AIP ADVANCES **5**, 057148 (2015)

Estimating the CCSD basis-set limit energy from small basis sets: basis-set extrapolations vs additivity schemes

Peter R. Spackman and Amir Karton[a]
School of Chemistry and Biochemistry, The University of Western Australia, Perth, WA 6009, Australia

(Received 28 February 2015; accepted 13 May 2015; published online 20 May 2015)

Coupled cluster calculations with all single and double excitations (CCSD) converge exceedingly slowly with the size of the one-particle basis set. We assess the performance of a number of approaches for obtaining CCSD correlation energies close to the complete basis-set limit in conjunction with relatively small DZ and TZ basis sets. These include global and system-dependent extrapolations based on the A + B/L^{α} two-point extrapolation formula, and the well-known additivity approach that uses an MP2-based basis-set-correction term. We show that the basis set convergence rate can change dramatically between different systems(e.g.it is slower for molecules with polar bonds and/or second-row elements). The system-dependent basis-set extrapolation scheme, in which unique basis-set extrapolation exponents for each system are obtained from lower-cost MP2 calculations, significantly accelerates the basis-set convergence relative to the global extrapolations. Nevertheless, we find that the simple MP2-based basis-set additivity scheme outperforms the extrapolation approaches. For example, the following root-mean-squared deviations are obtained for the 140 basis-set limit CCSD atomization energies in the W4-11 database: 9.1 (global extrapolation), 3.7 (system-dependent extrapolation), and 2.4 (additivity scheme) kJ mol^{-1}. The CCSD energy in these approximations is obtained from basis sets of up to TZ quality and the latter two approaches require additional MP2 calculations with basis sets of up to QZ quality. We also assess the performance of the basis-set extrapolations and additivity schemes for a set of 20 basis-set limit CCSD atomization energies of larger molecules including amino acids, DNA/RNA bases, aromatic compounds, and platonic hydrocarbon cages. We obtain the following RMSDs for the above methods: 10.2 (global extrapolation), 5.7 (system-dependent extrapolation), and 2.9 (additivity scheme) kJ mol^{-1}. © *2015 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License.* [http://dx.doi.org/10.1063/1.4921697]

I. INTRODUCTION

Coupled-cluster theory is one of the most reliable, yet computationally affordable, methods for solving the nonrelativistic electronic Schrödinger equation.[1] Coupled-cluster theory entails a hierarchy of approximations that can be systematically improved towards the exact quantum mechanical solution, providing a roadmap for highly accurate chemical properties.[2–8] In particular, the CCSD(T) method (coupled-cluster with single, double, and quasiperturbative triple excitations) has been found to be a cost-effective approach for the calculation of reliable thermochemical and kinetic data(e.g. reaction energies and barrier heights) as well as molecular properties based on energy derivatives (e.g. equilibrium/transition structures, vibrational frequencies, and electrical properties).[9–14] However, one of the greatest weaknesses of the CCSD(T) method is that it converges exceedingly slowly to the complete basis set (CBS) limit. This is particularly true for the double excitations, as they reflect dynamical rather than static electron correlation effects.[2,3]

[a]E-Mail: amir.karton@uwa.edu.au

 ©Author(s) 2015

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

One way to accelerate the basis set convergence is to use basis set extrapolations that exploit the systematic behavior of the correlation-consistent basis sets of Dunning and coworkers.[15–17] There has been an extensive discussion in the literature on the form that this extrapolation should take,[18–41] where the focus has been on extrapolations that use relatively large basis sets of at least quadruple-zeta quality. References 32–41 provide a comprehensive review of the previous works. These extrapolations use a global extrapolation formula for all systems regardless of the unique electronic structure and bonding situations of each system. The use of a global extrapolation formula is valid when large basis sets are used, however, it becomes less valid for basis sets that are far from the asymptotic limit.

In the present work we consider two alternative approaches for obtaining CCSD atomization energies close to the basis-set limit in conjunction with relatively small basis sets. The first is a system-dependent extrapolation scheme that accounts for the different basis-set convergence rate between different systems. We show that the basis-set convergence rate of the CCSD correlation energy can differ substantially between different molecular systems, and we introduce a system-dependent extrapolation scheme to account for these differences. In this scheme, unique extrapolation exponents for each molecular system are obtained from lower-cost second-order Møller-Plesset perturbation theory (MP2) calculations. In this context, it is worth mentioning the work of Bakowies,[36] which highlights the limitations of conventional extrapolations and proposes an empirically motivated extrapolation formula for MP2 and CCSD correlation energies. The scaling factors in this extrapolation depend on the number of hydrogen atoms and the number of correlated electrons in each system.The second approach is anadditivity scheme in which the CCSD energy is calculated with a relatively small basis set and an MP2-based basis-set-correction term is added in order to approximate the CCSD/CBS energy.This cost-effective approach, which is widely used in conjunction with the correlation-consistent basis sets for obtaining noncovalent interaction energies at the CCSD(T)/CBS limit,[42–45] has been recently shown to give good performance for reaction energies[46,47] and barrier heights.[48]

II. COMPUTATIONAL METHODS

All calculations were carried out on the Linux cluster of the Karton group at UWA. All MP2 and CCSD calculations were carried out with the correlation-consistent basis sets, using version 2012.1 of the MOLPRO program suite.[49,50] The notation A′V*n*Z indicates the combination of the standard correlation-consistent cc-pV*n*Z basis sets on hydrogen,[15] the aug-cc-pV*n*Z basis sets on first-row elements,[16] and the aug-cc-pV(n+d)Z basis sets on second-row elements.[17] All correlated calculations were performed 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.We consider the two-point extrapolation formula of Halkier *et al*.[26]

$$E(L) = E_\infty + A/L^\alpha \tag{1}$$

where L is the highest angular momentum represented in the basis set, and α can be a global exponent (Section III A) or a system-dependent exponent obtained individually for each system from lower-cost MP2 calculations (Section III B). We note that the global two-point extrapolations are extensively used in highly accurate composite methods (e.g. W*n*,[2,4,10,51,52] W*n*-F12,[53] and HEAT)[5–7] in conjunction with relatively large basis set of at least quadruple-ζ quality. We also consider the empirical two-point extrapolation formula of Schwenke[34]

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

which does not involve L (where F is an empirical scaling factor). We note that this expression is equivalent to the above inverse power extrapolation with an empirical extrapolation exponent. The notation A′V{n,n+1}Z indicates extrapolation from the A′VnZ and A′V(n+1)Z basis sets using the aforementioned two-point formulas. Basis set extrapolations that use only a single extrapolation parameter (α or F) for all systems will be referred to as global extrapolations, whereas extrapolations that use a unique extrapolation exponent for each system will be referred to as system-dependent extrapolations.

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

We also assess the performance of an additivity-based approach for approximating the CCSD/CBS energy. In this approach the CCSD base-energy is calculated with a small basis set and an MP2-based basis-set-correction term (ΔMP2) is calculated with larger basis sets

$$\text{CCSD/CBS} \approx \text{CCSD/X(MP2/Y)} = \text{CCSD/A}'\text{VXZ} + \Delta\text{MP2} \tag{3}$$

where ΔMP2 is given by

$$\Delta\text{MP2} = \text{MP2/A}'\text{V}\{\text{Y}-1,\text{Y}\}\text{Z} - \text{MP2/A}'\text{VXZ} \tag{4}$$

and the MP2/A′V{Y–1, Y}Z energy is extrapolated to the basis set limit with a global two-point extrapolation formula (eq. (1)). We note that the popular G*n*(MP2) composite thermochemical methods use a similar approach in conjunction with Pople-type basis setsto approximate the CCSD(T)/TZ energy.[54–56]

The performance of the various basis-set extrapolations and additivity schemes are evaluated relative to the basis-set limit valence CCSD atomization energies in the W4-11 database.[4] This database includes 140 total atomization energies (TAEs) of species that cover a broad spectrum of bonding situations and as such it is an excellent benchmark set for the validation of basis-set extrapolation techniques. The W4-11 dataset includes the following species: closed shell (97), radicals (27), singlet carbenes (9), and triplet systems (7). In terms of elemental composition the dataset includes 100 first-row species, 19 second-row species, and 21 mixed first- and second-row species.[4] The CCSD atomization energies in the W4-11 dataset were obtained by means of the W4 thermochemical protocol.[2] In particular, they are extrapolated from the A′V5Z and A′V6Z basis sets. Following the suggestion of Klopper,[32] they are partitioned into singlet-coupled pair energies, triplet-coupled pair energies, and T_1 terms. The singlet-coupled and triplet-coupled pair energies are extrapolated using eq. (1) with $\alpha_s = 3$ and $\alpha_t = 5$, respectively, and the T_1 term (which exhibits a weak basis-set dependence) is set equal to that in the largest basis set.

In addition,we evaluate the performance of the basis-set extrapolations and additivity schemes for aset of 20 TAEs of systems that are much larger than those in the W4-11 database. These include:

- Benzene and fulvene (C_6H_6), phenyl radical (C_6H_5), pyridine (C_5H_5N), pyrazine ($C_4H_4N_2$), furan (C_4H_4O), 1,3-butadiene (C_4H_6), and cyclobutene (*cyc*-C_4H_6). The CCSD correlation component of the TAEs for these systems are obtained from W3.2 theory (i.e. extrapolated from the A′V{Q,5}Z basis-set pair, see Ref. 10 for further details).
- The DNA/RNA bases adenine ($H_5C_5N_5$), cytosine ($H_5C_4N_3O$), thymine ($H_6C_5N_2O_2$), and uracil ($H_4C_4N_2O_2$). The amino acids alanine ($H_7C_3NO_2$), cysteine ($H_7C_3NO_2S$), glycine ($H_5C_2NO_2$), methionine ($H_{11}C_5NO_2S$), and serine ($H_7C_3NO_3$). The platonic hydrocarbon cages tetrahedrane (C_4H_4), triprismane (C_6H_6), and cubane (C_8H_8). The CCSD/CBS TAEs for these compounds are obtained from W2-F12 theory (i.e. extrapolated from the V{T,Q}Z-F12 basis-set pair). These reference values are taken from Ref. 53 (DNA/RNA bases), Ref. 57 (amino acids), and Ref. 58 (platonic hydrocarbons).

III. RESULTS AND DISCUSSION

A. Performance of conventional basis set extrapolations for obtaining the CCSD correlation component of the TAEs in the W4-11 dataset

Table I gives the error statistics for the CCSD component of the TAEs extrapolated from the A′V{D,T}Z, A′V{T,Q}Z, and A′V{Q,5}Z basis-set pairs. The individual errors can be found in Table S1 of the supplementary material.[59] As expected, extrapolating the CCSD energy from the A′V{D,T}Z basis sets with $\alpha = 3$ results in a very large RMSD of 13.5 kJ mol^{-1}. This level of theory tends to systematically and severely underestimate the basis-set-limit TAEs, as evident from the mean signed deviation (MSD) and mean absolute deviation (MAD) being practically the same (MAD ≈ MSD = –10.8 kJ mol^{-1}). Ad hoc optimization of α to minimize the RMSD over the W4-11 training set leads to $\alpha = 2.357$. Using this empirical extrapolation exponent alleviates the

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

TABLE I. **Error statistics for global extrapolations of the valence CCSD correlation component over the 140 total atomization energies in the W4-11 dataset (kJ mol^{-1}). The reference values are CCSD/A′V{5,6}Z total atomization energies from W4 theory.**[a]

Basis sets	α[d]	RMSD	MAD	MSD	LND	LPD
A′V{D,T}Z[b]	3	13.54	10.99	−10.83	−41.34 (SO_3)	3.81 (B_2H_6)
A′V{D,T}Z[b]	2.357[e]	9.12	6.86	2.19	−32.57 (SO_3)	20.67 (C_3H_8)
A′V{D,T}Z[b]	2.400[f]	9.16	6.89	−2.92	−32.32 (SO_3)	18.55 (C_3H_8)
A′V{D,T}Z[c]	[g]	9.30	6.99	−3.76	−34.16 (SO_3)	16.51 (B_2H_6)
A′V{T,Q}Z[b]	3	2.80	2.38	2.37	−0.84 (ClO_2)	6.66 ($AlCl_3$)
A′V{T,Q}Z[b]	3.403[e]	1.15	0.88	0.07	−3.57 (S_4)	3.08 (Si_2H_6)
A′V{T,Q}Z[c]	[g]	2.27	1.87	1.83	−1.35 (ClO_2)	5.78 ($AlCl_3$)
A′V{Q,5}Z[b]	3	0.81	0.54	0.51	−0.38 (CF_4)	3.59 ($AlCl_3$)
A′V{Q,5}Z[b]	3.220[e]	0.54	0.38	−0.02	−1.01 (HCNO	2.69 ($AlCl_3$)
A′V{Q,5}Z[b]	[g]	0.55	0.42	−0.14	−1.17 (HCNO	2.50 ($AlCl_3$)

[a]RMSD = root-mean-squared deviation, MAD = mean absolute deviation, MSD = mean signed deviation, LND = largest negative deviation, and LPD = largest positive deviation (the molecules associated with the LND and LPD are given in parenthesis).
[b]Using eq. (1).
[c]Using Schwenke's single parameter extrapolation formula eq. (2).
[d]Extrapolation exponent used in the two-point extrapolations.
[e]Optimized to minimize the RMSD over the 140 TAEs in the W4-11 dataset.
[f]Extrapolation exponent taken from Ref. 27.
[g]The extrapolation fitting constants from Ref. 34 are F = 1.5877616 (A′V{D,T}Z), 1.7001115 (A′V{T,Q}Z), and 1.9303174 (A′V{Q,5}Z).

bias toward underestimation of the basis-set limit TAEs (e.g. the MSD drops to –2.2 kJ mol^{-1}). However, an RMSD of 9.1 kJ mol^{-1} is still unacceptably large (Table I). Remarkably, Truhlar obtained a very similar extrapolation exponent of α = 2.400 by minimizing the RMSD over a much smaller set of correlation energies.[27] These two empirical exponents (α = 2.357 and 2.400) result in similar overall error statistics for the W4-11 dataset. Using Schwenke's extrapolation in conjunction with the A′V{D,T}Z basis sets leads to a slightly higher RMSD of 9.3 kJ mol^{-1}.[34] We note that Schwenke's extrapolation in this case is equivalent to an inverse power extrapolation with α = 2.451.

Using the A′V{T,Q}Z basis sets in the extrapolations significantly improves the performance. We note that these basis sets are used for extrapolating the CCSD correlation energy in W1w theory.[51] An extrapolation exponent of α = 3 results in an RMSD of 2.8 kJ mol^{-1} and a largest positive deviation of 6.7 kJ mol^{-1} ($AlCl_3$). Using an empirical extrapolation exponent of α = 3.403, which minimizes the RMSD over the TAEs in the W4-11 dataset, reduces the RMSD to merely 1.2 kJ mol^{-1} and the largest positive deviation to 3.08 kJ mol^{-1} (Si_2H_6). We note that Schwenke's extrapolation (which is equivalent to an inverse power extrapolation with α = 3.803) leads to an RMSD of 2.3 kJ mol^{-1} and to a largest positive deviation of 5.8 kJ mol^{-1} (Table I).

Using the A′V{Q,5}Z basis-set pair, which is used for the CCSD extrapolations in W2.2 theory,[2,51] results in excellent performance. For example, with α = 3 we obtain an RMSD of 0.8 kJ mol^{-1}, and anempirical extrapolation exponent of 3.220 results in an RMSD of only 0.5 kJ mol^{-1}. We note that in this case, Schwenke's extrapolation leads to practically the same RMSD (this extrapolation is equivalent to an inverse power extrapolation with α = 3.171).

B. System-dependent basis set extrapolations of the CCSD correlation energy

The global extrapolations considered in Section III A work well in conjunction with sufficiently large basis sets (e.g. A′V{T,Q}Z and larger). With the A′V{T,Q}Z basis-set pair the $1/L^{\alpha}$ extrapolation results in RMSDs of 2.8 (α = 3) and 1.2 kJ mol^{-1} (α = 3.403). However, using this extrapolation in conjunction with the A′V{D,T}Z basis setssignificantly worsens the performance, such that even an empirical exponent results in an unacceptably large RMSD of 9.1 kJ mol^{-1}. A

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

natural question that arises is: can a single extrapolation exponent adequately describe the basis set convergence of all systems in conjunction with small basis sets? For systems for which we have results at the one-particle basis-set limit, this question can easily be answeredby backtracking the ideal extrapolation exponents that will reproduce the CCSD basis-set limit energies. A narrow distribution of the ideal extrapolation exponents around a single value indicates a uniform basis set convergence, whereas a broad distribution of the ideal extrapolation exponents indicates that a global extrapolation cannot possibly do a good job for many systems.

The two-point extrapolation formula from the A′VnZ and A′V(n+1)Z basis sets can be written in the form

$$E_\infty = E_{n+1} + \frac{E_{n+1} - E_n}{\left(\frac{n+1}{n}\right)^\alpha - 1} \tag{5}$$

Where E_n, E_{n+1}, and E_∞ are the CCSD/A′VnZ, CCSD/A′V(n+1)Z, and CCSD/CBS energies, respectively. A simple rearrangementof eq. (1), gives the ideal extrapolation exponent(α_{ideal}) that will exactly reproduce E_∞ from E_n and E_{n+1}

$$\alpha_{ideal} = \frac{ln\left(\frac{E_{n+1}-E_n}{E_\infty - E_{n+1}}+1\right)}{ln\left(\frac{n+1}{n}\right)} \tag{6}$$

This ideal extrapolation exponent can be obtained for any system for which E_n, E_{n+1}, and E_∞ are known. Table S2 of the supplementary material[59] gives the ideal exponents that will exactly reproduce the CCSD/A′V{5,6}Z energies from the A′V{D,T}Z, A′V{T,Q}Z, and A′V{Q,5}Z basis-set pairs, whilst Figure 1 shows the standard distributions of these ideal exponents. We note three general observations:

- The Gaussian distributionfor the ideal exponents for the A′V{D,T}Z, A′V{T,Q}Z and A′V{Q,5}Z basis-set pairs are centered at values that are progressively closer to 3.
- The ideal exponents for the A′V{D,T}Z basis-set pair are distributed over a much wider interval than those for the A′V{T,Q}Z and A′V{Q,5}Z basis-set pairs.
- Hydrogen-rich molecules tend to converge faster to the basis set limit whereas molecules with polar bonds and/or second-row elements exhibit a slower basis set convergence.

The normal distribution functions for the ideal exponents are centered at 2.220 (A′V{D,T}Z), 3.462 (A′V{T,Q}Z), and 3.230 (A′V{Q,5}Z). These mean values arein close agreement with the empirical extrapolation exponents obtained for the global extrapolations by minimizing the RMSD over the W4-11 dataset (Table I). Specifically, the optimal global extrapolation exponents are

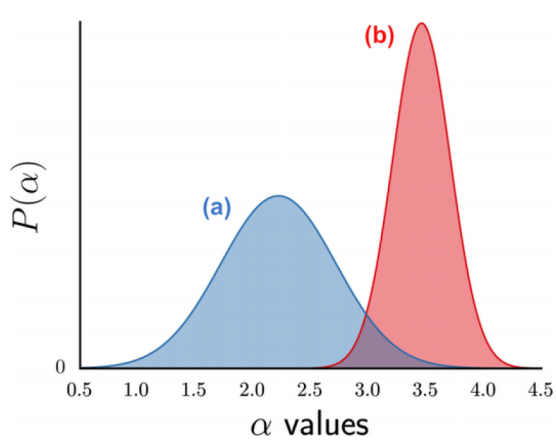

FIG. 1. Normal distributions of system-dependent extrapolation exponents (α_{ideal}, obtained from eq. (6)) for the 140 molecules in the W4-11 database. The ideal exponents reproduce the CCSD/CBS energy for each molecule from the (a) A′V{D,T}Z and (b) A′V{T,Q}Z basis-set pairs. The Gaussians are centered around 2.22 (A′V{D,T}Z) and 3.46 (A′V{T,Q}Z), and have standard deviations of 0.50 and 0.25, respectively.

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

2.357 (A′V{D,T}Z), 3.403 (A′V{T,Q}Z), and 3.220 (A′V{Q,5}Z). The ideal exponents for the A′V{D,T}Z basis-set pair span a wide range from 1.282 (SO_2) to 3.504 (H_2), with a standard deviation of 0.502. This illustrates that a global extrapolation exponent cannot possibly be a successful approach for estimating the CCSD basis-set limit from the A′VDZ and A′VTZ basis sets. The ideal exponents for the A′V{T,Q}Z basis-set pair span a narrower interval, ranging from of 2.871 (ClO_2) to 4.224 (AlH_3), with a standard deviation of 0.251. The narrower width of the normal distribution for the A′V{T,Q}Z basis sets indicates that a global extrapolation scheme is more suitable to be used in conjunction with these larger basis sets than with the A′V{D,T}Z basis-set pair. We note in passing that the ideal exponents for the A′V{Q,5}Z basis-set pair are asymmetrically distributed around the mean value and therefore do not fit sufficiently well to a normal distribution.[59] Nevertheless, they are narrowly distributed around a value of 3.220 (Figure S1, supplementary material). The above results illustrate the limitations of basis set extrapolations using a global exponent in conjunction with small basis sets and indicate that system-dependent extrapolations may result in performance improvements. This raises the question: how can one obtain ideal (or effective) extrapolation exponents without prior knowledge of the CCSD/CBS energy?

In the present work,we consider the possibility of estimating the exponents for each system from lower-cost MP2 calculations. Such a system-dependent approach is not intended to be used as an actual extrapolation—as it is not size consistent—but as a probe for the effective convergence rate. In particular, we will use the following procedureto obtain effective(α_{eff}) exponents for each system (atom or molecule):

i. Extrapolatethe MP2/CBS correlation energyfrom basis sets of at least A′V(n+1)Z and A′V(n+2)Z quality using eq. (5).
ii. Use eq. (6) to calculate an effective extrapolation exponent (α_{eff}) that will reproduce the MP2/CBS energy obtained in step (i) from the A′V{n,n+1}Z basis-set pair.
iii. Use the extrapolation exponent α_{eff} to extrapolate the CCSDcorrelation energy from the A′V{n,n+1}Z basis-set pair.

Table II gives error statistics for the CCSD TAEs extrapolated from the A′V{D,T}Z basis-set pair in conjunction with system-dependent exponents obtained via the above procedure. The supplementary material lists the errors for each molecule (Table S3) and the system-dependent extrapolation exponents (Table S4).[59] Calculating the MP2/CBS energy from the A′V{T,Q}Z basis-set pair in step (i), results in an RMSD of 4.4 kJ mol^{-1}. This represents a significant improvement over the global extrapolation from the same basis sets, which results in RMSDs of 13.5 (with α = 3) and 9.1–9.3 kJ mol^{-1} (with empirical exponents) (Table I).

TABLE II. **Error statistics for the system-dependent extrapolations of the valence CCSD correlation component over the 140 total atomization energies in the W4-11 test set (kJ mol^{-1}). The reference values are CCSD/A′V{5,6}Z total atomization energies from W4 theory**[a]

Basis sets[b]	λ[f]	RMSD	MAD	MSD	LND	LPD
A′V{D,T}Z[c]	N/A	4.42	3.47	1.80	−9.64 (SO_3)	12.35 (C_3H_8)
A′V{D,T}Z[c]	1.050	3.65	2.77	–0.38	−12.88 (SO_3)	8.02 (Si_2H_6)
A′V{D,T}Z[d]	1.050	3.18	2.52	–0.19	−10.68 (SO_3)	7.83 (Si_2H_6)
A′V{T,Q}Z[d]	1.169	0.90	0.60	–0.02	−1.62 (ClO_2)	5.36 ($AlCl_3$)
A′V{T,Q}Z[e]	1.188	0.77	0.48	–0.01	−2.66 (ClO_2)	3.40 ($AlCl_3$)
A′V{Q,5}Z[e]	1. 114	0.24	0.17	0.00	−0.51 (C_3H_8)	1.06 ($AlCl_3$)

[a]Footnote a to Table I applies here.
[b]Basis sets used in the CCSD extrapolations (eq. (5)).
[c]The system-dependent extrapolation exponents are obtained from MP2 calculations in conjunction with the A′VDZ, A′VTZ, and A′VQZ basis sets (eq. (6)).
[d]The system-dependent extrapolation exponents are obtained from MP2 calculations in conjunction with the A′VTZ, A′VQZ, and A′V5Z basis sets (eq. (6)).
[e]The system-dependent extrapolation exponents are obtained from MP2 calculations in conjunction with the A′VQZ, A′V5Z, and A′V6Z basis sets (eq. (6)).
[f]Global scaling factor optimized to minimize the RMSD over the 140 TAEs in the W4-11 dataset.

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

Note however, that our system-dependent approach is entirely devoid of empirical parameters. The performance of the system-dependent extrapolation can be improved by introducing a global scaling factor (λ) to account for systematic differences between the basis set convergence of the MP2 and CCSD correlation energies.[26,27,33,36] Introducing such an empirical scaling factor optimized to minimize the RMSD over the W4-11 training set (i.e. multiplying all the system-dependent extrapolation exponents by λ = 1.050) results in an RMSD of 3.7 kJ mol^{-1}. We also note that the largest negative and positive deviations for the system-dependent extrapolation (–12.9 and 8.0 kJ mol^{-1}, respectively) are significantly smaller than those obtained for the global extrapolation (–32.6 and 20.7 kJ mol^{-1}, respectively) (Table I). Finally, it is worth pointing out that the extra MP2/A′VQZ calculation that is neededfor obtaining thesystem-dependent extrapolation exponentsis computationally more economical than the CCSD/A′VTZ calculation. Therefore, the system-dependent approach does not significantly increase the computational cost over the conventional approach. Calculating the MP2/CBS energy from the larger A′VQZ and A′V5Z basis sets (eqs. (5) and (6)) results in an RMSD of 3.2 kJ mol^{-1}.

It is also of interest to assess the performance of the global and system-dependent CCSD/A′V{D,T}Z extrapolations together with the Hartree–Fock (HF) and (T) contributions against CCSD(T)/CBS reference values from the W4-11 database.[4] For this purposewe will takethe HF and (T) componentsfrom W1w theory[51,60] and add them to the CCSD/A′V{D,T}Z components calculated via the global and system-dependent extrapolations. The global extrapolations result in RMSDs that are close to those obtained for the CCSD component alone, namely they are 13.55 (α = 3) and 9.50 (α = 2.357) kJ mol^{-1}. The RMSDs for the system-dependent extrapolations increase by about 10% relative to the RMSDs for the CCSD component alone. In particular, calculating the CCSD/A′V{D,T}Z energy using the nonempirical system-dependent extrapolation results in an RMSD of 5.04 kJ mol^{-1}. The empirical system-dependent extrapolations result in RMSDs of 4.21 and 3.56 kJ mol^{-1} (depending on the basis sets used for extrapolating the MP2/CBS energy).

So far we have shown that extrapolating the CCSD energy from the A′VDZ and A′VTZ basis sets using system-dependent basis set extrapolations cuts the RMSD by over50% relative to the global basis set extrapolations.We now turn to CCSD extrapolations in conjunction with the larger A′VTZ and A′VQZ basis sets. The system-dependent basis set extrapolation (with an empirical scaling factor of λ = 1.169) results in an RMSD = 0.9 kJ mol^{-1}.For comparison, the global extrapolation with an optimized exponent of 3.403 results in an RMSD of 1.2 kJ mol^{-1} (Table I). Calculating the MP2/CBS energy from the A′V5Z and A′V6Z basis sets results in a slightly lower RMSD of 0.8 kJ mol^{-1}. Finally, extrapolating the CCSD energy from the A′V{Q,5}Z basis-set pair using the system-dependent extrapolation results in a near-zero RMSD of 0.2 kJ mol^{-1} (in conjunction with a scaling factor of 1.114). Furthermore, the largest deviations are very close to 1 kJ mol^{-1} (Table II). For comparison, the global extrapolation (with an empirical exponent of α = 3.220) results in an RMSD that is roughly twice as large (namely, 0.5 kJ mol^{-1}), a largest positive deviation of 2.7 kJ mol^{-1}, and a largest negative deviation of–1.0 kJ mol^{-1} (Table I).

C. Performance of basis-set additivity schemes for obtaining the CCSD correlation component of the TAEs in the W4-11 dataset

In the previous section we have shown that system-dependent extrapolations are superior to the global basis-set extrapolations, most notably in conjunction with the A′V{D,T}Z basis set pair. For example, the system-dependent approach results in RMSDs of 3.7 and 3.2 kJ mol^{-1} (Table II), whereas the global basis set extrapolation results in RMSDs of 9.1–13.5 kJ mol^{-1} (Table I). Another approach forestimating the CCSD/CBS energy through CCSD calculations with smaller basis sets and MP2 calculations with larger basis sets is the additivity scheme (eqs. (3) and (4)). Table III gives error statistics for the CCSD/X(MP2/Y) methods, the individual deviations for the systems in the W4-11 database are given in Table S5 of the supplementary material.[59] Note that the CCSD/X(MP2/Y) notation indicates that the A′VXZ and A′VYZ basis sets are the largest basis sets used in the CCSD and MP2 calculations, respectively.

Let us first consider the computationally most economical CCSD/D(MP2/Q) method. Using an exponent of 3 for extrapolating the MP2 energy gives a relatively large RMSD of 11.8 kJ mol^{-1}.

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

TABLE III. **Error statistics for the valence CCSD correlation component over the 140 total atomization energies in the W4-11 test set obtained via the CCSD/X(MP2/Y) additivity scheme (kJ mol^{-1}). The reference values are CCSD/A′V{5,6}Z total atomization energies from W4 theory[a,b]**

CCSD[c]	MP2[d]	α[e]	RMSD	MAD	MSD	LND	LPD
A′VDZ	A′V{T,Q}Z	3	11.76	10.23	10.20	–1.68 (BN)	29.54 (F_2O_2)
A′VDZ	A′V{T,Q}Z	5.75[f]	5.18	3.62	0.54	–11.94 ($AlCl_3$)	20.70 (F_2O_2)
A′VTZ	A′V{T,Q}Z	3	6.01	5.23	5.18	–2.44 ($AlCl_3$)	14.06 (C_3H_8)
A′VTZ	A′V{T,Q}Z	4.00[f]	2.37	1.77	–0.05	–9.02 ($AlCl_3$)	4.41 (F_2O_2)
A′VTZ	A′V{Q,5}Z	3	6.22	5.52	5.52	N/A	14.45 (P_4)
A′VTZ	A′V{Q,5}Z	6.50[f]	1.90	1.44	0.22	–7.40 ($AlCl_3$)	4.47 (F_2O_2)
A′VQZ	A′V{Q,5}Z	3	4.31	3.89	3.89	N/A	13.15 (P_4)
A′VQZ	A′V{Q,5}Z	5.00[f]	0.78	0.62	–0.12	–3.12 (SO_3)	1.70 (BN)

[a]Footnote a to Table I applies here.
[b]The notation CCSD/X(MP2/Y) indicates that the CCSD correlation energy is calculated with the A′VXZ basis set and the MP2 correlation energy is extrapolated from the A′V(Y–1)Z and A′VYZ basis sets.
[c]Basis set used in the CCSD calculations (eq. (3)).
[d]Basis sets used in the two-point MP2 extrapolations (eq. (4)).
[e]Extrapolation exponent used in the MP2 extrapolations (eq. (1)).
[f]Optimized to minimize the RMSD over the 140 TAEs in the W4-11 dataset.

Optimizing the exponent used in the MP2 extrapolation cuts the RMSD by more than 50% and leads to an RMSD of 5.2 kJ mol^{-1}. The performance of the CCSD/D(MP2/Q) method with an empirical exponent is quite good considering its low computational cost. The CCSD/T(MP2/Q) method results in an RMSD of 6.0 kJ mol^{-1} with a nonempirical exponent of $\alpha = 3$. However, using an optimized exponent of $\alpha = 4.00$ results in a remarkably low RMSD of merely 2.4 kJ mol^{-1}. For comparison, the system-dependent extrapolation from the A′V{D,T}Z basis sets, which has the same computational cost as the CCSD/T(MP2/Q) method, results in a larger RMSD of 3.7 kJ mol^{-1} (Table II).Extrapolating the MP2 energy from the larger A′V{Q,5}Z basis sets with an empirical extrapolation exponent further improves the performance (RMSD = 1.9 kJ mol^{-1}). Finally, calculating the CCSD energy in conjunction with the A′VQZ basis set and extrapolating the MP2 energy from the A′V{Q,5}Z basis sets with an empirical exponent of $\alpha = 5.00$ results in an RMSD of 0.78 kJ mol^{-1}. The system-dependent extrapolation with the same computational cost results in practically the same RMSD of 0.77 kJ mol^{-1} (Table II).

A useful way of looking at the performance of the CCSD/X(MP2/Y) additivity scheme is to rewrite eq. (3) as ΔCCSD $\approx$ ΔMP2, where ΔCCSD = CCSD/CBS – CCSD/A′VXZ and ΔMP2 is given by eq. (4). Table S6 of the supplementary material[59] lists the ΔMP2 and ΔCCSD terms for the molecules in the W4-11 database. The magnitude of the ΔCCSD term calculated in conjunction with the A′VDZ basis set spans a wide range from 11.5 (F_2) to 193.2 (P_4) kJ mol^{-1}. Inspection of Table S6 reveals that for many systems the ΔMP2 = MP2/A′V{T,Q}Z – MP2/A′VDZ basis-set correction term does not provide a very good approximation to this ΔCCSD term. For example, for nearly 10% of the systems the difference between the ΔMP2 and ΔCCSD terms is greater than 10 kJ mol^{-1} (most notably, for oxides such as Cl_2O, ClO_2, F_2O, FO_2, HO_3, O_3, and F_2O_2). On the other hand, the ΔMP2 = MP2/A′V{T,Q}Z – MP2/A′VTZ term provides an excellent approximation to the ΔCCSD term calculated in conjunction with the A′VTZ basis set. The difference between the ΔMP2 and ΔCCSD terms is smaller than 2 kJ mol^{-1} for 67% of the systems, and it exceeds 5 kJ mol^{-1} for only five molecules (S_4, SiF_4, SO_3, P_4, and $AlCl_3$). Upon removing these five second-row molecules from the evaluation set the RMSD for the CCSD/T(MP2/Q) method drops from 2.4 to 1.9 kJ mol^{-1}.

D. Performance of basis-set extrapolations and additivity schemes for the CCSD correlation energy for TAEs of larger molecules

The W4-11 dataset contains 140 first- and second-row species containing up to five non-hydrogen atoms. For example, the largest systems in the W4-11 dataset aremolecules such as acetic

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

acid ($H_4C_2O_2$), cyanogen (N_2C_2), dioxygen difluoride (F_2O_2), aluminium trichloride ($AlCl_3$), sulfur trioxide (SO_3), tetraphosphorus (P_4), tetrafluoromethane (CF_4), and tetrafluorosilane (SiF_4). It is also of interest to evaluate the performance of the basis-set extrapolations and additivity schemes for larger molecules. This is important since the CCSD component to the TAE increases with the size of the system. For example, the CCSD correlation component for the molecules in the W4-11 database ranges between 77.5 (AlH) and 874.7 (propane) kJ mol^{-1},whereas for the larger molecules considered in the present section (*vide infra*) it ranges between 906.2 (1,3-butadiene) and 2130.1 (adenine) kJ mol^{-1}.In addition, it is of interest to evaluate the performance of the empirical extrapolations in which a single adjustable parameter has been parameterized over the W4-11 dataset for molecules that are not included in the training set.

In particular, we will consider a set of 20 TAEs of larger molecules for which we have CCSD/CBS reference values from W3.2lite (i.e. CCSD/A′V{Q,5}Z) or W2-F12 (i.e. CCSD-F12/V{T,Q}Z-F12) theory, this database will be referred to as the TAE20dataset. These CCSD/CBS values allow us to evaluate the performance of conventional and system-dependent basis-set extrapolations in conjunction with the A′V{D,T}Z basis-set pair. The TAE20 dataset includes amino acids (alanine, cysteine, glycine, methionine, and serine); DNA/RNA bases (adenine, cytosine, thymine, and uracil); aromatic compounds (benzene, phenyl radical, pyridine, pyrazine, and furan); hydrocarbons (fulvene, 1,3-butadiene, and cyclobutene); and hydrocarbon cages (tetrahedrane, triprismane, and cubane).

Table IV gives error statistics for the CCSD/A′V{D,T}Z correlation energies extrapolated using the conventional basis-set extrapolations. Using the two-point extrapolation with an exponent of $\alpha = 3$ results in very poor performance, which is much worse than the performance for the systems

TABLE IV. **Error statistics for the global and system-dependent extrapolations of the valence CCSD correlation energies over the 20 total atomization energies in the TAE20 dataset (kJ mol^{-1}).**[a,b]

Basis sets[c]	[j]	RMSD	MAD	MSD	LND	LPD
			Global extrapolations			
A′V{D,T}Z[d]	3[k]	26.99	24.36	−24.36	−49.38 (adenine)	N/A
A′V{D,T}Z[d]	2.357[k]	15.13	12.83	12.68	−1.52 (uracil)	33.71 (cubane)
A′V{D,T}Z[d]	2.400[k]	12.60	10.03	9.52	−4.88 (uracil)	29.39 (cubane)
A′V{D,T}Z[e]	[l]	10.22	8.23	5.94	−8.69 (uracil)	24.48 (cubane)
			System-dependent extrapolations			
A′V{D,T}Z[f]	1.050[m]	5.70	5.27	5.27	N/A	9.70 (cubane)
A′V{D,T}Z[g]	1.050[m]	5.45	4.81	5.06	N/A	9.01 (cubane)
			Additivity schemes			
A′VDZ[h]	5.75[n]	14.61	12.69	−12.69	−29.60 (adenine)	N/A
A′VTZ[h]	4.00[n]	2.88	2.18	−1.28	−7.35 (adenine)	2.15 (methionine)
A′VTZ[i]	6.50[n]	2.53	1.93	−1.58	−5.32 (adenine)	1.20 (alanine)

[a]Footnote *a* to Table I applies here.
[b]The reference values in the TAE20 database are calculated at the CCSD/A′V{Q,5}Z (for benzene, fulvene, phenyl radical, pyridine, pyrazine, furan, 1,3-butadiene, and cyclobutene) or CCSD-F12/V{T,Q}Z-F12 (for adenine, cytosine, thymine, uracil, alanine, cysteine, glycine, methionine, serine, tetrahedrane, triprismane, and cubane) levels of theory.
[c]Basis sets used in the CCSD calculations.
[d]Extrapolated using eq. (1).
[e]Extrapolated using eq. (2).
[f]The system-dependent extrapolation exponents are obtained from MP2 calculations in conjunction with the A′VDZ, A′VTZ, and A′VQZ basis sets (eq. (6)).
[g]The system-dependent extrapolation exponents are obtained from MP2 calculations in conjunction with the A′VTZ, A′VQZ, and A′V5Z basis sets (eq. (6)).
[h]The ΔMP2 basis-set correction is calculated with the A′V{T,Q}Z basis sets (eq. (4)).
[i]The ΔMP2 basis-set correction is calculated with the A′V{Q,5}Z basis sets (eq. (4)).
[j]Single parameter used in the global and system-dependent extrapolations and in the additivity schemes (see text).
[k]Extrapolation exponent used in the two-point extrapolations (eq. (1)).
[l]The extrapolation fitting constant from Ref. 34 is $F = 1.5877616$ (eq. (2)).
[m]Scaling factor used in the sysrem-dependent extrapolations.
[n]Extrapolation exponent used in the MP2 extrapolations (eq. (4)).

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

in the W4-11 database. To be specific, the RMSD for the TAE20 set is 27.0 kJ mol^{-1}, compared to an RMSD of 13.5 kJ mol^{-1} for the W4-11 dataset. Using empirical exponents cuts the RMSD by about 60–70%, resulting in RMSDs of 15.1 (α = 2.357, this work), 12.6 (α = 2.357, Truhlar),[27] and 10.2 (α = 2.451, Schwenke) kJ mol^{-1}. In all cases the largest deviations of 24.5–33.7 kJ mol^{-1} are obtained for cubane.

Using the system-dependent extrapolations in conjunction with the A′V{D,T}Z basis-set pair results in RMSDs of 5.5–5.7 kJ mol^{-1} (depending on the basis sets used for extrapolating the MP2/CBS energy). These RMSDs are higher by about 2 kJ mol^{-1} relative to the RMSDs obtained for the W4-11 database (Table II), nevertheless, they represent a substantial improvement over the performance of the global extrapolations. The additivity approach significantly outperforms the system-dependent extrapolations at no increase in the computational cost. In particular, the CCSD/T(MP2/Q)method results in an RMSD of 2.9 kJ mol^{-1}, and the CCSD/T(MP2/5) additivity scheme results in an RMSD of 2.5 kJ mol^{-1}. We also note that the latter approach results in a largest deviation (in absolute value) of only 5.3 kJ mol^{-1}.

IV. CONCLUSIONS

We assess the performance of a number of approaches for obtaining CCSD correlation energies close to the complete basis set limit in conjunction with relatively small basis sets.These approaches are evaluated against two datasets of highly accurate CCSD/CBS limit total atomization energies. The first is the W4-11 database of 140 total atomization energies of small molecules. The species in the W4-11 database cover a broad spectrum of bonding situations and multireference character, and as such it is an excellent benchmark set for the validation of basis-set extrapolation techniques. The reference values in the W4-11 database are CCSD/A′ V{5,6}Z TAEs obtained from W4 theory. The second database is the TAE20 set of 20 TAEs of larger molecules. This set includes five amino acids, four DNA/RNA bases, six (hetero)aromatic compounds, three platonic hydrocarbon cages, and three simple hydrocarbons. The reference values in the TAE20 database are either CCSD/A′ V{Q,5}Z TAEs (from W3.2 theory) or CCSD-F12/V{T,Q}Z-F12 TAEs (from W2-F12 theory).

We consider the following approaches for obtaining CCSD/CBS energies in conjunction with relatively small basis sets: (i) global and system-dependent extrapolations based on the A + B/L^{α} two-point extrapolation formula, and (ii) the well-known additivity approach that uses an MP2-based basis-set-correction term. In the system-dependent extrapolationsthe CCSD energy is extrapolated using the two-point extrapolation formula $A + B/L\alpha_{eff}$, where α_{eff} is an effective extrapolation exponent that is uniquely determined for each molecule (or atom) from lower-cost MP2 calculations. We show that this system-dependent extrapolation scheme is superior to conventional basis-set extrapolations, which use a global extrapolation exponent for all systems.For example, for the W4-11 database we obtain the following RMSDs for the CCSD energies extrapolated from the aug′-cc-pV(D+d)Z and aug′-cc-pV(T+d)Z basis sets: 9.1 (best global extrapolation), 3.7 (best system-dependent extrapolation) kJ mol^{-1}. For the TAE20 database we obtain somewhat higher RMSDs, namely: 10.2 (best global extrapolation), 5.7 (best system-dependent extrapolation) kJ mol^{-1}. However, the additivity approach, which uses an MP2-based basis-set-correction term and has the same computational cost as the above system-dependent extrapolation, significantly outperforms the system-dependent extrapolation. In particular, it results in RMSDs of 2.4 (W4-11 dataset) and 2.9 (TAE20 dataset) kJ mol^{-1}. Both the system-dependent extrapolation and additivity approaches use a single adjustable parameter that minimizes the RMSD over the W4-11 database. The CCSD calculations in these two approaches are carried out in conjunction with basis sets of up to aug′-cc-pV(T+d)Z quality and the MP2 calculations with basis sets of up to aug'-cc-pV(Q+d)Z quality.

ACKNOWLEDGMENTS

We gratefully acknowledge the system administration support provided by the Faculty of Science at UWA to the Linux cluster of the Karton group, the financial support of Danish National

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

Research Foundation (Center for Materials Crystallography, DNRF-93) to PRS, and an Australian Research Council (ARC) Discovery Early Career Researcher Award (DE140100311) to AK.The authors would like to thank the referees of this manuscript for their valuable comments.

[1] I. Shavitt and R. J. Bartlett, *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory*, Cambridge molecular science series (Cambridge University Press, Cambridge, 2009).
[2] A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).
[3] A. Karton, P. R. Taylor, and J. M. L. Martin, Chem. Phys. **127**, 064104 (2007).
[4] A. Karton, S. Daon, and J. M. L. Martin, Chem. Phys. Lett. **510**, 165 (2011).
[5] 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).
[6] Y. J. Bomble, J. Vázquez, M. Kalláy, C. Michauk, P. G. Szalay, A. G. Cśaszár, J. Gauss, and J. F. Stanton, J. Chem. Phys. **125**, 064108 (2006).
[7] M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, J. Chem. Phys. **128**, 114111 (2008).
[8] W. Klopper, B. Ruscic, D. P. Tew, F. A. Bischoff, and S. Wolfsegger, Chem. Phys. **356**, 14 (2009).
[9] K. A. Peterson, D. Feller, and D. A. Dixon, Theor. Chem. Acc. **131**, 1079 (2012).
[10] A. Karton, I. Kaminker, and J. M. L. Martin, J. Phys. Chem. A **113**, 7610 (2009).
[11] T. Helgaker, W. Klopper, and D. P. Tew, Mol. Phys. **106**, 2107 (2008).
[12] J. M. L. Martin, *Annual Reports in Computational Chemistry* (Elsevier, New York, 2005), Vol. 1, pp. 31–43.
[13] A. Karton and J. M. L. Martin, J. Chem. Phys. **133**, 144102 (2010).
[14] J. M. L. Martin and M. K. Kesharwani, J. Chem. Theory Comput. **10**, 2085 (2014).
[15] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[16] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[17] T. H. Dunning, K. A. Peterson, and A. K. Wilson, J. Chem. Phys. **114**, 9244 (2001)
[18] D. Feller, J. Chem. Phys. **96**, 6104 (1992); **98**, 7059 (1993).
[19] K. A. Peterson, D. E. Woon, and T. H. Dunning, J. Chem. Phys. **100**, 7410 (1994).
[20] J. M. L. Martin, J. Chem. Phys. **100**, 8186 (1994).
[21] J. M. L. Martin, Chem. Phys. Lett. **259**, 669 (1996).
[22] J. M. L. Martin and T. J. Lee, Chem. Phys. Lett. **258**, 136 (1996).
[23] J. M. L. Martin and P. R. Taylor, J. Chem. Phys. **106**, 8620 (1997).
[24] A. K. Wilson and T. H. Dunning, J. Chem. Phys. **106**, 8718 (1997).
[25] T. Helgaker, W. Klopper, H. Koch, and J. Nago, J. Chem. Phys. **106**, 9639 (1997).
[26] A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).
[27] D. G. Truhlar, Chem. Phys. Lett. **294**, 45 (1998).
[28] D. Feller and K. A. Peterson, J. Chem. Phys. **108**, 154 (1998).
[29] P. L. Fast, M. L. Sánchez, and D. G. Truhlar, J. Chem. Phys. **111**, 2921 (1999).
[30] D. Feller and K. A. Peterson, J. Chem. Phys. **110**, 8384 (1999); **126**, 114105 (2007).
[31] A. J. C. Varandas, J. Chem. Phys. **113**, 8880 (2000).
[32] W. Klopper, Mol. Phys. **99**, 481 (2001).
[33] E. F. Valeev, W. D. Allen, R. Hernandez, C. D. Sherrill, and H. F. Schaefer, J. Chem. Phys. **118**, 8594 (2003).
[34] D. W. Schwenke, J. Chem. Phys. **122**, 014107 (2005).
[35] D. Feller, K. A. Peterson, and T. D. Crawford, J. Chem. Phys. **124**, 054107 (2006)
[36] D. Bakowies, J. Chem. Phys. **127**, 164109 (2007); **127**, 084105 (2007).
[37] E. C. Barnes, G. A. Petersson, D. Feller, and K. A. Peterson, J. Chem. Phys. **129**, 194115 (2008).
[38] D. Feller, K. A. Peterson, and J. G. Hill, J. Chem. Phys. **135**, 044102 (2011).
[39] D. Feller, J. Chem. Phys. **138**, 074103 (2013).
[40] D. S. Ranasinghe and G. A. Petersson, J. Chem. Phys. **138**, 144104 (2013).
[41] F. N. N. Pansini and A. J. C. Varandas, Chem. Phys. Lett. (2015), in press
[42] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[43] L. Goerigk, A. Karton, J. M. L. Martin, and L. Radom, Phys. Chem. Chem. Phys. **15**, 7028 (2013).
[44] D. G. Liakos and F. Neese, J. Phys. Chem. A **116**, 4801 (2012).
[45] B. Brauer, M. K. Kesharwani, and J. M. L. Martin, J. Chem. Theory Comput. **10**, 3791 (2014).
[46] A. Karton and L.-J. Yu, Chem. Phys. **441**, 166 (2014).
[47] X. He, L. Fusti-Molnar, and K. M. Merz, Jr., J. Phys. Chem. A **113**, 10096 (2009).
[48] A. Karton and L. Goerigk, J. Comp. Chem. **36**, 622 (2015).
[49] MOLPRO is a package of ab initio programs written by, H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, and G. Rauhut, *et al*. See: http://www.molpro.net.
[50] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242 (2012).
[51] J. M. L. Martin and G. de Oliveira, J. Chem. Phys. **111**, 1843 (1999).
[52] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).
[53] A. Karton and J. M. L. Martin, J. Chem. Phys. **136**, 124114 (2012).
[54] L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov, and J. A. Pople, J Chem Phys **110**, 4703 (1999).
[55] L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **127**, 124105 (2007).
[56] L. A. Curtiss, P. C. Redfern, and K. Raghavachari, WIREs Comput. Mol. Sci. **1**, 810 (2011).
[57] A. Karton, L.-J. Yu, M. K. Kesharwani, and J. M. L. Martin, Theor. Chem. Acc. **133**, 1483 (2014).
[58] A. Karton, P. R. Schreiner, and J. M. L. Martin, J. Comp. Chem. (2015) accepted.

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01

[59] See supplementary material at http://dx.doi.org/10.1063/1.4921697 for the individual errors for the global extrapolations (Table S1); ideal extrapolation exponents for system-dependent extrapolations that will reproduce the CCSD/CBS energies (Table S2 and Figure S1); individual errors for the system-dependent extrapolations (Table S3); extrapolation exponents for system-dependent extrapolations obtained from MP2 calculations (Table S4); individual errors for the CCSD/X(MP2/Y) additivity scheme (Table S5); and ΔCCSD and ΔMP2 basis-set-correction terms used in the CCSD/X(MP2/Y) additivity schemes (Table S6).

[60] In W1w theory the HF energy is extrapolated from the A′V{T,Q}Z basis sets using eq. (1) with $\alpha = 5$, and the (T) contribution is extrapolated from the A′V{D,T}Z basis sets using the same extrapolation with $\alpha = 3.22$.

All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported license. See: http://creativecommons.org/licenses/by/3.0/ Downloaded to IP: 203.142.140.128 On: Wed, 20 May 2015 13:50:01