

## On the effectiveness of CCSD(T) complete basis set extrapolations for atomization energies

David Feller, Kirk A. Peterson, and J. Grant Hill

Citation: *J. Chem. Phys.* **135**, 044102 (2011); doi: 10.1063/1.3613639

View online: <http://dx.doi.org/10.1063/1.3613639>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v135/i4>

Published by the [American Institute of Physics](#).

---

### Additional information on *J. Chem. Phys.*

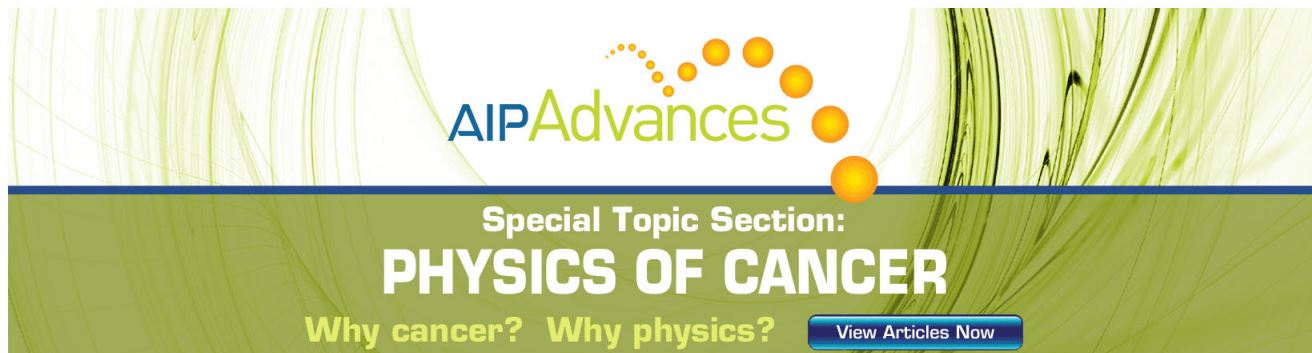
Journal Homepage: <http://jcp.aip.org/>

Journal Information: [http://jcp.aip.org/about/about\\_the\\_journal](http://jcp.aip.org/about/about_the_journal)

Top downloads: [http://jcp.aip.org/features/most\\_downloaded](http://jcp.aip.org/features/most_downloaded)

Information for Authors: <http://jcp.aip.org/authors>

## ADVERTISEMENT



**AIP Advances**

Special Topic Section:  
**PHYSICS OF CANCER**

Why cancer? Why physics? [View Articles Now](#)

# On the effectiveness of CCSD(T) complete basis set extrapolations for atomization energies

David Feller,<sup>a)</sup> Kirk A. Peterson,<sup>b)</sup> and J. Grant Hill<sup>c)</sup>*Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, USA*

(Received 2 May 2011; accepted 28 June 2011; published online 22 July 2011)

The leading cause of error in standard coupled cluster theory calculations of thermodynamic properties such as atomization energies and heats of formation originates with the truncation of the one-particle basis set expansion. Unfortunately, the use of finite basis sets is currently a computational necessity. Even with basis sets of quadruple zeta quality, errors can easily exceed 8 kcal/mol in small molecules, rendering the results of little practical use. Attempts to address this serious problem have led to a wide variety of proposals for simple complete basis set extrapolation formulas that exploit the regularity in the correlation consistent sequence of basis sets. This study explores the effectiveness of six formulas for reproducing the complete basis set limit. The W4 approach was also examined, although in lesser detail. Reference atomization energies were obtained from standard coupled-cluster singles, doubles, and perturbative triples (CCSD(T)) calculations involving basis sets of 6 $\zeta$  or better quality for a collection of 141 molecules. In addition, a subset of 51 atomization energies was treated with explicitly correlated CCSD(T)-F12b calculations and very large basis sets. Of the formulas considered, all proved reliable at reducing the one-particle expansion error. Even the least effective formulas cut the error in the raw values by more than half, a feat requiring a much larger basis set without the aid of extrapolation. The most effective formulas cut the mean absolute deviation by a further factor of two. Careful examination of the complete body of statistics failed to reveal a single choice that outperformed the others for all basis set combinations and all classes of molecules. © 2011 American Institute of Physics. [doi:10.1063/1.3613639]

## I. INTRODUCTION

The accuracy of standard wave function-based methods for solving the molecular Schrödinger equation hinge on two principal expansions that underlie most of the widely used approaches. The first expansion, which is associated with the one-particle basis set, has received much more attention in the chemical literature in recent years than the complementary  $n$ -particle expansion. To date, over 300 families of Gaussian basis sets have been proposed. This abundance reflects, in part, the wide variety of needs among computational chemists. Some researchers are interested in very high accuracy, while others seek only a qualitative understanding of molecular phenomena. Different properties also place different demands on the basis set. The strongest motivation driving the development of new basis sets is the steep rise in computational time as the size and quality of the basis set improves. This overarching characteristic of electronic structure methods places a premium on using the smallest possible basis set capable of providing the “right” answer. The present work will focus on techniques for obtaining the highest possible levels of basis set accuracy, while at the same time avoiding the expense of a brute force assault on the problem. The ultimate goal is represented by the complete basis set (CBS) limit where

all error attributable to the one-particle expansion vanishes. Achieving this goal is hindered by the simultaneous desire to combine large basis sets with a high quality  $n$ -particle expansion method. For example, density functional theory (DFT) is known to provide much faster convergence with respect to the basis set than wave function or orbital expansion methods such as coupled cluster theory. Presently, DFT is not yet capable of consistently providing high accuracy across a wide range of chemical systems and properties.

The cause of the slow basis set convergence is the widely recognized difficulty that orbital product expansion methods have in reproducing the electron-electron Coulomb cusps of the exact wave function, represented by<sup>1</sup>

$$\lim_{(r_{12} \rightarrow 0)} \Psi(r_{12}) \rightarrow 1 + 0.5r_{12}, \quad (1)$$

where  $r_{12}$  is the interelectronic distance. Explicitly correlated methods, especially those that involve nonlinear terms in the interelectronic distance,  $r_{ij}$ , offer more rapid basis set convergence.<sup>2–12</sup> While such methods offer much hope for ameliorating the one-particle basis set expansion problem, they are not the focus of the current investigation.

Fixed node, variational and diffusion quantum Monte Carlo (QMC) techniques also allow one to circumvent the slow basis set convergence and steep scaling problems of standard wave function techniques.<sup>13</sup> In a comparison with frozen core (FC) coupled cluster theory, the diffusion QMC method produced reaction enthalpies that were slightly less accurate than coupled cluster cc-pVTZ basis set results for a collection of 17 test cases involving small first row molecules.<sup>14</sup>

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: dfeller@owt.com.

<sup>b)</sup> Electronic mail: kipeters@wsu.edu.

<sup>c)</sup> Present address: School of Chemistry, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, United Kingdom. Electronic mail: grant.hill@glasgow.ac.uk.

The coupled cluster theory method included iterative single and double excitations plus a quasiperturbative treatment of “connected” triple excitations, a technique commonly known as coupled-cluster single double (triple) (CCSD(T)). In the worst case, the QMC value was in error by 28 kcal/mol. A more recent study by Nemeč *et al.* found that diffusion QMC was capable of reproducing the total atomization energies of 55 small molecules with a mean absolute deviation (MAD), or mean unsigned deviation, of 3.2 kcal/mol.<sup>15</sup> It should be noted that atomization energies have proven to be among the most difficult thermochemical properties to compute. In multiple instances the errors exceeded 5 kcal/mol. For the same collection of molecules, we have reported results that are within 0.5 kcal/mol of experiment.<sup>16</sup> Thus, at this stage in their development QMC techniques do not appear to offer the same level of thermochemical accuracy as coupled cluster theory.

Within the realm of coupled cluster theory, the  $n$ -particle expansion corresponds to a sequence of methods incorporating higher and higher excitation levels. If the expansion is truncated at double excitations, coupled cluster theory yields a relatively poor MAD of  $\sim 6$  kcal/mol with respect to high quality experimental results.<sup>17</sup> Since “chemical accuracy,” the generally accepted target for thermochemical properties, corresponds to an error of  $\pm 1$  kcal/mol (4.184 kJ/mol) or less, CCSD is inadequate for our purposes. Incorporating triple excitations via the CCSD(T)(FC) method cuts the MAD in half if sufficiently large basis sets can be used. Further cuts in the error are achievable by including a number of smaller but fairly expensive corrections. The perturbative (T) correction produces a result that falls fortuitously closer to the full configuration interaction (FCI) limit than the fully iterative inclusion of triples via CCSDT. Thus, CCSD(T) provides a good starting point for many approaches that aim for high accuracy, including the  $W_n$  methods of Martin and co-workers,<sup>18–21</sup> the HEATxyz methods,<sup>22–24</sup> the ccCA methods of Wilson, Cundari, and co-workers,<sup>25–27</sup> as well as our own composite approach.<sup>28–42</sup> The power of the last of these is demonstrated by its ability to reproduce experimental heats of formation with an accuracy of 0.19 kcal/mol ( $\epsilon_{\text{MAD}}$ ), 0.30 kcal/mol (root mean square deviation,  $\epsilon_{\text{RMS}}$ ) and  $-1.25$  kcal/mol (maximum deviation,  $\epsilon_{\text{MAX}}$ ). The statistical metrics were obtained from 123 comparisons involving molecules whose experimental uncertainties were  $\leq 1$  kcal/mol. Still closer agreement was found with a smaller set of 64 molecules whose experimental uncertainties are  $\leq 0.2$  kcal/mol. There we find  $\epsilon_{\text{MAD}} = 0.08$ ,  $\epsilon_{\text{RMS}} = 0.10$ , and  $\epsilon_{\text{MAX}} = -0.24$  kcal/mol. These statistics relied on information stored in the Computational Results Database (CRDB), which contains  $\sim 120\,000$  experimental and theoretical entries covering 363 molecules.<sup>43</sup>

The goal of the present study is to measure the effectiveness of different CCSD(T)(FC) complete basis set extrapolation formulas when used for predicting atomization energies. In our composite coupled cluster procedure, which was developed with high accuracy thermochemistry in mind, the CCSD(T)(FC)/CBS component strongly dominates the other components and at the same time is the most slowly convergent. This behavior is observed in other composite approaches that also incorporate a CCSD(T)(FC) step. Although our fo-

cus will be on frozen core atomization energies, many of the same extrapolation formulas can be applied to the more rapidly convergent core/valence (CV) correction. To the best of our knowledge no head-to-head, large scale study of the effectiveness of various CBS extrapolation formulas has been reported.

## II. SELECTED APPROACHES FROM THE LITERATURE

The critical importance of the one-particle expansion and the need to mitigate as much of the basis set truncation error as possible is reflected in the chemical literature. There are a large number of papers touching on this topic. We first discuss the six approaches to be treated in detail, followed by a brief survey of selected other formulas. Due to the large number of papers, a comprehensive review of the relevant literature is beyond the scope of the current work.

Shortly after the introduction of the diffuse function augmented correlation consistent basis sets<sup>44,45</sup> in 1992, they were used in a study of the binding energy of the water dimer.<sup>46</sup> Motivated by the observation that the first three members of this sequence produced atomic correlation energies that converged in roughly an exponential manner, a simple functional form,

$$E(n) = E_{\text{CBS}} + Ae^{-bn}, \quad (2)$$

was used to estimate the binding energy at the basis set limit. In this expression,  $n$  is the index of the basis set, i.e.,  $n = 2$  for aug-cc-pVDZ, 3 for aug-cc-pVTZ, etc. The three parameters  $E_{\text{CBS}}$ ,  $A$  and  $b$  were exactly fit to a sequence of three energies. While a  $5\zeta$  correlation consistent basis set was developed at approximately the same time, diffuse functions had not been reported. When the first row aug-cc-pV5Z basis sets were finally released, their use in the case of the water dimer would have required a 574 function calculation which would have been prohibitively expensive with the hardware available in 1992. Four years later, it became possible to carry out the aug-cc-pV5Z calculation using the resolution of the identity Møller-Plesset perturbation theory method.<sup>47</sup> In light of the scarcity of evidence available at the time, the use of an exponential function to estimate the basis set limit was largely a conjecture. To the best of our knowledge, this was the first attempt to extrapolate the correlation consistent basis set sequence to the CBS limit using a simple functional form. Other expressions would soon follow.

Shortly after the appearance of the exponential extrapolation formula, a three parameter, mixed Gaussian/exponential expression,

$$E(n) = E_{\text{CBS}} + Ae^{-(n-1)} + Be^{-(n-1)^2}, \quad (3)$$

was suggested by Peterson *et al.* in the context of a multireference configuration interaction study of the  $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$  exchange reaction.<sup>48</sup> The same expression was used by Woon and Dunning in a paper on first row/second row AB diatomics with various levels of perturbation theory and CCSD(T).<sup>49</sup> It was noted that the mixed formula fit the total energies through cc-pV5Z better than the straight exponential function, although it was not clear if Eq. (3) was being used in an exact fit to three energies or in a least squares fashion

to four energies. In the present study, the three parameters in Eqs. (2) and (3) will always be used in exact fits to three energies. While least squares fitting of lower index basis sets is possible, retaining the lower quality energies in the extrapolation has not been found to lead to improved results. All of the two parameter formulas to be discussed below were used with the two largest basis sets from any three basis set combination. Eqs. (2) and (3) are typically used with total energies, whereas many of the formulas to follow were initially intended to describe only the correlation energy. We will discuss the merits of separating the Hartree-Fock and correlation energies below.

The next two formulas involve inverse powers of the highest angular momentum value present in the basis set ( $\ell_{\max}$ ). They were motivated by the  $1/Z$  second order perturbation theory work of Schwartz, who analyzed the correlation energy in 2-electron, He-like systems.<sup>50,51</sup> In Schwartz's study the basis sets were saturated at every angular momentum level. A subsequent CI/natural orbital study of the He atom correlation energy by Carroll *et al.* showed that  $E(\ell_{\max})$  could be expanded in inverse powers of  $(\ell_{\max} + 1/2)$  beginning with the 3rd power and with energy increments,  $E(\ell_{\max}) - E(\ell_{\max}-1)$  beginning with the 4th power.<sup>52</sup> Hill<sup>53</sup> reached similar conclusions using the Rayleigh-Ritz variational method, as did Kutzelnigg and Morgan<sup>54</sup> in their partial-wave expansion covering arbitrary electronic states for 2-electron atoms. Martin<sup>55</sup> subsequently adopted this approach with the formula,

$$E(\ell_{\max}) = E_{\text{CBS}} + A/(\ell_{\max} + \frac{1}{2})^4, \quad (4)$$

where  $E(\ell_{\max})$  is the CCSD(T) correlation energy for a specific correlation consistent basis set characterized by a fixed value of  $\ell_{\max}$  and  $E_{\text{CBS}}$  is the corresponding complete basis set limit. The offset of  $1/2$  was a compromise choice between hydrogen, which would have had a value of 0, and first- and second-row elements, with a value of 1. With the correlation consistent basis set family, the basis set index,  $n$ , in Eqs. (2) and (3) is identical to  $\ell_{\max}$  for second and third period elements Li – Ar. Some ambiguity may arise if the basis set index does not correspond to a single value of  $\ell_{\max}$  for all elements in the chemical system. Examples would include simple systems such as  $\text{C}_2\text{H}_4$  and  $\text{ZnF}$ . In such cases we have chosen to use the highest  $\ell_{\max}$  value in the extrapolation formula.

Martin also proposed several other three parameter formulas. The first of these extended Eq. (4) with an additional  $1/(\ell_{\max} + 1/2)^6$  term,<sup>55</sup> and a second formula substituted the inverse fifth power for the inverse fourth power in Eq. (4).<sup>56</sup> In a 1997 paper, Martin and Taylor considered replacing the fixed 4th power in Eq. (4) with a variable exponent.<sup>56</sup> For a set of 15 small, first row molecules they found exponent values that ranged from 3.52 to 5.38, with an average value of 4.21. Although the collection of molecules was limited, the results suggest that an exponent of 4 should be close to the optimal value.

While some studies have emphasized this connection to partial-wave analyses in order to justify one extrapolation over another, the practical situation with finite basis sets differs considerably from a partial-wave expansion for 2-electron atoms. The correlation consistent basis sets are far

from  $\ell$ -space saturated even at the quadruple zeta level. In a molecular environment the wave function is also not an eigenfunction of the angular momentum operator, as is the case in atoms. It might seem a reasonable assumption that molecular correlation energies can be efficiently expanded in terms of  $\ell_{\max}$  (at least for atom-centered basis functions), however, we know of no direct theoretical justification for this conjecture. Klopper *et al.*<sup>57</sup> discussed the basis set convergence of the correlation consistent and atomic natural orbital basis sets<sup>58</sup> in terms of a “principal expansion,” by which they meant an expansion in  $n$ , the principal quantum number. Following work by Bunge<sup>59</sup> and Carroll *et al.*<sup>52</sup> on He atom, which showed that the energy contribution of a NO is proportional to  $1/n^6$ , Klopper *et al.* argued in favor of extrapolations based on inverse powers of the basis set index, beginning with the 3rd power. However, they conceded that their analysis should only be viewed as *motivating* such formulas and that it did not constitute a rigorous derivation for the cc basis sets because these sets are not NO-based. Consequently, the basis set convergence observed with real basis sets and chemical systems beyond 2-electron atoms is not well characterized. The numerous extrapolation formulas to be discussed in this work reflect both the diverse nature of chemical systems and the lack of a clear-cut favorite that satisfies all needs.

A 1999 study examined the effectiveness of Eqs. (2)–(4) for reproducing the experimental atomization energies of 57 small molecules.<sup>60</sup> When the extrapolations were limited to aug-cc-pVDZ through aug-cc-pVQZ basis sets, the mixed formula was found to produce slightly better agreement with experiment. With larger basis sets (aug-cc-pV5Z and aug-cc-pV6Z) the situation was less clear due to the limited number of examples (CCSD(T)/aug-cc-pV5Z calculations on polyatomic molecules were still very challenging at the time of the report) and the uncertainties in the experimental values. No single formula was found to work best across all molecules and basis sets.

The second formula involving the inverse power of  $\ell_{\max}$  is

$$E(\ell_{\max}) = E_{\text{CBS}} + A/\ell_{\max}^3, \quad (5)$$

and was originally presented in a paper discussing the basis set convergence of the water molecule by Helgaker *et al.*<sup>61</sup> A year later, a follow-up paper on Ne,  $\text{N}_2$ , and  $\text{H}_2\text{O}$  compared the effectiveness of Eq. (5) with quintuple and sextuple zeta quality basis sets to explicitly correlated R12 MP2, CCSD, and CCSD(T) calculations.<sup>62</sup> The authors concluded, “. . . that a two-parameter linear Schwartz extrapolation of the correlation energy of the form (2) gives excellent fits with R12 energies. . . .” In the text to follow, labels such as aVQ56Z will be used to indicate that the aug-cc-pVQZ through aug-cc-pV6Z sequence of basis sets was used with three parameter fits or the aug-cc-pV5Z and aug-cc-pV6Z sets were used for two parameters fits.

Another formula which will be considered in detail in this study is due to Schwenke

$$E_{\text{CBS}} = (E_{n+1} - E_n)F_{n+1}^{\text{C}} + E_n, \quad (6)$$

and involves parameters ( $F_{n+1}^{\text{C}}$ ) that are specific for the Hartree-Fock (HF), CCSD, and (T) pieces of the CCSD(T)

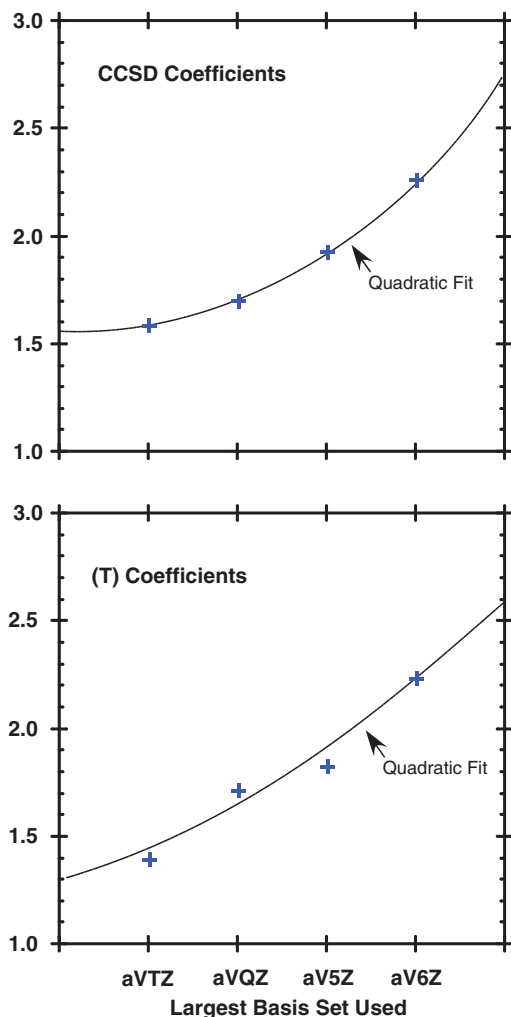


FIG. 1. Variation in the Schwenke extrapolation coefficients for the aug-cc-pVnZ basis sets.

energy. The parameters were determined by least squares fitting of reference energies for seven first row molecules (Ne, N<sub>2</sub>, CH<sub>2</sub>, H<sub>2</sub>O, CO, HF, and F<sub>2</sub>).<sup>63</sup> The reference HF and (T) energies were evaluated with very large, uncontracted “f-limit” basis sets. The details of the basis sets varied from element to element, but all involved even-tempered exponents with  $\ell_{\max} = 6$  (*i* functions) for first row elements. The CCSD piece of the reference energies were taken from the explicitly correlated CCSD-R12 energies of Klopper.<sup>6</sup> Despite the limitations of the training set, this formula and the accompanying coefficients have proven remarkably applicable beyond first row molecules. While the original parameterization was based on the cc-pVnZ and aug-cc-pVnZ sequences of basis sets, the method also works well with the aug-cc-pV(n+d)Z second row basis sets. In the present study we have applied Eq. (6) to molecules composed of even heavier elements, such as Br and I, with good results.

The original Schwenke extrapolation coefficients were only reported for basis sets up through aug-cc-pV6Z. We have extended the coefficients through the aug-cc-pV7Z basis by extrapolating their values, as seen in Figure 1. A somewhat conservative choice was made by taking the average of the aV6Z values and the aV7Z results predicted by a quadratic

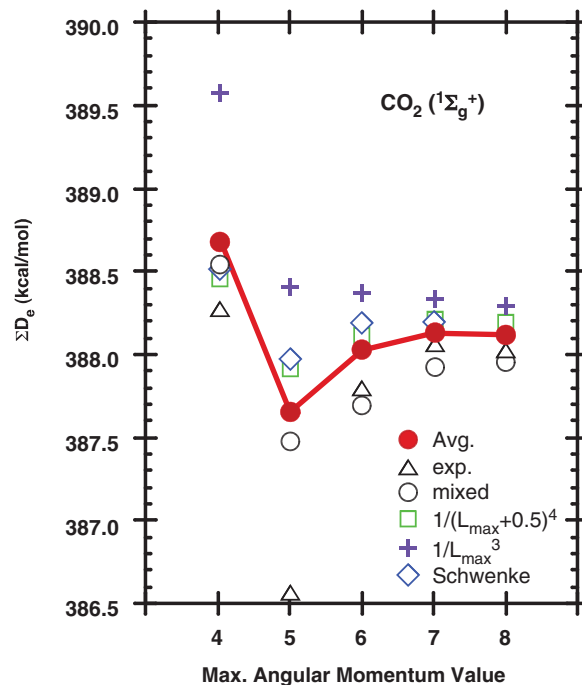


FIG. 2. Convergence of the CO<sub>2</sub> estimated complete basis set CCSD(T)(FC) atomization energies (kcal/mol) as a function of the maximum angular momentum ( $\ell_{\max}$ ) in the orbital basis set. The open shell atomic asymptotes were treated with the R/UCCSD(T) method.

extrapolation. Specifically, we used a CCSD coefficient of 2.49059 and a (T) coefficient of 2.41715. The self-consistent field (SCF) coefficient was taken from the published aV6Z value, namely, 1.119855. These choices are somewhat arbitrary but seemed reasonable in light of the very limited scope of the training set. In the case of CO<sub>2</sub>, use of the aV6Z coefficients with aV7Z energies results in an atomization energy of 388.15 kcal/mol. The coefficients we adopted yielded a value that was 0.02 kcal/mol larger, while the coefficients predicted by a quadratic fit yielded a value that was several hundredths of a kcal/mol larger yet.

Our own work has involved Eqs. (2)–(6), as well as their average. The latter choice was suggested by observations of the extrapolated atomization energy convergence patterns for many different molecules. Figure 2 shows a typical convergence pattern for the CO<sub>2</sub> molecule as a function of the underlying basis sets. Previously published CBS atomization energy convergence patterns bear a qualitative resemblance to Figure 2.<sup>16,39,64–66</sup> While the exact details of the convergence varies from molecule to molecule, the exponential and mixed formulas generally approach the best estimate of the CBS limit from below while the formulas involving the inverse powers of  $\ell_{\max}$  approach it from above. This pattern persists up through basis sets as large as aV7Z and aV8Z. The Schwenke values often follow the  $(\ell_{\max} + 1/2)^{-4}$  results, but can differ quantitatively by as much as 1 kcal/mol for larger molecules.

We have found that averaging the values from the five formulas helps to damp out some of the undesirable fluctuations in the individual extrapolations and the spread in the values can be interpreted as a crude estimate of the uncertainty in the extrapolation. The use of an average in

this situation is similar in spirit to the averaging employed by Samson and Klopper in their explicitly correlated second order Møller-Plesset perturbation theory (MP2-R12) study of C<sub>2</sub>H<sub>4</sub> and C<sub>2</sub>H<sub>6</sub>.<sup>67</sup> Of the two available MP2-R12 *Ansätze*, the /A method produced correlation energies that converged to the CBS limit from below, while the /B variant converged from above. Samson and Klopper proposed using a 60/40 weighted average as their best estimate of the true limit.

Among the many formulas that space limitations will not allow us to cover in detail is one by Martin and Lee utilizing a simple three parameter inverse power expression,

$$A(n) = A_{\text{CBS}} + BC^{-n}, \quad (7)$$

that was applied to harmonic frequencies obtained from CCSD(T) calculations using the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets.<sup>68</sup> The authors stated that an advantage of Eq. (7) is that it did not presuppose monotonic convergence and that it produced CBS frequencies accurate to about 0.1 cm<sup>-1</sup>. Huang and Lee<sup>69</sup> compared quartic force field frequencies produced by extrapolating the cc-pVTZ, cc-pVQZ, and cc-pV5Z grid energies with Martin's three-point formula<sup>55</sup> and with the  $1/\ell_{\text{max}}^3$  formula. Very recently, Huang *et al.* compared a three parameter basis set extrapolation against explicitly correlated results.<sup>70</sup> They concluded that a 5 $\zeta$  quality basis set was needed for accurate results.

Another formula that we will not be able to examine further was suggested by Gdanitz who applied it to the case of the helium dimer,

$$E_{\text{CBS}} = \left[ \left( n + \frac{3}{3} \right)^3 E_n - \left( n - 1 + \frac{3}{3} \right)^3 E_{n-1} \right] / \left[ \left( n + \frac{3}{3} \right)^3 - \left( n - 1 + \frac{3}{3} \right)^3 \right], \quad (8)$$

in a multireference averaged pair functional study published in 2000.<sup>71</sup>

Klopper considered separate extrapolations of the CCSD singlet and triplet pair correlation energies using  $1/\ell_{\text{max}}^3$  and  $1/\ell_{\text{max}}^5$  formulas, respectively.<sup>6</sup> He compared his total energies for six closed shell molecules (CH<sub>2</sub> <sup>1</sup>A<sub>1</sub>, H<sub>2</sub>O, HF, N<sub>2</sub>, CO, and F<sub>2</sub>) and the neon atom against explicitly correlated CCSD-R12/B calculations performed with a large uncontracted basis set. Valeev *et al.* found that introducing offsets of 1/2 into the singlet and triplet pair energy formulas, e.g.,  $1/(\ell_{\text{max}} + 1/2)^3$  for singlet pairs, dramatically improved results at the MP2 level.<sup>7</sup> For open shell systems, a third term associated with single excitations is required in the decomposition of the CCSD correlation energy. Martin and co-workers have adopted separate extrapolations of the Hartree-Fock, singlet and triplet pair and (T) energies in their W2.2, W3.2, and W4n methods.<sup>19,20</sup>

Although the focus in the present paper is on methods utilizing correlation consistent basis set energies of quadruple zeta or better quality, some researchers have proposed extrapolations with the much smaller cc-pVDZ and cc-pVTZ basis sets. Truhlar suggested variable exponent expressions of the form,

$$E^{\text{HF}}(n) = E_{\text{CBS}}^{\text{HF}} + A^{\text{HF}} n^{-\alpha}, \quad (9)$$

$$E^{\text{corr}}(n) = E_{\text{CBS}}^{\text{corr}} + A^{\text{corr}} n^{-\beta}, \quad (10)$$

for the HF and correlation energies, as opposed to the fixed exponent expressions in Eqs. (4) and (5).<sup>72</sup> The exponents  $\alpha$  and  $\beta$  were determined by minimizing  $\epsilon_{\text{RMS}}$  with respect to the estimated basis set limit energies for Ne, HF, and H<sub>2</sub>O using values given by Halkier *et al.*<sup>62</sup> Energy differences were not discussed and the basis set limit energies were only defined to 1–2 mE<sub>h</sub>. The RMS total energy accuracy of the CCSD(T) extrapolation was reported to be 2.4 kcal/mol, compared to a raw CCSD(T)/cc-pV6Z value of 3.1 kcal/mol. In a follow-up study a year later, the exponents were re-optimized with respect to reference atomization energies (as opposed to total energies) for 29 small molecules, the largest of which were C<sub>2</sub>H<sub>2</sub> and H<sub>2</sub>CO.<sup>73</sup> Fast *et al.*<sup>73</sup> generated their reference atomization energies from 3-point extrapolations with cc-pVTZ through cc-pV5Z basis sets. The resulting CCSD(T)  $\epsilon_{\text{MAD}}$ ,  $\epsilon_{\text{RMS}}$ , and  $\epsilon_{\text{MAX}}$  values were 1.79, 2.16, and 4.48 kcal/mol, respectively. For a smaller set of 19 molecules, the atomization energies obtained from the VDZ/VTZ extrapolation formulas of Fast *et al.* were found to be slightly closer to experiment than raw CCSD(T)(FC)/cc-pV5Z values. On the basis of experience with other calibration studies, we would expect the magnitude of the error for the VDZ/VTZ extrapolation to increase as the reference set encompasses a larger and more diverse set of molecules. To illustrate this point, we found errors with respect to our best values for n-propane and n-butane of 11.2 and 14.2 kcal/mol using R/UCCSD(T) atoms and MP2(FC)/cc-pVDZ optimized geometries. While it may be true that extrapolation leads to an improvement over the raw VTZ values, errors of this magnitude are not suitable for accurate work. Chuang and Truhlar advocated the use of Eqs. (9) and (10) with the VDZ/VTZ basis set combination for performing CBS geometry optimizations.<sup>74</sup> Their study was limited to just three molecules (H<sub>2</sub>O, NH<sub>3</sub>, and H<sub>2</sub>O<sub>2</sub>) at the MP2 and CCSD levels of theory. Halkier *et al.* criticized the VDZ/VTZ geometry extrapolation scheme of Truhlar and co-workers as being unreliable in the sense that it sometimes produced results that were worse than the raw cc-pVTZ results.<sup>75</sup> While acknowledging that the requirements for an extrapolation procedure were subject to personal tastes, these authors felt that an effective scheme should only rarely predict CBS results that were less accurate than the best raw results used in the extrapolation. In light of the very small number of molecules included in the Chuang/Truhlar study, their statistical comparison is necessarily limited. Nonetheless, using aug-cc-pV5Z and R12 results as their reference standard, Halkier *et al.* found that the Truhlar CBS scheme reduced the VTZ mean signed deviation from 0.0020 Å to -0.0011 Å, but  $\epsilon_{\text{MAD}}$  actually increased from 0.0020 Å to 0.0028 Å and  $\epsilon_{\text{RMS}}$  increased slightly from 0.0026 to 0.0029 Å.

Martin and Oliveira adopted Eq. (10) for estimating the CCSD/CBS limit in their W1 model chemistry. The empirical parameter  $\beta$  was chosen to minimize the mean absolute deviation with respect to 28 experimental atomization energies. Unlike Truhlar, they extrapolated from the much larger aVTZ+2d1f/aVQZ+2d1f combination of basis sets. The notation “+2d1f” indicates that two tight *d*'s and one tight *f* primitive were added to the regular correlation consistent sets in order to account for inner shell polarization effects.

At the low end of the basis set sequence, Schwenke's formulas also support a VDZ/VTZ extrapolation.<sup>63</sup> In our experience, the accuracy of the results is not comparable to the accuracy achievable with extrapolations based on aug-cc-pVQZ or better basis sets.<sup>64</sup> The Schwenke VDZ/VTZ scheme also often fails to reach the goal of "chemical accuracy."

Varandas discussed a CBS extrapolation expression of the form,

$$E^{\text{corr}}(n) = E_{\text{CBS}}^{\text{corr}} [1 + A_3(1 + A_4/n)/n^3], \quad (11)$$

where  $A_4 = a \exp(bA_3) + c$  and the three parameters ( $a$ ,  $b$ ,  $c$ ) were either globally fit to reference data on 20 molecules and one atom at the MP2, CCSD, and CCSD(T) levels of theory or to individual levels of theory using basis sets up through cc-pV6Z.<sup>76</sup> The single atom was Ne for which the MP2/cc-pVnZ correlation energies were arbitrarily given a weight of 5. The reference values were taken from the work of Klopper *et al.*<sup>57</sup> and Halkier *et al.*<sup>62</sup> The nonlinear system of equations associated with Eq. (11) must be solved iteratively. Varandas' work dealt strictly with the correlation component of the energy. Total energies or energy differences were not reported. Compared to estimates of the CCSD(T)/CBS correlation energies obtained from  $1/\ell_{\text{max}}^3$  and the V5Z/V6Z basis set combination, the formula of Varandas led to maximum errors of  $-4.8$  and  $-4.4$  mE<sub>h</sub> using (D,T) and (T,Q) combinations.

Varandas also proposed a formula to describe the basis set limit for multireference singles and doubles configuration interaction with the Davidson correction (MRSD-CI+Q),

$$E^{\text{corr}}(n) = E_{\text{CBS}}^{\text{corr}} + A_3/(n - \frac{3}{8})^3 [1 + \tau_{53}/(n - \frac{3}{8})^2], \quad (12)$$

where  $E^{\text{corr}}$  is the dynamic correlation energy component.<sup>76</sup> The non-dynamical correlation energy, which corresponds to the complete active space SCF (CASSCF) energy, was extrapolated with a four-point formula. The parameters  $A_3$  and  $\tau_{53}$  were fit to data for 23 atomic and molecular systems. As in the earlier study, only total energies were reported.

Huh and Lee focused on the CBS limits for hydrogen-bonded and van der Waals complexes.<sup>77</sup> They pointed out that the  $1/\ell_{\text{max}}^3$  formula underestimated the CCSD(T) correlation energy of 9 small reference systems by as much as 18.2 mE<sub>h</sub> with the aVDZ/aVTZ combination and *overshot* it by as much as  $-1.1$  mE<sub>h</sub> with the largest combinations they considered (aV5Z/aV6Z). Dropping the basis set combination from (aV5Z/aV6Z) to (aVQZ/aV5Z) caused the errors to double. Errors at the MP2 level of theory were even larger, hitting a maximum of 30.5 mE<sub>h</sub>. Based on their analysis, they proposed the expression,

$$E^{\text{corr}}(n) = E_{\text{CBS}}^{\text{corr}} + 1/(n + \gamma)^3, \quad (13)$$

for aVnZ/aV(n + 1)Z combinations where  $\gamma$  is specific for each combination and level of theory. In the case of CCSD(T),  $\gamma$  takes on values of 1/2 (aVDZ/aVTZ),  $-1/4$  (aVTZ/aVQZ), and  $-1/2$  (aVQZ/aV5Z).

A more recent paper by Bakowies advocates replacing the fixed exponent in Eq. (5) with an adjustable, noninteger parameter ( $\beta$ ),

$$E(\ell_{\text{max}}) = E_{\text{CBS}} + A/\ell_{\text{max}}^\beta, \quad (14)$$

$\beta$  is chosen to minimize the RMS deviation with respect to two types of reference systems.<sup>78</sup> One set consisted of 105 closed-shell, neutral first row molecules for which Bakowies ran MP2 and CCSD calculations with cc-pVnZ,  $n = \text{D, T, ... 6}$ . CBS limits were not estimated for these molecules. Limits for another 12 molecules were taken from the R12 work of Klopper and co-workers.<sup>79</sup> As pointed out by Hill *et al.*<sup>80</sup> and even earlier by Schwenke,<sup>63</sup> for particular values of the  $F$  coefficients in Schwenke's expression, see Eq. (6), it yields the same results as Eq. (14). Not surprisingly, the statistics in Table VIII of the Bakowies article were quite close to the results obtained from the Schwenke formulas in terms of predicting CCSD basis set limit correlation energies. Atomization energies were also reported, but the author felt that the very limited CBS reference data (7 molecules) was not representative enough or accurate enough to allow reliable exponents to be derived for (Q5) or (56) extrapolations.

In addition to the CCSD(T)-based approaches already discussed, the literature contains numerous "model chemistries," such as the Gn methods,<sup>81-89</sup> which incorporate empirical parameters in order to improve agreement with experiment. In a similar vein, Petersson and co-workers proposed a variety of CBS methods that likewise involved empirical corrections and attempt to estimate the one-particle and n-particle limits through extrapolations.<sup>90-92</sup> As an illustration of the capabilities of one of these models, the ROCBS-QB3 method was reported to produce an  $\epsilon_{\text{MAD}}$  of 0.91 kcal/mol for the heats of formation in the G2/97 test set.<sup>93</sup> This test set consists of 148 small molecules composed of first and second row elements. In previous studies, we have compared our coupled cluster approach to methods like the CBS-Q method of Petersson and co-workers.<sup>29,35,39,94-96</sup> Across a collection of more than 100 small molecules, we found an  $\epsilon_{\text{MAD}}(\text{CBS-Q})$  of 1.3 kcal/mol for the heats of formation. Klopper noted that the CBS-Atomic Pair Natural Orbital (APNO) level of theory,<sup>97</sup> another of the methods developed by Petersson and co-workers, produced MP2 correlation energies roughly on par with raw cc-pV6Z calculations.<sup>6</sup> For example, Klopper reported a Ne atom MP2 correlation energy of  $-0.3200$  E<sub>h</sub> obtained from an explicitly correlated calculation, compared to  $-0.3118$  E<sub>h</sub> for raw MP2/cc-pV6Z and  $-0.3097$  E<sub>h</sub> for CBS-APNO.

### III. APPROACH

Before discussing the details of our approach, we note that practical limitations preclude a brute force strategy for the creation of a broad collection of CBS reference atomization energies unambiguously accurate to  $\leq 0.1$  kcal/mol. A "broad collection" would ideally include a sufficient number of molecules to represent many of the myriad bonding situations observed in small chemical systems containing elements from throughout the periodic table. Even in the case of a simple first row diatomic molecule such as C<sub>2</sub> (<sup>1</sup>Σ<sub>g</sub><sup>+</sup>), where correlation consistent basis sets as large as aug-cc-pV8Z are available, it has not been possible to achieve 0.1 kcal/mol convergence in D<sub>e</sub> via direct computation.<sup>98</sup> Explicitly correlated methods provide much faster basis set convergence, but up through cc-pVQZ-F12 are still not capable of achieving 0.1 kcal/mol

TABLE I. List of molecules and corresponding largest basis sets.<sup>a</sup>

H2 ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) <sup>b</sup> aV8Z	OFO ( <sup>2</sup> B <sub>2</sub> ) aV8Z	CS <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) <sup>b</sup> aV8Z	FCN ( <sup>1</sup> Σ <sup>+</sup> ) aV7Z
CH ( <sup>2</sup> Π) aV8Z	CCl ( <sup>2</sup> Π) aV8Z	S2 ( <sup>3</sup> Σ <sub>g</sub> <sup>-</sup> ) <sup>b</sup> aV8Z	FNC ( <sup>1</sup> Σ <sup>+</sup> ) aV7Z
C2 ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) <sup>b</sup> aV8Z	CINO ( <sup>1</sup> A') aV7Z	Cl <sub>2</sub> O ( <sup>1</sup> A <sub>1</sub> ) aV7Z	FS ( <sup>2</sup> Π) aV7Z
O2 ( <sup>3</sup> Σ <sub>g</sub> <sup>-</sup> ) aV8Z	PS ( <sup>2</sup> Π) aV7Z	SO ( <sup>3</sup> Σ <sub>g</sub> <sup>-</sup> ) <sup>b</sup> aV8Z	CIF ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV8Z
H2O ( <sup>1</sup> A <sub>1</sub> ) aV8Z	B <sub>2</sub> ( <sup>3</sup> Σ <sub>g</sub> <sup>-</sup> ) <sup>b</sup> aV7Z	SO <sub>2</sub> ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV7Z	CHFO ( <sup>1</sup> A') aV7Z
CH <sub>2</sub> ( <sup>3</sup> B <sub>1</sub> ) aV8Z	BH <sub>3</sub> ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV7Z	SO <sub>3</sub> ( <sup>1</sup> A <sub>1</sub> ) aV7Z	HS ( <sup>2</sup> Π) <sup>b</sup> aV8Z
CH <sub>2</sub> ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV8Z	CH <sub>3</sub> ( <sup>2</sup> A <sub>2</sub> '') aV8Z	P2 ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV7Z	Br <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV6Z
CO ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV8Z	CH <sub>4</sub> ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV7Z	PN ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV7Z	HBr ( <sup>1</sup> Σ <sup>+</sup> ) aV6Z
C <sub>2</sub> H <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV8Z	NH ( <sup>3</sup> Σ <sup>-</sup> ) aV8Z	PO ( <sup>2</sup> Π) <sup>b</sup> aV7Z	CBr ( <sup>2</sup> Π) aV6Z
N <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV8Z	NH <sub>2</sub> ( <sup>2</sup> B <sub>1</sub> ) aV8Z	PH ( <sup>3</sup> Σ <sup>-</sup> ) aV7Z	BrO ( <sup>2</sup> Π) aV6Z
CN ( <sup>2</sup> Σ <sup>+</sup> ) <sup>b</sup> aV8Z	NH <sub>3</sub> ( <sup>1</sup> A <sub>1</sub> ) aV7Z	PH <sub>2</sub> ( <sup>2</sup> B <sub>1</sub> ) aV7Z	HO <sub>2</sub> ( <sup>2</sup> A'') aV7Z
CO <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) <sup>b</sup> aV8Z	C <sub>2</sub> H ( <sup>2</sup> Σ) aV8Z	PH <sub>3</sub> ( <sup>1</sup> A <sub>1</sub> ) aV7Z	AlN ( <sup>3</sup> Π) <sup>b</sup> aV6Z
NO ( <sup>2</sup> Π) aV8Z	HCO ( <sup>2</sup> A') aV7Z	CCl <sub>2</sub> ( <sup>1</sup> A <sub>1</sub> ) aV7Z	Si <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV7Z
HNO ( <sup>1</sup> A') aV8Z	H <sub>2</sub> CO ( <sup>1</sup> A <sub>1</sub> ) aV7Z	CCl <sub>3</sub> ( <sup>2</sup> A <sub>1</sub> ) aV7Z	SiO ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV6Z
F <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) <sup>b</sup> aV8Z	CF <sub>2</sub> O ( <sup>1</sup> A <sub>1</sub> ) aV7Z	BP ( <sup>3</sup> Π) aV7Z	AlH ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV6Z
CF ( <sup>2</sup> Π) aV8Z	CF ( <sup>4</sup> Σ <sup>-</sup> ) aV8Z	PF ( <sup>3</sup> Σ <sup>-</sup> ) aV7Z	BN ( <sup>3</sup> Π) <sup>b</sup> aV7Z
CF <sub>2</sub> ( <sup>1</sup> A <sub>1</sub> ) aV8Z	NO <sub>2</sub> ( <sup>2</sup> A <sub>1</sub> ) aV8Z	BCl ( <sup>1</sup> Σ <sup>+</sup> ) aV7Z	H <sub>2</sub> O <sub>2</sub> ( <sup>1</sup> A') <sup>b</sup> aV7Z
HF ( <sup>1</sup> Σ <sup>+</sup> ) aV8Z	O <sub>3</sub> ( <sup>1</sup> A <sub>1</sub> ) aV8Z	H <sub>2</sub> S ( <sup>1</sup> A <sub>1</sub> ) aV8Z	SiH ( <sup>2</sup> Π) aV7Z
FO ( <sup>2</sup> Π) aV8Z	F <sub>2</sub> O ( <sup>1</sup> A <sub>1</sub> ) aV7Z	Al <sub>2</sub> ( <sup>3</sup> Π <sub>u</sub> ) <sup>b</sup> aV7Z	SiH <sub>2</sub> ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV6Z
NF ( <sup>3</sup> Σ <sup>-</sup> ) aV8Z	FOO ( <sup>2</sup> A'') aV7Z	ClO <sub>2</sub> ( <sup>2</sup> B <sub>1</sub> ) aV7Z	SiH <sub>4</sub> ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV7Z
OH ( <sup>2</sup> Π) aV8Z	SiF ( <sup>2</sup> Π) aV7Z	AlF ( <sup>1</sup> Σ <sup>+</sup> ) aV7Z	C <sub>2</sub> H <sub>2</sub> F <sub>2</sub> 1,1-difluo aV6Z
C <sub>2</sub> H <sub>4</sub> ( <sup>1</sup> A <sub>g</sub> ) <sup>b</sup> aV8Z	FOOF ( <sup>1</sup> A) aV7Z	C <sub>3</sub> H <sub>4</sub> cyclopropene aV7Z	C <sub>2</sub> H <sub>2</sub> F <sub>2</sub> trans-1,2- aV6Z
Cl <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) <sup>b</sup> aV8Z	HCN ( <sup>1</sup> Σ <sup>+</sup> ) aV8Z	C <sub>2</sub> F <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV7Z	C <sub>2</sub> H <sub>2</sub> F <sub>2</sub> cis-1,2- aV6Z
HCl ( <sup>1</sup> Σ <sup>+</sup> ) aV8Z	HNC ( <sup>1</sup> Σ <sup>+</sup> ) aV8Z	N <sub>2</sub> O ( <sup>1</sup> Σ <sup>+</sup> ) aV7Z	SiS ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV6Z
ClO ( <sup>2</sup> Π) aV8Z	CS ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV8Z	CH ( <sup>4</sup> Σ <sup>-</sup> ) aV8Z	OCS ( <sup>1</sup> Σ <sup>+</sup> ) aV7Z
C <sub>3</sub> H <sub>6</sub> cyclop aV6Z	CH <sub>2</sub> Cl <sub>2</sub> aV6Z	C <sub>4</sub> H <sub>6</sub> trans-1,3-buta aV6Z	LiO ( <sup>2</sup> Π) aV6Z
CICN ( <sup>1</sup> Σ <sup>+</sup> ) <sup>b</sup> aV6Z	H <sub>2</sub> SiO ( <sup>1</sup> A <sub>1</sub> ) aV6Z	BrCl ( <sup>1</sup> Σ <sup>+</sup> ) aV6Z	I <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV6Z
HOCl ( <sup>1</sup> A') aV6Z	CH <sub>3</sub> Cl ( <sup>1</sup> A <sub>1</sub> ) aV6Z	Li <sub>2</sub> ( <sup>1</sup> Σ <sub>g</sub> <sup>+</sup> ) aV6Z	AlF <sub>3</sub> ( <sup>1</sup> A <sub>1</sub> ) aV6Z
CF <sub>3</sub> ( <sup>2</sup> A <sub>1</sub> ) aV6Z	LiCl ( <sup>1</sup> Σ <sup>+</sup> ) aV6Z	LiF ( <sup>1</sup> Σ <sup>+</sup> ) aV6Z	C <sub>3</sub> H <sub>4</sub> allene aV6Z
CF <sub>4</sub> ( <sup>1</sup> A <sub>1</sub> ) aV6Z	LiH ( <sup>1</sup> Σ <sup>+</sup> ) aV6Z	LiN ( <sup>3</sup> Σ <sup>-</sup> ) aV6Z	C <sub>2</sub> F <sub>4</sub> ( <sup>1</sup> A <sub>g</sub> ) aV6Z
CH <sub>3</sub> OH aV6Z	HNCO ( <sup>1</sup> A') aV7Z	CH <sub>3</sub> CHO acetald aV6Z	HOCN ( <sup>1</sup> A') aV6Z
CH <sub>3</sub> F ( <sup>1</sup> A <sub>1</sub> ) aV6Z	H <sub>2</sub> CS ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV7Z	AIP ( <sup>3</sup> Σ <sup>-</sup> ) <sup>b</sup> aV7Z	H <sub>2</sub> S <sub>2</sub> ( <sup>1</sup> A') <sup>b</sup> aV7Z
NS ( <sup>2</sup> Π) aV8Z	ZnP ( <sup>4</sup> Σ <sup>-</sup> ) aV6Z	ZnO ( <sup>1</sup> Σ <sup>+</sup> ) aV6Z	CH <sub>3</sub> NH <sub>2</sub> ( <sup>1</sup> A') aV6Z
C <sub>2</sub> H <sub>6</sub> ( <sup>1</sup> A <sub>g</sub> ) aV6Z	N <sub>2</sub> H <sub>2</sub> trans aV7Z	N <sub>2</sub> H <sub>2</sub> cis aV7Z	N <sub>2</sub> H <sub>2</sub> iso aV7Z
N <sub>2</sub> H <sub>4</sub> ( <sup>1</sup> A) aV6Z	NH <sub>2</sub> OH ( <sup>1</sup> A') aV7Z	SiF <sub>4</sub> ( <sup>1</sup> A <sub>1</sub> ) <sup>b</sup> aV7Z	NCO ( <sup>2</sup> Π) aV7Z
C <sub>2</sub> HF ( <sup>1</sup> Σ) aV7Z			

<sup>a</sup>For second row elements (Na Cl) aVDZ = aug-cc-pV(D + d)Z, aVTZ = aug-cc-pV(T + d)Z, etc.

<sup>b</sup>Member of the F12 reference set.

accuracy.<sup>80,98</sup> Clearly, if this level of accuracy is desired in a reference set obtained from standard CCSD(T), some type of extrapolation seems unavoidable. Regrettably, this requirement inherently introduces an element of uncertainty. Despite the difficulties, we propose to create a diverse reference set of ground state closed and open shell molecules with elements as heavy as iodine (see Table I). Our ultimate goal is to calibrate the effectiveness of selected CBS extrapolation formulas intended for use with standard CCSD(T). In order to avoid the complications arising from reference values derived from experimental data, we have adopted a two-pronged approach which relies on very large basis set standard CCSD(T) atomization energies combined with extrapolations and very large explicitly correlated CCSD(T)-F12b (Refs. 9–11) calculations that do not involve extrapolation.

### A. Reference values from standard CCSD(T)

Most high accuracy thermochemical studies in the recent literature employ the correlation consistent basis sets of Dunning, Peterson, and co-workers.<sup>44,45,60,99–105</sup> Due to the un-

derlying regularity inherent to these basis sets, the resulting total energies and atomization energies smoothly converge to the CBS limit. It is this critical behavior that enables simple extrapolation formulas to be effective. The diffuse function augmented sequence of basis sets is conventionally denoted aug-cc-pVnZ,  $n = D, T, Q, 5-8$ . For second row elements, an additional tight  $d$  function is known to be important for accelerating convergence to the basis set limit. The latter are referred to as the aug-cc-pV( $n+d$ )Z sequence of basis sets.<sup>106</sup> For the sake of brevity, we will use the notation aVDZ, aVTZ, etc. to represent collectively both the aug-cc-pVnZ and aug-cc-pV( $n+d$ )Z sequences of basis sets. Several new aV6Z, aV7Z, and aV8Z basis sets were developed for this work.

Unless otherwise noted, all calculations discussed in this paper utilized the frozen core approximation. CCSD(T) calculations that did not involve  $k$  ( $\ell = 7$ ) or  $l$  ( $\ell = 8$ ) functions were performed with MOLPRO 2010.1.<sup>107</sup> Calculations with  $k$  and  $l$  functions were performed with DALTON 2.0.<sup>108</sup> Most of the aV7Z and aV8Z results reported here did not explicitly include  $k$  and  $l$  functions due to excessively long run times with DALTON. Instead, the small contributions from the



TABLE II. Raw CCSD(T)(FC) atomization energies errors (kcal/mol).<sup>a</sup>

Basis set	Items	$\epsilon_{\text{MAD}}$	$\epsilon_{\text{STDEV}}$	$\epsilon_{\text{Max}}$
aug-cc-pVQZ	141	2.34	2.69	-8.08
aug-cc-pV5Z	141	1.05	1.22	-3.69
aug-cc-pV6Z	141	0.52	0.61	-1.86
aug-cc-pV7Z	99	0.27	0.31	-0.58
aug-cc-pV8Z	46	0.12	0.15	-0.33

<sup>a</sup>The errors are measured with respect the best available average estimate of the CBS limits. MAD = mean absolute deviation. STDEV = standard deviation. Max. = maximum error (negative sign indicates the raw value is smaller than the best estimate of the CBS limit).

missing functions were estimated via a short extrapolation as discussed elsewhere.<sup>64</sup> The accuracy of this approximation for atomization energies was tested on 7 systems ( $\text{C}_2$ ,  $\text{N}_2$ ,  $\text{F}_2$ ,  $\text{C}_2\text{H}_2$ ,  $\text{CH}_4$ ,  $\text{ClF}$ , and  $\text{Cl}_2$ ) and was found to result in errors of  $\leq 0.02$  kcal/mol relative to the exact values. The magnitude of this error may grow with the size of the molecule, but we are currently unable to calibrate that trend. For the molecules in the standard method reference set, we expect the error associated with this approximation to be small compared to the error associated with extrapolating to the basis set limit.

Our collective experience gained from studies on many different chemical species indicates that CBS extrapolations using a basis set sequence ending with  $\text{aV}n\text{Z}$  generally provides results comparable to or better than raw results with an  $\text{aV}(n+2)\text{Z}$  basis set. Since the smallest basis set used to produce the current reference atomization values is  $\text{aV}6\text{Z}$ , we expect the least accurate of our reference values to be of  $\text{aV}8\text{Z}$  quality or better. As seen in Table II, where various statistical measures of quality are shown for the raw (i.e., non-extrapolated) atomization energies as a function of basis set, the  $\text{aV}6\text{Z}$ -based reference energies should equate to a mean absolute deviation (or root mean square deviation) of  $\sim 0.1$  kcal/mol, assuming a  $\text{CBS}(n) \approx \text{Raw}(n+2)$  correspondence. All three metrics shown in Table II decrease by approximately a factor of 2 for every increment in the basis set index  $n$ . To the extent the correspondence between extrapolated and raw values continues to hold for the larger basis sets, we would expect the  $\text{aV}6\text{Z}$ -based reference values to be comparable in quality to raw  $\text{aV}10\text{Z}$  results, with  $\epsilon_{\text{MAD}}$  and  $\epsilon_{\text{RMS}}$  on the order of 0.02 to 0.03 kcal/mol and  $\epsilon_{\text{MAX}} \sim 0.06$  kcal/mol. Because the majority of reference values were based on  $\text{aV}5\text{Z}$  or  $\text{aV}6\text{Z}$  extrapolations, we conservatively estimate the overall uncertainty in the standard method reference set to be  $\pm 0.05$  kcal/mol. Therefore, statistical differences associated with this reference set of less than  $\sim 0.05$  kcal/mol are not likely to be significant.

Our reference set includes both ground and excited electronic states, open and closed shell species, and selected stationary points on the potential energy surface. Although it is fairly large and diverse, it is nonetheless weighted towards small systems and elements from the first few rows of the periodic table. As a result, the raw statistical metrics in Table II would likely increase somewhat as bigger chemical systems with larger correlation energies are added. For instance,  $\epsilon_{\text{MAX}}$  for the  $\text{aV}6\text{Z}$  basis set is given as  $-8.08$  kcal/mol, but if it had been possible to include a system such as octane ( $\text{C}_8\text{H}_{18}$ ) in

the reference set, the maximum error would have been  $\sim -11$  kcal/mol. Errors of this magnitude emphasize the importance of extrapolation methods that partially alleviate the basis set truncation problem.

As stated previously, some type of CBS extrapolation is essential for achieving the highest possible level of accuracy in our standard CCSD(T) reference values. For this purpose, we will use the average of the five formulas given in Eqs. (2)–(6). The sensitivity of the statistical analysis to that choice will be probed by performing a parallel analysis with the  $1/(\ell_{\text{max}}+1/2)^4$  formula, which in our experience also performs well across many different types of molecules.

Energies for the standard CCSD(T) reference values were evaluated at the optimal geometries for every basis set level, as is our customary practice. When coupled with CBS extrapolation, this provides atomization energy estimates at the basis set limit geometry. The use of optimized geometries produces slightly larger atomization energies than would be found if fixed geometries were used, especially if the fixed geometries were obtained at a low level of theory and differed significantly from the CBS limit structures. This procedure differs from the one used in creating the CCSD(T)-F12b reference values, where all energies were evaluated at CCSD(T)(CV)/cc-pCVQZ optimized geometries.

In the present work, open shell CCSD(T) energies were based on the R/UCCSD(T) method, which begins with restricted open-shell Hartree-Fock orbitals, but allows a small amount of spin contamination in the solution of the CCSD equations.<sup>109–112</sup> Full atomic symmetry was imposed in calculations on the atomic asymptotes. This matches the procedure normally followed in our thermochemical studies. With the CCSD(T)-F12 reference set, it was not possible to impose full atomic symmetry on the atoms. For small molecules, such as the ones considered in this study, the differences in the magnitude of the atomization energies are small. However, all comparisons will be made using similarly treated atoms, i.e., with the standard reference set the atoms will be symmetry equivalenced and with the F12b set they are not equivalenced.

## B. Reference values from explicitly correlated CCSD(T)-F12b

In addition to the large reference group of molecules treated with the standard CCSD(T) method, atomization energies were also compiled for a second set of 51 representative first and second row molecules with the explicitly correlated CCSD(T)-F12b method.<sup>8,11</sup> For this purpose we used the very large orbital basis sets recently reported by Hill *et al.*<sup>80</sup> and used in a subsequent calibration study of hydrocarbons.<sup>98</sup> It constitutes a subset of the list of molecules in Table I. Careful attention was paid to the F12 orbital basis sets with the goal of approximating the CBS limit as closely as possible. The  $(s, p)$  portion of the basis sets was taken from the correlation consistent  $\text{aV}6\text{Z}$  basis set, which includes contractions of  $(17s, 11p) \rightarrow [8s, 7p]$  (B–F) and  $(22s, 15p) \rightarrow [9s, 8p]$  (Al–Cl). Higher angular momentum even-tempered Gaussian primitives, with exponents optimized in explicitly correlated MP2 calculations on the atoms, were added in an attempt to saturate each l space. The final sets consisted of  $(9d, 9f, 8g,$

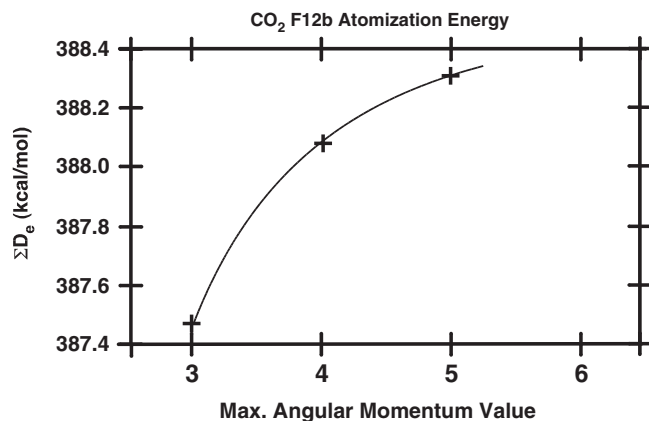


FIG. 3. Convergence of the  $\text{CO}_2$  CCSD(T)-F12b atomization energy (kcal/mol) as a function of the maximum angular momentum ( $\ell_{\max}$ ) in the orbital basis set ( $\ell_{\max}$ ).

7h) (B–F) and (12d, 10f, 8g, 6h) (A1–C1). Figure 3 shows the convergence of the  $\text{CO}_2$  atomization energy as a function of  $\ell_{\max}$  with the CCSD(T)-F12b method. The basis sets just described were used to estimate the CCSD component of the CCSD(T) energy. The noniterative triples portion of the energy was obtained from standard CCSD(T) calculations carried out with basis sets further augmented with a group of four i-type functions. Software limitations, which are discussed below, prevented the inclusion of i-type functions in the F12 orbital basis sets.

The CCSD(T)-F12b method is more computationally complex than standard CCSD(T). While much of the complexity can be hidden from the user by reasonable program defaults, it cannot be completely ignored in studies that push the theory to its limits. Besides the normal orbital basis set, the F12 method requires a complementary auxiliary basis set (CABS) for the resolution of the identity step and two density fitting (DF) basis sets for the Fock Coulomb and exchange matrices, as well as for other 2-electron integrals. Details of the DF auxiliary and MP2 basis sets, used for all integrals needed in F12 except the coulomb and exchange matrices, are given in the work of Hill *et al.*<sup>80</sup> Inclusion of i-type functions in the F12 orbital basis set would have required the presence of k-type functions in the auxiliary basis sets. MOLPRO does not currently support k-type functions. Studies of DF-MP2 theory indicate that truncating the angular momentum space of the auxiliary basis set can lead to overestimation of the correlation energy.<sup>113</sup> Unlike the reference values derived from standard CCSD(T), the F12b reference values do not involve any extrapolation. Reference Hartree-Fock energies for this portion of our study included a CABS-singles correction.<sup>8,10</sup> As previously mentioned, all energies were evaluated at CCSD(T)(CV)/cc-pCVQZ optimized geometries.

For the present study the 3C(FIX) ansatz was used in combination with a geminal exponent ( $\beta$ ) of 1.4 in order to maintain consistency with the reference data previously reported by Hill and Peterson.<sup>80</sup> We chose not to use energy-optimized  $\beta$  values since as discussed previously<sup>80</sup> this leads to values that are much larger than those obtained at the MP2-

F12 level of theory, and these were deemed unphysical. This was confirmed in the present work for both  $\text{CO}_2$  and  $\text{O}_3$ , where the optimal CCSD-F12b  $\beta$  values were found to be close to 2.4. A value this large resulted in correlation energies 0.4–0.5  $mE_h$  lower than were obtained with  $\beta = 1.4$ . The impacts on the atomization energy, however, were negligible ( $\sim 0.03$  kcal/mol) in both cases. Werner *et al.* recently warned about the undesirability of optimizing geminal exponents in CCSD(T)-F12b due to the inclusion of an approximate Lagrangian that is used to compute the energy.<sup>114</sup>

The uncertainty in the reference values derived from standard CCSD(T) have already been discussed. We now turn our attention to estimating the uncertainty in the F12 reference set of atomization energies. Due to the increased complexity of the F12 method, it presents additional challenges when trying to gauge the accuracy. The absolute accuracy of the F12b CCSD total energies for a set of 7 small molecules was estimated by Hill *et al.* to be on the order of 0.2 to 0.3  $mE_h$  (0.13 to 0.19 kcal/mol) relative to the CBS limit.<sup>80</sup> If the errors in the atomic asymptotes were negligibly small, we could therefore expect the F12 atomization energies to possess a similar accuracy. However, it seems likely that some cancellation of error occurs. The differences between F12a and F12b correlation energies may provide some clue about the absolute accuracy of the F12b reference energies. In the limit of a complete orbital basis, the two methods are expected to yield the same energy. Hill *et al.* noted that despite the use of very large reference basis sets they found energy differences of more than 1  $mE_h$ . For example, for  $\text{CO}_2$  the F12a–F12b difference is 2.3  $mE_h$ . It is unknown whether the true basis set limit lies between the F12a and F12b values. While this may provide some insight on the absolute accuracy, it tells us little about any cancellation of error.

Although the focus in the present work is on energy differences, we briefly examined the accuracy of the F12b correlation energies for a few closed shell chemical systems. A comparison of reference F12b and large basis set CCSD(T) values is presented in Table III. The CCSD/CBS estimates were based on separate extrapolations of the singlet and triplet pair energies using  $1/\ell_{\max}^3$  and  $1/\ell_{\max}^5$  formulas, respectively. The (T)/CBS component was handled with a  $1/\ell_{\max}^3$  extrapolation. All standard method energies explicitly included the contributions of k-type and l-type basis functions. For the four diatomic molecules in Table III, the F12b method consistently produces larger CCSD correlation energies than those obtained from extrapolations of the standard method and smaller (T) energies. Thus, differences between standard CCSD(T) and CCSD(T)-F12b partially cancel. This behavior has been observed in several other systems that are not included in the table for the sake of brevity. In the worse cases, the differences in the CCSD correlation energies amounted to 0.40 to 0.55  $mE_h$ . Due to the difficulty of achieving unequivocal accuracy in the correlation energies to better than 0.2  $mE_h$ , it is difficult to draw unassailable conclusions about which values are more accurate. In the case of the Ne ( $^1S$ ) atom, where Barnes *et al.* reported a CCSD correlation energy of  $-315.64 \pm 0.02$   $mE_h$ ,<sup>115</sup> the V7Z/V8Z extrapolation produces a value of  $-315.72$   $mE_h$ , a difference of  $-0.08$   $mE_h$ , while the F12b reference value<sup>80</sup> is  $-315.60$   $mE_h$ , a difference of

TABLE III. Comparison of F12b reference correlation energies and standard CCSD(T) values.<sup>a</sup>

$C_2 (^1\Sigma_g^+)$					
Basis sets <sup>b</sup>	Singlet pair Computed <sup>c</sup>	$E_{\text{corr.}}^{\text{CCSD}}$ $1/\ell_{\text{max}}^3$	Triplet pair Computed <sup>c</sup>	$E_{\text{corr.}}^{\text{CCSD}}$ $1/\ell_{\text{max}}^5$	CCSD/CBS $E_{\text{corr.}}$
aV7Z, aV8Z	-0.283175	-0.283742	-0.083551	-0.083581	-0.367323
F12b ref. value					-0.367438
Basis sets <sup>b</sup>	(T) Computed <sup>c</sup>	(T)/CBS $1/\ell_{\text{max}}^3$			
aV7Z, aV8Z	-0.036048	-0.036113			
F12b ref. value		-0.036020			
$N_2 (^1\Sigma_g^+)$					
Basis sets <sup>b</sup>	Singlet pair Computed <sup>c</sup>	$E_{\text{corr.}}$ $1/\ell_{\text{max}}^3$	Triplet pair Computed <sup>c</sup>	$E_{\text{corr.}}$ $1/\ell_{\text{max}}^5$	CCSD/CBS Computed <sup>c</sup>
aV7Z, aV8Z	-0.280741	-0.281892	-0.125508	-0.125559	-0.407451
F12b ref. value <sup>d</sup>					-0.40750
Basis sets <sup>b</sup>	(T) Computed <sup>c</sup>	(T)/CBS $1/\ell_{\text{max}}^3$			
aV7Z, aV8Z	-0.021256	-0.021335			
F12b ref. value <sup>d</sup>		-0.02119			
$F_2 (^1\Sigma_g^+)$					
Basis sets <sup>b</sup>	Singlet pair Computed <sup>c</sup>	$E_{\text{corr.}}$ $1/\ell_{\text{max}}^3$	Triplet pair Computed <sup>c</sup>	$E_{\text{corr.}}$ $1/\ell_{\text{max}}^5$	CCSD/CBS Computed <sup>c</sup>
aV7Z, aV8Z	-0.411239	-0.413956	-0.186800	-0.186914	-0.600870
F12b ref. value <sup>d</sup>					-0.60127
Basis sets <sup>b</sup>	(T) Computed <sup>c</sup>	(T)/CBS $1/\ell_{\text{max}}^3$			
aV7Z, aV8Z	-0.022752	-0.022889			
F12b ref. value <sup>d</sup>		-0.02274			
$Cl_2 (^1\Sigma_g^+)$					
Basis sets <sup>b</sup>	Singlet pair Computed <sup>c</sup>	$E_{\text{corr.}}$ $1/\ell_{\text{max}}^3$	Triplet pair Computed <sup>c</sup>	$E_{\text{corr.}}$ $1/\ell_{\text{max}}^5$	CCSD/CBS Computed <sup>c</sup>
aV7Z, aV8Z	-0.327905	-0.330031	-0.142795	-0.142821	-0.472852
F12b ref. value					-0.473397
Basis sets <sup>b</sup>	(T) Computed <sup>c</sup>	(T)/CBS $1/\ell_{\text{max}}^3$			
aV7Z, aV8Z	-0.023840	-0.024011			
F12b ref. value		-0.023708			

<sup>a</sup>The F12b reference energies were obtained with the large basis sets discussed in the text.<sup>b</sup>Basis sets used in the CBS extrapolation.<sup>c</sup>Raw value obtained with the largest basis set in this group.<sup>d</sup>Value quoted in Hill *et al.* found in Ref. 80.

+0.04 mE<sub>h</sub>. The earlier discussion of the sensitivity of the total energy to the choice of geminal exponent suggests that the use of optimized  $\beta$  values, which is not recommended, would increase the differences between F12 and standard CCSD values.

Our admittedly limited evidence suggests that the F12b reference energies *may* somewhat overshoot the CBS limit and lead to atomization energies that are somewhat larger than the values obtained from the standard method. Future work with even larger basis sets and shorter extrapolations could lead to more definitive conclusions. Based on the information on hand, we tentatively assign a crude uncertainty of  $\sim 0.1$  kcal/mol to the F12b reference atomization energies.

## IV. RESULTS AND DISCUSSION

### A. Separability of the Hartree-Fock and correlation components

It has been our practice to apply CBS extrapolation formulas (2)–(5) to the total CCSD(T) energy rather than perform separate extrapolations of the HF and correlation energy components. This choice was motivated by the belief that decreases in the CCSD(T) energy accompanying improvements in the basis set were dominated by the change in the correlation energy, at least for basis sets of aV5Z quality or better. Anecdotal evidence suggests that separate extrapolations of the HF and correlation components typically leads to

negligible differences in atomization energies using the average of multiple methods.<sup>16,65,98</sup> For example, a difference of only 0.07 kcal/mol was noted for the dissociation energy of C<sub>2</sub> with the aVTQ5Z basis set combination, an amount considerably smaller than several other sources of error. For larger molecules occasional instances were found where separate extrapolations led to differences on the order of 0.5 kcal/mol with the aVQZ basis set, but again other sources of error in the calculation were much larger.

Among the extrapolation formulas examined in detail in this study, only the Schwenke formulas inherently separate the HF and correlation energies. Many researchers perform separate extrapolations of the HF and correlation components when using Eq. (5),  $(1/\ell_{\max}^3)$ , which was intended strictly for the correlation energy. Other protocols for high accuracy thermochemistry, such as the Wn models, estimate the HF/CBS energy with a variety of formulas. W1, W2, and W3 used  $E_{\text{CBS}} + B/\ell_{\max}^5$ , whereas W4 used a formula to be discussed below.<sup>18–20</sup> The HEAT model<sup>22</sup> extrapolates the HF energy with an exponential. However, in a recent HEAT paper<sup>24</sup> it was noted, “in fact, it seems that there is little point in doing extrapolations of the HF-SCF energy at all.” Following a suggestion by Halkier *et al.*<sup>116</sup> the ccCA model<sup>25,117</sup> and the G4 model chemistry<sup>85</sup> used a linearized exponential  $E(n) = E_{\text{CBS}} + B e^{-1.63n}$  for estimating the HF limit. The Halkier *et al.* exponent was derived from a study of nine first row diatomic molecules for which numerical HF energies were available. Martin and Taylor<sup>56</sup> examined a three parameter expression of the form  $A + B/(\ell_{\max} + 1/2)^x$  and a two parameter function of the form  $A + B/(\ell_{\max} + 1/2)^{-5}$ . We find that, as with the correlation component, there is a general lack of consensus on how best to handle the HF energy, despite the relative ease of approaching within 1 kcal/mol via direct calculation. For example, in CO<sub>2</sub> the aVQZ basis set yields an atomization energy within  $\sim 0.1$  kcal/mol of the HF limit, without recourse to extrapolation.

A re-examination of this issue was undertaken in the present study using the 51 molecule F12 reference set. Two formulas were chosen for extrapolating the HF energy. The first was the three parameter exponential, given by Eq. (2). The second was a two parameter formula recommended by Karton and Martin,<sup>118</sup>

$$E(\ell_{\max}) = E_{\text{SCF}}(\text{CBS}) + A(\ell_{\max} + 1)e^{-B\sqrt{\ell_{\max}}}, \quad (15)$$

where  $B = 7$  for TQ and  $B = 9$  for Q5 and 56 combinations. Strictly speaking, Karton and Martin only proposed  $B = 9$  for the Q5 and 56 combinations, but we adopted  $B = 7$  from the data in their Table I. Both  $B$  values were determined from an analysis of numerical HF energies for 42 diatomic molecules published by Jensen.<sup>119</sup> The Karton and Martin formula is a variation on a three parameter formula originally suggested by Jensen,<sup>120</sup>

$$E(\ell_{\max}, n(s)) = E_{\text{SCF}}(\text{CBS}) + A(\ell_{\max} + 1)e^{-B\sqrt{n(s)}}, \quad (16)$$

where  $B$  is now an adjustable parameter and  $n(s)$  is the number of  $s$ -type functions in the basis set. Jensen also discussed the possibility of fixing  $B$  at some reasonable value and turning Eq. (16) into a two parameter fit. Jensen’s work, in turn,

was preceded by work on the hydrogen atom by Klopper and Kutzelnigg in which they described the Gaussian basis set convergence of the SCF energy to the exact value with the expression,

$$E(n) = E_{\text{SCF}}(\text{CBS}) + A e^{-B\sqrt{n}}, \quad (17)$$

where  $n$  is the number of  $s$  functions. Zhong *et al.* recommended Eq. (15) with  $B = 6.30$  and their newly developed nZAP basis sets, which are designed to converge systematically for both SCF and correlation energies.<sup>121</sup>

Neese and Valeev found that the SCF piece of the energy was surprisingly *more* difficult to handle than the correlation piece.<sup>122</sup> They adopted the same basic formula described above, but fitted  $B$  for every pair of basis sets. Very recently, Bischoff and Valeev proposed a multiresolution finite element technique for solving the SCF equations to a guaranteed precision of  $1 \mu\text{E}_h$  for systems up to 14 electrons.<sup>123</sup>

Among the four simple functional forms examined here (Eqs. (2)–(5)) the only one to show any notable effect from separate extrapolations of the HF and correlation pieces was the  $1/\ell_{\max}^3$  formula. In Table IV the impact of separate extrapolations is shown for combinations of  $1/\ell_{\max}^3$  and  $1/(\ell_{\max} + 1/2)^4$  for  $E_{\text{corr}}$  with the Karton/Martin and exponential functions for  $E_{\text{HF}}$ . The statistical errors ( $\varepsilon_{\text{MSD}}$ ,  $\varepsilon_{\text{MAD}}$ ,  $\varepsilon_{\text{STDEV}}$ ) for the  $1/\ell_{\max}^3$  formula are cut roughly in half with the (aVDTQ) and (aVTQ5) basis set combinations. The amount by which the  $1/\ell_{\max}^3$  formula typically overestimates the true CBS atomization limit is greatly reduced. Clearly, this expression does a particularly poor job of describing the HF basis set convergence. There is also a reduction for the (aVQ56) combination, but is statistically questionable in light of the  $\pm 0.1$  kcal/mol uncertainty in the F12 reference values. With the  $1/(\ell_{\max} + 1/2)^4$  formula there is no clear improvement for any of the basis set combinations. The same is true for the exponential and mixed formulas. If the analysis is expanded to the larger standard method reference set, the impact of separate extrapolations was found to be similar to what was observed with the F12 reference set, as will be discussed in a subsequent section.

These findings reflect the relatively minor role played by the HF energy in CCSD(T) basis set extrapolations. If we focused exclusively on HF atomization energies, we can use HF energies obtained with the reference basis sets augmented with the CABS-singles corrections. In this context, we find that the exp(aVDTQ) estimates yield  $\varepsilon_{\text{MSD}}$ ,  $\varepsilon_{\text{MAD}}$ , and  $\varepsilon_{\text{STDEV}}$  values that are  $\sim 0.2$  kcal/mol larger than the mixed, Karton/Martin or Schwenke values. Errors of this magnitude are small compared to the uncertainty inherent in CCSD(T)/CBS estimates with basis sets of this size. When larger basis sets are used, all of the extrapolation methods produce HF atomization error metrics that fall within 0.05 kcal/mol of the reference values.

Overall, the Karton and Martin Hartree-Fock formula was not found to provide any significant advantage over the exponential function for atomization energies in this collection of molecules. High absolute accuracy in the HF energies is not a prerequisite for achieving accuracy in the atomization energies. We only require that the error in the HF component be significantly smaller than the error in the

TABLE IV. Impact of separate HF/correlation CBS extrapolations on CCSD(T) atomization energies.<sup>a</sup>

Extrapolation method(s)								
Correlation	HF	Basis	Max. pos.	Max. neg.	MSD	MAD	STDEV	
$1/\ell_{\max}^3$	K/M <sup>b</sup>	(aVDTQ)	0.86	-0.58	0.31	0.35	0.40	
	Exponential		0.92	-1.55	0.05	0.28	0.38	
	Combined <sup>c</sup>		1.35	-0.05	0.60	0.60	0.65	
	K/M <sup>b</sup>	(aVTQ5)	0.49	-0.07	0.13	0.14	0.18	
	Exponential		0.51	-0.17	0.12	0.14	0.19	
	Combined <sup>c</sup>		0.68	-0.08	0.23	0.23	0.28	
$1/(\ell_{\max} + 1/2)^4$	K/M <sup>b</sup>	(aVQ56)	0.15	-0.03	0.03	0.04	0.05	
	Exponential		0.19	-0.04	0.04	0.05	0.06	
	Combined <sup>c</sup>		0.41	-0.01	0.09	0.09	0.12	
	K/M <sup>b</sup>	(aVDTQ)	0.44	-1.54	-0.17	0.27	0.37	
	Exponential		0.14	-0.53	-0.12	0.14	0.63	
	Combined <sup>c</sup>		0.78	-0.86	0.03	0.21	0.26	
$1/(\ell_{\max} + 1/2)^4$	K/M <sup>b</sup>	(aVTQ5)	0.11	-0.50	-0.11	0.13	0.18	
	Exponential		0.14	-0.53	-0.12	0.14	0.19	
	Combined <sup>c</sup>		0.24	-0.43	-0.05	0.10	0.14	
	K/M <sup>b</sup>	(aVQ56)	0.01	-0.26	-0.10	0.10	0.12	
	Exponential		0.03	-0.24	-0.09	0.09	0.11	
	Combined <sup>c</sup>		0.07	-0.22	-0.06	0.07	0.09	

<sup>a</sup>Based on the 51 molecules in the F12 reference set. Max. pos. = maximum positive deviation. Max. neg. = maximum negative deviation. MSD = mean signed deviation. MAD = mean absolute deviation. STDEV = standard deviation.

<sup>b</sup>K/M = Karton/Martin formula given in Ref. 118.

<sup>c</sup>Combined – the total CCSD(T) energies (Hartree-Fock and correlation energies) were extrapolated with the indicated correlation formula.

electron correlation component, a goal that is relatively easy to achieve given the difference in their respective convergence rates. The present findings are consistent with the atomization energy results reported in Jensen's 2005 study<sup>120</sup> with correlation consistent basis sets but differ somewhat from the much earlier study by Martin and Taylor who found improved results when using separate extrapolations for the SCF and correlation components.<sup>56</sup> The Martin-Taylor work used different formulas than were considered here.

## B. Statistical results utilizing the standard method reference set

The results of a statistical comparison based on the 141 member standard method reference set are presented in Table V. A graphical representation of these findings can be found in Figure 4. The CBS estimates were measured with respect to reference values obtained from basis sets that were at least one level higher in quality. For instance, an aVQ56Z extrapolation, which we will denote as CBS(aV6Z), would only be compared to a reference value obtained from at least an aV7Z basis set. As mentioned previously, while we sometimes refer to a three basis set combination, such as aVQ56Z, in the case of two parameter fits only the largest two sets were used. Before discussing the details of the analysis, several general observations will be made. First, at every basis set level even the least accurate formula yielded results roughly equivalent to raw values from the next higher level basis set, i.e.,  $\text{CBS}_{\text{worst}}(\text{aVnZ}) \approx \text{raw}(\text{aV}(n+1)\text{Z})$ . For example, the  $1/\ell_{\max}^3$  CBS(aVQZ) MAD value of 0.78 kcal/mol compares to a raw(aV5Z) MAD of 1.02 kcal/mol. The practical consequences of this are apparent when one considers that the computational cost of CCSD(T) scales as steeply as  $n^2N^4N_{\text{it}}$

(with a single  $n^3N^4$  step), where  $n$ ,  $N$ , and  $N_{\text{it}}$  are the number of occupied and unoccupied molecular orbitals and the number of CCSD iterations. Thus, for a fixed molecule CCSD(T) scales as  $N^4$  with respect to the size of the basis set. The best performing extrapolation formulas produce results that are superior to the raw results from basis sets two levels higher, i.e.,  $\text{CBS}_{\text{best}}(\text{aVnZ})$  was better than raw(aV(n+2)Z). Second, as the basis sets improve in quality all error metrics for the best formulas uniformly approach zero. Among the poorer performing formulas there were instances where some of the metrics increased slightly despite a step up in basis set quality.

Table V and Figure 4 illustrate the point that no single formula is optimal in all cases. Among the choices examined in detail, the mixed Gaussian/exponential formula proved most accurate with the small aVDTQZ basis set combination, but the  $1/(\ell_{\max} + 1/2)^4$  and 5-formula average were only slightly ( $\sim 0.05$  kcal/mol) worse. Recall that the estimated uncertainty in the standard method reference values is on the order of  $\pm 0.05$  kcal/mol. For the aVDTQZ combination the exponential and  $1/\ell_{\max}^3$  formulas displayed the largest errors, at least when the latter was used on total energies. However, when the Hartree-Fock energy is extrapolated separately from the CCSD(T) correlation energy, the  $1/\ell_{\max}^3$  formula does much better, roughly matching the accuracy of the Schwenke method. The values corresponding to separate extrapolations of the Hartree-Fock and correlation energies are listed in parentheses in Table V. As discussed in Sec. III A dealing with the F12 reference set, separate extrapolations roughly cut the  $1/\ell_{\max}^3$  error metrics in half.

The poor performance of the exponential comes as no surprise, since its tendency to underestimate the importance of higher angular momentum functions beyond  $\ell_{\max} = 4$  is well known, as reflected in a MSD of

TABLE V. CBS performance statistics of atomization energies obtained relative to the standard CCSD(T) reference set (kcal/mol).<sup>a</sup>

Basis sets	Formula	Max. pos.	Max. neg.	MSD	MAD	STDEV
aVDTQZ	Exp.	2.94	-2.75	-0.43	0.57	0.80
	Mixed	1.77	-1.27	0.12	0.27	0.39
	$1/(\ell_{\max} + 1/2)^4$	1.76	-0.84	0.20	0.33	0.44
	$1/\ell_{\max}^3$	3.15 (2.51) <sup>b</sup>	-0.53 (-1.45)	0.83 (0.38)	0.84 (0.50)	1.06 (0.67)
	Schwenke	1.77	-1.43	0.33	0.46	0.60
	Average	2.05	-1.17	0.22	0.33	0.45
aVTQ5Z	Exp.	0.21	-1.94	-0.53	0.53	0.67
	Mixed	0.09	-1.13	-0.31	0.31	0.38
	$1/(\ell_{\max} + 1/2)^4$	0.49	-0.47	-0.01	0.11	0.15
	$1/\ell_{\max}^3$	1.26 (0.81)	-0.10 (-0.77)	0.30 (0.12)	0.30 (0.16)	0.38 (0.22)
	Schwenke	0.53	-0.91	-0.04	0.13	0.19
	Average	0.92	-0.72	-0.11	0.16	0.23
aVQ56Z	Exp.	0.08	-0.79	-0.21	0.21	0.26
	Mixed	0.00	-0.80	-0.22	0.22	0.26
	$1/(\ell_{\max} + 1/2)^4$	0.49	-0.13	0.02	0.05	0.08
	$1/\ell_{\max}^3$	0.90 (0.91)	0.00 (-0.58)	0.18 (0.09)	0.18 (0.11)	0.23 (0.17)
	Schwenke	0.75	-0.62	0.03	0.08	0.14
	Average	0.42	-0.22	-0.04	0.07	0.10
aV567Z	Exp.	0.12	-0.23	-0.01	0.06	0.08
	Mixed	0.05	-0.22	-0.09	0.10	0.13
	$1/(\ell_{\max} + 1/2)^4$	0.18	-0.03	0.06	0.06	0.08
	$1/\ell_{\max}^3$	0.27 (0.29)	0.00 (-0.15)	0.12 (0.11)	0.12 (0.12)	0.14 (0.14)
	Average	0.27	-0.09	0.02	0.04	0.07
	Combined	Exp.			-0.37	0.42
Mixed				-0.12	0.26	0.34
$1/(\ell_{\max} + 1/2)^4$				0.07	0.16	0.27
$1/\ell_{\max}^3$				0.43 (0.20)	0.43 (0.26)	0.66 (0.42)
Schwenke				0.11	0.24	0.39
Average				0.03	0.18	0.29

<sup>a</sup>Based on the 141 member reference set. Max. pos. = maximum positive deviation (CBS extrapolated value is larger than the reference value). Max. neg. = largest magnitude negative deviation (CBS extrapolated value is smaller than the reference value). MSD = mean signed deviation. MAD = mean absolute deviation. STDEV = standard deviation. The reference set was based on the average of the available formulas obtained with the largest available basis sets. Unless otherwise noted, all of the CBS extrapolations were applied to the total CCSD(T) energies, except the Schwenke formula. The "Combined" values were based on the cumulative statistics for basis set combinations aVDTQZ through aV567Z.

<sup>b</sup>Values in parentheses are based on separate extrapolations of the Hartree-Fock and correlation energies. The Hartree-Fock extrapolation was performed with the Karton/Martin formula found in Ref. 118. An exponential function produced very similar results.

-0.43 kcal/mol. On the other hand, the  $1/\ell_{\max}^3$  formula tends to overestimate contributions from higher angular momentum functions, even if separate extrapolations are employed (MSD = 0.38 kcal/mol). The significance of Table V and Figure 4 is they provide for the first time a quantitative measure of the variations in extrapolation performance for a sizable sample set and across a wide range of basis sets. With the large aV567Z combination of basis sets all of the formulas are closely grouped. The difference between the  $1/\ell_{\max}^3$  values based on total energies and separate Hartree-Fock and correlation energies is negligible. Even with the slightly smaller aVQ56Z combination this difference is approaching the point where it is statistically questionable. The exponential and mixed formulas often underestimate the CBS limit while the  $\ell_{\max}$  formulas err on the opposite side, in a manner similar to the situation depicted in Figure 2 for CO<sub>2</sub>. Balancing these competing errors, which are most pronounced for smaller basis sets is what first suggested the use of an average.

For thermochemical studies it has been our practice to apply a single, consistent CBS extrapolation technique across all basis sets and molecules, in order to gauge the stability of the values. We tended to avoid switching formulas for different

situations. Therefore, it was of interest to examine the cumulative statistics across all basis sets. We find that the  $1/(\ell_{\max} + 1/2)^4$  and 5-formula average provide the smallest measures of error ( $\epsilon_{\text{MAD}} \sim 0.17$ ,  $\epsilon_{\text{STDEV}} \sim 0.27$  kcal/mol) on the basis of 417 comparisons. The largest combined errors were found with the exponential and the  $1/\ell_{\max}^3$  formulas, where  $\epsilon_{\text{MAD}}$  was more than twice as large as in the best cases. The Schwenke and mixed formulas fell somewhere in the middle. These results are insensitive to the choice of formula used in defining the reference values. Switching the reference values from the 5-formula average to the  $1/(\ell_{\max} + 1/2)^4$  formula produced shifts in the numerical values of the error metrics by just 0.02–0.04 kcal/mol.

While the extrapolation formulas were found to be very effective in general at reducing the one-particle truncation error, several instances were found where the extrapolated atomization energies overshoot the CBS limit by more than the raw value obtained with the largest basis set undershot the limit. We refer to such situations as "failures" of the methods. Figure 5 shows the number of failures for each extrapolation technique where a threshold of 0.05 kcal/mol was used to define a "failure." Most of the failures amounted to less than 0.1

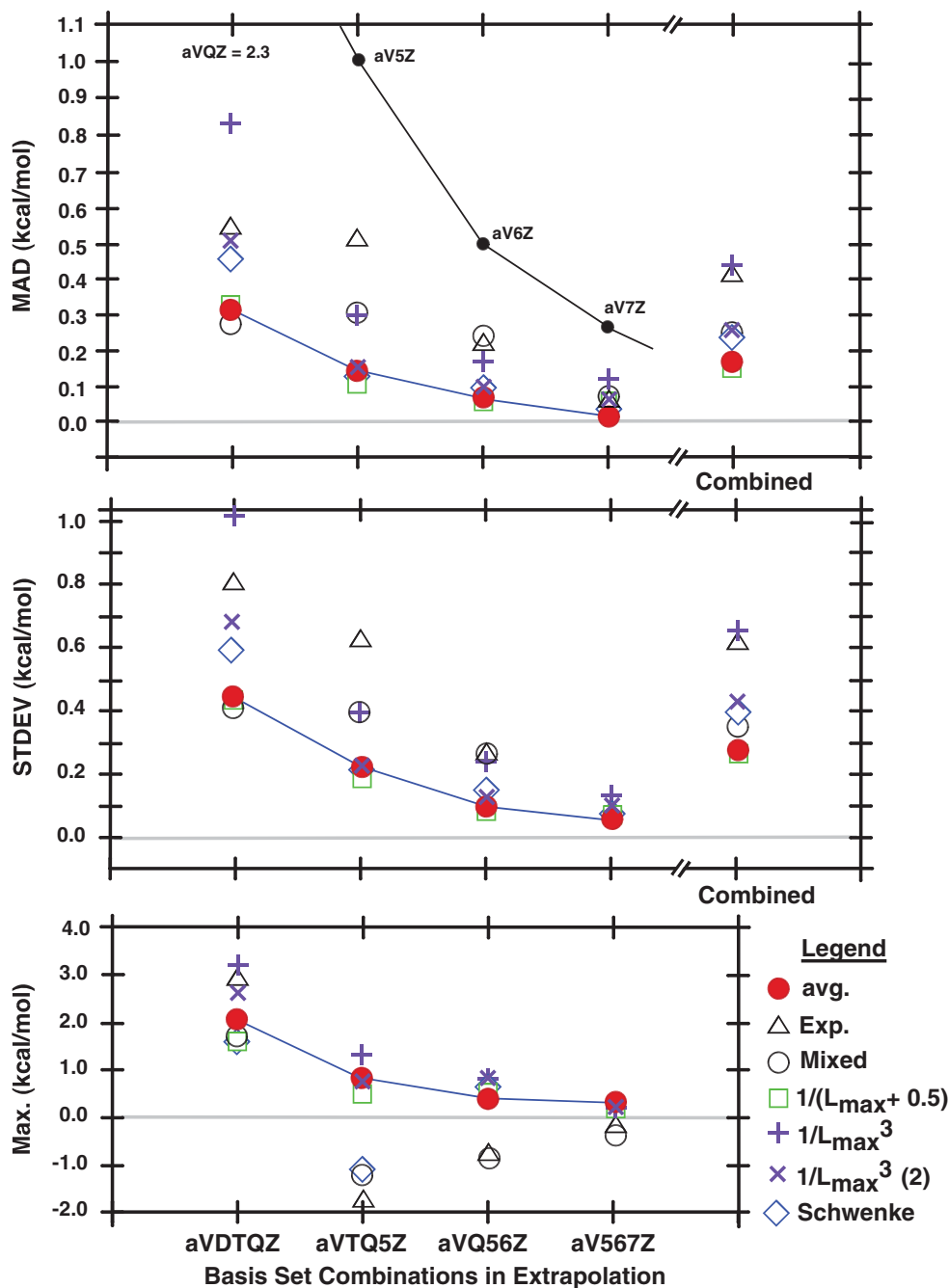


FIG. 4. Convergence of the CCSD(T)/CBS extrapolation statistics as a function of the basis set quality for the large standard method reference set. The  $1/\ell_{\max}^3$  (2) values were obtained by separately extrapolating the Hartree-Fock and correlation energies. All other values, except those labeled Schwenke, correspond to extrapolation of the total CCSD(T) energies.

kcal/mol, but in two of the cases the error slightly exceeded 1 kcal/mol. Both cases involved the exponential function and the aVDTQZ basis set combination with molecules containing the Zn atom. The cause for this anomaly was traced back to the aVDZ basis set for Zn which produced energies that were in some sense much better than is normally seen with a double zeta set. In fact, the aVDZ basis set for Zn was actually derived from the aVTZ set mainly by contracting the  $2f$  set to  $1f$  using atomic natural orbital coefficients.<sup>124</sup> This caused the aVDZ  $\rightarrow$  aVTZ energy difference to be similar in magnitude to the aVTZ  $\rightarrow$  aVQZ energy differences which in turn caused the exponential extrapolation to overestimate the

CBS limit by several millihartrees. With larger basis sets this anomalous behavior disappeared.

As mentioned earlier, the W4 thermochemical model of Karton *et al.*<sup>20</sup> decomposes the CCSD(T) extrapolation into four separate pieces (SCF, CCSD singlet pair, CCSD triplet pair, and (T)). In accord with most high accuracy computational strategies, the complete W4 method includes additional corrections to account for core/valence correlation, scalar relativistic and higher order correlation effects. The largest correlation contributions to the atomization energy come from the CCSD singlet pair and (T) components. Because W4 uses a  $1/\ell_{\max}^3$  formula for both, it

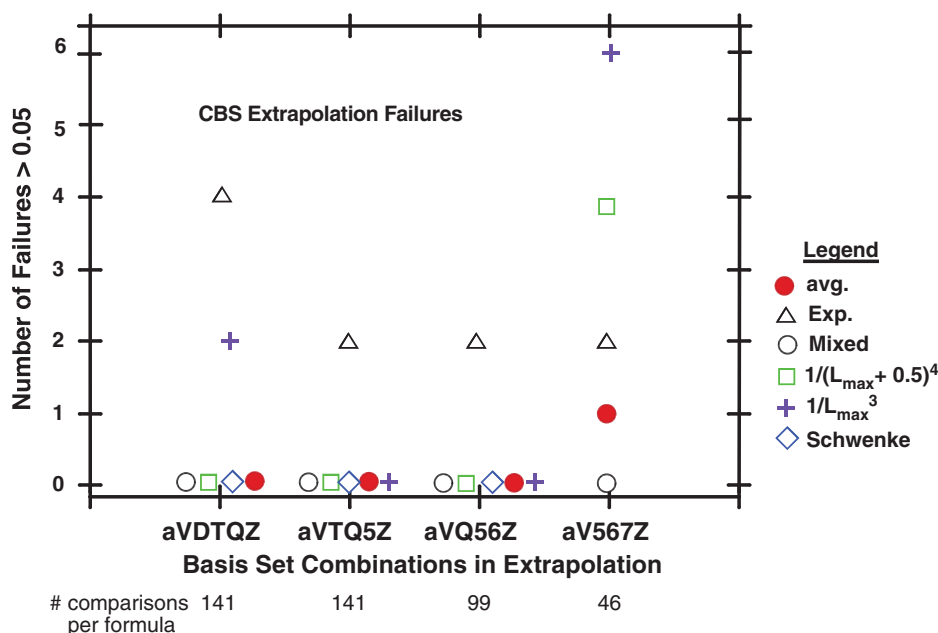


FIG. 5. Number of instances in which the CBS extrapolation error, measured with respect to the best available values, exceeds the error in the raw result by more than 0.05 kcal/mol.

seems plausible that W4 might mimic the error statistics associated with the  $1/\ell_{\max}^3(\text{aVQ56Z})$  formula, whose results are found in Table V. W4 utilizes a mixture of basis sets that include  $\text{cc-pVnZ}$  (H),  $\text{aug-cc-pVnZ}$  (B–F), and  $\text{aug-cc-pV}(n+d)\text{Z}$  (A1–C1), with the SCF, singlet and triplet pair energies handled by the  $\text{aV5Z/aV6Z}$  combination and the (T) component handled by a smaller  $\text{aVQZ/aV5Z}$  combination.

As already mentioned, the  $1/\ell_{\max}^3(\text{aVQ56Z})$  formula tends to overestimate the reference values. It has a mean signed error of 0.18 kcal/mol (extrapolation on the total energy) or 0.09 kcal/mol (separately extrapolating the SCF and correlation energies). However, Karton *et al.* reported very high levels of agreement between W4 atomization energies and experimental/theoretical values taken from the active thermochemical tables (ATcT).<sup>20,125</sup> For a collection of 25 small first and second row molecules the RMS error was 0.09 kcal/mol (MAD = 0.07 kcal/mol). The authors actually examined 26 systems, but one of them (C1CN) displayed a significantly larger error and the ATcT value for it was largely based on older theoretical numbers. If C1CN is included in the statistics, the RMS increased to 0.15 kcal/mol and MAD to 0.09. The mean signed deviation for W4 was a scant  $-0.04$  kcal/mol, indicative of very little systematic bias. Karton *et al.*<sup>20</sup> also reported that including “W3 species” and older experimental data in the statistics, but continuing to exclude C1CN, yielded RMS and MAD values of 0.10 kcal/mol. Although the  $\text{CCSD(T)(FC)/CBS}$  energy is the largest single piece of the multi-component W4 method, the associated error may not be the largest source of error in the final atomization energies. Unless there is a systematic cancellation of error, the high level of agreement with ATcT values would appear to imply an accuracy in the  $\text{CCSD(T)(FC)/CBS}$  piece of no more than  $\pm 0.1$  kcal/mol. Here, we are assuming the quoted uncertainties in the ATcT values ( $\epsilon_{\text{MAD}} = 0.05$  kcal/mol) are accurate.

In an effort to understand these differences, W4  $\text{CCSD(T)/CBS}$  values were obtained for a collection of 42 molecules chosen from the standard method reference set of this study. All of the molecules found in Table VI of the W4 paper were included in this comparison.<sup>20</sup> W4 uses two slightly different open shell coupled cluster methods, one for the singlet/triplet pair extrapolation and another for the (T) extrapolation. For the sake of simplicity and to make for a more direct comparison with the reference data, we have chosen to use the  $\text{R/UCCSD}$  definition in MOLPRO for both parts. We have also chosen to use symmetry equivalenced atomic asymptotes, in order to better match the reference set. These changes resulted in small increases (several hundredths of a kcal/mol) compared to the data for 30 molecules listed in the W4 paper. All calculations were performed at the optimal  $\text{CCSD(T)(FC)/VQZ}$  geometries, as specified in the W4 protocol. We will refer to these results as “W4-like.”

As expected, the W4-like atomization statistics closely track the  $1/\ell_{\max}^3(\text{aVQ56Z})$  statistics, with  $\epsilon_{\text{MSD}} = 0.15$  kcal/mol (W4) compared to 0.11 kcal/mol ( $1/\ell_{\max}^3$ ) with separate HF and correlation extrapolations, reflecting the same systematic bias towards overshooting our best estimates of the limit. The largest observed overestimation of the CBS limit was 0.5 kcal/mol. Expanding the  $1/\ell_{\max}^3$  statistics to the entire 141 molecule reference set increases  $\epsilon_{\text{MSD}}$  to 0.20 kcal/mol. Consequently, we tentatively assume the same would be true of the W4-like statistics. Offsetting this tendency to overestimate the CBS limit is the omission of two corrections in the W4 procedure that are included in the more computationally demanding W4.2 and W4.3 models. These include a consideration of higher order core/valence correlation effects and an improved accounting of higher order valence correlation. The magnitude of both corrections varies considerably from molecule to molecule, but in general both tend to increase the atomization energy. While many



TABLE VI. CBS performance statistics obtained relative to the Explicitly correlated CCSD(T) reference set (kcal/mol).<sup>a</sup>

Basis sets	Formula	Max. pos.	Max. neg.	MSD	MAD	STDEV
aVDTQZ	Exp.	0.36	-2.27	-0.42	0.45	0.61
	Mixed	0.72	-1.02	-0.01	0.18	0.26
	$1/(\ell_{\max} + 1/2)^4$	0.78	-0.86	0.02	0.20	0.26
	$1/\ell_{\max}^3$	1.35	NA	0.59	0.59	0.65
	Schwenke	0.83	-0.64	0.26	0.31	0.36
	Average	0.77	-0.88	0.09	0.20	0.27
aVTQ5Z	Exp.	0.01	-1.31	-0.47	0.47	0.55
	Mixed	0.05	-0.84	-0.31	0.31	0.38
	$1/(\ell_{\max} + 1/2)^4$	0.24	-0.43	-0.05	0.10	0.14
	$1/\ell_{\max}^3$	0.68	-0.08	0.22	0.23	0.28
	Schwenke	0.27	-0.29	-0.01	0.09	0.11
	Average	0.17	-0.54	-0.12	0.15	0.19
aVQ56Z	Exp.	NA	-0.74	-0.28	0.28	0.32
	Mixed	NA	-0.62	-0.27	0.27	0.31
	$1/(\ell_{\max} + 1/2)^4$	0.07	-0.22	-0.06	0.07	0.09
	$1/\ell_{\max}^3$	0.41	-0.01	0.08	0.09	0.12
	Schwenke	0.08	-0.09	-0.01	0.03	0.04
	Average	0.02	-0.30	-0.11	0.11	0.13

<sup>a</sup>Based on the 51 member reference set. Max. pos. = maximum positive deviation (CBS extrapolated value is larger than the reference value). Max. neg. = largest magnitude negative deviation (CBS extrapolated value is smaller than the reference value). NA indicates that no deviations of that type were found. MAD = mean absolute deviation. STDEV = standard deviation. The reference set was based on very large basis set CCSD(T)-F12b calculations and did not involve any extrapolation. All of the CBS extrapolations were applied to the total CCSD(T) energies, except the Schwenke formula.

of the higher order core/valence corrections for the W4 set of molecules are negligible ( $<0.05$  kcal/mol), examples of some that are not include: 0.11 ( $B_2$ ), 0.13 ( $O_3$  triples only), 0.27 ( $C_2$ ), 0.24 (BN), and 0.07 ( $CO_2$ ), all in kcal/mol. These values are taken from the columns labeled “MOLPRO  $\Delta\Delta[T_3-(T)]$ ” and “ $\Delta\Delta(Q)$ ” in Table IV of the W4 paper and from our own calculations. If instead of the MOLPRO  $\Delta\Delta[T_3-(T)]$  column entries, we used the ACES  $\Delta\Delta[T_3-(T)]$  entries (many of which are negative) there would be some cancellation with the positive  $\Delta\Delta(Q)$  corrections. The effect of improved accounting of valence correlation energy, taken as the difference between W4.2 and W4.3, is also generally small. However, some corrections are non-negligible, such as: 0.07 ( $Cl_2$ ), 0.07 (SO), 0.10 ( $F_2$ ), 0.08 (CS), and 0.10 ( $PH_3$ ), all values in kcal/mol. For particularly problematic molecules, we have seen corrections to CCSD(T) atomization energies as large as 0.37 kcal/mol obtained from explicit FCI calculations. In W4.3, where both corrections discussed above are included, the model overestimates the ATcT atomization energies in 10 out of 13 comparisons. Thus, it seems possible that the level of agreement between W4 and ATcT or other experimental values is partially due to a fortuitous cancellation of errors.

### C. Statistical results from the explicitly correlated method reference set

Table VI contains the results of a statistical analysis based on the F12b reference set. As already discussed, this reference set was created without reliance on extrapolation formulas. While the numerical values differ somewhat from the values presented in Table V, the overall picture that emerges

TABLE VII. Raw CCSD(T)-F12b and F12b/CBS performance statistics obtained relative to the explicitly correlated CCSD(T) reference set (kcal/mol).<sup>a</sup>

Basis set	Formula	Max. pos.	Max. neg.	MSD	MAD	STDEV
VDZ-F12	None	0.08	-4.73	-2.27	2.27	2.53
VTZ-F12	None	NA	-1.74	-0.88	0.88	0.96
VQZ-F12	None	NA	-0.61	-0.28	0.28	0.31
VDTZ-F12	Hill <i>et al.</i>	0.29	-1.76	-0.34	0.37	0.53
VTQZ-F12	Hill <i>et al.</i>	0.19	-0.13	0.01	0.06	0.06

<sup>a</sup>Based on the 51-member reference set. Max. pos. = maximum positive deviation (CBS extrapolated value is larger than the reference value). Max. neg. = largest magnitude negative deviation (CBS extrapolated value is smaller than the reference value). NA indicates that no deviations of that type were found. MAD = mean absolute deviation. STDEV = standard deviation. The reference set was based on very large basis set CCSD(T)-F12b calculations and did not involve any extrapolation. Hill *et al.* of Ref. 80.

is quite similar. Recall that the estimated uncertainty in the F12b reference values is on the order of  $\pm 0.1$  kcal/mol. For small basis set combinations, the mixed formula displayed the best performance by a very small margin over the  $1/(\ell_{\max} + 1/2)^4$  and 5-formula average, with  $\epsilon_{\text{MAD}} = 0.18$  and  $\epsilon_{\text{STDEV}} = 0.26$  kcal/mol. These findings are in semi-quantitative agreement with the standard method reference set results, where  $\epsilon_{\text{MAD}} = 0.26$  and  $\epsilon_{\text{STDEV}} = 0.37$  kcal/mol. The exponential and  $1/\ell_{\max}^3$  formulas displayed significantly larger errors, with the Schwenke formula falling somewhere in between.

With the aVTQ5Z basis set combination, the  $1/(\ell_{\max} + 1/2)^4$  and Schwenke formulas generated the smallest error metrics, with the errors for the 5-formula average coming in only slightly larger. Finally, with the large aVQ56Z combination, the Schwenke formula errors were roughly half the size of the  $1/(\ell_{\max} + 1/2)^4$  and 5-formula average, but the differences among all formulas were below the  $\pm 0.1$  threshold at which they are likely to be significant. With the standard method reference set results in Table V, the Schwenke,  $1/(\ell_{\max} + 1/2)^4$  and 5-formula average gave more-or-less equivalent statistics. The aVQ56Z standard method and F12b reference sets also differ on the relative performance of the  $1/\ell_{\max}^3$  formula, which fares significantly better with the explicitly correlated reference set. However, even with the F12b reference set, the  $1/\ell_{\max}^3$  formula exhibits a large maximum positive error of 0.4 kcal/mol.

Newly developed CCSD(T)-F12b/CBS extrapolation formulas were also tested on the 51-member reference set. Recall that the extrapolation coefficients of Hill *et al.* were optimized for a subset of the molecules in this reference set.<sup>80</sup> The results are shown in Table VII. Relative to the raw CCSD(T) results in Table II for the 141 member reference set, the raw CCSD(T)-F12b/VDZ-F12 atomization energies are roughly comparable to standard CCSD(T)/aVQZ values. At the high end, the raw CCSD(T)-F12b/VQZ-F12 level of theory yields a MAD comparable to CCSD(T)/aV7Z. These findings are consistent with the findings reported in an earlier study that was limited to hydrocarbons.<sup>98</sup> CBS extrapolation with the VDTZ-F12 basis set combination cuts the MSD, MAD, and standard deviation by a factor of two, but the maximum negative error and the standard deviation are largely unaffected.

Switching to the larger VTQZ-F12 basis set combination dramatically reduces most of the error metrics, relative to the raw VQZ-F12 results, and provides an overall level of accuracy equivalent to the best standard aVQ56Z extrapolated values reported in Table VI, but at a much reduced computational cost.

## V. CONCLUSION

The simple question that is raised most often about CBS extrapolation formulas is “Which one is best?.” Sadly, as the results of this study and previous studies reveal, there is no simple answer. In terms of reducing the basis set truncation error, the performance statistics show that any of the techniques are quite effective. Formulas with the largest errors still yielded a level of accuracy comparable to raw results obtained with basis sets at least one level better in quality. The  $1/(\ell_{\max} + 1/2)^4$  and 5-formula average (both using total energies) were found to yield the best statistics, while the exponential and  $1/\ell_{\max}^3$  (HF and correlation combined) formulas produced the largest errors. The Schwenke, mixed and  $1/\ell_{\max}^3$  (the latter with separate HF and correlation) approaches fell somewhere in between. For the small aVDTQZ sequence of basis sets the mixed formula posted the smallest error metrics. In light of these findings, any of the formulas considered here might be preferred on formal grounds or because their accuracy for a specific collection of molecules has been documented. The difference in accuracy between the best and the worst across all basis set combinations is approximately a factor of two. The literature already contains a large number of proposed extrapolation procedures and each year brings new offerings. We are under no illusions that a single calibration study will sway many opinions away from or towards any particular formula, but it is our hope that future suggestions will be more thoroughly vetted.

Ideally, the reference set by which the extrapolation formulas are judged would be devoid of ambiguity and accurate to several hundredths of a kcal/mol. While that may become possible in the future, the current state-of-the-art in electronic structure methods is unable to reduce the uncertainties in the reference data to less than 0.05-0.10 kcal/mol without invoking extrapolation procedures and thereby introducing an additional element of uncertainty. Judging by the progression of errors for the raw atomization energies shown in Table II, it would require an ability to perform standard CCSD(T)/aug-cc-pV9Z quality calculations on the collection of molecules in this study in order to avoid completely the use of extrapolations. In a future study we hope to examine the possibility of using the data accumulated in this study to improve upon the formulas presented here.

See supplementary material for the new basis sets and SCF and CCSD(T)(FC) energies.<sup>126</sup>

## ACKNOWLEDGMENTS

K.A.P. and J.G.H. gratefully acknowledge the support of the National Science Foundation (NSF) (Grant No. CHE-0723997).

- <sup>1</sup>T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- <sup>2</sup>W. Kutzelnigg, *Theor. Chim. Acta* **68**, 445 (1985).
- <sup>3</sup>W. Kutzelnigg and W. Klopper, *J. Chem. Phys.* **94**, 1985 (1991).
- <sup>4</sup>W. Klopper and W. Kutzelnigg, *Chem. Phys. Lett.* **134**, 17 (1987).
- <sup>5</sup>W. Klopper and W. Kutzelnigg, *J. Phys. Chem.* **94**, 5625 (1990).
- <sup>6</sup>W. Klopper, *Mol. Phys.* **99**, 481 (2001).
- <sup>7</sup>E. F. Valeev, W. D. Allen, R. Hernandez, C. D. Scherrill, and H. F. Schaefer, III, *J. Chem. Phys.* **118**, 8594 (2003).
- <sup>8</sup>T. B. Adler, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **127**, 221106 (2007).
- <sup>9</sup>H.-J. Werner, T. B. Adler, and F. R. Manby, *J. Chem. Phys.* **126**, 164102 (2007).
- <sup>10</sup>G. Knizia and H. J. Werner, *J. Chem. Phys.* **128**, 154103 (2008).
- <sup>11</sup>G. Knizia, T. B. Adler, and H.-J. Werner, *J. Chem. Phys.* **130**, 054104 (2009).
- <sup>12</sup>G. Rauhut, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **130**, 054105 (2009).
- <sup>13</sup>B. J. Hammond, W. A. Lester, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).
- <sup>14</sup>S. Manten and A. Lüchow, *J. Chem. Phys.* **115**, 5362 (2001).
- <sup>15</sup>N. Nemeč, M. D. Towler, and R. J. Needs, *J. Chem. Phys.* **132**, 034111 (2010).
- <sup>16</sup>D. Feller, K. A. Peterson, and D. A. Dixon, *J. Chem. Phys.* **129**, 204105 (2008).
- <sup>17</sup>D. Feller and K. A. Peterson, *J. Chem. Phys.* **108**, 154 (1998).
- <sup>18</sup>J. M. L. Martin and G. de Oliveira, *J. Chem. Phys.* **111**, 1843 (1999).
- <sup>19</sup>A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kallay, and J. Gauss, *J. Chem. Phys.* **120**, 4129 (2004).
- <sup>20</sup>A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, *J. Chem. Phys.* **125**, 144108 (2006).
- <sup>21</sup>A. Karton, P. R. Taylor, and J. M. L. Martin, *J. Chem. Phys.* **127**, 064104 (2007).
- <sup>22</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>23</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>24</sup>M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, *J. Chem. Phys.* **128**, 114111 (2008).
- <sup>25</sup>N. J. DeYonker, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **124**, 114104 (2006).
- <sup>26</sup>N. J. DeYonker, T. Grimes, S. Yokel, A. Dinescu, B. Mintz, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **125**, 104111 (2006).
- <sup>27</sup>N. J. DeYonker, K. A. Peterson, G. Steyl, A. K. Wilson, and T. R. Cundari, *J. Phys. Chem. A* **111**, 11269 (2007).
- <sup>28</sup>D. Feller, D. A. Dixon, and K. A. Peterson, *J. Phys. Chem.* **102**, 7053 (1998).
- <sup>29</sup>D. A. Dixon and D. Feller, *J. Phys. Chem. A* **102**, 8209 (1998).
- <sup>30</sup>D. A. Dixon, D. Feller, and G. Sandrone, *J. Phys. Chem. A* **103**, 4744 (1999).
- <sup>31</sup>D. Feller and D. A. Dixon, *J. Phys. Chem. A* **103**, 6413 (1999).
- <sup>32</sup>D. Feller and D. A. Dixon, *J. Phys. Chem. A* **104**, 3048 (2000).
- <sup>33</sup>D. Feller and J. A. Franz, *J. Phys. Chem. A* **104**, 9017 (2000).
- <sup>34</sup>B. Ruscic, D. Feller, D. A. Dixon, K. A. Peterson, L. B. Harding, R. L. Asher, and A. F. Wagner, *J. Phys. Chem. A* **105**, 1 (2001).
- <sup>35</sup>D. Feller and D. A. Dixon, *J. Chem. Phys.* **115**, 3484 (2001).
- <sup>36</sup>B. Ruscic, A. F. Wagner, L. B. Harding, R. L. Asher, D. Feller, D. A. Dixon, K. A. Peterson, Y. Song, X. Qian, C. Ng, J. Liu, W. Chen, and D. W. Schwenke, *J. Phys. Chem. A* **106**, 2727 (2002).
- <sup>37</sup>D. A. Dixon, W. A. deJong, K. A. Peterson, and J. S. Francisco, *J. Phys. Chem. A* **106**, 4725 (2002).
- <sup>38</sup>D. Feller, K. A. Peterson, W. A. deJong, and D. A. Dixon, *J. Chem. Phys.* **118**, 3510 (2003).
- <sup>39</sup>D. Feller, D. A. Dixon, and J. S. Francisco, *J. Phys. Chem. A* **107**, 1604 (2003).
- <sup>40</sup>D. Feller and D. A. Dixon, *J. Phys. Chem. A* **107**, 9641 (2003).
- <sup>41</sup>N. B. Balabanov and K. A. Peterson, *J. Phys. Chem. A* **107**, 7465 (2003).
- <sup>42</sup>D. A. Dixon, D. Feller, K. O. Christe, W. W. Wilson, A. Vij, V. Vij, H. D. B. Jenkins, R. Olson, and M. S. Gordon, *J. Am. Chem. Soc.* **126**, 834 (2004).
- <sup>43</sup>D. Feller, *J. Comput. Chem.* **17**, 1571 (1996).
- <sup>44</sup>T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).

- <sup>45</sup>R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- <sup>46</sup>D. Feller, *J. Chem. Phys.* **96**, 6104 (1992).
- <sup>47</sup>M. W. Feyereisen, D. Feller, and D. A. Dixon, *J. Phys. Chem.* **100**, 2993 (1996).
- <sup>48</sup>K. A. Peterson, D. E. Woon, and T. H. Dunning, Jr., *J. Chem. Phys.* **100**, 7410 (1994).
- <sup>49</sup>D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **101**, 8877 (1994).
- <sup>50</sup>C. Schwartz, *Phys. Rev.* **126**, 1015 (1962).
- <sup>51</sup>C. Schwartz, in *Methods in Computational Physics*, edited by B. J. Alder, S. Fernbach, and M. Rotenberg (Academic, New York, 1963), Vol. 2, pp. 262–265.
- <sup>52</sup>D. P. Carroll, H. J. Silverstone, and R. M. Metzger, *J. Chem. Phys.* **71**, 4142 (1979).
- <sup>53</sup>R. N. Hill, *J. Chem. Phys.* **83**, 1173 (1985).
- <sup>54</sup>W. Kutzelnigg and J. D. Morgan, *J. Chem. Phys.* **96**, 4484 (1992).
- <sup>55</sup>J. M. L. Martin, *Chem. Phys. Lett.* **259**, 669 (1996).
- <sup>56</sup>J. M. L. Martin and P. R. Taylor, *J. Chem. Phys.* **106**, 8620 (1997).
- <sup>57</sup>W. Klopper, K. L. Bak, P. Jørgensen, J. Olsen, and T. Helgaker, *J. Phys. B* **32**, R103 (1999).
- <sup>58</sup>J. Almlöf and P. R. Taylor, *J. Chem. Phys.* **86**, 4070 (1987).
- <sup>59</sup>C. F. Bunge, *Theor. Chim. Acta* **16**, 126 (1970).
- <sup>60</sup>D. Feller and K. A. Peterson, *J. Chem. Phys.* **110**, 8384 (1999).
- <sup>61</sup>T. Helgaker, W. Klopper, H. Koch, and J. Nago, *J. Chem. Phys.* **106**, 9639 (1997).
- <sup>62</sup>A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, *Chem. Phys. Lett.* **286**, 243 (1998).
- <sup>63</sup>D. W. Schwenke, *J. Chem. Phys.* **122**, 014107 (2005).
- <sup>64</sup>D. Feller, K. A. Peterson, and T. D. Crawford, *J. Chem. Phys.* **124**, 054107 (2006).
- <sup>65</sup>D. Feller and K. A. Peterson, *J. Chem. Phys.* **126**, 114105 (2007).
- <sup>66</sup>D. Feller, K. A. Peterson, and D. A. Dixon, *J. Phys. Chem. A* **114**, 613 (2010).
- <sup>67</sup>C. C. M. Samson and W. Klopper, *Mol. Phys.* **102**, 2499 (2004).
- <sup>68</sup>J. M. L. Martin and T. J. Lee, *Chem. Phys. Lett.* **258**, 129 (1996).
- <sup>69</sup>X. Huang and T. J. Lee, *J. Chem. Phys.* **129**, 044312 (2008).
- <sup>70</sup>X. Huang, E. F. Valeev, and T. J. Lee, *J. Chem. Phys.* **133**, 244108 (2010).
- <sup>71</sup>R. J. Gdanitz, *J. Chem. Phys.* **113**, 5145 (2000).
- <sup>72</sup>D. G. Truhlar, *Chem. Phys. Lett.* **294**, 45 (1998).
- <sup>73</sup>P. L. Fast, M. L. Sánchez, and D. G. Truhlar, *J. Chem. Phys.* **111**, 2921 (1999).
- <sup>74</sup>Y. Chuang and D. G. Truhlar, *J. Phys. Chem. A* **103**, 651 (1999).
- <sup>75</sup>A. Halkier, T. Helgaker, W. Klopper, P. Jørgensen, and A. G. Császár, *Chem. Phys. Lett.* **310**, 385 (1999).
- <sup>76</sup>A. J. C. Varandas, *J. Chem. Phys.* **113**, 8880 (2000).
- <sup>77</sup>S. B. Huh and J. S. Lee, *J. Chem. Phys.* **118**, 3035 (2003).
- <sup>78</sup>D. Bakowies, *J. Chem. Phys.* **127**, 084105 (2007).
- <sup>79</sup>H. Fliegl, W. Klopper, and C. Hättig, *J. Chem. Phys.* **122**, 084107 (2005).
- <sup>80</sup>J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **131**, 194105 (2009).
- <sup>81</sup>J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989).
- <sup>82</sup>L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **93**, 2537 (1990).
- <sup>83</sup>L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.* **94**, 7221 (1991).
- <sup>84</sup>L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.* **109**, 7764 (1998).
- <sup>85</sup>L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **126**, 084108 (2007).
- <sup>86</sup>C. W. Bauschlicher, Jr. and H. Partridge, *J. Chem. Phys.* **103**, 1788 (1995).
- <sup>87</sup>L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov, and J. A. Pople, *J. Chem. Phys.* **110**, 4703 (1999).
- <sup>88</sup>L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **114**, 108 (2001).
- <sup>89</sup>D. J. Henry, M. B. Sullivan, and L. Radom, *J. Chem. Phys.* **118**, 4849 (2003).
- <sup>90</sup>G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley, and J. Mantzaris, *J. Chem. Phys.* **89**, 2193 (1988).
- <sup>91</sup>G. A. Petersson and M. A. Al-Laham, *J. Chem. Phys.* **94**, 6081 (1991).
- <sup>92</sup>G. A. Petersson, T. G. Tensfeldt, and J. A. Montgomery, Jr., *J. Chem. Phys.* **94**, 6091 (1991).
- <sup>93</sup>G. P. F. Wood, L. Radom, G. A. Petersson, E. C. Barnes, M. J. Frisch, and J. A. Montgomery, Jr., *J. Chem. Phys.* **125**, 094106 (2006).
- <sup>94</sup>D. Feller, *Chem. Phys. Lett.* **322**, 543 (2000).
- <sup>95</sup>D. Feller and J. A. Sordo, *J. Chem. Phys.* **112**, 5604 (2000).
- <sup>96</sup>D. J. Grant, D. A. Dixon, J. S. Francisco, D. Feller, and K. A. Peterson, *J. Phys. Chem. A* **113**, 11343 (2009).
- <sup>97</sup>J. A. Montgomery, Jr., J. W. Ochterski, and G. A. Petersson, *J. Chem. Phys.* **101**, 5900 (1994).
- <sup>98</sup>D. Feller, K. A. Peterson, and J. G. Hill, *J. Chem. Phys.* **133**, 184102 (2010).
- <sup>99</sup>D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **100**, 2975 (1994).
- <sup>100</sup>D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **103**, 4572 (1995).
- <sup>101</sup>K. A. Peterson and T. H. Dunning, Jr., *J. Chem. Phys.* **117**, 10548 (2002).
- <sup>102</sup>J. Koput and K. A. Peterson, *J. Phys. Chem. A* **106**, 9595 (2002).
- <sup>103</sup>D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **98**, 1358 (1993).
- <sup>104</sup>A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning Jr., *J. Chem. Phys.* **110**, 7667 (1999).
- <sup>105</sup>N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.* **123**, 064107 (2005).
- <sup>106</sup>T. H. Dunning, Jr., K. A. Peterson, and A. K. Wilson, *J. Chem. Phys.* **114**, 9244 (2001).
- <sup>107</sup>MOLPRO, a package of *ab initio* programs designed by H.-J. Werner and P. J. Knowles, version 2010.1, F. R. Manby, M. Schütz, P. Celani *et al.*; see <http://www.molpro.net>.
- <sup>108</sup>DALTON, C. Angeli, K. L. Bak, V. Bakken, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, T. Enevoldsen, B. Fernandez, C. Hättig, K. Hald, A. Halkier, H. Heiberg, T. Helgaker, H. Hettema, H. J. A. Jensen, D. Jonsson, P. Jørgensen, S. Kirpekar, W. Klopper, R. Kobayashi, H. Koch, A. Ligabue, O. B. Lutnaes, K. V. Mikkelsen, P. Norman, J. Olsen, M. J. Packer, T. B. Pedersen, E. R. Z. Rinkevicius, T. A. Ruden, K. Ruud, P. Salek, A. S. d. Meras, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. O. Sylvester-Hvid, P. R. Taylor, O. Vahtras, D. J. Wilson, and H. Ågren, 2005.
- <sup>109</sup>M. Rittby and R. J. Bartlett, *J. Phys. Chem.* **92**, 3033 (1988).
- <sup>110</sup>G. E. Scuseria, *Chem. Phys. Lett.* **176**, 27 (1991).
- <sup>111</sup>P. J. Knowles, C. Hampel, and H.-J. Werner, *J. Chem. Phys.* **99**, 5219 (1994).
- <sup>112</sup>P. J. Knowles, C. Hampel, and H.-J. Werner, *J. Chem. Phys.* **112**, 3106 (2000).
- <sup>113</sup>J. G. Hill and J. A. Platts, *J. Chem. Theory Comput.* **5**, 500 (2009).
- <sup>114</sup>H.-J. Werner, G. Knizia, and F. R. Manby, *Mol. Phys.* **109**, 407 (2011).
- <sup>115</sup>E. C. Barnes, G. A. Petersson, D. Feller, and K. A. Peterson, *J. Chem. Phys.* **129**, 194115 (2008).
- <sup>116</sup>A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, *Chem. Phys. Lett.* **302**, 437 (1999).
- <sup>117</sup>N. J. DeYonker, T. G. Williams, A. E. Imel, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **131**, 024106 (2009).
- <sup>118</sup>A. Karton and J. M. L. Martin, *Theor. Chem. Acc.* **115**, 330 (2006).
- <sup>119</sup>F. Jensen, *Theor. Chem. Acc.* **113**, 187 (2005).
- <sup>120</sup>F. Jensen, *Theor. Chem. Acc.* **113**, 267 (2005).
- <sup>121</sup>S. Zhong, E. C. Barnes, and G. A. Petersson, *J. Chem. Phys.* **129**, 184116 (2008).
- <sup>122</sup>F. Neese and E. F. Valeev, *J. Chem. Theory Comput.* **7**, 33 (2011).
- <sup>123</sup>F. A. Bischoff and E. F. Valeev, *J. Chem. Phys.* **134**, 104104 (2011).
- <sup>124</sup>N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.* **125**, 074110 (2006).
- <sup>125</sup>B. Ruscic, R. E. Pinzon, M. L. Morton, G. V. Lazzewski, S. Bittner, S. G. Nijssure, K. A. Amin, M. Minkoff, and A. F. Wagner, *J. Phys. Chem. A* **108**, 9979 (2004).
- <sup>126</sup>See supplementary material at <http://dx.doi.org/10.1063/1.3613639> for the new basis sets and SCF and CCSD(T)(FC) energies.