MP2-F12 Basis Set Convergence near the Complete Basis Set Limit: Are \( h \) Functions Sufficient?

Nisha Mehta and Jan M. L. Martin*

**ABSTRACT:** We have investigated the title question for the W4-08 thermochemical benchmark using \( l \)-saturated truncations of a large reference (REF) basis set, as well as for standard F12-optimized basis sets. With the REF basis set, the root-mean-square (RMS) contribution of \( i \) functions to the MP2-F12 total atomization energies (TAEs) is about 0.01 kcal/mol, the largest individual contributions being 0.04 kcal/mol for \( P_2 \) and \( P_4 \). However, even for these cases, basis set extrapolation from \( \{g,h\} \) basis sets adequately addresses the problem. Using basis sets insufficiently saturated in the \( spdfgh \) angular momenta may lead to exaggerated \( i \) function contributions. For extrapolation from \( spdf \) and \( spdfgh \) basis sets, basis set convergence appears to be quite close to the theoretical asymptotic \( \propto L^{-7} \) behavior. We hence conclude that \( h \) functions are sufficient even for highly demanding F12 applications. With one-parameter extrapolation, \( spdf \) and \( spdfgh \) basis sets are adequate, aug-cc-pV\( [T,Q] \)Z-F12 yielding a RMSD = 0.03 kcal/mol. A limited exploration of CCSD(F12\( ^* \)) and CCSD-F12b suggests our conclusions are applicable to higher-level F12 methods as well.

**INTRODUCTION**

Explicitly correlated quantum chemistry methods (see refs 1–3 for reviews) get their name from the inclusion of basis functions that involve explicit interelectronic distances (so-called “geminal” functions, as distinct from “orbital” functions that only involve a single electronic position).

The \textit{de facto} standard at this point are the F12 geminals introduced by Ten-no:4

\[
F(r_{12}) = \frac{1 - \exp(-r_{12})}{\gamma} \tag{1}
\]

which in the limit for small \( r_{12} \) corresponds to adding \( r_{12} \), and hence to satisfying the Kato cusp condition.5 (For reasons of computational convenience, the Slater function is typically approximated by a linear combination of Gaussians, which actually is reminiscent of the Gaussian geminal approach of Persson and Taylor6 a decade earlier.)

Exigencies for the underlying orbital basis set are quite different from those in a pure orbital calculation, as not so much effort needs to be invested in describing correlation near the interelectronic cusp. Hence basis sets specifically optimized for F12 calculations, such as the cc-pVnZ-F12 (\( n = D,T,Q,S \)) family by Peterson and co-workers,7,8 or their anion-friendly aug-cc-pVnZ-F12 variants,9 perform much better in an F12 setting than similar-sized basis sets for orbital calculations (such as the correlation consistent family of Dunning and co-workers). Indeed, both their contraction patterns and their exponents are markedly different7 from those for the corresponding orbital-optimized basis sets. In fact, it has been shown10 that non-F12 basis sets in thermochemical applications lead to erratic, nonmonotonic basis set convergence due to elevated basis set superposition error.

It is well-known that F12 calculations exhibit fairly rapid basis set convergence in terms of the maximum angular momentum \( L \) in the basis set: for two-electron model systems, Kutzelnigg and Morgan11 showed \( \propto L^{-7} \) for explicitly correlated calculations, compared to \( L^{-3} \) for singlet-coupled pair correlation energies in pure orbital calculations.

Hence, few F12 practitioners go beyond \( L = 4 \) (i.e., \( g \) functions), that is, beyond cc-pVQZ-F12 or aug-cc-pVQZ-F12 basis sets. In accurate thermochemical applications, one may find (e.g., refs 10 and 12) cc-pVSZ-F12 or aug-cc-pwCVSZ\( ^{13–15} \) applied, both of which go up to \( h \) functions.

At least one major electronic structure code often used for F12 calculations, MOLPRO,16 has a hard limit of \( i \) functions overall, and hence (because of the need for one extra \( L \) step in the RI-MP2 auxiliary basis set) in practice the orbital basis set in F12 calculations tops out at \( h \) functions. Two other codes,
Table 1. RMSD (kcal/mol) for the TAEs of the W4-08 Dataset

| truncated REF basis sets | Ordinary MP2 | MP2-F12 | AV6Z(h) | aug-cc-pV6Z(d) |
|--------------------------|--------------|---------|---------|----------------|
| $a$                      | 13.136       | 4.965   | 2.205   | 1.203          | 0.789 |
| $b$                      | 13.137       | 4.981   | 2.221   | 1.219          | 0.791 |

Extrapolations MP2-F12

| Relative to REF-({d,f,g,h,i}) limit (all of W4-08, 96 systems). |
|---------------------------------------------------------------|
| $a$ | 1.373 | 0.027 | 0.006 | REF |
| $b$ | 0.127 | 0.022 | REF   | 0.006 |

F12 optimized correlation consistent

| VDZ-F12 | VTZ-F12 | AVTZ-F12 | VQZ-F12 | AVQZ-F12 | VSZ-F12 |
|---------|---------|----------|---------|----------|---------|
| $a$ | 1.437 | 0.325 | 0.315 | 0.084 | 0.063 | 0.044 |
| $b$ | 1.444 | 0.328 | 0.317 | 0.085 | 0.065 | 0.046 |

core–valence correlation consistent

| ACV6Z(f) | ACV6Z(g) | ACV6Z(h) | ACV6Z | augCVSZ | ditto, AV6Z(h) |
|----------|----------|----------|-------|---------|----------------|
| $a$ | 0.375 | 0.082 | 0.028 | 0.019 | 0.030 | 0.033 |
| $b$ | 0.368 | 0.081 | 0.028 | 0.018 | 0.027 | 0.031 |

Valence aug-cc-pV(n + d)Z

| AVSZ | AV6Z(h) | AV6Z |
|------|---------|------|
| $a$ | 0.055 | 0.041 | 0.020 |
| $b$ | 0.059 | 0.051 | 0.022 |

“Relative to REF-({h,i}) (largest systems omitted, 90 systems retained).” Relative to REF-({g,h}) limit (all of W4-08, 96 systems).

TurboMole\textsuperscript{17} and the most recent versions of ORCA, are capable of going up to at least $i$ functions in an F12 context.

$i$ functions are routinely employed in accurate orbital-only calculations, usually with basis set extrapolation as in the W4\textsuperscript{16–20}, HEAT\textsuperscript{21,22} and PDP\textsuperscript{23–25} thermochemical protocols. One study in our group\textsuperscript{27} on the W4–17 thermochemical benchmark\textsuperscript{26} went up to $k$ functions.

In the present note, we will investigate basis set convergence at the MP2-F12 level for the total atomization energies (TAEs) in the W4-08 subset\textsuperscript{27} of W4-17. We will show that, while $i$ functions may still make some contributions to the atomization energies of some molecules (particularly, those featuring multiple bonds between second-row elements), this can be adequately addressed through $\propto L^2$ basis set extrapolation, and de facto F12 basis set convergence is achieved with $h$ functions.

## Computational Methods

All calculations were carried out using ORCA\textsuperscript{5.28–30} with density fitting for HF and MP2 disabled through the NoCoSX and NoRI keywords. This leaves only the CABS (complementary auxiliary basis set for F12) as a fitting basis set, thus eliminating the Coulomb-exchange and RI-MP2 fitting basis sets as possible “confounding factors”. For technical reasons, UHF references were used for open-shell species.

The largest $spdfgh$ basis set we considered was the reference basis set from Hill et al.\textsuperscript{31} which was also used in refs 8 and 10, for calibrating the VSZ-F12 basis set. The $sp$ part of this is the aug-cc-pV6Z (AV6Z for short) basis set from two (first row) or one (second row) additional primitives were decontracted; the $dfgh$ part is made up of large even-tempered sequences. This basis set we denote REF-h. For its CABS, we used the large uncontracted “reference-ri” basis set associated with it. REF-f and REF-g basis sets were generated by simple truncation of REF-h at $f$ and $g$ functions, respectively. (The linear dependency threshold for CABS was left at its default value of $10^{-8}$ throughout. The MP2-F12 ansatz used in ORCA corresponds to version D, which is a slight simplification\textsuperscript{32} of ansatz C\textsuperscript{33,34} together with fixed geminal amplitudes\textsuperscript{35,36} determined from the Kato cusp condition. This is basically equivalent to the default of “3C(Fix)” in MOLPRO.\textsuperscript{19})

In addition, for a large subset of molecules, we considered an even larger REF-i basis set to which four $i$ functions have been added. Here, for the CABS basis set, we added a $k$ function with the same exponent to every $i$ function.

Aside from VTZ-F12, VQZ-F12, and VSZ-F12 basis sets, we considered aug-cc-pV$(n+d)Z$ ($n = T, Q, S; AVnZ+d$ for short)\textsuperscript{35–38} as well as the core–valence correlation basis sets aug-cc-pwCVnZ ($n = T, Q, S$; awCVnZ for short)\textsuperscript{15}, and aug-cc-pCVnZ ($n = Q, S, 6; ACVnZ$)\textsuperscript{15}. (We are only correlating valence electrons here, but it has been shown\textsuperscript{8,10,12} that the additional radial flexibility of core–valence basis sets is beneficial for F12 calculations as well.) For VnZ-F12 ($n = D, T, Q, S$; VnZ-F12 for short) and AVnZ ($n = T, Q, S$), standard CABS “OptRI” basis sets\textsuperscript{39,40} are available. For AV6Z+d we carried out two sets of calculations; in the first, we repurposed Häggå’s unpublished DF-AV6Z basis set from the TurboMole library (downloaded from the Basis Set Exchange\textsuperscript{41}) as the CABS; in the second, we employed the reference-RI from ref 31 instead. For the AV6Znol basis set, which is a simple truncation of ACV6Z at $h$ functions, we employed reference-RI as the CABS.

As for the geminal exponent, for VTZ-F12 and VQZ-F12 we set $\gamma = 1.0$, and for VSZ-F12 $\gamma = 1.2$, as recommended in the literature for these respective basis sets.\textsuperscript{7,8} For the REF-n and ACVnZ, a fixed $\gamma = 1.4$ was used throughout as per common
practice for large orbital basis sets. (An elevated gamma restricts the geminal to the closer-in part of the cusp.)

Throughout this paper, we will refer to two-point basis set extrapolation using the braces notation, for example, V{T,Q}Z-F12 refers to extrapolation from VTZ-F12 and VQZ-F12 basis sets. See ref 42 for a discussion on how all two-point extrapolation schemes (e.g., refs 43–46) are interrelated.

## RESULTS AND DISCUSSION

Convergence for L-Saturated Basis Sets through L = 6, REF-i. Performance statistics with truncations of the large REF-i basis set at different angular momenta can be found in Table 1. We were able to obtain TAEs for all 96 species in W4-08 through REF-h, plus REF-i for all species except six: B,F6, C,H10, C,H12N2, C,H2NH2, C,H2NH, and NCCN (dicyano-

As mentioned in the introduction, according to Kutzelnigg and Morgan,17 the L-saturated basis set convergence behavior in an R12 calculation, for a closed-shell pair energy, should asymptotically be $\alpha L^{-7}$. Is this indeed the case here? One simple test would be to consider the RMS difference between MP2-F12 atomization energies obtained by $A + B$, L-F12 extrapolation from the REF-{g,h} basis set pair with those obtained in the same manner from the REF-{h,i} pair. As can be seen in Table 1, the RMS difference between these two columns is just 0.006 kcal/mol. We obtain the same value to three decimal places if we minimize RMSD with respect to the REF-{g,h} extrapolation exponent, for which we find 6.514 as the optimum value. Both observations indicate that for these large L-saturated basis sets we are essentially in the $\alpha L^{-7}$ regime (note that even for REF-{g} we already obtain 6.232 as the optimum extrapolation exponent). Incidentally, they also suggest that the MP2-F12/REF-{h,i} values are as close as we can reasonably hope to get to a CBS limit reference. The largest individual differences are for cyclic S,Fo 0.020 kcal/mol, and Si,Hfo $-0.020$ kcal/mol, followed by 0.017 kcal/mol for Al,Fo, 0.0152 kcal/mol for N,Fo, and 0.0012 kcal/mol for C,Fo, Si,Fo, and N,Fo. Convergence for some representative molecules is summarized in Table 2.

Using MP2-F12/REF-{l−1,l} extrapolation, and minimizing RMSD with respect to the extrapolation exponents, we obtain REF-{d,f} 0.13, REF-{f,g} 0.027, and as already mentioned REF-{g,h} 0.006 kcal/mol; with the respective exponents 4.543, 6.232, and 6.514.

Let us now consider the TAE increments $\Delta h = TAE[MP2-F12 / REF-h] - TAE[MP2-F12 / REF-g]$, and $\Delta i = TAE[MP2-F12 / REF-i] - TAE[MP2-F12 / REF-h]$. The RMS $\Delta a = 0.046$ kcal/mol, but individual values can reach as high as 0.23 kcal/mol for P,Fo, 0.18 kcal/mol for P, and 0.13 kcal/mol for S,Fo, which is much smaller, 0.011 kcal/mol RMS, but for P,Fo and P, it reaches 0.042 and 0.039 kcal/mol, respectively. The latter are definitely amounts of interest for high-accuracy electronic structure calculations (as the MP2-F12 basis set incomplete-

| basis | MP2 AV{5,6}Z + d | MP2-F12 truncated REF basis sets | differences | extrapolations |
|-------|------------------|---------------------------------|-------------|---------------|
|       | REF-d | REF-F | REF-g | REF-h | REF-i | $\Delta h$ | $\Delta i$ | $\Delta (h) - (g)$ | $\Delta (f)$ | $\Delta (g)$ | $\Delta (h)$ |
| BF,Fo | 496.38 | 496.45 | 496.56 | 496.53 | 496.52 | 496.52 | -0.01 | 0.00 | 0.00 | 496.58 | 496.53 | 496.52 | 496.52 |
| Al,Fo | 330.07 | 326.78 | 329.82 | 330.35 | 330.40 | 330.40 | 0.13 | 0.01 | -0.01 | 330.39 | 330.46 | 330.41 | 330.39 |
| Si,Fo | 260.73 | 256.22 | 259.64 | 260.53 | 260.66 | 260.67 | -0.13 | -0.04 | -0.00 | 256.56 | 256.39 | 256.32 | 256.30 |
| Si,Fo | 526.53 | 528.99 | 526.94 | 526.49 | 526.36 | 526.31 | 0.23 | 0.04 | 0.01 | 526.56 | 526.39 | 526.32 | 526.30 |
| P,Fo | 308.18 | 296.18 | 306.05 | 297.91 | 308.14 | 308.18 | 0.18 | 0.04 | 0.00 | 308.20 | 308.23 | 308.21 | 308.20 |
| Cl,Fo | 150.24 | 116.67 | 119.42 | 120.26 | 120.44 | 120.48 | 0.18 | 0.04 | 0.00 | 119.94 | 120.43 | 120.50 | 120.50 |
| SO,Fo | 382.96 | 383.30 | 382.82 | 382.61 | 382.53 | 382.51 | 0.01 | 0.00 | 0.00 | 382.73 | 382.57 | 382.51 | 382.50 |
| C,Fo | 414.26 | 413.81 | 414.16 | 414.25 | 414.26 | 414.26 | 0.01 | 0.00 | 0.00 | 414.22 | 414.26 | 414.26 | 414.26 |
| C,Fo | 566.53 | 567.32 | 566.64 | 566.51 | 566.48 | 566.47 | 0.01 | 0.00 | 0.00 | 566.52 | 566.48 | 566.47 | 566.47 |
| C,Fo | 712.78 | 713.87 | 712.91 | 712.74 | 712.70 | 712.70 | -0.04 | -0.02 | -0.00 | 712.73 | 712.70 | 712.69 | 712.69 |
| CO,Fo | 415.47 | 415.17 | 415.47 | 415.56 | 415.57 | 415.57 | 0.01 | 0.00 | 0.00 | 415.53 | 415.57 | 415.57 | 415.57 |
| N,Fo | 396.09 | 395.34 | 395.99 | 396.17 | 396.21 | 396.22 | 0.04 | 0.02 | 0.00 | 396.11 | 396.20 | 396.22 | 396.23 |
| Si,Fo | 381.13 | 381.74 | 381.38 | 381.08 | 381.00 | 371.98 | -0.08 | -0.02 | -0.00 | 381.13 | 381.02 | 371.98 | 371.97 |
| P,Fo | 343.51 | 336.24 | 337.53 | 343.46 | 343.40 | 343.38 | -0.07 | -0.02 | -0.00 | 343.47 | 343.40 | 343.38 | 343.37 |
| CH,Fo | 417.87 | 418.47 | 417.93 | 417.84 | 417.82 | 417.81 | -0.02 | 0.00 | 0.00 | 417.83 | 417.82 | 417.81 | 417.81 |
| H,Fo | 373.53 | 373.98 | 373.63 | 373.56 | 373.56 | 373.56 | -0.01 | 0.00 | 0.00 | 373.57 | 373.55 | 373.57 | 373.57 |

$\Delta h = TAE[REF-h] - TAE[REF-g]$. $\Delta i = TAE[REF-i] - TAE[REF-h]$.

Table 2. TAE (kcal/mol) as a Function of the Basis Set for Some Representative Molecules

https://doi.org/10.1021/acs.jpca.2c02494

J. Phys. Chem. A 2022, 126, 3964–3971
illustrated by considering the \( d, f, \) and \( g \) populations in a natural population analysis (NPA) \(^{50} \) presented in Table 3 for selected molecules at the HF/AVQZ+d and CCSD(T)/AVQZ+d levels. One sees a high \( d \) population already at the HF level for cases like \( SO_3 \), which is a paradigmatic example of "inner polarization" in which the oxygen lone pairs back-donate into the empty 3\( d \) shell of sulfur (and similarly for other second row elements in high oxidation states; see ref 51 and references therein), which thereby becomes an "honorary valence orbital of the second kind." \(^{52} \) It should be noted that this is primarily an SCF-level effect, and hence the increase in 3\( d \) population upon introducing correlation is quite modest in comparison. For cases like \( P_2 \) and \( P_4 \), however, this effect is clearly not operative—one does see a significant \( d \) population even at the HF level, but it is much enhanced by correlation, especially when compared to the isovalent first-row species \( N_2 \) and \( N_4 \). Moreover, this is not just limited to the \( d \) orbitals: natural orbital occupations for \( f \) and \( g \) shells clearly decay more slowly for the second-row species than for their first-row cognates. Similar observations can be made for \( Cl_2 \) vs \( F_2 \), \( S_3 \) vs \( O_3 \), and the like.

**cc-pVnZ-F12 and Related Basis Sets.** How does the cc-pVnZ-F12 series perform compared to the REF-{h,i} basis set limit estimate? RMSDs are 1.44, 0.33, 0.084, and 0.044 for \( n = D,T,Q,5, \) respectively. The latter is close to the "no extrapolation required" goal, but there is still some room for improvement in high-accuracy thermochemistry applications, where one would like a 3\( \sigma \) = 0.1 kcal/mol error bar. AVTZ-F12 performs only marginally better (0.31 kcal/mol) than the underlying VTZ-F12, while AVQZ-F12, at 0.063 kcal/mol, does represent a modest gain over VQZ-F12. It was previously shown \(^{51,2} \) that the aug-cc-pwCV5Z core–valence basis set (awCV5Z for short) performs remarkably well for F12 calculations, despite not being developed for this purpose at all: the RMD = 0.030 kcal/mol we find here is consistent with that observation. (Expanding the hydrogen basis set from AVSZ to AV6Z yields the same performance to within 0.03 kcal/mol, while AVQZ performs only marginally better (0.063 kcal/mol) than the underlying VTZ, while AVQZ-F12, at 0.063 kcal/mol, does represent a modest gain over VQZ-F12. It was previously shown that the aug-cc-pwCV5Z core–valence basis set (awCV5Z for short) performs remarkably well for F12 calculations.)

Table 3. NBO Angular Momentum Populations at the HF/cc-pV(Q+d)Z and CCSD(T)/cc-pV(Q+d)Z Levels for Selected Molecules

|      | \( d \) | \( f \) | \( g \) | \( d \) | \( f \) | \( g \) |
|------|--------|--------|--------|--------|--------|--------|
| \( N_2 \) | N      | 0.02493 | 0.00186 | 0.00047 | 0.04513 | 0.00509 | 0.00110 |
| P_2  | P      | 0.05648 | 0.00407 | 0.00127 | 0.11985 | 0.01390 | 0.00295 |
| N_4  | N      | 0.02957 | 0.00297 | 0.00177 | 0.05414 | 0.00719 | 0.00322 |
| P_4  | P      | 0.07726 | 0.00759 | 0.00321 | 0.14830 | 0.01982 | 0.00591 |
| F_2  | F      | 0.00417 | 0.00306 | 0.00008 | 0.03159 | 0.00442 | 0.00079 |
| Cl_2 | Cl     | 0.02517 | 0.00232 | 0.00029 | 0.13406 | 0.01664 | 0.00315 |
| CIF  | Cl     | 0.02472 | 0.00114 | 0.00009 | 0.12730 | 0.01451 | 0.00250 |
| BF_3 | B      | 0.01056 | 0.00148 | 0.00000 | 0.02739 | 0.00487 | 0.00027 |
| AlF_3| Al     | 0.00153 | 0.00098 | 0.00000 | 0.04272 | 0.00457 | 0.00090 |
| ClF  | Cl     | 0.00046 | 0.00050 | 0.00000 | 0.04518 | 0.00445 | 0.00092 |
| Cl   | Cl     | 0.00282 | 0.00199 | 0.00000 | 0.11392 | 0.01235 | 0.00359 |
| O_3  | O_{center} | 0.00952 | 0.00077 | 0.00000 | 0.03595 | 0.00464 | 0.00092 |
| SO_2 | S      | 0.13220 | 0.00186 | 0.00000 | 0.12065 | 0.01544 | 0.00274 |
| CF_2 | C      | 0.01101 | 0.00076 | 0.00000 | 0.03359 | 0.00379 | 0.00106 |
| CCl_2| C      | 0.02264 | 0.00187 | 0.00000 | 0.06138 | 0.01049 | 0.00395 |
| SO_3 | S      | 0.07474 | 0.00573 | 0.00000 | 0.06700 | 0.00794 | 0.00200 |
| SF_4 | S_{reduc} | 0.00890 | 0.00034 | 0.00000 | 0.02492 | 0.00480 | 0.00080 |
| HF   | F      | 0.00300 | 0.00040 | 0.00000 | 0.03680 | 0.00404 | 0.00078 |
| HCl  | Cl     | 0.00163 | 0.00089 | 0.00000 | 0.12527 | 0.01291 | 0.00252 |

"Populations refer to unique atoms (e.g., to one of the four equivalent N atoms in tetrahedral \( N_4 \)). CCSD(T)/cc-pVQZ geometry for \( N_4 \) \( r_{NN} = 1.4551 \) Å, was taken from ref 49."
reach 0.1 kcal/mol, such as P\textsubscript{6}, S\textsubscript{6}, and AlCl\textsubscript{3}, plus another 0.07 kcal/mol for S\textsubscript{6}. However, we have already established with the REF-h and REF-i basis set that the true \iota\textsubscript{contribution is much smaller; it is a well-known basis set convergence phenomenon (see, e.g., the older review of basis sets by Davidson and Feller\textsuperscript{53}) that insufficiently saturating the basis in lower angular momenta can exaggerate the impact of the highest angular momentum (through basis set superposition error). If we instead compare the core–valence versions of these basis sets, the \iota\textsubscript{i}-increments are substantially reduced, with P\textsubscript{4} and P\textsubscript{3} now remaining as the principal outliers. Finally, if one wants a single "prepackaged" basis set without extrapolation, ACV6Z has the smallest RMSD (0.019 kcal/mol) of them all.

**A Comment about Basis Set Extrapolation.** We have considered three sets of extrapolation exponents:

1. From the literature, as optimized for the very small training set of Hill et al.\textsuperscript{31} (that paper itself for V[D,T]Z-F12 and V[T,Q]Z-F12), ref \textsuperscript{54} for V[Q,S]-Z-F12, and ref \textsuperscript{55} for AV[T,Q]Z-F12.
2. By minimization of RMSD from REF-{\hbar,\iota} total energies for such W4-08 species for which we have REF-\iota energies.
3. By minimization of RMSD from REF-{\hbar,\iota} TAEs for the same species.

The three sets of extrapolation exponents are compared in Table 4 below, as are the RMSDs for total atomization energies.

**Table 4. Two-Point Extrapolation Exponents and RMSD from REF-{\hbar,\iota} (kcal/mol) for Different Basis Set Pairs**

| Basis Set Pairs | Extrapolation Exponents | RMSD with These Exponents (kcal/mol) | Regular MP2 for comparison |
|-----------------|--------------------------|--------------------------------------|-----------------------------|
|                 | V[D,T]Z-F12              | V[T,Q]Z-F12                          | AV[T,Q]Z-F12                |
|                 | V[Q,S]Z-F12              |                                      | AV[Q,5]Z-F12               |
|                 |                           |                                      | AV[5,6]Z-d                  |
| Extrapolation Exponents |                   |                                      |                            |
| lit.\textsuperscript{31,54,55} | 3.0878                  | 4.3548                               | 4.6324                     |
| W4-08 E\textsubscript{total} | 3.45                     | 5.03                                 | 5.64                        |
| W4-08 TAE       | 3.80                     | 5.11                                 | 5.97                        |
|                 |                          |                                      | 4.36                        |
| RMSD with These Exponents (kcal/mol) |                   |                                      |                            |
| lit.            | 0.178                    | 0.046                                | 0.046                       |
| W4-08 E\textsubscript{total} | 0.123                    | 0.039                                | 0.030                       |
| W4-08 TAE       | 0.107                    | 0.038                                | 0.029                       |
|                 |                          |                                      | 0.033                       |

*Extrapolation exponents from Table 8 of ref \textsuperscript{31}.

with them. The TAE-optimized exponent V[Q,S]Z-F12 appears anomalously small, but this is an artifact of the very flat surface there, which precludes a "clean" optimization. E\textsubscript{total} gives a somewhat better defined minimum, though even there the surface is quite flat. In terms of RMSD, for V[Q,S]Z-F12 all three possible exponents yield the same error. For AV[T,Q]Z-F12, both optimized values from the present work yield superior RMSDs to the value from ref \textsuperscript{55}; for VTZ-F12 the same holds, although the difference here is quite modest. Finally, for the V[D,T]Z-F12 pair, the two optimized values from the present work are clearly superior to the one from Hill et al.\textsuperscript{31} For reasons of numerical stability (notably because it eliminates the anomaly for V[Q,S]Z-F12), we favor the set optimized from total energies. The difference in RMSD is negligible; however, this is good news, since ideally one wants to be in a scenario where basis set extrapolation is only a minor component of the final result and is relatively insensitive to details of the extrapolation procedure. Note that, while all basis set pairs considered here are clearly some distance away from the asymptotic L\textsuperscript{-7} regime, the exponents do increase in that direction as L gets larger.

For some perspective, we add the RMSDs for regular MP2 with aug-cc-pV(L-1+d)Z and aug-cc-pV(L+d)Z basis sets. It is sobering to see that even AV[5,6]Z-d only reaches the accuracy level of V[D,T]Z-F12, and that V[T,Q]Z-F12 already markedly exceeds it.

**What Do These Results Imply for CCSD-F12 and CCSD(T)-F12?** At the request of a reviewer, we will address what we may infer for the basis set convergence of higher-level methods like CCSD(T)-(F12\*) or CCSD(T)-F12b.

First of all, the treatment of parenthetical triples (T) does not benefit in any way from the F12 treatment,\textsuperscript{56} so the convergence of that contribution is effectively the same as in an orbital calculation. The latter has been addressed in great detail in a recent book chapter by one of us:\textsuperscript{57} suffice to say here that basis set convergence of (T) with angular momentum is actually fairly rapid, and that radial flexibility of the basis set is actually more important than angular flexibility.

This leaves us then with CCSD-F12, or rather with the various practical approximations to it such as CCSD-F12b,\textsuperscript{58} CCSD\textsubscript{TT},\textsuperscript{59,60} and CCSD(F12\*)\textsuperscript{61}. Of these, CCSD(F12\*) has been shown\textsuperscript{32} to be the most rigorous approximation that still is computationally feasible, while CCSD-F12b is widely used owing to its implementation for both closed-shell and open-shell cases in the MOLPRO\textsuperscript{16} program system.

The basis set convergence of differences between different CCSD-F12 approximations has been studied in great detail in ref \textsuperscript{12}. In a nutshell: if correlation is predominantly dynamic then these differences decay quickly with increasing basis sets, but even moderate amounts of static correlation can cause nontrivial differences (as large as 0.3 kcal/mol) to persist even for spin\textsubscript{dgh} basis sets.

As additional complications, all of the available (to us) implementations of approximate CCSD-F12 methods require DFMP2 and JKFit auxiliary basis sets (further complicating comparisons), and the only CCSD(F12\*) implementation at our disposal with which we were able to get enough calculations in REF-h basis sets to converge, i.e., that in MOLPRO, is limited to closed shell. Hence we resorted to calculating a number of closed-shell reaction energies instead. The relevant basis set increments are reported in Table 5. Auxiliary basis sets were taken from the Supporting Information of ref \textsuperscript{31}.

The basis set increments given there for g and h layers are obtained directly. For the i layer, the ORCA MP2-F12 values are calculated directly while the MOLPRO values are obtained by L\textsuperscript{-7} extrapolation. (It can be easily shown that an estimate for the next basis set increment after E(L)−E(L-1) is given by the following formula:)

\[
E(L) - E(L-1) \approx \left[ \frac{1 - \left( \frac{L}{L+1} \right)^\alpha}{\left( \frac{L}{L-1} \right)^\alpha - 1} \right]
\]

(2)
For the sake of the estimate, we considered \( \alpha = 7 \) (the asymptotic convergence rate) as a best-case scenario, and \( \alpha = 5 \) as a worst-case scenario (for VQZ-F12 and VSZ-F12, we have previously found\(^4\) the intermediate value \( \alpha \approx 6 \)). The average of both extrapolated values, plus or minus half the difference between them, is taken as our approximate estimate.

As can be seen in Table 5, basis set convergence of especially CCSD(F12\(^*\)) actually seems, depending on the reaction, comparable to or faster than that of MP2-F12. And for those reactions where \( \Delta i \) is still somewhat significant at the MP2-F12 level, the corresponding estimated CCSD-F12b and especially CCSD(F12\(^*\)) increments are actually up to 3 times smaller.

We hence conclude that our conclusions about the lack of necessity of \( i \) functions in MP2-F12 calculations also apply to CCSD-F12b, CCSD(F12\(^*\)), and other approximations to CCSD-F12.

As an aside, we note that the CCSD(F12\(^*\)) − CCSD-F12b differences, while nontrivial, are conspicuously smaller than what was observed for cc-pVQZ-F12, cc-pVSZ-F12, and aug-cc-pCVSZ in ref \(^\text{12}\). Clearly, here too, the greater radial flexibility of the REF-\( n \) basis sets puts them at an advantage, even as they are unwieldy for practical production calculations.

## CONCLUSIONS

We may conclude that for basis sets that are adequately saturated in the angular momenta represented, the contribution of \( i \) functions to total atomization energies is minimal (about 0.01 kcal/mol RMS): the largest individual contributions we have found here only reach 0.04 kcal/mol (for \( P_4 \) and \( P_2 \)), and even that is adequately captured by \( L^{-7} \) extrapolation from REF-\( \{g,h\} \), which is closer than 0.01 kcal/mol RMS to the CBS limit. (In this \( L \) region, basis set convergence behavior for MP2-F12 is found to be quite close to the asymptotic\(^\text{11}\) \( L^{-7} \).

Comparison of AV6Z with and without \( i \) functions indicates more significant \( i \) contributions, owing to insufficient radial saturation. The \( h \) contribution is more significant, reaching 0.23 kcal/mol for \( P_4 \). REF-\{gh\} extrapolation is within less than 0.01 kcal/mol RMS of REF-\{hi\}, the largest single difference being 0.02 kcal/mol for \( S_4 \). This means that the REF-\{gh\} \( L^{-7} \) extrapolation is an acceptable proxy for the F12 basis set limit, and that \( h \) functions are sufficient even for highly demanding F12 applications.

If one-parameter extrapolation with an adjustable exponent is permissible, \( spdf \) and \( spdfg \) basis sets are adequate, with aug-cc-pV(T,Q)Z-F12 yielding the RMSD = 0.03 kcal/mol. Finally, the somewhat slower basis set convergence in the second row—especially when there are multiple adjacent such elements—can be rationalized to some degree in terms of lower-lying 3d and 4f orbitals.

A limited investigation for closed-shell reactions indicates that basis set convergence of CCSD(F12\(^*\)) and CCSD-F12b is comparable to or faster than that of MP2-F12, and that hence our conclusions about the relative insignificance of \( i \)-function contributions apply to F12 methods more broadly.
Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Work on this paper was supported in part by the Israel Science Foundation (Grant 1969/20) and by the Minerva Foundation (Grant 2020/05). We thank Prof. Amir Karton for helpful discussions. This paper is dedicated to Prof. Peter R. Taylor in honor of his 70th birthday, and of his landmark contributions (e.g., refs 6 and 63−65) to the theory and practice of explicitly correlated quantum chemistry.

REFERENCES

(1) Kong, L.; Bischoff, F. A.; Valeev, E. F. Explicitly correlated R12/F12 methods for electronic structure. Chem. Rev. 2012, 112, 75−107.
(2) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly correlated electrons in molecules. Chem. Rev. 2012, 112, 4−74.
(3) Ten-no, S.; Noga, J. Explicitly correlated electronic structure theory from R12/F12 ansätze. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2012, 2, 114−125.
(4) Ten-no, S. Initiation of explicitly correlated Slater-type geminal theory. Chem. Phys. Lett. 2004, 398, 56−61.
(5) Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. Commun. Pure Appl. Math. 1957, 10, 151−177.
(6) Persson, B. J.; Taylor, P. R. Accurate quantum-chemical calculations: The use of Gaussian-type geminal functions in the treatment of electron correlation. J. Chem. Phys. 1996, 105, 5915−5926.
(7) Petersen, K. A.; Adler, T. B.; Werner, H.-J. Systematically convergent basis sets for explicitly correlated wavefunctions: the atoms H, He, B-Ne, and Al-Ar. J. Chem. Phys. 2008, 128, 084102.
(8) Petersen, K. A.; Kesharwani, M. K.; Martin, J. M. L. The aug-cc-pV5Z-F12 basis set: reaching the basis set limit in explicitly correlated calculations. Mol. Phys. 2015, 113, 1551−1558.
(9) Sylvestry, N.; Kesharwani, M. K.; Martin, J. M. L. The aug-cc-pVnZ-F12 basis set family: Correlation consistent basis sets for explicitly correlated benchmark calculations on anions and non-covalent complexes. J. Chem. Phys. 2017, 147, 134106.
(10) Sylvestry, N.; Peterson, K. A.; Karton, A.; Martin, J. M. L. Toward a W4-F12 approach: Can explicitly correlated and orbital-based ab initio CCSD(T) limits be reconciled? J. Chem. Phys. 2016, 144, 214101.
(11) Kutzelnigg, W.; Morgan, J. D. Rates of convergence of the partial-wave expansions of atomic correlation energies. J. Chem. Phys. 1992, 96, 4484−4508.
(12) Kesharwani, M. K.; Sylvestry, N.; Köhn, A.; Tew, D. P.; Martin, J. M. L. Do CCSD and approximate CCSD-F12 variants converge to the same basis set limits? The case of atomization energies. J. Chem. Phys. 2018, 149, 154109.
(13) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. J. Chem. Phys. 1992, 96, 6796−6806.
(14) Woon, D. E.; Dunning, T. H. Gaussian-Basis Sets for Use in Correlated Molecular Calculations 0.3. The Atoms Aluminum through Argon. J. Chem. Phys. 1993, 98, 1358−1371.
(15) Peterson, K. A.; Dunning, T. H. Accurate correlation consistent basis sets for molecular core−valence correlation effects: The second row atoms Al−Ar, and the first row atoms B−Ne revisited. J. Chem. Phys. 2002, 117, 10548−10560.
(16) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Black, J. A.; Doll, K.; Heßelmann, A.; Kats, D.; Köhn, A.; Korona, T.; Kreplin, D. A.; et al. The Molpro quantum chemistry package. J. Chem. Phys. 2020, 152, 144107.
(17) Balasubramanis, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y. J.; Furchc, F.; Grotjahn, R.; Harding, M. E.; Hättig, C.; et al. TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. J. Chem. Phys. 2020, 152, 184107.
(18) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kl/mol predictions. J. Chem. Phys. 2006, 125, 144108.
(19) Karton, A.; Taylor, P. R.; Martin, J. M. L. Basis set convergence of post-CCSD contributions to molecular atomization energies. J. Chem. Phys. 2007, 127, 064104.
(20) Karton, A.; Daon, S.; Martin, J. M. L. W4−11: a high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. Chem. Phys. Lett. 2011, 510, 165−178.
(21) Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. J. Chem. Phys. 2008, 128, 114111.
(22) Thorpe, J. H.; Lopez, C. A.; Nguyen, T. L.; Baraban, J. H.; Bross, D. H.; Ruscic, B.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. IV. A modified recipe for computational efficiency. J. Chem. Phys. 2019, 150, 224102.
(23) Feller, D.; Peterson, K. A.; Ruscic, B. Improved accuracy benchmarks of small molecules using correlation consistent basis sets. Theor. Chem. Acc. 2014, 133, 1407.
(24) Dixon, D.; Feller, D.; Peterson, K. A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table. Ann. Rep. Comput. Chem. 2012, 8, 1−28.
(25) Feller, D.; Peterson, K.; Dixon, D. The Impact of Larger Basis Sets and Explicitly Correlated Coupled Cluster Theory on the Feller−Peterson−Dixon Composite Method. Ann. Rep. Comput. Chem. 2016, 12, 47−78.
(26) Karton, A.; Sylvestry, N.; Martin, J. M. L. W4−17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. J. Comput. Chem. 2017, 38, 2063−2075.
(27) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics. J. Phys. Chem. A 2008, 112, 12868−12886.
(28) Neese, F. The ORCA program system. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2012, 2, 73−78.
(29) Neese, F. Software update: the ORCA program system, version 4.0. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2018, 8, 4−9.
