

Effective basis set extrapolations for CCSDT, CCSDT(Q), and CCSDTQ correlation energies

Amir Karton

School of Molecular Sciences, The University of Western Australia, Perth, WA 6009,
Australia.

Cite as: A. Karton. J. Chem. Phys. 153, 024102 (2020).
<https://doi.org/10.1063/5.0011674>

ABSTRACT

It is well established that extrapolating the CCSD and (T) correlation energies using empirically motivated extrapolation exponents can accelerate the basis set convergence. Here we consider the extrapolation of coupled-cluster expansion terms beyond the CCSD(T) level to the complete basis set (CBS) limit. We obtain reference CCSDT–CCSD(T) (T_3 –(T)), CCSDT(Q)–CCSDT ((Q)), and CCSDTQ–CCSDT(Q) (T_4 –(Q)) contributions from cc-pV{5,6}Z extrapolations for a diverse set of 16 first- and second-row systems. We use these basis-set limit results to fit extrapolation exponents in conjunction with the cc-pV{D,T}Z, cc-pV{T,Q}Z, and cc-pV{Q,5}Z basis set pairs. The optimal extrapolation exponents result in noticeable improvements in performance (relative to $\alpha = 3.0$) in conjunction with the cc-pV{T,Q}Z basis set pair, however, smaller improvements are obtained for the other basis sets. These results confirm that the basis sets and basis set extrapolations used for obtaining post-CCSD(T) components in composite thermochemical theories such as Weizmann-4 and HEAT are sufficiently close to the CBS limit for attaining sub-kJ-per-mole accuracy. The fitted extrapolation exponents demonstrate that the T_3 –(T) correlation component converges more slowly to the CBS limit than the (Q) and T_4 terms. A systematic investigation of the effect of

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50011674

diffuse functions shows that it diminishes (i) in the order $T_{3-(T)} > (Q) > T_{4-(Q)}$ and (ii) with the size of the basis set. Importantly, we find that diffuse functions tend to systematically reduce the $T_{3-(T)}$ contribution but systematically increases the (Q) contribution. Thus, the use of the cc-pVnZ basis sets benefits from a certain degree of error cancellation between these two components.

*E-Mail: amir.karton@uwa.edu.au

1. Introduction

The solution to the nonrelativistic Schrödinger equation can be viewed as a two-dimensional problem in which both the one-particle space (“basis set completeness”) and n -particle space (“correlation treatment”) have to be converged to similar levels of accuracy. For example, in order to obtain thermochemical properties with benchmark accuracy (i.e., with 95% confidence intervals from accurate thermochemical data below ~ 1 kJ mol⁻¹) both spaces have to converge to a sub-kJ-per-mole level of accuracy.^{1,2,3,4,5,6,7,8,9,10} Coupled-cluster (CC) theory provides a road map for converging the n -particle space,^{11,12} where the base level is usually the so-called ‘gold standard’ of computational chemistry – the CCSD(T) method^{13,14} (i.e., coupled-cluster with single, double, and quasiperturbative triple excitations).^{2,6,7} However, it is well established that the CCSD(T) method cannot generally achieve benchmark accuracy for (i) systems dominated by moderate-to-severe non-dynamical correlation effects, and/or (ii) challenging thermochemical properties such as heats of formation.^{5,7,9,10,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33} In such cases post-CCSD(T) contributions, most importantly up to CCSDT(Q) or CCSDTQ,^{5,7,9,10,16,20,22,24} have to be added for achieving high accuracy. The shortcoming of the CCSD(T) method can be illustrated by examining the magnitude of post-CCSD(T) contributions to the heats of formation in the W4-17 database (see Table S1 of the supplementary material).¹⁶ Of the 200 highly accurate heats of formation in this representative database, post-CCSD(T) contributions exceed 0.24 kcal mol⁻¹ (or 1 kJ mol⁻¹) for 50% of the species, 0.5 kcal mol⁻¹ for 24% of the species, and 1 kcal mol⁻¹ for 7% of the species. Inspections of species with sizeable post-CCSD(T) contributions reveals that, in addition to multireference systems, they include challenging systems such as radicals and molecules with highly polar and/or multiple bonds.

The basis set convergence of the CCSD and CCSD(T) correlation energies has been extensively investigated in conjunction with the correlation-consistent basis sets by Dunning

and co-workers.^{2,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56} Owing to the systematic convergence of these basis sets, basis set extrapolations are an effective approach for converging the n -particle space while keeping the computational cost at a minimum. However, the optimal extrapolation parameters may depend on the correlation component being extrapolated (e.g., CCSD or (T)) and the specific basis sets being used in the extrapolation.^{48,55} A number of extensive works have obtained optimal extrapolation exponents for the CCSD and (T) correlation components in conjunction with the correlation-consistent basis sets,^{2,34,35,38,46,41,48,49,53,54} where reference data at the infinite basis set limit is obtained from explicitly correlated (R12 or F12) techniques or extrapolated from very large Gaussian basis sets.

Less attention has been devoted to obtaining optimal extrapolation exponents for post-CCSD(T) correlation contributions. Martin and co-workers extrapolated the CCSDT–CCSD(T) (T_3 –(T)) correlation component from the cc-pVTZ and cc-pVQZ basis sets for a set of 15 diatomic molecules and water.²² They found that, relative to cc-pV{Q,5}Z (or where available cc-pV{5,6}Z) reference data, using an optimized extrapolation exponent of $\alpha = 2.5$ results in a root-mean-square deviation (RMSD) of 0.013 kcal mol⁻¹. This RMSD is nearly half of that obtained with the asymptotic exponent of $\alpha = 3.0$,⁵⁷ for which an RMSD of 0.021 kcal mol⁻¹ is obtained. These results show that extrapolating the T_3 –(T) correlation component using effective decay exponents can significantly accelerate the basis set convergence.

Over the past three decades, there has been a proliferation of basis set extrapolation formulas for the correlation energy. For a historical overview of the various basis set extrapolations see for example refs. 34, 41, and 43. Here we will primarily focus on the two-point $A + B \times L^{-3}$ expression of Halkier *et al.*,⁵⁸ which is motivated by the partial-wave expansion of pair correlation energies in helium-like atoms^{57,58,59} and is widely used in

composite thermochemical protocols, such as the $Wn^{2,7,15,20,23,27,53}$ Wn -F12,^{16,42} and HEAT.^{21,25,26,60} This leads to the following expression for the extrapolated energy (E_∞):

$$E_\infty = E(L + 1) + \frac{E(L+1) - E(L)}{\left(\frac{L+1}{L}\right)^\alpha - 1} \quad (1)$$

where L is the highest angular momentum present in the basis set. We will use this formula with the asymptotic extrapolation exponent of $\alpha = 3.0$ and with an effective exponent α_{eff} which is optimized to minimize the RMSD for a given coupled-cluster expansion term over our test set. Schwenke proposed rearranging Eq. 1 to a simplified two-point linear extrapolation which does not explicitly involve L :

$$E_\infty = E(L) + F[E(L + 1) - E(L)] \quad (2)$$

where F is an empirical scaling factor, which is related to α via the following expression:

$$F = 1 + \frac{1}{\left(\frac{L+1}{L}\right)^\alpha - 1} \quad (3)$$

In the present work we further explore this concept for higher-level correlation components CCSDT(Q)-CCSDT ((Q)) and CCSDTQ-CCSDT(Q) (T₄-(Q)) for which effective extrapolation coefficients have not been previously obtained. We also re-examine the previous results²² for the T₃-(T) correlation contribution against the larger training set considered in this work. We begin by calculating the T₃-(T), (Q), and T₄-(Q) correlation contributions to the total atomization energies (TAEs) near the one-particle basis set limit (i.e., extrapolated from the cc-pV{5,6}Z basis set pair) for a representative set of 16 molecules: BH₃, CH, CH₂(¹A₁), CH₃, OH, H₂O, HF, AlH₃, H₂S, HCl, B₂, C₂(¹Σ⁺), CO, N₂, CS, and P₂. Our set of molecules includes first and second-row hydride and non-hydride systems with 2–4 atoms, including closed-shell singlet, singlet diradical, radical, and triplet systems. Using these basis set limit values we obtain effective decay exponent for the T₃-(T), (Q), and T₄-(Q) correlation

components in conjunction with the cc-pV{D,T}Z, cc-pV{T,Q}Z, and cc-pV{Q,5}Z basis set pairs. We also address questions such as:

- How does the effective convergence rate vary between the T_3 -(T), (Q), and \hat{T}_4 correlation components?
- How do the effective decay exponents compare with the asymptotically limiting L^{-3} convergence behavior?
- How does the effective convergence rate differ between hydride and nonhydride systems?
- To what extent does the addition of diffuse functions affect the post-CCSD(T) contributions?

2. Computational Methods

All calculations were carried out using the MRCC program suite^{61,62} with the standard correlation-consistent basis sets of Dunning and co-workers.^{63,64,65,66} For the sake of brevity, the cc-pVnZ basis sets ($n = D, T, Q, 5, 6$) are denoted by VnZ. We also consider the aug'-cc-pVnZ basis sets, which combine the cc-pVnZ basis sets on H and the aug-cc-pVnZ basis sets on first- and second-row atoms (denoted by A'VnZ).⁶⁷ Basis set extrapolations using the VnZ and V($n+1$)Z basis sets are denoted by V{ $n,n+1$ }Z. All calculations are nonrelativistic and are carried out within the frozen-core approximation, i.e., the 1s orbitals for first-row atoms and the 1s, 2s, and 2p orbitals for second-row atoms are constrained to be doubly occupied in all configurations. All the geometries were optimized at the CCSD(T)/cc-pV(Q+d)Z level⁶⁸ of theory and were taken from the W4-17 database¹⁶ (the geometries are given in Table S2 of the supplementary material). This level of theory has been shown to yield geometries that are in close agreement with CCSD(T) geometries near the complete basis set (CBS) limit, e.g., with mean absolute deviations of 0.001 Å from CCSD(T)/cc-pV(6+d)Z geometries.⁶⁹

In the present study we examine the basis set convergence of post-CCSD(T) contributions to molecular atomization energies near the one-particle basis set limit. We consider iterative and perturbative coupled cluster contributions up to connected quadruples (CCSDTQ) for a set of 16 first- and second- row molecules with up to four atoms. Table 1 lists the systems considered here along with the %TAE[(T)] diagnostics for nondynamical correlation effects.^{7,10,24} The %TAE[(T)] diagnostics is defined as the percentage of the total atomization energy accounted for by parenthetical connected triple excitations and has been found to be a reliable energy-based diagnostic for the importance of non-dynamical correlation effects.^{7,10,24} The %TAE[(T)] values suggest that the chosen set of molecules spans the gamut from systems dominated by dynamical correlation (e.g., H₂O and HF), moderate nondynamical correlation (e.g., N₂ and P₂), and strong nondynamical correlation (e.g., B₂ and C₂).^{7,10,24}

Table 1. Overview of the molecules considered in the present work with %TAE[(T)] diagnostics for nondynamical correlation effects.

Name	%TAE[(T)]
BH ₃	0.3
HC•	1.0
H ₂ C(¹ A ₁)	1.0
H ₃ C•	0.6
HO•	1.6
H ₂ O	1.5
HF	1.5
AlH ₃	0.1
H ₂ S	1.2
HCl	1.4
B ₂ (³ Σ _g)	14.7
C ₂ (¹ Σ _g ⁺)	13.3
CO	3.1
N ₂	4.2
CS	5.6
P ₂	8.3

Table 2 gives an overview of the coupled cluster excitations that are considered in the present work along with the abbreviations that are used. In particular, we consider the following

post-CCSD(T) contributions: CCSDT–CCSD(T) (T_3 –(T)), CCSDT(Q)–CCSDT ((Q)), CCSDTQ–CCSDT(Q) (T_4 –(Q)), and CSDTQ–CCSDT (T_4).

Table 2. Overview of the post-CCSD(T) contributions discussed in the present work.

Name	Definition	Abbreviation
Higher-order connected triples	CCSDT–CCSD(T)	T_3 –(T)
Noniterative connected quadruples	CCSDT(Q)–CCSDT	(Q)
Higher-order connected quadruples	CCSDTQ–CCSDT(Q)	T_4 –(Q)
Iterative connected quadruples	CCSDTQ–CCSDT	T_4

Finally, we note that the present work represents an extensive computational effort for obtaining basis set limit values for post-CCSD(T) contributions for systems with 2–4 atoms. We note that many of the calculations reported here strained our computational resources to the absolute limit. For example, the fully iterative CCSDTQ/V6Z calculations involve 0.7×10^9 (CH), 2.0×10^9 (B₂), 4.7×10^9 (OH), 8.3×10^9 (CH₂), 9.3×10^9 (HF and HCl), and 1.0×10^{10} (C₂) amplitudes. These calculations ran on dual Intel Xeon machines with up to 1024 GB of RAM for 3 (CH), 16 (HF), 22 (HCl), 24 (OH), 28 (B₂), 42 (CH₂), and 46 (C₂) days.

3. Results and Discussion

3.1 Reference CBS limit values. Table 3 lists the T_3 –(T), (Q), and T_4 –(Q), and T_4 CBS reference values used throughout this work for the parameterization of basis set extrapolation exponents. The T_3 –(T) and (Q) contributions are extrapolated to the infinite basis set limit from the V{5,6}Z basis set pair. For seven systems (CH, CH₂, OH, HF, HCl, B₂, and C₂) we were able to extrapolate the T_4 –(Q) contribution from the V{5,6}Z basis set pair. For these seven systems, the V{Q,5}Z extrapolation results in an RMSD of merely 0.0006 kcal mol^{–1} relative to the V{5,6}Z basis set limit values. Where the largest deviation being of 0.0010 kcal mol^{–1} is obtained for B₂. In light of these results, our best T_4 –(Q) reference values are calculated with

the V{5,6}Z basis set pair for the above seven systems and with the V{Q,5}Z basis set pair for the rest of the systems.

Table 3. Complete basis set limit reference values for the T_3 –(T), (Q), T_4 –(Q), and T_4 contributions to total atomization energies (TAEs) extrapolated from the V{5,6}Z basis set pair using eq. (1) with an asymptotic extrapolation exponent of $\alpha = 3.0$ (in kcal mol⁻¹).^a

Molecule	T_3 –(T)	(Q)	T_4 –(Q) ^a	\hat{T}_4 ^b
BH ₃	-0.024	0.035	[0.004]	0.038
CH	0.100	0.034	0.003	0.037
CH ₂ (¹ A ₁)	0.180	0.117	0.012	0.129
CH ₃	-0.037	0.064	[-0.001]	0.063
OH	-0.035	0.097	-0.007	0.090
H ₂ O	-0.243	0.231	[-0.025]	0.206
HF	-0.165	0.129	-0.015	0.114
AlH ₃	-0.016	0.030	[0.006]	0.036
H ₂ S	-0.110	0.166	[-0.002]	0.164
HCl	-0.141	0.117	-0.002	0.115
B ₂	0.077	1.253	-0.012	1.242
C ₂ (¹ Σ ⁺)	-2.291	3.426	-1.150	2.276
CO	-0.581	0.727	[-0.101]	0.626
N ₂	-0.778	1.197	[-0.176]	1.021
CS	-0.663	1.153	[-0.134]	1.019
P ₂	-0.974	1.649	[-0.189]	1.460

^aValues in square brackets are extrapolated from the V{Q,5}Z basis set pair. For the seven systems for which we have both V{Q,5}Z and V{5,6}Z values the RMSD between the two amounts to merely 0.0006 kcal mol⁻¹ (see text). ^bObtained from the values in the (Q) and T_4 –(Q) columns.

The T_3 –(T), (Q), and T_4 –(Q) reference values in Table 3 are extrapolated to the infinite basis set limit using eq. (1) where in all cases an asymptotic extrapolation exponent of $\alpha = 3.0$ has been used. Apart from the T_4 –(Q) component for the larger systems, we were able to obtain the reference values from the V{5,6}Z basis set pair. Nevertheless, an optimal empirical exponent for this basis set pair is still likely to deviate from an exponent of $\alpha = 3.0$ and possibly to vary between each of the correlation contributions (T_3 –(T), (Q), and T_4 –(Q)). It is therefore of interest to estimate by how much the V{5,6}Z CBS values change by varying the exponent from the asymptotic value? A practical approach for examining the effect of using $\alpha = 3$ in the

$V\{5,6\}Z$ extrapolations (rather than an optimal exponent), is to use the optimal exponents for the $V\{Q,5\}Z$ extrapolations: 2.7342 ($T_3\text{--}(T)$), 3.4216 ((Q)), and 3.1381 ($T_4\text{--}(Q)$) (Table 5, *vide infra*). The effect of varying the exponents for the $V\{5,6\}Z$ extrapolations for each of the post-CCSD(T) correlation contributions is presented in Table S3 of the Supporting Information. For the $T_3\text{--}(T)$ component changing α from 3.0 to 2.734 results in changes in the CBS values ranging from 0.0003 (CH) to 0.007 (C_2) kcal mol⁻¹. These changes represent a fraction of a percent of the $T_3\text{--}(T)/V\{5,6\}Z$ CBS values listed in Table 3. For the (Q) component changing α from 3.0 to 3.422 results in changes in the CBS values ranging from 0.0001 (AlH_3) to 0.004 (P_2) kcal mol⁻¹. Again, these changes represent a fraction of a percent of the (Q)/ $V\{5,6\}Z$ CBS values listed in Table 3. The changes for the $T_4\text{--}(Q)$ component are smaller than 0.001 kcal mol⁻¹. These small changes in the extrapolated $V\{5,6\}Z$ CBS values are expected since the effect of the extrapolation exponent on the extrapolated energy becomes less pronounced as we approach the infinite basis set limit (or as $E(L) \rightarrow E(L+1)$ in eqs. (1) and (2)).

It is of interest to examine the magnitude of the basis set limit $T_3\text{--}(T)$ and T_4 contributions (Table 3). It is well known that the $T_3\text{--}(T)$ contribution tends to universally decrease the TAEs, whereas the T_4 contribution tends to increase them.^{7,10,15,16,23,24,25} With three exceptions (B_2 , C_2 , and CH_2), the $T_3\text{--}(T)$ contribution decreases the TAEs by amounts ranging from -0.016 (AlH_3) to -2.291 (C_2) kcal mol⁻¹. With no exception, the T_4 contribution increases the TAEs by amounts ranging from +0.036 (AlH_3) to +2.276 (C_2) kcal mol⁻¹. Overall, due to error cancellation between the $T_3\text{--}(T)$ and T_4 components, post-CCSD(T) contributions tend to increase the atomization energies by up to 0.486 (P_2) kcal mol⁻¹. A striking exception to this is B_2 , for which there is no error cancellation between the $T_3\text{--}(T)$ and T_4 components, and as a result post-CCSD(T) contributions amount to as much as +1.318 kcal mol⁻¹.

Inspection of the CBS values in Table 3 reveals that the magnitude of the $T_3\text{--}(T)$, (Q), and $T_4\text{--}(Q)$ components are systematically larger for non-hydride systems than for hydride

systems. The magnitude of the T_3 -(T) component (in absolute value) ranges between 0.016–0.243 kcal mol⁻¹ for the hydride systems and between 0.077–2.291 kcal mol⁻¹ for the non-hydride systems. The (Q) component ranges between 0.030–0.231 (hydrides) and 0.727–3.426 (non-hydrides) kcal mol⁻¹. The T_4 -(Q) component ranges between 0.001–0.025 (hydrides) and 0.012–1.150 (non-hydrides) kcal mol⁻¹.

3.2 Overview of basis set convergence of post-CCSD(T) excitations. Figure 1 depicts the overall RMSDs for the V_nZ basis sets ($n = D-6$) for the T_3 -(T), (Q), and T_4 -(Q) components. This plot demonstrates the decrease in RMSD for successively higher coupled-cluster expansion terms. In particular, for any given basis set size (D, T, Q, 5, and 6), the RMSD for the (Q) component is roughly half of that for the T_3 -(T) component, and the RMSD for the T_4 -(Q) component is roughly one tenth of that for the T_3 -(T) component.

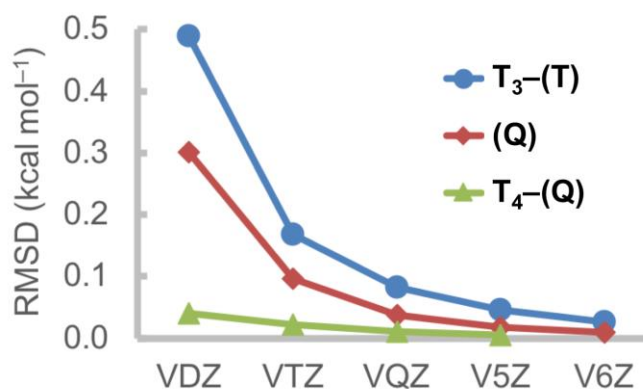


Figure 1. Basis set truncation errors for the T_3 -(T), (Q), and T_4 -(Q) components calculated in conjunction with the V_nZ basis sets ($n = D, T, Q, 5, 6$). The tabulated values are RMSDs over the entire set of 16 molecules relative to the basis set limit values in Table 3. The RMSDs are given in Table 4.

Table 4 lists the RMSDs for the VnZ basis sets ($n = D-6$) for the $T_3-(T)$, (Q), $T_4-(Q)$, and T_4 components. These RMSDs show that the (Q) component converges somewhat faster than the $T_3-(T)$ component. It is instructive to look at the ratios between the RMSDs for consecutive basis sets, i.e., $\text{RMSD}[V(n+1)Z]/\text{RMSD}[VnZ]$ (Table 4). For the $T_3-(T)$ component, these ratios are 0.34 (VTZ/VDZ), 0.49 (VQZ/VTZ), 0.55 (V5Z/VQZ), and 0.58 (V6Z/V5Z). For the (Q) component, these ratios are somewhat smaller indicating faster convergence, namely they are 0.32 (VTZ/VDZ), 0.38 (VQZ/VTZ), 0.47 (V5Z/VQZ), and 0.58 (V6Z/V5Z). Similar conclusions regarding the faster convergence of the (Q) component are drawn from examining the optimized extrapolation exponents in Table 5 (*vide infra*). For the $T_4-(Q)$ component the ratio between the RMSDs for consecutive basis sets is nearly constant at ~ 0.5 (Table 4). Namely, the RMSD is roughly halved with each increase in the highest angular momentum present in the basis set.

Table 4. RMSDs (in kcal mol⁻¹) over the entire set of 16 molecules relative to the basis set limit values in Table 3 for the $T_3-(T)$, (Q), $T_4-(Q)$, and T_4 components calculated in conjunction with the VnZ basis sets ($n = D, T, Q, 5, 6$).

Basis set	$T_3-(T)$	(Q)	$\hat{T}_4-(Q)$	\hat{T}_4
VDZ	0.490	0.301	0.040	0.265
VTZ	0.168	0.097	0.022	0.077
VQZ	0.083	0.037	0.011	0.027
V5Z	0.045	0.017	0.005	0.012
V6Z	0.026	0.010		
	RMSD Ratios ^a			
VTZ/VDZ	0.34	0.32	0.55	0.29
VQZ/VTZ	0.49	0.38	0.50	0.35
V5Z/VQZ	0.55	0.47	0.45	0.44
V6Z/V5Z	0.58	0.58		

^aRatio between the RMSDs calculated with the $V(n+1)Z$ and VnZ basis sets.

3.3 Effective exponents for higher-order connected triple excitations. Table 5 summarizes the optimal exponents for the post-CCSD(T) contributions ($T_3-(T)$, (Q), $T_4-(Q)$, and T_4). Table 6 lists deviations for the VnZ basis sets ($n = D-6$) and $V\{D,T\}Z$, $V\{T,Q\}Z$, and $V\{Q,5\}Z$ basis

set extrapolations from the basis set limit $T_{3-(T)}/V\{5,6\}Z$ values listed in Table 3. The $T_{3-(T)}$ component converges smoothly but slowly to the CBS limit. For example, the following RMSDs are obtained relative to the CBS reference values: 0.490 (VDZ), 0.168 (VTZ), 0.083 (VQZ), 0.045 (V5Z), and 0.026 (V6Z) kcal mol⁻¹ (Table 6). Inspection of the individual errors for each of the molecules reveals that, with the exception of BH₃ and AlH₃, all systems exhibit monotonic basis set convergence. For BH₃ and AlH₃ the VDZ basis set exhibits anomalous basis set convergence, which results in large errors for the V{D,T}Z extrapolation. For both systems, however, the basis set convergence becomes monotonic from the VTZ basis set onwards.

Table 5. Optimal exponents for the $A + B \times L^{-\alpha}$ extrapolation (eq. 1) and scaling factors (F) for the Schwenke-type linear extrapolation (eq. 2) for the $T_{3-(T)}$, (Q), $T_{4-(Q)}$, and \hat{T}_4 contributions.

Basis set	$T_{3-(T)}$	(Q)	$T_{4-(Q)}$	T_4
$A + B \times L^{-\alpha}$ extrapolation (eq. 1)				
V{D,T}Z	2.7174	2.9968	1.7139	3.3199
V{T,Q}Z	2.4807	3.3831	2.6072	3.6401
V{Q,5}Z	2.7342	3.4216	3.1381 ^a	4.0408 ^a
Schwenke-type extrapolation (eq. 2)				
V{D,T}Z	1.4976	1.4218	1.9964	1.3518
V{T,Q}Z	1.9602	1.6073	1.8952	1.5407
V{Q,5}Z	2.1896	1.8728	1.9860 ^a	1.6832 ^a

^aObtained for a subset of seven systems for which we have both V{Q,5}Z and V{5,6}Z values.

Table 6. Convergence of the T_3 –(T) contribution to the total atomization energy for the set of 16 molecules relative to basis set limit values from V{5,6}Z extrapolations (deviations and error statistics are given in kcal mol⁻¹).

Basis set	VDZ	VTZ	VQZ	V5Z	V6Z	V{D,T}Z	V{D,T}Z	V{T,Q}Z	V{T,Q}Z	V{Q,5}Z	V{Q,5}Z
α^a						3.0000	2.7174 ^c	3.0000	2.4807 ^c	3.0000	2.7342 ^c
F^b						1.4211	1.4976 ^c	1.7297	1.9602 ^c	2.0492	2.1896 ^c
BH ₃	0.016	0.023	0.010	0.005	0.003	0.025	0.026	0.001	-0.001	0.000	-0.001
CH	0.025	0.012	0.006	0.004	0.002	0.007	0.006	0.002	0.001	0.001	0.001
CH ₂ (¹ A ₁)	0.075	0.038	0.020	0.011	0.006	0.022	0.019	0.007	0.003	0.002	0.000
CH ₃	0.095	0.040	0.018	0.010	0.006	0.017	0.013	0.002	-0.003	0.001	0.000
OH	0.085	0.038	0.019	0.010	0.006	0.018	0.015	0.005	0.000	0.001	-0.001
H ₂ O	0.274	0.108	0.051	0.025	0.014	0.039	0.026	0.010	-0.003	-0.003	-0.007
HF	0.174	0.072	0.033	0.016	0.009	0.029	0.021	0.005	-0.004	-0.002	-0.004
AlH ₃	0.026	-0.019	-0.011	-0.009	-0.005	-0.038	-0.042	-0.006	-0.004	-0.007	-0.007
H ₂ S	0.206	0.059	0.025	0.017	0.010	-0.003	-0.015	0.001	-0.006	0.007	0.006
HCl	0.163	0.055	0.023	0.014	0.008	0.010	0.002	-0.001	-0.008	0.005	0.003
B ₂	0.563	0.282	0.140	0.073	0.042	0.164	0.142	0.036	0.004	0.003	-0.007
C ₂ (¹ Σ^+)	1.048	0.379	0.185	0.097	0.056	0.097	0.045	0.043	-0.002	0.004	-0.008
CO	0.603	0.193	0.090	0.045	0.026	0.020	-0.011	0.014	-0.009	-0.002	-0.008
N ₂	0.660	0.195	0.095	0.051	0.030	0.000	-0.036	0.022	-0.001	0.005	-0.001
CS	0.775	0.249	0.121	0.071	0.041	0.028	-0.012	0.028	-0.002	0.018	0.011
P ₂	0.906	0.252	0.132	0.082	0.048	-0.023	-0.073	0.044	0.016	0.030	0.023
RMSD _{all} ^{d,e}	0.490	0.168	0.083	0.045	0.026	0.052	0.046	0.021	0.006	0.010	0.008
MAD _{all} ^{d,e}	0.356	0.126	0.061	0.034	0.020	0.034	0.032	0.014	0.004	0.006	0.006
MSD _{all} ^{d,e}	0.356	0.124	0.060	0.033	0.019	0.026	0.008	0.013	-0.001	0.004	0.000
RMSD _{hyd} ^{d,f}	0.141	0.054	0.025	0.013	0.008	0.024	0.021	0.005	0.004	0.004	0.004
RMSD _{nonhyd} ^{d,g}	0.779	0.266	0.131	0.072	0.042	0.079	0.070	0.033	0.008	0.015	0.012

^aExtrapolation exponent (α) used in two-point $L^{-\alpha}$ extrapolation eq. (1). ^bExtrapolation coefficient (F) used in two-point Schwenke extrapolation eq. (2). ^cUsing empirical α and F parameters optimized to minimize the RMSD. ^dError statistics with respect to the V{5,6}Z reference values in Table 3. RMSD = root-mean-square deviation, MAD = mean absolute deviation, MSD = mean signed deviation. ^eOver all species. ^fOver hydride species. ^gOver non-hydride species.

The two-point extrapolation formula with $\alpha = 3.0$ results in RMSDs of 0.052 (V{D,T}Z), 0.021 (V{T,Q}Z), and 0.010 (V{Q,5}Z) kcal mol⁻¹. Minimizing these RMSDs by varying the extrapolation exponent results in modest improvements in the RMSDs for the V{D,T}Z and V{Q,5}Z basis set pairs, namely the RMSDs are reduced by 0.006 and 0.002 kcal mol⁻¹, respectively (Table 6). However, the RMSD for the V{T,Q}Z basis set pair is reduced from 0.021 (with $\alpha = 3.0$) to merely 0.006 kcal mol⁻¹ (with $\alpha = 2.4807$ or equivalently $F = 1.9602$). This empirical extrapolation exponent is in excellent agreement with the empirical extrapolation exponent of $\alpha = 2.5$ obtained previously using T_3 –(T)/V{Q,5}Z (and where available V{5,6}Z) reference data.²² We note that in contrast to the optimized extrapolation exponents, which do not vary monotonically along the series V{D,T}Z \rightarrow V{T,Q}Z \rightarrow

$V\{Q,5\}Z$, the Schwenke the extrapolation parameter clearly reflects the monotonic convergence of the extrapolation parameters along this series (Table 5).

The bottom two lines of Table 6 give the RMSDs over the subset of 10 hydride (RMSD_{hyd}) and 6 non-hydride (RMSD_{nonhyd}) systems. Inspection of these RMSDs reveals that for all the VnZ basis sets ($n = D-6$), the RMSDs over the hydride systems (RMSD_{hyd}) are consistently about ~ 5 times smaller than the RMSDs over the non-hydride systems (RMSD_{nonhyd}). Thus, the basis set convergence for the hydride systems is significantly faster than for non-hydride systems. The stark difference between the basis set convergence of the $T_3-(T)$ component for the hydride and non-hydride systems is depicted in Figure 2a.

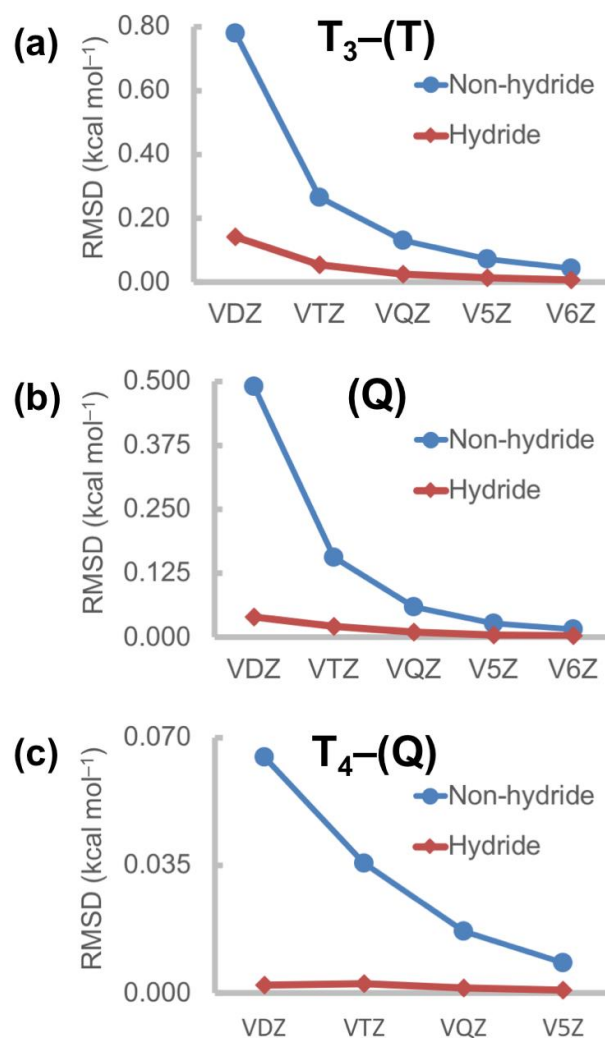


Figure 2. Basis set truncation errors for the T_3 -(T), (Q), and T_4 -(Q) components calculated in conjunction with the V_nZ basis sets ($n = D, T, Q, 5, 6$). The tabulated values are RMSDs from the basis set limit values in Table 3.

3.4 Effective exponents for noniterative connected quadruples excitations. Table 7 lists deviations for the V_nZ basis sets ($n = D-6$) and $V\{D,T\}Z$, $V\{T,Q\}Z$, and $V\{Q,5\}Z$ basis set extrapolations from the basis set limit (Q)/ $V\{5,6\}Z$ values listed in Table 3. As discussed in Section 3.2, the (Q) component converges somewhat faster to the CBS limit than the T_3 -(T) component. The following RMSDs are obtained relative to the CBS reference values: 0.301 (VDZ), 0.097 (VTZ), 0.037 (VQZ), 0.017 (V5Z), and 0.010 (V6Z) kcal mol⁻¹ (Table 7).

Table 7. Convergence of the (Q) contribution to the total atomization energy for the set of 16 molecules relative to basis set limit values from V{5,6}Z extrapolations (deviations and error statistics are given in kcal mol⁻¹).

Basis set	VDZ	VTZ	VQZ	V5Z	V6Z	V{D,T}Z	V{D,T}Z	V{T,Q}Z	V{T,Q}Z	V{Q,5}Z	V{Q,5}Z
α^a						3.0000	2.9968 ^c	3.0000	3.3831 ^c	3.0000	3.4216 ^c
F^b						1.4211	1.4218 ^c	1.7297	1.6073 ^c	2.0492	1.8728 ^c
BH ₃	0.013	-0.008	-0.002	-0.001	0.001	-0.006	-0.006	0.002	0.001	0.000	0.000
CH	0.008	-0.007	-0.003	-0.001	0.001	-0.007	-0.007	0.000	-0.001	0.000	0.000
CH ₂ (¹ A ₁)	0.035	-0.019	-0.008	-0.004	0.002	-0.013	-0.013	0.001	-0.001	0.001	0.000
CH ₃	0.006	-0.015	-0.005	-0.002	0.001	-0.019	-0.019	0.002	0.001	0.001	0.000
OH	0.016	-0.019	-0.009	-0.004	0.002	-0.034	-0.034	-0.002	-0.004	0.001	0.000
H ₂ O	0.031	-0.039	-0.018	-0.008	0.005	-0.069	-0.069	-0.003	-0.005	0.003	0.001
HF	0.060	-0.020	-0.009	-0.004	0.002	-0.053	-0.053	-0.002	-0.003	0.001	0.000
AlH ₃	0.014	-0.007	-0.002	-0.001	0.001	-0.003	-0.003	0.001	0.000	0.000	0.000
H ₂ S	0.082	-0.033	-0.016	-0.008	0.005	-0.013	-0.013	-0.003	-0.006	0.000	-0.002
HCl	0.050	-0.025	-0.013	-0.006	0.004	-0.014	-0.014	-0.004	-0.005	0.000	-0.001
B ₂	0.345	-0.091	-0.034	-0.014	0.008	0.016	0.017	0.008	0.001	0.006	0.003
C ₂ (¹ Σ^+)	0.770	-0.206	-0.074	-0.032	0.019	0.032	0.032	0.022	0.006	0.012	0.004
CO	0.093	-0.075	-0.027	-0.012	0.007	-0.067	-0.067	0.009	0.003	0.004	0.001
N ₂	0.169	-0.108	-0.040	-0.019	0.011	-0.082	-0.082	0.009	0.001	0.003	-0.001
CS	0.563	-0.175	-0.071	-0.034	0.020	-0.012	-0.012	0.005	-0.008	0.005	-0.001
P ₂	0.609	-0.218	-0.082	-0.041	0.024	-0.054	-0.053	0.017	0.001	0.003	-0.005
RMSD_{all}^{d,e}	0.301	0.097	0.037	0.017	0.010	0.040	0.040	0.008	0.004	0.004	0.002
MAD_{all}^{d,e}	0.179	0.067	0.026	0.012	0.007	0.031	0.031	0.006	0.003	0.003	0.001
MSD_{all}^{d,e}	0.166	-0.067	-0.026	-0.012	0.007	-0.025	-0.025	0.004	-0.001	0.003	0.000
RMSD_{hyd}^{d,f}	0.040	0.022	0.010	0.005	0.003	0.031	0.031	0.002	0.003	0.001	0.001
RMSD_{nonhyd}^{d,g}	0.489	0.156	0.059	0.028	0.016	0.051	0.051	0.013	0.004	0.006	0.003

^aExtrapolation exponent (α) used in two-point $L^{-\alpha}$ extrapolation eq. (1). ^bExtrapolation coefficient (F) used in two-point Schwenke extrapolation eq. (2). ^cUsing empirical α and F parameters optimized to minimize the RMSD.

^dError statistics with respect to the V{5,6}Z reference values in Table 3. RMSD = root-mean-square deviation, MAD = mean absolute deviation, MSD = mean signed deviation. ^eOver all species. ^fOver hydride species. ^gOver non-hydride species.

The two-point extrapolation formula with $\alpha = 3.0$ results in RMSDs of 0.040 (V{D,T}Z), 0.008 (V{T,Q}Z), and 0.004 (V{Q,5}Z) kcal mol⁻¹. Interestingly, the optimal exponent for the V{D,T}Z extrapolation ($\alpha = 2.9968$ or equivalently $F = 1.4218$) is practically

identical to the ideal exponent. Minimizing the RMSDs by varying the extrapolation exponents for the V{T,Q}Z and V{Q,5}Z basis set pairs results in an extrapolation exponent of ~ 3.4 in both cases. In particular, we obtain $\alpha = 3.3831$ or equivalently $F = 1.6073$ for the V{T,Q}Z extrapolation and $\alpha = 3.4216$ or equivalently $F = 1.8728$ for the V{Q,5}Z extrapolation. In both cases, varying the extrapolation exponent from 3.0 to ~ 3.4 halves the RMSD, and results in RMSDs of merely 0.004 and 0.002 kcal mol⁻¹ for the V{T,Q}Z and V{Q,5}Z extrapolations, respectively (Table 7).

In computationally economical post-CCSD(T) methods (e.g., W4lite theory and HEAT-345(Q))^{24,25} T₄ term is simply calculated as (Q)/VDZ. In W4lite theory this term is scaled by 1.1 to compensate for basis set incompleteness and the complete neglect of the T₄-(Q) component. As mentioned above, the RMSD for the (Q)/VDZ term relative to our (Q)/V{5,6}Z values is 0.301 kcal mol⁻¹, scaling the (Q)/VDZ term by a scaling factor of 4/3 optimized to minimize the RMSD results in an RMSD of 0.136 kcal mol⁻¹. Thus, for large systems, for which the (Q)/VTZ calculation is not feasible, we recommend scaling the (Q)/VDZ term by a scaling factor of 4/3. In case the (Q)/VDZ term is used for approximating the overall T₄/CBS energy, we find a slightly lower optimal scaling factor of 1.25, which results in an RMSD of 0.129 kcal mol⁻¹ relative to the T₄/CBS values. We note that the highly multireference C₂ molecule has been excluded from this parameterization.

Inspection of the RMSDs over the subset of 10 hydride (RMSD_{hyd}) and 6 non-hydride (RMSD_{nonhyd}) systems in Table 7 reveals that the latter are 5–12 times larger than the former. Thus, similarly to the T₃-(T) component, basis set convergence of the (Q) component for the hydride systems is significantly faster than for non-hydride systems (Figure 2b). However, in contrast to the T₃-(T) component where RMSD_{nonhyd}/RMSD_{hyd} ≈ 5 for all basis sets, for the (Q) component this ratio is systematically reduced with the size of the basis set. In particular, we obtain RMSD_{nonhyd}/RMSD_{hyd} = 12.2 (VDZ), 7.1 (VTZ), 5.9 (VQZ), 5.6 (V5Z), and 5.3 (V6Z).

3.5 Effective exponents for higher-order connected quadruple excitations. Table 5 summarizes the optimal exponents for the $\hat{T}_4\text{--}(Q)$ component. Table 8 lists deviations for the VnZ basis sets ($n = D-5$) and the $V\{D,T\}Z$ and $V\{T,Q\}Z$ basis set extrapolations from the basis set limit $\hat{T}_4\text{--}(Q)/V\{5,6\}Z$ (or $\hat{T}_4\text{--}(Q)/V\{Q,5\}Z$) values listed in Table 3. As briefly discussed in Section 3.3, the $T_4\text{--}(Q)$ component converges smoothly to the basis set limit with a reduction of the RMSD by 50% with each increase in the highest angular momentum present in the basis set. In particular, we obtain the following RMSDs: 0.040 (VDZ), 0.022 (VTZ), 0.011 (VQZ), and 0.005 (V5Z) kcal mol⁻¹ (Table 8). The two-point extrapolation formula with $\alpha = 3.0$ results in RMSDs of 0.016 ($V\{D,T\}Z$) and 0.003 ($V\{T,Q\}Z$) kcal mol⁻¹. Minimizing these RMSDs by varying the extrapolation exponent results in relatively small improvements in the RMSDs, namely the RMSDs are reduced to 0.011 ($V\{D,T\}Z$) and 0.002 ($V\{T,Q\}Z$) (Table 8).

Table 8. Convergence of the $T_4\text{--}(Q)$ contribution to the total atomization energy for the set of 16 molecules relative to basis set limit values from $V\{5,6\}Z$ or $V\{Q,5\}Z$ extrapolations (see Table 3) (deviations and error statistics are given in kcal mol⁻¹).

Basis set	VDZ	VTZ	VQZ	V5Z	$V\{D,T\}Z$	$V\{D,T\}Z$	$V\{T,Q\}Z$	$V\{T,Q\}Z$
α^a					3.0000	1.7139	3.0000	2.6072
F^b					1.4211	1.9964	1.7297	1.8952
BH ₃	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000
CH	0.000	0.001	0.000	0.000	0.001	0.001	0.000	0.000
CH ₂ (¹ A ₁)	0.002	0.003	0.002	0.001	0.004	0.005	0.001	0.001
CH ₃	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
OH	0.007	0.002	0.002	0.001	-0.001	-0.004	0.002	0.002
H ₂ O	0.000	0.003	0.003	0.001	0.004	0.006	0.002	0.002
HF	-0.001	-0.001	0.001	0.001	-0.001	-0.001	0.003	0.003
AlH ₃	0.000	0.000	0.001	0.000	0.001	0.001	0.001	0.001
H ₂ S	0.002	0.006	0.003	0.001	0.008	0.011	0.000	0.000
HCl	0.000	0.004	0.002	0.001	0.005	0.007	0.002	0.001
B ₂	0.096	0.043	0.021	0.010	0.021	-0.010	0.004	0.001
C ₂ (¹ Σ ⁺)	0.083	0.048	0.022	0.011	0.033	0.013	0.003	-0.001
CO	0.000	0.006	0.003	0.002	0.008	0.012	0.001	0.001
N ₂	0.019	0.025	0.010	0.005	0.028	0.031	-0.001	-0.004
CS	0.046	0.031	0.017	0.009	0.025	0.017	0.007	0.004
P ₂	0.082	0.044	0.021	0.011	0.028	0.006	0.004	0.000

RMSD_{all}^{d,e}	0.040	0.022	0.011	0.005	0.016	0.011	0.003	0.002
MAD_{all}^{d,e}	0.021	0.014	0.007	0.004	0.011	0.008	0.002	0.001
MSD_{all}^{d,e}	0.021	0.013	0.007	0.004	0.010	0.006	0.002	0.001
RMSD_{hyd}^{d,f}	0.002	0.003	0.002	0.001	0.004	0.005	0.002	0.002
RMSD_{nonhyd}^{d,g}	0.065	0.036	0.017	0.009	0.025	0.017	0.004	0.002

^aExtrapolation exponent (α) used in two-point $L^{-\alpha}$ extrapolation eq. (1). ^bExtrapolation coefficient (F) used in two-point Schwenke extrapolation eq. (2). ^cUsing empirical α and F parameters optimized to minimize the RMSD. ^dError statistics with respect to the V{5,6}Z reference values in Table 3. RMSD = root-mean-square deviation, MAD = mean absolute deviation, MSD = mean signed deviation. ^eOver all species. ^fOver hydride species. ^gOver non-hydride species.

The magnitude of the \hat{T}_4 -(Q) component for hydrides (in absolute value) ranges between 0.001 (CH₃) and 0.025 (H₂O) kcal mol⁻¹ and accordingly the RMSDs for the various basis sets are negligibly small (Table 8). The RMSDs over the subset of 6 non-hydride systems are significantly larger as illustrated in Figure 2c, specifically RMSD_{nonhyd} = 0.065 (VDZ), 0.036 (VTZ), 0.017 (VQZ), and 0.009 (V5Z) kcal mol⁻¹.

3.6 Effective exponents for iterative connected quadruple excitations. Let us now consider the basis set convergence of the \hat{T}_4 component as a whole. These results are summarized in Table 9. Overall, the basis set convergence of the \hat{T}_4 component is similar to that of the (Q) component. In particular, we obtain RMSDs of 0.265 (VDZ), 0.077 (VTZ), 0.027 (VQZ), and 0.012 (V5Z). Thus, the smallest basis set that achieves sub-kJ/mol accuracy is the VTZ basis set. The two-point extrapolation formula with $\alpha = 3.0$ results in RMSDs of 0.037 (V{D,T}Z) and 0.010 (V{T,Q}Z) kcal mol⁻¹. Similarly to the (Q) component, optimizing the exponent for the V{D,T}Z extrapolation results in marginal improvement in performance and leads to an RMSD of 0.034 kcal mol⁻¹. However, optimizing the exponent for the V{T,Q}Z extrapolation reduces the RMSD from 0.010 to 0.003 kcal mol⁻¹.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI:10.1063/1.50011674

Table 9. Convergence of the T_4 contribution to the total atomization energy for the set of 16 molecules relative to basis set limit values from V{5,6}Z or V{Q,5}Z extrapolations (see Table 3) (deviations and error statistics are given in kcal mol⁻¹).

Basis set	VDZ	VTZ	VQZ	V5Z	V{D,T}Z	V{D,T}Z	V{T,Q}Z	V{T,Q}Z
α^a					3.0	3.3199	3.0	3.6401
F^b					1.4211	1.3518	1.7297	1.5407
BH ₃	-0.012	-0.008	-0.002	-0.001	-0.006	-0.006	0.002	0.001
CH	-0.008	-0.006	-0.003	-0.001	-0.006	-0.006	0.000	-0.001
CH ₂ (¹ A ₁)	-0.033	-0.016	-0.006	-0.002	-0.009	-0.010	0.002	0.000
CH ₃	-0.006	-0.015	-0.005	-0.002	-0.018	-0.018	0.002	0.001
OH	0.023	-0.017	-0.008	-0.003	-0.035	-0.032	-0.001	-0.003
H ₂ O	0.031	-0.036	-0.016	-0.007	-0.065	-0.060	0.000	-0.004
HF	0.059	-0.021	-0.008	-0.003	-0.054	-0.049	0.001	-0.001
AlH ₃	-0.014	-0.006	-0.001	-0.001	-0.003	-0.003	0.002	0.001
H ₂ S	-0.080	-0.027	-0.013	-0.007	-0.005	-0.009	-0.003	-0.006
HCl	-0.050	-0.021	-0.010	-0.005	-0.009	-0.011	-0.002	-0.004
B ₂	-0.249	-0.048	-0.013	-0.004	0.037	0.023	0.013	0.006
C ₂ (¹ Σ^+)	-0.687	-0.158	-0.052	-0.021	0.065	0.028	0.026	0.006
CO	-0.093	-0.069	-0.023	-0.010	-0.059	-0.061	0.010	0.001
N ₂	-0.150	-0.083	-0.030	-0.014	-0.054	-0.059	0.008	-0.002
CS	-0.517	-0.144	-0.054	-0.025	0.013	-0.013	0.011	-0.006
P ₂	-0.527	-0.175	-0.061	-0.030	-0.026	-0.051	0.021	0.000
RMSD_{all}^{d,e}	0.265	0.077	0.027	0.012	0.037	0.034	0.010	0.003
MAD_{all}^{d,e}	0.159	0.053	0.019	0.008	0.029	0.027	0.007	0.003
MSD_{all}^{d,e}	-0.145	-0.053	-0.019	-0.008	-0.015	-0.021	0.006	-0.001
RMSD_{hyd}^{d,f}	0.039	0.020	0.008	0.004	0.030	0.028	0.002	0.003
RMSD_{nonhyd}^{d,g}	0.430	0.123	0.043	0.020	0.046	0.043	0.016	0.004

^aExtrapolation exponent (α) used in two-point $L^{-\alpha}$ extrapolation eq. (1). ^bExtrapolation coefficient (F) used in two-point Schwenke extrapolation eq. (2). ^cUsing empirical α and F parameters optimized to minimize the RMSD. ^dError statistics with respect to the V{5,6}Z reference values in Table 3. RMSD = root-mean-square deviation, MAD = mean absolute deviation, MSD = mean signed deviation. ^eOver all species. ^fOver hydride species. ^gOver non-hydride species.

3.7 The effect of diffuse functions on post-CCSD(T) contributions. The steep computational cost of post-CCSD(T) calculations (particularly CCSDT(Q) and beyond) makes the use of the aug-cc-pVnZ basis sets unfeasible for systems with more than a handful of atoms. Therefore, post-CCSD(T) composite theories such as Weizmann-4 and HEAT use augmented basis sets for the CCSD(T) calculations and regular cc-pVnZ basis sets for the post-CCSD(T) contributions. Nevertheless, it is of interest to examine the effect of diffuse functions on the post-CCSD(T) contributions. Here we will consider the T₃–(T), (Q), and T₄–(Q) components with basis sets of up to A'V6Z, A'V6Z, and A'VQZ quality, respectively. In addition, it is important to estimate to what extent the addition of diffuse function affects the CBS values

reported in Table 3. For this purpose, we consider a subset of nine systems containing electronegative atoms (N, O, F, P, S, and Cl). Table 10 gives an overview of the effects of adding diffuse functions on the post-CCSD(T) contributions. The tabulated values are the differences between the post-CCSD(T) contribution calculated with the VnZ and $A'VnZ$ basis sets, i.e., $VnZ - A'VnZ$. Two key points that can be drawn from Table 10 are that the effect of adding diffuse functions tends to diminish (i) in the order $T_{3-(T)} > (Q) > T_{4-(Q)}$ and (ii) with each increase in the highest angular momentum represented in the basis set. For any given basis set, the RMSD for the (Q) contribution is smaller by ~50% than that for the $T_{3-(T)}$ contribution. Similarly, the RMSD for the $T_{4-(Q)}$ contribution is smaller by about one order of magnitude than that for the (Q) contribution. For any given correlation component ($T_{3-(T)}$, (Q), and $T_{4-(Q)}$), the RMSD is reduced by ~50% with each increase in the highest angular momentum present in the basis set. The RMSDs for the largest basis set considered are 0.006 ($T_{3-(T)}/A'V6Z$), 0.002 ($(Q)/A'V6Z$), and 0.003 ($T_{4-(Q)}/A'VQZ$) kcal mol⁻¹. These RMSDs indicate that the CBS reference values reported in Table 3 should be little affected by the addition of diffuse function.

Table 10. Effect of diffuse functions on post-CCSD(T) contributions. The tabulated values are the differences between the post-CCSD(T) contribution to the total atomization energy calculated with the VnZ and $A'VnZ$ basis sets, i.e., $TAE(VnZ) - TAE(A'VnZ)$ (in kcal mol⁻¹).

	Basis set	OH	H ₂ O	HF	H ₂ S	HCl	CO	N ₂	CS	P ₂	RMSD ^a
T ₃ -(T)	A'VDZ	0.005	0.065	0.028	0.047	0.033	0.165	0.167	0.213	0.167	0.123
	A'VTZ	0.008	0.045	0.026	0.018	0.017	0.076	0.079	0.094	0.096	0.061
	A'VQZ	0.002	0.017	0.011	0.006	0.005	0.025	0.026	0.032	0.033	0.021
	A'V5Z	0.001	0.005	0.004	0.004	0.004	0.009	0.012	0.020	0.023	0.012
	A'V6Z	0.001	0.002	0.002	0.002	0.002	0.004	0.006	0.010	0.014	0.006
(Q)	A'VDZ	-0.005	-0.021	0.004	-0.014	-0.010	-0.049	-0.107	-0.129	-0.100	0.068
	A'VTZ	-0.005	-0.012	0.000	-0.009	-0.005	-0.016	-0.046	-0.044	-0.045	0.027
	A'VQZ	-0.004	-0.008	-0.003	-0.004	-0.003	-0.005	-0.019	-0.017	-0.018	0.011
	A'V5Z	-0.002	-0.003	-0.001	-0.002	-0.002	-0.002	-0.008	-0.007	-0.008	0.005
	A'V6Z	-0.001	-0.001	0.000	N/A	0.000	N/A	-0.003	N/A	-0.004	0.002
T ₄ -(Q)	A'VDZ	0.003	0.008	0.002	0.004	0.002	-0.006	0.012	0.014	0.023	0.011
	A'VTZ	0.001	0.002	-0.001	0.001	0.001	-0.004	0.007	0.007	0.007	0.004
	A'VQZ	0.001	0.002	0.000	0.001	0.000	-0.001	0.003	0.005	0.004	0.003
(Q)-(T)	A'VDZ	0.000	0.043	0.032	0.033	0.023	0.116	0.060	0.084	0.067	0.061
	A'VTZ	0.003	0.033	0.027	0.009	0.012	0.060	0.033	0.050	0.051	0.036
	A'VQZ	-0.002	0.009	0.008	0.001	0.002	0.020	0.007	0.015	0.015	0.011
	A'V5Z	-0.001	0.002	0.002	0.002	0.002	0.007	0.004	0.013	0.015	0.007

^aRMSD over the $TAE(VnZ) - TAE(A'VnZ)$ differences for the nine molecules.

Another important observation is that the addition of diffuse functions tends to systematically reduce the T₃-(T) contribution but systematically increases the (Q) contribution. This is a significant result since it indicates that the use of the regular VnZ basis sets for the calculation of the CCSDT(Q)-CCSD(T) contribution should benefit from a certain degree of error cancellation between the T₃-(T) and (Q) components (see e.g., RMSDs in Table 10).

3.8 Transferability of the optimized exponents. Finally, a comment is due on the transferability of the optimized exponents in Table 5 to systems outside the training set. In general, the transferability of the optimal exponents hinges on the selected training set being reasonably wide and diverse. For example, the selected set of 16 molecules includes first-row, second-row, hydride, non-hydride, closed-shell singlet, singlet diradical, radical, and triplet systems. Nevertheless, it is important to test the extendibility of the extrapolation parameters in Table 5 to systems outside the training set. One way to examine this is by removing six of

the 16 systems from the training set, reoptimizing the extrapolation exponents, and using the omitted systems as a test of the extendibility of the extrapolation parameters. We note that the reduced training set of ten systems is still reasonably diverse and includes BH_3 , CH_3 , OH , H_2O , AlH_3 , H_2S , B_2 , C_2 , N_2 , and P_2 . The RMSDs over the omitted systems using the exponents optimized over the reduced and entire training sets are compared in Table S4 of the Supporting Information. Using the exponents optimized over the reduced training set affects the overall RMSDs by less than $0.001 \text{ kcal mol}^{-1}$. Therefore, we conclude that the extrapolation parameters in Table 5 should be extendible to similar systems outside our training set of 16 systems.

4. Conclusions

The CCSDT-CCSD(T) ($T_3\text{-(T)}$), CCSDT(Q)-CCSDT(Q) ((Q)), CCSDTQ-CCSDT(Q) ($T_4\text{-(Q)}$), and $\text{CCSDTQ-CCSDT(T}_4)$ correlation components are extrapolated to the one-particle basis set limit from the $\text{cc-pV}\{5,6\}\text{Z}$ basis set pair for a diverse set of 16 molecules. The selected set of molecules includes first-row, second-row, hydride, non-hydride, closed-shell singlet, singlet diradical, radical, and triplet systems. Using these basis set limit values we obtain effective decay exponent for the $T_3\text{-(T)}$, (Q) , $T_4\text{-(Q)}$, and T_4 correlation components in conjunction with the $\text{cc-pV}\{D,T\}\text{Z}$, $\text{cc-pV}\{T,Q\}\text{Z}$, and $\text{cc-pV}\{Q,5\}\text{Z}$ basis set pairs. The following major conclusions can be drawn from the present study:

- The fitted extrapolation exponents demonstrate that the $T_3\text{-(T)}$ correlation component converges more slowly to the infinite basis set limit than the (Q) and T_4 terms.
- The optimal extrapolation exponents result in significant improvements in performance (relative to $\alpha = 3.0$) for the $T_3\text{-(T)}$, (Q) , and T_4 components in conjunction with the $\text{cc-pV}\{T,Q\}\text{Z}$ basis set pair. However, smaller improvements in performance are obtained with the other basis set pairs.

- For the T_3 -(T) component convergence is slower than the asymptotically limiting L^{-3} convergence behavior, with effective exponents $\alpha_{eff} = 2.7174$ (cc-pV{D,T}Z), 2.4807 (cc-pV{T,Q}Z), and 2.7342 (cc-pV{Q,5}Z).
- The (Q) component tends to converge faster than the L^{-3} convergence behavior as the size of the basis sets increases, with effective exponents of $\alpha_{eff} = 2.9968$ (cc-pV{D,T}Z), 3.3831 (cc-pV{T,Q}Z), and 3.4216 (cc-pV{Q,5}Z).
- In cases where only the (Q)/cc-pVDZ calculation is feasible, scaling this component by 1.25 results in an RMSD of 0.13 kcal mol⁻¹ relative to our best T_4 /CBS values.
- The T_4 component converges faster than the L^{-3} convergence behavior and the convergence rate increases with the size of the basis sets, with effective exponents of $\alpha_{eff} = 3.3199$ (cc-pV{D,T}Z), 3.6401 (cc-pV{T,Q}Z), and 4.0408 (cc-pV{Q,5}Z).
- The effect of adding diffuse functions tends to diminish (i) in the order T_3 -(T) > (Q) > T_4 -(Q) and (ii) with the highest angular momentum represented in the basis set.
- The addition of diffuse functions tends to systematically reduces the T_3 -(T) contribution and systematically increases the (Q) contribution. Thus, the use of the regular cc-pVnZ basis sets benefits from a certain degree of error cancellation between these two components.
- Taken together, these results confirm that the basis sets and basis set extrapolations used for obtaining post-CCSD(T) components in composite thermochemical theories such as Weizmann-4 and HEAT are sufficiently converged to the complete basis set limit for attaining sub-kJ-per-mole accuracy.
- Across all post-CCSD(T) contributions, the basis set convergence for the hydride systems is significantly faster than for non-hydride systems.

Supplementary Material

See supplementary material for post-CCSD(T) contributions for the 200 species in the W4-17 database (Table S1); optimized geometries for all the species considered in this work (Table S2); effect of varying the extrapolation exponent of $\alpha = 3.0$ for obtaining the CBS limit reference values in Table 3 (Table S3); examination of the transferability of the optimized exponents reported in Table 5 (Table S4); CCSDT–CCSD(T) contributions to the total atomization energies (Table S5); CCSDT(Q)–CCSDT contributions to the total atomization energies (Table S6); CCSDTQ–CCSDT(Q) contributions to the total atomization energies (Table S7); and CCSD(T), CCSDT, CCSDT(Q), and CCSDTQ absolute energies calculated in this work (Table S8).

Acknowledgments

This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We also acknowledge system administration support provided by the Faculty of Science at the University of Western Australia to the Linux cluster of the Karton group. I would like to thank the reviewers of this manuscript for their constructive comments and suggestions. I gratefully acknowledge an Australian Research Council (ARC) Future Fellowship (FT170100373).

Data Availability Statement

The data that supports the findings of this study are available within the article [and its supplementary material]. Any additional data are available from the corresponding author upon reasonable request.

References

- ¹ T. Helgaker, W. Klopper, K. L. Bak, A. Halkier, P. Jørgensen, and J. Olsen, “Highly accurate ab initio computation of thermochemical data,” in *Quantum-Mechanical Prediction of Thermochemical Data*, Understanding Chemical Reactivity, edited by J. Cioslowski (Kluwer, Dordrecht, 2001), Vol. 22, pp. 1–30.
- ² J. M. L. Martin and S. Parthiban, “W1 and W2 theory and their variants: Thermochemistry in the kJ/mol accuracy range,” in *Quantum-mechanical Prediction of Thermochemical Data*, Understanding Chemical Reactivity, edited by J. Cioslowski (Kluwer, Dordrecht, 2001), Vol. 22, pp. 31–65.
- ³ J. M. L. Martin, *Annu. Rep. Comput. Chem.* **1**, 31 (2005).
- ⁴ T. Helgaker, W. Klopper, and D. P. Tew, *Mol. Phys.* **106**, 2107 (2008).
- ⁵ D. Feller, K. A. Peterson, and D. A. Dixon, *J. Chem. Phys.* **129**, 204105 (2008).
- ⁶ N. DeYonker, T. R. Cundari, and A. K. Wilson, in *Advances in the Theory of Atomic and Molecular Systems*, Progress in Theoretical Chemistry and Physics, edited by P. Piecuch, J. Maruani, G. Delgado-Barrio, and S. Wilson (Springer Netherlands, Dordrecht, 2009), Vol. 19, pp. 197–224.
- ⁷ A. Karton, S. Daon, and J. M. L. Martin, *Chem. Phys. Lett.* **510**, 165 (2011).
- ⁸ L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *WIREs Comput. Mol. Sci.* **1**, 810 (2011).
- ⁹ K. A. Peterson, D. Feller, and D. A. Dixon, *Theor. Chem. Acc.* **131**, 1079 (2012).
- ¹⁰ A. Karton, *WIREs Comput. Mol. Sci.* **6**, 292 (2016).
- ¹¹ I. Shavitt and R. J. Bartlett, in *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge Molecular Science Series (Cambridge University Press, Cambridge, 2009).
- ¹² R. J. Bartlett and M. Musial, *Rev. Mod. Phys.* **79**, 291 (2007).

- ¹³ K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479 (1989).
- ¹⁴ J. D. Watts, J. Gauss, and R. J. Bartlett, *J. Chem. Phys.* **98**, 8718 (1993).
- ¹⁵ A. Karton, *J. Chem. Phys.* **149**, 034102 (2018).
- ¹⁶ A. Karton, N. Sylvetsky, and J. M. L. Martin, *J. Comput. Chem.* **38**, 2063 (2017).
- ¹⁷ N. Sylvetsky, K. A. Peterson, A. Karton, and J. M. L. Martin, *J. Chem. Phys.* **144**, 214101 (2016).
- ¹⁸ D. Feller, K. A. Peterson, and B. Ruscic, *Theor. Chem. Acc.* **133**, 1407 (2014).
- ¹⁹ A. Karton, S. Parthiban, and J. M. L. Martin, *J. Phys. Chem. A* **113**, 4802 (2009).
- ²⁰ A. Karton, I. Kaminker, and J. M. L. Martin, *J. Phys. Chem. A* **113**, 7610 (2009).
- ²¹ M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, *J. Chem. Phys.* **128**, 114111 (2008).
- ²² A. Karton, P. R. Taylor, and J. M. L. Martin, *J. Chem. Phys.* **127**, 064104 (2007).
- ²³ D. Feller and K. A. Peterson, *J. Chem. Phys.* **126**, 114105 (2007).
- ²⁴ A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, *J. Chem. Phys.* **125**, 144108 (2006).
- ²⁵ 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).
- ²⁶ 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).
- ²⁷ A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, and M. Kállay, J. Gauss, *J. Chem. Phys.* **120**, 4129 (2004).
- ²⁸ T. A. Ruden, T. Helgaker, P. Jørgensen, and J. Olsen, *Chem. Phys. Lett.* **371**, 62 (2003).
- ²⁹ A. Karton, P. R. Schreiner, and J. M. L. Martin, *J. Comput. Chem.* **37**, 49 (2016).
- ³⁰ R. O. Ramabhadran and K. Raghavachari, *Acc. Chem. Res.* **47**, 3596 (2014).
- ³¹ R. O. Ramabhadran and K. Raghavachari, *J. Phys. Chem. A* **116**, 7531 (2012).

- ³² S. E. Wheeler, WIREs Comput. Mol. Sci. **2**, 204 (2012).
- ³³ S. E. Wheeler, K. N. Houk, P. V. R. Schleyer, and W. D. Allen, J. Am. Chem. Soc. **131**, 2547 (2009).
- ³⁴ J. M. L. Martin, AIP Conf. Proc. **2040**, 020008 (2018).
- ³⁵ A. J. C. Varandas, Annu. Rev. Phys. Chem. **69**, 177 (2018).
- ³⁶ F. N. N. Pansini, A. C. Neto, and A. J. C. Varandas, Chem. Phys. Lett. **641**, 90 (2015).
- ³⁷ F. N. N. Pansini and A. J. C. Varandas, Chem. Phys. Lett. **631-632**, 70 (2015).
- ³⁸ A. J. C. Varandas and F. N. N. Pansini, J. Chem. Phys. **141**, 224113 (2014).
- ³⁹ D. S. Ranasinghe and E. C. Barnes, J. Chem. Phys. **140**, 184116 (2014).
- ⁴⁰ D. S. Ranasinghe and G. A. Petersson, J. Chem. Phys. **138**, 144104 (2013).
- ⁴¹ D. Feller, J. Chem. Phys. **138**, 074103 (2013).
- ⁴² A. Karton and J. M. L. Martin, J. Chem. Phys. **136**, 124114 (2012).
- ⁴³ D. Feller, K. A. Peterson, J. G. Hill, J. Chem. Phys. **135**, 044102 (2011).
- ⁴⁴ J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, J. Chem. Phys. **131**, 194105 (2009).
- ⁴⁵ D. Bakowies, J. Chem. Phys. **127**, 164109 (2007).
- ⁴⁶ D. Bakowies, J. Chem. Phys. **127**, 084105 (2007).
- ⁴⁷ D. Feller, K. A. Peterson, and T. D. Crawford, J. Chem. Phys. **124**, 054107 (2006).
- ⁴⁸ D. W. Schwenke, J. Chem. Phys. **122**, 014107 (2005).
- ⁴⁹ E. F. Valeev, W. D. Allen, R. Hernandez, C. D. Sherrill, and H. F. Schaefer, J. Chem. Phys. **118**, 8594 (2003).
- ⁵⁰ D. Feller and D. A. Dixon, J. Chem. Phys. **115**, 3484 (2001).
- ⁵¹ W. Klopper, Mol. Phys. **99**, 481 (2001).
- ⁵² K. L. Bak, P. Jørgensen, J. Olsen, T. Helgaker, and W. Klopper, J. Chem. Phys. **112**, 9229 (2000).
- ⁵³ J. M. L. Martin and G. de Oliveira, J. Chem. Phys. **111**, 1843 (1999).

- ⁵⁴ D. G. Truhlar, Chem. Phys. Lett. **294**, 45 (1998).
- ⁵⁵ A. G. Császár, W. D. Allen, and H. F. Schaefer III, J. Chem. Phys. **108**, 9751 (1998).
- ⁵⁶ W. Klopper, J. Noga, H. Koch, and T. Helgaker, Theor. Chem. Acc. **97**, 164 (1997).
- ⁵⁷ W. Kutzelnigg and J. D. Morgan III, J. Chem. Phys. **96**, 4484 (1992); J. Chem. Phys. **97**, 8821(E) (1992).
- ⁵⁸ A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).
- ⁵⁹ R. N. Hill, J. Chem. Phys. **83**, 1173 (1985).
- ⁶⁰ A. Ganyecz, M. Kállay, and J. Csontos, J. Chem. Theory Comput. **13**, 4193 (2017).
- ⁶¹ MRCC, a quantum chemical program suite written by M. Kállay, *et al.*, See also: <http://www.mrcc.hu>.
- ⁶² Z. Rolik, L. Szegedy, I. Ladjanszki, B. Ladoczki, and M. Kállay, J. Chem. Phys. **139**, 094105 (2013).
- ⁶³ T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
- ⁶⁴ R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
- ⁶⁵ A. K. Wilson, T. van Mourik, and T. H. Dunning, J. Mol. Struct.: THEOCHEM **388**, 339 (1996).
- ⁶⁶ T. van Mourik and T. H. Dunning, Int. J. Quantum Chem. **76**, 205 (2000).
- ⁶⁷ R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
- ⁶⁸ T. H. Dunning, K. A. Peterson, and A. K. Wilson, J. Chem. Phys. **114**, 9244 (2001).
- ⁶⁹ P. R. Spackman, D. Jayatilaka, and A. Karton, J. Chem. Phys. **145**, 104101 (2016).