

# Supplemental Material

## I. SPIN-ORBIT LOWERING

In the caption to Table III of the main paper we said that details about the derivation of the spin-orbit lowering value we used ( $-12.702 \text{ cm}^{-1}$ ) would be presented in the Supplemental Material. All numbers in the weighted averaging of our Table below come from [1].

Table I. Details about the spin-orbit lowering value.

System	Weighted Averaging Calculation	Result [ $\text{cm}^{-1}$ ]
$\text{C}(^3P)$	$(0.00 \times 1 + 16.40 \times 3 + 43.40 \times 5) / 9$	29.578
$\text{C}^+(^2P)$	$(0.00 \times 2 + 63.42 \times 4) / 6$	42.280
Difference		-12.702

After the completion of this work, Haris and Kramida published more accurate results which lead to the following [2]:

Table II. Details about the spin-orbit lowering value of [2].

System	Weighted Averaging Calculation	Result [ $\text{cm}^{-1}$ ]
$\text{C}(^3P)$	$(0.00 \times 1 + 16.4167130 \times 3 + 43.4134567 \times 5) / 9$	29.590 825
$\text{C}^+(^2P)$	$(0.00 \times 2 + 63.395 \times 4) / 6$	42.263 333
Difference		-12.672 508

## II. EXPERIMENTAL ENERGIES

Table III. Details about the experimental energies (in  $\text{cm}^{-1}$ ). The line labelled by [\*] is the Chang & Geller value but using 90883.854(15)-63.39509(2) as described in [2] but never calculated/reported there.

Author(s)	Year	Reference	$^3P_0 \rightarrow ^2P_{1/2}$	SO Lowering	Difference in COGs
Johansson	1966	[3]	90820.420(100)	-	-
Chang & Geller	1998	[1]	90820.469(015)	12.702	90833.171
[*]	[*]	[*]	90820.458(015)	-	-
Haris and Kramida	2017	[2]	90820.310(030)	-	-
Haris and Kramida	2017	[2]	90820.348(009)	12.672508	90833.021

## III. ESTIMATED UNCERTAINTIES FROM THE FCIQMC CALCULATIONS

In Section 1C of the main paper, we said that we would present details about how we arrived at our FCIQMC energies and uncertainties in them, for the aug-cc-pCV7Z and aug-cc-pCV8Z basis sets. Our best FCIQMC calculations (largest number of walkers, and largest number of configuration interaction Slater determinants included in  $|\psi_{\text{trial}}\rangle$ ) for the neutral atom and cation, for each basis set, are in red font in Table IV, and the stochastic uncertainties on their last two digits are in parentheses.

Table IV. Details about the FCIQMC calculations.

			$E = \frac{\langle \psi_{N_{\text{trial}}}   H_{\text{NR-CPN}}   \psi_{\text{FCIQMC}} \rangle}{\langle \psi_{N_{\text{trial}}}   \psi_{\text{FCIQMC}} \rangle}$	$E_{\text{ionization}}$
$N_{\text{walkers}}$	$N_{\text{trial}}$	Uncertainty	C( $2^3P$ ) [ $E_{\text{Hartree}}$ ]	C $^+$ ( $2^2P$ ) [ $E_{\text{Hartree}}$ ] [ $\text{cm}^{-1}$ ]
aug-cc-pCV7Z				
$64 \times 10^6$	1		-37.844 251 5(9)	-37.430 345 0(3) 90 841.955
	1000		-37.844 251 5(8)	-37.430 345 0(1) 90 841.977
		$\Delta E_{\text{trial}}$	< 0.000 000 1	< 0.000 000 1 < 0.02
$128 \times 10^6$	1		-37.844 251 5(8)	-37.430 345 0(2) 90 841.977
	1000		-37.844 251 5(5)	-37.430 345 1(1) 90 841.955
		$\Delta E_{\text{trial}}$	< 0.000 000 1	< 0.000 000 1 < 0.02
		$\Delta E_{\text{initiator}}$	< 0.000 000 1	< 0.000 000 1 < 0.02
aug-cc-pCV8Z				
$64 \times 10^6$	1		-37.844 355 5(9)	-37.430 412 3(5) 90 850.031
	1000		-37.844 355 5(9)	-37.430 412 3(5) 90 850.031
		$\Delta E_{\text{trial}}$	< 0.000 000 1	< 0.000 000 1 < 0.02
$128 \times 10^6$	1		-37.844 355 5(8)	-37.430 412 5(5) 90 849.987
	1000		-37.844 355 5(8)	-37.430 412 5(5) 90 849.987
		$\Delta E_{\text{trial}}$	< 0.000 000 1	< 0.000 000 1 < 0.02
		$\Delta E_{\text{initiator}}$	< 0.000 000 1	< 0.000 000 2 < 0.05

In all cases displayed in Table IV, the difference between projecting onto only the Hartree-Fock determinant, or projecting onto a trial wavefunction with the leading 1000 Slater determinants is at most 0.000 000 1  $E_{\text{Hartree}}$ . Therefore we assign  $\Delta E_{\text{trial}}$  to be no larger than 0.000 000 1  $E_{\text{Hartree}}$  or 0.02  $\text{cm}^{-1}$ .

In all cases displayed in Table IV, the difference between using 64 million walkers and 128 million walkers was no larger than 0.000 000 2  $E_{\text{Hartree}}$ , and since the initiator approximation vanishes in the limit of large walker number, we assign  $\Delta E_{\text{initiator}}$  to be no larger than 0.000 000 2  $E_{\text{Hartree}}$  or 0.05  $\text{cm}^{-1}$ . For aug-cc-pCV7Z, the difference between 64 million and 128 million walkers was no larger than 0.000 000 1  $E_{\text{Hartree}}$ , so  $\Delta E_{\text{initiator}}$  was assigned to be no larger than 0.02  $\text{cm}^{-1}$ .

The numbers in parentheses, which denote the stochastic fluctuations in each number of Table IV, were smaller than  $\Delta E_{\text{initiator}}$  and  $\Delta E_{\text{trial}}$  in all cases. So the final energy ionization energies have uncertainties dominated by initiator error. To be extra safe, we further increased our uncertainty estimates based on the formula for adding two independent uncertainties:

$$\sqrt{\Delta_{\text{trial}}^2 + \Delta_{\text{initiator}}^2}, \quad (1)$$

and this increases the ionization energy uncertainties to 0.028 for aug-cc-pCV7Z and 0.054 for aug-cc-pCV8Z. Our final energies are therefore reported in Table IV of the main paper to be **90841.955(028)** for aug-cc-pCV7Z and **90 849.987(054)** for aug-cc-pCV8Z and if there are any further doubts about the accuracy of the FCIQMC energies, we note that both of these uncertainties are more than an order of magnitude smaller than our final discrepancy with experiment of 0.872  $\text{cm}^{-1}$ , which is primarily from neglect of higher-order QED terms and from basis set incompleteness.

#### IV. BASIS SET EXTRAPOLATION

The next table shows, for each basis set size  $X$ , the values for the correlation energies ( $E_X$ ) which are needed to extrapolate to the CBS (complete basis set) limit using the formulas given in Eqs (1-4). Hartree-Fock energies are labeled ROHF to denote that we used Restricted Open Shell Hartree-Fock.

Table V. Correlation energies (in  $E_{\text{Hartree}}$ ) used for the basis set extrapolations.

Basis set	ROHF	FCI	$E_X$
$\text{C}(2^3P)$			
aug-cc-pCV5Z	-37.688 648 2	-37.842 955	-0.154 267 695
aug-cc-pCV6Z	-37.688 687 3	-37.843 840 5	-0.155 153 195
aug-cc-pCV7Z	-37.688 692 7	-37.844 251 5	-0.155 558 777
aug-cc-pCV8Z	-37.688 693 6	-37.844 355 5	-0.155 661 899
$\text{C}^+(2^2P)$			
aug-cc-pCV5Z	-37.292 242 5	-37.429 265 3	-0.137 022 886
aug-cc-pCV6Z	-37.292 283 0	-37.430 004	-0.137 720 959
aug-cc-pCV7Z	-37.292 289 3	-37.430 345 05	-0.138 062 009
aug-cc-pCV8Z	-37.292 290 6	-37.430 412 5	-0.138 123 149

Next, for each basis set size  $X$ , we use Eqs (1) and (3) with  $n = 3.5$  to obtain one CBS estimate of the correlation energy; and Eqs (2) and (4) with  $n = 4$ , to obtain a second CBS estimate of the correlation energy. These correlation energies are added to the CBS estimate of the Hartree-Fock energy, which we estimate to be the aug-cc-pCV8Z Hartree-Fock energy, to get CBS estimates of the total energies of C and  $\text{C}^+$ . While it might seem from the Table that the ROHF energies at aug-cc-pCV8Z are still not good estimates of the CBS values, we note that the energy *differences* between the C and  $\text{C}^+$  ROHF energies, *are* indeed converged much more than the total energies (a phenomenon quite common with Gaussian basis sets since the errors in describing the shape of the electronic wavefunction for C are similar to the errors for  $\text{C}^+$ , which is why our ionization energy agrees with experiment by more than an order of magnitude better than the diffusion Monte Carlo result despite diffusion Monte Carlo having lower variational total energies). Therefore for each  $X > 5$ , we have a CBS estimate of the ionization energy obtained with one formula (Eqs. 1 and 3 with  $n = 3.5$ ) and a CBS estimate of the ionization energy obtained with another formula (Eqs. 2 and 4 with  $n = 4$ ). For each  $X$ , these two estimates of the ionization energy can be averaged, and the unbiased variance around the mean can be calculated. This is what we present in the next table.

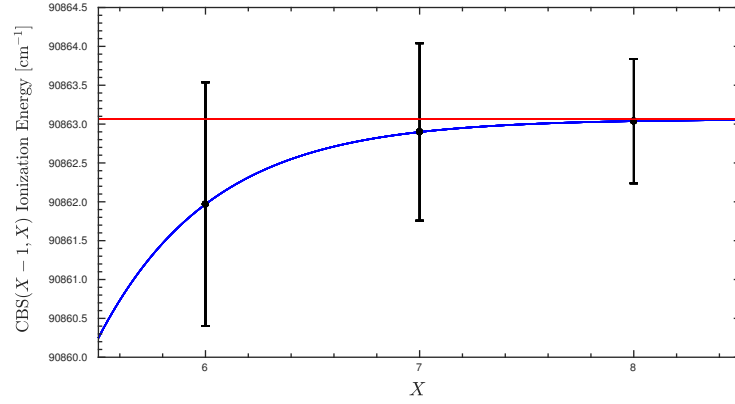
Table VI. Extrapolated ionization energies.

	Ionization Energy [ $\text{cm}^{-1}$ ]	Unbiased Variance [ $\text{cm}^{-1}$ ]
CBS(5,6)	90861.9685	$\pm 1.5682$
CBS(6,7)	90862.8988	$\pm 1.1407$
CBS(7,8)	90863.0375	$\pm 0.8004$

These three CBS ionization energy estimates have been fitted to an exponential of the form  $A - Be^{-CX}$  and plotted in the Figure below.

The fact that all points lie roughly between  $90862 \text{ cm}^{-1}$  and  $90863 \text{ cm}^{-1}$  makes us doubt that the final point  $90863.0375(8004) \text{ cm}^{-1}$  is incorrect by more than  $1 \text{ cm}^{-1}$ .

Figure 1. Fit of the extrapolated ionization energies to the function  $A - Be^{-CX}$ .



## V. X2C AND DBOC

In section 1E we claimed that our X2C (exact 2-component) calculations and DBOC (diagonal Born-Oppenheimer correction) calculations were converged with respect to the size of the basis set and level of correlation used. The table below shows that all numbers are converged on the scale of about  $0.05 \text{ cm}^{-1}$ .

Table VII. Basis set and correlation convergence of the X2C and DBOC corrections to the  $\text{C}(^3P) \rightarrow \text{C}^+(^2P)$  ionization energy in  $\text{cm}^{-1}$ . The difference between CCSDT and FCI (for X2C) and CCSDT and CCSDTQ (for DBOC) for the smallest basis sets is denoted by  $\Delta_{\text{correlation}}$  and the values at aug-cc-pCV3Z-unc are added to the CCSDT energies for all larger basis sets for X2C (and the same is done with aug-cc-pCV4Z for DBOC), resulting in FCI estimates presented in *italic* font. The numbers in parentheses for the CBS (complete basis set) estimates denote our assigned uncertainties in the last digits presented. CBS values and uncertainties are estimated using the difference between the values obtained at the two largest basis sets used (this uncertainty is likely conservative for X2C but too lenient for DBOC, based on how the numbers were changing in smaller basis sets. The X2C calculations are done with basis sets that were uncontracted. X2C uses ROHF, whereas DBOC uses UHF.

	aug-cc-pCV2Z	aug-cc-pCV3Z	aug-cc-pCV4Z	aug-cc-pCV5Z	aug-cc-pCV6Z	CBS
X2C						
CCSDT	-31.055	-30.144	-30.037	-30.004	-30.001	-29.999(003)
FCI	-31.074	-30.168	<i>-30.060</i>	<i>-30.028</i>	<i>-30.025</i>	<i>-30.023(050)</i>
$\Delta_{\text{correlation}}$	-0.019	-0.024	<i>-0.024</i>	<i>-0.024</i>	<i>-0.024</i>	<i>-0.024(050)</i>
DBOC						
CCSDT	-0.147	0.185	0.223	0.225		-0.227(002)
CCSDTQ	-0.142	0.193	0.231	<i>-0.233</i>		<i>-0.235(002)</i>
$\Delta_{\text{correlation}}$	0.007	0.008	0.008	<i>0.008</i>		<i>0.008(001)</i>

The DBOC calculations were done for the  $^{12}\text{C}$  isotopologue. However, the weighted average of the DBOCs for  $^{12}\text{C}$  and  $^{13}\text{C}$  is about the same:

Table VIII. Weighted averaging of DBOC correction in  $\text{cm}^{-1}$  at UHF-CCSD/aug-cc-pCV2Z level.

	DBOC correction	Abundance	Product of previous two columns
$^{12}\text{C}$	-0.217	0.9893	-0.21458587565661874
$^{13}\text{C}$	-0.200	0.0107	-0.00214195608061821
Sum of previous two rows			-0.21672783173723695

## VI. FINITE NUCLEAR-SIZE EFFECTS

We estimate the size of the point-nucleus approximation at the CCSD(T)/aug-cc-pCV6Z-unc level. As in the case of the X2C and DBOC corrections, the correlation and basis set dependence is much smaller than  $1 \text{ cm}^{-1}$ , and our final result is that the effect is far smaller than  $1 \text{ cm}^{-1}$ :

Table IX. Finite nuclear size effects are small.

Radius of nucleus [fm]	Ionization energy [ $\text{cm}^{-1}$ ]
0	90753.38064
2.7	90753.38607
Difference in Energy	0.00543

## VII. TIGHT FUNCTION EXPONENTS OPTIMIZED IN THIS WORK

Finally, we provide the tight exponent functions for our aug-cc-pCV7Z and aug-cc-pCV8Z basis sets in case others wish to use them for their own applications.

Table X. Tight exponents optimized in our work.

	aug-cc-pCV7Z	aug-cc-pCV8Z
<i>s</i> -type	276.1200 158.3000 90.7500 52.0260 29.8260 17.0990	365.5200 232.2000 147.5100 93.7070 59.5290 37.8170 24.0240
<i>p</i> -type	299.2000 149.4600 74.6570 37.2930 18.6290 9.3054	372.1000 198.9500 106.3700 56.8680 30.4050 16.2560 8.6911
<i>d</i> -type	255.6300 111.5700 55.1700 27.2820 13.4910	337.9800 191.2000 108.1700 61.1930 34.6180 19.5840
<i>f</i> -type	132.2600 56.3750 24.0300 10.2430	207.9100 103.1600 51.1850 25.3970 12.6010
<i>g</i> -type	94.4880 36.7620 14.3020	155.5200 72.6010 33.8910 15.8210
<i>h</i> -type	66.2270 22.6800	100.2300 39.5520 15.6080
<i>i</i> -type	48.3200	78.6910 28.9410
<i>k</i> -type	N/A	51.2410

- 
- [1] E. S. Chang and M. Geller, *Physica Scripta* **58**, 326 (1998).
  - [2] K. Haris and A. Kramida, (2017), arXiv:1704.07474.
  - [3] L. Johansson, *Arkiv Fysik* **31**, 201 (1966).