

## Explicitly correlated Wn theory: W1-F12 and W2-F12

Amir Karton and Jan M. L. Martin

Citation: *J. Chem. Phys.* **136**, 124114 (2012); doi: 10.1063/1.3697678

View online: <http://dx.doi.org/10.1063/1.3697678>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v136/i12>

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

---

### Additional information on *J. Chem. Phys.*

Journal Homepage: <http://jcp.aip.org/>

Journal Information: [http://jcp.aip.org/about/about\\_the\\_journal](http://jcp.aip.org/about/about_the_journal)

Top downloads: [http://jcp.aip.org/features/most\\_downloaded](http://jcp.aip.org/features/most_downloaded)

Information for Authors: <http://jcp.aip.org/authors>

#### ADVERTISEMENT

**AIP**Advances

*Submit Now*

### Explore AIP's new open-access journal

- Article-level metrics now available
- Join the conversation! Rate & comment on articles

## Explicitly correlated $W_n$ theory: W1-F12 and W2-F12

Amir Karton<sup>1</sup> and Jan M. L. Martin<sup>2</sup>

<sup>1</sup>*School of Chemistry, University of Sydney, Sydney, New South Wales 2006, Australia*

<sup>2</sup>*Department of Chemistry, University of North Texas, Denton, Texas 76203-5017, USA*

(Received 19 January 2012; accepted 8 March 2012; published online 29 March 2012)

In an attempt to extend the applicability of the W1 and W2 *ab initio* computational thermochemistry methods, we propose explicitly correlated versions thereof, denoted W1-F12 and W2-F12. In W2-F12, we can “save” one cardinal number (viz., angular momentum) in the basis set sequences without loss in accuracy; in W1-F12, we can do so for first-row compounds but not for second-row compounds. At a root mean square deviation (RMSD) = 0.19 kcal/mol for the first-row molecules in the W4-11 benchmark dataset, W1-F12 is in fact superior to ordinary W1 theory. For the entire W4-11 set, W2-F12 yields a RMSD = 0.20 kcal/mol, comparable to 0.19 kcal/mol from ordinary W2 theory. The extended applicability ranges of W1-F12 and W2-F12 are not just due to the lower computational cost but also to greatly reduced memory and especially storage requirements. They are illustrated through applications to nucleic acids and to polyacenes (with up to four rings), for which the following revised gas-phase heats of formation are found:  $\Delta_f H_{298}^\circ = 19.6$  (benzene), 34.94 (naphthalene), 53.9, (anthracene), 73.9 (naphthacene/tetracene), 54.9 (adenine),  $-16.3$  (cytosine), 5.1 (guanine),  $-80.6$  (thymine), and  $-71.6$  (uracil) kcal/mol. Our theoretical values for the DNA/RNA bases largely confirm recent predictions based on much lower-level calculations. The W1-F12 theoretical values for benzene, naphthalene, and anthracene are in very good to reasonable agreement with experiment. However, both W1-F12 and other computational estimates show that the accepted experimental value for naphthacene cannot be reconciled with those for the lower acenes: we suggest that  $\Delta_f H_{298}^\circ[\text{naphthacene(g)}] = 74.25 \pm 1$  kcal/mol is a more realistic estimate.

© 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.3697678>]

### I. INTRODUCTION

Nowadays nonempirical, *ab initio*, composite procedures are widely used in accurate determination of thermochemical data. In particular, over the past decade such methods have been used for the generation of accurate thermochemical benchmark data to be used for the parametrization and testing of more approximate electronic structure methods, such as density functional theory (DFT) exchange-correlation functionals. For instance, about two thirds of the 30 reference datasets in the GMTKN30 general-purpose database of Goerigk and Grimme<sup>1</sup> were obtained by composite approaches that aim at reproducing the CCSD(T)/CBS energy (complete basis set limit coupled cluster with singles, doubles, and quasiperturbative triple excitations). Examples of such procedures include: (i) the Weizmann- $n$  theories ( $n = 1$  and  $2$ );<sup>2,3</sup> and (ii) the correlation consistent composite approach (ccCA).<sup>4,5</sup> For a recent comprehensive review of composite procedures see Peterson *et al.*,<sup>6</sup> other reviews include Refs. 7 and 8), while a paper by Feller *et al.*<sup>9</sup> on reaching the CCSD(T) basis set limit must also be mentioned.

For very small molecules (typically, systems including up to four non-hydrogen atoms with some symmetry) it is possible to calculate thermochemical properties close to the full configuration interaction (FCI) basis set limit using state of the art composite procedures such as Weizmann-4 (W4) theory,<sup>10,11</sup> the focal point analysis (FPA) approach,<sup>12,13</sup> and the high-accuracy extrapolated *ab initio* thermochemistry (HEAT) protocol.<sup>14–16</sup> The accuracies achieved by these

methods rival all but the most reliable experimental data, such as those obtained from the active thermochemical tables (ATcTs) thermochemical network developed by Ruscic and co-workers.<sup>17–19</sup> In a recent study<sup>20</sup> we have shown that for a set of 35 highly accurate ATcT total atomization energies (TAEs), all with associated experimental uncertainties equal or smaller than  $\pm 0.05$  kcal/mol, W4 affords a root mean square deviation (RMSD) of 0.085 kcal/mol and a mean absolute deviation (MAD) of 0.065 kcal/mol. These imply a 95% confidence interval of only 0.17 kcal/mol. Moreover, the largest deviations from experiment (e.g., for ozone) are still below the threshold of “benchmark accuracy” (arbitrarily defined as 1 kJ/mol = 0.24 kcal/mol).

For larger systems one has to resort to calculating thermochemical data at the CCSD(T) level. Nevertheless, in the absence of strong nondynamical correlation (a.k.a. “multireference character”), the CCSD(T)/CBS energy should, generally, be well within  $\sim 1$  kcal/mol from the FCI/CBS energy. For example, in the above-mentioned study<sup>20</sup> we have shown that for a subset of the 25 TAEs, of systems dominated by mild nondynamical correlation effects, W2 theory affords a RMSD and MAD of 0.25 and 0.18 kcal/mol, respectively (implying a respectable 95% confidence interval of about half a kcal/mol).

The ability of the original W1/2 theories to successfully reproduce the CCSD(T)/CBS energy relies on judicious combination of very large Gaussian basis sets with extrapolation techniques (see Sec. II B for specific details). In practice,

for systems consisting of more than 7–8 non-hydrogen atoms (with C1 symmetry), even the computationally more economical W1 theory becomes prohibitively expensive with current mainstream server hardware.

Coming to the subject of the present work, it is well established that the inclusion of terms that explicitly depend on the interelectronic distance into the wavefunction drastically accelerates the basis set convergence.<sup>21–23</sup> Experience with CCSD-F12 has shown that basis set convergence is faster than for conventional orbital calculations:<sup>24,25</sup> typically, the “gain” amounts to at least one angular momentum. This suggested to us the possibility that the use of the CCSD-F12 method might accelerate the basis set convergence up to the point where basis sets of no larger than *spdf* quality would be used in an “explicitly correlated” version of W1 theory. We shall show below that for systems containing first-row elements (and H) not only this is indeed the case, but the RMSD (relative to the CCSD(T) infinite basis set limit) is improved by more than 50%. However, for systems that contain second-row elements basis sets of up to *spdfg* quality still have to be employed, ostensibly due to the somewhat anemic character of the VDZ basis set. (We have in the past noted similar issues for the post-CCSD(T) contribution in W3.2lite.<sup>26</sup>)

Klopper *et al.*<sup>27</sup> have calculated the atomization energies of 106 first-row molecules using a nonempirical composite scheme (see also Ref. 28) in which the CCSD correlation energy was obtained using the explicitly correlated CCSD(F12) approach.<sup>29</sup> Including higher-order excitations (up to perturbative quadruple excitations) and other secondary contributions, such as, core-valence (CV), scalar relativistic, spin-orbit, and anharmonic zero-point vibrational energy (ZPVE) corrections, they obtained a RMSD and MAD of 0.4 and 0.3 kcal/mol, respectively, for a subset of 77 systems for which ATcT atomization energies are available. Using this approach Klopper and Haunschuld have also obtained reference values for the AE6 and BH6 test sets.<sup>30</sup> The authors note that an explicitly correlated version of the composite ccCA procedure has recently been developed by Mahler and Wilson.<sup>31</sup>

## II. COMPUTATIONAL DETAILS

### A. Hardware and software

Initial calculations were performed on the Martin group Linux cluster and on the Faculty of Chemistry HPC facility at the Weizmann Institute of Science, while the remainder were carried out on the cluster of the Center for Advanced Scientific Computing and Modeling (CASCaM) at the University of North Texas. The MOLPRO 2010.1 program suite was used throughout for the Hartree–Fock (HF) and coupled cluster calculations.<sup>32</sup> The geometry optimizations, harmonic frequency calculations, and some composite thermochemistry calculations on the DNA/RNA bases and polyacenes considered in the “applications” part of this work (Sec. III H) were performed using the GAUSSIAN 09 program suite.<sup>33</sup> The geometries of the five DNA/RNA bases and polyacenes were optimized at the B3LYP-D/Def2-TZVP level of theory<sup>34</sup> and are available as supplementary material.<sup>35</sup>

### B. Standard W1 and W2 calculations

W1 and W2 theories represent a layered extrapolation to the all-electron, relativistic CCSD(T) basis set limit energy, and can achieve an accuracy in the sub-kcal/mol range for molecules whose wave functions are dominated by dynamical correlation.<sup>20</sup> The computational protocols of W1 and W2 theories have been specified in great detail elsewhere.<sup>2,3</sup> In the present work the W1w and W2w variants are employed throughout.<sup>10</sup> In brief, in W2w the Hartree–Fock component is extrapolated from the A'VQZ and A'V5Z basis sets, using the  $E(L) = E_\infty + A/L^\alpha$  two-point extrapolation formula (with  $\alpha = 5$ ). Where the shorthand notation A'VnZ indicates the combination of the regular correlation consistent cc-pVnZ basis set on hydrogen,<sup>36</sup> aug-cc-pVnZ on first-row elements,<sup>37</sup> and aug-cc-pV(n+d)Z on second-row elements.<sup>38</sup> The valence CCSD correlation energy is obtained using the same basis sets using the aforementioned two-point extrapolation formula (with  $\alpha = 3$ ), while the (T) valence correlation energy is extrapolated (with  $\alpha = 3$ ) from the A'VTZ and A'VQZ basis sets. The CCSD(T) inner-shell contribution is obtained at the CCSD(T) level using the MTsmall basis set (where MTsmall denotes a completely uncontracted cc-pVTZ basis set with additional *2d1f* high-exponent functions<sup>2</sup>). As in the GAUSSIAN 09 implementation<sup>39</sup> of W1 theory, the scalar relativistic contribution is obtained from second-order Douglas–Kroll CCSD(T) calculations using the MTsmall basis set.<sup>40,41</sup> The more economical W1w theory uses smaller basis sets for the HF, CCSD, (T) contributions. In particular, the HF, CCSD, and (T) contributions are extrapolated from the A'V{T,Q}Z, A'V{T,Q}Z, and A'V{D,T}Z basis set pairs, respectively, using the two-point extrapolation formula with  $\alpha = 5, 3.22$ , and  $3.22$ , respectively.

### C. W4-11 reference data

The W4-11 database<sup>20</sup> is a set of 140 highly accurate total atomization energies of small first- and second-row molecules. The species in the W4-11 database cover a broad spectrum of bonding situations and multireference character, and as such it is an excellent benchmark set for the validation of approximate composite (and DFT) methods. In particular, the W4-11 dataset includes the following species: closed shell (97), singlet carbenes (9), radicals (27), and triplet systems (7). In terms of elemental composition the dataset includes 100 first-row species, 19 second-row species, and 21 mixed first- and second-row species. Fourteen of the species have %TAE<sub>e</sub>[(T)] diagnostics<sup>10</sup> of over 10%, indicating severe multireference character, while 25 more have %TAE<sub>e</sub>[(T)] diagnostics between 5% and 10%, indicating moderate such character (%TAE<sub>e</sub>[(T)] indicates the percentage of the total atomization energy accounted for by parenthetical connected triple excitations and has been shown to be a reliable energy-based diagnostic for the importance of nondynamical correlation effects, for further details see also Ref. 20).

The atomization energies in the W4-11 dataset were obtained by means of the W4 (or higher) thermochemical protocol. Weizmann-4 theory represents a layered extrapolation

to the all-electron, relativistic FCI infinite basis set limit energy, and has been specified and rationalized in great detail elsewhere.<sup>10,11,20</sup> In short, the SCF and valence CCSD components are extrapolated from the A'V5Z and A'V6Z basis sets. The valence (T) component is extrapolated from the A'VQZ and A'V5Z basis sets. The higher-order connected triples,  $\hat{T}_3$ -(T), valence correlation contribution is extrapolated from the cc-pVDZ and cc-pVTZ basis sets. The (Q) and  $\hat{T}_4$ -(Q) contributions are calculated with the cc-pVTZ and cc-pVDZ basis sets, respectively. The connected quintuples,  $\hat{T}_5$ , contribution is calculated with the *sp* part of the cc-pVDZ basis set. For a detailed, yet concise summary of the steps involved in the W4 computational protocol see reference.<sup>20</sup>

Finally, note that three systems (Be<sub>2</sub>, BeF<sub>2</sub>, and BeCl<sub>2</sub>) were not considered in the present work since the VnZ-F12 basis sets are not currently available for beryllium.

#### D. Explicitly correlated calculations

In all of the explicitly correlated coupled cluster calculations the diagonal, fixed-amplitude 3C(FIX) ansatz<sup>42-44</sup> and the CCSD-F12b approximation<sup>45,46</sup> are employed. (The CCSD-F12a/3C(FIX) approximation is also tested, however, it does not lend itself to effective basis set extrapolations.)

The so-called complementary auxiliary basis (CABS) singles correction<sup>44,45,47</sup> was added to the SCF energies in the explicitly correlated calculations. This correction allows the HF orbitals to be relaxed in the CABS space, and has been shown to substantially reduce the basis set error in the HF energy. The Hartree–Fock orbitals were relaxed within the frozen core approximation. In the present work, the resultant SCF energy is denoted by HF\*.

The quasiperturbative triples, (T), corrections were calculated in the same way as in standard CCSD(T) calculations (i.e., without inclusion of F12 terms). Scaling the (T) component by the factor  $f = \frac{E_{corr}^{MP2-F12}}{E_{corr}^{MP2}}$  has been shown to accelerate the basis set convergence. This rather simplistic approach assumes that  $f \approx \frac{E^{(T)-F12}}{E^{(T)}}$ , nevertheless, it has been shown to work rather well in most cases.<sup>46</sup> The scaled parenthetical contribution is denoted by (T\*).

In the explicitly correlated valence calculations the cc-pVnZ-F12 basis sets of Peterson *et al.*<sup>24</sup> ( $n = D, T, Q$ ), which were developed with explicitly correlated calculations in mind (in the paper, these are denoted by VnZ-F12). In the explicitly correlated core-valence calculations the core-valence cc-pCVnZ-F12 basis sets of Hill *et al.*<sup>48</sup> were employed. Optimal values for the geminal Slater exponents ( $\beta$ ) used in conjunction with the VnZ-F12 and CVnZ-F12 basis sets were taken from Refs. 24 and 48. The RI approximation was applied using the OptRI auxiliary basis sets of Yousaf and Peterson<sup>49</sup> within the CABS approach of Valeev.<sup>50</sup> The JKFIT fitting basis sets of Weigend<sup>51</sup> were employed for the density fitting in the Hartree–Fock calculations, while the MP2FIT set of Hättig and co-workers<sup>52</sup> was employed for the density fitting of the remaining two-electron integrals in the MP2-F12 and CCSD-F12 calculations.

### III. RESULTS AND DISCUSSION

In Secs. III A–III E we compare the basis set convergence of the SCF, valence CCSD, valence (T), core-valence, and scalar relativistic components of the TAEs in the W4-11 dataset in standard and explicitly correlated calculations. Section III F summarizes the final steps in explicitly correlated versions of W1 and W2 theories. Finally, in Sec. III H the heats of formation of the DNA and RNA bases are calculated as an illustrative example, followed by some conclusions offered in Sec. IV.

#### A. SCF component

Table I gives an overview of the basis set convergence of the HF/A'VnZ and HF\*/VnZ-F12 contributions to the TAEs in the W4-11 dataset. The reference values are HF/A'V{5,6}Z TAEs obtained from W4 theory (see Ref. 20 for further details). The RMSDs of the HF\*/VnZ-F12 calculations are smaller by 1–2 orders of magnitude than those obtained in the conventional HF/A'VnZ calculations with the same cardinal number  $n$  ( $n = D, T, \text{ and } Q$ ). Likewise, the RMSD obtained from the HF\*/V{D,T}Z-F12 extrapolation (0.04 kcal/mol) is 1 order of magnitude smaller than that obtained from the HF/A'V{D,T}Z extrapolation (0.43 kcal/mol). Quite remarkably, to reproduce the said RMSD obtained from the HF\*/V{D,T}Z-F12 extrapolation one has to extrapolate the conventional HF energy from the A'VQZ and A'V5Z basis sets. These results are consistent with the results reported by Knizia *et al.*<sup>46</sup> for a set of 104 reactions involving closed- and open-shell species.

Thus, it is evident that including the perturbative singles correction in the SCF energy using the CABS orbitals drastically accelerates the basis set convergence of the Hartree–Fock energy.<sup>45,47</sup> This fast basis set convergence obviates the need for SCF basis set extrapolations when an *spdfg* quality basis set is used. However, when basis sets of up to *spdfg* quality are used, basis set extrapolations still reduce RMSDs noticeably (e.g., RMSDs of 0.08 and 0.04 kcal/mol are obtained for the HF\*/VTZ-F12 and HF\*/V{D,T}Z-F12 levels of theory, respectively). Also note that the HF\* TAEs calculated in conjunction with the VDZ-F12 and VTZ-F12 basis sets systematically underestimate the basis set limit SCF TAEs (as is evident from MSD =  $-1 \times \text{MAD}$ ). However, for the HF\*/VQZ-F12 and HF\*/V{D,T}Z-F12 levels of theory the MSDs are essentially zero.

Finally, we note that the VTZ-F12 and VQZ-F12 basis sets (as well as the V{D,T}Z-F12 extrapolations) show monotonous performance for both first- and second-row species. However, for a number of second-row systems large deviations are obtained with the smaller VDZ-F12 basis set, in particular large underestimations ranging between 1 and 2 kcal/mol are found for S<sub>2</sub>O, SO<sub>2</sub>, SO<sub>3</sub>, S<sub>3</sub>, S<sub>4</sub>, P<sub>4</sub>, OClO, and SiF<sub>4</sub>, see Table SI of the supplementary material<sup>35</sup> for the specific deviations). Consequently, the RMSD obtained for the first-row species (0.46 kcal/mol) is significantly smaller than that obtained for the systems including second-row elements (0.71 kcal/mol).

TABLE I. Statistical analysis of the basis set convergence in conventional and CABS-corrected SCF calculations carried out in conjunction with the  $A'VnZ$  and  $VnZ$ -F12 basis sets, respectively, for the TAEs in the W4-11 test set (in kcal/mol).<sup>a,b</sup>

Basis set	$\alpha^d$	Everything (137 systems)			First-row (97 systems)			Second-row <sup>c</sup> (40 systems)		
		RMSD	MAD	MSD	RMSD	MAD	MSD	RMSD	MAD	MSD
Conventional orbital calculations										
A'VDZ		8.80	7.39	-7.39	8.01	7.18	-7.18	10.47	7.88	-7.88
A'VTZ		1.17	0.97	-0.97	0.94	0.85	-0.85	1.59	1.26	-1.26
A'VQZ		0.26	0.17	-0.17	0.13	0.11	-0.11	0.43	0.31	-0.31
A'V5Z		0.09	0.05	-0.05	0.03	0.02	-0.02	0.16	0.12	-0.12
A'V{D,T}Z	5.20	0.43	0.30	-0.09	0.39	0.27	0.02	0.52	0.38	-0.35
A'V{T,Q}Z	6.10	0.15	0.09	0.00	0.08	0.06	0.05	0.26	0.16	-0.11
A'V{Q,5}Z	5.00	0.04	0.03	0.00	0.03	0.02	0.02	0.06	0.04	-0.02
Explicitly correlated calculations <sup>e</sup>										
VDZ-F12		0.55	0.45	-0.45	0.46	0.41	-0.41	0.71	0.53	-0.53
VTZ-F12		0.08	0.06	-0.06	0.07	0.06	-0.06	0.10	0.07	-0.07
VQZ-F12		0.02	0.01	0.00	0.01	0.01	0.00	0.03	0.02	0.00
V{D,T}Z-F12 <sup>f</sup>	5.00	0.04	0.03	0.00	0.03	0.03	0.00	0.04	0.03	0.00

<sup>a</sup>RMSD: root mean square deviation; MSD: mean signed deviation; MAD: mean absolute deviation.

<sup>b</sup>Reference values at the basis set limit are taken from the W4-11 database, (Ref. 20), i.e., they were obtained at the W4 (or higher) level.

<sup>c</sup>System including at least one second-row element.

<sup>d</sup> $\alpha$  is the extrapolation exponent used in the two point extrapolations, unless otherwise noted it is optimized to minimize the RMSD over the entire W4-11 dataset.

<sup>e</sup>The perturbative CABS singles correction is included in the SCF energies.

<sup>f</sup>Note that extrapolating the SCF component from the V{T,Q}Z-F12 basis set pair does not improve on the VQZ-F12 results.

## B. Valence CCSD component

Table II gives the statistical analysis for the CCSD correlation energy obtained from explicitly correlated and standard calculations (the deviations for each molecule are given in Table SII of the supplementary material<sup>35</sup>). The reference values are CCSD/A'V{5,6}Z TAEs obtained from W4 theory (see Ref. 20 for further details). For starters, the RMSDs over the entire W4-11 test set obtained in conjunction with the VDZ-F12, VTZ-F12, and VQZ-F12 basis sets are 2.45, 0.71, and 0.18 kcal/mol, respectively. Compared to standard

CCSD/A'VnZ calculations with the corresponding cardinal number ( $n = D, T, Q$ ), these RMSDs are smaller by 0.8, 1.1, and 1.6 orders of magnitude, respectively.

Inspection of the deviations for systems containing second-row elements (see Table SII (Ref. 35)) reveals that the VDZ-F12 basis set behaves in an erratic manner. This is also indicated by a MSD of  $-0.16$  kcal/mol (rather than  $\text{MSD} \approx -1 \times \text{MAD} = -1.33$  kcal/mol, see Table II). Thus, the VDZ-F12 basis set does not lend itself for basis set extrapolations of second-row systems. Nevertheless, for the first-row

TABLE II. Statistical analysis of the basis set convergence in conventional and explicitly correlated CCSD calculations carried out in conjunction with the  $A'VnZ$  and  $VnZ$ -F12 basis sets, respectively, for the TAEs in the W4-11 test set (in kcal/mol).<sup>a</sup>

Basis set	$\alpha$	Everything (137 systems)			First-row (97 systems)			Second-row (40 systems)		
		RMSD	MAD	MSD	RMSD	MAD	MSD	RMSD	MAD	MSD
A'VDZ		19.18	17.01	-17.01	20.45	18.46	-18.46	15.71	13.48	-13.48
A'VTZ		7.65	6.86	-6.86	7.69	7.03	-7.03	7.55	6.45	-6.45
A'VQZ		2.88	2.57	-2.57	2.86	2.61	-2.61	2.95	2.48	-2.48
A'V5Z		1.41	1.25	-1.25	1.43	1.30	-1.30	1.36	1.14	-1.14
A'V{D,T}Z	2.38	2.18	1.64	-0.62	1.58	1.27	0.00	3.21	2.54	-2.13
A'V{T,Q}Z	3.46	0.28	0.21	-0.05	0.20	0.16	-0.01	0.43	0.34	-0.15
A'V{Q,5}Z	3.33	0.14	0.12	-0.06	0.15	0.12	-0.12	0.14	0.10	0.07
VDZ-F12		2.45	2.10	-1.76	2.70	2.42	-2.41	1.71	1.33	-0.16
VTZ-F12		0.71	0.59	-0.58	0.70	0.62	-0.61	0.72	0.53	-0.51
VQZ-F12		0.18	0.13	-0.10	0.17	0.14	-0.11	0.19	0.12	-0.08
V{D,T}Z-F12	3.67	0.45	0.28	-0.24	0.17	0.14	-0.09	0.80	0.61	-0.61
V{D,T}Z-F12	3.38 <sup>b</sup>	0.46	0.25	-0.18	0.13	0.10	0.00	0.83	0.63	-0.62
V{T,Q}Z-F12	5.94	0.13	0.10	0.01	0.11	0.09	0.01	0.15	0.12	0.02

<sup>a</sup>Footnotes a, b, c, and d to Table I apply.

<sup>b</sup>The extrapolation exponent is optimized over the first-row systems only.

TABLE III. Statistical analysis of the basis set convergence in conventional and explicitly correlated (T) calculations carried out in conjunction with the A'VnZ and VnZ-F12 basis sets, respectively, for the TAEs in the W4-11 test set (in kcal/mol).<sup>a</sup>

Basis set	$\alpha$	Everything (137 systems)			First-row (97 systems)			Second-row (40 systems)		
		RMSD	MAD	MSD	RMSD	MAD	MSD	RMSD	MAD	MSD
A'VDZ		3.03	2.68	-2.68	3.17	2.87	-2.87	2.68	2.20	-2.20
A'VTZ		0.83	0.73	-0.73	0.81	0.74	-0.74	0.87	0.71	-0.71
A'VQZ		0.36	0.31	-0.31	0.34	0.30	-0.30	0.41	0.33	-0.33
A'V{D,T}Z	3.33	0.17	0.07	-0.05	0.04	0.03	0.01	0.30	0.19	-0.19
A'V{T,Q}Z	2.93	0.05	0.03	0.00	0.03	0.03	0.03	0.07	0.05	-0.05
VDZ-F12		0.40	0.27	0.19	0.26	0.19	0.10	0.62	0.48	0.40
VTZ-F12		0.18	0.13	0.13	0.17	0.13	0.12	0.20	0.14	0.13
VQZ-F12		0.10	0.08	0.08	0.10	0.08	0.08	0.11	0.08	0.07
V{D,T}Z-F12	3.12	0.15	0.12	0.10	0.16	0.13	0.13	0.10	0.08	0.03
V{T,Q}Z-F12	2.14	0.04	0.03	0.02	0.04	0.03	0.03	0.04	0.03	-0.01
VDZ-F12 <sup>b</sup>		0.30	0.22	0.00	0.21	0.16	-0.09	0.44	0.35	0.22
VTZ-F12 <sup>b</sup>		0.08	0.06	0.00	0.07	0.05	-0.01	0.10	0.08	0.02
VQZ-F12 <sup>b</sup>		0.04	0.03	0.00	0.03	0.02	0.00	0.05	0.03	0.00
V{D,T}Z-F12 <sup>c</sup>	4.54	0.06	0.04	0.00	0.05	0.04	0.01	0.08	0.05	-0.02
V{T,Q}Z-F12 <sup>c</sup>	3.87	0.02	0.02	0.00	0.02	0.02	0.00	0.03	0.02	-0.01

<sup>a</sup>Footnotes *a*, *b*, *c*, and *d* to Table I apply.

<sup>b</sup>Scaling the (T\*)/VnZ-F12 corrections by 0.981, 0.987, and 0.992 for *n* = D, T, and Q, respectively, see text.

<sup>c</sup>Extrapolating the scaled (T\*) corrections.

systems extrapolation from the VDZ-F12 and VTZ-F12 basis sets results in a spectacularly small RMSD of only 0.17 kcal/mol. This RMSD is identical to that obtained with the VQZ-F12 basis set, and is also on par with that obtained from the CCSD/A'V{Q,5}Z extrapolations in conventional CCSD calculations (0.15 kcal/mol).

The VTZ-F12 and VQZ-F12 basis sets behave in a consistent manner for both first- and second-row systems. Extrapolation from the VTZ-F12 and VQZ-F12 basis sets results in RMSDs of 0.11 and 0.15 kcal/mol for the first- and second-row systems, respectively. The RMSD over the entire W4-11 test set (0.13 kcal/mol) is on par with that obtained from CCSD/A'V{Q,5}Z extrapolations in conventional CCSD calculations (0.14 kcal/mol).

### C. Valence (T) component

Table III presents the error statistics for the standard “parentetical” triples, (T), and the “parentetical” triples scaled by the explicit correlation factor  $f = \frac{E_{cor}^{MP2-F12}}{E_{cor}^{MP2}}$ , (T\*). The deviations for each molecule are given as supplementary material<sup>35</sup> (Table SIII). The reference values are basis set limit (T) contributions obtained from W4 theory (or higher), i.e., extrapolated from the A'VQZ and A'V5Z (or larger) basis sets.<sup>20</sup> The first thing that can be noted in Table III is that the standard (T) component systematically converges to the basis set limit from below, whereas the (T\*) component tends to converge to the basis set limit from above (see also Table SIII (Ref. 35)). For the (T\*) component, RMSDs of 0.40, 0.18, and 0.10 kcal/mol are obtained in conjunction with the VDZ-F12, VTZ-F12, and VQZ-F12 basis sets, respectively. For comparison, RMSDs of 3.03, 0.83, and 0.36 kcal/mol are obtained in the standard (T)/A'VnZ calculations with *n* = D, T, and Q, respectively. However, when comparing the

extrapolated results the advantages of scaling are no longer apparent. Specifically, RMSDs of 0.15 and 0.17 kcal/mol are obtained for the (T\*)/V{D,T}Z-F12 and (T)/A'V{D,T}Z extrapolations, respectively. Likewise, the (T\*)/V{T,Q}Z-F12 and (T)/A'V{T,Q}Z extrapolations result in nearly identical RMSDs (0.04 and 0.05 kcal/mol, respectively). Nevertheless, it should be noted that for systems containing second-row elements, the (T\*)/V{D,T}Z-F12 extrapolation performs much better than the (T)/A'V{D,T}Z extrapolation (RMSD = 0.10 and 0.30 kcal/mol, respectively), whereas for first-row systems (T)/A'V{D,T}Z extrapolation outperforms the (T\*)/V{D,T}Z-F12 extrapolation (RMSD = 0.04 and 0.16 kcal/mol, respectively).

Finally, one can consider scaling the (T\*) contribution by an empirical factor which is optimized to minimize the RMSD over the entire W4-11 dataset. The optimized scaling factors obtained (see the lower pane of Table III are 0.981, 0.987, and 0.992 for *n* = D, T, and Q, respectively, which appear to indicate that on balance, the “explicit correlation factor” overcorrects for basis set incompleteness in (T). The rescaling reduces RMSDs by 25–60%. In particular, note that the RMSD obtained for 0.987 × (T\*)/VTZ-F12 (0.08 kcal/mol) is about half of that obtained from the (T\*)/V{D,T}Z-F12 extrapolation (0.15 kcal/mol), where essentially all of the improvement is due to the better performance for the first-row systems.

### D. Core-valence correction

Table IV gives the error statistics for the CV contributions to the TAEs. These are broken down to the CCSD and (T) components in the standard calculations, and to the CCSD-F12 and (T\*) components in explicitly correlated calculations. The reference CV corrections are obtained from W4 theory (i.e., calculated at the CCSD(T)/A'PWCV{T,Q}Z level

TABLE IV. Statistical analysis of the basis set convergence of the CCSD and (T) components (in standard calculations), and of the CCSD-F12 and (T\*) components (in explicitly correlated calculations) of the core-valence contribution to the TAEs (in kcal/mol).<sup>a</sup>

	Basis set	Everything (137 systems)			First-row (97 systems)			Second-row (40 systems)		
		RMSD	MAD	MSD	RMSD	MAD	MSD	RMSD	MAD	MSD
CCSD	A'PWCVDZ	0.50	0.38	-0.07	0.49	0.38	-0.10	0.52	0.39	0.02
	A'PWCVTZ(no <i>f</i> )	0.39	0.27	-0.25	0.23	0.18	-0.16	0.63	0.48	-0.47
	PWCVTZ	0.20	0.15	-0.10	0.18	0.14	-0.12	0.26	0.18	-0.06
	1.1×PWCVTZ	0.15	0.08	-0.03	0.06	0.05	-0.02	0.27	0.17	-0.05
	A'PWCVTZ	0.17	0.13	-0.05	0.15	0.12	-0.08	0.20	0.15	0.02
	1.1×A'PWCVTZ	0.12	0.08	0.02	0.06	0.04	0.02	0.21	0.16	0.03
	A'PWCVQZ	0.07	0.06	-0.02	0.06	0.05	-0.04	0.08	0.06	0.01
	MTsmall(no <i>f</i> )	0.41	0.27	-0.26	0.20	0.17	-0.15	0.69	0.51	-0.51
	MTsmall(1 <i>f</i> )	0.31	0.20	-0.18	0.16	0.13	-0.11	0.52	0.35	-0.33
Mtsmall	0.15	0.09	-0.04	0.06	0.05	-0.02	0.26	0.17	-0.10	
CCSD-F12	CVDZ-F12	0.20	0.16	-0.08	0.21	0.18	-0.13	0.19	0.12	0.04
	CVTZ-F12	0.08	0.06	-0.03	0.07	0.06	-0.05	0.09	0.07	0.00
	CVQZ-F12	0.05	0.03	0.00	0.03	0.03	0.00	0.07	0.05	0.00
(T)	A'PWCVDZ	0.19	0.16	-0.15	0.17	0.15	-0.15	0.22	0.18	-0.15
	PWCVTZ(no <i>f</i> )	0.06	0.05	-0.04	0.04	0.04	-0.04	0.08	0.06	-0.06
	A'PWCVTZ(no <i>f</i> )	0.06	0.05	-0.05	0.05	0.04	-0.04	0.08	0.06	-0.06
	A'PWCVTZ	0.06	0.05	-0.05	0.05	0.04	-0.04	0.08	0.06	-0.06
	A'PWCVQZ	0.02	0.02	-0.02	0.02	0.02	-0.02	0.03	0.03	-0.03
	MTsmall(no <i>f</i> )	0.05	0.04	-0.04	0.04	0.03	-0.03	0.08	0.06	-0.05
	MTsmall(1 <i>f</i> )	0.05	0.04	-0.04	0.04	0.03	-0.03	0.08	0.06	-0.05
	Mtsmall	0.05	0.04	-0.03	0.03	0.03	-0.03	0.07	0.05	-0.05
(T*)	CVDZ-F12	0.24	0.11	0.09	0.07	0.04	0.02	0.43	0.28	0.28
	CVTZ-F12	0.20	0.11	0.11	0.07	0.05	0.04	0.36	0.27	0.27
	CVQZ-F12	0.23	0.13	0.13	0.08	0.06	0.06	0.41	0.30	0.30

<sup>a</sup>Footnotes *a*, *b*, and *c* to Table I apply.

of theory). Beginning with the standard CCSD calculations, the MTsmall and A'PWCVQZ basis sets give satisfactory results for the first-row systems (RMSD = 0.06 kcal/mol), whereas for second-row systems only the A'PWCVQZ basis set achieves this goal. For large systems, such as those considered in Sec. III H, the CV step can quickly become the bottleneck in terms of CPU time even for first-row compounds. Therefore, any reduction in the basis set size for this step is highly desirable. We note that the PWCVTZ basis set results in a RMSD of 0.18 kcal/mol over the first-row systems. However, scaling the CCSD/PWCVTZ CV component by 1.1 reduces the RMSD to only 0.06 kcal/mol. The (T) component of the CV correction converges faster to the basis set limit, such that both the MTsmall(no *f*) and PWCVTZ(no *f*) basis sets yield sufficiently small RMSDs (specifically, of 0.04 kcal/mol for first-row systems and 0.08 kcal/mol for second-row systems).

Moving on to the explicitly correlated CCSD-F12 calculations, the CVTZ-F12 basis set performs acceptably well for both first- and second-row systems, with RMSDs of 0.07 and 0.09 kcal/mol, respectively. The (T\*)/CV $n$ Z-F12 component, however, behaves rather uniformly for  $n = D, T,$  and  $Q$ . Moreover, it seems to work fairly well for first-row systems, but not for systems containing second-row elements. Note that MAD/RMSD ratios of 0.65–0.75 obtained for the second-row systems, indicate an error distribution that is not strongly dominated by a small number of outliers.<sup>53</sup> Inspection of the

errors (see Table SV of the supplementary material<sup>35</sup>) reveals that the largest errors (of 0.6–1.4 kcal/mol) are obtained for CS<sub>2</sub>, S<sub>3</sub>, P<sub>4</sub>, and S<sub>4</sub> (for all the basis sets). After elimination of these four species from the error statistics, the RMSDs are still unacceptably large (0.26–0.29 kcal/mol).

## E. Scalar relativistic

The scalar relativistic contribution (in the second-order Douglas–Kroll–Hess approximation<sup>40,41</sup>) is obtained as the difference between non-relativistic CCSD(T)/A'V $n$ Z and relativistic CCSD(T)/A'V $n$ Z-DK calculations.<sup>54</sup> We compare our results to the scalar relativistic correction from W4 theory (i.e., obtained with  $n = Q$ ). For first-row systems exceptionally good results are obtained even with double-zeta basis sets. Specifically, a RMSD of 0.01 kcal/mol is obtained, and with the exception of one outlier (the formyl radical for which a deviation of 0.11 kcal/mol is obtained) all the deviations are smaller or equal to 0.02 kcal/mol. For systems containing second-row elements a RMSD of 0.06 kcal/mol is obtained. However, significant underestimations are obtained for a number of systems, namely by: 0.17 (P<sub>4</sub> and SO<sub>3</sub>), 0.16 (SO<sub>2</sub>), 0.11 (S<sub>2</sub>O and AlCl<sub>3</sub>), and 0.08 kcal/mol (CS<sub>2</sub> and SO). When the triple-zeta basis set is used all these errors are brought to below 0.03 kcal/mol, and a RMSD of 0.01 kcal/mol is obtained for both the first- and second-row species.

TABLE V. Overall error statistics for the explicitly correlated W1-F12 and W2-F12 protocols as well as for the standard W1w and W2w protocols over the TAEs in the W4-11 dataset (in kcal/mol).<sup>a</sup>

	Everything (137 systems)			First-row (97 systems)			Second-row (40 systems)		
	RMSD	MAD	MSD	RMSD	MAD	MSD	RMSD	MAD	MSD
	Non-relativistic, valence CCSD(T) <sup>b</sup>								
W1w	0.58	0.49	0.32	0.53	0.48	0.48	0.67	0.52	-0.04
W2w	0.17	0.14	0.12	0.13	0.11	0.10	0.25	0.20	0.19
W1-F12	0.58	0.25	-0.24	0.15	0.10	-0.03	1.05	0.60	-0.76
W2-F12	0.17	0.13	0.01	0.16	0.12	0.00	0.18	0.14	0.04
	Relativistic, all-electron CCSD(T) <sup>c</sup>								
W1w	0.51	0.43	0.25	0.48	0.43	0.43	0.58	0.43	-0.18
W2w	0.19	0.12	0.05	0.09	0.07	0.05	0.32	0.25	0.04
W1-F12	0.65	0.32	-0.33	0.19	0.13	-0.09	1.16	0.76	-0.90
W2-F12	0.20	0.15	-0.01	0.18	0.14	-0.02	0.24	0.18	0.03
	Relativistic, all-electron CCSD(T) <sup>d</sup>								
W1w	0.63	0.48	0.11	0.57	0.44	0.25	0.74	0.58	-0.25
W2w	0.42	0.26	-0.14	0.43	0.26	-0.15	0.37	0.28	-0.13
W3w <sup>e</sup>	0.19	0.14	0.01	0.13	0.10	0.06	0.30	0.24	-0.14
W1-F12	0.74	0.51	-0.50	0.45	0.32	-0.31	1.19	0.99	-0.99
W2-F12	0.42	0.30	-0.21	0.44	0.31	-0.24	0.35	0.27	-0.14
W3-F12 <sup>e</sup>	0.27	0.20	-0.07	0.21	0.17	-0.03	0.37	0.26	-0.15

<sup>a</sup>Footnotes *a*, *b*, and *c* to Table I apply.

<sup>b</sup>Bottom of the well, non-relativistic, valence CCSD(T) atomization energies (i.e., considering only the SCF, valence CCSD, and valence (T) contributions in W1w, W1-F12, and in the reference values from the W4-11 dataset).

<sup>c</sup>Bottom of the well, relativistic, all-electron CCSD(T) atomization energies (i.e., considering the SCF, valence CCSD, valence (T), core-valence, relativistic, spin-orbit, and DBOC contributions in W1w, W1-F12, and in the reference values from the W4-11 dataset).

<sup>d</sup>Same as footnote *c*, but here the reference values are bottom of the well, relativistic, all-electron FCI atomization energies from the W4-11 dataset.

<sup>e</sup>Adding the post-CCSD(T) contributions from W3.2 theory to the final W2w and W2-F12 energies.

## F. Explicitly correlated thermochemical protocol

The results presented in Secs. III A–III E can be incorporated into two explicitly correlated thermochemical protocols (labeled W1-F12 and W2-F12). W1-F12 is designed with an eye towards applicability to larger first-row systems, whereas W2-F12 should yield satisfactory results for both first- and second-row species.

1. *Reference geometry and ZPVE*: In both variants are obtained at the B3LYP/cc-pV(T+d)Z level of theory (where the ZPVE contribution is scaled by 0.985, as is done in the original W1 theory<sup>2</sup>).
2. *SCF energy*: Extrapolated from the HF\*/VDZ-F12 and HF\*/VTZ-F12 energies with an exponent of 5.0 in W1-F12, and calculated at the HF\*/VQZ-F12 level of theory in W2-F12.
3. *Valence CCSD*: In W1-F12 extrapolated from the V{D,T}Z-F12 basis set pair with an exponent of 3.67, and in W2-F12 extrapolated from the V{T,Q}Z-F12 basis set pair with an exponent of 5.94.
4. *Valence (T)*: In W1-F12 the (T) contribution from W1w theory is used (i.e., extrapolated from the A'V{D,T}Z basis set pair using the two-point extrapolation formula with  $\alpha = 3.22$ , see Sec. II B). In W2-F12 it is taken as the (T\*)/VTZ-F12 correlation energy scaled by 0.987.
5. *Inner-shell*: In both variants the CCSD and (T) components are obtained from standard calculations. The CCSD component is calculated with the PWCVTZ ba-

sis set (in W1-F12) and A'PWCVTZ basis sets (in W2-F12), where in both cases the CCSD component is scaled by 1.1. The (T) component is calculated with the PWCVTZ(no *f*) basis set in both W1-F12 and W2-F12.

6. *Scalar relativistic correction*: Obtained as the difference between non-relativistic CCSD(T)/A'VnZ and relativistic CCSD(T)/A'VnZ-DK calculations (with  $n = D$  and  $T$  in W1-F12 and W2-F12, respectively).
7. *Spin-orbit and DBOC*: In both variants these are obtained in the same way as in W2.2 theory.<sup>10</sup> Namely, the spin-orbit coupling terms are taken from the experimental fine structure, and the diagonal Born–Oppenheimer correction is calculated at the HF/A'VTZ level of theory.

Table V gives the error statistics for the explicitly correlated thermochemical protocols as well as for W1w and W2w theories<sup>10</sup> (the individual errors can be found in Table SVI of the supplementary material<sup>35</sup>). Let us start by considering the performance for the valence, non-relativistic, CCSD(T) TAEs (i.e., only the SCF, valence CCSD, and valence (T) components). These results are given in the upper pane of Table V. For the first-row systems, W1-F12 theory results in a RMSDs of 0.15 kcal/mol, which represents an improvement of more than 50% over W1w theory (which attains a RMSD of 0.53 kcal/mol). W2-F12 results in a RMSD of 0.16 kcal/mol; however, the extra computational cost of this method does not seem to be warranted for systems containing only first-row elements. For the second-row systems, W1-F12 yields an unacceptably large RMSD of 1.05 kcal/mol: the majority of the



error can be traced to the poor performance of the  $V\{D,T\}Z$ -F12 extrapolations for the second-row species. This source of error is eliminated in W2-F12, which attains RMSDs of 0.18 and 0.17 kcal/mol for the second-row species and for the entire set, respectively.

The middle pane of Table V depicts the error statistics for the W1-F12, W2-F12, W1w, and W2w protocols, with all the components included (except for the ZPVE). The reference values are relativistic, all-electron CCSD(T) basis set limit TAEs from the W4-11 database. For the first-row systems the RMSDs are very similar to those obtained for the valence, non-relativistic CCSD(T) energies (specifically, they are: 0.19, 0.18, 0.48, and 0.09 kcal/mol for W1-F12, W2-F12, W1w, and W2w, respectively). For the second-row species W2-F12 attains a RMSD of 0.24 kcal/mol. However, this RMSD is still much lower than that obtained for W1w (0.58 kcal/mol) and slightly lower than that obtained for W2w (0.32 kcal/mol). Over the entire set W2-F12 gives a RMSD of 0.20, for comparison W1w and W2w give RMSDs of 0.51 and 0.19 kcal/mol, respectively.

So far we have evaluated the performance of W1-F12, W2-F12, W1w, and W2w against reference values at the CCSD(T) infinite basis set limit. However, it is also of interest to see what sort of deviations are obtained against reference data at the FCI basis set limit. Essentially, this is equivalent to evaluating the performance of the  $W_n$  procedures against experimental reference values. These error statistics (against relativistic, all-electron FCI basis set limit values from the W4-11 database) are presented in the lower pane of Table V. For the purpose of this evaluation we have eliminated 13 pathologically multi-reference systems from the W4-11 dataset.<sup>55</sup> For the first-row systems W1-F12, W2-F12, W1w, and W2w attain RMSDs of 0.45, 0.44, 0.57, and 0.43 kcal/mol, respectively. For the second-row systems, W2-F12, W1w, and W2w result in RMSDs of 0.35, 0.74, and 0.37 kcal/mol, respectively. For the entire W4-11 set, RMSDs of 0.42, 0.63, and 0.42 kcal/mol are obtained for W2-F12, W1w, and W2w, respectively. If we include the post-CCSD(T) corrections obtained from W3.2 theory<sup>10</sup> to W2-F12 and W2w (these are labeled W3w and W3-F12 in Table V) then RMSDs of 0.27 and 0.19 kcal/mol are obtained for the entire W4-11 set, respectively. We note that if the CCSD-F12 component in W3-F12 is replaced with the CCSD/A/V{Q,5}Z component from W3w then the RMSD of W3-F12 goes down to 0.17 kcal/mol. Thus, the inferior performance of W3-F12 relative to W3w can be largely attributed to the performance of the CCSD-F12 component.

### G. Computational efficiency

It is instructive to compare the cpu times required to run a W1-F12 and a W1w calculation for a relatively large molecule: let us, for example, consider uracil. Assuming that in both cases the core-valence correction is obtained at the same level of theory. The cpu times for calculating the valence CCSD(T)/VDZ-F12 and CCSD(T)/VTZ-F12 energies in W1-F12 are 14 and 109 minutes, respectively, the corresponding cpu times for calculating the valence CCSD(T)/A/VTZ and CCSD/A/VQZ energies in W1w are 52 and 958 min, respec-

tively. (Both calculations ran on 16 Intel Nehalem 8837 cores at 2.67 GHz.) Furthermore, the W1-F12 and W1w calculations used 81.7 and 357.2 GB of scratch disk, respectively, and with very large calculations, disk space requirements often become the narrowest bottleneck, unless solid state storage is procured at (as of this writing still) ruinous expense. Thus, W1-F12 offers a substantial price-to-performance improvement and, because of the reduction in other resource requirements, a substantially increased applicability range.

For the aromatic hydrocarbons the valence CCSD(T)-F12 steps in W1-F12 used the following computational resources (cpu time, disk space): benzene (6 min, 12.1 GB), naphthalene (68 min, 76.5 GB), and anthracene (440 min, 247.4 GB).

### H. Some illustrative examples

In this section, we apply W1-F12 and W2-F12 to the DNA/RNA bases: adenine (A), cytosine (C), guanine (G), thymine (T), and Uracil (U), as well as to polyacenes (with up to four rings). Dorofeeva and Vogt (DV) (Ref. 56) recently reviewed the available experimental data and carried out G3X calculations;<sup>57</sup> the main problem with the available experimental data is reliable estimation of the phase transition enthalpies required to convert them to the gas phase. Dorofeeva and Vogt proposed revised values based on a combination of theory and experiment that are substantial upward revisions from the accepted wisdom. These systems are (by high-accuracy *ab initio* standards) rather large, the largest purine (G) containing 11 nonhydrogen atoms without spatial symmetry. Therefore, only the said G3X investigation has hitherto been reported. In Table VI we report W1-F12 and W2-F12 values (only W1-F12 for guanine), as well as their component breakdown and additional heats of formation from empirical thermochemical procedures. The HF\* contribution to the TAEs are practically indistinguishable between W1-F12 and W2-F12, which is consistent with our findings in Sec. III A. The CCSD-F12 contributions for cytosine, thymine, and uracil are very similar between W1-F12 and W2-F12 (they differ by 0.04, 0.09, and 0.15 kcal/mol, respectively); however, for adenine they differ by 0.4 kcal/mol (see Table VI). The “parentetical” triples contributions range between 37.05 (in U) to 54.64 kcal/mol (in G). As the %TAE<sub>e</sub>[(T)] diagnostics of: 3.0 (A), 2.7 (C), 3.1 (G), 2.5 (T), and 2.8% (U) only indicate mild nondynamical correlation effects, we can expect<sup>10,11,20</sup> good error cancellation between higher-order triple excitations,  $\hat{T}_3$ -(T) (which almost universally decrease TAEs (Refs. 10 and 58) and connected quadruple excitations,  $\hat{T}_4$ , (which universally increase TAEs (Refs. 10 and 58)). We note that similar %TAE<sub>e</sub>[(T)] values are obtained for the considered aromatic hydrocarbons: benzene (2.0%), naphthalene (2.2%), anthracene (2.3%), and naphthacene a.k.a. tetracene (2.4%).

Summing up all the contributions and including a ZPVE from a B3LYP-D/Def2-TZVP harmonic calculation (scaled by 0.985), the following atomization energies are obtained at 0 K: A 1609.59, C 1345.65, G 1717.75, T 1577.40, and U 1296.38 kcal/mol. Converting these to heats of formation at 298 K (see Table VI) using RRHO thermodynamic functions,

TABLE VI. Component breakdown of the  $W_n$ -F12 ( $n = 1$  and  $2$ ) total atomization energies (TAE) for the DNA/RNA bases (A, C, G, T, and U), benzene, naphthalene, anthracene, and naphthacene (in kcal/mol).

Component	$W_n$ -F12	A	C	G	T	U	$C_6H_6$	$C_{10}H_8$	$C_{14}H_{10}$	$C_{18}H_{12}$
SCF/V{D,T}Z-F12	1	1112.03	957.72	1175.64	1155.64	926.16	1045.18	1643.02	2235.72	2825.85
SCF/VQZ-F12	2	1112.03	957.73		1155.64	926.15				
CCSD/V{D,T}Z-F12	1	509.51	405.65	553.06	446.30	382.57	290.67	464.58	639.47	815.03
CCSD/V{T,Q}Z-F12	2	509.11	405.61		446.41	382.72				
$0.987 \times (T^*)^a$	1 and 2	50.74	38.06	54.64	41.09	37.05	26.69	46.78	67.31	88.12
Core-valence <sup>b</sup>	1 and 2	8.62	6.64	8.93	7.46	6.29	6.97	11.66	16.32	21.56
Relativity	1 and 2	-2.02	-1.66	-2.30	-1.83	-1.63	-0.99	-1.66	-2.32	-2.98
Spin-orbit	1 and 2	-0.42	-0.56	-0.65	-0.87	-0.78	-0.51	-0.85	-1.18	-1.52
DBOC <sup>c</sup>	1 and 2	0.40	0.36	0.45	0.38	0.32	0.15	0.24	0.35	0.44
ZPVE <sup>d</sup>	1 and 2	68.86	60.53	72.03	70.88	53.73	62.08	91.14	120.01	148.79
TAE <sub>e</sub>	1	1678.84	1406.22	1789.78	1648.17	1349.98	1368.17	2163.77	2955.67	3746.50
	2	1678.44	1406.18		1648.28	1350.11				
TAE <sub>0</sub>	1	1609.98	1345.69	1717.75	1577.29	1296.24	1306.09	2072.63	2835.66	3597.71
	2	1609.59	1345.65		1577.40	1296.38				
$\Delta H_{f,0K}^\circ$ <sup>e</sup>	1	60.65	-11.02	11.87	-74.44	-66.68	23.86	40.67	61.00	82.33
	2	61.05	-10.98		-74.55	-66.81				
$\Delta H_{f,298K}^\circ$ <sup>f</sup>	1	54.53	-16.30	5.09	-80.47	-71.43	19.64	34.94	53.88	73.86
	2	54.93	-16.26		-80.57	-71.56				
					$\Delta H_{f,298K}^\circ$ from other composite procedures <sup>g</sup>					
CBS-QB3		53.6	-16.5	3.4	-82.1	-73.0	21.4	37.7	57.0	78.1
G4		54.3	-16.0	4.7	-80.8	-71.6	20.2	35.1	53.6	73.6
G4(MP2)		55.9	-14.6	7.0	-79.2	-70.1	19.0	33.2	51.0	70.2
G4(MP2)-6X		55.1	-15.6	5.7	-81.0	-71.5	18.5	32.7	50.7	70.0
G3X <sup>h</sup>		53.8	-17.1	3.8	-81.9	-72.6				
					$\Delta H_{f,298K}^\circ$ from experiment					
	Pedley <sup>i</sup>	$48.9 \pm 2.0$	$-14.2 \pm 2.4$	...	$-78.6 \pm 1.0$	$-72.9 \pm 0.6$	$19.7 \pm 0.2$	$35.9 \pm 0.4$	$55.2 \pm 0.5$	$69.6 \pm 2.2$
	Pedley revised (DV) <sup>h</sup>	$53.9 \pm 0.8$	$-16.6 \pm 0.8$	...	$-80.8 \pm 0.6$	$-72.1 \pm 0.6$				
	RTCNI <sup>j</sup>						$19.8 \pm 0.2$	$36.0 \pm 0.4$	$54.8 \pm 0.7$	$81.9 \pm 1.4$ <sup>k</sup>

<sup>a</sup>The (T) component is taken from W2-F12 theory (except for naphthacene where it is taken from W1-F12 theory).

<sup>b</sup>The CV contribution is calculated as CCSD-F12/CVTZ-F12 + (T)/MTsmall(no  $f$ ) (except for naphthacene where it is taken from W1-F12 theory), for first-row systems these result in very similar CV corrections, see Sec. III D and Table IV.

<sup>c</sup>For the DNA/RNA bases the DBOC is calculated at the HF/A'VDZ level of theory, whereas for the aromatic hydrocarbons the best values from Ref. 64 are used.

<sup>d</sup>DNA/RNA bases: B3LYP-D/Def2-TZVP harmonic frequencies scaled by 0.985;  $C_6H_6$ : anharmonic ZPVE taken from Ref. 26;  $C_{10}H_8$ ,  $C_{14}H_{10}$ , and  $C_{18}H_{12}$ :  $0.985 \times B3LYP-D/Def2-TZVP$  harmonic frequencies scaled by  $c$  (where  $c = ZPVE_{anhar}[\text{benzene}]/ZPVE_{har}[\text{benzene}] = 1.003$ ).

<sup>e</sup>The total atomization energies at 0 K (TAE<sub>0</sub>) were converted to heats of formation at 0 K using ATcT atomic heats of formation at 0 K (H  $51.633 \pm 0.000$ , C  $170.024 \pm 0.014$ , N  $112.469 \pm 0.007$ , and O  $58.997 \pm 0.000$  kcal/mol).<sup>65</sup>

<sup>f</sup>Heats of formation at 298 K were converted to 0 K using the CODATA (Ref. 66) enthalpy functions,  $H_{298} - H_0$ , for the elemental reference states ( $H_2(g) = 2.024 \pm 0.000$ , C(cr,graphite) =  $0.251 \pm 0.005$ ,  $N_2(g) = 2.072 \pm 0.000$ , and  $O_2(g) = 2.075 \pm 0.000$  kcal/mol), whereas the enthalpy functions for the DNA and RNA bases were obtained within the rigid rotor-harmonic oscillator (RRHO) approximation from the B3LYP-D/Def2-TZVP calculated geometry and harmonic frequencies.

<sup>g</sup>The TAE<sub>0</sub> obtained from these thermochemical protocols were converted to  $\Delta H_{f,298K}^\circ$  values as prescribed in footnotes *b* and *c*.

<sup>h</sup>From Ref. 56.

<sup>i</sup>From Ref. 60.

<sup>j</sup>From Ref. 61 (the ATcT value for benzene is  $19.89 \pm 0.07$  kcal/mol).

<sup>k</sup>The experimental solid-phase enthalpy of combustion used to derive this value is questionable, see text.

our best  $\Delta H_{f,298K}^\circ$  are: 54.93 (A), -16.26 (C), 5.09 (G), -80.57 (T), and -71.56 kcal/mol (U). The theoretical heat of formation for uracil is in reasonably good agreement with both the experimental heat of formation ( $-72.9 \pm 0.6$  kcal/mol) and DV's revised value ( $-72.1 \pm 0.6$  kcal/mol). For thymine, the theoretical heat of formation is 2.0 kcal/mol more exothermic than the experimental value ( $-78.6 \pm 1.0$ ) but in excellent agreement with DV's revision ( $-80.8 \pm 0.6$  kcal/mol) suggesting that the experimental value should be revised downwards. And while the theoretical value for adenine is more endothermic than the original experiment ( $48.9 \pm 2.0$  kcal/mol) by as much as 6.0 kcal/mol, the gap shrinks to 1 kcal/mol with DV's revised

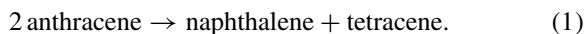
value. Finally, for cytosine, our calculation is once again in excellent agreement with DV's revised value. Given that post-CCSD(T) excitations are not expected to have significant contributions (*vide supra*) the weakest part in our theoretical model is the neglect of explicit anharmonicity in the ZPVE. Nevertheless, the use of a scaling factor to implicitly accounts for anharmonicity cannot plausibly account for such a large discrepancy between theory and experiment.

We note that the heats of formation obtained from G4 and G4(MP2)-6X theories are in reasonably good agreement with our best  $W_n$ -F12 values, the largest discrepancies being of 0.7 kcal/mol. However, larger deviations of, typically, 1–2 kcal/mol are obtained for the other composite

procedures that were considered (namely, CBS-QB3, G4(MP2), and G3X).<sup>59</sup>

In addition, we have calculated the heats of formation at 298 K for benzene, naphthalene, anthracene, and naphthacene (a.k.a., tetracene). These results are shown in Table VI. Our W1-F12 value for benzene (19.64 kcal/mol) is in good agreement with the experimental values ( $19.89 \pm 0.07$  kcal/mol, from ATcT, and  $19.8 \pm 0.2$  kcal/mol, from the 2008 Roux *et al.* (RTCN) compilation<sup>61</sup>). For naphthalene and anthracene the W1-F12 values (34.94 and 53.88 kcal/mol, respectively) are  $\sim 1$  kcal/mol lower than the experimental values ( $36.0 \pm 0.4$  and  $54.8 \pm 0.7$  kcal/mol, respectively, from RTCN). However, for naphthacene our W1-F12 value is higher by as much as 8 kcal/mol (!) than the experimental value reported in RTCN (and adopted by the NIST-JANAF thermochemical tables<sup>62</sup>).

A sanity check on the theoretical and experimental gas-phase values is afforded by considering the reaction enthalpy of:



At 298 K, the W1-F12 reaction energy of 1.0 kcal/mol is in close agreement with values of 1.5 kcal/mol (G4), 1.4 kcal/mol (G4MP2), 1.3 kcal/mol (G4MP2-6X), and 1.8 kcal/mol (CBS-QB3) obtained by other composite thermochemistry schemes. All of these values are irreconcilable with +8.3 kcal/mol obtained from RTCN,<sup>61</sup> or for that matter  $-4.9$  kcal/mol obtained from the older Pedley *et al.* compilation.<sup>60</sup>

Based on the heats of formation for the homologous series, Slayden and Liebman (SL)<sup>63</sup> estimated a heat of formation for naphthacene (72.2 kcal/mol, see further details in Ref. 63) which is in better agreement with our W1-F12 value (i.e., lower by 1.7 kcal/mol). Note that the solid-phase enthalpy of formation recommended by SL (42.3 kcal/mol) is lower by 7.1 kcal/mol than the 49.4 kcal/mol used by RTCN (Ref. 61) (marked “questionable” in that work). In addition, the enthalpy of sublimation used in Ref. 61 for obtaining the gaseous heat of formation is taken as the average of five values that differ from each other by as much as 3.3 kcal/mol. Combining the RTCN data for naphthalene and anthracene with the W1-F12 reaction energy for Eq. (1) leads to  $\Delta_f H_{298}^\circ[\text{naphthacene(g)}] = 74.6$  kcal/mol, just 0.75 kcal/mol higher than the directly computed W1-F12 value and within the 1.0 kcal/mol uncertainty in the former number based on just the constituent experimental uncertainties. Averaging between both numbers, we obtain  $74.25 \pm 1$  kcal/mol, to which we have attached a conservative uncertainty estimate.

#### IV. DISCUSSION AND CONCLUSIONS

Basis set convergence of the HF\*, CCSD-F12b, and (T\*) components of the total atomization energies in explicitly correlated calculations have been studied for the 137 species in the W4-11 database. Comparisons with the basis set convergence in standard HF, CCSD, and (T) calculations are made. While in conventional calculations basis set extrapolations (of very large one-particle Gaussian basis sets) play a key role in

approaching the infinite basis set limit, in explicitly correlated calculations basis set extrapolations play a much lesser role.

The HF\* energy (Hartree–Fock energy with the CABS singles correction) converges much more rapidly to the basis set limit than the standard HF energy. For example, the same RMSD (0.04 kcal/mol) is obtained from HF\*/V{D,T}Z-F12 and HF/A'V{Q,5}Z extrapolations. This fast basis set convergence of the HF\* component completely obviates the need for basis set extrapolations when an *spdfg* quality basis set is used. However, when basis sets of up to *spdf* quality are used, basis set extrapolations are still meaningful (e.g., RMSDs of 0.08 and 0.04 kcal/mol are obtained for the HF\*/VTZ-F12 and HF\*/V{D,T}Z-F12 levels of theory, respectively).

Similarly, the CCSD-F12b energy converges more rapidly to the basis set limit than the CCSD energy. For first-row systems the CCSD-F12/V{D,T}Z-F12 extrapolation results in a RMSD of only 0.17 kcal/mol, for comparison a similar RMSD (0.15 kcal/mol) is obtained from the A'V{Q,5}Z extrapolation in conventional CCSD calculations. However, we find that the VDZ-F12 basis set behaves in an erratic manner for systems containing second-row elements, and thus does not lend itself for effective basis set extrapolations in such cases. The CCSD-F12/V{T,Q}Z-F12 extrapolation results in RMSDs of 0.15 and 0.13 kcal/mol over the second-row systems and over the entire set, respectively. For comparison the A'V{Q,5}Z extrapolation in conventional CCSD calculations results in a RMSD of 0.14 kcal/mol over the second-row systems and over the entire set.

The (T\*)/V{D,T}Z-F12 and (T)/A'V{D,T}Z extrapolations yield very similar RMSD over the entire W4-11 set (namely, of 0.15 and 0.17 kcal/mol, respectively). However, it is found that the (T\*)/V{D,T}Z-F12 extrapolation performs better than the (T)/A'V{D,T}Z extrapolation for second-row systems (RMSD = 0.10 and 0.30 kcal/mol, respectively). Whereas for first-row systems (T)/A'V{D,T}Z extrapolation outperforms the (T\*)/V{D,T}Z-F12 extrapolation (RMSD = 0.04 and 0.16 kcal/mol, respectively). Nevertheless, scaling the (T\*)/VTZ-F12 by an empirical factor of 0.987 results in RMSDs of 0.07, 0.10, and 0.08 kcal/mol for the first- and second-row species and the entire set, respectively.

Based on these findings we propose two explicitly correlated variants of W1 and W2 theories (W1-F12 and W2-F12). The computationally more economic W1-F12 theory shows excellent performance for systems containing only first-row elements (and H). Specifically, over the 97 first-row systems in the W4-11 dataset, W1-F12 attains a RMSD of 0.19 kcal/mol against all-electron, relativistic CCSD(T) reference atomization energies at the infinite basis set limit. When considering reference atomization energies at the FCI basis set limit a RMSD of 0.45 kcal/mol is obtained. The W2-F12 protocol shows excellent performance for both first- and second-row systems. In particular, against all-electron, relativistic CCSD(T) reference atomization energies a RMSD of 0.20 kcal/mol is obtained over the entire W4-11 test set (for comparison for the W1w and W2w protocols RMSDs of 0.51 and 0.19 kcal/mol are obtained, respectively). Thus, W1-F12 clearly outperforms W1w, whereas W2-F12 and W2w show similar performance. When considering reference atomization energies at the FCI basis set limit, RMSDs of 0.42,

0.63, and 0.42 kcal/mol are obtained for W2-F12, W1w, and W2w respectively (over the entire W4-11 test set).

Finally, the proposed W1-F12 and W2-F12 protocols have been successfully applied to the DNA/RNA bases adenine, cytosine, guanine, thymine, and uracil. The W1-F12 and W2-F12 predictions for the heats of formation of cytosine, thymine, and uracil are in agreement with the experimental values. However, the theoretical  $\Delta H_{f,298K}^\circ$  for adenine is impossible to reconcile with experiment and suggest that re-measurement may be in order. In addition, the W1-F12 protocol was applied to the homologous series (benzene, naphthalene, anthracene, and naphthacene/tetracene). The W1-F12 theoretical values for benzene, naphthalene, and anthracene are in very good to reasonable agreement with experiment. However, the W1-F12 value for naphthacene represents a substantial revision from experiment; combination of a direct and an indirect estimate suggests that  $\Delta_f H_{298}^\circ[\text{naphthacene(g)}] = 74.25 \pm 1$  kcal/mol is a more realistic estimate.

## ACKNOWLEDGMENTS

J.M.L.M. is on leave of absence as the Thatcher Professor of Chemistry at the Weizmann Institute of Science. This research was supported in part by the Weizmann AERI (Alternative Energy Research Initiative) and by a startup grant from the University of North Texas. A.K. was supported by Australian Research Council Discovery Project Grant No. DP110102336. The authors would like to acknowledge a generous grant of computer time by the Faculty of Chemistry at the Weizmann Institute of Science. This paper is dedicated to Peter R. Taylor on the occasion of his 60th birthday.

<sup>1</sup>L. Goerigk and S. Grimme, *J. Chem. Theory Comput.* **7**, 291 (2011).  
<sup>2</sup>J. M. L. Martin and G. de Oliveira, *J. Chem. Phys.* **111**, 1843 (1999).  
<sup>3</sup>S. Parthiban and J. M. L. Martin, *J. Chem. Phys.* **114**, 6014 (2001).  
<sup>4</sup>N. J. DeYonker, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **124**, 114104 (2006).  
<sup>5</sup>N. J. DeYonker, T. Grimes, S. Yockel, A. Dinescu, B. Mintz, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **125**, 104111 (2006).  
<sup>6</sup>K. A. Peterson, D. Feller, and D. A. Dixon, *Theor. Chem. Acc.* **131**, 1079 (2012).  
<sup>7</sup>T. Helgaker, W. Klopper, and D. P. Tew, *Mol. Phys.* **106**, 2107 (2008).  
<sup>8</sup>J. M. L. Martin and S. Parthiban, *Computational Thermochemistry. Prediction and Estimation of Molecular Thermodynamics*, Understanding Chemical Reactivity Vol. 22, edited by J. Cioslowski (Kluwer, Dordrecht, 2001), pp. 31–65.  
<sup>9</sup>D. Feller, K. A. Peterson, and J. G. Hill, *J. Chem. Phys.* **135**, 044102 (2011).  
<sup>10</sup>A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, *J. Chem. Phys.* **125**, 144108 (2006).  
<sup>11</sup>A. Karton, P. R. Taylor, and J. M. L. Martin, *J. Chem. Phys.* **127**, 064104 (2007).  
<sup>12</sup>W. D. Allen, A. L. L. East, and A. G. Császár, *Structures and Conformations of Non-Rigid Molecules*, edited by J. Laane, M. Dakkouri, B. van der Veken, and H. Oberhammer (Kluwer, Dordrecht, 1993), p. 343.  
<sup>13</sup>A. G. Császár, W. D. Allen, and H. F. Schaefer, *J. Chem. Phys.* **108**, 9751 (1998).  
<sup>14</sup>A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, J. F. Stanton, *J. Chem. Phys.* **121**, 11599 (2004).  
<sup>15</sup>Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, and J. F. Stanton, *J. Chem. Phys.* **125**, 064108 (2006).  
<sup>16</sup>M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, *J. Chem. Phys.* **128**, 114111 (2008).

<sup>17</sup>B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszewski, S. Bittner, S. G. Nijssure, K. A. Amin, M. Minkoff, and A. F. Wagner, *J. Phys. Chem. A* **108**, 9979 (2004).  
<sup>18</sup>B. Ruscic, *Encyclopedia of Science and Technology (2005 Yearbook of Science and Technology)* (McGraw-Hill, New York, 2004), pp. 3–7.  
<sup>19</sup>B. Ruscic, R. E. Pinzon, M. L. Morton, N. K. Srinivasan, M.-C. Su, J. W. Sutherland, and J. V. Michael, *J. Phys. Chem. A* **110**, 6592 (2006).  
<sup>20</sup>A. Karton, S. Daon, and J. M. L. Martin, *Chem. Phys. Lett.* **510**, 165 (2011).  
<sup>21</sup>E. A. Hylleraas, *Z. Phys.* **1929**, 54 (347); For an english translation, see H. Hettema, *Quantum Chemistry: Classic Scientific Papers* (World Scientific, Singapore, 2000), Vol. 8, p. 81.  
<sup>22</sup>W. Klopper and W. Kutzelnigg, *Chem. Phys. Lett.* **134**, 17 (1987).  
<sup>23</sup>W. Kutzelnigg and W. Klopper, *J. Chem. Phys.* **94**, 1985 (1991).  
<sup>24</sup>A. K. Peterson, T. B. Adler, and H.-J. Werner, *J. Chem. Phys.* **128**, 084102 (2008).  
<sup>25</sup>J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **131**, 194105 (2009).  
<sup>26</sup>A. Karton, I. Kaminker, and J. M. L. Martin, *J. Phys. Chem. A* **113**, 7610 (2009).  
<sup>27</sup>W. Klopper, R. Bachorz, C. Hättig, and D. P. Tew, *Theor. Chim. Acta* **126**, 289 (2010).  
<sup>28</sup>W. Klopper, B. Ruscic, D. P. Tew, F. A. Bischoff, and S. Wolfsegger, *Chem. Phys.* **356**, 14 (2009).  
<sup>29</sup>H. Fliegl, W. Klopper, and C. Hättig, *J. Chem. Phys.* **122**, 084107 (2005).  
<sup>30</sup>R. Haunschuld and W. Klopper, *Theor. Chem. Acc.* **131**, 1112 (2012).  
<sup>31</sup>A. Mahler and A. K. Wilson (unpublished).  
<sup>32</sup>H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz *et al.*, PLBIBSC-MOLPRO, version 2010.1, a package of *ab initio* programs, 2010, see <http://www.molpro.net>.  
<sup>33</sup>M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 09, Revision A.1, Gaussian, Inc., Wallingford, CT, 2009.  
<sup>34</sup>F. Weigend, *Phys. Chem. Chem. Phys.* **8**, 1057 (2006), and references therein.  
<sup>35</sup>See supplementary material at <http://dx.doi.org/10.1063/1.3697678> for the basis set convergence of the SCF, CCSD, and (T) components in standard and explicitly correlated calculations for all of the molecules in the W4-11 dataset are presented in Tables SI-I, SI-II, and SI-III, respectively. Basis set convergence of the CCSD and (T) components of the core-valence corrections for all of the molecules in the W4-11 dataset are presented in Table SI-IV and SI-V. The deviations from basis set limit reference values of the W1-F12, W2-F12, W1w, and W2w protocols for the molecules in the W4-11 dataset are given in Table SI-VI. B3LYP-D/Def2-TZVP optimized geometries of adenine, cytosine, guanine, thymine, uracil, benzene, naphthalene, anthracene, and naphthacene are also provided.  
<sup>36</sup>T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).  
<sup>37</sup>R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).  
<sup>38</sup>T. H. Dunning, K. A. Peterson, and A. K. Wilson, *J. Chem. Phys.* **114**, 9244 (2001).  
<sup>39</sup>E. C. Barnes, G. A. Petersson, J. A. Montgomery, M. J. Frisch, and J. M. L. Martin, *J. Chem. Theory Comput.* **5**, 2687 (2009).  
<sup>40</sup>M. Douglas and N. M. Kroll, *Ann. Phys.* **82**, 89 (1974).  
<sup>41</sup>B. A. Heß, *Phys. Rev. A* **33**, 3742 (1986).  
<sup>42</sup>This work was inspired by the following pioneering paper on static Gaussian geminals: B. J. Persson and P. R. Taylor, *J. Chem. Phys.* **105**, 5915 (1996); S. Ten-no and J. Noga, *WIREs Comput. Mol. Sci.* **2**, 114 (2012).  
<sup>43</sup>H.-J. Werner, T. B. Adler, and F. R. Manby, *J. Chem. Phys.* **126**, 164102 (2007).  
<sup>44</sup>G. Knizia and H.-J. Werner, *J. Chem. Phys.* **128**, 154103 (2008).  
<sup>45</sup>T. B. Adler, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **127**, 221106 (2007).  
<sup>46</sup>G. Knizia, T. B. Adler, and H.-J. Werner, *J. Chem. Phys.* **130**, 054104 (2009).  
<sup>47</sup>J. Noga, S. Kedžuch, and J. Šimunek, *J. Chem. Phys.* **127**, 034106 (2007).  
<sup>48</sup>G. J. Hill, S. Mazumder, and K. A. Peterson, *J. Chem. Phys.* **132**, 054108 (2010).  
<sup>49</sup>K. E. Yousaf and K. A. Peterson, *J. Chem. Phys.* **129**, 184108 (2008).  
<sup>50</sup>E. F. Valeev, *Chem. Phys. Lett.* **395**, 190 (2004).  
<sup>51</sup>F. Weigend, *Phys. Chem. Chem. Phys.* **4**, 4285 (2002).

- <sup>52</sup>F. Weigend, A. Köhn, and C. Hättig, *J. Chem. Phys.* **116**, 3175 (2002); C. Hättig, *Phys. Chem. Chem. Phys.* **7**, 59 (2005).
- <sup>53</sup>For a normal distribution without bias (systematic error),  $\text{MAD/RMSD} = \sqrt{2/\pi} = 0.7979 \dots \approx 4/5$ , see R. C. Geary, *Biometrika* **27**, 310 (1935). As discussed in Ref. 20 and references therein, a few large outliers will cause a significant drop from that idealized ratio.
- <sup>54</sup>W. A. de Jong, R. J. Harrison, and D. A. Dixon, *J. Chem. Phys.* **114**, 48 (2001).
- <sup>55</sup>The 13 highly multireference systems, for which the  $\%TAE_e[\text{T}]$  diagnostic is in excess of 10%, are:  $\text{B}_2$ ,  $\text{C}_2(^1\Sigma^+)$ ,  $\text{BN}(^1\Sigma^+)$ , OF,  $\text{F}_2\text{O}$ , FOO,  $\text{FOOF}$ ,  $\text{Cl}_2\text{O}$ , CIOO, OCIO,  $\text{O}_3$ ,  $\text{S}_3$ , and  $\text{S}_4$ , see Ref. 20 for further details.
- <sup>56</sup>O. V. Dorofeeva and N. Vogt, *J. Chem. Eng. Data* **54**, 1348 (2009).
- <sup>57</sup>For a review of G3, G3X, G4, and related empirical composite wavefunction methods, see L. A. Curtiss, P. C. Redfern, K. Raghavachari, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **1**, 810 (2011).
- <sup>58</sup>A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss, *J. Chem. Phys.* **120**, 4129 (2004).
- <sup>59</sup>L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **126**, 084108 (2007); **127**, 124105 (2007); B. Chan, J. Deng, L. Radom, *J. Chem. Theory Comput.* **7**, 112 (2011); J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **110**, 2822 (1999).
- <sup>60</sup>J. B. Pedley, *Thermochemical Data and Structures of Organic Compounds* (Thermodynamics Research Center, College Station, TX, 1994), Vol. I.
- <sup>61</sup>M. Victoria Roux, M. Temprado, J. S. Chickos, and Y. Nagano, *J. Phys. Chem. Ref. Data* **37**, 1855 (2008); See also Y. Nagano, *J. Chem. Thermodyn.* **34**, 377 (2002); W. Acree and J. S. Chickos, *J. Phys. Chem. Ref. Data* **39**, 043101 (2010).
- <sup>62</sup>M. W. Chase, C. A. Davies, J. R. Downey, D. J. Frurip, R. A. McDonald, A. N. Syverud, *JANAF Thermochemical Tables*, 3rd ed. (American Chemical Society and American Institute of Physics for the National Bureau of Standards, 1985), [J. Phys. Chem. Ref. Data **14**, (Suppl. 1) (1985)]; M. W. Chase, *NIST-JANAF Thermochemical Tables*, 4th ed. (American Institute of Physics, 1998), [J. Phys. Chem. Ref. Data 1998, Monograph No. 9]. Data are also accessible online via <http://webbook.nist.gov/chemistry>.
- <sup>63</sup>S. W. Slayden and J. F. Liebman, *Chem. Rev.* **101**, 1541 (2001).
- <sup>64</sup>A. Tajti, P. G. Szalay, and J. Gauss, *J. Chem. Phys.* **127**, 014102 (2007).
- <sup>65</sup>Based on early beta version 1.110 of the Core (Argonne) Thermochemical Network, see also <http://atct.anl.gov/Thermochemical%20Data/version%20earlyBeta%201.110/index.html>.
- <sup>66</sup>J. D. Cox, D. D. Wagman, and V. A. Medvedev, *CODATA Key Values for Thermodynamics* (Hemisphere Publishing Corp., New York, 1989), see also <http://www.codata.org/resources/databases/key1.html>.