	4.01	-0.55	2.45	-0.48
pcSseg-1	-0.88	2.23	3.49	-1.74	1.13	-1.19
pcSseg-2	0.38	2.77	3.84	-0.88	2.07	-0.84
pcSseg-3	0.85	2.98	3.99	-0.59	2.42	-0.55
pcSseg-4	1.00	3.06	4.04	-0.49	2.56	-0.26
pcJ-1	-0.80	2.24	3.51	-1.72	1.18	-1.13
pcJ-2	0.53	2.81	3.88	-0.83	2.17	-0.91
pcJ-3	0.93	3.00	4.01	-0.55	2.48	-0.43
pcJ-4	1.04	3.07	4.04	-0.47	2.59	-0.21
ccX-DZ	0.36	2.52	3.76	-1.15	1.90	-1.08
ccX-TZ	0.73	2.86	3.92	-0.72	2.28	-0.62
ccX-QZ	0.95	3.01	4.01	-0.55	2.49	-0.37
ccX-5Z	1.05	3.07	4.04	-0.47	2.59	-0.22

the def2- $nzvp$ family, or might yield underestimates due to the very slow convergence seen with the cc-p VnZ family.

Exposing all the core basis functions by uncontracting either basis set adds enough flexibility to the calculation, shifting the binding energies to higher values and showing much faster convergence behavior. The same behavior can be observed with partially uncontracted basis sets, like pcJ- n , pcSseg- n , and ccX- n families. In particular, both pcJ- n and ccX- n seem very well suited to study GW core-level binding energies.

The quality of GW valence charged excitations does not degrade by using one of the core-rich basis sets. In fact, vertical electron affinities achieve faster convergence rates.

We therefore recommend the use of either pcJ- n or ccX- nZ families for core-level and valence

Table 3: Diffuse exponents present in each double- ζ basis set tested as a fraction of the corresponding def2-svp exponent.

	Boron			Carbon		
	S	P	D	S	P	D
def2-svp	1.00	1.00	1.00	1.00	1.00	1.00
cc-pVDZ	1.25	0.99	0.69	1.22	0.99	0.69
pcSseg-1	1.19	0.90	1.40	1.14	0.89	1.00
pcJ-1	1.27	0.90	1.40	1.23	0.89	1.00
ccX-DZ	0.89	0.52	0.69	0.85	0.50	0.55

	Nitrogen			Oxygen		
	S	P	D	S	P	D
def2-svp	1.00	1.00	1.00	1.00	1.00	1.00
cc-pVDZ	1.20	1.00	0.82	1.18	1.00	0.99
pcSseg-1	1.12	0.89	0.90	1.12	0.89	0.83
pcJ-1	1.22	0.89	0.90	1.20	0.89	0.83
ccX-DZ	0.83	0.49	0.82	0.81	0.44	0.99

	Fluorine		
	S	P	D
def2-svp	1.00	1.00	1.00
cc-pVDZ	1.17	1.00	1.17
pcSseg-1	1.12	0.89	0.79
pcJ-1	1.20	0.89	0.79
ccX-DZ	0.80	0.45	1.17

GW calculations. However, it is important to note that both basis sets do impact the speed of the calculation. A more cost-effective choice might be the use of an uncontracted standard basis set, like cc-pVnZ and def2-nzvp. Note that standard J, JK, an RI fitting bases are not recommended when using uncontracted basis sets.

Another fact worth noting is that the current study focused only on the elements C to F (present in the CORE65 database); therefore the conclusions of this paper might not necessarily hold when applying GWA to molecules containing heavier elements.

Acknowledgement

The authors acknowledge funding from the Center for Scalable and Predictive methods for Excitation and Correlated phenomena (SPEC), which is funded by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, the Division of Chemical Sciences, Geosciences, and Biosciences. This research also benefited from computational resources provided by EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at the Pacific Northwest National Laboratory (PNNL). PNNL is operated by Battelle Memorial Institute for the United States Department of Energy under DOE contract number DE-AC05-76RL1830. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

Supporting Information Available

Tables with all individual core-level binding energies (in eV) used to make the plots shown in Figures 1–5. Comparison of mean errors and mean absolute errors with respect to experiment.

References

- (1) Norman, P.; Dreuw, A. Simulating X-ray spectroscopies and calculating core-excited states of molecules. *Chemical Reviews* **2018**, *118*, 7208–7248.
- (2) Milne, C. J.; Penfold, T. J.; Chergui, M. Recent experimental and theoretical developments in time-resolved X-ray spectroscopies. *Coordination Chemistry Reviews* **2014**, *277*, 44–68.
- (3) Halbert, L.; Vidal, M. L.; Shee, A.; Coriani, S.; Pereira Gomes, A. S. Relativistic EOM-CCSD for core-excited and core-ionized state energies based on the four-component Dirac-Coulomb(-Gaunt) Hamiltonian. *Journal of Chemical Theory and Computation* **2021**, *17*, 3583–3598.
- (4) Siegbahn, K.; Nordling, C.; Fahlman, A.; Nordberg, R.; Hamrin, K.; Hedman, J.; Johansson, G.; Bergmark, T.; Karlsson, S.-E.; Lindgren, I. et al. *ESCA, Atomic Molecular and Solid State Structure Studied by Means of Electron Spectroscopy*; Almqvist and Wiksells: Uppsala, Sweden, 1967.
- (5) Hedin, L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Physical Review* **1965**, *139*, A796.
- (6) van Setten, M. J.; Costa, R.; Viñes, F.; Illas, F. Assessing *GW* approaches for predicting core level binding energies. *Journal of Chemical Theory and Computation* **2018**, *14*, 877–883.
- (7) Golze, D.; Wilhelm, J.; van Setten, M. J.; Rinke, P. Core-level binding energies from *GW*: An efficient full-frequency approach within a localized basis. *Journal of Chemical Theory and Computation* **2018**, *14*, 4856–4869.
- (8) Golze, D.; Keller, L.; Rinke, P. Accurate absolute and relative core-level binding energies from *GW*. *Journal of Physical Chemistry Letters* **2020**, *11*, 1840.

- (9) Mejia-Rodriguez, D.; Kunitsa, A.; Aprà, E.; Govind, N. Scalable Molecular GW Calculations: Valence and Core Spectra. *Journal of Chemical Theory and Computation* **2021**, *17*, 7504–7517.
- (10) Keller, L.; Blum, V.; Rinke, P.; Golze, D. Relativistic correction scheme for core-level binding energies from *GW*. *The Journal of Chemical Physics* **2020**, *153*, 114110.
- (11) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297–3305.
- (12) Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007.
- (13) Feller, D. Applications of systematic sequences of wave functions to the water dimer. *The Journal of Chemical Physics* **1992**, *96*, 6104–6114.
- (14) Feller, D. The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water. *The Journal of Chemical Physics* **1993**, *98*, 7059–7071.
- (15) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *The Journal of Chemical Physics* **1997**, *106*, 9639–9646.
- (16) Schwartz, C. Importance of angular correlations between atomic electrons. *Physical Review* **1962**, *126*, 1015.
- (17) Kutzelnigg, W.; Morgan III, J. D. Rates of convergence of the partial-wave expansions of atomic correlation energies. *The Journal of Chemical Physics* **1992**, *96*, 4484.
- (18) Burke, K.; Cancio, A.; Gould, T.; Pittalis, S. *J. Chem. Phys.* **2016**, *145*, 054112.
- (19) Cancio, A.; Chen, G. P.; Krull, B. T.; Burke, K. *J. Chem. Phys.* **2018**, *149*, 084116.

- (20) Hung, L.; Bruneval, F.; Baishya, K.; Ögüt, S. Benchmarking the *GW* approximation and Bethe-Salpeter equation for groups IB and IIB atoms and monoxides. *Journal of Chemical Theory and Computation* **2017**, *13*, 2135–2146.
- (21) Bruneval, F.; Maliyov, I.; Lapointe, C.; Marinica, M.-C. Extrapolating unconverged *GW* energies up to the complete basis set limit with linear regression. *Journal of Chemical Theory and Computation* **2020**, *16*, 4399–4407.
- (22) Jensen, F. Segmented contracted basis sets optimized for nuclear magnetic shielding. *Journal of Chemical Theory and Computation* **2015**, *11*, 132–138.
- (23) Ambroise, M. A.; Dreuw, A.; Jensen, F. Probing basis set requirements for calculating core ionization and core excitation spectra using correlated wave function methods. *Journal of Chemical Theory and Computation* **2021**, *17*, 2832–2842.
- (24) Sarangi, R.; Vidal, M. L.; Coriani, S.; Krylov, A. I. On the basis set selection for the calculation of core-level states: different strategies to balance cost and accuracy. *Molecular Physics* **2020**, *118*, e1769872.
- (25) Fouda, A. E. A.; Besley, N. A. Assessment of basis sets for density functional theory-based calculations of core-electron spectroscopies. *Theoretical Chemistry Accounts* **2018**, *137*, 6.
- (26) Loos, P.-F.; Pradines, B.; Scemama, A.; Giner, E.; Toulouse, J. Density-based basis-set incompleteness correction for *GW* methods. *Journal of Chemical Theory and Computation* **2020**, *16*, 1018–1028.
- (27) Jensen, F. The basis set convergence of spin-spin coupling constants calculated by density functional methods. *Journal of Chemical Theory and Computation* **2006**, *2*, 1360–1369.
- (28) Adler, S. L. Quantum theory of the dielectric constant in real solids. *Physical Review* **1962**, *126*, 413.
- (29) Wiser, N. Dielectric constant with local field effect included. *Physical Review* **1963**, *129*, 62.

- (30) Holzer, C.; Klopper, W. Ionized, electron-attached, and excited states of molecular systems with spin-orbit coupling: Two-component GW and Bethe-Salpeter implementations. *The Journal of Chemical Physics* **2019**, *150*, 204116.
- (31) Aprà, E.; Bylaska, E. J.; de Jong, W. A.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; van Dam, H. J. J.; Alexeev, Y.; Anchell, J. et al. NWChem: Past, present, and future. *The Journal of Chemical Physics* **2020**, *152*, 184102.
- (32) Pritchard, B. P.; Chow, E. Horizontal vectorization of electron repulsion integrals. *Journal of Computational Chemistry* **2016**, *37*, 2537–2546.
- (33) Huang, H.; Chow, E. Accelerating Quantum Chemistry with Vectorized and Batched Integrals. SC18: International Conference for High Performance Computing, Networking, Storage and Analysis. 2018.
- (34) Atalla, V.; Yoon, M.; Caruso, F.; Rinke, P.; Scheffler, M. Hybrid density functional theory meets quasiparticle calculations: A consistent electronic structure approach. *Physical Review B* **2013**, *88*, 165122.
- (35) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic generation of auxiliary basis sets. *Journal of Chemical Theory and Computation* **2017**, *13*, 554–562.
- (36) Jensen, F. Basis set convergence of nuclear magnetic shielding constants calculated by density functional methods. *Journal of Chemical Theory and Computation* **2008**, *4*, 719–727.
- (37) van Setten, M. J.; Caruso, F.; Sharifzadeh, S.; Ren, X.; Scheffler, M.; Liu, F.; Lischner, J.; Lin, L.; Deslippe, J. R.; Louie, S. G. et al. $GW100$: Benchmarking G_0W_0 for molecular systems. *Journal of Chemical Theory and Computation* **2015**, *11*, 5665.
- (38) van Setten, M. $GW100$. <https://www.github.com/setten/GW100>, [Online; accessed 09-March-2021].

- (39) Govoni, M.; Galli, G. *GW100: Comparison of methods and accuracy of results obtained with the WEST code. Journal of Chemical Theory and Computation* **2018**, *14*, 1895–1909.
- (40) Trickl, T.; Cromwell, E. F.; Lee, Y. T.; Kung, A. H. State-selective ionization of nitrogen in the $X^2\Sigma_g^+$ $v^+ = 0$ and $v^+ = 1$ states by two-color (1 + 1) photon excitation near threshold. *The Journal of Chemical Physics* **1989**, *91*, 6006–6012.
- (41) Van Lonkhuyzen, H.; De Lange, C. High-resolution UV photoelectron spectroscopy of diatomic halogens. *Chemical Physics* **1984**, *89*, 313–322.
- (42) Ramanathan, R.; Zimmerman, J. A.; Eyler, J. R. Ionization potentials of small carbon clusters. *The Journal of Chemical Physics* **1993**, *98*, 7838–7845.
- (43) Bieri, G.; Åsbrink, L.; Von Niessen, W. 30.4-nm He(II) Photoelectron spectra of organic molecules: Part IV. Fluoro-compounds (C, H, F). *Journal of Electron Spectroscopy and Related Phenomena* **1981**, *23*, 281–322.
- (44) Farber, M.; Srivastava, R. D. Electron and thermal dissociation of $\text{BF}_3(\text{g})$. *The Journal of Chemical Physics* **1984**, *81*, 241–244.
- (45) Eland, J. H. D.; Berkowitz, J. Photoionization mass spectrometry of HI and DI at high resolution. *The Journal of Chemical Physics* **1977**, *67*, 5034–5039.
- (46) Potts, A.; Williams, T. The observation of “forbidden” transitions in He II photoelectron spectra. *Journal of Electron Spectroscopy and Related Phenomena* **1974**, *3*, 3–17.
- (47) Katsumata, S.; Shiromaru, H.; Kimura, T. Photoelectron Angular Distribution and Assignment of Photoelectron Spectrum of Ozone. *Bulletin of the Chemical Society of Japan* **1984**, *57*, 1784–1788.
- (48) Ayala, J. A.; Wentworth, W. E.; Chen, E. C. M. Electron attachment to halogens. *The Journal of Physical Chemistry* **1981**, *85*, 768–777.

- (49) Arnold, D. W.; Bradforth, S. E.; Kitsopoulos, T. N.; Neumark, D. M. Vibrationally resolved spectra of C₂–C₁₁ by anion photoelectron spectroscopy. *The Journal of Chemical Physics* **1991**, *95*, 8753–8764.
- (50) Arnold, D. W.; Xu, C.; Kim, E. H.; Neumark, D. M. Study of low-lying electronic states of ozone by anion photoelectron spectroscopy of O₃⁻. *The Journal of Chemical Physics* **1994**, *101*, 912–922.
- (51) Eustis, S. N.; Wang, D.; Bowen, K. H.; Naresh Patwari, G. Photoelectron spectroscopy of hydrated hexafluorobenzene anions. *The Journal of Chemical Physics* **2007**, *127*, 114312.