

# Computer-predicted ionization energy of carbon within $1\text{ cm}^{-1}$ of the best experiment

Nike Dattani\*

*Harvard-Smithsonian Center for Astrophysics, Atomic and Molecular Physics Division, 02138, Cambridge, MA, USA,  
McMaster University, Department of Chemistry and Chemical Biology, 606-8103, Hamilton, ON, Canada, and  
Kyoto University, Fukui Institute for Fundamental Chemistry, 606-8103, Kyoto, Japan.*

Giovanni LiManni†

*Max Planck Institute for Solid State Systems, Department of Electronic Structure Theory, Stuttgart, Germany.*

David Feller‡

*Washington State University, Department of Chemistry, Pullman, Washington 99164-4630, USA, and  
University of Alabama, Department of Chemistry, Tuscaloosa, Alabama 35487-0336, USA,*

Jacek Koput§

*Adam Mickiewicz University, Department of Chemistry, 61-614 Poznan, Poland.*

(Dated: 28th January 2018)

We show that we can predict the first ionization energy of carbon to within  $0.994\text{ cm}^{-1}$  of the experimental value. This is an improvement of more than a factor of 5 over the preceding best prediction in [Phys. Rev. A **81**, 022503] and comparable to the accuracies achieved for the smaller atoms up to boron.

In the last seven years, ionization energies (IEs) have been calculated with unprecedented precision for the Li atom [1], Be atom [2] and B atom [3]. Tight variational bounds for non-relativistic ground state energies assuming a clamped, point-sized nucleus have reached 49 digits in units of Hartree for He [4], 19 digits for Li [5], 12 digits for Be [6], and 11 digits for B [3] (see Table I). Calculated IEs have been made in agreement with experiment to within  $10^{-3}\text{ cm}^{-1}$  for  $^7\text{Li}$ ,  $10^{-1}\text{ cm}^{-1}$  for  $^9\text{Be}$  and  $1\text{ cm}^{-1}$  for  $^{11}\text{B}$  (see Table II).

For the C atom, before this present study, no high-precision calculation had been reported to predict an IE to within  $\sim 1\text{ cm}^{-1}$  agreement with experiment. Table I shows that the method used for the smaller atoms up to boron has not had success for carbon. With twice as many variationally optimizable parameters, one fewer digit was obtained for the B atom than for the Be atom, which also suggests that it would be very difficult to variationally optimize a fully explicitly correlated wavefunction ansatz for atoms and molecules coming from the rest of the periodic table.

The best known variational bound for the non-relativistic, clamped, point-sized nucleus (NR,CPN) ground state energy for C was calculated in 2015 using fixed-node diffusion Monte Carlo (FN-DMC) with the nodes of the electronic wavefunction fixed at the locations of a CISD/cc-pV5Z wavefunction, and the statistical uncertainty based on the stochastic fluctuations was  $\pm 20\text{ }\mu E_{\text{Hartree}}$  [7]. However, the IE for C predicted by FN-DMC was in discrepancy with experiment by more

than  $40\text{ cm}^{-1}$ . In this paper, the approach we use to calculate the NR,CPN energy of the ground state of C is FCIQMC (full configuration interaction quantum Monte Carlo) with basis sets as large as aug-cc-pCV8Z. Table I shows that our NR,CPN energy is at least  $76\text{ }\mu E_{\text{Hartree}}$  higher than the variational upper bound obtained from FN-DMC; but since in our approach, imperfections in the description of the wavefunction for the neutral atom are almost the same as in the cation, the individual errors almost completely cancel when calculating the energy difference. Therefore, with our approach we achieve agreement with experiment that is comparable to what has been seen with the explicitly correlated approach for atoms as big as (but not exceeding) boron.

After adding relativistic and quantum electrodynamics (QED) corrections, and corrections to the clamped nucleus approximation, we obtained an IE for the ground state of C which is in only  $0.994\text{ cm}^{-1}$  disagreement with the best known experimental estimate. While this is not as impressive as the method of variationally optimizing parameters in an explicitly correlated wavefunction ansatz has proven to be for Li and Be, the disagreement with experiment has the same order of magnitude as the latter approach for B (see Table II). We finally note that the approach used in this paper, of calculating FCIQMC on a basis set of non-explicitly correlated orbitals has successfully treated systems with far more electrons (transition metal atoms [8], diatomics [9], larger molecules such as butadiene [10], and even solid state systems [11]), so it is conceivable that the approach used in this paper may in the near future be able to determine (with fair accuracy) the IEs which at present remain experimentally elusive or poorly known. These include arsenic (whose experimental IE has an uncertainty of  $\pm 2\text{ cm}^{-1}$ ), Pm, Pa, Fm, Md, No, Sg, Bh and Hs (whose IEs are only known based on extrapolations of other experimental data and have

\* n.dattani@cfa.harvard.edu

† g.limanni@fkf.mpg.de

‡ dfeller@owt.com

§ koput@amu.edu.pl

uncertainties between  $\pm 140 \text{ cm}^{-1}$  and  $\pm 4000 \text{ cm}^{-1}$ , Rf and Db (whose IEs are only known from theoretical calculations), and Mt, Ds, Rg, Cn, Nh, Fl, Mc, Lv, Ts, and Og (for which no IE is given in NIST’s most recent databases).

## I. METHODOLOGY

We begin with our main result in Table III, which shows that our computer-predicted ionization energy comes mainly from the NR-CPN Hamiltonian. This energy was calculated in four stages which we describe in the sub-sections below: (A) We developed larger core-valence (CV) basis sets than previously available for carbon, (B) we calculated the 1- and 2-electron integrals in these basis sets, (C) we solved the NR,CPN Schrödinger equation at the FCI level in our finite-sized basis sets of two different sizes, and (D) we extrapolated the finite basis set results to estimate the energies at the complete basis set (CBS) limits. Finally, sub-section (E) describes how we added the corrections due to special relativity, QED, and due to the atom having an unclamped nucleus.

Table III. Summary of our main result. All energies are between the hyperfine centre of gravity (hcog) of  $\text{C}(^3P)$  and the hcog of  $\text{C}^+(^2P)$  so the experimental spin-orbit lowering of  $12.672508 \text{ cm}^{-1}$  needs to be subtracted from all numbers to obtain the  $\text{C}(^3P_0) \leftarrow \text{C}^+(^2P_{1/2})$  energy (see Supplemental Material). The experimental uncertainty is a 68% confidence interval, meaning that there is a 32% chance that the true energy is outside the range spanned by the uncertainty.

Hamiltonian		Ionization Energy (Calc - Obs)	
		[ $\text{cm}^{-1}$ ]	[ $\text{cm}^{-1}$ ]
NR-CPN		90 863.037(800)	
X2C		-30.023(050)	
Breit & QED		-0.5	
DBOC		-0.5	
Total (theory)	Present	<b>90 832.014</b>	-0.994
Experiment	2017 [21]	90 833.008(009)	
Theory	2010 [22]	90 838.75	5.74
Theory	2017 [23]	90 840.16	7.15
Theory	2015 [7]	90 786.66	-46.35

### A. Optimization of ‘tight function’ exponents for the aug-cc-pCV7Z and aug-cc-pCV8Z basis sets.

The largest basis sets known for C prior to this work were the (aug-cc-pVXZ,  $X=7,8,9$ ) sets used by Feller in 2016 [24]. These basis sets did not contain ‘tight’ exponent functions for capturing the effects of the correlation

between the core ( $1s^2, 2s^2$ ) electrons and the valence electrons ( $2p^2$ ). The largest known basis set for carbon prior to this work including the CV (core-valence) correction was the aug-cc-pCV6Z [25] set. In this work we start by optimizing the ‘tight’ exponents for the CV correction to Feller’s 2016 aug-cc-pV7Z and aug-cc-pV8Z basis sets, yielding the first aug-cc-pCV7Z basis set for carbon, and the first aug-cc-pCV8Z basis set known.

The final aug-cc-pCVXZ basis sets have  $X$  new tight functions of  $s$ -type,  $X-1$  of  $p$ -type,  $X-2$  of  $d$ -type, and so forth, up to the final  $i$ -type function for  $X=7$  and the final  $k$ -type function for  $X=8$ . The  $j^{\text{th}}$  exponent corresponding to a function of type  $L$  is named  $\gamma_{X,L,j}$ , and is assumed to follow an “even-tempered” model:  $\gamma_{X,L,j} = \alpha_{X,L,j} \beta_{X,L,j}^{j-1}$ .

In the non-linear optimization procedure to obtain  $\alpha_{7,L,j}$  and  $\beta_{7,L,j}$ , the starting values were chosen to be the  $\alpha_{6,L,j}$  and  $\beta_{6,L,j}$  values that were already optimized in [25]. These were then treated as free parameters to minimize the difference between the frozen core and all-electron CISD energies of the carbon atom with all other exponent functions fixed. The MOLPRO program [26] was used to calculate the CISD energies, and the L-BFGS-B program of [27] was used to optimize the free parameters. The  $s$ -type functions were added first, then once they were optimized they were held fixed while the  $p$ -type functions were added and optimized. Then both the  $s$ - and  $p$ -type functions were held fixed while the  $d$ -type functions were added, and so on up to the single  $i$ -type function. The procedure for  $X=8$  was the same, except the procedure continued to  $k$ -type functions, and the starting values came from the newly optimized  $X=7$  case rather than the  $X=6$  case from [25].

The tight exponents optimized in this work for aug-cc-pCV7Z and aug-cc-pCV8Z are presented in the Supplemental Material.

### B. Calculation of 1- and 2-electron integrals including $k$ - and $l$ - functions

The calculation of the 1- and 2-electron integrals for (aug)-cc-p(C)VXZ basis sets with  $X \geq 7$  is not possible with most quantum chemistry packages, since very few software packages support  $k$ - and  $l$ - functions, but for first row elements,  $k$ -functions appear in  $X=7$  basis sets and  $l$ -functions appear when  $X=8$ . To calculate these integrals, we have used a locally modified version of MOLCAS 8.0 [28] which supports larger basis sets. The 1- and 2-electron integrals for C and  $\text{C}^+$  were evaluated in the basis of the optimized CASSCF(6,5) and CASSCF(5,5) orbitals respectively, with the five active orbitals being the  $1s$ ,  $2s$ ,  $2p_x$ ,  $2p_y$  and  $2p_z$  of the C atom/ion. This active space is the minimal active space including all electrons, that is able to provide balanced orbitals for the three degenerate states of the  $^3P$  state of the C atom, or the  $^2P$  state of the  $\text{C}^+$  ion.

Table I. Tight variational upper bounds for total electronic energies. For H, the listed energy is exact, since the Hartree energy unit is defined as twice this energy, and can be related to fundamental SI units through an analytic solution to the relevant Schrödinger equation. VO stands for variational optimization (parameters in a wavefunction ansatz are optimized in attempt to obtain the lowest possible energy). Hylleraas-Log indicates the use of Hylleraas functions supplemented with auxiliary log functions, and ECG( $N$ ) indicates the use of explicitly correlated Gaussians with  $N$  variationally optimizable parameters. FN-DMC stands for fixed node diffusion Monte Carlo.

Total non-relativistic clamped point-nucleus (NR,CPN) energy [Hartree]		Method/Ansatz type	Reference
H 1	-0.5	Analytic	1926 Schrödinger
He 2	-2.903 724 377 034 119 598 311 159 245 194 404 446 696 925 309 838	VO/Hylleraas-Log	2006 Schwartz [4]
Li 3	-7.478 060 323 910 134 843	VO/Hylleraas	2011 Wang [5]
Be 4	-14.667 356 494 9	VO/ECG(4096)	2013 Puchalski [2]
B 5	-24.653 867 537	VO/ECG(8192)	2015 Puchalski [3]
C 6	-37.844 48(2)	FN-DMC	2015 Yang [7]
C 6	<b>-37.844 355 5(08)</b>	FCIQMC/aug-cc-pCV8Z	Present work -
C 6	-37.843 333	VO/ECG(1000)	2013 Bubin [12]

Table II. The most precisely calculated ionization energies for the first 6 atoms, compared to the best known experimental measurements to date. The last column indicates that if aiming for the best precision, an experimental measurement is still the best way to obtain the energy for most atoms, but for Be, the energy has been obtained more precisely *in silico* than in any experiment to date. The value for carbon of **90 832.014** cm<sup>-1</sup> was calculated in the present work.

Transition		Experiment [cm <sup>-1</sup> ]	Theory [cm <sup>-1</sup> ]	Calc - Obs [cm <sup>-1</sup> ]	$\left  \frac{\text{Calc} - \text{Obs}}{\text{Uncertainty in obs}} \right $	More precise
<sup>1</sup> H	H <sup>+</sup> (1 <sup>1</sup> S) ← H (1 <sup>2</sup> S)	-	109 678.771 743 07(10) <sup>a</sup>	-	-	Theory
<sup>4</sup> He	He <sup>+</sup> (1 <sup>2</sup> S) ← He (1 <sup>1</sup> S)	198 310.666 37(2) [15]	198 310.665 07(1) [16]	-0.001 3	65.00	Theory
<sup>3</sup> Li	Li <sup>+</sup> (1 <sup>1</sup> S) ← Li (2 <sup>2</sup> S)	3 487.159 40(18) [17]	43 487.159 0(8) [18]	-0.000 4	2.22	Experiment
<sup>9</sup> Be	Be <sup>+</sup> (2 <sup>2</sup> S) ← Be (2 <sup>1</sup> S)	75 192.64(6) [19]	75 192.699(7) [2]	0.059	0.98	Theory
<sup>11</sup> B	B <sup>+</sup> (2 <sup>1</sup> S) ← B (2 <sup>2</sup> P)	66 928.036(22) [20]	66 927.91(21) [3]	-0.126	5.73	Experiment
<sup>12</sup> C	C <sup>+</sup> (2 <sup>2</sup> P) ← C (2 <sup>3</sup> P)	90 833.008(009) [21]	<b>90 832.014</b>	- 0.994	110.44	Experiment

<sup>a</sup> This number is based on the data in [13] although it is not explicitly written anywhere there. Two of the authors of [13] have presented the number in Table III of [14].

### C. Calculation of NR-CPN energies in finite basis sets without truncating the possible excitation levels (FCIQMC)

A deterministic FCI (full configuration interaction) calculation for the 5e<sup>-</sup> C<sup>+</sup> ion in the aug-cc-pCV7Z basis set would require almost 55 TB of RAM, and for the neutral atom would require more. Therefore we use FCIQMC for all NR-CPN calculations. The method was introduced in [29], and we use the initiator method first described in [30], and the semi-stochastic method as described in [31]. The calculations are performed using the developer version of the software NECI [32].

Within a given Hamiltonian (in this case the NR-CPN Hamiltonian) and basis set, there are three sources of error in the FCIQMC energy calculations:

1. Trial wavefunction error ( $\Delta E_{\text{trial}}$ ), which approaches zero in the limit where the number of determinants used in the trial wavefunction

approaches the number of determinants in the FCIQMC wavefunction;

2. Initiator error ( $\Delta E_{\text{initiator}}$ ), which approaches zero in the limit where the number of walkers  $N_{\text{walkers}}$  gets sufficiently large; and
3. Stochastic error ( $\Delta E_{\text{stoch}}$ ), which for a given number of walkers is estimated as the square root of the unbiased variance among different estimates  $E_i$  of the energy from their mean  $\bar{E}$  after different numbers  $N$  of Monte Carlo macro-iterations (determined using the Flyvbjerg-Petersen blocking analysis [33]) after the walkers have reached equilibrium:

$$\Delta E_{\text{stochastic}} \approx \sqrt{\frac{\sum_{i=1}^N (E_i - \bar{E})^2}{N-1}} = \mathcal{O}(1/\sqrt{N}).$$

Our goal was to obtain all energies to a precision of  $\pm \epsilon$  where  $\epsilon \leq 1\mu E_{\text{Hartree}} \approx 0.2$  cm<sup>-1</sup> (within the basis sets used). To ensure that  $\Delta E_{\text{initiator}}$  can be neglected,

we used a sufficiently large value of  $N_{\text{walkers}}$  for every energy calculation, so that the energy difference between using  $N_{\text{walkers}}$  and  $\frac{1}{2}N_{\text{walkers}}$  was smaller than  $1 \mu E_{\text{Hartree}}$ . Likewise, to ensure that  $\Delta E_{\text{trial}}$  can be neglected, we used a sufficiently large number of determinants in the trial wavefunction for every energy calculation, such that  $\Delta E_{\text{trial}}$  would also be smaller than  $1 \mu E_{\text{Hartree}}$ . We then ran every calculation for enough macro-iterations  $N$  such that  $\Delta E_{\text{stoch}}$  was smaller than  $\Delta E_{\text{trial}}$  and  $\Delta E_{\text{initiator}}$ . Further details are presented in the Supplemental Material, including tables which show that all three sources of error in our final numbers are not larger than we claim.

Table IV. Final NR-CPN energies. The break-down of how these energies were obtained, and the Hartree-Fock energies that were used for the extrapolations are available in the Supplemental Material. Numbers in parenthesis indicate uncertainties in the last digit(s) shown, and their determination is described in the Supplemental Material.

	$C(^3P)$ [ $E_{\text{Hartree}}$ ]	$C(^2P)$ [ $E_{\text{Hartree}}$ ]	$2^3P \rightarrow 2^2P$ [ $\text{cm}^{-1}$ ]
aug-cc-pCV7Z	-37.844 251 5(05)	-37.430 345 1(01)	90 841.955(028)
aug-cc-pCV8Z	-37.844 355 5(08)	-37.430 412 5(05)	90 849.987(054)
Eq.(3), $n = 3.5$	-37.844 528 6	-37.430 523 6	90 863.604
Eq.(4), $n = 4$	-37.844 514 2	-37.430 514 3	90 862.471
Mean ( $\sigma$ )	-37.844 521 4	-37.843 051 9	90 863.037

#### D. Extrapolations to the CBS (complete basis set) limit

We use two different families of formulas to extrapolate the correlation energies (FCIQMC energies with the ROHF energies subtracted out) from  $E_{X-1}$  and  $E_X$  to  $E_{\text{CBS}}$ :

$$E_{\text{CBS}} = E_X - \frac{A}{X^n}, \quad (1)$$

$$E_{\text{CBS}} = E_X - \frac{A}{(X + 1/2)^n}. \quad (2)$$

If we set  $n = 3$  in Eq. (1), we recover the formula originally proposed in [34]. If we set  $n = 4$  in Eq. (2), we recover the formula originally proposed in [35]. If we have values for  $E_X$  at two different  $X$  values, we can eliminate  $A$  in both cases, so Eq.(1) leads to Eq.(3) and Eq.(2) leads to Eq.(4):

$$E_{\text{CBS}} = \frac{X^n E_X - (X-1)^n E_{X-1}}{X^n - (X-1)^n}, \quad (3)$$

$$E_{\text{CBS}} = E_X + \frac{(2X-1)^n (E_X - E_{X-1})}{(1+2X)^n - (2X-1)^n}. \quad (4)$$

As explained on page 5 of [24], extrapolations to the CBS limit using  $n = 3$  in Eq.(1) tend to over-shoot the CBS limit. The value of  $n = 3.5$  was therefore used in [24], and we have used it in this present study. The values of  $E_{\text{CBS}}$  obtained from using  $n = 3.5$  in Eq. (3) and  $n = 4$  in Eq.(4), for  $X = 8$  were added to the ROHF energies for  $X = 8$  and are presented in Table IV. The final NR-CPN energy was taken as the mean of both values obtained from extrapolating the correlation energy and adding it to the ROHF energy with  $X = 8$ , and the uncertainty in this mean was estimated as the square root of the unbiased variance of the two values.

#### E. Estimation of relativistic, QED, and finite nuclear mass corrections

Scalar relativistic corrections were calculated by comparing the energies using the spin-free version of the  $1e^-$  X2C (exact 2-component) Hamiltonian, to the energies of the NR-CPN Hamiltonian. The integrals of our X2C Hamiltonian with ROHF orbitals were done in the CF0UR program, and were calculated at various levels of coupled cluster theory with the MRCC program [36].

Further scalar relativistic effects were included by adding the Breit and QED corrections (including the vacuum polarization and the self-energy terms that together comprise the Lamb shift) from the state-averaged Dirac-Fock calculations done in [22]. The overall contribution from the Breit and QED corrections to the IE for C was  $-0.48 \text{ cm}^{-1}$ .

Diagonal Born-Oppenheimer breakdown corrections (DBOC) to the clamped nucleus approximation were calculated using CF0UR, with MRCC used for the coupled-cluster part. Our value of  $-0.50 \text{ cm}^{-1}$  is more than double the value of  $-0.08 \text{ cm}^{-1}$  estimated in [22], due to including higher levels of correlation.

The basis set and correlation convergence of the X2C and DBOC corrections is shown in the Supplemental Material and the final corrections that contributed to our final computer-predicted ionization energy are presented in Table III.

## II. CONCLUSION

Table III summarizes the various contributions to our value of the IE, and compares our final value to experiment and to three recent theoretical estimates. Our value is  $0.994 \text{ cm}^{-1}$  smaller than the best experimental value.

The best theoretical estimate of the IE before this work was in [22], and was in disagreement with experiment by more than a factor of 5 more than our present result. We believe that this could have been due to any or all of three things: (1) approximations inherent to the F12 approach used for their NR-CPN energy, (2) the perturbative nature of their scalar relativistic corrections (ie. use the mass velocity and Darwin terms, rather than the X2C Hamiltonian used in the present work) and (3) the CCSD approximation made in their DBOC correction (as opposed to the CCSDTQ used in the present work which we have shown appears to be converged to the FCI limit).

- 
- [1] M. Puchalski, D. Kędziera, and K. Pachucki, *Phys. Rev. A* **82**, 062509 (2010).
- [2] M. Puchalski, J. Komasa, and K. Pachucki, *Physical Review A* **87**, 030502 (2013).
- [3] M. Puchalski, J. Komasa, and K. Pachucki, *Physical Review A* **92**, 062501 (2015).
- [4] C. Schwartz, *ArXiv Mathematical Physics e-prints* (2006), math-ph/0605018.
- [5] L. M. Wang, Z.-C. Yan, H. X. Qiao, and G. W. F. Drake, *Phys. Rev. A* **83**, 034503 (2011).
- [6] M. Puchalski, D. Kędziera, and K. Pachucki, *Physical Review A* **87**, 032503 (2013).
- [7] Y. Yang, I. Kylänpää, N. M. Tubman, J. T. Krogel, S. Hammes-Schiffer, and D. M. Ceperley, *The Journal of Chemical Physics* **143**, 124308 (2015).
- [8] R. E. Thomas, G. H. Booth, and A. Alavi, *Physical Review Letters* **114**, 033001 (2015).
- [9] D. Cleland, G. H. Booth, C. Overy, and A. Alavi, *Journal of Chemical Theory and Computation* **8**, 4138 (2012).
- [10] C. Daday, S. Smart, G. H. Booth, A. Alavi, and C. Filippi, *Journal of Chemical Theory and Computation* **8**, 4441 (2012).
- [11] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2012).
- [12] S. Bubin, M. Pavanello, W.-C. Tung, K. L. Sharkey, and L. Adamowicz, *Chemical Reviews* **113**, 36 (2013).
- [13] U. D. Jentschura, S. Kotochigova, E.-O. Le Bigot, P. J. Mohr, and B. N. Taylor, *Phys. Rev. Lett.* **95**, 163003 (2005).
- [14] P. J. Mohr, D. B. Newell, and B. N. Taylor, *Rev. Mod. Phys.* **88**, 035009 (2016).
- [15] D. Z. Kandula, C. Gohle, T. J. Pinkert, W. Ubachs, and K. S. E. Eikema, *Physical Review A* **84**, 062512 (2011).
- [16] K. Pachucki, V. Patkőš, and V. A. Yerokhin, *Physical Review A* **95**, 062510 (2017).
- [17] B. A. Bushaw, W. Nörtershäuser, G. W. F. Drake, and H.-J. Kluge, *Phys. Rev. A* **75**, 052503 (2007).
- [18] M. Puchalski and K. Pachucki, *Physical Review A* **78**, 052511 (2008).
- [19] R. Beigang, D. Schmidt, and P. J. West, *Le Journal de Physique Colloques* **44**, C7 (1983).
- [20] A. E. Kramida and A. N. Ryabtsev, *Physica Scripta* **76**, 544 (2007).
- [21] K. Haris and A. Kramida, (2017), arXiv:1704.07474.
- [22] W. Kloppe, R. A. Bachorz, D. P. Tew, and C. Hättig, *Phys. Rev. A* **81**, 022503 (2010).
- [23] D. Feller, *The Journal of Chemical Physics* **147**, 34103 (2017).
- [24] D. Feller, *The Journal of Chemical Physics* **144**, 014105 (2016).
- [25] A. K. Wilson, T. van Mourik, and T. H. Dunning, *Journal of Molecular Structure: THEOCHEM* **388**, 339 (1996).
- [26] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, “Molpro, version 2015.1, a package of ab initio programs,” (2015), see <http://www.molpro.net>.
- [27] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, *ACM Trans. Math. Softw.* **23**, 550 (1997).
- [28] F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P.-Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Velyazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, *Journal of Computational Chemistry* **37**, 506 (2016).
- [29] G. H. Booth, A. J. W. Thom, and A. Alavi, *The Journal of Chemical Physics* **131**, 054106 (2009).
- [30] D. Cleland, G. H. Booth, and A. Alavi, *The Journal of Chemical Physics* **132**, 041103 (2010).
- [31] N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, *The Journal of Chemical Physics* **142**, 184107 (2015).
- [32] “Public version is available here: [https://github.com/ghb24/NECI\\_STABLE](https://github.com/ghb24/NECI_STABLE).”
- [33] H. Flyvbjerg and H. G. Petersen, *The Journal of Chemical Physics* **91**, 461 (1989).
- [34] W. Kutzelnigg and J. D. Morgan, *The Journal of Chemical Physics* **96**, 4484 (1992).
- [35] J. M. Martin, *Chemical Physics Letters* **259**, 669 (1996).
- [36] M. Kallay, Z. Rolik, J. Csontos, P. Nagy, G. Samu, D. Mester, I. Ladjanszki, L. Szegedy, B. Ladoczki, K. Petrov, M. Farkas, and B. Hegely, “MRCC, a quantum chemical program suite.”

# Supplemental Material

## I. SPIN-ORBIT LOWERING

In the caption to Table III we said that details about the derivation of the spin-orbit lowering value we used ( $-12.672\,508$ ) 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 [ $\text{cm}^{-1}$ ]
$\text{C}(^3P)$	$(0.00 \times 1 + 16.4167130 \times 3 + 43.4134567 \times 5) / 9$	29.590 825
$\text{C}^+(^2P)$	$(0.00 \times 2 + 63.395 \times 4) / 6$	42.263 333
Difference		-12.672 508

## II. ESTIMATED UNCERTAINTIES FROM THE FCIQMC CALCULATIONS

In Section 1C, 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 determinants included in  $|\psi_{\text{trial}}\rangle$ ) for the neutral atom and cation, for each basis set, are in red font in Table II, and the stochastic uncertainties on their last two digits are in parentheses.

Table II. 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	$\text{C}(2^3P) [E_{\text{Hartree}}]$	$\text{C}^+ (2^2P) [E_{\text{Hartree}}]$	$[\text{cm}^{-1}]$
aug-cc-pCV7Z					
$64 \times 10^6$	1		-37.844 251 5(09)	-37.430 345 0(03)	90 841.955
	1000		-37.844 251 5(08)	-37.430 345 0(01)	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(08)	-37.430 345 0(02)	90 841.977
	1000		-37.844 251 5(05)	-37.430 345 1(01)	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(09)	-37.430 412 3(05)	90 850.031
	1000		-37.844 355 5(09)	-37.430 412 3(05)	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(08)	-37.430 412 5(05)	90 849.987
	1000		-37.844 355 5(08)	-37.430 412 5(05)	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 II, 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 II, 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 II, 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-pCV7Z 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  $0.994\,\text{cm}^{-1}$ , which is primarily from neglect of higher-order QED terms and from basis set extrapolation.

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

Basis set	ROHF	FCI	$E_X$
$\text{C}(2^3P)$			
aug-cc-pCV <b>5</b> Z	-37.688 648 2	-37.842 955	-0.154 267 695
aug-cc-pCV <b>6</b> Z	-37.688 687 3	-37.843 840 5	-0.155 153 195
aug-cc-pCV <b>7</b> Z	-37.688 692 7	-37.844 251 5	-0.155 558 777
aug-cc-pCV <b>8</b> Z	-37.688 693 6	-37.844 355 5	-0.155 661 899
$\text{C}^+(2^2P)$			
aug-cc-pCV <b>5</b> Z	-37.292 242 5	-37.429 265 3	-0.137 022 886
aug-cc-pCV <b>6</b> Z	-37.292 283 0	-37.430 004	-0.137 720 959
aug-cc-pCV <b>7</b> Z	-37.292 289 3	-37.430 345 05	-0.138 062 009
aug-cc-pCV <b>8</b> Z	-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.

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

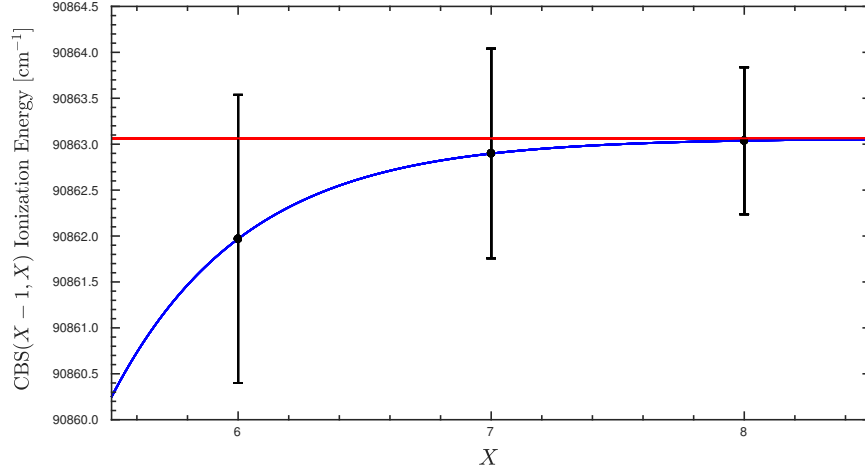
The fact that all points lie roughly between  $90862\,\text{cm}^{-1}$  and  $90863\,\text{cm}^{-1}$  makes us doubt that the final point  $90863.0375(8004)\,\text{cm}^{-1}$  is incorrect by more than  $1\,\text{cm}^{-1}$ , but to the best of our knowledge, this is the first time

Table IV. 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	$\pm 0.8004$

anyone has fitted CBS extrapolations this way and tried to do an “extrapolation of an extrapolation”, so we will see how it looks for nitrogen and oxygen.

Figure 1. Fit of the extrapolated ionization energies.



#### IV. 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 V. 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 the smallest two 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 (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, which we believe to be very conservative. The “unc” in the basis set names is used to denote that the basis sets were uncontracted.

	aug-cc-pCV2Z-unc	aug-cc-pCV3Z-unc	aug-cc-pCV4Z-unc	aug-cc-pCV5Z-unc	aug-cc-pCV6Z-unc	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	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5
FCI	-0.5	-0.5	<i>-0.5</i>	<i>-0.5</i>	<i>-0.5</i>	<i>-0.5</i>
$\Delta_{\text{correlation}}$	0.0	0.0	<i>0.0</i>	<i>0.0</i>	<i>0.0</i>	<i>0.0</i>



## V. FINITE NUCLEAR-SIZE EFFECTS

In the spirit of being careful, 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 VI. Finite nuclear size effects are small.

	Ionization energy [ $\text{cm}^{-1}$ ]
Point Nucleus	90753.38064
Finite Nucleus	90753.38607
Difference	0.00543

## VI. 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 which to use them for their own applications.

Table VII. 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		51.2410

- 
- [1] K. Haris and A. Kramida, (2017), arXiv:1704.07474.