Computer predicted ionization energy of carbon accurate to within ± 1 cm^{-1}

Nike Dattani*

Harvard Smithsonian Center for Astrophysics, 02138, Cambridge, MA, USA.

Giovanni LiManni†

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

David Feller‡

Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, USA,

Jacek Koput§

Department of Chemistry, Adam Mickiewicz University, 60–780 Poznan, Poland.

(Dated: 24th November 2018)

We show that we can predict the first ionization energy in carbon to within 1.076 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. The sum of all error contributions in our calculations up to the fifth power of the fine structure constant is estimated to be at most 0.806 cm^{-1}, which yields our estimate of contributions in the sixth and higher powers to be: 1.076 ± 0.806 cm^{-1}.

In the last seven years, ionization energies (IEs) have been calculated with unprecedented precision for the Li atom [?], Be atom [?] and B atom [?]. 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, 19 digits for Li, 12 digits for Be, and 11 digits for B (see Table I). Calculated IEs have been made in agreement with experiment to within 10^{-3} cm^{-1} for ^{7}Li, 10^{-1} cm^{-1} for ^{9}Be and 1 cm^{-1} for ^{11}B (see Table II).

For the C atom, before this present study, no high-precision calculation had been reported to predict an IE to ~ 1 cm^{-1} agreement with experiment.

Table I demonstrates that with twice as many variationally optimizable parameters, 1 fewer digit was obtained for the B atom than for the Be atom. This suggests that it would be very difficult to variationally optimize a fully explicitly correlated wavefunction ansatz for atoms and molecules coming from most of the periodic table. The best known variational bound for the 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 on the Monte Carlo was ± 20 μE_{Hartree} [?]. However, the IE for C predicted by FN-DMC was in discrepancy with experiment by more than 40 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) in an aug-cc-pCV8Z basis set. Table I shows that our NR,CPN

energy is at least 76 μE_h 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 errors are almost completely eliminated when taking the energy difference, therefore bringing us to 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 sate of C which is within 1.076 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 wavefunciton ansatz has proven to be for Li and Be, the disagreement 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 [?], diatomics [?], larger molecules such as butadiene [?], and even solid state systems [?]), 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 ± 2 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 uncertainties between ± 140 cm^{-1} and ± 4000 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).

---

* ndattani@cfa.harvard.edu
† g.limanni@fkf.mpg.de
‡ dfeller@owt.com
§ koput@amu.edu.pl

Table I. Tight variational upper bounds for atomic 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 relavent Schroedinger 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 Schroedinger	
He	2	-2.903 724 377 034 119 598 311 159 245 194 404 446 696 925 309 838	VO/Hylleraas-Log	2006 Schwartz	[?]
Li	3	-7.478 060 323 910 134 843	VO/Hylleraas	2011 Wang	[?]
Be	4	-14.667 356 494 9	VO/ECG(4096)	2013 Puchalski	[?]
B	5	-24.653 867 537	VO/ECG(8192)	2015 Puchalski	[?]
C	6	-37.844 48(2)	FN-DMC	2015 Yang	[?]
C	6	-37.844 355 5(08)	FCIQMC/aCV8Z	Present work	-
C	6	−37.843 333	VO/ECG(1000)	2013 Bubin	[?]

I. METHODOLOGY

Our calculations in this paper can be divided into five stages: (A) We develop larger core-valence (CV) basis sets than previously available for carbon, (B) we calculate the 1- and 2-electron integrals in these basis sets, (C) we attempt to solve the NR,CPN Schroedinger equation at the FCI level in finite basis sets, (D) we extrapolate the finite basis set results to estimate the energies at the complete basis set (CBS) limtis, and (E) we estimate the size of corrections due to special relativity, QED, and due to the atom having an unclamed nucleus.

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 [?]. 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 [?] 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 jth 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 [?]. 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 [?] was used to calculate the CISD energies, and the L-BFGS-B program of [?] 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 [?].

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 [?] which supports larger basis sets. The 1- and 2-electron integrals for C and 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 ac-

Table II. The most precisely calculated electronic excitation 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 90 832.034(802) cm^{-1} was calculated in the present work.

	Transition	Experiment [cm^{-1}]		Theory [cm^{-1}]		Calc - Obs [cm^{-1}]	$\left\|\frac{\text{Calc - Obs}}{\text{Uncertainty in obs}}\right\|$	More precise
H								
He								
^{7}Li	Li^{+} $(1^1S) \leftarrow$ Li(2^2S)	43 487.159 40(18)	[?]	43 487.159 0(8)	[?]	-0.000 4	57.1	Experiment
^{9}Be	Be^{+} $(2^2S) \leftarrow$ Be(2^1S)	75 192.64(6)	[?]	75 192.699(7)	[?]	0.059	0.98	Theory
^{11}B	B^{+} $(2^1S) \leftarrow$ B(2^2P)	66 928.036(22)	[?]	66 927.91(21)	[?]	-0.126	5.7	Experiment
^{12}C	C^{+} $(2^2P) \leftarrow$ C(2^3P)	90 833.11(10)		90 832.034(802)		1.076	11.08	Experiment

Table III. Summary of our main result.

Hamiltonian	Hamiltonian	Energy in cm^{-1}	(Obs - Calc) in cm^{-1}
NR-CPN	NR-CPN	90 863.324(810)	
X2C	X2C	-30.023(050)	
Breit & QED	Breit & QED	-0.5	
DBOC	DBOC	-0.5	
Present	Theory (total)	90 832.301(810)	0.707
2017	Experiment [?]	90 833.008(009)	
2010	Theory [?]	90 838.75	-5.74
2017	Theory [?]	90 840.16	-7.15
2015	Theory [?]	90 786.66	46.35

tive 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 C$^+$ ion.

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

Deterministic FCI calculations were possible for the 5e$^-$ system C$^+$ in aug-cc-pCV6Z with about 130 GB of RAM to store the Davidson vectors, but were not possible for the 6e$^-$ neutral C atom in this basis set (which would require almost 9 TB of RAM), nor for 5e$^-$ in the aug-cc-pCV7Z basis set (which would require almost 55 TB of RAM).

We therefore use FCIQMC for all NR-CPN calculations. The wavefunction of the relevant system is expanded as a sum of all possible Slater determinants given the basis set (full configuration interaction), and the coefficients are determined by the number of walkers standing on each determinant after a Monte Carlo sampling using the NR-CPN Hamiltonian. The method was introduced in [?], and we use the initiator method first described in [?], and the semi-stochastic method as described in [?]. The calculations are performed using the developer version of the software NECI [?].

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_{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 [?]) after the walkers have reached equilibrium:

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

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^{-1} (within the basis sets used). To ensure that $\Delta E_{\text{initiator}}$ can be neglected, we used a sufficiently large value of N_{walkers} for every energy calculation, so that the energy difference between using N_{walkers} and $\frac{1}{2}N_{\text{walkers}}$ was smaller than 1 μE_{Hartree}. Likewise, to ensure that ΔE_{trial} can be neglected, we used a sufficiently large number of determinants in the trial wavefunction for every energy calculation, such that ΔE_{trial} would also be smaller than 1 μE_{Hartree}. We then ran every calculation for enough macro-iterations N such that ΔE_{stoch} was smaller than ΔE_{trial} and $\Delta E_{\text{initiator}}$.