

# Performance of W4 theory for spectroscopic constants and electrical properties of small molecules

Amir Karton and Jan M. L. Martin\*

*Department of Organic Chemistry, Weizmann*

*Institute of Science, IL-76100 Rehovot, Israel<sup>\*†</sup>*

(Dated: JCP MS# A10.07.0057: submitted July 4, 2010; revised draft August 21, 2010)

Cite as:

A. Karton, J. M. L. Martin, *J. Chem. Phys.* 133, 144102 (2010).

<http://dx.doi.org/10.1063/1.3489113>

---

\* Present address: Department of Chemistry, University of North Texas, Denton, TX 76203-5017. Email: gershom@unt.edu

## Abstract

Accurate spectroscopic constants and electrical properties of small molecules are determined by means of W4 and post-W4 theories. For a set of 28 first- and second-row diatomic molecules for which very accurate experimental spectroscopic constants are available, W4 theory affords near-spectroscopic or better predictions. Specifically, the root-mean-square deviations (RMSD) from experiment are 0.04 pm for the equilibrium bond distances ( $r_e$ ), 1.03 cm<sup>-1</sup> for the harmonic frequencies ( $\omega_e$ ), 0.20 cm<sup>-1</sup> for the first anharmonicity constants ( $\omega_e x_e$ ), 0.10 cm<sup>-1</sup> for the second anharmonicity constants ( $\omega_e y_e$ ), and 0.001 cm<sup>-1</sup> for the vibration-rotation coupling constants ( $\alpha_e$ ). These RMSD imply 95% confidence intervals of about 0.1 pm for  $r_e$ , 2.0 cm<sup>-1</sup> for  $\omega_e$ , 0.4 cm<sup>-1</sup> for  $\omega_e x_e$ , and 0.2 cm<sup>-1</sup> for  $\omega_e y_e$ . We find that post-CCSD(T) contributions are essential to achieve such narrow confidence intervals for  $r_e$  and  $\omega_e$ , but have little effect on  $\omega_e x_e$  and  $\alpha_e$ , and virtually none on  $\omega_e y_e$ . Higher-order connected triples,  $\hat{T}_3$ -(T), improve agreement with experiment for the hydride systems, but their inclusion (in the absence of  $\hat{T}_4$ ) tends to worsen agreement with experiment for the nonhydride systems. Connected quadruple excitations,  $\hat{T}_4$ , have significant and systematic effects on  $r_e$ ,  $\omega_e$ , and  $\omega_e x_e$ , in particular they universally increase  $r_e$  (by up to 0.5 pm), universally reduce  $\omega_e$  (by up to 32 cm<sup>-1</sup>), and universally increase  $\omega_e x_e$  (by up to 1 cm<sup>-1</sup>). Connected quintuple excitations,  $\hat{T}_5$ , are spectroscopically significant for  $\omega_e$  of the nonhydride systems, affecting  $\omega_e$  by up to 4 cm<sup>-1</sup>. Diagonal Born–Oppenheimer corrections have systematic and spectroscopically-significant effects on  $r_e$  and  $\omega_e$  of the hydride systems, universally increasing  $r_e$  by 0.01–0.06 pm and decreasing  $\omega_e$  by 0.3–2.1 cm<sup>-1</sup>. Obtaining  $r_e$  and  $\omega_e$  of the pathologically multireference BN and BeO systems with near-spectroscopic accuracy requires large basis sets in the core-valence CCSD(T) step and augmented basis sets in the valence post-CCSD(T) steps in W4 theory. The triatomic molecules H<sub>2</sub>O, CO<sub>2</sub>, and O<sub>3</sub> are also considered. The equilibrium geometries and harmonic frequencies (with the exception of the asymmetric stretch of O<sub>3</sub>) are obtained with near-spectroscopic accuracy at the W4 level. The asymmetric stretch of ozone represents a severe challenge to W4 theory, in particular the connected quadruple contribution converges very slowly with the basis set size. Finally, the importance of post-CCSD(T) correlation effects for electrical properties, namely dipole moments ( $\mu$ ), polarizabilities ( $\alpha$ ), and first hyperpolarizabilities ( $\beta$ ) is evaluated.

## I. INTRODUCTION

Since the early 1990s, there has been growing interest in calculating accurate *ab initio* anharmonic force fields for small molecules. There are two aspects to be considered: (i) the nuclear vibrational analysis, and (ii) the level of the electronic structure calculations. For diatomic molecules (i) is relatively straightforward using a Dunham analysis, and for rigid triatomic molecules, such as those considered in the present work, (i) entails no serious problems. As for (ii), for systems with mild to moderate nondynamical correlation effects, the accuracy of CCSD(T) basis set limit results is on the order of 5–10  $\text{cm}^{-1}$  for harmonic frequencies<sup>1–5</sup> and 0.1–0.5 pm for bond lengths.<sup>1,2,6,7</sup> To surpass this level of accuracy (or when considering molecules with more pronounced multireference character such as  $\text{C}_2$ , BN, BeO, and  $\text{O}_3$ ) one should consider multireference methods<sup>8–10</sup> or post-CCSD(T) contributions.<sup>4–7,11</sup> Recently, Mintz *et al.*<sup>12</sup> have applied the multireference ccCA method<sup>13</sup> to compute the potential energy curves of  $\text{N}_2$  and  $\text{C}_2$ ; they obtained errors of  $\sim 0.06$  pm for the bond lengths and  $\sim 2$   $\text{cm}^{-1}$  for the harmonic frequencies.

The determination of very precise molecular energies (atomization energies or enthalpies of formation) has been one of the primary goals of composite *ab initio* methods. It has been shown that highly accurate thermochemical protocols such as HEAT<sup>14–16</sup> and W4<sup>17,18</sup> are capable of sub-kJ/mol accuracy on average (see Ref.<sup>19</sup> for a recent review). In particular, W4 theory obtains a root mean square deviation (RMSD) of 0.08 kcal/mol against a test set of 25 first- and second-row small molecules for which very accurate ATcT<sup>20</sup> atomization energies are available. This implies a 95% confidence interval of  $\sim 0.16$  kcal/mol; the mean signed deviation (MSD) of just -0.01 kcal/mol suggests that W4 is free of systematic bias. Worst-case errors for problematic molecules are  $< 1$  kJ/mol: for example, for  $\text{F}_2\text{O}_2$ ,<sup>21</sup>  $\text{C}_2$ ,<sup>18</sup>  $\text{F}_2\text{O}$ ,<sup>21</sup> and  $\text{O}_3$ ,<sup>17</sup> W4 is 0.09, 0.17, 0.18, and 0.23 kcal/mol away from the ATcT atomization energies (the ATcT value for  $\text{C}_2$  is available in Ref.<sup>22</sup> and for the remaining species in Ref.<sup>23</sup>). In our continued work in this field we have shown that contributions from successively higher cluster expansion terms converge increasingly faster with the basis set, as they increasingly reflect nondynamical rather than dynamical correlation.<sup>17,18</sup> Indeed, the fact that W4 theory and related methods can be carried out at all at a realistic cost hinges on this behavior.

In view of the success of W4 theory for thermochemical properties<sup>17,18,21</sup> it is of interest to establish the reliability of W4 (and related) methods for spectroscopic properties based

on energy derivatives with respect to the nuclear coordinates (*e.g.*  $r_e$ ,  $\omega_e$ , and  $\omega_e x_e$ ) and for electrical properties based on energy derivatives with respect to an external static electric field (*e.g.*  $\mu$ ,  $\alpha$ , and  $\beta$ ).

The present study considers a chemically diverse data set of 31 first- and second-row diatomic molecules as well as the triatomic molecules H<sub>2</sub>O, CO<sub>2</sub>, and O<sub>3</sub>. The chosen set, which includes radicals, polar systems, hydrides, and nonhydrides with single and multiple bonds, evidently spans a wide gamut from systems dominated by dynamical correlation (*e.g.* BH and H<sub>2</sub>O) to systems with pathological nondynamical correlation (*e.g.* C<sub>2</sub>, BN, BeO, and O<sub>3</sub>) and all shades in between.

## II. COMPUTATIONAL METHODS

### A. Electronic structure

All the SCF, CCSD, and CCSD(T) calculations were carried out using MOLPRO 2009.1<sup>24</sup> running on the Martin group Linux cluster at the Weizmann Institute. The post-CCSD(T) calculations were carried out using MRCC interfaced to the MOLPRO program suite.<sup>25</sup> The diagonal Born–Oppenheimer correction (DBOC) calculations were carried out using the CFOUR program system.<sup>26</sup> All basis sets employed belong to the correlation consistent family of Dunning and co-workers.<sup>27–31</sup> The notation aug'-cc-pV( $n+d$ )Z indicates the combination of regular cc-pV $n$ Z on hydrogen and aug-cc-pV( $n+d$ )Z on other elements.

The computational protocols of W4, W4.2, and W4.3 theories have been specified and rationalized in great detail elsewhere.<sup>17,18</sup> In the present work the steps involved in W4 theory are divided in the following manner:

- W4[up to CCSD(T)<sub>*fc,nr*</sub>] represents the clamped-nuclei, non-relativistic CCSD(T) infinite basis set limit energy in the frozen-core approximation (in which the 1s orbitals for first-row atoms and the 1s, 2s, and 2p orbitals for second-row atoms are constrained to be doubly occupied in all configurations). The *fc* and *nr* subscripts stand for “frozen-core” and “non-relativistic”. The following extrapolations are used for the HF, CCSD, and (T) contributions:
  - The ROHF-SCF contribution is extrapolated from the aug'-cc-pV(5+d)Z and aug'-cc-pV(6+d)Z basis sets using the Karton-Martin modification<sup>32</sup> of Jensen’s

extrapolation formula.<sup>33</sup>

- The RCCSD valence correlation energy is calculated from these same basis sets. Following the suggestion of Klopper,<sup>34</sup> it is partitioned into singlet-coupled pair energies, triplet-coupled pair energies, and  $\hat{T}_1$  terms. The singlet-coupled and triplet-coupled pair energies are extrapolated using the  $A + B/L^\alpha$  two-point extrapolation formula (where  $L$  is the highest angular momentum present in the basis set) with  $\alpha_S=3$  and  $\alpha_T=5$ , respectively, and the  $\hat{T}_1$  term (which exhibits very weak basis set dependence) is set equal to that in the largest basis set.
- The (T) valence correlation energy is extrapolated from the aug'-cc-pV(Q+d)Z and aug'-cc-pV(5+d)Z basis sets using the  $A + B/L^3$  two-point extrapolation formula. For open-shell systems the Werner-Knowles-Hampel<sup>35</sup> (a.k.a. MOLPRO) definition of the restricted open-shell CCSD(T) energy is employed throughout, rather than the original Watts-Gauss-Bartlett<sup>36</sup> (a.k.a. ACES II/CFour) definition.
- W4[up to CCSD(T)] in addition includes inner-shell correlation and scalar relativistic contributions. The former is extrapolated from RCCSD(T)/aug'-cc-pwCVTZ and RCCSD(T)/aug'-cc-pwCVQZ energies using the  $A + B/L^3$  two-point extrapolation formula, and the latter (in the second-order Douglas-Kroll-Hess approximation<sup>37,38</sup>) is obtained from the difference between non-relativistic CCSD(T)/aug'-pV(Q+d)Z and relativistic CCSD(T)/aug'-cc-pV(Q+d)Z-DK calculations.<sup>31</sup>
- W4[up to CCSDT] additionally includes higher-order connected triples,  $\hat{T}_3-(T)$ , valence correlation contribution extrapolated from the cc-pVDZ and cc-pVTZ basis sets using the  $A + B/L^3$  two-point extrapolation formula.
- In W4[up to CCSDTQ] the connected quadruples,  $\hat{T}_4$ , term is also included. The valence correlation (Q) and  $\hat{T}_4-(Q)$  contributions are calculated with the cc-pVTZ and cc-pVDZ basis sets, respectively. In Refs.<sup>17,18</sup> we found that scaling their sum by 1.10 offers a very reliable (as well as fairly cost-effective) estimate of the basis set limit  $\hat{T}_4$  contribution.
- Adding the connected quintuple,  $\hat{T}_5$ , valence correlation contribution calculated with the *sp* part of the cc-pVDZ basis set — denoted cc-pVDZ(no *d*) — results in full W4

theory. The  $\hat{T}_5$  contribution converges very rapidly with the basis set as it primarily represents static correlation.<sup>17,18</sup>

W4 represents an approximation to the relativistic basis-set limit CCSDTQ5 energy. The DBOC contributions, calculated at the ROHF/aug'-cc-pVTZ level of theory, are reported separately in Tables I–V, and are included in the final error statistics presented in these tables.

For the smaller systems we also consider the post-W4 methods (W4.2 and W4.3) as defined in Ref.<sup>17</sup> The changes in W4.2 and W4.3 relative to W4 are summarized as follows:

- W4.2 theory in addition takes account of the  $\hat{T}_3$ –(T) correction to the core-valence contribution obtained using the cc-pwCVTZ basis set.
- In W4.3 all the valence post-CCSD(T) corrections are additionally upgraded: the  $\hat{T}_3$ –(T) and (Q) corrections are extrapolated from the cc-pVTZ and cc-pVQZ basis sets, the  $\hat{T}_4$ –(Q) and  $T_5$  corrections are calculated with the cc-pVTZ and cc-pVDZ basis sets, respectively, and the  $\hat{T}_6$  correction is calculated with the cc-pVDZ(no  $d$ ) basis set.

Finally, for the pathologically multireference systems in Table VI ( $C_2$ , BN, and BeO) we also consider two additional extensions of W4 theory:

- Replacing the regular correlation-consistent basis sets on electronegative atoms (N and O) with their augmented versions in all the valence post-CCSD(T) steps in W4 theory.
- Extrapolating the CCSD(T) inner-shell correlation contribution from larger basis sets, namely the aug-cc-pwCVQZ and aug-cc-pwCV5Z basis sets.

## B. Vibrational analysis

For each of the diatomic molecules, a 21-point potential energy curve was calculated using the above-mentioned W4 (and related) theories. The single point energy calculations are carried out at bond distances  $r_x = r_e^{expt.} + x$  ( $x = 0, \pm 1, \pm 2, \dots, \pm 10$  pm), where  $r_e^{expt.}$  is the experimental equilibrium bond length. The spectroscopic constants ( $r_e$ ,  $\omega_e$ ,  $\omega_e x_e$ ,  $\omega_e y_e$ , and  $\alpha_e$ ) are obtained from a 6<sup>th</sup>–8<sup>th</sup> degree Dunham fit of the potential energy curves.<sup>39</sup>

For the triatomics CO<sub>2</sub>, H<sub>2</sub>O, and O<sub>3</sub> grids of all points required to generate the nonvanishing quadratic, cubic, and quartic force constants in symmetry-adapted internal coordinates are calculated using the above-mentioned W4 (and related) theories.

The convergence criteria throughout are tightened such that the SCF and coupled-cluster energies are converged to at least 10<sup>-12</sup> a.u.

### C. Response properties

The dipole moments ( $\mu$ ), static polarizabilities ( $\alpha$ ), and static hyperpolarizabilities ( $\beta$ ) are calculated at the experimental structures via numerical differentiation of energies with respect to a uniform electric field. A field strength of  $h=0.0025$  a.u. was chosen as the basic grid size: the values reported were obtained by means of Richardson extrapolation<sup>40</sup> from successive multiples of that grid size. The first-order ( $\mu_z$ ) and third-order ( $\beta_{zzz}$ ) partial derivatives are obtained by means of a four point formula using static field strengths of  $\pm h$  and  $\pm 2h$ . The second-order derivatives ( $\alpha_{xx}$ ,  $\alpha_{yy}$ , and  $\alpha_{zz}$ ) are obtained through a five point formula using field strengths of 0,  $\pm h$ , and  $\pm 2h$ . The mixed third-order derivatives ( $\beta_{jjz}$ ,  $j = x, y$ ) are computed using a six point formula with field strengths of  $\pm 2h\hat{\mathbf{z}}$ ,  $\pm 2h(\hat{\mathbf{z}} + \hat{\mathbf{j}})$ , and  $\pm 2h(\hat{\mathbf{z}} - \hat{\mathbf{j}})$ .

## III. RESULTS AND DISCUSSION

### A. Multireference considerations

In reference<sup>17</sup> we proposed a number of energy-based diagnostics for the importance of nondynamical correlation that are specifically designed for thermochemical purposes. In particular, the %TAE<sub>e</sub>[(T)] diagnostic, *i.e.* the percentage of the total atomization energy accounted for by parenthetical connected triple excitations, proved to be a very efficient and cost-effective *a priori* diagnostic:

$$\%TAE_e[(T)] = 100 \times \frac{TAE_e[CCSD(T)] - TAE_e[CCSD]}{TAE_e[CCSD(T)]} \quad (1)$$

where TAE<sub>e</sub>[CCSD] and TAE<sub>e</sub>[CCSD(T)] represent, respectively, the non-relativistic, clamped-nuclei CCSD and CCSD(T) total atomization energies at the bottom of the well. Values of %TAE<sub>e</sub>[(T)] are interpreted as follows:<sup>17</sup> below 2% indicates systems dominated by

dynamical correlation; 2–5% mild nondynamical correlation; 5–10% moderate nondynamical correlation; and in excess of 10% severe nondynamical correlation.

Table S-I of the Supporting Information<sup>46</sup> provides the percentage of the non-relativistic, clamped-nuclei total atomization energy at the bottom of the well accounted for by SCF, (T) triples, post-CCSD(T), and  $\hat{T}_4 + \hat{T}_5$  contributions. Table S-I also lists the coupled cluster  $\mathcal{T}_1$  diagnostic,<sup>41</sup>  $\mathcal{D}_1$  diagnostic,<sup>42</sup> and the largest CCSD  $T_2$  amplitudes. However, the use of these diagnostics as a measure of multireference effects has been criticized by us<sup>17,21</sup> and by others.<sup>43–45</sup> In the context of the present work, for example, a  $\mathcal{T}_1$  diagnostic of 0.03 is obtained for CS and SiO as well as for O<sub>3</sub>, although the latter is a notoriously multireference system. The  $\mathcal{D}_1$  diagnostic of O<sub>2</sub> (0.01) is the same as, or even smaller than, that of the hydride systems, despite the former having almost an order of magnitude greater %TAE<sub>e</sub>[ $\hat{T}_4 + \hat{T}_5$ ]. The largest  $T_2$  cluster amplitudes are deceptively high for BH (0.11) and deceptively low for BeO (0.04). In practice, the %TAE[(T)] indicator proves to be a very useful predictor for the importance of post-CCSD(T) correlation effects (see relevant discussions in Refs.<sup>17,21</sup>).

As expected, the hydride systems considered in the present study exhibit very mild nondynamical correlation effects and can be regarded as dominated by dynamical correlation: 60–80% of the atomization energy is accounted for at the SCF level, and only 0–2% by the (T) triples. The nonhydrides BF, BCl, AlF, AlCl, CO, SiO, SiS, and N<sub>2</sub> are characterized by mild nondynamical correlation: at least 50% of the TAE<sub>e</sub> is accounted for at the SCF level, and  $2 \leq \%TAE_e[(T)] \leq 5$ . The homonuclear diatomics (O<sub>2</sub>, Si<sub>2</sub>, P<sub>2</sub>, S<sub>2</sub>, and Cl<sub>2</sub>), as well as CS, PN, SO, ClF, and BeO are characterized by moderate nondynamical correlation effects,  $20 \leq \%TAE_e[SCF] \leq 60$  and  $6 \leq \%TAE_e[(T)] \leq 8$ . Finally, C<sub>2</sub>, BN, and O<sub>3</sub> are pathologically multireference: at the SCF level BN and O<sub>3</sub> are unbound, and only 10% of the atomization energy of C<sub>2</sub> is accounted for, while the (T) triples account for as much as 13–19% of the binding energies of these systems.

## B. Equilibrium geometries ( $r_e$ ) of diatomic molecules

Table I compares the equilibrium bond distances obtained at the W4 (and related) levels with experimental bond lengths for our set of 10 monohydride and 18 nonhydride diatomic molecules. The equilibrium bond lengths determined at the clamped-nuclei, frozen-core, non-relativistic CCSD(T) basis set limit (W4[up to CCSD(T)]<sub>fc,nr</sub> method)



almost universally overestimate the experimental bond distances (by  $\sim 0.3$  pm on average). Inclusion of the core-valence and scalar relativistic corrections consistently reduces  $r_e$  by amounts ranging from  $\sim 0.1$  pm (for HF, OH, F<sub>2</sub>, and NH) to  $\sim 1.0$  pm (for AlCl and Si<sub>2</sub>). Consequently, at the clamped-nuclei, all-electron, relativistic CCSD(T) basis set limit (W4[up to CCSD(T)] method)  $r_e$  are universally underestimated (by  $\sim 0.2$  pm on average). The largest underestimations are seen for the nonhydrides, *e.g.* by 0.26, 0.28, 0.31, 0.32, 0.33, 0.35, 0.38, 0.39, and 0.52 pm, for SiS, PN, ClF, Cl<sub>2</sub>, O<sub>2</sub>, S<sub>2</sub>, F<sub>2</sub>, P<sub>2</sub>, and Si<sub>2</sub>, respectively. For the subset of 18 nonhydride systems the W4[up to CCSD(T)] method attains an RMSD of 0.27 pm. As expected, the performance for the subset of monohydride systems, which are mostly dominated by dynamical correlation ( $0.1 \leq \%TAE[(T)] \leq 1.6$ ), is much better with an RMSD of only 0.06 pm. For the entire set of 28 diatomics W4[up to CCSD(T)] theory attains an RMSD of 0.22 pm.

Higher-order triple excitations,  $\hat{T}_3-(T)$ , have a relatively small effect on the bond lengths:  $|r_e|$  changes by  $\leq 0.1$  pm (with the exception of P<sub>2</sub> and PN for which  $r_e$  decreases by 0.2 pm). For the hydride systems the  $\hat{T}_3-(T)$  contribution consistently increases the bond lengths by 0.01–0.06 pm thereby improving agreement with experiment. For the nonhydride systems the  $\hat{T}_3-(T)$  excitations have a mixed effect, bettering agreement with experiment for some while worsening it for others. For the entire set of 28 diatomics, the effect of including the  $\hat{T}_3-(T)$  contribution is manifested by a slight increase in the RMSD from 0.22 to 0.25 pm. Note that, with very few exceptions (namely BH, CH, and AlH), the W4[up to CCSDT] method still universally underestimates the experimental bond lengths.

Connected quadruple excitations,  $\hat{T}_4$ , universally increase the bond distances, thereby substantially improving the agreement with experiment, particularly for the nonhydride systems. Upon inclusion of the  $\hat{T}_4$  excitations the RMSD over the nonhydrides is reduced from 0.30 to 0.06 pm (!), where in most cases the bond lengths increase by  $> 0.2$  pm (*e.g.* by 0.28, 0.29, 0.30, 0.33, 0.37, 0.42, and 0.46 for SO, S<sub>2</sub>, Si<sub>2</sub>, O<sub>2</sub>, PN, F<sub>2</sub>, and P<sub>2</sub>, respectively). For the monohydrides, the effect of the  $\hat{T}_4$  excitations is rather modest, the bond lengths increasing by only 0.00–0.03 pm and the RMSD being reduced from 0.04 to 0.03 pm.

As expected, connected quintuple excitations have essentially no effect on the internuclear distances of the hydride systems. For the nonhydride systems the bond lengths generally increase by 0.00–0.05 pm upon inclusion of the  $\hat{T}_5$  excitations.

At the full W4 level the RMSD from experiment are {0.03, 0.05, 0.04} pm for the hydrides,

nonhydrides, and the entire set, respectively. The largest deviations (underestimations) are 0.07 pm (for P<sub>2</sub>, SiH, SiS, and ClF) and 0.09 pm (for Cl<sub>2</sub>).

The trends discussed above can be qualitatively understood by looking at how the various energy contributions ( $E^{\text{core-valence}}$ ,  $E^{\hat{T}_3-(\text{T})}$ ,  $E^{\hat{T}_4}$ , and  $E^{\hat{T}_5}$ ) change as a function of  $r$ . Tables SII–SV of the Supporting Information<sup>46</sup> give these energy contributions as a function of  $r$ . The said energy contributions (with the exception of  $E^{\hat{T}_3-(\text{T})}$  for BF and AlF; *vide infra*) show a nearly linear dependence on  $r$  in the interval  $|r - r_e| \leq 10$  pm. Addition of a linear term ( $E \approx ar$ ) to a simple harmonic potential results in an increase in  $r_e$  if  $a < 0$  and a decrease in  $r_e$  if  $a > 0$ . For all the systems, the connected quadruples energy contribution,  $E^{\hat{T}_4}$ , decreases with  $r$  (in a nearly linear manner:  $R^2 \geq 0.99$ ) and thereby inclusion of the  $\hat{T}_4$  excitations universally lengthens the bond distances. The core-valence energy contribution,  $E^{\text{core-valence}}$ , systematically increases with  $r$  (in a fairly linear fashion,  $R^2 \geq 0.98$ ), thus universally shortening the bond lengths. For the hydride systems  $E^{\hat{T}_3-(\text{T})}$  decreases linearly with  $r$  ( $R^2 \geq 0.98$ ), therefore lengthening the equilibrium bond lengths. For the nonhydrides, the  $E^{\hat{T}_3-(\text{T})}$  energy contribution does not behave in a consistent manner: for some (namely Si<sub>2</sub>, Cl<sub>2</sub>, BCl, ClF, AlCl, and SiS) it decreases with  $r$  therefore increasing  $r_e$ , while for others it increases with  $r$ , thus decreasing  $r_e$ . Interestingly, for the highly polar BF and AlF systems  $E^{\hat{T}_3-(\text{T})}$  varies quadratically with  $r$  in the scanned interval ( $R^2 \geq 0.998$ , see Table S-II of the Supporting Information<sup>46</sup>) and  $r_e$  is not affected by the  $\hat{T}_3-(\text{T})$  correction. For most of the systems  $E^{\hat{T}_5}$  decreases with  $r$  (exceptions are: BF, BCl, AlF, AlCl, CS, SiS, and SiO), therefore inclusion of  $\hat{T}_5$  excitations increases  $r_e$  for all but the said exceptions. In the next section we shall see that the said energy contributions that increase  $r_e$  decrease  $\omega_e$  and vice versa — as expected since  $\omega \propto \frac{1}{r}$ .

For a few systems we were able to obtain potential curves at the W4.2 and W4.3 levels (see Table I). Adding the  $\hat{T}_3-(\text{T})$  correction to the core-valence contribution in W4.2 theory has little effect on the bond distances, namely  $r_e$  is increased by 0.02 and 0.03 pm for F<sub>2</sub> and P<sub>2</sub>, respectively. Likewise, computing all the valence post-CCSD(T) corrections with larger basis sets and adding a  $\hat{T}_6$  correction in W4.3 theory increases the bond length of F<sub>2</sub> by merely 0.03 pm.

Finally, DBOC corrections universally increase the bond distances and, as expected, only have a perceptible effect on the hydride systems, something more pronounced for the lighter hydrides within each row. For example, the DBOC correction increases  $r_e$  by 0.06, 0.04,

0.03, 0.02, and 0.01 pm for the first-row hydrides BH, CH, NH, OH, and HF, respectively, and by 0.05, 0.04, 0.03, 0.02, and 0.01 pm for the second-row hydrides AlH, SiH, PH, SH, and HCl, respectively.

### C. Harmonic frequencies ( $\omega_e$ ) of diatomic molecules

The harmonic frequencies at the W4 (and related) levels are given in Table II together with the experimental values. At the clamped-nuclei, all-electron, relativistic CCSD(T) basis set limit (W4[up to CCSD(T)] method) the harmonic frequencies are universally overestimated by amounts ranging from  $\sim 1$   $\text{cm}^{-1}$  (for AlF and AlCl) up to  $29$   $\text{cm}^{-1}$  (for  $\text{O}_2$ ). Not surprisingly, the smallest errors (of  $1$ – $8$   $\text{cm}^{-1}$ ) are observed for the systems for which  $\% \text{TAE}_e[(\text{T})] < 4$  (see section III A). The RMSD of the W4[up to CCSD(T)] method are  $\{5.1, 11.2, 9.8\}$   $\text{cm}^{-1}$  for the hydrides, nonhydrides, and the entire set, respectively.

For most systems W4[up to CCSD(T) $_{fc,nr}$ ] and W4[up to CCSD(T)] err on the same side of experiment, the exceptions are AlF, CO, BCl, CH, BF, and BH for which W4[up to CCSD(T) $_{fc,nr}$ ] underestimates experiment by  $2.5$ – $8.6$   $\text{cm}^{-1}$ . We note that with few exceptions (namely BH, AlH, BF, AlF, CH, and SiH) W4[up to CCSD(T) $_{fc,nr}$ ] is closer to experiment than W4[up to CCSD(T)], as manifested in an RMSD for the entire set smaller by  $2$   $\text{cm}^{-1}$ .

Higher-order triple excitations,  $\hat{T}_3$ –(T), consistently reduce the harmonic frequencies of the monohydride compounds by  $1.4$ – $5.0$   $\text{cm}^{-1}$ , thus cutting the RMSD for the monohydrides by almost 50% from  $5.1$   $\text{cm}^{-1}$  (W4[up to CCSD(T)]) to  $2.8$   $\text{cm}^{-1}$  (W4[up to CCSDT]). For the nonhydrides, the higher-order triple excitations have a mixed effect, reducing  $\omega_e$  for some systems (e.g., by  $\sim 1$   $\text{cm}^{-1}$  for  $\text{Si}_2$ , BCl, and  $\text{Cl}_2$ ) and increasing  $\omega_e$  for others (most notably by  $\sim 10$   $\text{cm}^{-1}$  for  $\text{N}_2$  and PN). In effect, the RMSD for the nonhydrides increases upon inclusion of the  $\hat{T}_3$ –(T) correction (from  $11.2$   $\text{cm}^{-1}$  for W4[up to CCSD(T)] to  $15.2$   $\text{cm}^{-1}$  for W4[up to CCSDT]).

Connected quadruple excitations,  $\hat{T}_4$ , universally reduce  $\omega_e$  by amounts ranging from  $0.0$ – $5.0$   $\text{cm}^{-1}$  for the monohydride systems and  $0.3$ – $31.5$   $\text{cm}^{-1}$  for the nonhydride systems. The largest decreases, of over  $10$   $\text{cm}^{-1}$ , are seen for  $\text{P}_2$ , SO,  $\text{F}_2$ ,  $\text{N}_2$ , PN, and  $\text{O}_2$  (specifically by  $13.4$ ,  $16.1$ ,  $16.7$ ,  $20.4$ ,  $21.8$ , and  $31.5$   $\text{cm}^{-1}$ , respectively). Inclusion of the  $\hat{T}_4$  excitations reduces the RMSD over the set of hydrides from  $2.8$   $\text{cm}^{-1}$  (W4[up to CCSDT]) to  $0.6$

$\text{cm}^{-1}$  (W4[up to CCSDTQ]), and over the set of nonhydrides from  $15.2 \text{ cm}^{-1}$  (W4[up to CCSDT]) to  $2.4 \text{ cm}^{-1}$  (W4[up to CCSDTQ]). Over the entire set of 28 diatomics the RMSD is reduced from  $12.7 \text{ cm}^{-1}$  (W4[up to CCSDT]) to  $2.0 \text{ cm}^{-1}$  (W4[up to CCSDTQ]), an improvement by nearly one order of magnitude (!). This clearly demonstrates that connected quadruple excitations are essential for obtaining the harmonic frequencies with near-spectroscopic accuracy.

As expected, connected quintuple excitations,  $\hat{T}_5$ , have little or no effect for the hydrides: the biggest changes are seen for OH and HF for which  $\omega_e$  is reduced by merely  $0.1 \text{ cm}^{-1}$ . For the nonhydrides  $\hat{T}_5$  excitations change  $\omega_e$  by up to  $4 \text{ cm}^{-1}$ . The biggest changes are observed for  $\text{N}_2$ , PN, and  $\text{O}_2$ , namely,  $\omega_e$  is reduced by 3.3, 3.5, and  $4.0 \text{ cm}^{-1}$ , respectively.

Upon inclusion of the  $\hat{T}_5$  excitations the RMSDs for the nonhydrides (and for the entire set) are reduced by 50%, thus at the W4 level RMSDs of  $\{0.6, 1.2, 1.0\} \text{ cm}^{-1}$  are obtained for the hydrides, nonhydrides, and the entire set, respectively. The largest deviation being  $2.1 \text{ cm}^{-1}$  for  $\text{F}_2$  and  $\text{P}_2$ . We note that these errors drop to  $0.9$  and  $1.3 \text{ cm}^{-1}$ , respectively, upon adding the  $\hat{T}_3$ -(T) correction to the core-valence contribution in W4.2 theory. The error for  $\text{F}_2$  is further reduced to  $0.5 \text{ cm}^{-1}$  by upgrading the valence post-CCSD(T) corrections and adding a  $\hat{T}_6$  correction in W4.3 theory.

Finally, a word is due on the effect of the DBOC correction on the harmonic frequencies. For the monohydride systems the DBOC reduces  $\omega_e$  across the board, again more appreciably so for the lighter hydrides within each row. Specifically, the DBOC reduces  $\omega_e$  by 2.1, 1.8, 1.4, 0.9, and  $0.3 \text{ cm}^{-1}$  for the first-row hydrides BH, CH, NH, OH, and HF, respectively, and by 1.1, 1.0, 0.9, 0.7, and  $0.5 \text{ cm}^{-1}$  for the second-row hydrides AlH, SiH, PH, SH, and HCl, respectively. For the nonhydrides the DBOC negligibly increases the harmonic frequencies by  $<0.01 \text{ cm}^{-1}$ , the biggest change of  $+0.08 \text{ cm}^{-1}$  being found for  $\text{O}_2$ .

#### D. First anharmonicity constant ( $\omega_e x_e$ ) of diatomic molecules

Table III lists the theoretical and experimental first anharmonicity constants. Monohydrides are well known to exhibit relatively strong anharmonicity effects:  $\omega_e x_e$  ranges between  $30$ – $90 \text{ cm}^{-1}$  for the subset of monohydrides, and between  $2$ – $14 \text{ cm}^{-1}$  for the subset of nonhydrides.

The W4[up to CCSD(T)] method almost universally underestimates  $\omega_e x_e$  by amounts

ranging from  $0.01 \text{ cm}^{-1}$  (BF) to  $1.0 \text{ cm}^{-1}$  (NH). At this level RMSDs of  $\{0.6, 0.3, 0.5\} \text{ cm}^{-1}$  are obtained for the hydrides, nonhydrides, and the entire set, respectively. We note that the W4[up to CCSD(T) $_{fc,nr}$ ] method shows very similar performance for most of the systems, the core-valence correction tends to slightly increase  $\omega_e x_e$  but overall offers little improvement.

Inclusion of higher-order triple excitations improves agreement with experiment for the monohydride systems, and slightly worsens agreement with experiment for the nonhydride systems. For the monohydride systems  $\hat{T}_3$ -(T) excitations increase  $\omega_e x_e$  by up to  $0.5 \text{ cm}^{-1}$  (for NH), whereas for the nonhydrides they tend to decrease  $\omega_e x_e$  by up to  $0.3 \text{ cm}^{-1}$  (for  $\text{N}_2$  and  $\text{F}_2$ ). The RMSD of the W4[up to CCSDT] method is  $0.4 \text{ cm}^{-1}$  for both the hydride and nonhydride subsets.

Inclusion of connected quadruple excitations in the W4[up to CCSDTQ] method universally increases  $\omega_e x_e$  by up to  $1.0 \text{ cm}^{-1}$ . Thus, improving agreement with experiment across the board, something more pronounced for the nonhydride systems. The largest improvements are seen for  $\text{N}_2$  and  $\text{O}_2$ : the error for  $\text{N}_2$  is reduced from  $0.7 \text{ cm}^{-1}$  to  $0.2 \text{ cm}^{-1}$ , and that for  $\text{O}_2$  from  $1.1 \text{ cm}^{-1}$  to  $0.1 \text{ cm}^{-1}$ .

Contributions from connected quintuple excitations are generally negligible (for most of the systems  $<0.01 \text{ cm}^{-1}$ ) the largest changes being  $+0.06$ ,  $+0.11$ , and  $+0.15 \text{ cm}^{-1}$  for PN,  $\text{N}_2$ , and  $\text{O}_2$ , respectively.

For the nonhydride systems agreement between W4 and experiment is as good as one can hope for with an RMSD of  $0.04 \text{ cm}^{-1}$ . For the hydride systems the RMSD is almost one order of magnitude larger ( $0.32 \text{ cm}^{-1}$ ) but still well within the goal of “spectroscopic accuracy”.

W4.2 and W4.3 theories offer no significant improvement over W4 theory. Likewise, DBOC corrections are negligible (the largest correction of  $-0.02 \text{ cm}^{-1}$  is obtained for BH).

### E. Second anharmonicity constant ( $\omega_e y_e$ ) of diatomic molecules

Table IV reports the theoretical and experimental second anharmonicity constants. The second-order anharmonicity constants are typically 2–3 orders of magnitude smaller than the first-order anharmonicity constants. For the monohydrides  $\omega_e y_e$  ranges from  $0.1 \text{ cm}^{-1}$  (for SiH, PH, and SH) to  $0.9 \text{ cm}^{-1}$  (for HF). For the nonhydrides, which exhibit weaker

anharmonicity effects,  $\omega_e y_e$  are typically  $< 0.01 \text{ cm}^{-1}$ , and are given in Table S-VI of the Supporting Information.

The W4[up to CCSD(T) $_{fc,nr}$ ] method attains near-zero RMSD of 0.003, 0.09, and  $0.06 \text{ cm}^{-1}$  for the nonhydrides, hydrides, and the entire set, respectively. For both subsets core-valence, post-CCSD(T), and DBOC corrections have very little effect on the second anharmonicity constants.

Figure 1 shows the normal distribution functions for the errors between the theoretical and experimental  $r_e$ ,  $\omega_e$ ,  $\omega_e x_e$ , and  $\omega_e y_e$ . It illustrates that  $r_e$  and  $\omega_e$  are much more sensitive to post-CCSD(T) correlation effects, in particular to the  $\hat{T}_4$  excitations, than the anharmonic corrections. This was suggested long ago, based on CEPA<sup>47</sup> calculations on the methane surface by Pulay and coworkers.<sup>48</sup> The Gaussians for the W4[up to CCSDT] and W4[up to CCSDTQ] methods are centered, respectively, around  $-0.17$  and  $-0.04 \text{ pm}$  for  $r_e$ , around  $8.1$  and  $1.0 \text{ cm}^{-1}$  for  $\omega_e$ , around  $-0.3$  and  $-0.1 \text{ cm}^{-1}$  for  $\omega_e x_e$ , and around  $-0.03$  and  $-0.04 \text{ cm}^{-1}$  for  $\omega_e y_e$ .

#### F. Vibration-rotation coupling constant ( $\alpha_e$ ) of diatomic molecules

Table V depicts the theoretical and experimental vibration-rotation coupling constants. The RMSD of the W4[up to CCSD(T) $_{fc,nr}$ ], W4[up to CCSD(T)], W4[up to CCSDT], W4[up to CCSDTQ], and W4 methods are 0.0034, 0.0024, 0.0017, 0.0012, and  $0.0012 \text{ cm}^{-1}$ , respectively. Hence it can be concluded that, while the  $\alpha_e$  are reasonably well converged at the clamped-nuclei, frozen-core, non-relativistic CCSD(T) basis set limit (W4[CCSD(T) $_{fc,nr}$ ] method), significant improvements can still be obtained beyond it.

#### G. Difficult 8-valence electron systems: C<sub>2</sub>, BN, and BeO

The low-lying  $^1\Sigma^+$  electronic states of the twelve-electron isoelectronic diatomics C<sub>2</sub>, BN, and BeO are known to have significant contributions from singly and doubly excited determinants, and thus represent an extremely challenging test for single reference electron correlation methods.<sup>43</sup> Table VI lists the spectroscopic constants at the W4 (and related) levels together with the experimental values. The harmonic frequencies are severely overestimated at the W4[up to CCSD(T)] level, specifically by 17, 50, and  $10 \text{ cm}^{-1}$  for

C<sub>2</sub>, BN, and BeO, respectively. Higher-order triples,  $\hat{T}_3$ -(T), contributions increase the errors for C<sub>2</sub> and BeO to about 20 cm<sup>-1</sup>, but reduce the error for BN by 36 cm<sup>-1</sup> (!) to 14 cm<sup>-1</sup>. (Interestingly, the same decrease of 36 cm<sup>-1</sup> was computed nearly two decades ago by one of us using a double zeta quality basis set.<sup>8</sup>) As expected, connected quadruple excitations reduce  $\omega_e$  by substantial amounts, namely 15.5, 6.6, and 12.6 cm<sup>-1</sup> for C<sub>2</sub>, BN, and BeO, respectively. Note however, that for BN the  $\hat{T}_4$  excitations have a lesser effect on  $\omega_e$  than the  $\hat{T}_3$ -(T) excitations.  $\hat{T}_5$  excitations further reduce the errors for C<sub>2</sub> and BN by 1.5 and 2.0 cm<sup>-1</sup>, respectively, but increase the error for BeO by 7.2 cm<sup>-1</sup>. Thus, at the W4 level the harmonic frequencies of C<sub>2</sub>, BN, and BeO are overestimated by 2.6, 5.7, and 14.3 cm<sup>-1</sup>, respectively.

We suspected that the large discrepancies seen for the ionic BN and BeO molecules are partly due to the difficulty of the unaugmented basis sets to describe the charge distribution in the valence post-CCSD(T) steps. Adding diffuse functions on the electronegative atoms (N and O) in all the valence post-CCSD(T) steps in W4 theory (see Computational Methods section and footnote *b* in Table VI) reduces the errors for BN and BeO to 2.1 and 6.1 cm<sup>-1</sup>, respectively. In addition, because of the fairly small core-valence gap in BeO, we also considered extrapolating the CCSD(T) inner-shell correlation contribution from larger basis sets (see Computational Methods section and footnote *c* in Table VI), which results in further reduction of the error for BeO to 1.2 cm<sup>-1</sup>. (The reduction to 1.9 cm<sup>-1</sup> for BN is not significant.)

As mentioned in section III B, the W4[up to CCSD(T)] method consistently underestimates the experimental bond lengths of the diatomic molecules in Table I (by  $\sim 0.2$  pm on average). Similarly, the bond distances of C<sub>2</sub> and BeO are underestimated by 0.2 pm, but that of BN is underestimated by as much as 1.0 pm. Higher-order triples,  $\hat{T}_3$ -(T), have little effect on  $r_e$  of C<sub>2</sub> and BeO, but increase  $r_e$  of BN by as much as 0.7 pm (!) narrowing the gap between theory and experiment to 0.3 pm.  $\hat{T}_4$  contributions further reduce the discrepancy between theory and experiment: W4[up to CCSDTQ] underestimates the experimental bond lengths of C<sub>2</sub>, BN, and BeO by 0.05, 0.12, 0.11 pm, respectively.  $\hat{T}_5$  excitations increase  $r_e$  of C<sub>2</sub> and BN by 0.03 and 0.05 pm, but decrease  $r_e$  of BeO by 0.11 pm. Thus, at the W4 level the C<sub>2</sub> and BN bond lengths are in relatively good agreement with experiment (underestimating by 0.03 and 0.07 pm, respectively), but  $r_e$  of BeO is underestimated by as much as 0.21 pm. Again, augmenting the basis sets on N and O in all

the valence post-CCSD(T) steps in W4 theory reduces the discrepancy for BeO to 0.08 pm, and for BN to 0.03 pm. Extrapolating the inner-shell correlation from RCCSD(T)/aug-cc-pwCVQZ and RCCSD(T)/aug-cc-pwCV5Z calculations results in further reduction of the error for BN and BeO to merely 0.02 pm.

Turning our attention to the first anharmonicity constants, for  $C_2$  errors of 1.1 and 0.1  $\text{cm}^{-1}$  are obtained at the W4[up to CCSD(T)] and W4 levels, where the biggest improvement comes from the  $\hat{T}_4$  excitations (see Table VI). For BN errors of 5.5 and 1.2  $\text{cm}^{-1}$  are obtained at the W4[up to CCSD(T)] and W4 levels, but here the improvement essentially stems from the  $\hat{T}_3$ -(T) contribution. For BeO errors of 0.3 and 0.5  $\text{cm}^{-1}$  are obtained at the W4[up to CCSD(T)] and W4 levels. We note that by adding diffuse functions on N and O in all the post-CCSD(T) steps in W4 theory the errors for BN and BeO are further reduced to 0.85 and 0.05  $\text{cm}^{-1}$ , respectively.

## H. Triatomic systems

The W4 equilibrium geometries, harmonic frequencies, and fundamental frequencies of  $H_2O$ ,  $CO_2$ , and  $O_3$  are given in Table VII together with the available experimental values. The W4 equilibrium geometries are in close agreement with the experimental ones. The equilibrium bond distances of  $CO_2$  and  $O_3$  are both 0.02 pm lower than experiment, and the bond angle of  $O_3$  is 0.07° higher than that of Barbe *et al.*<sup>49</sup> For  $H_2O$  the W4 bond distance (95.76 pm) is bracketed in between the experimental value of Benedict *et al.*<sup>50</sup> (95.72 pm) and the more recent value of Jensen<sup>51</sup> (95.84 pm), and the W4 bond angle (104.50°) is in between the said experimental values (104.52°<sup>50</sup> and 104.44°<sup>51</sup>, respectively). The Born–Oppenheimer correction is expected to be negligible for  $CO_2$  and  $O_3$ , and is on the order of 0.003 pm and 0.02° for the bond distance and bond angle of water, respectively.<sup>52</sup> We note that our W4  $r_e$  and  $\theta_e$  for water differ by 0.02 pm and 0.02°, respectively, from the *ab initio* Born–Oppenheimer geometry reported in Ref.<sup>52</sup>.

The W4 harmonic frequencies for  $CO_2$  and  $H_2O$  are in relatively good agreement with experiment. In particular, the symmetric, bending, and antisymmetric vibrational frequencies of  $CO_2$  underestimate the experimental values by 1.9, 2.4, and 0.1  $\text{cm}^{-1}$ , respectively, and those of  $H_2O$  overestimate experiment<sup>51</sup> by 2.8, 0.5, and 3.3  $\text{cm}^{-1}$ , respectively.



Ozone represents an extremely challenging system for both single reference and multireference methods (see, *e.g.*<sup>53-55</sup> and references therein). The theoretical symmetric stretch converges smoothly with the level of theory, *i.e.* it overestimates the experimental value of Barbe *et al.*<sup>49</sup> by 20.3, 3.7, and 0.6  $\text{cm}^{-1}$ , respectively, at the W4[up to CCSDT], W4[up to CCSDTQ], and W4 levels. The harmonic bending frequency converges more rapidly, in particular it overestimates the experimental one by 2.9, 0.5, and 0.2  $\text{cm}^{-1}$  at the same levels, respectively. However, the asymmetric stretch ( $\omega_3$ ), which exhibits severe nondynamical correlation character<sup>56-58</sup>, behaves in a less systematic manner. It underestimates the experimental value by 1.6, 17.1, and 20.2  $\text{cm}^{-1}$ , respectively, at the said sequence of levels (where the good agreement between the W4[up to CCSDT] value and experiment is clearly fortuitous). It is likely that the counterintuitive deterioration in agreement with experiment upon inclusion of the connected quadruple excitations is due to basis set deficiencies. The hypersensitivity of  $\omega_3$  to the treatment of connected quadruple excitations has been noted in Ref.<sup>58</sup> Determination of  $\omega_3$  requires evaluation of the energy at only one point (in  $C_s$  symmetry) aside from the equilibrium  $C_{2v}$  structure. We were able to obtain the CCSDTQ/cc-pVTZ(no  $f$   $1d$ ) energies at these two points — cc-pVTZ(no  $f$   $1d$ ) denotes the  $sp$  part of the cc-pVTZ basis set combined with the  $d$  function from the cc-pVDZ basis set. Replacing the  $\hat{T}_4$  term in W4 theory with  $\hat{T}_4/\text{cc-pVTZ}(\text{no } f \text{ } 1d)$  results in an asymmetric stretch of 1078.1  $\text{cm}^{-1}$ , *i.e.* the error from experiment is reduced by more than 50%. Additionally upgrading the basis set for the parenthetical connected quadruple excitations to cc-pVTZ (*i.e.*,  $\hat{T}_4=\text{CCSDT}(\text{Q})/\text{cc-pVTZ}-\text{CCSDT}/\text{cc-pVTZ}+\text{CCSDTQ}/\text{cc-pVTZ}(\text{no } f \text{ } 1d)-\text{CCSDT}(\text{Q})/\text{cc-pVTZ}(\text{no } f \text{ } 1d)$ ) results in a minor lowering of the asymmetric stretch by  $\sim 1$   $\text{cm}^{-1}$ , and a further improvement in the basis set to the cc-pVQZ(no  $g$   $1f$ ) basis set has a similar effect. Unfortunately, carrying out the fully iterative CCSDTQ/cc-pVTZ calculations is beyond the capabilities of our current computational resources, these calculations involving  $5 \times 10^9$  and  $10.5 \times 10^9$  amplitudes in  $C_{2v}$  and  $C_s$  symmetries, respectively.

It is appropriate to note, at this point, that — in spite of the difficulties the potential energy surface of ozone presents for W4 theory — W4 performs very well for the thermochemistry of  $\text{O}_3$ . In particular, the post-CCSD(T) components are practically converged at the W4 level, as evident from very small differences between W4 and W4.3 theories. The post-CCSD(T) contributions to the total atomization energy of ozone in W4

and W4.3 theories are:  $\hat{T}_3-(T) = -1.34$  and  $-1.37$ ;  $\hat{T}_4 = 3.81$  and  $3.84$ ; and  $\hat{T}_5 = 0.41$  and  $0.43$ .<sup>59</sup> In addition, W4.3 includes a  $\hat{T}_6$  contribution of  $0.05$  kcal/mol. This suggests that while basis set convergence is faster for second-order properties (*e.g.* frequencies) than for the energy, n-particle treatment convergence is arguably slower.

## I. Electrical properties

Fairly accurate experimental dipole moments are available for some of the molecules considered in the present work; however, for the higher-order properties ( $\alpha$  and  $\beta$ ) accurate experimental data are much more scarce. Furthermore, direct comparison between theory and experiment is hampered by the fact that the theoretical values correspond to the static limit, whereas the experimental values are generally frequency-dependent. Thus, we will not attempt to quantify the reliability of W4 theory against experimental data; our intent is rather to investigate the effects of core-valence and post-CCSD(T) contributions on the electrical properties.

Table VIII gives the theoretical and experimental dipole moments. Core-valence and post-CCSD(T) corrections have no noticeable effect on systems dominated by dynamical correlation effects, and relatively little effects on the strongly multireference systems. The dipole moments at the frozen-core, non-relativistic CCSD(T) basis set limit (W4[up to CCSD(T)<sub>fc,nr</sub>] method) agree very well with the full W4 values (RMSD=0.005 a.u., and the largest deviation of 0.01 a.u. is seen for BeO). At the W4[up to CCSD(T)] level, the RMSD from W4 is 0.002 a.u., while the largest error of 0.005 a.u. is found for O<sub>3</sub>. Adding the higher order triple excitations increases the errors for the nonhydrides: in particular, a dramatic increase is seen for BN (from -0.003 to 0.05 a.u.). We note that for BN, the  $\hat{T}_3-(T)$  excitations raise  $\mu$  by 0.05 a.u. and the  $\hat{T}_4$  excitations reduce it by 0.04 a.u. At the W4[up to CCSDTQ] level, the largest errors of -0.014 and +0.012 a.u. are seen for BeO and BN, respectively, and the RMSD from W4 is 0.06 a.u.

Table IX shows the theoretical polarizabilities together with the available experimental data. The RMSDs from the full W4 values are 2.6, 2.9, 0.5, and 0.2 a.u. for the W4[up to CCSD(T)<sub>fc,nr</sub>], W4[up to CCSD(T)], W4[up to CCSDT], and W4[up to CCSDTQ] methods, respectively. However, for most of the systems in Table IX, the polarizabilities at the W4[up to CCSD(T)<sub>fc,nr</sub>] level agree to within 0.1 a.u. with the W4 values. The main exceptions are

$\alpha_{zz}$  of BeO, and  $\alpha_{zz}$  and  $\alpha_{xx}$  of C<sub>2</sub>, which are overestimated by as much as 1.5, 5.4, and 13.0 a.u., respectively. Upon excluding C<sub>2</sub> and BeO the RMSD of the W4[up to CCSD(T)<sub>fc,nr</sub>] method drops to 0.13 a.u. The W4[up to CCSD(T)<sub>fc,nr</sub>] and W4[up to CCSD(T)] methods show very similar performance. Inclusion of  $\hat{T}_3$ -(T) excitations dramatically reduces the said errors for C<sub>2</sub> to 0.1 and 1.0 a.u., respectively, and the RMSD for the entire set drops to 0.5 a.u. Nevertheless, at this level errors of -1.0 and 2.1 a.u. are still seen for  $\alpha_{xx}$  and  $\alpha_{zz}$  of BeO, respectively. These are reduced to -0.3 and 1.0 a.u., respectively, by inclusion of the  $\hat{T}_4$  excitations in the W4[up to CCSDTQ] method.

Table X lists the theoretical W4 (and related) hyperpolarizabilities. Similarly to the polarizabilities, for most systems, the first hyperpolarizabilities at the W4[up to CCSD(T)<sub>fc,nr</sub>] level agree with the W4 values to within 0.1 a.u. The main exceptions are O<sub>3</sub> ( $\beta_{yyz}$ ) and CS ( $\beta_{zzz}$ ); excluding these, the W4[up to CCSD(T)<sub>fc,nr</sub>] method attains an RMSD of 0.3 a.u. from the W4 values. Inclusion of the core-valence and relativistic contributions (in the W4[up to CCSD(T)] method) reduces the said RMSD to 0.1 a.u., and inclusion of post-CCSD(T) excitations offers little improvement. Turning our attention to the more problematic systems, we note that at the W4[up to CCSD(T)<sub>fc,nr</sub>] level  $\beta_{zzz}$  of CS is  $\sim 3$  a.u. removed from the W4 value and that core-valence and post-CCSD(T) corrections do not significantly affect this result. The W4[up to CCSD(T)<sub>fc,nr</sub>] method gives the wrong sign for the  $\beta_{yyz}$  component of O<sub>3</sub>, *i.e.* it overestimates the W4 value by as much as  $\sim 9$  a.u. Inclusion of  $\hat{T}_3$ -(T) excitations reduces this error to  $\sim 2$  a.u. and further inclusion of  $\hat{T}_4$  excitations reduces the error to -0.2 a.u.

## J. Conclusions

For a chemically diverse set of 28 first- and second-row diatomic molecules for which very accurate experimental spectroscopic constants are available W4 theory is capable of spectroscopically accurate predictions. Specifically, the RMSD from experiment are 0.04 pm for  $r_e$ , 1.03 cm<sup>-1</sup> for  $\omega_e$ , 0.20 cm<sup>-1</sup> for  $\omega_e x_e$ , 0.10 cm<sup>-1</sup> for  $\omega_e y_e$ , and 0.001 cm<sup>-1</sup> for  $\alpha_e$ .

Core-valence and post-CCSD(T) contributions have little effect on  $\omega_e x_e$  and  $\alpha_e$ , and especially on  $\omega_e y_e$ . As for  $r_e$  and  $\omega_e$ , higher-order triple excitations,  $\hat{T}_3$ -(T), improve agreement with experiment for the monohydride systems, but their inclusion (in the absence of  $\hat{T}_4$ ) generally worsens agreement with experiment for the nonhydride diatomics. For

instance, upon inclusion of the  $\hat{T}_3$ -(T) corrections the RMSD in  $\omega_e$  reduces from 5.1 to 2.8  $\text{cm}^{-1}$  for the subset of monohydride systems, and increases from 11.2 to 15.2  $\text{cm}^{-1}$  for the subset of nonhydride systems.

Connected quadruple excitations,  $\hat{T}_4$ , are essential to obtain  $r_e$  and  $\omega_e$  with RMSD of  $\sim 0.05$  pm and  $\sim 2$   $\text{cm}^{-1}$ , respectively. For example, inclusion of the  $\hat{T}_4$  excitations reduces the RMSD in  $\omega_e$  from 2.8 to 0.6  $\text{cm}^{-1}$  for the subset of monohydrides, and from 15.2 to 2.4  $\text{cm}^{-1}$  for the subset of nonhydrides. The RMSD in  $r_e$  are reduced from 0.04 to 0.03 pm for the hydrides subset, and from 0.30 to 0.06 pm for the nonhydrides subset. Moreover, inclusion of  $\hat{T}_4$  excitations results in systematic trends: namely  $r_e$  universally increase (by up to 0.5 pm),  $\omega_e$  universally decrease (by up to 31.5  $\text{cm}^{-1}$ ), and  $\omega_e x_e$  universally increase (by up to 1.0  $\text{cm}^{-1}$ ).

Connected quintuple excitations,  $\hat{T}_5$ , are of minor importance for  $r_e$  and  $\omega_e x_e$ , but are spectroscopically significant for  $\omega_e$ . Inclusion of a CCSDTQ5/cc-pVDZ(no  $d$ ) correction term cuts the RMSD in the harmonic frequencies for our set of 28 diatomic molecules by 50% from 2.0 to 1.0  $\text{cm}^{-1}$ , practically all of the improvement being due to the nonhydride systems.

DBOC corrections result in systematic trends and are of importance mainly for  $r_e$  and  $\omega_e$  of the hydride systems, something more pronounced for the lighter hydrides within each row. For example, for our set of monohydrides DBOC universally increase  $r_e$  by 0.06–0.01 pm and universally decrease  $\omega_e$  by 2.1–0.3  $\text{cm}^{-1}$ .

For the pathologically multireference systems (BeO and BN) obtaining near-spectroscopic accuracy requires: (i) augmenting the correlation-consistent basis sets on N and O in the valence post-CCSD(T) steps in W4 theory, and (ii) extrapolating the CCSD(T) inner-shell contribution in W4 theory from the aug-cc-pwCV{Q,5}Z basis set pair.

The present study also considers water, carbon dioxide, and ozone. The experimental geometries and harmonic frequencies (with the exception of the asymmetric mode of ozone) are reproduced relatively accurately at the W4 level. However, the asymmetric stretch of ozone, which exhibits severe nondynamical character, is underestimated by as much as 20  $\text{cm}^{-1}$  at the W4 level. This deviation is reduced by more than 50% when the connected quadruples,  $\hat{T}_4$ , term in W4 theory is calculated with the cc-pVTZ(no  $f$  1d) basis set.

We also calculated the dipole moments ( $\mu$ ), polarizabilities ( $\alpha$ ), and first hyperpolarizabilities ( $\beta$ ). We find that: (i) excluding the pathologically multireference systems, the post-

CCSD(T) contributions are of little importance for these electrical properties; (ii) the core-valence and DBOC contributions are also of little importance in most cases; and (iii) that the sensitivity to post-CCSD(T) effects, particularly for systems with significant nondynamical correlation, increases with increasing order of derivative with respect to the static electric field, *i.e.*, in the order:  $\mu < \alpha < \beta$ .

### Acknowledgments

Research at Weizmann was funded by the Israel Science Foundation (grant 709/05), the Helen and Martin Kimmel Center for Molecular Design, and the Weizmann Alternative Energy Research Initiative (AERI). JMLM is the incumbent of the Baroness Thatcher Professorial Chair of Chemistry and a member *ad personam* of the Lise Meitner-Minerva Center for Computational Quantum Chemistry.

### Supporting Information

Diagnostics for the importance of nondynamical correlation can be found in Table S-I. Tables S-II through S-V give the  $E^{\text{core-valence}}$ ,  $E^{\hat{T}_3-(\text{T})}$ ,  $E^{\hat{T}_4}$ , and  $E^{\hat{T}_5}$  energy contributions as a function of  $r$ . Table S-VI gives theoretical and experimental second-order anharmonic corrections for the set of nonhydride diatomic molecules.

TABLE I: Equilibrium bond distances ( $r_e$ , pm) of diatomic molecules obtained from a Dunham analysis of potential energy curves computed at the W4 and related levels.

	W4[up to CCSD(T) $f_{c,nr}$ ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	W4.2	W4.3	$\Delta$ DBOC <sup>a</sup>	Expt.	Ref.
BH	123.25	122.90	122.91	122.91	122.91	122.91	122.92	0.060	122.95	<sup>60</sup>
CH	111.96	111.75	111.77	111.77	111.78			0.041	111.777	<sup>2</sup>
NH	103.68	103.55	103.58	103.60	103.60			0.026	103.655	<sup>2</sup>
OH	97.01	96.92	96.93	96.96	96.96			0.015	96.966	<sup>61</sup>
HF	91.73	91.67	91.66	91.69	91.69	91.69	91.69	0.005	91.6984	<sup>2</sup>
AlH	164.99	164.48	164.53	164.54	164.54			0.053	164.5362	<sup>62</sup>
SiH	152.18	151.81	151.85	151.86	151.86			0.039	151.966	<sup>63</sup>
PH	142.35	142.06	142.11	142.13	142.13			0.028	142.14	<sup>64</sup>
SH	134.21	133.96	133.99	134.01	134.01			0.020	134.0614	<sup>65</sup>
HCl	127.62	127.41	127.42	127.44	127.44	127.44	127.45	0.013	127.46149	<sup>66</sup>
N <sub>2</sub>	109.89	109.65	109.58	109.73	109.75	109.75	109.75	0.001	109.768 <sub>5</sub>	<sup>61</sup>
O <sub>2</sub>	120.61	120.42	120.33	120.66	120.70			0.000	120.752	<sup>61</sup>
F <sub>2</sub>	140.93	140.81	140.75	141.17	141.18	141.20	141.23	0.000	141.193	<sup>61</sup>
Si <sub>2</sub>	225.09	224.08	224.22	224.52	224.54			0.002	224.6	<sup>61</sup>
P <sub>2</sub>	189.76	188.95	188.77	189.23	189.27	189.30		0.001	189.34	<sup>61</sup>
S <sub>2</sub>	189.14	188.57	188.55	188.85	188.87			0.001	188.92	<sup>61</sup>
Cl <sub>2</sub>	198.87	198.47	198.55	198.70	198.70	198.71		0.001	198.7 <sub>9</sub>	<sup>61</sup>
BF	126.66	126.21	126.21	126.24	126.23			0.001	126.25 <sub>9</sub>	<sup>61</sup>
CO	113.02	112.75	112.72	112.80	112.80	112.80	112.81	0.001	112.82427	<sup>66</sup>
BCl	172.19	171.41	171.45	171.49	171.48			0.000	171.5283	<sup>67</sup>
CS	153.86	153.31	153.30	153.45	153.45	153.45		0.001	153.4941	<sup>61</sup>
SiO	151.42	150.82	150.78	150.95	150.93			0.001	150.9739	<sup>61</sup>
PN	149.35	148.81	148.63	149.00	149.05			0.001	149.0866	<sup>61</sup>
SO	148.23	147.87	147.77	148.05	148.08			0.000	148.1087	<sup>61</sup>
AlF	166.19	165.40	165.40	165.42	165.42			0.001	165.4369	<sup>61</sup>
ClF	162.74	162.52	162.55	162.75	162.76			0.001	162.8313	<sup>61</sup>
AlCl	213.87	212.91	212.94	212.96	212.96			0.001	213.0113	<sup>61</sup>
SIS	193.49	192.67	192.68	192.88	192.85			0.001	192.9264	<sup>68</sup>
error statistics for hydrides (10 systems) <sup>b</sup>										
MSD	0.21	-0.04	-0.01	0.00	0.00					
MAD	0.21	0.04	0.03	0.03	0.03					
RMSD	0.25	0.06	0.04	0.03	0.03					
error statistics for nonhydrides (18 systems) <sup>b</sup>										
MSD	0.30	-0.23	-0.26	-0.05	-0.04					
MAD	0.36	0.23	0.26	0.05	0.04					
RMSD	0.43	0.27	0.30	0.06	0.05					
error statistics for everything (28 systems) <sup>b</sup>										
MSD	0.27	-0.16	-0.17	-0.04	-0.03					
MAD	0.30	0.17	0.18	0.05	0.04					
RMSD	0.33	0.22	0.25	0.05	0.04					

<sup>a</sup>DBOC correction at the HF/AVTZ level. <sup>b</sup>Mean signed deviations (MSD), mean absolute deviation (MAD), and root mean squared deviation (RMSD) from experiment, note that the DBOC corrections (cfr. footnote *a*) are included in the error statistics.

TABLE II: Harmonic frequencies ( $\omega_e$ ,  $\text{cm}^{-1}$ ) of diatomic molecules obtained from a Dunham analysis of potential energy curves computed at the W4 and related levels.

	W4[up to CCSD(T) $f_{c,nr}$ ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	W4.2	W4.3	$\Delta\text{DBOC}^a$	Expt. <sup>b</sup>
BH	2360.24	2370.79	2369.13	2369.08	2369.08	2369.10	2368.69	-2.08	2366.7296
CH	2858.08	2865.99	2863.57	2862.53	2862.51			-1.83	2860.7508
NH	3286.50	3292.03	3287.33	3284.65	3284.62			-1.43	3282.7200
OH	3742.91	3746.58	3744.10	3739.55	3739.48			-0.90	3737.7610
HF	4141.64	4143.57	4143.71	4138.87	4138.77	4138.79	4138.51	-0.31	4138.3850
AlH	1688.45	1685.17	1682.54	1682.14	1682.14			-1.10	1682.3747
SiH	2048.29	2047.07	2043.63	2042.91	2042.90			-1.04	2042.5229
PH	2370.75	2370.79	2365.80	2364.41	2364.41			-0.91	2363.7740
SH	2700.78	2701.91	2698.64	2696.42	2696.40			-0.73	2696.2475
HCl	2994.17	2995.65	2994.23	2991.50	2991.48	2991.60	2990.94	-0.53	2990.9248
N <sub>2</sub>	2362.34	2371.91	2382.27	2361.83	2358.54	2357.89	2358.21	0.04	2358.5700
O <sub>2</sub>	1605.48	1609.22	1617.24	1585.76	1581.78			0.08	1580.1610
F <sub>2</sub>	931.17	932.34	936.12	919.44	918.96	917.82	917.44	0.02	916.9290
Si <sub>2</sub>	516.41	517.67	516.23	511.23	510.78			0.00	510.9800
P <sub>2</sub>	787.79	792.23	797.62	784.26	782.87	782.10		0.00	780.7700
S <sub>2</sub>	734.42	735.99	736.93	728.16	727.48			0.00	725.7102
Cl <sub>2</sub>	564.79	565.65	564.48	561.28	561.24	560.94		0.00	559.7510
BF	1394.96	1405.01	1404.45	1403.02	1403.25			0.04	1402.1587
CO	2166.60	2176.25	2178.86	2170.22	2170.37	2170.16	2170.35	0.05	2169.7559
BCl	836.91	843.27	841.90	840.74	840.93			0.03	840.2947
CS	1285.98	1292.50	1293.26	1284.95	1285.24	1285.45		0.02	1285.1546
SiO	1241.66	1248.53	1250.98	1241.56	1242.79			0.01	1241.5439
PN	1344.72	1352.54	1363.39	1341.62	1338.13			0.02	1336.9480
SO	1160.68	1163.56	1168.78	1152.64	1151.10			0.00	1150.7913
AlF	799.79	803.09	802.84	802.36	802.44			0.00	802.3245
ClF	791.38	792.11	791.38	784.96	784.81			0.00	783.4534
AlCl	481.75	482.80	482.42	482.11	482.12			0.00	481.7747
SiS	752.61	755.69	755.59	750.30	751.09			0.00	749.6456
error statistics for hydrides (10 systems) <sup>c</sup>									
MSD	1.88	4.66	1.96	-0.10	-0.13				
MAD	4.49	4.66	2.15	0.48	0.46				
RMSD	4.81	5.09	2.80	0.61	0.60				
error statistics for nonhydrides (18 systems) <sup>c</sup>									
MSD	4.61	9.11	11.57	1.67	0.97				
MAD	6.41	9.11	11.57	1.69	1.00				
RMSD	8.67	11.22	15.17	2.35	1.21				
error statistics for everything (28 systems) <sup>c</sup>									
MSD	3.64	7.52	8.14	1.04	0.58				
MAD	5.73	7.52	8.21	1.26	0.80				
RMSD	7.79	9.78	12.69	1.99	1.03				

<sup>a</sup>DBOC correction at the HF/AVTZ level. <sup>b</sup>From Ref.<sup>69</sup>. <sup>c</sup>See footnote *b* of Table I.

TABLE III: First-order anharmonic corrections ( $\omega_e x_e$ ,  $\text{cm}^{-1}$ ) of diatomic molecules obtained from a Dunham analysis of potential energy curves computed at the W4 and related levels.

	W4[up to CCSD(T) $f_{c,nr}$ ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	W4.2	W4.3	$\Delta\text{DBOC}^a$	Expt. <sup>b</sup>
BH	48.92	49.16	49.33	49.33	49.33	49.34	49.33	-0.02	49.33983
CH	63.66	63.89	64.19	64.26	64.26			-0.01	64.4387
NH	77.93	78.07	78.53	78.74	78.74			0.01	79.04
OH	84.38	84.48	84.72	85.01	85.02			0.00	84.8813
HF	89.73	89.83	89.87	90.09	90.10	90.10	90.13	0.01	89.9432
AlH	28.75	28.45	28.59	28.61	28.61			-0.01	29.05098
SiH	35.67	35.59	35.84	35.89	35.89			-0.01	36.0552
PH	43.10	43.09	43.46	43.58	43.58			0.00	43.907
SH	47.98	48.01	48.25	48.40	48.40			0.00	48.742
HCl	51.89	51.92	52.02	52.18	52.18	52.18	52.16	0.00	52.8
N <sub>2</sub>	13.86	13.90	13.64	14.14	14.24	14.26	14.24	0.00	14.324
O <sub>2</sub>	11.03	11.07	10.85	11.81	11.96			0.00	11.95127
F <sub>2</sub>	11.45	11.41	11.07	11.40	11.40	11.45	11.44	0.00	11.3221
Si <sub>2</sub>	1.93	1.94	1.92	2.02	2.03			0.00	2.02
P <sub>2</sub>	2.72	2.71	2.68	2.83	2.84	2.86		0.00	2.835
S <sub>2</sub>	2.73	2.73	2.70	2.85	2.86			0.00	2.8582
Cl <sub>2</sub>	2.61	2.59	2.62	2.67	2.67	2.68		0.00	2.69427
BF	11.71	11.81	11.85	11.86	11.86			0.00	11.82106
CO	13.11	13.17	13.12	13.33	13.30	13.30	13.27	0.00	13.28803
BCl	5.37	5.42	5.44	5.47	5.47			0.00	5.4917
CS	6.32	6.37	6.36	6.54	6.52	6.51		0.00	6.502605
SiO	5.84	5.90	5.83	6.02	5.99			0.00	5.97437
PN	6.56	6.57	6.44	6.76	6.81			0.00	6.8958
SO	6.16	6.17	6.09	6.47	6.51			0.00	6.4096
AlF	4.78	4.86	4.86	4.86	4.86			0.00	4.849915
ClF	4.86	4.85	4.84	4.95	4.95			0.00	4.9487
AlCl	2.08	2.10	2.10	2.11	2.11			0.00	2.101811
SiS	2.52	2.53	2.52	2.61	2.59			0.00	2.58623
error statistics for hydrides (10 systems) <sup>c</sup>									
MSD	-0.62	-0.57	-0.34	-0.21	-0.21				
MAD	0.62	0.57	0.34	0.27	0.27				
RMSD	0.68	0.64	0.41	0.32	0.32				
error statistics for nonhydrides (18 systems) <sup>c</sup>									
MSD	-0.18	-0.15	-0.22	-0.01	0.01				
MAD	0.19	0.17	0.22	0.05	0.03				
RMSD	0.28	0.26	0.35	0.07	0.04				
error statistics for everything (28 systems) <sup>c</sup>									
MSD	-0.34	-0.30	-0.26	-0.08	-0.07				
MAD	0.35	0.31	0.27	0.13	0.12				
RMSD	0.48	0.45	0.39	0.21	0.20				

<sup>a</sup>DBOC correction at the HF/AVTZ level. <sup>b</sup>From Ref.<sup>69</sup>. <sup>c</sup>See footnote *b* of Table I.



TABLE IV: Second-order anharmonic corrections ( $\omega_{ey_e}$ ,  $\text{cm}^{-1}$ ) of diatomic molecules obtained from a Dunham analysis of potential energy curves computed at the W4 and related levels.

	W4[up to CCSD(T) $f_{e,nr}$ ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	W4.2	W4.3	$\Delta\text{DBOC}^a$	Expt. <sup>b</sup>
BH	0.384	0.369	0.359	0.360	0.360	0.359	0.367	0.000	0.362
CH	0.322	0.314	0.301	0.298	0.300			-0.002	0.3634
NH	0.290	0.138	0.113	0.096	0.096			0.001	0.367
OH	0.654	0.518	0.515	0.507	0.507			0.003	0.5409
HF	1.064	1.052	1.063	1.056	1.056	1.056	1.052	0.002	0.92449
AlH	0.238	-0.030	-0.031	-0.032	-0.032			-0.001	0.24762
SiH	0.149	0.104	0.095	0.092	0.092			-0.002	0.1254
PH	0.029	0.012	-0.007	-0.015	-0.014			0.001	0.1059
SH	0.065	0.061	0.050	0.040	0.040			0.000	0.1124
HCl	0.021	-0.018	-0.020	-0.027	-0.027	-0.028	-0.028	0.001	0.21803
error statistics for hydrides (10 systems) <sup>c</sup>									
MSD	-0.015	-0.084	-0.093	-0.099	-0.099				
MAD	0.075	0.112	0.121	0.126	0.125				
RMSD	0.094	0.147	0.155	0.160	0.160				
error statistics for everything (28 systems) <sup>c</sup>									
MSD	-0.005	-0.032	-0.034	-0.038	-0.038				
MAD	0.030	0.045	0.049	0.051	0.051				
RMSD	0.061	0.095	0.100	0.103	0.103				

<sup>a</sup>DBOC correction at the HF/AVTZ level. <sup>b</sup>From Ref.<sup>69</sup>. <sup>c</sup>See footnote *b* of Table I.

TABLE V: Vibration-rotation coupling constants ( $\alpha_e$ ,  $\text{cm}^{-1}$ ) of diatomic molecules obtained from a Dunham analysis of potential energy curves computed at the W4 and related levels.

	W4[up to CCSD(T) $f_{c,nr}$ ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	W4.2	W4.3	$\Delta\text{DBOC}^a$	Expt. <sup>b</sup>
BH	0.41852	0.42106	0.42207	0.42205	0.42205	0.42207	0.42216	-0.00027	0.421565
CH	0.53115	0.53345	0.53503	0.53543	0.53544			-0.00026	0.53654
NH	0.64301	0.64513	0.64754	0.64869	0.64871			-0.00018	0.65038
OH	0.71515	0.71697	0.71830	0.72005	0.72009			-0.00016	0.7242
HF	0.78693	0.78854	0.78893	0.79043	0.79047	0.79048	0.79048	-0.00002	0.7933704
AlH	0.18372	0.18561	0.18625	0.18636	0.18636			-0.00010	0.1870527
SiH	0.21455	0.21641	0.21741	0.21761	0.21761			-0.00006	0.21814
PH	0.24869	0.25042	0.25180	0.25223	0.25223			-0.00005	0.25339
SH	0.27622	0.27774	0.27865	0.27922	0.27923			-0.00005	0.2799
HCl	0.30344	0.30488	0.30532	0.30594	0.30595	0.30593	0.30588	-0.00005	0.3069985
N <sub>2</sub>	0.01697	0.01703	0.01682	0.01722	0.01729	0.01731	0.01729	0.00000	0.017318
O <sub>2</sub>	0.01512	0.01518	0.01501	0.01578	0.01590			0.00000	0.0159305
F <sub>2</sub>	0.01244	0.01244	0.01225	0.01259	0.01260	0.01264	0.01263	0.00000	0.0125952
Si <sub>2</sub>	0.00128	0.00130	0.00130	0.00134	0.00134			0.00000	0.00135
P <sub>2</sub>	0.00142	0.00143	0.00141	0.00146	0.00147	0.00148		0.00000	0.00149
S <sub>2</sub>	0.00153	0.00155	0.00154	0.00159	0.00159			0.00000	0.00159754
Cl <sub>2</sub>	0.00145	0.00145	0.00145	0.00147	0.00147	0.00148		0.00000	0.001516
BF	0.01888	0.01901	0.01905	0.01906	0.01906			0.00000	0.01904848
CO	0.01733	0.01741	0.01737	0.01754	0.01752	0.01752	0.01750	0.00000	0.01750513
BCl	0.00667	0.00674	0.00677	0.00679	0.00679			0.00000	0.0068124
CS	0.00580	0.00585	0.00585	0.00594	0.00593	0.00593		0.00000	0.00591835
SiO	0.00494	0.00499	0.00496	0.00505	0.00504			0.00000	0.00503784
PN	0.00534	0.00537	0.00528	0.00547	0.00550			0.00000	0.0055337
SO	0.00557	0.00560	0.00556	0.00575	0.00577			0.00000	0.0057508
AlF	0.00490	0.00499	0.00500	0.00500	0.00500			0.00000	0.004984261
ClF	0.00425	0.00426	0.00426	0.00433	0.00433			0.00000	0.0043385
AlCl	0.00159	0.00161	0.00161	0.00162	0.00162			0.00000	0.001611082
SiS	0.00143	0.00145	0.00145	0.00148	0.00147			0.00000	0.00147313
error statistics for hydrides (10 systems) <sup>c</sup>									
MSD	-0.00514	-0.00325	-0.00214	-0.00147	-0.00146				
MAD	0.00514	0.00325	0.00219	0.00152	0.00150				
RMSD	0.00549	0.00379	0.00279	0.00192	0.00190				
error statistics for nonhydrides (18 systems) <sup>c</sup>									
MSD	-0.00016	-0.00012	-0.00016	-0.00002	-0.00001				
MAD	0.00016	0.00012	0.00016	0.00003	0.00002				
RMSD	0.00024	0.00021	0.00028	0.00005	0.00002				
error statistics for everything (28 systems) <sup>c</sup>									
MSD	-0.00194	-0.00124	-0.00087	-0.00054	-0.00053				
MAD	0.00194	0.00124	0.00089	0.00056	0.00055				
RMSD	0.00341	0.00236	0.00174	0.00119	0.00118				

<sup>a</sup>DBOC correction at the HF/AVTZ level. <sup>b</sup>From Ref.<sup>69</sup>. <sup>c</sup>See footnote *b* of Table I.

TABLE VI: Spectroscopic constants ( $r_e$  in pm, all the rest in  $\text{cm}^{-1}$ ) for the 12-electron isoelectronic series  $\text{C}_2$ ,  $\text{BN}$ , and  $\text{BeO}$  obtained from a Dunham analysis of potential energy curves computed at the W4 and related levels.<sup>a</sup>

	W4[up to CCSD(T) $f_{c,nr}$ ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	W4 <sup>b</sup>	W4 <sup>bc</sup>	W4.2 <sup>bc</sup>	W4.3 <sup>c</sup>	Expt. <sup>d</sup>
$\text{C}_2 (X^1\Sigma_g^+)$										
$r_e$	124.40	124.04	124.00	124.19	124.21		124.22	124.22	124.22	124.2440
$\omega_e$	1861.29	1872.11	1874.62	1859.09	1857.63		1857.43	1856.55	1855.84	1855.0142
$\omega_e x_e$	12.51	12.48	12.63	13.44	13.45		13.45	13.50	13.53	13.5547
$\omega_e y_e$	-0.020	-0.019	-0.039	-0.081	-0.082		-0.082	-0.083	-0.088	-0.1321
$\alpha_e$	0.01724	0.01726	0.01728	0.01783	0.01783		0.01783	0.01788	0.01789	0.018013
$\text{BN} (a^1\Sigma^+)$										
$r_e$	126.85	126.43	127.17	127.33	127.38	127.42	127.43	127.43	127.33	127.45081
$\omega_e$	1742.85	1755.27	1719.62	1713.06	1711.08	1707.50	1707.28	1707.64	1709.81	1705.4032
$\omega_e x_e$	16.33	16.08	12.29	11.99	11.76	11.40	11.40	11.37	11.68	10.55338
$\omega_e y_e$	-0.505	-0.511	-0.004	-0.013	-0.015	-0.015	-0.015	-0.024	0.036	
$\alpha_e$	0.01669	0.01663	0.01665	0.01638	0.01627	0.01618	0.01617	0.01611	0.01632	0.013857
$\text{BeO} (X^1\Sigma^+)$										
$r_e$	133.63	132.89	132.85	132.98	132.88	133.01	133.07	133.07		133.09
$\omega_e$	1479.86	1497.59	1507.02	1494.41	1501.65	1493.43	1488.55	1488.55		1487.32
$\omega_e x_e$	12.05	12.14	10.95	12.12	12.34	11.78	11.69	11.66		11.83
$\alpha_e$	0.01889	0.01900	0.01810	0.01893	0.01889	0.01881	0.01884	0.01882		0.0190

<sup>a</sup>For  $\text{C}_2$ ,  $\text{BN}$ , and  $\text{BeO}$  DBOC contributions are found to be insignificant and are not included in the theoretical values.

<sup>b</sup>Diffuse functions are added to electronegative atoms (N and O) in all the valence post-CCSD(T) steps, see text. <sup>c</sup>The inner-shell correlation contribution is extrapolated from RCCSD(T)/aug-cc-pwCVQZ and RCCSD(T)/aug-cc-pwCV5Z calculations, see text. <sup>d</sup> $\text{C}_2$  from Ref.<sup>70</sup>,  $\text{BN}$  from Ref.<sup>71</sup> and  $\text{BeO}$  from Ref.<sup>61</sup>.

TABLE VII: Theoretical<sup>a</sup> and experimental equilibrium geometries, harmonic frequencies, and fundamental frequencies (in pm, degrees, and cm<sup>-1</sup>) for H<sub>2</sub>O, CO<sub>2</sub>, and O<sub>3</sub>.

	$r_e$	$\theta_e$	$\omega_1$	$\omega_2$	$\omega_3$	$\nu_1$	$\nu_2$	$\nu_3$	Ref.
CO <sub>2</sub>									
W4	115.98		1351.9	670.5	2396.4	1331.5	666.4	2350.2	
Expt.	116.00		1353.8	672.9	2396.5	1332.9	672.9	2349.1	<sup>72</sup>
H <sub>2</sub> O									
W4	95.76	104.50	3834.6	1648.3	3945.7	3655.8	1595.3	3754.3	
W4 <sup>b</sup>			3835.2	1648.0	3946.3	3656.4	1595.2	3755.0	
Expt.	95.72	104.52	3832.2	1648.5	3942.5				<sup>50</sup>
Expt.	95.84	104.44	3831.8	1647.8	3942.4				<sup>51</sup>
Expt.						3657.1	1594.7	3755.9	<sup>73</sup>
O <sub>3</sub>									
W4[up to CCSDT]			1153.6	717.9	1085.7	1126.5	704.9	1042.1	
W4[up to CCSDTQ]			1137.0	715.5	1070.2	1105.3	701.5	1020.3	
W4	127.15	116.82	1133.9	715.2	1067.1	1100.4	701.5	1016.0	
W4 <sup>*c</sup>					1078.1				
Expt.	127.17	116.78	1134.9	716.0	1089.2				<sup>74</sup>
Expt.	127.17	116.75	1133.3	715.0	1087.3				<sup>49</sup>

<sup>a</sup>All the theoretical frequencies are computed at the W4-optimized geometries; unless otherwise indicated the DBOC contributions are not included in the theoretical values, cfr. footnote *b*. <sup>b</sup>Including a DBOC correction at the HF/aug'-cc-pVTZ level of theory. <sup>c</sup>W4\*=W4[up to CCSDT]+ $\hat{T}_4$ /cc-pVTZ(no *f* *1d*)+ $\hat{T}_5$ /cc-pVDZ(no *d*) see text.

TABLE VIII: Dipole moments (in a.u., 1 a.u. = 2.5416 Debye).

	W4[up to CCSD(T) $f_c, nr$ ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	Expt.	Ref.
BH	0.550	0.552	0.551	0.551		0.50	<sup>61</sup>
HF	0.709	0.709	0.709	0.709	0.709	0.7185 <sup>a</sup>	<sup>61</sup>
HCl	0.433	0.430	0.429	0.429	0.429	0.4362±0.0001	<sup>75</sup>
CO	0.044	0.046	0.044	0.048	0.049	0.0432	<sup>61</sup>
CS	0.774	0.777	0.771	0.774	0.778	0.770	<sup>61</sup>
BeO	2.496	2.484	2.469	2.470	2.484	2.2680	<sup>61</sup>
BN	0.796	0.796	0.850	0.811	0.799		
H <sub>2</sub> O	0.729	0.730	0.730	0.729	0.729	0.730±0.002	<sup>75</sup>
O <sub>3</sub>	0.218	0.218	0.222	0.214	0.213	0.2100	<sup>75</sup>
MSD <sup>b</sup>	0.035	0.034	0.031	0.031	0.033		
MAD <sup>b</sup>	0.038	0.038	0.035	0.035	0.038		
RMSD <sup>b</sup>	0.083	0.079	0.074	0.074	0.079		
MSD <sup>c</sup>	0.001	0.000	0.004	-0.001			
MAD <sup>c</sup>	0.004	0.002	0.010	0.004			
RMSD <sup>c</sup>	0.005	0.002	0.018	0.006			

<sup>a</sup>Ref.<sup>76</sup> gives 0.707 a.u. <sup>b</sup>Error statistics with respect to experiment. <sup>c</sup>Error statistics with respect to the W4 values.

TABLE IX: Static Polarizabilities (in a.u., 1 a.u. =  $1.48185 \times 10^{-25}$  cm<sup>3</sup>)<sup>a</sup>.

		W4[up to CCSD(T) <i>f<sub>c,nr</sub></i> ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	Expt.	Ref.
BH	$\alpha_{xx}$	20.87	20.78	20.78	20.77			
	$\alpha_{zz}$	23.14	23.03	23.05	23.06			
	$\bar{\alpha}$	21.63	21.53	21.53	21.53			
HF	$\alpha_{xx}$	5.15	5.14	5.14	5.14	5.14	5.08	<sup>b</sup>
	$\alpha_{zz}$	6.29	6.28	6.28	6.29	6.29	6.40	<sup>b</sup>
	$\bar{\alpha}$	5.53	5.52	5.52	5.53	5.53	5.52 <sup>c</sup>	<sup>b</sup>
HCl	$\alpha_{xx}$	16.68	16.65	16.65	16.65	16.65		
	$\alpha_{zz}$	18.32	18.30	18.30	18.30	18.30		
	$\bar{\alpha}$	17.23	17.20	17.20	17.20	17.20	17.39 <sup>d</sup>	<sup>77</sup>
CO	$\alpha_{xx}$	11.85	11.81	11.80	11.81	11.81	11.86	<sup>78</sup>
	$\alpha_{zz}$	15.42	15.38	15.36	15.40	15.42	15.51	<sup>78</sup>
	$\bar{\alpha}$	13.04	13.00	12.99	13.01	13.01	13.08	<sup>78</sup>
CS	$\alpha_{xx}$	23.74	23.62	23.63	23.62	23.63		
	$\alpha_{zz}$	37.93	37.81	37.79	37.87	37.85		
	$\bar{\alpha}$	28.47	28.35	28.35	28.37	28.37		
P <sub>2</sub>	$\alpha_{xx}$	39.91	39.66	39.66	39.65	39.65		
	$\alpha_{zz}$	67.86	67.69	67.77	67.52	67.68		
	$\bar{\alpha}$	49.23	49.00	49.03	48.94	49.00		
Cl <sub>2</sub>	$\alpha_{xx}$	24.86	24.81	24.81	24.81	24.81	24.43	<sup>79</sup>
	$\alpha_{zz}$	41.74	41.66	41.68	41.65	41.65	44.54	<sup>79</sup>
	$\bar{\alpha}$	30.48	30.43	30.44	30.43	30.43	31.11 <sup>e</sup>	<sup>79</sup>
H <sub>2</sub> O	$\alpha_{xx}$	9.20	9.18	9.18	9.18	9.18		
	$\alpha_{yy}$	9.84	9.83	9.83	9.84	9.84		
	$\alpha_{zz}$	9.48	9.47	9.47	9.48	9.48		
	$\bar{\alpha}$	9.50	9.49	9.49	9.50	9.50	9.642 <sup>f</sup>	<sup>77</sup>
CO <sub>2</sub>	$\alpha_{xx}$	12.74	12.74	12.73	12.75	12.75		
	$\alpha_{zz}$	26.55	26.55	26.51	26.57	26.61		
	$\bar{\alpha}$	17.34	17.34	17.32	17.36	17.37	17.51 <sup>g</sup>	<sup>77</sup>
N <sub>2</sub>	$\alpha_{xx}$	10.19	10.15	10.15	10.16	10.16		
	$\alpha_{zz}$	14.78	14.73	14.76	14.71	14.72		
	$\bar{\alpha}$	11.72	11.68	11.69	11.68	11.68	11.744 <sup>h</sup>	<sup>75</sup>
F <sub>2</sub>	$\alpha_{xx}$	6.38	6.37	6.37	6.37	6.37		
	$\alpha_{zz}$	12.34	12.32	12.36	12.26	12.26		
	$\bar{\alpha}$	8.37	8.35	8.36	8.33	8.33	8.38	<sup>80</sup>
C <sub>2</sub>	$\alpha_{xx}$	36.88	38.68	22.86	23.82	23.86		
	$\alpha_{zz}$	31.80	31.81	26.36	26.33	26.42		
	$\bar{\alpha}$	35.18	36.39	24.03	24.66	24.71		
BeO	$\alpha_{xx}$	31.46	31.36	30.63	31.30	31.59		
	$\alpha_{zz}$	33.39	33.05	34.01	32.91	31.90		
	$\bar{\alpha}$	32.10	31.92	31.76	31.84	31.70		
O <sub>3</sub>	$\alpha_{xx}$	11.75	11.72	11.72	11.73	11.73		
	$\alpha_{yy}$	31.26	31.27	31.03	30.81	30.74		
	$\alpha_{zz}$	14.18	14.14	14.10	14.11	14.11		
	$\bar{\alpha}$	19.06	19.04	18.95	18.88	18.86	21.7	<sup>75</sup>
	MSD <sup>i</sup>	0.72	0.72	0.01	0.01			
	MAD <sup>i</sup>	0.73	0.75	0.17	0.06			
	RMSD <sup>i</sup>	2.59	2.89	0.47	0.20			

<sup>a</sup> $\bar{\alpha} = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$ . <sup>b</sup>Expt. value with zero-point vibrational corrections taken from Ref.<sup>81</sup>. <sup>c</sup>Ref.<sup>75</sup> gives:  $\bar{\alpha}=5.4$ , and Ref.<sup>77</sup> gives  $\bar{\alpha}=5.601$  a.u. <sup>d</sup>Ref.<sup>82</sup> gives:  $\bar{\alpha}=17.55$  a.u., Ref.<sup>79</sup> gives:  $\alpha_{xx}=16.13$ ,  $\alpha_{zz}=21.12$ , and  $\bar{\alpha}=17.75$  a.u. <sup>e</sup>Ref.<sup>82</sup> gives:  $\bar{\alpha}=31.1$  a.u. <sup>f</sup>Ref.<sup>75</sup> gives:  $\bar{\alpha}=9.8$  a.u. <sup>g</sup>Ref.<sup>82</sup> gives:  $\bar{\alpha}=17.7$  a.u., Ref.<sup>79</sup> gives:  $\bar{\alpha}=17.88$  a.u., and Ref.<sup>75</sup> gives:  $\bar{\alpha}=19.64$  a.u. <sup>h</sup>Ref.<sup>79</sup> gives:  $\alpha_{xx} = 9.79$ ,  $\alpha_{zz}=16.06$ , and  $\bar{\alpha}=11.88$  a.u., and Ref.<sup>82</sup> gives:  $\bar{\alpha}=11.92$  a.u. <sup>i</sup>Error statistics with respect to the W4 values.

TABLE X: Static Hyperpolarizabilities (in a.u., 1 a.u. =  $8.6392 \times 10^{-33} \text{cm}^4 \text{statvolt}^{-1}$ )<sup>a</sup>.

	W4[up to CCSD(T) <i>f<sub>c,nr</sub></i> ]	W4 [up to CCSD(T)]	W4 [up to CCSDT]	W4 [up to CCSDTQ]	W4	
BH	$\beta_{zzz}$	21.81	22.39	22.05	22.16	
	$\beta_{xxz}$	69.95	68.92	68.88	69.06	
	$\bar{\beta}$	70.85	69.27	69.43	69.57	
HF	$\beta_{zzz}$	8.75	8.77	8.79	8.82	8.83
	$\beta_{xxz}$	1.03	1.03	1.04	1.05	1.05
	$\bar{\beta}$	6.49	6.50	6.52	6.55	6.55
HCl	$\beta_{zzz}$	9.77	9.78	9.81	9.81	9.81
	$\beta_{xxz}$	0.29	0.33	0.32	0.31	0.31
	$\bar{\beta}$	6.21	6.26	6.28	6.26	6.26
CO	$\beta_{zzz}$	28.63	28.45	28.32	28.67	28.5
	$\beta_{xxz}$	7.66	7.60	7.56	7.64	7.67
	$\bar{\beta}$	26.37	26.19	26.06	26.38	26.31
CS	$\beta_{zzz}$	15.95	15.22	15.62	15.90	12.81
	$\beta_{xxz}$	4.79	4.48	4.38	4.61	4.72
	$\bar{\beta}$	15.31	14.50	14.63	15.07	13.35
H <sub>2</sub> O	$\beta_{zzz}$	11.86	11.90	11.90	11.96	11.96
	$\beta_{xxz}$	4.62	4.62	4.63	4.66	4.66
	$\beta_{yyz}$	9.48	9.51	9.53	9.56	9.56
	$\bar{\beta}$	15.58	15.62	15.64	15.70	15.70
O <sub>3</sub>	$\beta_{zzz}$	-17.32	-17.22	-17.40	-17.43	-17.44
	$\beta_{xxz}$	-3.61	-3.59	-3.61	-3.62	-3.63
	$\beta_{yyz}$	1.66	1.92	-5.32	-7.22	-7.04
	$\bar{\beta}$	-11.56	-11.33	-15.80	-16.96	-16.87
	MSD <sup>b</sup>	0.77	0.70	0.22	0.18	
	MAD <sup>b</sup>	0.77	0.70	0.32	0.20	
	RMSD <sup>b</sup>	2.26	2.26	0.81	0.75	

<sup>a</sup> $\bar{\beta} = \frac{3}{5}(\beta_{xxz} + \beta_{yyz} + \beta_{zzz})$ . <sup>b</sup>Error statistics with respect to the W4 values.

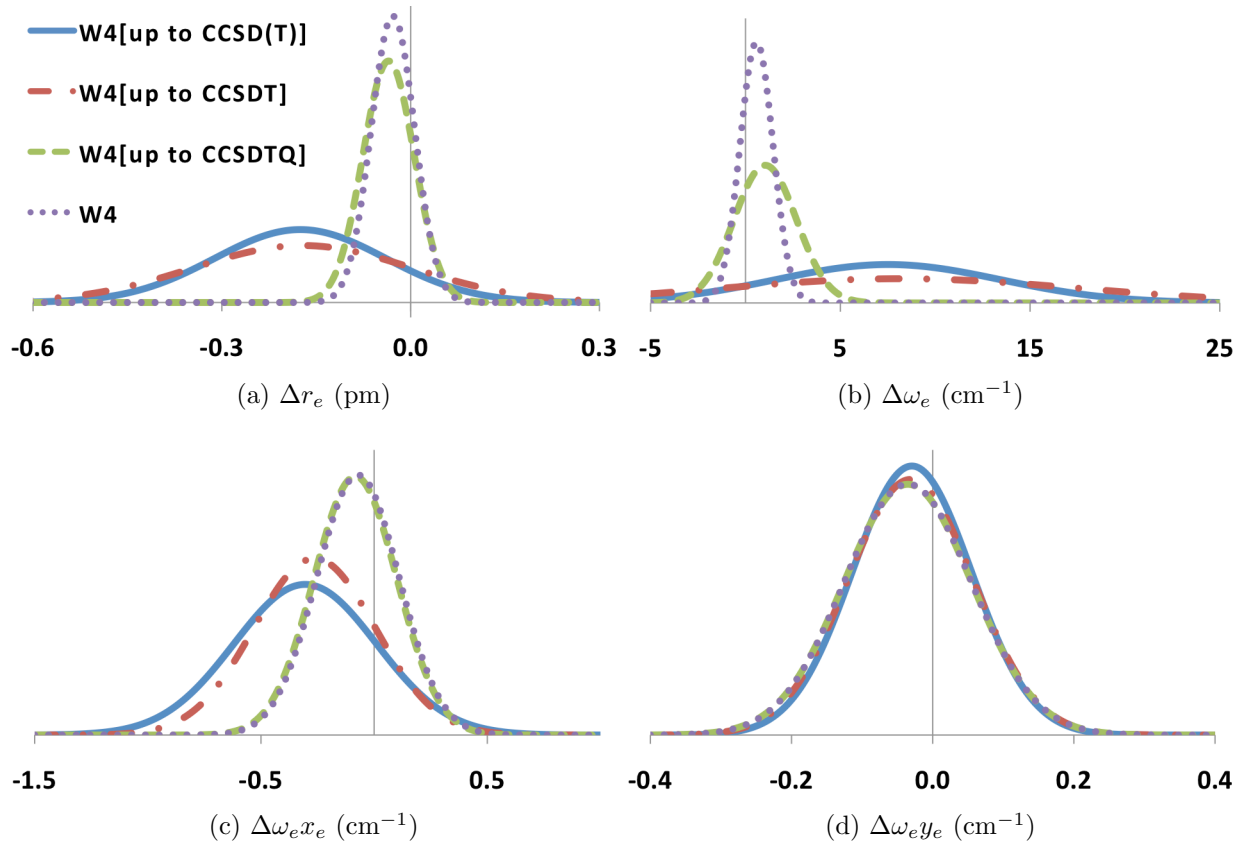


FIG. 1: Normal distribution of errors between theory and experiment for  $r_e$ ,  $\omega_e$ ,  $\omega_e x_e$ , and  $\omega_e y_e$ . The centers of the Gaussians for the W4[up to CCSD(T)], W4[up to CCSDT], W4[up to CCSDTQ], and W4 methods, respectively, are given in parenthesis. (a) equilibrium bond lengths (-0.16, -0.17, -0.04, -0.03 pm), (b) harmonic frequencies (7.5, 8.1, 1.0, and 0.6  $\text{cm}^{-1}$ ), (c) first-order anharmonic corrections (-0.30, -0.26, -0.08, -0.07  $\text{cm}^{-1}$ ), and (d) second-order anharmonic corrections (-0.03, -0.03, -0.04, -0.04  $\text{cm}^{-1}$ ).



- 
- † Electronic address: gershom@weizmann.ac.il
- <sup>1</sup> J. M. L. Martin, T. J. Lee, P. R. Taylor, and J. P. François, *J. Chem. Phys.* **103**, 2589 (1995).
  - <sup>2</sup> J. M. L. Martin, *Chem. Phys. Lett.* **292** 411 (1998).
  - <sup>3</sup> F. Pawłowski, A. Halkier, P. Jørgensen, K. L. Bak, T. Helgaker, and W. Klopper, *J. Chem. Phys.* **118**, 2539 (2003).
  - <sup>4</sup> T. A. Ruden, T. Helgaker, J. Gauss, P. Jørgensen, and J. Olsen, *J. Chem. Phys.* **121**, 5874 (2004).
  - <sup>5</sup> D. P. Tew, W. Klopper, M. Heckert, and J. Gauss, *J. Phys. Chem. A* **111**, 11242 (2007).
  - <sup>6</sup> M. Heckert, M. Kállay, and J. Gauss, *Mol. Phys.* **103**, 2109 (2005).
  - <sup>7</sup> M. Heckert, M. Kállay, D. P. Tew, W. Klopper, and J. Gauss, *J. Chem. Phys.* **125**, 044108 (2006).
  - <sup>8</sup> J. M. L. Martin, T. J. Lee, G. E. Scuseria, and P. R. Taylor, *J. Chem. Phys.* **97**, 6549 (1992).
  - <sup>9</sup> R. Shepard, G. S. Kedziora, H. Lischka, I. Shavitt, T. Müller, P. G. Szalay, M. Kállay, and M. Seth, *Chem. Phys.* **349**, 37 (2008).
  - <sup>10</sup> X. Huang and T. J. Lee, *J. Chem. Phys.* **129**, 044312 (2008); *ibid.* **131**, 104301 (2009); X. Huang, D. W. Schwenke, and T. J. Lee, *J. Chem. Phys.* **129**, 214304 (2008).
  - <sup>11</sup> C. Michauk and J. Gauss, *J. Chem. Phys.* **127**, 044106 (2007).
  - <sup>12</sup> B. Mintz, T. G. Williams, L. Howard, and A. K. Wilson, *J. Chem. Phys.* **130**, 234104 (2009).
  - <sup>13</sup> N. J. DeYonker, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **124**, 42 114104 (2006).
  - <sup>14</sup> A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, *J. Chem. Phys.*, **121**, 11599 (2004).
  - <sup>15</sup> Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, and J. F. Stanton, *J. Chem. Phys.*, **125**, 064108 (2006).
  - <sup>16</sup> M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, *J. Chem. Phys.* **128**, 114111 (2008).
  - <sup>17</sup> A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, *J. Chem. Phys.* **125**, 144108 (2006).
  - <sup>18</sup> A. Karton, P. R. Taylor, and J. M. L. Martin, *J. Chem. Phys.* **127**, 064104 (2007).
  - <sup>19</sup> T. Helgaker, W. Klopper, and D. P. Tew, *Mol. Phys.* **106**, 2107 (2008).
  - <sup>20</sup> B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszewski, S. Bittner, S. G. Nijssure, K.

- A. Amin, M. Minkoff, and A. F. Wagner, *J. Phys. Chem. A* **108**, 9979 (2004); B. Ruscic, "Active Thermochemical Tables", in *2005 Yearbook of Science and Technology* (annual update to *McGraw-Hill Encyclopedia of Science and Technology*), McGraw-Hill, New York (2004), pp. 3-7; B. Ruscic, R. E. Pinzon, M. L. Morton, N. K. Srinivasan, M.-C. Su, J. W. Sutherland, and J. V. Michael, *J. Phys. Chem. A* **110**, 6592 (2006).
- <sup>21</sup> A. Karton, S. Parthiban, and J. M. L. Martin, *J. Phys. Chem. A* **113**, 4802 (2009).
- <sup>22</sup> D. Feller and K. A. Peterson, *J. Chem. Phys.* **126**, 114105 (2007).
- <sup>23</sup> W. Klopper, R. A. Bachorz, C. Hättig, and D. P. Tew, *Theor. Chem. Acc.*, Online First (2010) <http://dx.doi.org/10.1007/s00214-010-0733-7>.
- <sup>24</sup> MOLPRO, version 2009.1, a package of *ab initio* programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, P. Celani, T. Korona, A. Mitrushenkov, G. Rauhut, 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, G. Hetzer, T. Hrenar, G. Knizia, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and A. Wolf, see <http://www.molpro.net>.
- <sup>25</sup> MRCC, a string-based general coupled cluster program suite written by M. Kállay, See also M. Kállay and P. R. Surján, *J. Chem. Phys.* **115**, 2945 (2001), as well as: <http://www.mrcc.hu>.
- <sup>26</sup> CFOUR, a quantum chemical program package written by J. F. Stanton, J. Gauss, M. E. Harding, P. G. Szalay, with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, Y. J. Bomble, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W. J. Lauderdale, D. A. Matthews, T. Metzroth, D. P. O'Neill, D. R. Price, E. Prochnow, K. Ruud, F. Schiffmann, S. Stopkowicz, M. E. Varner, J. Vázquez, F. Wang, J. D. Watts, and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see <http://www.cfour.de>.
- <sup>27</sup> T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>28</sup> R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- <sup>29</sup> T. H. Dunning, K. A. Peterson, and A. K. Wilson, *J. Chem. Phys.* **114**, 9244 (2001).
- <sup>30</sup> K. A. Peterson and T. H. Dunning, *J. Chem. Phys.* **117**, 10548 (2002).

- <sup>31</sup> W. A. de Jong, R. J. Harrison, and D. A. Dixon, *J. Chem. Phys.* **114**, 48 (2001).
- <sup>32</sup> A. Karton and J. M. L. Martin, *Theor. Chem. Acc.* **115**, 330 (2006).
- <sup>33</sup> F. Jensen, *Theor. Chem. Acc.* **113**, 267 (2005).
- <sup>34</sup> W. Klopper, *Mol. Phys.* **99**, 481 (2001).
- <sup>35</sup> P. J. Knowles, C. Hampel, and H.-J. Werner, *J. Chem. Phys.* **99**, 5219 (2000); *erratum* **112**, 3106 (2000).
- <sup>36</sup> J. D. Watts, J. Gauss, and R. J. Bartlett, *J. Chem. Phys.* **98**, 8718 (1993).
- <sup>37</sup> M. Douglas and N. M. Kroll, *Ann. Phys.* **82**, 89 (1974).
- <sup>38</sup> B. A. Heß, *Phys. Rev. A* **33**, 3742 (1986).
- <sup>39</sup> J. L. Dunham, *Phys. Rev.* **41**, 721 (1932); *ibid* **41**, 713 (1932).
- <sup>40</sup> A. Ralston and P. Rabinowitz, *A first course in numerical analysis*, (Dover Publications, Mineola, NY, 2000), pp. 94-95 and pp. 148-149.
- <sup>41</sup> T. J. Lee and P. R. Taylor, *Int. J. Quantum Chem. Symp.* **23**, 199 (1989); T. J. Lee, J. E. Rice, G. E. Scuseria, and H. F. Schaefer III, *Theor. Chim. Acta* **75**, 81 (1989).
- <sup>42</sup> T. J. Lee, *Chem. Phys. Lett.* **372**, 362 (2003); M. L. Leininger, I. M. B. Nielsen, T. D. Crawford, and C. L. Janssen, *Chem. Phys. Lett.* **328**, 431 (2000); C. L. Janssen and I. M. B. Nielsen, *Chem. Phys. Lett.* **290**, 423 (1998); I. M. B. Nielsen and C. L. Janssen, *Chem. Phys. Lett.* **310**, 568 (1999).
- <sup>43</sup> J. D. Watts and R. J. Bartlett, *Int. J. Quantum Chem.* **52**, 195 (1994).
- <sup>44</sup> J. D. Watts, M. Urban, and R. J. Bartlett, *Theor. Chim. Acta* **90**, 341 (1995).
- <sup>45</sup> J. Zheng, Y. Zhao, and D. G. Truhlar, *J. Phys. Chem. A* **111**, 4632 (2007).
- <sup>46</sup> See Supplementary Material Document No. \_\_\_\_\_ Tables SI–SVI. For information on Supplementary Material, see <http://www.aip.org/pubservs/epaps.html>.
- <sup>47</sup> For a review see: R. Ahlrichs, *Comp. Phys. Commun.* **17**, 31 (1979).
- <sup>48</sup> P. Pulay, W. Meyer, and J. E. Boggs, *J. Chem. Phys.* **68**, 5077 (1978).
- <sup>49</sup> A. Barbe, A. Chichery, T. Cours, V. G. Tyuterev, and J. J. Plateaux, *J. Mol. Struct. (Theochem)* **616**, 55 (2002).
- <sup>50</sup> S. Benedict, N. Gailar, and E. K. Plyler, *J. Chem. Phys.* **24**, 1139 (1956).
- <sup>51</sup> P. Jensen, *J. Mol. Spectrosc.* **128**, 478 (1988).
- <sup>52</sup> A. G. Császár, G. Czakó, T. Furtenbacher, J. Tennyson, V. Szalay, S. V. Shirin, N. F. Zobov, and O. L. Polyansky, *J. Chem. Phys.* **122**, 214305 (2005).

- <sup>53</sup> I. Ljubić and A. Sabljčić, *Chem. Phys. Lett.* **385** 214 (2004).
- <sup>54</sup> K. W. Sattelmeyer, H. F. Schaefer III, and J. F. Stanton, *Chem. Phys. Lett.* **378** 42 (2003).
- <sup>55</sup> R. J. Bartlett and J. F. Stanton, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1994), vol. 5, ch. 2, pp. 65.
- <sup>56</sup> T. J. Lee, W. D. Allen, and H. F. Schaefer III, *J. Chem. Phys.* **87**, 7063 (1987).
- <sup>57</sup> J. F. Stanton, W. N. Lipscomb, D. H. Magers, and R. J. Bartlett, *J. Chem. Phys.* **90**, 1077 (1989).
- <sup>58</sup> S. A. Kucharski and R. J. Bartlett, *J. Chem. Phys.* **110**, 8233 (1999).
- <sup>59</sup> The W4 values are taken from Ref.<sup>17</sup>, and due to computational resources constraints the CCSDTQ–CCSDT(Q) difference in W4.3 theory is calculated with the *spd* part of the cc-pVTZ basis set instead of the regular cc-pVTZ basis set.
- <sup>60</sup> J. M. L. Martin, *J. Mol. Spectrosc.* **145** 392 (1991).
- <sup>61</sup> K. P. Huber and G. Herzberg, *Constants of Diatomic Molecules*, (Van Nostrand Reinhold, New York, 1979).
- <sup>62</sup> J. B. White, M. Dulick, P. F. Bernath, *J. Chem. Phys.* **99**, 8371 (1993).
- <sup>63</sup> M. Betrencourt, D. Boudjaadar, P. Chollet, G. Guelachvili, and M. Morillon-Chapey, *J. Chem. Phys.* **84** 4121 (1986).
- <sup>64</sup> N. Ohashi, K. Kawaguchi, and E. Hirota, *J. Mol. Spectrosc.* **103**, 337 (1984).
- <sup>65</sup> R. S. Ram, P. F. Bernath, R. Engleman, Jr. and J. W. Brault, *J. Mol. Spectrosc.* **172**, 34 (1995).
- <sup>66</sup> J. K. G. Watson, *J. Mol. Spectrosc.* **45**, 99 (1973).
- <sup>67</sup> A. G. Maki, F. J. Lovas, R. D. Suenram *J. Mol. Spectrosc.* **91**, 424 (1982).
- <sup>68</sup> H. S. P. Müller, M. C. McCarthy, L. Bizzocchi, H. Gupta, S. Esser, H. Lichau, M. Caris, F. Lewen, J. Hahn, C. Degli Esposti, S. Schlemmer, and P. Thaddeus, *Phys. Chem. Chem. Phys.* **9**, 1579 (2007).
- <sup>69</sup> K. K. Irikura, *J. Phys. Chem. Ref. Data* **36**, 389 (2007).
- <sup>70</sup> M. Douay, R. Nietmann, and P. F. Bernath, *J. Mol. Spectrosc.* **131**, 250 (1988).
- <sup>71</sup> R. S. Ram and P. F. Bernath, *J. Mol. Spectrosc.* **180**, 414 (1996).
- <sup>72</sup> J. L. Teffo, O. N. Sulakshina, V. I. Perevalov, *J. Mol. Spectrosc.* **156**, 48 (1992).
- <sup>73</sup> S. V. Shirin, N. F. Zobov, R. I. Ovsyannikov, O. L. Polyansky, and J. Tennyson, *J. Chem. Phys.* **128**, 224306 (2008).

- <sup>74</sup> A. Barbe, C. Secroun, P. Jouve, *J. Mol. Spectrosc.* **49**, 171 (1974).
- <sup>75</sup> D. R. Lide, *CRC Handbook of Chemistry and Physics*, 85th ed., (CRC Press/Taylor and Francis, Boca Raton, FL., 2004-2005).
- <sup>76</sup> J. S. Muentner and W. Klemperer, *J. Chem. Phys.* **52**, 6033 (1970).
- <sup>77</sup> M. A. Spackman, *J. Chem. Phys.* **94**, 1288 (1991).
- <sup>78</sup> K. K. Sunil, K. D. Jordan, *Chem. Phys. Lett.* **145**, 377 (1988).
- <sup>79</sup> J. O. Hirschfelder, C. F. Curtis, and R. B. Bird, *Molecular Theory of Gases and Liquids*, (Wiley, New York, 1954) pp. 950.
- <sup>80</sup> G. C. Straty and B. A. Younglove, *J. Chem. Phys.* **57**, 2255 (1972).
- <sup>81</sup> G. H. F. Diercksen, A. J. Sadlej, *J. Chem. Phys.* **75**, 1253 (1981).
- <sup>82</sup> N. J. Bridge and A. D. Buckingham, *Proc. R. Soc. Lond.* **295**, 334 (1966).