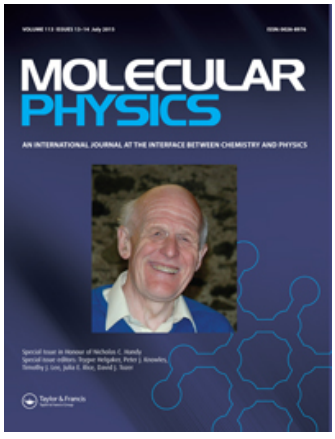
This article was downloaded by: [Weizmann Institute of Science]
On: 15 July 2015, At: 03:36
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London, SW1P 1WG



Molecular Physics: An International Journal at the Interface Between Chemistry and Physics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tmph20

The cc-pV5Z-F12 basis set: reaching the basis set limit in explicitly correlated calculations

Kirk A. Peterson[a], Manoj K. Kesharwani[b] & Jan M.L. Martin[b]
[a] Department of Chemistry, Washington State University, Pullman, WA, USA
[b] Department of Organic Chemistry, Weizmann Institute of Science, Reḥovot, Israel
Published online: 06 Dec 2014.

CrossMark
Click for updates

To cite this article: Kirk A. Peterson, Manoj K. Kesharwani & Jan M.L. Martin (2015) The cc-pV5Z-F12 basis set: reaching the basis set limit in explicitly correlated calculations, Molecular Physics: An International Journal at the Interface Between Chemistry and Physics, 113:13-14, 1551-1558, DOI: 10.1080/00268976.2014.985755

To link to this article: http://dx.doi.org/10.1080/00268976.2014.985755

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Molecular Physics, 2015
Vol. 113, Nos. 13–14, 1551–1558, http://dx.doi.org/10.1080/00268976.2014.985755

Taylor & Francis
Taylor & Francis Group

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

INVITED ARTICLE

The cc-pV5Z-F12 basis set: reaching the basis set limit in explicitly correlated calculations

Kirk A. Peterson[a], Manoj K. Kesharwani[b] and Jan M.L. Martin[b,*]

[a]*Department of Chemistry, Washington State University, Pullman, WA, USA;* [b]*Department of Organic Chemistry, Weizmann Institute of Science, Reḥovot, Israel*

(*Received 18 September 2014; accepted 4 November 2014*)

We have developed and benchmarked a new extended basis set for explicitly correlated calculations, namely cc-pV5Z-F12. It is offered in two variants, cc-pV5Z-F12 and cc-pV5Z-F12(rev2), the latter of which has additional basis functions on hydrogen not present in the cc-pVnZ-F12 (n = D,T,Q) sequence. A large uncontracted 'reference' basis set is used for benchmarking. cc-pVnZ-F12 (n = D–5) is shown to be a convergent hierarchy. Especially the cc-pV5Z-F12(rev2) basis set can yield the valence CCSD (coupled cluster with all single and double substitutions) component of total atomisation energies, *without any extrapolation*, to an accuracy normally associated with aug-cc-pV{5,6}Z extrapolations. Hartree-Fock self-consistent field (SCF) components are functionally at the basis set limit, while the MP2 limit can be approached to as little as 0.01 kcal/mol without extrapolation. The determination of (T) appears to be the most difficult of the three components and cannot presently be accomplished without extrapolation or scaling. (T) extrapolation from cc-pV{T,Q}Z-F12 basis sets, combined with CCSD-F12b/cc-pV5Z-F12 calculations, appears to be an accurate combination for explicitly correlated thermochemistry. For accurate work on noncovalent interactions, the basis set superposition error with the cc-pV5Z-F12 basis set is shown to be so small that counterpoise corrections can be neglected for all but the most exacting purposes.

Keywords: explicitly correlated; basis sets; coupled cluster; thermochemistry; correlation consistent

Introduction

Wavefunction *ab initio* calculations are for the most part a two-dimensional convergence problem (the 'Pople diagram' [1]) with basis set convergence on one axis and the electron correlation method (a.k.a. n-particle treatment) on the other. For heavy elements, relativity [2] becomes the third major dimension, leading to the Császár cube [3].

For systems with a mild nondynamical correlation, the 'gold standard' CCSD(T) [4,5] (i.e., coupled cluster [6] with all single and double substitutions, augmented with quasiperturbative triples) method is quite close to the full configuration interaction limit, leaving basis set convergence as the main bottleneck. With conventional one-particle basis sets, convergence is quite slow [7–9], although its comparatively slow and monotonic character naturally suggests extrapolation procedures when correlation-consistent basis sets are used [10–12].

Explicitly correlated methods [13–20], in which some 'geminal' terms that explicitly depend on the interelectronic distance are added to the one-particle basis set, exhibit greatly accelerated basis set convergence. (Early work, including his own, has been reviewed by Handy [21]; for later reviews, see Refs. [22–26].)

A number of forays into explicitly correlated computational thermochemistry were recently made by the present authors [27–30] and by others [31,32]. The somewhat disappointing conclusion that emerged is while F12 methods allow one to quickly reach the general neighbourhood of the basis set limit, approaching more closely in a consistent way becomes quite challenging. Specifically, while basis set convergence in pure one-particle approaches (e.g., in W4 theory [33,34], the HEAT approach [35,36], and the FPD approach [30,37–40]) tends to be smooth with monotonically increasing total atomisation energies (TAEs), basis set convergence of coupled cluster TAEs from F12 methods can be oscillatory or even (anomalously) monotonically decreasing.

One reason for this problem resides in the use of basis sets not specifically optimised for explicitly correlated calculations. A team involving one of us sought to address this problem by the development of the cc-pVnZ-F12 and cc-pCVnZ-F12 basis sets [41–43] (n = D,T,Q), which were optimised at the MP2-F12 level in the presence of the appropriate geminal terms. These basis sets do appear to have a smoother convergence behaviour than orbital-optimised aug-cc-pVnZ basis sets (n = D,T,Q,5,6) but obviously do not reach as far, and basis set extrapolation was still found to be necessary [44]. A crude 'rule of thumb' has emerged [22–24] from practical applications, namely, that F12 calculations gain the user between two and three 'zetas' over their conventional counterparts.

* Corresponding author. Email: gershom@weizmann.ac.il

© 2015 Taylor & Francis

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

If that is so, expanding the cc-pVnZ-F12 basis set by an additional member (n = 5) may bring us within the subchemical accuracy of the basis set limit, and indeed altogether eliminate the need for basis set extrapolation. In the present paper, we will develop and present such a basis set for H and B–Ne, and by thorough benchmarks show that it is indeed sufficiently close to the basis set limit for this purpose.

Computational details

Most calculations were carried out using MOLPRO version 2012.1 [45] running on the Faculty of Chemistry HPC cluster at the Weizmann Institute of Science. Some additional CCSD(F12) and open-shell atomic CCSD(F12*) calculations were carried out using TURBOMOLE version 6.4 [46] on the same hardware.

For the cc-pVnZ-F12 correlation consistent basis sets (n = D,T,Q) optimised for F12 calculations [41], we employed the auxiliary basis sets [47] and complementary auxiliary basis sets (CABS) [48] developed for use with them, as well as the Weigend [49,50] JK-fitting basis sets which are the MOLPRO default. Unless indicated otherwise, the geminal exponent values β recommended in Ref. [44] (rather than those of Ref. [41]) were used: $\beta = 0.9$ for cc-pVDZ-F12, $\beta = 1.0$ for cc-pVTZ-F12, and $\beta = 1.0$ for cc-pVQZ-F12. The Hartree-Fock self-consistent field (SCF) component was improved through the 'CABS correction' [19,51].

For the reference (REF) basis sets (see below), we employed very large uncontracted auxiliary basis sets previously reported in Ref. [44]. For the V5Z-F12 basis set, we considered two combinations of auxiliary basis sets: the first is that used for the REF basis sets and the second is a combination of Weigend's aug-cc-pV5Z/JKFIT basis set [50] for the Coulomb and exchange elements and Hättig's aug-cc-pwCV5Z/MP2FIT basis set [52] for both the RI-MP2 parts and for the CABS. For comparison, some F12 calculations were run with ordinary aug-cc-pVnZ basis sets [53], where the JKFIT basis set [50] was extended by a single even-tempered layer of diffuse functions, the RI-MP2 basis set was again taken from Ref. [52], and the CABS basis sets of Yousaf and Peterson [54] were employed.

Our discussion focuses on the CCSD-F12b approximation [19,20], but we also performed some exploratory calculations using the CCSD(F12*) (a.k.a. CCSD-F12c) [55] approximation. F12 approaches as presently practised do not directly affect the connected quasiperturbative triples, so the basis set convergence behaviour of the (T) contribution is effectively that of a conventional calculation. Marchetti and Werner [56] proposed a convergence acceleration by scaling the (T) contribution by the MP2-F12/MP2 correlation energy ratio, and found that this considerably improves calculated interaction energies for noncovalent complexes. Such scaling will be indicated by the notation (T*) instead of (T). (In two recent studies [57,58], we found it to be beneficial for F12 harmonic frequency calculations and for noncovalent interaction energies [58] as well.)

The MP2-F12 correlation energies discussed are those obtained with the 3C ansatz [18] with fixed amplitudes [16], a.k.a. '3C(Fix)'.

The 'frozen core' approximation was applied, i.e., all inner-shell orbitals were constrained to be doubly occupied.

For reference purposes, we took the large even-tempered uncontracted spdfgh basis set already used in previous work by Peterson and co-workers [59], expanded with four additional *i* functions [44] (denoted by REF). In this work, truncations at a given angular momentum will be indicated as REF-*f* (keeping up to *f* functions inclusive), REF-*g*, REF-*h*, and REF-*i*. For F12 calculations, at most REF-*h* is used in conjunction with very large uncontracted RI, JK, and CABS basis sets [44]; for conventional calculations, up to REF-*i* inclusive is considered.

For comparison, conventional orbital-based SCF, CCSD, and CCSD(T) results were obtained using the aug-cc-pV5Z [53] and aug-cc-pV6Z [60,61] basis sets (AV5Z and AV6Z for short), as well as with the aug-cc-pCV5Z [62,63] and aug-cc-pCV6Z [64] core-valence basis sets (ACV5Z and ACV6Z for short) and the core-valence-weighted aug-cc-pwCVQZ and aug-cc-pwCV5Z basis sets [63]. In the conventional calculations, we omitted diffuse functions on hydrogen, a common practice variously denoted by aug'-cc-pVnZ [65], jul-cc-pVnZ [66], or heavy-aug-cc-pVnZ [67]. No such omission was made in the F12 calculations with ordinary aug-cc-pVnZ basis sets, which were carried out purely for comparison purposes.

The two principal data-sets for benchmarking were derived from the W4-11 benchmark of Karton *et al*. [68], the reference geometries having been taken from the Supporting Information of that work. The 'validation set' TAE71 consists of 71 closed-shell first-row molecules among the 140 in the W4-11 set. Out of these, we selected a TAE28 'training set' of 28 molecules for which we were able to obtain CCSD-F12b total energies with the REF-*h* basis set, and CCSD(T) with REF-*i*. They are: BF, BH_3, BH, BN, C_2H_2, C_2H_4, C_2, CF_2, CH_2NH, CH_4, CO_2, CO, F_2O, F_2, H_2CO, H_2O_2, H_2O, H_2, HCN, HF, HNC, HNO, HOF, N_2O, N_2, NH_3, O_3, and $CH_2(^1A_1)$.

An additional benchmark is considered in the guise of the A24 weak interaction data-set [69] of Hobza and co-workers; reference geometries for such sets were downloaded from the BEGDB website [70]. (We note that a subset of the S22 × 5 [71,72] weak interaction benchmark was also very recently treated by Brauer *et al*. [58] with the present V5Z-F12 basis sets.)

Basis set optimisation

The development of the new cc-pV5Z-F12 orbital basis sets was similar to the previous optimisations of the n = D–Q sets [41]. In the present work, the HF basis sets (s functions

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

for H and He, s and p for B−Ne) were taken from the standard contracted cc-pV6Z (H) and aug-cc-pV6Z (He, B−Ne) basis sets [60]. Correlating functions optimised for the MP2-F12/3C total energy were then added to these HF sets: (5p3d2f1g) for H and He, and (5d4f3g2h) for B−Ne. In each case, the exponents were constrained to follow an even-tempered sequence. For B–Ne, a single p function was also uncontracted from the HF aug-cc-pV6Z primitive sets (seventh most diffuse). Except for He and Ne, the optimisations were carried out on the electronic ground states of the homonuclear diatomics using the geometries of Ref. [48] As in Ref. [41], a slightly stretched ($1.25 \times r_e$) geometry was utilised for H_2. For consistency with Ref. [41], all optimisations employed a geminal exponent of 1.4 with the reference DF and RI basis sets of Ref. [44]. (This choice of the geminal exponent keeps the optimised orbital exponents somewhat more diffuse so that the F12 factor covers the short-range correlation, leaving the basis set to take care of the long range.)

An alternative series of basis sets was also determined for the hydrogen atom, denoted by cc-pVnZ-F12(rev2), for n = D–5. These sets were optimised as above, but with a larger number of correlating functions in analogy to B–Ne (at TZ and above), i.e., (2p) for DZ, (3p2d) for TZ, (4p3d2f) for QZ, and (5p4d3f2g) for 5Z.

The geminal exponent β was optimised at the MP2-F12 level following the same procedure as in Ref. [44], found to be $\beta = 1.2$.

Results and discussion

The SCF component of TAE

In orbital-based *ab initio* calculations, convergence of the correlation energy is so much slower compared to that of the Hartree–Fock energy that with the sizes of basis sets (like aug-cc-pV6Z) routinely used in such calculations, the Hartree–Fock component is typically well converged (see, however, Ref. [73] for a caveat). In explicitly correlated calculations, on the other hand, the Hartree–Fock component typically needs to be refined with a CABS-type correction [19,51] in order to obtain the SCF component at a similar accuracy to the MP2-F12 or CCSD-F12b correlation energies.

We thus consider the TAEs of the TAE28 set. Our reference data are orbital SCF/REF-*i* calculations. Root-mean-square deviations (RMSDs) at various levels of theory are given in Table 1. As the SCF/REF-*h* results differ by less than 0.001 kcal/mol RMSD, we can safely consider the SCF/REF-*i* values to represent the basis set limit.

For conventional AV{5,6}Z, the RMSDs are 0.029 and 0.010 kcal/mol, respectively. Note, however, that switching to the core-valence versions of these basis sets, ACV{5,6}Z, drastically reduces these errors to 0.008 and 0.002 kcal/mol, respectively, illustrating the importance of additional radial flexibility for accurate SCF energies. It has been noted by both groups involved that W4 (which takes valence extrapolation as the starting point) and HEAT (which takes all-electron calculations as the starting point) sometimes yield nontrivially different values for the same molecule at the CCSD(T) limit: it appears this must be ascribed to the additional flexibility of the core-valence basis set for the SCF and valence correlation components, rather than the inner-shell correlation energy. (By way of illustration: the SCF/ACVQZ–SCF/AVQZ difference can become as large as 0.19 kcal/mol for CO_2 and 0.18 kcal/mol for N_2O.)

For HF + CABS/cc-pV5Z-F12, the RMSD is just 0.001 kcal/mol, functionally equivalent to the HF limit. For cc-pVQZ-F12, this increases to 0.012 kcal/mol, and for cc-pVTZ-F12 to 0.052 kcal/mol. We can thus conclude that HF + CABS is effectively at the basis set limit for cc-pV5Z-F12 (and *a fortiori* for V5Z-F12(rev2)).

Table 1. RMSDs (kcal/mol) for SCF and MP2-F12 valence correlation components of total atomisation energies for the TAE28 set.

	HF + CABS	MP2-F12 $\beta = 1.4$	MP2-F12 β opt
VTZ-F12	0.052	0.157	0.166
VQZ-F12	0.012	0.031	0.050
V{T,Q}Z-F12		0.038	0.020
V5Z-F12	0.001	0.009	0.014
V5Z-F12(rev2)	0.001	0.011	0.011
REF-*f*	0.006	0.100	
REF-*g*	0.000	0.019	
REF-*h*	0.000	0.004	
awCVQZ	0.061	0.050	
awCV5Z	0.003	0.012	
AVQZ	0.038	0.120	
AV5Z	0.013	0.061	
	Orbital-HF		
AV5Z	0.029		
AV6Z	0.010		
ACV5Z	0.008		
ACV6Z	0.002		

The valence MP2-F12 component

RMSDs for the MP2 components for the TAE28 set can again be found in Table 1. As the reference values, we carried out two-point extrapolations from REF-{*g*,*h*} values assuming an L^{-7} dependence. The extrapolation covers just 0.004 kcal/mol, and even the REF-*g* results are within 0.02 kcal/mol RMSD of the reference.

The cc-pV5Z-F12 basis set is within 0.01 kcal/mol of the reference values, on average. For comparison, the similar-sized AV5Z basis set still has an RMSD error of 0.06 kcal/mol, which drops down to 0.01 kcal/mol when core-valence functions are added. Progressively deleting those

reveals that almost all the improvement resides in greater radial flexibility in the s and p spaces: d-type core-valence functions have little effect, and f and higher, basically none.

From V{T,Q}Z-F12 results and using the extrapolation of Hill *et al*. [44], an RMSD of just 0.02 kcal/mol can be obtained at a much lower expense: the attraction of V5Z-F12 then merely rests in being able to forgo extrapolation entirely.

The valence CCSD-F12b component

Error statistics are given in Table 2; the final column, the number of basis functions for one molecule (C_2H_4), is given as an approximate indicator of computational cost.

As was seen repeatedly before, the CCSD valence correlation component actually displays *slower* basis set convergence in F12 calculations than the MP2 component.

For the TAE28 'training set', we were able to obtain REF-*g* and REF-*h* values with both $\beta = 1.2$ and $\beta = 1.4$. The RMSD between the two sets is just 0.004 kcal/mol, confirming that the sensitivity toward the geminal exponent β is very weak as long as the orbital basis is sufficiently large. We have (somewhat arbitrarily) used the $\beta = 1.4$ values as our reference.

Performance of conventional AVnZ basis sets clearly is unacceptably slow through AVQZ inclusive: the RMSD for AVQZ still exceeds 0.5 kcal/mol, even as that for AV5Z drops down to 0.09 kcal/mol. In contrast, awCVQZ and awCV5Z have RMSD errors of just 0.06 and 0.02 kcal/mol: for comparison, the Schwenke extrapolation [74] from conventional A'V{5,6}Z results yields RMSD = 0.06 kcal/mol. Like for the HF components, stripping core-valence basis functions one angular momentum at a time revealed that the main improvement from AVQZ to awCVQZ results from greater radial flexibility in the (s,p) sets, followed by the d functions.

Considering the VnZ-F12 sequence with recommended geminal exponents {0.9,1.0,1.0,1.2}, we find RMSDs tapering off rapidly from 2.0 (D) via 0.6 (T) and 0.14 (Q) to 0.05 (n = 5) kcal/mol. When instead $\beta = 1.4$ (often favoured in

Table 2. RMSDs (kcal/mol) for CCSD-F12b valence correlation components of total atomisation energies.

	TAE28	TAE71	TAE71	C_2H_4
	With respect to different reference levels			
	REF{*g*,*h*} $\beta = 1.4$	awCV5Z $\beta = 1.4$	AV{5,6}Z Schwenke	Virtuals (d)
VDZ-F12 $\beta = 0.9$	1.991	3.184	3.159	88
VTZ-F12 $\beta = 1.0$	0.589	0.927	0.900	170
VTZ-F12 $\beta = 1.4$	0.418	0.664	2.242	170
VQZ-F12 $\beta = 1.0$	0.143	0.223	0.214	302
VQZ-F12 $\beta = 1.4$	0.063	0.087	0.107	302
V5Z-F12 $\beta = 1.2$	0.052	0.073	0.090	496
V5Z-F12 $\beta = 1.4$	0.032	0.036	0.078	496
V5Z-F12(rev2) $\beta = 1.2$	0.033	0.044	0.066	580
V5Z-F12(rev2) $\beta = 1.4$	0.021	0.017	0.069	580
V{T,Q}Z-F12 $\beta = 1.0$ (a)	0.125	0.183	0.234	302
V{T,Q}Z-F12 $\beta = 1.4$ (a)	0.066	0.081	0.141	302
Schwenke A'V{5,6}Z (c)	0.058	0.071	REFERENCE	734
awCVQZ $\beta = 1.4$	0.056	0.066	0.096	394
awCV5Z $\beta = 1.4$	0.013	REFERENCE	0.018	674
AVQZ $\beta = 1.4$	0.543	0.443	0.487	336
AV5Z $\beta = 1.4$	0.086	0.139	0.192	466
AV{Q,5}Z $\beta = 1.4$ (a)	0.228	0.154	0.170	466
REF-*f* $\beta = 1.2$	0.313			670
REF-*f* $\beta = 1.4$	0.287			670
REF-*g* $\beta = 1.2$	0.067			958
REF-*g* $\beta = 1.4$	0.054			958
REF-*h* $\beta = 1.2$	0.017			1244
REF-*h* $\beta = 1.4$	0.011			1244
REF-*i*	—			1348
REF-{*g*,*h*} $\beta = 1.2$ (b)	0.004			F12b (b)
REF-{*g*,*h*} $\beta = 1.4$ (b)	REFERENCE			F12b (b)

Note: ROHF-UCCSD(T) used for atomic energies throughout.
(a) Extrapolated by $E_\infty = E(L) + [E(L) - E(L-1)]/[(L/L-1)^\alpha - 1]$ with α for that basis set pair taken from Ref. [44].
(b) Extrapolated by $E_\infty = E(L) + [E(L) - E(1-1)]/[(L/L-1)^7 - 1]$.
(c) Conventional orbital-based CCSD calculation.
(d) Indicative of basis set size. Cost can be assumed to scale as N^4.

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

large basis set studies) was employed, these RMSDs were, at least relatively speaking, significantly reduced, to 0.4 kcal/mol for VTZ-F12 via 0.11 kcal/mol for VQZ-F12 to 0.03 kcal/mol for V5Z-F12.

By way of an additional perspective, with $\beta = 1.4$, REF-*f* reaches 0.29 kcal/mol and REF-*g* 0.05 kcal/mol, suggesting that VTZ-F12 and VQZ-F12 still are some distance removed from spdf and spdfg radial saturation, respectively.

The VnZ-F12(rev2) basis sets, with their additional polarisation functions for hydrogen, actually yield slightly worse error statistics for VDZ-F12 (due to 'error decompensation') and only marginal improvements compared to VTZ-F12 and VQZ-F12, but for V5Z-F12(rev2) the improvement is more notable in relative terms. Especially for such species as C_2H_4, the difference is significant.

As REF-*h* is not a practical option for most species in the TAE71 set, a secondary standard needs to be chosen. In view of the very low RMSD of 0.013 kcal/mol for awCV5Z ($\beta = 1.4$), this will be our first choice; a second choice, less as an actual secondary standard than as a comparison point to conventional orbital-based calculations, is AV{5,6}Z with Schwenke's extrapolation [74].

Compared to awCV5Z ($\beta = 1.4$), for TAE71, VnZ-F12 RMSDs are 3.2, 0.93, 0.22, and 0.07 kcal/mol for n = D, T, Q, and 5, respectively; for V5Z-F12(rev2), a further lowering to 0.04 kcal/mol is seen which (naturally) derives from the many species with several hydrogens that are part of TAE71 but not its subset TAE28. For $\beta = 1.4$, we can reduce these values to 0.66 for VTZ-F12, 0.09 for VQZ-F12, and 0.04 kcal/mol for V5Z-F12, with V5Z-F12(rev2) reaching a low of 0.017 kcal/mol that is similar to the uncertainty in the secondary standard itself.

What about extrapolation from V{T,Q}Z-F12 results? With the extrapolation exponent 4.596 given in Ref. [44], RMSD = 0.18 kcal/mol for $\beta = 1.0$ and 0.08 kcal/mol for $\beta = 1.4$, which are relatively small improvements over the raw VQZ-F12 results. For V5Z-F12, extrapolation appears to be entirely redundant.

The bottom line from this entire section is that with the V5Z-F12 basis set and especially the V5Z-F12(rev2) basis set, the valence CCSD atomisation energies of small molecules can be obtained *without any extrapolation* at an accuracy comparable or superior to the conventional extrapolation techniques involving 5Z and 6Z basis sets used in methods like W4 theory [33,34], FPD [30,37–40], and HEAT [35,36].

In terms of relative computational cost, V5Z-F12 or even V5Z-F12(ref2) will be an order of magnitude cheaper than REF-*h*. V5Z-F12(ref2) is of course equivalent to V5Z-F12 if no hydrogen atoms are present, but for ethylene, for example, the computational cost will approximately double, but the additional energy lowering is significant in high-accuracy computational thermochemistry. It is still only about half as expensive as the awCV5Z basis set being used as a secondary standard.

As an aside, we considered the difference between the CCSD-F12b [19,20] and the more rigorous CCSD(F12*) [55] (a.k.a. CCSD-F12c) methods, as it converges along the cc-pVnZ-F12 series. (As MOLPRO has no open-shell CCSD-F12c implementation, the F12c–F12b differences for the atoms were calculated using TURBOMOLE [46].) Compared to the orbital-based A'V{5,6}Z extrapolated contributions to the TAE71 atomisation energies (which we estimate to be reliable to only about 0.1 kcal/mol, but obviously are not biased toward one F12 approximation or another), CCSD(F12*) is noticeably closer than CCSD-F12b for n = D (by 1.2 kcal/mol) and n = T (by 0.4 kcal/mol), but for n = Q and n = 5, the RMSD values are comparable for both approaches. For molecules dominated by dynamical correlation, the differences taper off to nearly nothing for n = 5. For ozone and other molecules with significant post-CCSD(T) correlation terms [34,68], however, nontrivial differences (0.1–0.3 kcal/mol) between CCSD-F12b and CCSD(F12*) appear to remain at the basis set limit. We speculate that this may reflect the limitations of an MP2-F12 based *ansatz* for systems where MP2 is a poor approximation to the correlation energy. CCSD(F12*) entails fewer approximations than CCSD-F12b and thus should in principle be more rigorous. In practice, however, the presently available data are inconclusive.

At the request of a reviewer, we considered the effect of not neglecting the CABS contributions to the projector in the CCSD-F12 coupling term in the CCSD-F12b calculations (IXPROJ = 1). For cc-pVDZ-F12, the effect on the TAE71 atomisation energies is 0.48 kcal/mol RMSD, which drops to 0.048 kcal/mol for cc-pVTZ-F12 and to just 0.004 kcal/mol for cc-pVQZ-F12.

The valence (T) component

In Table 3 can be found error statistics for the (T) component, both RMSDs and (for the sake of detecting systematic bias) mean-signed deviations (MSDs). As this contribution does not benefit from the geminal terms, basis set convergence of the unscaled quasiperturbative triples is expected to be similar to that in a conventional calculation.

As the reference, we employ REF-*h* and REF-*i* conventional calculations, extrapolated to the complete basis limit using $1/L^{-3}$. The extrapolation in this case covers just 0.022 kcal/mol RMS.

It should be noted that conventional AV5Z and AV6Z calculations extrapolated using the expression of Ranasinghe and Petersson [75] agree with our reference values to 0.01 kcal/mol RMSD (−0.007 signed).

From the MSDs, it is clear that unscaled (T)-F12b corrections with the VnZ-F12 basis sets yield unacceptable systematic underestimates of (T), even with the V5Z-F12 and V5Z-F12(rev2) basis sets. Marchetti–Werner scaling [56], indicated by the symbol (T*), leads instead to mild

Table 3. MSD and RMSD for the (T) components of the TAE28 atomisation energies for different basis sets and approximations (kcal/mol), as well as global (T) scale factors (dimensionless) for the (Ts) approximation.

	MSD Orbital	MSD (T)F12b	MSD (T*)F12b	MSD (Ts)F12b	RMSD Orbital	RMSD (T)F12b	RMSD (T)*F12b	RMSD (Ts)F12b	Scale factor (Ts)F12b
REF-*f*	−0.171	−0.228	+0.179		0.199	0.267	0.237		
REF-*g*	−0.084	−0.112	+0.081		0.086	0.118	0.118		
REF-*h*	−0.048	−0.063	+0.046		0.040	0.058	0.074		
REF-*i*	−0.018	–	–		0.022	–	–		
REF-{*h*,*i*}	REF	–	–		REF	–	–		
VDZ-F12	–	−1.217	+0.149	−0.135	–	1.358	0.355	0.436	1.1413
VTZ-F12	–	−0.492	+0.135	−0.050	–	0.532	0.202	0.133	1.0527
VQZ-F12	–	−0.227	+0.090	−0.026	–	0.235	0.131	0.054	1.0232
V5Z-F12	–	−0.137	+0.065	−0.018	–	0.135	0.094	0.027	1.0136
V5Z-F12r2	–	−0.128	+0.062	−0.013	–	0.129	0.092	0.024	1.0131
{T,Q}	–	−0.011	–		–	0.024	–		
{Q,5}	–	−0.024	–		–	0.036	–		
AV{5,6}Z Petersson	−0.007	–	–		0.010	–	–		

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

systematic overestimates, with (T*) coming closer to the basis set limit as the basis set is increased. It appears to matter little for the statistics whether the atomic (T) contributions are scaled by their own E_{corr}(MP2-F12)/E_{corr}(MP2) ratios or that of the molecule, even though only the latter choice is strictly size consistent.

Another approach would be to use a single global scale factor for each basis set, fitted to minimise the RMSD. The error statistics obtained in this manner are much more favourable, especially for the larger basis sets: for V5Z-F12, the RMSD drops to 0.027 kcal/mol and is reduced slightly further to 0.024 kcal/mol for V5Z-F12(rev2). The fitted scale factors are noticeably smaller than those obtained from E_{corr}(MP2-F12)/E_{corr}(MP2) according to the Marchetti–Werner prescription, but fairly similar to those obtained from E_{corr}(CCSD-F12b)/E_{corr}(CCSD).

What about basis set extrapolation from smaller basis sets? For V{T,Q}Z-F12 using the extrapolation exponent 2.895 given in Table XI of Ref. [44], a pleasingly low RMSD = 0.024 kcal/mol for the (T) contribution is obtained. Fitting an extrapolation exponent 2.61 for V{Q,5}Z-F12, we in fact obtain a slightly poorer RMSD = 0.036 kcal/mol.

All this hints at a computational thermochemistry approach in which CCSD(T)-F12b/VTZ-F12 and VQZ-F12 calculations are combined with CCSD-F12b/V5Z-F12 or, better still, CCSD-F12b/V5Z-F12(rev2).

Weak molecular interactions

Results have already been reported for dissociation curves of a seven-system subset of the S22 × 5 data-set [71,72] in a previous paper [58] on the desirability of counterpoise corrections (CP) in explicitly correlated studies of weak molecular interaction. In short, we found that with the V5Z-F12 basis set, we could effectively obtain basis set limit values.

Presently, we shall consider the more recent A24 data-set of Řezáč and Hobza [69]. Two systems containing argon needed to be deleted. RMSDs with and without CP are given in Table 4. The reference geometries from Ref. [69] were downloaded from begdb.com [70] and used without further optimisation.

If we assume the reference values from Ref. [69] to be accurate to about 0.01 kcal/mol, then VQZ-F12 can reach that accuracy level with a CP, while V5Z-F12 can do so even

Table 4. RMSD (kcal/mol) at the CCSD(T)-F12b level for the A24 set of weak interactions using different counterpoise choices.

	VDZ-F12	VTZ-F12	VQZ-F12	V5Z-F12	V5Z-F12(rev2)
With (T*) scaling					
Raw	0.050	0.036	0.018	0.012	0.011
Counterpoise	0.113	0.036	0.007	0.007	
Half–half	0.063	0.016	0.009	0.009	
Without (T) scaling					
Raw	0.061	0.022	0.011	0.009	0.008
Counterpoise	0.159	0.058	0.015	0.004	
Half–half	0.104	0.027	0.007	0.005	

Note: Reference data were taken from [69].

Table 5. Comparison of RMS CP (kcal/mol) at the CCSD(T)/aug'-cc-pVnZ and CCSD(T)-F12b/cc-pVnZ-F12 levels for the A24 set of weak interactions.

CCSD(T)			CCSD(T*)-F12b	CCSD(T)-F12b
		cc-pVDZ-F12	0.122	0.120
aug'-cc-pVTZ	0.254	cc-pVTZ-F12	0.065	0.066
aug'-cc-pVQZ	0.095	cc-pVQZ-F12	0.020	0.023
aug'-cc-pV5Z	0.035	cc-pV5Z-F12	0.006	0.009

without counterpoise. For weak interaction work, at least, V5Z-F12(rev2) does not appear to offer any advantages over V5Z-F12. In fact, the values for the A24 data-set agree to ±0.001–3 kcal/mol RMS between V5Z-F12 and V5Z-F12(rev2).

What about basis set superposition error (BSSE) relative to conventional calculations? RMS values including CP for the A24 set with similar-sized basis sets for orbital and F12 calculations are compared in Table 5. For CCSD(T*)-F12b/cc-pVTZ-F12, the RMS BSSE is about one-quarter that for CCSD(T)/aug-cc-pVTZ; for cc-pV5Z-F12, the corresponding ratio goes down to about one-sixth, favouring the F12 calculation even further. In fact, the value of 0.006 kcal/mol for cc-pV5Z-F12 is small enough that, for most purposes, the CP can be neglected in a CCSD(T*)-F12b/cc-pV5Z-F12 calculation.

As a further illustration, we have calculated the CP for $(H_2O)_2$ and $(HF)_2$ (systems 2 and 4 of the A24 set) at the CCSD(T)/aug-cc-pV6Z level. The respective CPs are still 0.028 and 0.033 kcal/mol, compared to 0.055 and 0.065 kcal/mol at the CCSD(T)/aug-cc-pV5Z level but just 0.010 and 0.011 kcal/mol at the CCSD(T*)-F12b/cc-pV5Z-F12 level.

Conclusions

We have developed and benchmarked a new extended basis set for explicitly correlated calculations involving H, He, and B–Ne, namely, cc-pV5Z-F12. It is offered in two variants, cc-pV5Z-F12 and cc-pV5Z-F12(rev2), the latter of which has additional basis functions on hydrogen not offered in the cc-pVnZ-F12 (n = D,T,Q) sequence. These rev2 sets on H are also developed for n = D–Q. A large uncontracted 'reference' basis set is used for benchmarking cc-pVnZ-F12 (n = D–5) is shown to be a convergent hierarchy. Especially the cc-pV5Z-F12(rev2) basis set can yield the valence CCSD component of TAEs, *without any extrapolation*, to an accuracy normally associated with aug-cc-pV{5,6}Z extrapolations. SCF components are effectively at the basis set limit, while MP2 is within as little as 0.01 kcal/mol. However, the determination of (T) cannot presently be accomplished without extrapolation or scaling.

A composite scheme in which SCF and valence CCSD components are evaluated with the cc-pV5Z-F12(rev2) basis set and (T) extrapolated from CCSD(T)-F12b/cc-pV{T,Q}Z-F12 calculations appears to be very promising for accurate thermochemical calculations.

In addition, the aug-cc-pwCV5Z basis set appears to be large enough that it can be used as a reference standard when CCSD-F12b/REF-*h* calculations are not technically feasible and still a greater accuracy than CCSD-F12b/V5Z-F12(rev2) is required.

For weak interactions, cc-pV5Z-F12 effectively may be considered as within 0.01–0.02 kcal/mol of the basis set limit.

Acknowledgements

JML Martin would like to thank Dr Amir Karton (University of Western Australia, Perth), Dr Sebastian Kozuch (Ben Gurion University of the Negev, Beer-Sheva, Israel), and Dr Martin Suhm (University of Göttingen, Germany) for helpful discussions.

Funding

Research at Weizmann was supported by the Minerva Foundation, Munich, Germany, the Lise Meitner-Minerva Center for Computational Quantum Chemistry, and the Helen and Martin Kimmel Center for Molecular Design.

Supplemental data

Supplemental data for this article can be accessed http://dx.doi.org/10.1080/00268976.2014.985755.

References

[1] W.J. Hehre, L. Radom, P. von, R. Schleyer, and J.A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).
[2] K.G. Dyall and K. Faegri, *Introduction to Relativistic Quantum Chemistry* (Oxford University Press, New York, 2007).
[3] G. Tarczay, A.G. Császár, W. Klopper, and H.M. Quiney, Mol. Phys. **99**, 1769 (2001).
[4] K. Raghavachari, G.W. Trucks, J.A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).
[5] J.D. Watts, J. Gauss, and R.J. Bartlett, J. Chem. Phys. **98**, 8718 (1993).
[6] I. Shavitt and R.J. Bartlett, *Many-Body Methods in Chemistry and Physics* (Cambridge University Press, Cambridge, 2009).
[7] W. Kutzelnigg and J.D. Morgan, J. Chem. Phys. **96**, 4484 (1992).
[8] C. Schwartz, Phys. Rev. **126**, 1015 (1962).
[9] R.N. Hill, J. Chem. Phys. **83**, 1173 (1985).
[10] J.M.L. Martin, Chem. Phys. Lett. **259**, 669 (1996).
[11] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A.K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).
[12] W. Klopper, Mol. Phys. **99**, 481 (2001).
[13] W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

Downloaded by [Weizmann Institute of Science] at 03:36 15 July 2015

[14] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991).
[15] D. Tew and W. Klopper, J. Chem. Phys. **123**, 74101 (2005).
[16] S. Ten-no, Chem. Phys. Lett. **398**, 56 (2004).
[17] W. Klopper and C.C.M. Samson, J. Chem. Phys. **116**, 6397 (2002).
[18] H.-J. Werner, T.B. Adler, and F.R. Manby, J. Chem. Phys. **126**, 164102 (2007).
[19] T.B. Adler, G. Knizia, and H.-J. Werner, J. Chem. Phys. **127**, 221106 (2007).
[20] G. Knizia, T.B. Adler, and H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009).
[21] N.C. Handy, in *Comput. Tech. Quantum Chem. Mol. Phys. [NATO ASI Ser. Vol. 15]*, edited by G.H.F. Diercksen, B.T. Sutcliffe, and A. Veillard (Springer, Dordrecht, 1975), pp. 425–433.
[22] C. Hättig, W. Klopper, A. Köhn, and D.P. Tew, Chem. Rev. **112**, 4 (2012).
[23] L. Kong, F.A. Bischoff, and E.F. Valeev, Chem. Rev. **112**, 75 (2012).
[24] S. Ten-no and J. Noga, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 114 (2012).
[25] S. Ten-no, Theor. Chem. Acc. **131**, 1070 (2012).
[26] W. Klopper, F.R. Manby, S. Ten-No, and E.F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).
[27] D. Feller and K.A. Peterson, J. Chem. Phys. **139**, 084110 (2013).
[28] D. Feller, K.A. Peterson, and J.G. Hill, J. Chem. Phys. **133**, 184102 (2010).
[29] A. Karton and J.M.L. Martin, J. Chem. Phys. **136**, 124114 (2012).
[30] D.H. Bross, J.G. Hill, H.-J. Werner, and K.A. Peterson, J. Chem. Phys. **139**, 094302 (2013).
[31] A. Mahler and A.K. Wilson, J. Chem. Theory Comput. **9**, 1402 (2013).
[32] B. Chan and L. Radom, J. Chem. Theory Comput. **8**, 4259 (2012).
[33] A. Karton, E. Rabinovich, J.M.L. Martin, and B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).
[34] A. Karton, P.R. Taylor, and J.M.L. Martin, J. Chem. Phys. **127**, 064104 (2007).
[35] 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).
[36] M.E. Harding, J. Vázquez, B. Ruscic, A.K. Wilson, J. Gauss, and J.F. Stanton, J. Chem. Phys. **128**, 114111 (2008).
[37] D. Feller, K.A. Peterson, and D.A. Dixon, J. Chem. Phys. **129**, 204105 (2008).
[38] S. Li, J.M. Hennigan, D.A. Dixon, and K.A. Peterson, J. Phys. Chem. A **113**, 7861 (2009).
[39] D. Dixon, D. Feller, and K. Peterson, Annu. Rep. Comput. Chem. **8**, 1 (2012).
[40] D. Feller, K.A. Peterson, and B. Ruscic, Theor. Chem. Acc. **133**, 1407 (2013).
[41] K.A. Peterson, T.B. Adler, and H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).
[42] J.G. Hill, S. Mazumder, and K.A. Peterson, J. Chem. Phys. **132**, 054108 (2010).
[43] J.G. Hill and K.A. Peterson, Phys. Chem. Chem. Phys. **12**, 10460 (2010).
[44] J.G. Hill, K.A. Peterson, G. Knizia, and H.-J. Werner, J. Chem. Phys. **131**, 194105 (2009).
[45] H.-J. Werner, MOLPRO, version 2012.1, a package of ab initio programs. http://www.molpro.net
[46] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, Wiley Interdiscip. Rev. Comput. Mol. Sci. **4**, 91 (2013).
[47] F. Weigend, A. Kohn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).
[48] K.E. Yousaf and K.A. Peterson, J. Chem. Phys. **129**, 184108 (2008).
[49] F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).
[50] F. Weigend, J. Comput. Chem. **29**, 167 (2008).
[51] J. Noga and J. Šimunek, Chem. Phys. **356**, 1 (2009).
[52] C. Hättig, Phys. Chem. Chem. Phys. **7**, 59 (2005).
[53] R.A. Kendall, T.H. Dunning, and R.J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[54] K.E. Yousaf and K.A. Peterson, Chem. Phys. Lett. **476**, 303 (2009).
[55] C. Hättig, D.P. Tew, and A. Köhn, J. Chem. Phys. **132**, 231102 (2010).
[56] O. Marchetti, and H.-J. Werner, Phys. Chem. Chem. Phys. **10**, 3400 (2008).
[57] J.M.L. Martin and M.K. Kesharwani, J. Chem. Theory Comput. **10**, 2085 (2014).
[58] B. Brauer, M.K. Kesharwani, and J.M.L. Martin, J. Chem. Theory Comput. **10**, 3791 (2014).
[59] D. Feller, K.A. Peterson, and J. Grant Hill, J. Chem. Phys. **135**, 044102 (2011).
[60] A.K. Wilson, T. van Mourik, and T.H. Dunning, J. Mol. Struct. THEOCHEM **388**, 339 (1996).
[61] T. Van Mourik, A.K. Wilson, and T.H. Dunning, Mol. Phys. **96**, 529 (1999).
[62] D.E. Woon and T.H. Dunning, J. Chem. Phys. **103**, 4572 (1995).
[63] K.A. Peterson and T.H. Dunning, J. Chem. Phys. **117**, 10548 (2002).
[64] K.A. Peterson, A.K. Wilson, D.E. Woon, and T.H. Dunning Jr., Theor. Chem. Acc. **97**, 251 (1997).
[65] J.E. Del Bene, J. Phys. Chem. **97**, 107 (1993).
[66] E. Papajak and D.G. Truhlar, J. Chem. Theory Comput. **7**, 10 (2011).
[67] J. Řezáč, K.E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 3466 (2011).
[68] A. Karton, S. Daon, and J.M.L. Martin, Chem. Phys. Lett. **510**, 165 (2011).
[69] J. Řezáč and P. Hobza, J. Chem. Theory Comput. **9**, 2151 (2013).
[70] J. Řezáč, P. Jurečka, K.E. Riley, J. Černý, H. Valdes, K. Pluháčková, K. Berka, T. Řezáč, M. Pitoňák, J. Vondrášek, and P. Hobza, Collect. Czechoslov. Chem. Commun. **73**, 1261 (2008).
[71] P. Jurecka, J. Sponer, J. Cerný, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[72] L. Gráfová, M. Pitoňák, J. Řezáč, and P. Hobza, J. Chem. Theory Comput. **6**, 2365 (2010).
[73] J.M.L. Martin, J. Mol. Struct. THEOCHEM **771**, 19 (2006).
[74] D.W. Schwenke, J. Chem. Phys. **122**, 14107 (2005).
[75] D.S. Ranasinghe and G.A. Petersson, J. Chem. Phys. **138**, 144104 (2013).