

Do CCSD and approximate CCSD-F12 variants converge to the same basis set limits? The case of atomization energies

Manoj K. Kesharwani, Nitai Sylvetsky, Andreas Köhn, David P. Tew, and Jan M. L. Martin

Citation: *The Journal of Chemical Physics* **149**, 154109 (2018); doi: 10.1063/1.5048665

View online: <https://doi.org/10.1063/1.5048665>

View Table of Contents: <http://aip.scitation.org/toc/jcp/149/15>

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

PHYSICS TODAY

WHITEPAPERS

ADVANCED LIGHT CURE ADHESIVES

Take a closer look at what these environmentally friendly adhesive systems can do

READ NOW

PRESENTED BY
MASTERBOND
ADHESIVES | SEALANTS | COATINGS

Do CCSD and approximate CCSD-F12 variants converge to the same basis set limits? The case of atomization energies

Manoj K. Kesharwani,^{1,2} Nitai Sylvetsky,¹ Andreas Köhn,² David P. Tew,³ and Jan M. L. Martin^{1,a)}

¹Department of Organic Chemistry, Weizmann Institute of Science, 76100 Rehovot, Israel

²Institute for Theoretical Chemistry, University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany

³Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany

(Received 17 July 2018; accepted 20 September 2018; published online 18 October 2018)

While the title question is a clear “yes” from purely theoretical arguments, the case is less clear for practical calculations with finite (one-particle) basis sets. To shed further light on this issue, the convergence to the basis set limit of CCSD (coupled cluster theory with all single and double excitations) and of different approximate implementations of CCSD-F12 (explicitly correlated CCSD) has been investigated in detail for the W4-17 thermochemical benchmark. Near the CBS ([1-particle] complete basis set) limit, CCSD and CCSD(F12*) agree to within their respective uncertainties (about ± 0.04 kcal/mol) due to residual basis set incompleteness error, but a nontrivial difference remains between CCSD-F12b and CCSD(F12*), which is roughly proportional to the degree of static correlation. The observed basis set convergence behavior results from the superposition of a rapidly converging, attractive, CCSD[F12]–CCSD-F12b difference (consisting mostly of third-order terms) and a more slowly converging, repulsive, fourth-order difference between CCSD(F12*) and CCSD[F12]. For accurate thermochemistry, we recommend CCSD(F12*) over CCSD-F12b if at all possible. There are some indications that the nZaPa family of basis sets exhibits somewhat smoother convergence than the correlation consistent family. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5048665>

INTRODUCTION

Explicitly correlated (R12 and F12) electron correlation methods (see, e.g., Refs. 1–4 for reviews) greatly accelerate basis set convergence of wavefunction *ab initio* methods. Experience from thermochemistry (e.g., Refs. 5–7), noncovalent interactions (e.g., Refs. 8–12), and vibrational frequencies^{13,14} shows that explicitly correlated methods gain two,¹⁵ or for larger basis sets even three, angular momentum increments on their conventional correlated counterparts. Because the cost of coupled cluster methods scales steeply with the number of basis functions—which scales approximately as $\propto L^3$ with angular momentum L —this can mean a cost difference of an order of magnitude.

In current practice, Slater-type geminals,¹⁶ better known as F12 geminals, have become the *de facto* standard for explicitly correlated methods. (For a recent comparison of alternative geminal forms, see Ref. 17.)

CCSD(T), coupled cluster theory¹⁸ with all singles and doubles¹⁹ plus a quasiperturbative correction for triple excitations,^{20,21} is widely considered the “gold standard” of *ab initio* quantum chemistry. The triples component does not benefit from F12 functions (despite attempts^{22–24} to compensate by scaling; for attempts to incorporate F12 into triples, see Refs. 25 and 26), and hence we found^{7,12} that it is preferable

to obtain the triples contribution from conventional CCSD(T) calculations. At any rate, the (T) component is typically an order of magnitude smaller than the CCSD valence correlation, and higher order correlation effects, another order of magnitude smaller still (see the supplementary material of Refs. 27 and 28 for tabulation for 140 and 200 molecules, respectively). Furthermore, (T) converges faster with the basis set^{29–32} than CCSD, which leaves basis set convergence in the latter as the one “accuracy bottleneck,” where F12 has the most to offer.

Recently, however, claims have been proffered that CCSD and common explicitly correlated methods like CCSD-F12b do not converge to the same basis set limit. Such claims were made by Cremer and co-workers³³ for the formic acid dimer, and by Feller³⁴ for total atomization energies of small molecules. While the formic acid discrepancy was ultimately ascribed to other sources,³⁵ it bears emphasizing that CCSD-F12b^{36,37} and its more rigorous companion CCSD(F12*),³⁸ also known as CCSD-F12c in MOLPRO-speak, are *approximations* to the full CCSD-F12 method. From a purely formal viewpoint, the presence of the strong orthogonality projector guarantees that all F12 terms should eventually vanish as the one-electron basis becomes complete, and hence CCSD and all variants of CCSD-F12 should converge to the same complete basis set (CBS) limit. This does not guarantee, however, that this would be the case for finite basis sets small enough to be practically usable.

During our work on the W4-17 benchmark,²⁸ where the reference values were obtained by basis set extrapolation from

^{a)}Author to whom correspondence should be addressed: gershom@weizmann.ac.il

conventional calculations, we also carried out explicitly correlated calculations using large basis sets for certain molecules. For many systems, we found excellent agreement between approaches, but for some others—notably those with significant nondynamical correlation³⁹—we found discrepancies between CCSD and CCSD-F12b that appeared to persist even with large basis sets. Moreover, for those systems we found discrepancies *between different approximate F12 methods* that persisted even with large basis sets.

We then embarked on a comprehensive investigation.

COMPUTATIONAL METHODS

Choice of benchmark

We chose the W4-17 benchmark²⁸ of 200 small first- and second-row molecules, which is an expanded update of the earlier W4-11 benchmark,²⁷ which in turn represents an expansion of the W4-08 benchmark⁴⁰ used for parametrizing the B2GP-PLYP double hybrid density functional.⁴⁰ The W4-17 dataset is chemically diverse in that it spans a broad variety of bonding situations (covalent, ionic, strained, “hypervalent,” etc.) as well as a broad range of electronic correlation from predominantly dynamical to pathological static correlation. The reference geometries were taken from the supplementary material of Ref. 28 and used without further modification. The W4-08 subset (96 molecules) will also be referred to in a few places in this paper.

To avoid clouding the issue, we focus exclusively on valence correlation in this paper. The basis set convergence of subvalence correlation for the W4-17 benchmark has very recently been studied in great detail;⁴¹ it was found there that even conventional CCSD(T) extrapolated from aug-cc-pwCV{T,Q}Z basis sets⁴² can reproduce this contribution to within 0.03 kcal/mol root mean square (RMS) and CCSD(T)/aug-cc-pwCV{Q,5}Z to about 0.01 kcal/mol RMS. The use of core-valence F12 basis sets⁴³ for the CCSD component was also discussed in Ref. 41 (The comparatively fast convergence of this term owes much to the near-perfect cancellation of the core-core correlation energy, and the still substantial cancellation of the core-valence correlation energy, between the molecule and its proatoms.) Hence, this is not the accuracy-limiting factor in thermochemical studies: as pointed out repeatedly (e.g., Refs. 7 and 44), that honor belongs to the valence correlation energy.

Choice of basis sets

For the explicitly correlated calculations, we considered three basis sets: the cc-pVTZ-F12 and cc-pVQZ-F12 basis set of Peterson and co-workers,⁴⁵ and the aug-cc-pwCV5Z basis set.⁴² In a previous study,²⁴ we were able to show that this latter basis set yields results very close to reference data obtained with very large spd²gh uncontracted basis sets⁴⁶ (REF-h, in the notation of Ref. 24). In addition, we considered the cc-pV5Z-F12 basis set^{7,24} for a subset of molecules; as our available central processing unit (CPU) time resources forced us to choose between either a complete set of aug-cc-pwCV5Z data

(and partial cc-pV5Z-F12 results) or a complete cc-pV5Z-F12 data set and partial aug-cc-pwCV5Z, we decided to prioritize aug-cc-pwCV5Z, which is the larger and more complete of the two basis sets.

For the conventional calculations, we considered two families of basis sets. One is the correlation-consistent family aug-cc-pV(n+d)Z, $n = T, Q, 5, 6, 7$, taken from Ref. 47 and references therein. The “+d” indicates that a tight d function is added⁴⁸ to second-row atoms to assist with the description of the 3d orbital, which in high oxidation states of these elements sinks low enough to become a back-donation acceptor.⁴⁹ As is our custom, we do not place diffuse functions on hydrogen, both in order to save CPU time and to enhance numerical stability. These basis sets are denoted by the shorthand haVnZ [heavy atom-augmented (correlation consistent) valence n-tuple zeta] basis set throughout the paper.

The second family of basis sets are the nZaPa (uniformly convergent n-tuple zeta augmented polarized) of Petersson and co-workers,^{32,50,51} which were designed for smooth uniform convergence of both SCF and valence correlation energies. (Unlike the correlation consistent family,⁵² but akin to the Weigend-Ahlrichs basis sets,⁵³ the number of primitives in nZaPa varies from left to right in a row of the Periodic Table to ensure that the basis set incompleteness error stays approximately constant.) After some initial experimentation, we applied the same approximation for hydrogen as in haVnZ, namely, that we are using the un-augmented basis sets on H. This is indicated by the notation nZaPha.

Electronic structure codes

Unless otherwise indicated, all conventional calculations were carried out using Gaussian 09 Rev.E01⁵⁴ and Gaussian 16 Rev.B01,⁵⁵ while all explicitly correlated calculations were carried out using TURBOMOLE 6.6 and 7.2.1.⁵⁶ For open-shell species, the Watts-Gauss-Bartlett version of ROCCSD was used throughout; i.e., orbitals were semicanonicalized before transformation (unlike in MOLPRO,⁵⁷ where semicanonicalization happens after integral transformation—see the Appendix to Ref. 58 for a brief explanation of the differences for open-shell cases). The options `int(nobasistransform, acc2e=13) iop(3/59=7, 8/11=1)` were used in the Gaussian calculations.

For the F12 calculations, the RI-MP2 auxiliary basis sets of Refs. 59 and 60, and the JKFit basis sets of Ref. 61, together with the CABS (complementary auxiliary basis sets) of Refs. 62 and 63 have been employed in the cc-pVnZ-F12 calculations, while the aug-cc-pwCV5Z/MP2FIT basis set was used for both RI-MP2 and CABS in the aug-cc-pwCV5Z calculations.⁶⁰ As in our previous thermochemical papers, the geminal exponent β was set to $1.4 a_0^{-1}$ throughout. (In response to an early preprint of this work, the question arose whether the large differences for VTZ-F12 and VQZ-F12 were an artifact of our choice of geminal exponent, rather than the MP2-F12 optimized “recommended” values of 1.0 for those basis sets. As shown in Figs. S.1 and S.2 in Ref. 84, recalculation does not qualitatively change our conclusion, and in fact the box plot ranges become *larger* at $\beta = 1.0 a_0^{-1}$.)

TABLE I. Schwenke coefficients and equivalent Petersson shifts for different basis set pairs.

	Basis sets	Schwenke coefficients A_L				Equivalent Petersson shifts a			
		{6,7}	{5,6}	{4,5}	{3,4}	{6,7}	{5,6}	{4,5}	{3,4}
L^{-3} pure	Generic	1.701	1.374	1.049	0.730	0.00	0.00	0.00	0.00
$(L+3/2)^{-3}$ pure	Generic	2.194	1.865	1.537	1.211	1.50	1.50	1.50	1.50
Martin ^a	MP2/AVnZ	1.852	1.503	1.127		0.46	0.40	0.24	
Petersson ³² opt.	MP2/nZaPa	1.865	1.519	1.185	0.886	0.50	0.45	0.42	0.49
Hill <i>et al.</i> ⁴⁶	MP2/AVnZ	N/A	1.478	1.186	0.933	N/A	0.32	0.42	0.64
Martin ^b	CCSD/AVnZ	1.602	1.283	0.932		-0.30	-0.28	-0.36	
Martin ^b	CCSD/nZaPa	1.605	1.232	0.917		-0.29	-0.44	-0.41	
Varandas ^{66,67}	CCSD/AVnZ	N/A	1.295	0.912	0.665	N/A	-0.24	-0.43	-0.21
Schwenke ³⁰	CCSD/AVnZ	N/A	1.266	0.930	0.700	N/A	-0.33	-0.37	-0.09

^aReference 64 fitted to MP2-F12/REF-h⁴⁶ data obtained using MOLPRO 2015.⁵⁷ Auxiliary basis sets from Ref. 46.

^bReference 64 fitted to CCSD-F12 data for 12 closed-shell species in supplementary material of Ref. 13 at ref. geoms. *ibid.* Original aug-cc-pV7Z basis sets taken from Ref. 47 and references therein; updates courtesy of Dr. David Feller (PNL).

For the larger basis sets, coupled cluster convergence problems were encountered in some instances: in most cases, these could be remedied by tightening the SCF convergence criterion and setting $\$scftol=1.0E-16$ (which tightens several associated parameters in the `ccsdf12` module of Turbomole) or even $\$scftol=1.0E-18$.

Basis set extrapolation

For the nZaPa basis sets at the MP2 level, we used the extrapolation recommended by Ranasinghe and Petersson,³² expressed here in the form of Schwenke,³⁰

$$E_\infty = E(L) + A_L[E(L) - E(L-1)]. \quad (1)$$

Here the Schwenke coefficient A_L is specific to the basis set pair and level of electronic structure theory. For the other levels of theory, we use a reparametrization of the same formula (Table I). As shown in Ref. 64, this expression can be related quite simply to the more familiar extrapolation formulas,

$$E_L = E_\infty + \frac{B}{L^\alpha} \quad \text{if } \alpha = \frac{\log\left(1 + \frac{1}{A_L}\right)}{\log\left(\frac{L}{L-1}\right)}, \quad (2)$$

$$E_L = E_\infty + \frac{D}{(L+a)^3} \quad \text{if } a = \frac{1}{\left(1 + \frac{1}{A_L}\right)^{1/3} - 1} + 1 - L, \quad (3)$$

and conversely,

$$A_L = \frac{1}{\left(\frac{L+a}{L-1+a}\right)^\alpha - 1}. \quad (4)$$

Petersson and co-workers observed repeatedly^{32,51,65} that the CCSD-MP2 difference, at least for the nZaPa basis sets, asymptotically converges approximately as $(L+3/2)^{-3}$. Thus, CCSD-MP2 differences at the basis set limit can be obtained in two ways: using this formula and as differences of extrapolated CCSD and MP2 CBS limit. Comparison offers one estimate of the residual uncertainty in these quantities.

Another estimate of the uncertainty derives from Eq. (1): the uncertainty in the extrapolation parameter times the RMS atomization energy difference between the two largest basis

sets. As shown in Table I, Schwenke coefficients between different basis set families such as nZaPa and haVnZ, or between different calibrations, are actually remarkably consistent. We believe that an uncertainty of ± 0.1 in A_L is actually a somewhat conservative estimate; for the {6,7}ZaPha and haV{6,7}Z+d basis set pairs, this would translate into a RMS uncertainty of about 0.06 kcal/mol at the MP2 level and 0.04 kcal/mol at the CCSD level. For the CCSD-MP2 difference, this would be an even smaller 0.02 kcal/mol, but we would approximately double that amount on account of the greater uncertainty in the Schwenke parameter A_L . If we had limited ourselves to the {5,6} basis set pairs, all uncertainties would be approximately double those we have currently. For the discussion at hand, it means that the additional computational effort of the {6,7} basis set pairs is justified.

RESULTS AND DISCUSSION

Quality and convergence of the conventional reference data

MP2 convergence behavior using different approaches

We were able to obtain MP2/haV{6,7}Z+d data for 188 out of the 200 species, and MP2/{6,7}ZaPha as well as MP2-F12/awCV5Z data for the complete set of 200. The root mean square difference (RMSD) between MP2-F12/awCV5Z and MP2/haV{6,7}Z+d is 0.095 kcal/mol, but the RMSD between MP2-F12/awCV5Z and MP2/{6,7}ZaPha is only 0.064 kcal/mol. For the W4-08 subset of molecules, both RMSDs are comparable: 0.067 and 0.065 kcal/mol, respectively. The main sources of the difference appear to be a number of chlorine compounds such as CCl_4 and C_2Cl_6 , for which the haVnZ+d basis sets appear to exhibit oscillatory convergence.

As can be seen in Fig. 1, this is reflected in surprisingly large errors for some of these systems even with basis sets as large as haV{5,6}Z+d. Also by way of illustration, the RMSD between MP2/{5,6}ZaPha and MP2/{6,7}ZaPha is 0.13 kcal/mol but increases to 0.17 kcal/mol between MP2/haV{5,6}Z+d and MP2/haV{6,7}Z+d. We have previously⁷ discussed the overcontraction issues of the haVnZ+d

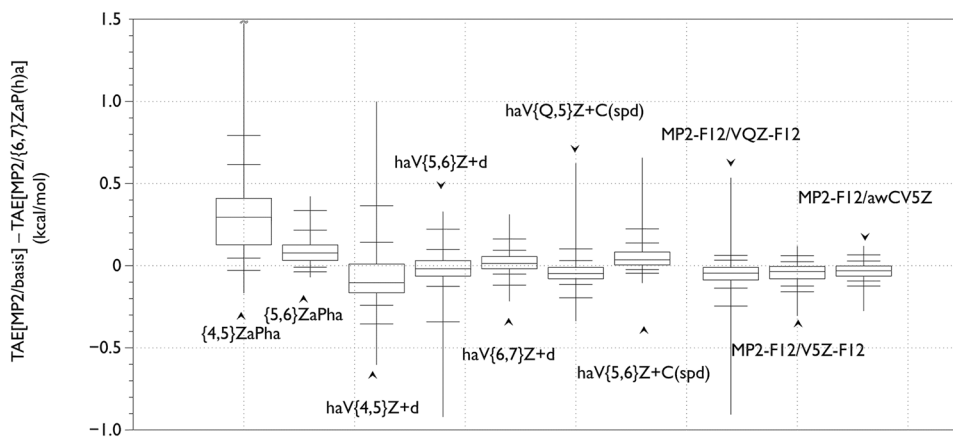


FIG. 1. Box plot of deviations from MP2/{6,7}ZaP(h)a reference values (kcal/mol) for the valence MP2 correlation components of the W4-17 atomization energy benchmark. The outer fences encompass the middle 95% of the distribution, the inner fences 80%, and the box 50%. Vertical lines span from population minimum to maximum.

basis sets for second-row atoms. As further evidence, for the {5,6} basis set pair, the RMSD between extrapolated MP2 values with n ZaPha and $haVnZ+d$ basis sets is 0.20 kcal/mol, which drops to 0.07 kcal/mol for the {6,7} pair.

In Ref. 7, we observed that increasing the radial flexibility of the aug-cc-pV(5+d)Z and aug-cc-pV(6+d)Z basis sets, e.g., by adding core-valence basis functions of the lower angular momenta, greatly remediated the oscillatory convergence observed especially for chlorine compounds. We have taken the “maximalist” of the different approaches discussed there and added *spd* core-valence functions from the corresponding aug-cc-pCVnZ basis set—this combination we have denoted $haVnZ+C(sp)$. (As the core-valence d functions already cover the tight d required for 3d orbital back-bonding at the SCF level, the “+d” addition was not required.) As can be seen in Fig. 1, this does bring some succor, but not as much as a large basis set MP2-F12 calculation.

CCSD-MP2 differences considered

We were able to obtain CCSD/{6,7}ZaPha values for 192 out of 200 molecules, and CCSD(F12*)/awCV5Z values for all species except C_6H_6 . Comparison of individual differences between MP2 and MP2-F12 for molecules that ought not present convergence issues with the {6,7} pair, such as alkanes, suggests that the hydrogen basis set in MP2-F12/awCV5Z may still be insufficiently saturated, although adequate saturation does appear to have been achieved for nonhydrogen species. We do not, however, expect the CCSD-MP2 difference to be appreciably affected by this, as these small discrepancies (about 0.01 kcal/mol per hydrogen, cf. Ref. 24) can reasonably be expected to cancel between CCSD and MP2. For instance, for propane, the MP2-F12/awCV5Z correlation contribution of 224.42 kcal/mol is more than 0.1 kcal/mol below the MP2/ $haV\{6,7\}Z$ and MP2/{6,7}ZaPha values of 224.56 kcal/mol and 224.53 kcal/mol, respectively. Yet for the CCSD(F12*)-MP2-F12 difference with the awCV5Z basis set, we obtain -15.61 kcal, which is considerably closer to the conventional {6,7}ZaPha values of -15.66 kcal/mol computed as CCSD/CBS-MP2/CBS, and -15.62 kcal/mol using Petersson’s $(L+3/2)^{-3}$ formula applied to the CCSD-MP2 differences of the {6,7} pair. The corresponding conventional

values for the $haV\{6,7\}Z$ pair are -15.68 and -15.66 kcal/mol, respectively.

Hence, we have compared the CCSD-MP2 differences between {6,7}ZaPha and $haV\{6,7\}Z+d$, and found the two sets of data to differ by just 0.034 kcal/mol RMSD. In addition, {5,6}ZaPha and {6,7}ZaPha differ by just 0.06 kcal/mol RMSD, and a similar small RMSD of 0.07 kcal/mol is seen between $haV\{5,6\}Z+d$ and $haV\{6,7\}Z+d$. This bolsters our confidence in the quality of the CCSD-MP2 differences with the {6,7} basis set pairs.

Can we legitimately omit the diffuse functions on H in {6,7}ZaPha? Particularly for the Al and Si compounds, placing no diffuse functions on H while attaching them to the metal and metalloid, respectively, seems dubious, and even for B this choice can be called into question. For the W4-08 subset, we performed calculations using {6,7}ZaPa in which diffuse functions were *not* omitted on hydrogen. MP2/{6,7} differences between the {6,7}ZaPha and {6,7}ZaPa basis set pairs were found to be 0.01 kcal/mol or less for the HBC-NOFPSCl compounds but reached 0.05 kcal/mol for Si_2H_6 , 0.04 kcal/mol for SiH_4 , and 0.03 kcal/mol for AlH_3 . The only Al or Si hydride in W4-17 that is not an element of the W4-08 subset is SiH_3F , for which we found a difference of 0.03 kcal/mol upon recalculation. We have hence selected, as our reference level, {6,7}ZaPa for SiH_3F and the W4-08 subset, and {6,7}ZaPha for the remainder. This also circumvents the oscillatory convergence issues we noted for the $haVnZ+d$ sequence.

The CCSD-MP2 values were obtained by taking the differences between extrapolated CCSD and MP2 limits. In principle, we could also consider the CCSD-MP2 differences directly, and extrapolate them according to the $(L+3/2)^{-3}$ formula observed empirically by Petersson and co-workers.^{65,68} For the {6,7} pair, the RMSD from the difference-of-limits values is 0.05 kcal/mol for the n ZaPha basis sets.

We also investigated the use of “interference corrections,”⁶⁹ but the results were erratic, and we have not retained them for our analysis.

Based on the RMS differences between the {6,7}ZaPha and $haV\{6,7\}Z+d$ data, as well as between extrapolation of differences and differences of extrapolations, we would estimate the remaining uncertainty of our chosen reference

TABLE II. RMSD (kcal/mol) for the W4-17 TAE benchmark from our best CCSD/{6,7}ZaP(h)a-MP2/{6,7}ZaP(h)a limits for different CCSD-F12 approximations. The aug-cc-pwCV5Z basis set with $\beta = 1.4$ was used throughout the F12 calculations.

	CCSD(F12*)-MP2-F12	CCSD-F12b-MP2-F12	CCSD[F12]-MP2-F12	CCSD-F12a-MP2-F12
RMSD	0.035	0.087	0.137	0.470
MSD	0.008	0.071	0.116	0.424
StDev ^a	0.034	0.050	0.073	0.202

^aCalculated as $(\text{RMSD}^2 - \text{MSD}^2)^{1/2}$.

level, i.e., [CCSD-MP2]/{6,7}ZaPha, to be on the order of 0.03–0.04 kcal/mol.

Agreement with approximate CCSD-F12 data near the basis set limit

We are now in a position to compare our best CCSD-MP2 values with the CCSD-F12-MP2-F12 differences obtained using various approximate CCSD-F12 methods. The aug-cc-pwCV5Z basis set was employed throughout.

First, we can see in Table II and Fig. 2 that CCSD(F12*) has an RMSD of only 0.035 kcal/mol, comparable to the uncertainty in the reference values.

CCSD(F12*) is itself an approximation to the more rigorous (and much more costly) CCSD(F12) method.⁷⁰ It has been shown before³⁸ that the difference between them is very small, but it does bear verifying this for the present W4-17 atomization energy benchmark. For the cc-pVTZ-F12 basis set, the difference between CCSD(F12*) and CCSD(F12) atomization energies is just 0.01 kcal/mol RMS, reaching a maximum of 0.035 kcal/mol for HClO₄. These differences (Fig. 3) become negligible altogether when expanding the basis set to cc-pVQZ-F12: RMS 0.002 kcal/mol, with a maximum of 0.008 kcal/mol for C₂Cl₆. For the W4-08 subset, we finally considered the cc-pV5Z-F12 dataset and found differences of less than 0.001 kcal/mol RMS that should be considered thermochemically equivalent within the numerical noise of the calculation. We conclude that for cc-pVQZ-F12 or larger basis sets, CCSD(F12*) is functionally equivalent to CCSD(F12).

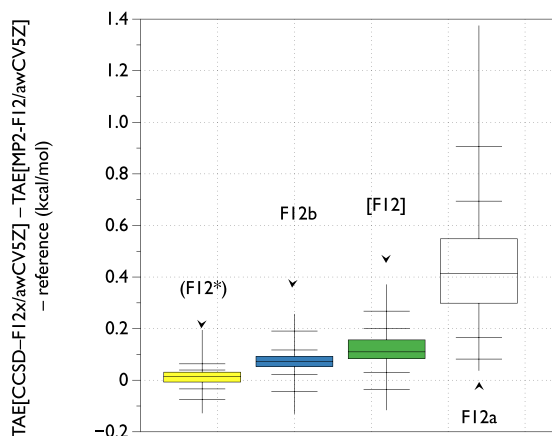


FIG. 2. Box plot of deviations from our best estimates for CCSD-MP2 components of W4-17 atomization energies (kcal/mol) using the awCV5Z basis set and different approximations to CCSD-F12.

While we cannot definitely rule out that post-CCSD(F12) corrections might be nontrivial for *some* molecules, verifying this is not technically feasible at present, both because of code limitations and because the changes would be below the resolution level of our reference data. [A few examples have been investigated in Ref. 71 using the cc-pVTZ-F12 basis set. It was found that while the individual contributions from terms beyond CCSD(F12) can be sizable, they cancel very systematically.]

By contrast, for CCSD-F12b we see an RMSD that is about 2.5 times as large and cannot easily be blamed on uncertainty in the reference values anymore. (In fact, 0.08 kcal/mol is about the RMSD of W4 theory for its training set.) This result is consistent with that of earlier, less rigorous, benchmark studies on smaller samples and using smaller basis sets.^{38,72} CCSD-F12b does represent a dramatic improvement over the still more approximate CCSD-F12a method, which has an error of 0.43 kcal/mol RMS even with such large basis sets.

Köhn and Tew⁷¹ carried out a detailed multiple perturbation theory analysis of approximations to CCSD-F12 in terms of five coupling strengths: single (σ) and double (λ) excitation amplitudes into the virtual orbitals, single (τ) and double (μ) excitations into the auxiliary basis set, and finally ν , which corresponds to mixed double excitations (one electron into virtual, the other into auxiliary space). The key points of the analysis are summarized in Table III: for mathematical details, we refer to Ref. 71 itself, specifically Tables I, II, and VII and the surrounding discussion.

The multiple perturbation expansion is winnowed by restricting the total order in μ and ν to at most two. Retaining then all remaining terms through third order corresponds to

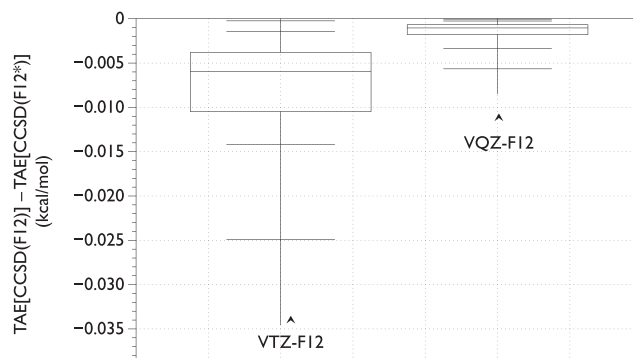


FIG. 3. Box plot of basis set convergence of CCSD(F12)-CCSD(F12*) differences (kcal/mol) for the W4-17 dataset.

TABLE III. Terms included in CCSD(F12) and approximations thereto. For further details, see Tables I and II in Ref. 71 and the surrounding discussion.

Parent expression ^a	Label	Order	(F12)	(F12*)	[F12]	F12b	F12a	(2) _{F12}	MP2-F12 + ΔCCSD
$\langle 0 G\bar{R} + \bar{R}^\dagger G + \bar{R}^\dagger F\bar{R} 0 \rangle$	$E_{\text{MP2}}^{\text{unc}}$	μ^2	✓	✓	✓	✓	✓	✓	✓
$\langle 0 \bar{R}^\dagger F^{(0)} T_2 0 \rangle$	R.1	λv	✓	✓	✓	✓	✓		
$\langle 0 \bar{R}^\dagger [G, T_2] 0 \rangle$	R.2	$\lambda \mu^2$	✓	✓	✓	✓		✓	
	R.3	λv^2	✓	✓	✓				
$\langle 0 \bar{R}^\dagger [G, T_1] 0 \rangle$	R.4	$\sigma \mu^2$	✓	✓	✓	✓		✓	
	R.5	σv^2	✓	✓	✓				
$\langle 0 \bar{R}^\dagger [[G, T_1], T_2] 0 \rangle$	R.6	$\sigma \lambda v^2$	✓						
$\frac{1}{2} \langle 0 \bar{R}^\dagger [[G, T_1], T_1] 0 \rangle$	R.7	$\sigma^2 \mu^2$	✓	✓		✓			
	R.8	$\sigma^2 v^2$	✓						
$\frac{1}{6} \langle 0 \bar{R}^\dagger [[[G, T_1], T_1], T_1] 0 \rangle$	R.9	$\sigma^3 v^2$	✓						
$\langle 0 \Lambda_1 [G, \bar{R}] 0 \rangle$	L1.1	$\sigma \mu^2$	✓	✓	✓	✓	✓	✓	
	L1.2	σv^2	✓	✓	✓				
$\langle 0 \Lambda_1 [F^{(1)}, \bar{R}] 0 \rangle$	L1.3	$\sigma \tau v$	✓	✓	✓				
$\langle 0 \Lambda_1 [[G, T_1], \bar{R}] 0 \rangle$	L1.4	$\sigma^2 \mu^2$	✓	✓					
	L1.5	$\sigma^2 v^2$	✓						
$\langle 0 \Lambda_2 F^{(0)} \bar{R} 0 \rangle$	L2.1	λv	✓	✓	✓	✓	✓		
$\langle 0 \Lambda_2 [G, \bar{R}] 0 \rangle$	L2.2	$\lambda \mu^2$	✓	✓	✓	✓	✓	✓	
	L2.3	λv^2	✓	✓	✓				
$\langle 0 \Lambda_2 [[G, T_2], \bar{R}] 0 \rangle$	L2.4	$\lambda^2 \mu^2$	✓	✓					
	L2.5	$\lambda^2 v^2$	✓						
$\langle 0 \Lambda_2 [[G, T_1], \bar{R}] 0 \rangle$	L2.6	$\sigma \lambda \mu^2$	✓	✓		✓	✓		
	L2.7	$\sigma \lambda v^2$	✓						
$\langle 0 \Lambda_2 [[F^{(1)}, T_1], \bar{R}] 0 \rangle$	L2.8	$\sigma \tau \lambda v$	✓						
$\frac{1}{2} \langle 0 \Lambda_2 [[[G, T_1], T_1], \bar{R}] 0 \rangle$	L2.9	$\sigma^2 \lambda \mu^2$	✓	✓					
	L2.10	$\sigma^2 \lambda v^2$	✓						

^a G , T_1 , and T_2 have their usual respective meanings of the two-electron repulsion operator, the single substitution operator, and the double substitution operator, respectively. The matrix elements of \bar{R} are those of the geminal function, while Λ_1 and Λ_2 are de-excitation operators associated with the Lagrange multipliers for the singles and doubles coupled-cluster residual equation. $F^{(0)}$ is the Fock operator and is only non-zero if the Brillouin or the generalized Brillouin theorem are not fulfilled $F^{(1)}$. For further details, see Ref. 71.

the CCSD[F12] method,⁷¹ which was not recommended for practical use (by its authors) but turns out to be useful in our analysis. The CCSD(F12*) method is obtained (see Table VII of Ref. 71) by adding a group of fourth-order terms quadratic in μ , namely, of orders $\sigma^2 \mu^2$, $\sigma \lambda \mu^2$, and $\lambda^2 \mu^2$, plus a single fifth-order term of order $\sigma^2 \lambda \mu^2$. By contrast, beyond second order, CCSD-F12b^{36,37} and Valeev's CCSD(2)_{F12} method⁷³ only contain the third-order contributions of orders $\sigma \mu^2$ and $\lambda \mu^2$, plus fourth-order contributions scaling as $\sigma^2 \mu^2$ and $\sigma \lambda \mu^2$. Finally, CCSD-F12a discards all CCSD-F12b contributions higher than second order to the energy equation, but not in the amplitude equations.

In the present work, we found CCSD[F12] to have an RMSD slightly higher than CCSD-F12b, 0.11 kcal/mol. Some additional light is shed by the mean signed difference

(MSD), as well as by the standard deviation about it (StDev). While CCSD(F12*) pleasingly has an RMSD close to zero (indicating an absence of systematic bias), CCSD-F12b, CCSD[F12], and CCSD-F12a have progressively larger systematic biases. In fact, the StDev values of CCSD-F12b and CCSD[F12] are somewhat comparable, while that for CCSD-F12a is about three times as large. The statistical behaviors of the various approximations are depicted as a box plot in Fig. 2.

Basis set convergence of differences between various F12 methods

Figure 4 illustrates the basis set convergence of the CCSD-F12b–CCSD-F12a difference. Said difference corresponds, in Table VII of Ref. 71, to the terms R.2($\lambda \mu^2$) + R.4($\sigma \mu^2$) + R.7

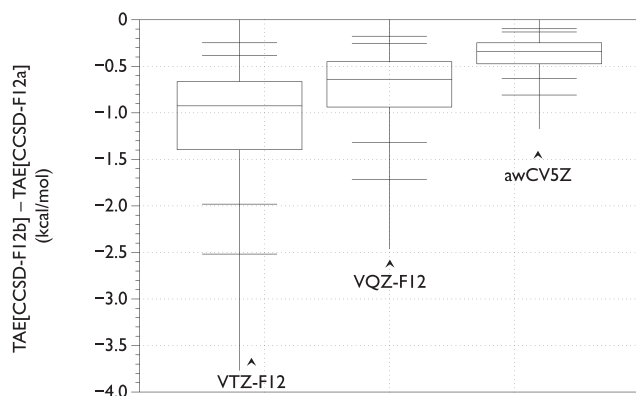


FIG. 4. Box plot of the CCSD-F12b–CCSD-F12a difference (kcal/mol) for the W4-17 dataset and different basis sets.

($\sigma^2\mu^2$), where the orders in perturbation theory are indicated in parentheses as powers of the coupling strengths.

For cc-pVTZ-F12, which is the smallest basis set one would normally consider for thermochemical production applications, the median difference is close to 1 kcal/mol, and for 2.5% of our test species it actually exceeds 2.5 kcal/mol. For cc-pVQZ-F12, which is the largest basis set used in typical CCSD-F12x applications, the median F12b–F12a difference is still close to 0.7 kcal/mol, and values in excess of 1.5 kcal/mol can still be found. These numbers can be cut in half with awCV5Z, but this is hardly an acceptable error level for such large (by F12 standards) basis sets. We conclude that CCSD-F12a cannot be recommended at all for production work, although for small basis sets the results may seem superior to CCSD-F12b due to an error compensation³⁸ between the missing terms (which are antibonding) and basis set incompleteness error (see also Ref. 12, for example, and Refs. 3 and 4 for further discussion).

It turns out to be enlightening to decompose the CCSD(F12*)–CCSD-F12b difference into two components: CCSD(F12*)–CCSD[F12] and CCSD[F12]–CCSD-F12b. Energetically speaking (see Figs. 3 and 4 in Ref. 71), the most important third-order contributions are (following the notation in Table VII of Ref. 71) the ladder terms R.2(λ,μ^2) and L2.2(λ,μ^2), followed by the ring terms R.3(λ,ν^2) and L2.3(λ,ν^2). (In this notation, R, L1, and L2 refer to the energy, singles amplitudes, and doubles amplitudes equations, respectively.) R.2 and L2.2 are already present in CCSD-F12b (in fact, L2.2 even in CCSD-F12a!), while CCSD[F12] adds in R.3(λ,ν^2) and L2.3(λ,ν^2), besides the smaller terms R.5 ($\sigma\nu^2$), L1.2 ($\sigma\nu^2$), and L1.3 ($\sigma\tau\nu$). The fourth-order terms R.7 ($\sigma^2\mu^2$) and L2.6 ($\sigma\lambda\mu^2$) from CCSD-F12b are omitted.

As can be seen in Fig. 5, the CCSD[F12]–CCSD-F12b difference (a) is universally attractive; (b) decays very rapidly with the basis set. As the F12 corrections should in principle vanish in the complete basis set limit, basis set extrapolation should yield results close to zero. Such extrapolation from V{T, Q}Z-F12 using the formula of Hill *et al.*⁴⁶ shows a very narrow box with a median close to zero (Fig. 5, far right): while there are clearly some outliers, the extrapolated values on the whole are somewhere between V5Z-F12 and awCV5Z in quality. This suggests that the convergence behavior is pretty

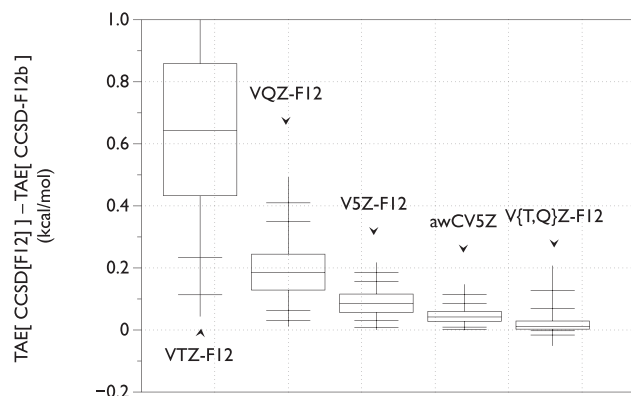


FIG. 5. Box plot of the CCSD[F12]–CCSD-F12b difference (kcal/mol) for the W4-17 dataset and different basis sets.

regular even with fairly modest basis sets. (There are solid theoretical grounds to believe that each ring term converges rapidly. For atoms, the partial wave expansion of terms with a single contraction over the CABS space was shown to be rapidly converging by Noga and Kutzelnigg (Appendix C of Ref. 74). This is the reason why they are neglected in the Standard Approximation.)

While it might be remotely possible that, for large basis sets, the seven terms in the CCSD[F12]–CCSD-F12b difference might simply benefit from an unusually felicitous form of mutual error cancellation, Occam’s razor explanation for the rapid tapering off of their sum with increasing basis set would seem to be that the individual terms converge fairly rapidly.

Basis set convergence of the CCSD(F12*)–CCSD[F12] difference is depicted in Fig. 6. As can be seen there, these contributions are universally repulsive and do taper off with the basis set: median values are about halved with each successive basis set, from about 0.4 kcal/mol for cc-pVTZ-F12 to 0.2 kcal/mol for cc-pVQZ-F12 to 0.1 kcal/mol for awCV5Z. Crucially, however—even for basis sets as large as awCV5Z—they remain nontrivial (0.1 kcal/mol median), exceeding 0.2 kcal/mol for about 10% of the sample and 0.3 kcal/mol for about 2.5%. What is more, the distribution is strongly skewed/asymmetric. Furthermore, extrapolation does not help for this term: the error distribution of V{T, Q}Z-F12 looks no better than that of VQZ-F12 itself, and likewise

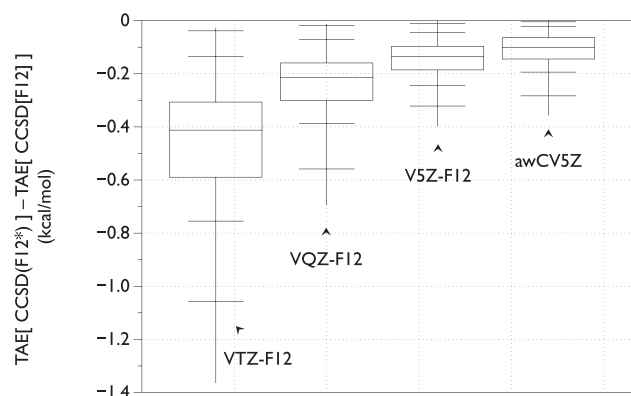


FIG. 6. Box plot of the CCSD(F12*)–CCSD[F12] difference (kcal/mol) for the W4-17 dataset and different basis sets.

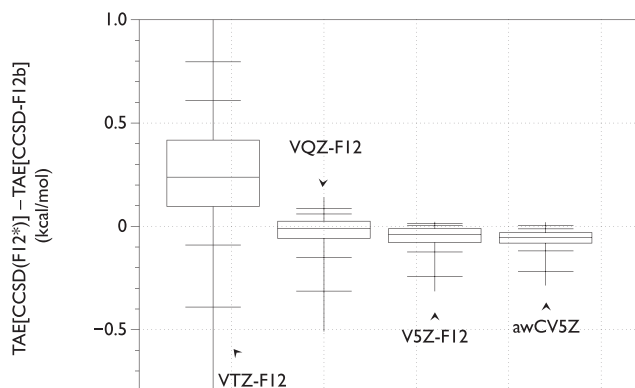


FIG. 7. Box plot of the CCSD(F12*)–CCSD-F12b difference (kcal/mol) for the W4-17 dataset and different basis sets.

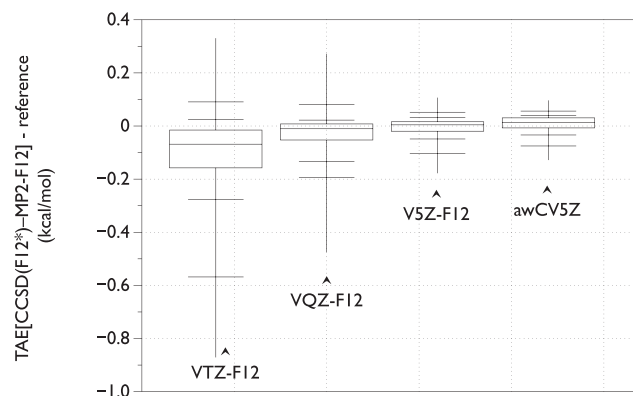


FIG. 8. Box plot of CCSD(F12*)–MP2-F12 difference (kcal/mol) for the W4-17 dataset and different basis sets.

for $V\{Q,5\}Z$ -F12 using the extrapolation from Table 2 in Ref. 11.

In the notation of Ref. 71, the difference corresponds to the following fourth-order terms: R.7 ($\sigma^2\mu^2$) in the energy expression, L1.4 ($\sigma^2\mu^2$) in the singles amplitudes equation, and L2.4 ($\lambda^2\mu^2$) + L2.6 ($\sigma\lambda\mu^2$) + L2.9 ($\sigma^2\lambda\mu^2$) in the doubles amplitude equation. Out of these, R.7 and L2.6 correspond to restoring the two F12b terms deleted in CCSD[F12] (which can hence be assumed to be small), which leaves L1.4, L2.4, and L2.6. For a much smaller sample of molecules and just the cc-pVDZ-F12 and cc-pVTZ-F12 basis sets, Fig. 3 in Ref. 71 shows a plot of the contribution of the various terms to the correlation energy (normalized by the number of valence electrons): L2.4 is by far the most important, followed by L1.4 two orders of magnitude lower and L2.6 another order of magnitude below.

Detailed inspection of the molecules for which the CCSD(F12*)–CCSD[F12] difference is large reveals an intriguing pattern: among small species, notorious multireference cases like $C_2(^1\Sigma_g^+)$, $BN(^1\Sigma^+)$, O_3 , FO_2 , $CIOO$, immediately jump out. If we normalize the difference by the number of valence electrons, then all the values in excess of 10 cal/(mol/e⁻) are seen for molecules with significant static correlation, such as N_2O_4 , HO_3 , FO_2 , F_2O_2 , P_4 , S_4 , $CIOO$, NO_2 , N_2O , O_3 , $BN(^1\Sigma^+)$, $C_2(^1\Sigma_g^+)$, $B_2(^3\Sigma_g^-)$, P_2 , and CIO .

L2.4 is a first-order geminal correction to a disconnected quadruples ($\hat{T}_2^2/2$) term, as is L1.4 to the disconnected doubles

$\hat{T}_1^2/2$. It stands to reason that in situations where some doubles amplitudes T_2 are large, L2.4 will become important, and to a lesser extent, so will L1.4 if some singles amplitudes T_1 are large. (If some of both are large, the T_1T_2 -geminals coupling term L2.6 could become significant.) Such scenarios occur, of course, when there is significant static correlation in the molecule.

Coming back to the CCSD(F12*)–CCSD-F12b difference, the superposition of two terms with opposite sign and different (fast vs. relatively slow) convergence rates leads to the convergence behavior seen in Fig. 7.

The reader may wonder how these various terms evolve as a molecule grows, e.g., along the n-alkane sequence $CH_3(CH_2)_nCH_3$ ($n=0, 1, 2, 3$) or the sequence ethylene, *trans*-butadiene, and 1,3,5-hexatriene. As seen in Table IV for the cc-pVTZ-F12 and cc-pVQZ-F12 basis sets, the [F12]–F12b and (F12*)–[F12] differences grow almost perfectly linearly along the sequences, as expected for strictly size extensive methods like coupled-cluster theory. In turn, size extensivity means that all conclusions from this paper carry over to arbitrarily large systems, as long as strong long-range correlations do not play a role (in which case single-reference CCSD breaks down anyway).

Finally, let us address the basis set convergence of the most rigorous approximation to CCSD–MP2, i.e., CCSD(F12*)–MP2-F12. This is depicted in Fig. 8. As can be seen there, one needs at least a cc-pVQZ-F12 basis set for a satisfactory error distribution.

TABLE IV. Chain length dependence of differences (kcal/mol) between different CCSD(F12) approximations.

	VTZ-F12			VQZ-F12		
	[F12]-F12b	(F12*)-[F12]	(F12*)-F12b	[F12]-F12b	(F12*)-[F12]	(F12*)-F12b
n-pentane	1.597	-0.581	1.016	0.441	-0.300	0.140
n-butane	1.293	-0.467	0.826	0.356	-0.242	0.114
Propane	0.988	-0.353	0.634	0.270	-0.183	0.087
Ethane	0.682	-0.239	0.443	0.184	-0.124	0.060
Methane	0.372	-0.127	0.245	0.096	-0.066	0.030
Hexatriene	1.653	-0.706	0.947	0.455	-0.368	0.088
Butadiene	1.123	-0.475	0.648	0.308	-0.247	0.060
Ethylene	0.593	-0.244	0.349	0.160	-0.127	0.033

Relationship to static correlation diagnostics

It would, of course, be helpful to have an *a priori* prediction for whether the $(F12^*)-F12b$ difference is non-trivial. Thus, establishing a link with static correlation character would serve a pragmatic as well as an interpretative purpose.

Figure 9 depicts the Pearson correlation coefficients R between various static correlation diagnostics and the differences between various F12 approximations. In this table, the cc-pVQZ-F12 basis set was used for the F12 results, as the $(F12^*)-[F12]$ term, in particular, was deemed too small with the awCV5Z basis set to yield useful conclusions.

It should be kept in mind that the various diagnostics do not all measure the same thing. Hollett and Gill³⁹ (see also Scuseria and Tsuchimochi⁷⁵) distinguish between “type A static correlation” (absolute near-degeneracy, with dissociating H_2 as the paradigmatic example) and “type B static correlation” (relative near-degeneracy, with Be-like ions (Z) (Z^{-4})⁺ as a textbook case). The von Neumann correlation entropy S_{corr} , for instance (a.k.a., “entanglement entropy”), is strongly linked to type A static correlation, while pragmatic diagnostics like the percentage of connected triples in the total atomization energy, $\%TAE[(T)]$,⁵⁸ or the percentage of post-CCSD(T) correction are primarily concerned with thermochemical importance. (Note that these latter diagnostics remain identically zero for the Be-like ions, as CCSD is an exact solution for the two-electron problem.) The density functional theory (DFT)-based $A_{25}[PBE]$ diagnostic,⁷⁶ which is based on the slope of the DFT atomization energy as a function of the HF-like exchange percentage, likewise is more pragmatic in character, although it has been argued⁷⁶ that it primarily samples type B static correlation.

Very recently, Matito and co-workers⁷⁷ proposed two new diagnostics for the importance of nondynamical and dynamical correlation that are based on the natural orbital

occupations. While I_{ND} and I_{D} are not intensive but extensive (their values for a dimer at infinite separation are exactly the sums of the respective monomer values), the quantity $I_{\text{ND}}/(I_{\text{ND}} + I_{\text{D}})$ can be considered as an intensive quantity (like the popular T_1 diagnostic,⁷⁸ which is the Euclidian norm of the single substitution vector divided by the square root of the number of correlated electrons).

While none of the correlations in Table III are good enough to permit quantitative estimation by linear regression, we *can* identify a few correlations of 0.8 or better. In particular, the “pragmatic” thermochemical diagnostics $\%TAE[(T)]$ and $A_{25}[PBE]$ have correlation coefficients of about -0.8 with the $(F12^*)-F12b$ difference. [$\%TAE[(T_4+T_5)]$, unless needed anyway as part of a W4, HEAT,⁷⁹ or Feller-Peterson-Dixon (FPD)⁸⁰ calculation, is simply too costly to serve as an *a priori* estimate.] Matito’s diagnostic and the correlation entropy⁸¹ S_{corr} , on the other hand, have equally good positive correlations with the $[F12]-F12b$ difference. (We note that for the W4-17 dataset, we find $R = 0.994$ between Matito’s I_{ND} and S_{corr} , indicating that they largely tell the same story.) If, on the other hand, we normalize the $(F12^*)-F12b$ difference by the number of valence electrons, we get a negative correlation $R = -0.83$ with the T_1 diagnostic.

It should be clarified that the minus sign in front of the larger negative correlations reflects negative (antibonding) contributions to the atomization energy. For total energies, the signs would be reversed.

The D_1 ⁸² and especially D_2 diagnostics⁸³ also bear mentioning. While T_1 corresponds to a vector norm divided by the square root of the number of valence electrons, D_1 corresponds to a matrix norm, i.e., the square root of the largest eigenvalue of $T_1 \cdot T_1^\dagger$. Similarly, D_2 is obtained as the root of the largest eigenvalue of the double excitations amplitude matrix multiplied by its transpose, $T_2 \cdot T_2^\dagger$. Because of the mathematical properties of matrix norms, D_1 and D_2 are much more

	unnormalized				Normalized by number of valence electrons			
	F12b-F12a	(F12*)-F12b	[F12]-F12b	(F12*)-[F12]	F12b-F12a	(F12*)-F12b	[F12]-F12b	(F12*)-[F12]
$\%TAE[(T)]$	0.125	-0.796	0.001	-0.631	0.308	-0.761	-0.205	-0.664
$\%TAE[\text{post-}(T)]$	0.248	-0.632	-0.226	-0.400	0.279	-0.614	-0.293	-0.489
$\%TAE[T_4+T_5]$	0.112	-0.785	-0.037	-0.597	0.222	-0.797	-0.150	-0.707
$A_{25}[PBE]$	0.191	-0.776	-0.074	-0.590	0.353	-0.721	-0.282	-0.615
T_1 diagnostic	0.059	-0.644	-0.059	-0.463	0.020	-0.834	0.013	-0.768
D_1 diagnostic	0.004	-0.721	0.043	-0.573	0.090	-0.799	0.003	-0.749
Truhlar M_{diag}	0.058	-0.562	-0.057	-0.394	0.026	-0.729	-0.031	-0.640
Matito I_{ND}	-0.510	-0.524	0.792	-0.726	0.209	-0.171	0.160	-0.207
Matito I_{D}	-0.539	-0.374	0.841	-0.643	0.257	0.059	0.123	-0.003
D_2 diagnostic	-0.013	-0.780	0.104	-0.652	0.132	-0.842	-0.826	-0.826
S_{corr}	-0.525	-0.460	0.815	-0.690	0.225	-0.075	-0.122	-0.122
$r_{\text{ND}} = I_{\text{ND}} / (I_{\text{ND}} + I_{\text{D}})$	0.093	-0.369	-0.113	-0.214	-0.061	-0.640	0.027	-0.564
$r_{\text{ND}} E_{\text{corr}}(\text{CCSD})$	0.341	0.623	-0.683	0.804	-0.395	0.248	0.023	0.270

FIG. 9. Pearson correlation coefficients R between F12x differences and various nondynamical correlation diagnostics. In the red-white-blue color “heatmap,” red refers to large positive R , blue to large negative R , and white to near-zero R .

likely to reflect the part of the molecule with the most severe static correlation. (This becomes apparent from considering the didactic example of $\text{BN} \cdots n\text{-octane}$ at long separation. The large T_1 diagnostic of BN will be “diluted” by the small one of the alkane, while D_1 and D_2 are identical to those of BN monomer.) It would seem to be reasonable that a diagnostic for important double excitation amplitudes would be at least a semiquantitative predictor for $\text{CCSD}(\text{F12}^*)\text{--CCSD-F12b}$ differences, considering that they appear to be driven by the L2.4 double excitations coupling term.

We also note in the full correlation matrix between the diagnostics (see Table S1 in Ref. 84) that D_2 correlates considerably better with the energetic diagnostics like $\%TAE(T)$ than does T_1 or D_1 .

PERSPECTIVE AND CONCLUSIONS

Following an extensive survey for a large thermochemical benchmark, we are in a position to conclude the following:

- (a) Near the one-particle basis set limit, the difference between $\text{CCSD}(\text{F12}^*)$ and conventional CCSD/CBS is comparable to the uncertainty in the latter (less than 0.04 kcal/mol RMS).
- (b) The difference between $\text{CCSD}(\text{F12}^*)$ and $\text{CCSD}(\text{F12})$ becomes thermochemically negligible (0.002 kcal/mol RMS) already with the cc-pVQZ-F12 basis set.
- (c) We hence have no indication that CCSD-F12 terms beyond $\text{CCSD}(\text{F12})$ would play a significant thermochemical role with large basis sets.
- (d) By contrast, a thermochemically significant difference between $\text{CCSD}(\text{F12}^*)$ and the more widely used CCSD-F12b approximation remains even with basis sets as large as aug-cc-pwCV5Z .
- (e) Said difference is largest in molecules with significant degrees of static correlation and negligible in molecules dominated by dynamical correlation.
- (f) The (mostly third-order) terms that make up the difference between CCSD-F12b and $\text{CCSD}[\text{F12}]$ converge rapidly with the basis set and are universally bonding. They can also be greatly reduced through basis set extrapolation.
- (g) The fourth-order terms that make up the difference between $\text{CCSD}(\text{F12}^*)$ and $\text{CCSD}[\text{F12}]$ converge more slowly and are universally antibonding. These terms cannot be well reduced through basis set extrapolation, at least not in practical basis set regimes.
- (h) The sometimes nonmonotonic basis set convergence of the $\text{CCSD}(\text{F12}^*)\text{--CCSD-F12b}$ difference results from the different rates of convergence of these last two opposing terms.
- (i) The CCSD-F12a approximation continues to have unacceptably large errors even near the basis set limit.
- (j) There is some evidence that, for $n = \{5, 6, 7\}$, the $n\text{ZaPa}$ family of basis sets exhibits somewhat smoother basis set convergence than the $\text{aug-cc-pV}(n+d)\text{Z}$ sequence.
- (k) If at all possible, $\text{CCSD}(\text{F12}^*)$ is to be preferred over CCSD-F12b . For codes that only have a closed-shell implementation of $\text{CCSD}(\text{F12}^*)$, such as MOLPRO,

the use of closed-shell reaction cycles may represent a workaround.

ACKNOWLEDGMENTS

The authors would like to thank Professor Amir Karton (U. of Western Australia, Perth) and Professor John F. Stanton (Quantum Theory Project, U. of Florida) for helpful discussions. J.M.L.M. would also like to thank Professor Hans-Joachim Werner (U. Stuttgart, Germany) for helpful comments on a conference poster based on an early draft of this presentation.

This research was supported by the Israel Science Foundation (Grant No. 1358/15) and by the Minerva Foundation, Munich, Germany, as well as by two internal Weizmann Institute funding sources: the Helen and Martin Kimmel Center for Molecular Design and a research grant from the estate of Emile Mimran. N.S. and M.K.K. acknowledge doctoral and post-doctoral fellowships, respectively, from the Feinberg Graduate School (WIS).

- ¹S. Ten-no, “Explicitly correlated wave functions: Summary and perspective,” *Theor. Chem. Acc.* **131**(1), 1070 (2012).
- ²S. Ten-no and J. Noga, “Explicitly correlated electronic structure theory from R12/F12 ansätze,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**(1), 114–125 (2012).
- ³L. Kong, F. A. Bischoff, and E. F. Valeev, “Explicitly correlated R12/F12 methods for electronic structure,” *Chem. Rev.* **112**(1), 75–107 (2012).
- ⁴C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, “Explicitly correlated electrons in molecules,” *Chem. Rev.* **112**(1), 4–74 (2012).
- ⁵K. D. Vogiatzis, R. Haunschuld, and W. Klopper, “Accurate atomization energies from combining coupled-cluster computations with interference-corrected explicitly correlated second-order perturbation theory,” *Theor. Chem. Acc.* **133**(3), 1–12 (2014).
- ⁶W. Klopper, B. Ruscic, D. P. Tew, F. A. Bischoff, and S. Wolfsegger, “Atomization energies from coupled-cluster calculations augmented with explicitly-correlated perturbation theory,” *Chem. Phys.* **356**(1-3), 14–24 (2009).
- ⁷N. Sylvetsky, K. A. Peterson, A. Karton, and J. M. L. Martin, “Toward a W4-F12 approach: Can explicitly correlated and orbital-based *ab initio* $\text{CCSD}(T)$ limits Be reconciled?,” *J. Chem. Phys.* **144**(21), 214101 (2016).
- ⁸L. A. Burns, M. S. Marshall, C. D. Sherrill, L. A. Burns, M. S. Marshall, and C. D. Sherrill, “Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches appointing silver and bronze standards for noncovalent interactions,” *J. Chem. Phys.* **141**, 234111 (2014).
- ⁹D. A. Sirianni, L. A. Burns, and C. D. Sherrill, “Comparison of explicitly correlated methods for computing high-accuracy benchmark energies for noncovalent interactions,” *J. Chem. Theory Comput.* **13**(1), 86–99 (2017).
- ¹⁰M. K. Kesharwani, A. Karton, and J. M. L. Martin, “Benchmark *ab initio* conformational energies for the proteinogenic amino acids through explicitly correlated methods. Assessment of density functional methods,” *J. Chem. Theory Comput.* **12**(1), 444–454 (2016).
- ¹¹B. Brauer, M. K. Kesharwani, S. Kozuch, and J. M. L. Martin, “The S66x8 benchmark for noncovalent interactions revisited: Explicitly correlated *ab initio* methods and density functional theory,” *Phys. Chem. Chem. Phys.* **18**(31), 20905–20925 (2016).
- ¹²D. Manna, M. K. Kesharwani, N. Sylvetsky, and J. M. L. Martin, “Conventional and explicitly correlated *ab initio* benchmark study on water clusters: Revision of the BEGDB and WATER27 data sets,” *J. Chem. Theory Comput.* **13**(7), 3136–3152 (2017).
- ¹³D. P. Tew, W. Klopper, M. Heckert, and J. Gauss, “Basis set limit $\text{CCSD}(T)$ harmonic vibrational frequencies †,” *J. Phys. Chem. A* **111**(44), 11242–11248 (2007).
- ¹⁴J. M. L. Martin and M. K. Kesharwani, “Assessment of $\text{CCSD}(T)\text{--F12}$ approximations and basis sets for harmonic vibrational frequencies,” *J. Chem. Theory Comput.* **10**(5), 2085–2090 (2014).

- ¹⁵D. P. Tew, W. Klopper, C. Neiss, and C. Hättig, "Quintuple- ζ quality coupled-cluster correlation energies with triple- ζ basis sets," *Phys. Chem. Chem. Phys.* **9**(16), 1921–1930 (2007).
- ¹⁶S. Ten-no, "Initiation of explicitly correlated slater-type geminal theory," *Chem. Phys. Lett.* **398**(1-3), 56–61 (2004).
- ¹⁷C. M. Johnson, S. Hirata, and S. Ten-no, "Explicit correlation factors," *Chem. Phys. Lett.* **683**, 247–252 (2017).
- ¹⁸I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics* (Cambridge University Press, Cambridge, 2009).
- ¹⁹G. D. Purvis and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," *J. Chem. Phys.* **76**(4), 1910–1918 (1982).
- ²⁰K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.* **157**(6), 479–483 (1989).
- ²¹J. D. Watts, J. Gauss, and R. J. Bartlett, "Coupled-cluster methods with non-iterative triple excitations for restricted open-shell Hartree–Fock and other general single determinant reference functions. Energies and analytical gradients," *J. Chem. Phys.* **98**(11), 8718–8733 (1993).
- ²²O. Marchetti and H.-J. Werner, "Accurate calculations of intermolecular interaction energies using explicitly correlated coupled cluster wave functions and a dispersion-weighted MP2 method," *J. Phys. Chem. A* **113**(43), 11580–11585 (2009).
- ²³O. Marchetti and H.-J. Werner, "Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions," *Phys. Chem. Chem. Phys.* **10**(23), 3400–3409 (2008).
- ²⁴K. A. Peterson, M. K. Kesharwani, and J. M. L. Martin, "The cc-pV5Z-F12 basis set: Reaching the basis set limit in explicitly correlated calculations," *Mol. Phys.* **113**(13-14), 1551–1558 (2015).
- ²⁵A. Köhn, "Explicitly correlated connected triple excitations in coupled-cluster theory," *J. Chem. Phys.* **130**(13), 131101 (2009).
- ²⁶A. Köhn, "Explicitly correlated coupled-cluster theory using cusp conditions. II. Treatment of connected triple excitations," *J. Chem. Phys.* **133**(17), 174118 (2010).
- ²⁷A. Karton, S. Daon, and J. M. L. Martin, "W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data," *Chem. Phys. Lett.* **510**, 165–178 (2011).
- ²⁸A. Karton, N. Sylvetsky, and J. M. L. Martin, "W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods," *J. Comput. Chem.* **38**(24), 2063–2075 (2017).
- ²⁹W. Klopper, J. Noga, H. Koch, and T. Helgaker, "Multiple basis sets in calculations of triples corrections in coupled-cluster theory," *Theor. Chem. Acc. Theory. Comput. Model. (Theor. Chim. Acta)* **97**(1-4), 164–176 (1997).
- ³⁰D. W. Schwenke, "The extrapolation of one-electron basis sets in electronic structure calculations: How it should work and how it can be made to work," *J. Chem. Phys.* **122**(1), 014107 (2005).
- ³¹J. M. L. Martin and G. de Oliveira, "Towards standard methods for benchmark quality *ab initio* thermochemistry—W1 and W2 theory," *J. Chem. Phys.* **111**(5), 1843–1856 (1999).
- ³²D. S. Ranasinghe and G. A. Petersson, "CCSD(T)/CBS atomic and molecular benchmarks for H through Ar," *J. Chem. Phys.* **138**(14), 144104 (2013).
- ³³R. Kalescky, E. Kraka, and D. Cremer, "Accurate determination of the binding energy of the formic acid dimer: The importance of geometry relaxation," *J. Chem. Phys.* **140**(8), 084315 (2014).
- ³⁴D. Feller, "Statistical electronic structure calibration study of the CCSD(T*)-F12b method for atomization energies," *J. Phys. Chem. A* **119**(28), 7375–7387 (2015).
- ³⁵E. Miliordos and S. S. Xantheas, "On the validity of the basis set superposition error and complete basis set limit extrapolations for the binding energy of the formic acid dimer," *J. Chem. Phys.* **142**, 094311 (2015).
- ³⁶T. B. Adler, G. Knizia, and H.-J. Werner, "A simple and efficient CCSD(T)-F12 approximation," *J. Chem. Phys.* **127**(22), 221106 (2007).
- ³⁷G. Knizia, T. B. Adler, and H.-J. Werner, "Simplified CCSD(T)-F12 methods: Theory and benchmarks," *J. Chem. Phys.* **130**(5), 054104 (2009).
- ³⁸C. Hättig, D. P. Tew, and A. Köhn, "Communications: Accurate and efficient approximations to explicitly correlated coupled-cluster singles and doubles, CCSD-F12," *J. Chem. Phys.* **132**(23), 231102 (2010).
- ³⁹J. W. Hollett and P. M. W. Gill, "The two faces of static correlation," *J. Chem. Phys.* **134**(11), 114111 (2011).
- ⁴⁰A. Karton, A. Tarnopolsky, J.-F. Lamère, G. C. Schatz, and J. M. L. Martin, "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," *J. Phys. Chem. A* **112**(50), 12868–12886 (2008).
- ⁴¹N. Sylvetsky and J. M. L. Martin, "Probing the basis set limit for thermochemical contributions of inner-shell correlation: Balance of core-core and core-valence contributions," *Mol. Phys.* (published online).
- ⁴²K. A. Peterson and T. H. Dunning, "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.* **117**(23), 10548–10560 (2002).
- ⁴³J. G. Hill, S. Mazumder, and K. A. Peterson, "Correlation consistent basis sets for molecular core-valence effects with explicitly correlated wave functions: The atoms B–Ne and Al–Ar," *J. Chem. Phys.* **132**(5), 054108 (2010).
- ⁴⁴J. M. L. Martin and S. Parthiban, "W1 and W2 theories, and their variants: Thermochemistry in the KJ/mol accuracy range," in *Quantum-Mechanical Prediction of Thermochemical Data*, Understanding Chemical Reactivity Vol. 22, edited by J. Cioslowski (Kluwer Academic Publishers, Dordrecht, 2002), pp. 31–65.
- ⁴⁵K. A. Peterson, T. B. Adler, and H.-J. Werner, "Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B–Ne, and Al–Ar," *J. Chem. Phys.* **128**(8), 084102 (2008).
- ⁴⁶J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, "Extrapolating MP2 and CCSD explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets," *J. Chem. Phys.* **131**(19), 194105 (2009).
- ⁴⁷D. Feller, K. A. Peterson, and T. D. Crawford, "Sources of error in electronic structure calculations on small chemical systems," *J. Chem. Phys.* **124**(5), 054107 (2006).
- ⁴⁸T. H. Dunning, K. A. Peterson, and A. K. Wilson, "Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited," *J. Chem. Phys.* **114**(21), 9244–9253 (2001).
- ⁴⁹J. M. L. Martin, "Heats of formation of perchloric acid, HClO₄, and perchloric anhydride, Cl₂O₇. Probing the limits of W1 and W2 theory," *J. Mol. Struct.: THEOCHEM* **771**(1-3), 19–26 (2006).
- ⁵⁰S. Zhong, E. C. Barnes, and G. A. Petersson, "Uniformly convergent *n*-tuple- ζ augmented polarized (*n*ZaP) basis sets for complete basis set extrapolations. I. Self-consistent field energies," *J. Chem. Phys.* **129**(18), 184116 (2008).
- ⁵¹E. C. Barnes and G. A. Petersson, "MP2/CBS atomic and molecular benchmarks for H through Ar," *J. Chem. Phys.* **132**(11), 114111 (2010).
- ⁵²K. A. Peterson, "Gaussian basis sets exhibiting systematic convergence to the complete basis set limit," *Ann. Rep. Comput. Chem.* **3**, 196–206 (2007).
- ⁵³F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," *Phys. Chem. Chem. Phys.* **7**(18), 3297–3305 (2005).
- ⁵⁴M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, M. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, GAUSSIAN 09, Revision 01, Gaussian, Inc., Wallingford, CT, 2012.
- ⁵⁵M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin,

- V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, GAUSSIAN 16, Revision B01, Gaussian, Inc., Wallingford, CT, 2016.
- ⁵⁶F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, "TurboMole," *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**(2), 91–100 (2013).
- ⁵⁷H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselman, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. M. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, MOLPRO, version 2015.1, a package of *ab initio* programs, University of Cardiff Chemistry Consultants (UC3), Cardiff, Wales, UK, 2015, see <http://www.molpro.net>.
- ⁵⁸A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, "W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions," *J. Chem. Phys.* **125**(14), 144108 (2006).
- ⁵⁹F. Weigend, A. Köhn, and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," *J. Chem. Phys.* **116**(8), 3175–3183 (2002).
- ⁶⁰C. Hättig, "Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple-zeta basis sets for H to Ar and QZVPP basis sets for Li to Kr," *Phys. Chem. Chem. Phys.* **7**(1), 59–66 (2005).
- ⁶¹F. A. Weigend, "Fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency," *Phys. Chem. Chem. Phys.* **4**(18), 4285–4291 (2002).
- ⁶²K. E. Yousaf and K. A. Peterson, "Optimized complementary auxiliary basis sets for explicitly correlated methods: aug-cc-pVnZ orbital basis sets," *Chem. Phys. Lett.* **476**(4-6), 303–307 (2009).
- ⁶³K. E. Yousaf and K. A. Peterson, "Optimized auxiliary basis sets for explicitly correlated methods," *J. Chem. Phys.* **129**(18), 184108 (2008).
- ⁶⁴J. M. L. Martin, "A simple 'range extender' for basis set extrapolation methods for MP2 and coupled cluster correlation energies," AIP Conf. Proc. (in press); e-print [arXiv:1804.08844](https://arxiv.org/abs/1804.08844).
- ⁶⁵E. C. Barnes, G. A. Petersson, D. Feller, and K. A. Peterson, "The CCSD(T) complete basis set limit for Ne revisited," *J. Chem. Phys.* **129**(19), 194115 (2008).
- ⁶⁶A. J. C. Varandas and F. N. N. Pansini, "Narrowing the error in electron correlation calculations by basis set Re-hierarchization and use of the unified singlet and triplet electron-pair extrapolation scheme: Application to a test set of 106 systems," *J. Chem. Phys.* **141**, 224113 (2014).
- ⁶⁷F. N. N. Pansini, A. C. Neto, and A. J. C. Varandas, "On the performance of various hierarchized bases in extrapolating the correlation energy to the complete basis set limit," *Chem. Phys. Lett.* **641**, 90–96 (2015).
- ⁶⁸D. S. Ranasinghe, M. J. Frisch, and G. A. Petersson, "A density functional for core-valence correlation energy," *J. Chem. Phys.* **143**(21), 214111 (2015).
- ⁶⁹M. R. Nyden and G. A. Petersson, "Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions," *J. Chem. Phys.* **75**(4), 1843 (1981).
- ⁷⁰H. Fliegler, W. Klopper, and C. Hättig, "Coupled-cluster theory with simplified linear-R12 corrections: The CCSD(R12) model," *J. Chem. Phys.* **122**(8), 084107 (2005).
- ⁷¹A. Köhn and D. P. Tew, "Explicitly correlated coupled-cluster theory using cusp conditions. I. Perturbation analysis of coupled-cluster singles and doubles (CCSD-F12)," *J. Chem. Phys.* **133**(17), 174117 (2010).
- ⁷²D. P. Tew, "Explicitly correlated coupled-cluster theory with Brueckner orbitals," *J. Chem. Phys.* **145**(7), 074103 (2016).
- ⁷³E. F. Valeev, "Coupled-cluster methods with perturbative inclusion of explicitly correlated terms: A preliminary investigation," *Phys. Chem. Chem. Phys.* **10**(1), 106–113 (2008).
- ⁷⁴J. Noga and W. Kutzelnigg, "Coupled cluster theory that takes care of the correlation cusp by inclusion of linear terms in the interelectronic coordinates," *J. Chem. Phys.* **101**(9), 7738–7762 (1994).
- ⁷⁵G. E. Scuseria and T. Tsuchimochi, "Constrained-pairing mean-field theory. II. Exact treatment of dissociations to nondegenerate orbitals," *J. Chem. Phys.* **131**(16), 164119 (2009).
- ⁷⁶U. R. Fogueri, S. Kozuch, A. Karton, and J. M. L. Martin, "A simple DFT-based diagnostic for nondynamical correlation," *Theor. Chem. Acc.* **132**(1), 1291 (2012).
- ⁷⁷E. Ramos-Cordoba, P. Salvador, and E. Matito, "Separation of dynamic and nondynamic correlation," *Phys. Chem. Chem. Phys.* **18**(34), 24015–24023 (2016).
- ⁷⁸T. J. Lee and P. R. Taylor, "A diagnostic for determining the quality of single-reference electron correlation methods," *Int. J. Quantum Chem.* **36**(S23), 199–207 (1989).
- ⁷⁹M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, "High-accuracy extrapolated *ab initio* thermochemistry. III. Additional improvements and overview," *J. Chem. Phys.* **128**(11), 114111 (2008).
- ⁸⁰D. A. Dixon, D. Feller, and K. A. Peterson, "A practical guide to reliable first principles computational thermochemistry predictions across the periodic table," *Ann. Rep. Comput. Chem.* **8**, 1–28 (2012).
- ⁸¹P. Zanardi, "Quantum entanglement in fermionic lattices," *Phys. Rev. A* **65**(4), 042101 (2002).
- ⁸²C. Janssen, "New diagnostics for coupled-cluster and Møller-Plesset perturbation theory," *Chem. Phys. Lett.* **290**(4-6), 423–430 (1998).
- ⁸³I. M. B. Nielsen and C. L. Janssen, "Double-substitution-based diagnostics for coupled-cluster and Møller-Plesset perturbation theory," *Chem. Phys. Lett.* **310**(5-6), 568–576 (1999).
- ⁸⁴Our best reference MP2, MP2-F12, CCSD, and CCSD(F12*) data for W4-17 are freely available online (in Microsoft Excel format) in the FigShare data repository at <http://doi.org/10.6084/m9.figshare.6818564>, reference number 6818564.