W1X-1 and W1X-2: W1-Quality Accuracy with an Order of Magnitude Reduction in Computational Cost

Bun Chan* and Leo Radom*

School of Chemistry and ARC Centre of Excellence for Free Radical Chemistry and Biotechnology, University of Sydney, Sydney, NSW 2006, Australia

ⓢ Supporting Information

ABSTRACT: We have examined a number of approaches for reducing the computational requirements of the W1w and W1-F12 procedures, while maintaining the accuracy. A key finding is that MP2/cc-pCVTZ provides a reliable means for the evaluation of core-correlation effects at a cost that is negligible in the context of W1-type procedures. This greatly reduces the overall computational cost, since calculations for core-correlation represent the most time-consuming steps for both W1w and W1-F12. For the evaluation of valence CCSD(T)/CBS, we find that truncation of the sets of diffuse functions leads to a significant savings in time, with only a minor deterioration in the performance. In order to eliminate the need to carry out CCSD calculations with a quadruple-ζ basis set, we have maintained the approach employed in W1-F12, namely the use of explicitly correlated procedures. Our resulting procedures are termed W1X-1 and W1X-2 (where the X refers to the eXplicitly correlated procedures). Of these, the W1X-1 protocol requires two CCSD-F12b plus two CCSD(T) calculations to obtain the valence CCSD(T)/CBS energy component, as with W1-F12. The W1X-2 procedure, on the other hand, requires only two CCSD(T)-F12b calculations for the evaluation of CCSD(T)/CBS and is therefore less expensive than W1X-1. Indeed, the W1X-2 protocol is an order of magnitude less computationally demanding than W1w and ∼80% less costly than W1-F12. Extensive assessment of the W1X-1 and W1X-2 procedures shows that W1X-1 performs well and comparably to W1w and W1-F12 in virtually all cases. The W1X-2 method gives equivalently good results for most thermochemical properties, but the heats of formation of fluorocarbons and complexation energies of hydrogen fluoride clusters represent notable exceptions.

INTRODUCTION

Computational chemists need constantly to make decisions on the appropriate compromise between accuracy and computational expense. The now ubiquitous CCSD(T) methodology,[1] when used in conjunction with a large basis set that is sufficiently close to the complete-basis-set (CBS) limit, generally provides a reliable means for obtaining accurate energies. However, the combination of CCSD(T) with a very large basis set is currently prohibitively expensive for medium- to large-sized molecules. The range of applicability of high-level methods can be extended through the use of composite procedures,[2] which operate via the appropriate blending of a number of lower-level methods.

Among the various composite schemes, the W1 procedure[3] is notable for generally providing a reliable means for estimating the CCSD(T)/CBS energy. It is among the first of the composite procedures to attain chemical accuracy (1 kcal mol^{-1}, 4.2 kJ mol^{-1}), while at the same time being applicable in a practical sense to medium-sized systems. For instance, W1 has been employed to evaluate the heats of formation of phenol[4] and Cl_2O_7.[5] Nonetheless, application of W1 to larger systems remains costly. It is therefore desirable to further reduce the cost, in order to broaden the range of application. To this end, we note that the W1w variant[6] utilizes smaller but somewhat more balanced aug-cc-pV(n+d)Z basis sets instead of the aug-cc-pVnZ+2df basis sets used in the original W1, which leads to an improved accuracy together with lower computational requirements. The use of the aug-cc-pV(n+d)Z basis sets has also been adopted in the closely related W2w procedure[6] and

the higher-level W3 protocol.[6] During the course of the present investigation, Karton and Martin proposed the W1-F12 procedure as an economical alternative to W1w.[7] It employs the explicitly correlated CCSD-F12b procedure to avoid the use of a large quadruple-ζ basis set in W1w.

We have been interested for some time in the development of cost-effective computational chemistry procedures for theoretical thermochemistry.[8] In the present study, we examine modifications to W1w and W1-F12, with the aim of further reducing the computational cost, while retaining the accuracy, thus enabling the study of larger systems with a sub-kcal mol^{-1} accuracy methodology.

COMPUTATIONAL DETAILS

Standard ab initio molecular orbital theory and density functional theory (DFT) calculations[9] were carried out with the Gaussian 09[10] and Molpro 2010[11] programs. Unless otherwise noted, geometries were optimized with the B3-LYP/cc-pVTZ +d procedure. Following each geometry optimization, harmonic frequency analysis was carried out to confirm the nature of the stationary point as an equilibrium structure or a transition structure, as appropriate. For all W1-type methods, improved single-point energies for open-shell species were obtained using restricted-open-shell reference wave functions.

Two-point extrapolation schemes with the formula $E_L = E_{CBS} + AL^{-\alpha}$ were used in the evaluation of the CBS limit,[12] where L is

the cardinal number of the basis set, i.e., 2 for double-ζ and 3 for triple-ζ basis sets, etc., and α is an adjustable parameter. In the W1w scheme, α is taken as 5 for Hartree−Fock energies, and this value was also employed throughout the present study. For correlation energies, a value of 3.22 for α is used in W1w and, unless otherwise noted, was also employed in the present study. In the W1-F12 procedure, α values of 3.67 and 3.22 are employed for the CCSD-F12b and conventional (T) correlation energy components, respectively. In a number of cases in the present study, the value of α was reoptimized for the correlation energy components using the G2/97 training set.[13] Following previous practice,[8] our optimization procedures involved minimizing the average of the mean absolute deviation (MAD) from benchmark values and the standard deviation (SD) of these deviations.

Literature values for corrections for spin−orbit coupling, where available, were included for atomic species and a number of molecules.[13] To obtain the zero-point vibrational energies (ZPVEs) and thermal corrections for enthalpies at 298 K (ΔH_{298}), we used B3-LYP/cc-pVTZ+d harmonic vibrational frequencies, scaled by 0.985 as prescribed for W1w. All relative energies are given in kJ mol^{-1}.

■ THEORETICAL APPROACH

The baseline method that we employed in the present study is the W1w procedure[6] for all but a small number of species that contain Na and Mg, for which a minor modification is employed. For the sake of simplicity, we will continue to refer to this composite scheme as W1w. It comprises the following single-point energy components:

- An HF/CBS energy is obtained from calculations with the aug′-cc-pV(T+d)Z and aug′-cc-pV(Q+d)Z basis sets, and the extrapolation formula $E_L = E_{CBS} + AL^{-5}$. The use of aug′ denotes the inclusion of diffuse functions only for p-block elements.[14]
- A CCSD correlation energy component (ΔCC) is extrapolated to the CBS limit using the aug′-cc-pV-(T+d)Z and aug′-cc-pV(Q+d)Z basis sets and the formula $E_L = E_{CBS} + AL^{-3.22}$. The chemical core (i.e., the largest noble-gas core) is frozen for the evaluation of the correlation energy, except for Na and Mg, where the inner-valence, i.e., the 2s and 2p, orbitals are also correlated in conjunction with cc-pCVnZ basis sets.[14,15]
- A (T) component [Δ(T)] of the CCSD(T) correlation energy is extrapolated using the aug′-cc-pV(D+d)Z and aug′-cc-pV(T+d)Z basis sets with the formula[14,15] $E_L = E_{CBS} + AL^{-3.22}$.
- A combined core-correlation and scalar-relativistic component [Δ(C+R)] is obtained as the difference between the frozen-core and all-electron scalar-relativistic (second-order Douglas−Kroll−Hess) CCSD(T)/MTsmall energies.

Thus, to obtain the total composite W1w energy, five calculations are required: (1) CCSD(T)/aug′-cc-pV(D+d)Z, (2) CCSD(T)/aug′-cc-pV(T+d)Z, (3) CCSD/aug′-cc-pV(Q+d)Z, (4) CCSD(T)/MTsmall, and (5) CCSD(T)(Full,DKH)/MTsmall.

It has been pointed out that, for medium-sized molecules such as benzene, the evaluation of the core-correlation and scalar-relativistic effects with CCSD(T)/MTsmall calculations often represents the most time-consuming step.[16] To this end, the W1c method[17] was introduced that makes use of an empirical

scheme for the evaluation of core-correlation and scalar-relativistic effects. While W1c addresses the cost issue, the use of an empirical correction may lead to larger uncertainty. In addition, one can envisage limitations in its applicability, associated with the limitations in the empirical scheme. We note that the core-correlation component in the W1-F12 protocol is already less expensive than that in W1w.[7] However, it still represents a substantial bottleneck in the W1-F12 procedure. We also note that alternative Gn(MP2)-type composite procedures[8,18] do not explicitly include a core-correlation component, but they still perform reasonably well. We therefore believe that there is room for further substantial reduction in cost, associated with a more efficient evaluation of the core-correlation component, and this is one of the key objectives of the present investigation.

In a recent study, Papajak and Truhlar[19] examined the use of truncated sets of diffuse functions within the aug-cc-pV(n+d)Z basis sets. Their jul-cc-pV(n+d)Z basis sets are equivalent to the aug′-cc-pV(n+d)Z sets employed in the present study. In comparison, the jun-cc-pV(n+d)Z basis sets are further reduced in size, by the removal of the highest angular momentum diffuse functions. It has been pointed out that the jun-cc-pV(n+d)Z basis sets are capable of systematically approaching the CBS limit.[19] In the present study, we investigate the use of both the jul-cc-pV(n+d)Z and the smaller jun-cc-pV(n+d)Z basis sets. For the sake of compactness, the jul-cc-pV(n+d)Z and jun-cc-pV(n+d)Z basis sets are abbreviated hereafter as 7-nZ and 6-nZ.

The use of a quadruple-ζ basis set for the CCSD calculation in W1w often contributes significantly to the computational cost. Nevertheless, this is required due to the slow convergence of the CCSD correlation energy, rendering a [D,T] extrapolation inadequate and necessitating the [T,Q] scheme to be used in W1w. In recent years, however, advances in explicitly correlated methods,[20] e.g., CCSD-F12-type procedures,[20] have enabled a more rapid convergence of correlation energies using smaller basis sets. As noted earlier, the CCSD-F12b methodology has recently been employed by Karton and Martin in the implementation of the W1-F12 procedure.[7] In the present study, we also incorporate CCSD-F12b and, in some cases, CCSD(T)-F12b into our composite procedures.

■ RESULTS AND DISCUSSION

Modifications to the Core-Correlation and Scalar-Relativistic Components. A brief summary of the components of the various modified W1-type procedures is presented in Table 1, while their performance for the G2/97 test set, and the timing comparisons, are shown in Table 2. We first note that W1w (Table 2, column 1) performs well for the complete G2/97 set, which is consistent with the results of a previous study[21] on a truncated G2/97 set. The overall mean absolute deviation (MAD) from experimental values is 3.0 kJ mol^{-1}. Among the four subsets of thermochemical properties, the largest MAD of 3.7 kJ mol^{-1} is found for heats of formation ($\Delta H_f'$s). For W1w, the time required for obtaining the core-correlation and scalar-relativistic components [Δ(C+R)] corresponds, on average for the G2/97 set, to 69% of the total time required for obtaining the complete W1w energy. We also note that, while the Δ(C+R) component in W1-F12 (column 2) is noticeably less expensive than that for W1w, it still represents the most costly constituent in the W1-F12 protocol, even for the G2/97 set that comprises relatively small molecules (the largest being toluene). Therefore, we first focus on reducing the cost for the calculation of Δ(C+R).

Table 1. Components of the Various W1-Type Procedures[a,b]

	1	2	3	4	5	6	7	8
	W1w	W1-F12[c]	W1(C1)[d]	W1(C2)[d]	W1(C3)[d]	6-W1(C3)[d]	W1X-1	W1X-2
HF	7-[T,Q]Z	HF(CABS)/V[D,T]Z-F12	7-[T,Q]Z	7-[T,Q]Z	7-[T,Q]Z	6-[T,Q]Z	HF(CABS)/ V[D,T]Z-F12	HF(CABS)/ A′V[D,T]Z
ΔCC	7-[T,Q]Z	CCSD-F12b/V[D,T]Z-F12	7-[T,Q]Z	7-[T,Q]Z	7-[T,Q]Z	6-[T,Q]Z	CCSD-F12b/ V[D,T]Z-F12	CCSD-F12b/ A′V[D,T]Z
Δ(T)	7-[D,T]Z	7-[D,T]Z	7-[D,T]Z	7-[D,T]Z	7-[D,T]Z	6-[D,T]Z	7-[D,T]Z	CCSD(T)-F12b/ A′V[D,T]Z
Δ(C+R)	CCSD(T)/ MTSmall	ΔC: ΔC[CCSD/WCVTZ] + ΔC[(T)/WCVTZ(−f)] ΔR: CCSD(T)(DKH)/7-DZ-DK − CCSD(T)/7-DZ	CCSD(T)/ CVTZ	CCSD/ CVTZ	MP2/ CVTZ	MP2/ CVTZ	MP2/CVTZ	MP2/CVTZ

[a]HF(CABS) = HF energy with complementary auxiliary basis singles correction. Unless otherwise indicated, ΔCC = CCSD − HF, extrapolated to the CBS limit, Δ(T) = CCSD(T) − CCSD, extrapolated to the CBS limit, Δ(C+R) = all-electron DKH energy − frozen-core energy. [b]7-nZ = jul-cc-pV(nd)Z (= aug′-cc-pV(nd)Z), 6-nZ = jun-cc-pV(nd)Z, VnZ-F12 = cc-pVnZ-F12, A′VnZ = aug′-cc-pVnZ, CVTZ = cc-pCVTZ, WCVTZ = cc-pWCVTZ, WCVTZ(−f) = WCVTZ with the exclusion of f functions. [c]See ref 7 for details. [d]C1, C2, and C3 denote the use of CCSD(T)/CVTZ, CCSD/CVTZ, and MP2/CVTZ, respectively, for the evaluation of Δ(C+R).

Table 2. Mean Absolute Deviations (MAD, kJ mol^{-1}) from Experimental Values for the G2/97 Set and its Subsets, and Timing Comparisons for the Various W1-Type Procedures[a]

	1	2	3	4	5	6	7	8
	W1w	W1-F12	W1(C1)	W1(C2)	W1(C3)	6-W1(C3)	W1X-1	W1X-2
					performance			
G2/97[b]	3.0	2.7	2.8	2.8	2.6	2.8	2.4	2.5
ΔH_f	3.7	3.1	3.3	3.3	2.9	3.2	2.4	2.7
IE	2.7	2.5	2.7	2.8	2.8	2.9	2.8	2.6
EA	1.7	2.1	1.7	1.8	1.8	1.8	1.9	1.7
PA	1.8	2.1	1.9	1.8	1.9	2.0	2.2	2.2
					timing[c,d]			
DZ	0.004	0.011	0.004	0.004	0.004	0.002	0.011	0.005
TZ	0.074	0.125	0.074	0.074	0.074	0.045	0.124	0.075
QZ	0.231		0.231	0.231	0.231	0.165		
Δ(C+R)	0.691	0.280	0.262	0.072	0.003	0.003	0.003	0.003
total	1.000	0.416	0.571	0.381	0.312	0.215	0.138	0.083

[a]See Table 1 for details of the various theoretical procedures. [b]ΔH_f = heats of formation, IE = ionization energies, EA = electron affinities, PA = proton affinities. [c]Total aggregate CPU time adjusted to a single processor for all calculations for the G2/97 set, relative to the corresponding total time required for this set for W1w calculations. [d]DZ = CCSD(T)/DZ, TZ = CCSD(T)/TZ, and QZ = CCSD/QZ calculations.

The MTsmall basis set is a completely decontracted cc-pVTZ basis set augmented with a small number of extra polarization functions.[3] While complete decontraction ensures an almost complete recovery of core-correlation energy, it adds substantially to the cost of the calculation. For the evaluation of relative energies, systematic error cancellation may allow for the use of a more economical methodology. To this end, we note that the less expensive nonempirical ccCA procedure[2c] makes use of MP2/aug-cc-pCVTZ for the evaluation of the core-correlation component. In addition, the W1-F12 procedure employs the cc-pWCVTZ basis set, which is closely related to cc-pCVTZ, for the evaluation of core-correlation energies. In the present study, we have examined the use of the cc-pCVTZ basis set (CVTZ) as an alternative to MTsmall.[22] In addition to CCSD(T)/CVTZ, we have also investigated the use of CCSD and MP2 as less costly alternatives to CCSD(T). The use of CCSD(T)/CVTZ, CCSD/CVTZ, and MP2/CVTZ for the evaluation of Δ(C+R) is denoted as C1, C2, and C3, respectively (Table 1, columns 3−5).

The use of CCSD(T)/CVTZ [Table 2, column 3, W1(C1)] instead of CCSD(T)/MTsmall (column 1) actually leads to a somewhat lower MAD for ΔH_f's, while the MADs for ionization energies (IEs), electron affinities (EAs), and proton

affinities (PAs) are comparable to those for W1w. In terms of timing, the Δ(C+R) component of W1(C1) requires only ~40% of the time when compared with that for W1w, and this leads to W1(C1) being significantly less expensive than W1w. We note that, for the G2/97 set, the costs associated with the Δ(C+R) term for W1(C1) and W1-F12 are comparable with one another. The use of CCSD/CVTZ [column 4, W1(C2)] for the evaluation of Δ(C+R) offers a significant reduction in cost, while maintaining the performance in comparison with W1(C1).

The use of MP2/CVTZ for the evaluation of Δ(C+R) *essentially eliminates the cost* associated with the calculation of this component within the composite scheme. In addition, the resultant W1(C3) method [Table 2, column 5] provides a somewhat improved performance for ΔH_f's, with an MAD of 2.9 kJ mol^{-1} (versus 3.7 [W1w], 3.1 [W1-F12], 3.3 [W1(C1)], and 3.3 [W1(C2)] kJ mol^{-1}). The performance for IEs, EAs, and PAs is similar to that for W1w, W1(C1), and W1(C2). We conclude by noting that, while further savings can be made in the calculation of the Δ(C+R) term, e.g., by the use of the resolution-of-the-identity (RI) approximation,[23] it would be more beneficial at this point to seek cost reduction in other components.

Truncated Diffuse Functions for the Valence CCSD(T)/ CBS Component. The highest cost component involved in the evaluation of the valence CCSD(T)/CBS energy is by far the CCSD/7-QZ calculation (Table 2), which is followed by the CCSD(T)/7-TZ calculation. The 7-TZ basis set includes a set of 2df polarization functions and a set of spdf diffuse functions, while 7-QZ contains sets of 3d2fg polarization and spdfg diffuse functions. Mintz et al. have investigated the possibility of savings in computational cost by using truncated polarization functions in the correlation-consistent basis sets.[24] They found that the results systematically converge to the CBS limit. However, in our brief examination of the use of these basis sets, we find that such truncation of the polarization functions leads to a deterioration in the performance ($\sim$1 kJ mol^{-1}) that is significant in the context of highly accurate W1-type procedures.

We now turn our attention to an alternative approach for reducing basis set size, namely by truncation of the diffuse functions through the use of the partially augmented[19] 6-nZ basis sets (Table 1, column 6). In these basis sets, the highest angular momentum diffuse function is removed from the set. Thus, 6-DZ contains sp diffuse functions as opposed to spd diffuse functions in 7-DZ, while 6-TZ and 6-QZ contain spd and spdf diffuse functions, respectively. It can be seen from Table 2 that the use of 6-nZ (column 6) in the valence CCSD(T) and CCSD calculations results in only slightly larger MADs for the G2/97 set and its subsets, in comparison with the 7-nZ analog (column 5), which is also comparable to the baseline W1w method (column 1). In terms of timing, the use of the 6-nZ basis sets reduces the computational time by about 30% for each of the three CCSD(T)/CBS components, namely CCSD(T)/DZ, CCSD(T)/TZ, and CCSD/QZ, leading to a total time for 6-W1(C3) (column 6) that is also roughly 30% less than that for W1(C3) (column 5).

W1X-1 and W1X-2: Methodology. While the use of the 6-nZ basis sets offers notable cost savings, it does come at the expense of a slight deterioration in performance across the board. We now examine an alternative approach to savings in computational resources. The W1w scheme employs two-point extrapolations for the Hartree−Fock, CCSD, and (T) energies. These components can in principle be obtained from CCSD(T)/DZ and CCSD(T)/TZ calculations. The inclusion of the expensive CCSD/QZ step is necessary because of the slow basis-set convergence for CCSD. However, this difficulty can be overcome with explicitly correlated methods[20] such as CCSD-F12, which display much more rapid basis-set convergence. In its original formulation, the CCSD-F12 component introduces additional terms to the conventional CCSD method, and this in turn leads to a notable increase in cost over CCSD. Recent simplifications to CCSD-F12 yield more economical procedures. For example, the computational costs of the CCSD-F12a and CCSD-F12b methodologies[25] are only marginally greater than that of conventional CCSD.[26] The related CCSD(T)-F12a and CCSD(T)-F12b procedures comprise CCSD-F12a or CCSD-F12b, and a conventional perturbative (T)-correction.

The W1-F12 protocol[7] employs the HF(CABS) procedure, i.e., HF with complementary auxiliary basis singles correction, and also uses the CCSD-F12b components of two CCSD-F12b calculations with cc-pVnZ-F12 basis sets[27] (VnZ-F12; n = D and T), to obtain the HF/CBS and ΔCC/CBS values. The VnZ-F12 basis sets are specifically designed for F12-type procedures. For the perturbative triples contribution to the correlation energy, i.e., Δ(T)/CBS, two regular CCSD(T) calculations

(with the 7-DZ and 7-TZ basis sets) are used. Such separation of the CCSD and (T) components allows for the use of the somewhat better performing VnZ-F12 basis sets for the CCSD component, and the smaller 7-nZ basis sets for the faster-converging but more computationally demanding Δ(T) term.

In the present study, we also focus on the F12b approximation. The F12b-based methods (Table 1, columns 7 and 8) utilize DZ and TZ calculations and do not involve calculations using a QZ basis set. We note that, together with a DZ or a TZ basis set, the alternative F12a variant has been shown to yield correlation energies that are often closer to the CBS limit than the corresponding F12b values. However, the F12b form is found to possess smoother convergence behavior toward the CBS limit than F12a, which we deem a more desirable feature for the purpose of basis-set extrapolation.[25b] The new W1-type procedures are denoted W1(C3)-X1 and W1(C3)-X2 in full or, for the sake of simplicity, W1X-1 and W1X-2 in short, where X signifies the use of the eXplicitly correlated procedure. In the W1X-1 procedure, we simply adopt the set of four calculations in W1-F12 for the valence CCSD(T)/CBS component. For W1X-2, we aim for a further reduction in cost by simplifying the protocol to include only two CCSD(T)-F12b calculations using the aug′-cc-pVnZ (A′VnZ) basis sets (n = D and T) that are smaller than the VnZ-F12 basis sets. The A′VnZ basis sets are closely related to the 7-nZ basis sets employed in W1w, the difference between A′VnZ and 7-nZ being that 7-nZ includes additional tight-d functions for second-row p-block elements.[28,29]

As F12b methods converge differently when compared with conventional coupled-cluster procedures, one can contemplate a different optimal exponent α in the CBS extrapolation formula $E_L = E_{CBS} + AL^{-\alpha}$. Indeed, a recent study[30] obtained optimal α values for MP2-F12 that vary considerably from those for regular MP2. The study also demonstrated the different requirements for α when different types of basis sets are used, i.e., cc-pVnZ-F12 versus aug-cc-pVnZ. In addition, the W1-F12 procedure employs a different α value for the ΔCCSD-F12b component (3.67) from the one used for ΔCCSD (3.22) in W1w.

For W1X-1, we initially assumed the same α values as those used for W1-F12. Curiously, we find (results not shown) that this leads to a noticeable deterioration in the performance when compared with W1(C3) (column 5). This indicates that the valence CCSD(T)/CBS component in W1-F12 is somewhat different from that for W1(C3), the latter being the same as that for W1w (column 1). Presumably, the interdependence between the valence CCSD(T)/CBS and Δ(C+R) components in W1-F12 plays an important role in the overall good performance of this procedure (column 2). Taking all of the above into account, we reoptimized the α values for the extrapolation of the CCSD-F12b and (T) energies for both the W1X-1 and W1X-2 procedures. The optimal α values are shown in Table 3. We note that the optimized α value for the

Table 3. The Optimized α Values in the Formula $E_L = E_{CBS} + AL^{-\alpha}$ for the Extrapolation of CCSD and (T) Correlation Energies for the F12b-Based W1X-type Procedures

	W1X-1	W1X-2
	method	
ΔCC	CCSD-F12/V[D,T]Z-F12	CCSD(T)-F12/A′V[D,T]Z
Δ(T)	CCSD(T)/7-[D,T]Z	CCSD(T)-F12/A′V[D,T]Z
	coefficient α	
ΔCC	3.67	4.74
Δ(T)	2.04	2.09

ΔCC component for W1X-1 (3.67) turns out to be the same as the value used in W1-F12. However, the remaining optimized α values differ significantly.

Results for the G2/97 Set. How do the F12b-based W1X-type methods perform? We can see that W1X-1 (Table 2, column 7) has a somewhat improved overall performance for the G2/97 set when compared with W1(C3) (column 5). There is a somewhat lower MAD for ΔH_f for W1X-1 [2.4 versus 2.9 kJ mol^{-1} for W1(C3)]. The VnZ-F12 basis sets were developed with the inclusion of only a minimal set of sp diffuse functions, and their potential shortcoming for properties sensitive to diffuse functions has been previously commented on.[27] Despite this, the MADs for EAs and PAs are only very slightly larger than those for W1(C3). The A'VnZ-based method (W1X-2, column 8) also offers a consistently good performance across the different properties. Thus, the MAD for the G2/97 set is 2.5 kJ mol^{-1}, while the values for the subsets are 2.7 (ΔH_f), 2.6 (IE), 1.7 (EA), and 2.2 (PA) kJ mol^{-1}.

How do the two explicitly correlated W1-type methods fare in timing comparisons? The removal of the CCSD/QZ component from W1(C3) leads to a substantial cost reduction. Such a savings has been discussed by Karton and Martin.[7] For the complete G2/97 set, we find that the total time required for W1X-1 is approximately half of that for W1(C3). For W1X-2, the use of only two calculations with the A'VnZ basis sets offers further savings in computer time. A comparison of the timings for CCSD(T)-F12b/A'VnZ (column 8) and that for CCSD(T) with the 7-nZ basis sets (columns 1, 3–5), which are comparable in size to A'VnZ, shows that the CCSD(T)-F12b method is only marginally more expensive than CCSD(T). This observation is in accord with that of Werner and co-workers.[25] Overall, the timing for the W1X-1 method is ~14% of that for our baseline W1w method, and ~33% of that for W1-F12. The timing for W1X-2 is ~8% of W1w and 20% of W1-F12. Furthermore, the removal of the CCSD/QZ calculation also significantly reduces the requirements for scratch disk space. A more detailed discussion on the computational resource requirements for the various methods will be presented in the following section.

Resource Considerations for a Medium-Sized System: $C_6H_5Me^+$. The toluene radical cation is one of the largest molecules in the G2/97 set. We now use it to illustrate the reduced computational resources required for W1X-1 and W1X-2 computations compared with those for W1w and W1-F12 (Table 4) for a prototypical medium-sized system. We can see that, for the W1w procedure, the all-electron scalar-relativistic CCSD(T)/MTSmall calculation (CCSD(T)(Full,DKH)/MTSmall) is by far the most time-consuming step (111.1 CPU hours, versus 50.8 h for the second most time-consuming CCSD/7-QZ step). However, the use of the somewhat larger 7-QZ basis set in the CCSD step leads to the requirements for larger memory (15.2 GB, versus 11.3 GB for CCSD(T)(Full,DKH)/MTSmall) and (especially) disk (336.8 GB, versus 79.8 GB for CCSD(T)-(Full,DKH)/MTSmall).

As noted elsewhere,[7] the W1-F12 procedure substantially reduces the demands on the various computational resources. In the case of $C_6H_5Me^+$, the most demanding component for memory (9.3 GB, versus 15.2 GB for the CCSD/7-QZ step for W1w) and disk (71.1 GB, versus 336.8 GB for the CCSD/7-QZ step for W1w) is CCSD-F12b/VTZ-F12, while the CCSD-(T)(Full)/WCVTZ(−f) step took the longest time of 25.1 CPU hours to complete (versus 111.1 h for the CCSD(T)(Full,DKH)/MTSmall step for W1w). We note that, while there is a significant

Table 4. Comparison of Resource Usage for the Components of W1X-2, W1X-1, W1-F12, and W1w Procedures for $C_6H_5Me^+$

procedure	component[a,b]	CPU time (hour)[c]	memory (GB)	disk (GB)
W1X-2	CCSD(T)(FC)-F12b/A'VDZ	1.2	4.5	13.5
W1X-2	CCSD(T)(FC)-F12b/A'VTZ	20.1	6.4	57.8
W1X-2, W1X-1	MP2(FC)/CVTZ	0.2	0.6	16.2
W1X-2, W1X-1	MP2(Full,DKH)/CVTZ	0.2	0.6	17.3
W1X-1, W1-F12	CCSD(FC)-F12b/VDZ-F12	1.6	4.2	19.6
W1X-1, W1-F12	CCSD(FC)-F12b/VTZ-F12	12.2	9.3	71.1
W1X-1, W1-F12	CCSD(T)(FC)/7-DZ	1.0	1.0	4.0
W1X-1, W1-F12	CCSD(T)(FC)/7-TZ	19.2	6.3	35.0
W1-F12	CCSD(FC)/WCVTZ	4.8	4.8	19.8
W1-F12	CCSD(Full)/WCVTZ	11.4	6.7	32.3
W1-F12	CCSD(T)(FC)/WCVTZ(−f)	9.6	4.1	19.3
W1-F12	CCSD(T)(Full)/WCVTZ(−f)	25.1	6.1	31.0
W1-F12	CCSD(T)(FC,DKH)/7-DZ-DK	1.0	1.0	3.9
W1w	CCSD(T)(FC)/7-DZ	0.9	1.0	4.0
W1w	CCSD(T)(FC)/7-TZ	19.6	6.2	35.0
W1w	CCSD(FC)/7-QZ	50.8	15.2	336.8
W1w	CCSD(T)(FC)/MTS	38.6	10.4	52.8
W1w	CCSD(T)(Full,DKH)/MTS	111.1	11.3	79.8
W1X-2	total	21.8	6.4	57.8
W1X-1	total	34.4	9.3	71.1
W1-F12	total	85.8	9.3	71.1
W1w	total	220.9	15.2	336.8

[a]VnZ = cc-pVnZ, A'VnZ = aug'-cc-pVnZ, CVTZ = cc-pCVTZ, 7-nZ = jul-cc-pV(n+d)Z (= aug'-cc-pV(n+d)Z), WCVTZ = cc-pWCVTZ, WCVTZ(−f) = WCVTZ without f polarization functions, MTS = MTSmall. [b]FC = frozen-core approximation, Full = all-electron calculation. [c]Aggregate CPU time adjusted to a single processor.

savings for the evaluation of the core-correlation component, it still represents the bottleneck in the W1-F12 calculation in terms of CPU time.

The W1X-1 protocol shares the same components for the evaluation of the valence CCSD(T)/CBS limit as W1-F12. These represent substantial savings on the memory, disk, and time requirements when compared with W1w, as discussed above. Furthermore, we note that the MP2 calculations used to obtain scalar-relativistic and core-correlation effects consume a virtually negligible amount of resources, with the MP2-(Full,DKH)/CVTZ step utilizing 0.2 CPU hours, 0.6 GB of memory, and 17.3 GB of disk. For the W1X-2 procedure, the CCSD(T)-F12b/A'VTZ calculation represents the most computationally intense step, in terms of CPU time (20.1 CPU hours), memory (6.4 GB), and disk (57.8 GB) requirements. These represent additional savings in memory and disk when compared with W1X-1, for which the corresponding requirements are 9.3 and 71.1 GB, respectively. In terms of timing comparisons, the use of only two calculations (versus four for W1X-1) for the evaluation of the valence CCSD(T)/CBS contribution leads to a reduction of approximately 36% in the total CPU time used.

Table 5. Comparison of the MADs (kJ mol⁻¹) for the E2 Set (and Its Subsets) and the BDE261 Set for the W1w, W1-F12, W1X-1, and W1X-2 Protocols and the More Economical G4(MP2)-6X and G4(MP2) (Composite) and DuT-D3 (Double-Hybrid DFT) Procedures

test set	property	W1w	W1-F12	W1X-1	W1X-2	G4(MP2) -6X[b]	G4 (MP2)[b]	DuT-D3[c]
E2[a]	E0 + G2/97′ + E1		2.9	2.4	2.7	3.6	4.4	5.9
E0	W4/08 + DBH24 + HB16 + WI9/04	2.7	3.0	2.1	2.9	3.4	4.6	6.5
W4/08	atomization energies	3.7	4.1	2.7	3.8	4.0	4.6	8.5
DBH24	reaction barriers	0.8	0.9	1.3	1.3	3.0	5.7	3.7
HB16	hydrogen bonds	0.4	0.6	0.6	1.0	1.8	3.2	2.0
WI9/04	weak interactions	0.7	0.7	0.7	0.7	1.3	4.1	0.8
G2/97′	G2/97′ ΔH_f + IE + EA + PA	3.1	2.7	2.5	2.6	4.1	4.4	6.2
G2/97′ ΔH_f	heats of formation	4.4	3.2	2.7	3.0	3.0	3.6	4.6
IE	ionization energies	2.7	2.5	2.8	2.7	4.7	4.8	7.7
EA	electron affinities	1.7	2.1	1.9	1.7	5.3	5.3	6.5
PA	proton affinities	1.8	2.1	2.2	2.2	2.8	3.1	4.9
E1	G3/99′ ΔH_f + ADD + ABS + PR8		3.3	2.4	2.6	2.9	4.2	4.7
G3/99′ ΔH_f	heats of formation		4.9	3.5	3.7	3.4	4.1	6.5
ADD	radical addition energies		1.3	1.0	1.4	2.5	2.7	2.0
ABS	radical abstraction energies		1.2	0.8	0.9	1.7	5.8	1.6
PR8	pericyclic reaction barriers		2.1	1.6	1.4	4.4	3.7	6.5
BDE261	bond dissociation energies		1.0	0.9	1.0	4.8	5.9	6.4

[a]Literature[8b] BMK/6-31+G(2df,p) geometries, ZPVE, and ΔH_{298} are used except for the species in the G2/97′ set, for which B3-LYP/cc-pVTZ+d quantities are used. [b]MAD values obtained from ref 8b. [c]Reference 8c.

Assessment on the Larger E2 Test Set. Although the empiricism in the W1X-1 and W1X-2 protocols is minimal with only two fitted parameters, namely the exponents for the extrapolation of the ΔCCSD and Δ(T) energies, it is important to ask, how well do these "slightly-empirical" procedures perform for systems that are not included in the training set? To answer this question, we have evaluated their performance on the E2 set of 526 accurate experimental and theoretical relative energies (see ref 8b and citations therein). The E2 set is comprised of the G2/97 set of thermochemical quantities plus additional ΔH_f's from the G3/99 set, as well as reaction barriers, radical reaction energies, hydrogen-bond energies, and other weak interaction energies. For comparison purposes, we have also obtained W1-F12 values for the entire E2 set and W1w values for the smaller systems in the E2 set, namely the E0 set of energies. In addition, we have also evaluated the performance of these procedures for the BDE261 set of W1w bond dissociation energies (BDEs),[31] which covers a diverse range of 45 bond types and a total of 261 BDEs.

The statistical performance of the W1w, W1-F12, W1X-1, and W1X-2 procedures for the E2 and BDE261 sets is given in Table 5, together with those for the computationally less demanding G4(MP2)[32] and G4(MP2)-6X[8b] composite procedures and the DuT-D3[8c] double-hybrid DFT procedure for comparison. We can see that, while the computationally less demanding methods, in particular the two G4(MP2)-type procedures, already offer very good performance, notable improvements can still be achieved with the W1-type procedures. We can see that all the W1-type methods give MAD values that are less than ~5 kJ mol⁻¹ for all the subsets of E2 that have been assessed. Importantly, the MADs for the W1-type procedures for the subsets are less than 3 kJ mol⁻¹ with the exception of those for atomization energies or ΔH_f, whereas for the more economical G4(MP2)-type and DuT-D3 methods, somewhat larger deviations can be seen for several of the thermochemical properties.

For the W4/08 set of atomization energies, W1X-1 gives an MAD of 2.7 kJ mol⁻¹, while W1w, W1-F12, and W1X-2 yield somewhat larger MAD values of ~4 kJ mol⁻¹. All the W1-type procedures perform quite well for the DBH24 set of barriers, the HB16 set of hydrogen-bond energies, and the WI9/04 set of weak interactions, with W1w and W1-F12 somewhat outperforming W1X-1 and W1X-2. A comparison between W1w and the other W1-type procedures shows that the more economical W1-F12, W1X-1, and W1X-2 methods actually perform somewhat better for the G2/97′ ΔH_f set. They perform comparably to W1w for other properties in the G2/97 set. For G3/99′ ΔH_f, W1-F12 gives a larger MAD when compared with W1X-1 and W1X-2. For the radical reactions (ADD and ABS sets), pericyclic reaction barriers (PR8), and BDEs (BDE261), all of these economical W1-type procedures perform well and comparably to one another. Overall, it appears that the introduction of two parameters in the W1X-1 and W1X-2 protocols does not lead to any significant loss of robustness when compared with W1w and W1-F12 and has the advantage of being less demanding on computational resources.

Barrier Heights for Proton Exchange and Complexation Energies. In a recent study,[33] it was shown that several G4(MP2)-type procedures give rather poor complexation energies for NH_3 and especially H_2O and HF clusters, as well as poor barrier heights for proton exchange within these clusters. Given the uncharacteristically large deviations for the G4(MP2)-type procedures for such seemingly straightforward systems, we feel that it is also important to investigate the performance of W1X-1 and W1X-2 for these properties.

We have obtained W1w and W2w values to use as benchmark complexation energies and barriers, together with the previously reported W1-F12 values. We will focus only on the HF clusters, as they represent some of the most difficult cases, as well as being smaller in size (compared with H_2O and NH_3) so that W1w and W2w calculations are more attainable. We can see from Table 6 that all the economical W1-type procedures, namely W1-F12, W1X-1, and W1X-2, produce complexation energies and barriers that are in reasonable to excellent agreement with the benchmark W2w values. While G4(MP2) understabilizes the aggregates, i.e., the clusters and the

Table 6. Energies of Complexation for $(HF)_n$ Clusters and Barriers for Proton Exchange for These Clusters (kJ mol^{-1})

n	G4(MP2)[a]	W1X-2	W1X-1	W1-F12[a]	W1w	W2w
			complexation energies			
2	−17.9	−19.6	−19.1	−19.2	−19.0	−19.0
3	−56.0	−65.3	−63.7	−63.6	−63.5	−63.3
4	−105.6	−118.8	−116.1	−116.2	−115.9	−115.3
5	−146.4	−161.6	−158.3	−158.4	−158.1	−157.1
6	−180.6	−197.9	−194.0	−194.2	−193.8	
		barriers relative to n isolated HF molecules				
2	165.8	155.5	156.6	157.4	158.4	158.4
3	37.1	18.8	22.0	22.7	23.4	24.0
4	−36.7	−60.1	−55.5	−54.9	−54.4	−52.9
5	−76.8	−103.5	−98.0	−97.6	−97.2	−94.7
6	−101.2	−131.8	−125.3	−124.8	−124.7	
		barriers relative to $(HF)_n$ clusters				
2	183.7	175.1	175.8	176.6	177.4	177.4
3	93.1	84.1	85.7	86.3	86.9	87.3
4	68.9	58.7	60.6	61.3	61.4	62.3
5	69.7	58.1	60.3	60.8	60.9	62.4
6	79.4	66.1	68.7	69.4	69.1	

[a]Reference 33.

transition structures, relative to the isolated HF molecules, W1-F12, W1X-1, and W1X-2 tend to overstabilize these structures. This is most apparent for W1X-2, where the largest deviation from the benchmark value is −8.8 kJ mol^{-1} for the $(HF)_5$ transition structure. The corresponding deviations for W1X-1 and W1-F12 are −3.3 and −2.9 kJ mol^{-1}, respectively.

In addition to listing the proton-exchange barriers relative to isolated HF molecules, we have also calculated the barriers relative to the energies of the corresponding $(HF)_n$ clusters. This approach benefits from some systematic cancellation of errors, such that the deviations for the $(HF)_5$ transition structure are reduced to −4.3 (W1X-2), −2.1 (W1X-1), and −1.6 (W1-F12) kJ mol^{-1}, respectively. Nonetheless, since the deviations for barriers are generally larger than those for the corresponding complexation energies, the cancellation is by no means perfect. In addition, there is a general trend for the deviation to grow with the number of HF molecules in the cluster. Therefore, we caution against the use of the W1X-2 procedure for the study of medium-sized to large HF (and by inference, H_2O) clusters, if highly accurate energies are required. For these systems, the W1X-1 and W1-F12 procedures appear to represent a better compromise between computational cost and accuracy.

Application to Larger Systems. We now probe the applicability of W1X-1 and W1X-2 to systems that are larger than those found in the E2 and BDE261 sets. Thus, we apply these procedures, where feasible, to two sets of chemical systems. These include the set of 11 pericyclic reaction barriers of Houk and co-workers,[34] of which eight smaller systems are included in the PR8 subset of the E2 set, and the set of ΔH_f's of 14 medium-sized molecules of Curtiss et al.[35] We have included 10 out of the 11 pericyclic reaction barriers and the ΔH_f's of all 14 medium-sized molecules in our analysis, and we will refer to the pericyclic reaction barriers and the medium-sized molecule ΔH_f's as the PR10 and MS14 sets, respectively. Where appropriate, we use the W1X-1 and W1X-2 results to suggest improved theoretical benchmark values for some of these larger systems.

For the barriers in the PR10 set, previously reported W1 values[36] are used as the best theoretical references for the eight smaller reactions in the PR10 set (equivalent to the PR8 subset of the E2 set), while CBS-QB3 values have been previously obtained for 10 reactions.[34] In the present study, we are able to calculate W1X-1 and W1X-2 barriers for these same 10 reactions. In addition, we have obtained W1-F12 values for the eight smaller reactions. A comparison of the W1, W1-F12, W1X-1, W1X-2, CBS-QB3, and experimental values is shown in Table 7.

We can see that the W1-type procedures give barriers that are generally comparable with one another. The largest difference can be seen for reaction 10, where the W1X-1 and W1X-2 values differ by 5.5 kJ mol^{-1}. While the agreement between the various W1-type barriers and the experimental values is typically adequate, relatively large discrepancies are found in a number of cases. Notably, for reactions 1, 3, 5, and 8, the differences between theory and experiment are larger than 5 kJ mol^{-1} for all the W1-type procedures. This may be indicative of larger uncertainties for the experimental barriers than those quoted. For reaction 9 where the W1 barrier is not available, the W1X-1 (63.6 kJ mol^{-1}) and W1X-2 (62.3 kJ mol^{-1}) values are reasonably close to one another, and we propose the W1X-1 value of 63.6 kJ mol^{-1} as our highest-level theoretical benchmark value. It represents a significant improvement over the CBS-QB3 value of 45.5 kJ mol^{-1}. Likewise, for reaction 10 we note that the W1X-1 barrier of 100.5 kJ mol^{-1} represents a significant improvement over the CBS-QB3 value of 90.8 kJ mol^{-1}, relative to the experimental 102.9 kJ mol^{-1}.

We now turn our attention to the MS14 set of medium-sized molecules (Table 8).[37,38] For this set, we have obtained G4(MP2)-type[8a,b,32] ΔH_f values for all the species, and W1X-1

Table 7. Comparison of Theoretical and Experimental Barriers for the PR10 Set of Pericyclic Reactions (298 K, kJ mol^{-1})

		CBS-QB3[a]	W1X-2	W1X-1	W1-F12	W1[a]	exptl[b]
1	cyclobutene → cis-butadiene	134.1	139.2	138.9	137.5	140.9	133.5 ± 0.8
2	cis-1,3,5-hexatriene →1,3-cyclohexadiene	117.5	122.5	122.5	123.0	126.4	124.7 ± 2.1
3	o-xylylene → benzocyclobutene	103.3	110.7	110.8	110.9	113.3	118.4 ± 1.3
4	cis-1,3-pentadiene sigmatropic shift	150.1	151.8	151.3	153.7	152.9	148.1 ± 2.1
5	cyclopentadiene sigmatropic shift	108.2	109.9	110.3	111.1	108.5	98.7 ± 2.1
6	cis-1,5-hexadiene sigmatropic shift	133.5	144.9	140.7	140.8	143.3	140.2 ± 2.1
7	ethene + cis-butadiene → cyclohexene	90.7	94.0	94.0	93.6	96.7	101.3 ± 8.4
8	ethene + cyclopentadiene Diels−Alder reaction	67.6	78.2	79.0	78.0	80.8	94.6 ± 6.7
9	cyclopentadiene dimerization	45.5	62.3	63.6			63.6 ± 2.5
10	cis-triscyclopropanocyclohexane →1,4,7-cyclononatriene	90.8	106.0	100.5			102.9 ± 12.6

[a]Obtained from vibrationless values quoted in ref 34b, adjusted to 298 K enthalpic barriers using B3-LYP/cc-pVTZ+d frequencies. [b]Reference 36.

4265

dx.doi.org/10.1021/ct300632p | J. Chem. Theory Comput. 2012, 8, 4259−4269

Table 8. Comparison of Theoretical and Experimental Heats of Formation for the MS14 Set of Medium-Sized Molecules (298 K, kJ mol^{-1})

		G4(MP2)	G4(MP2)-5H	G4(MP2)-6X	W1X-2	W1X-1	exptl[a]
C$_6$F$_6$		−949.8	−950.0	−953.0	−963.8	−957.4	−955.42
C$_6$F$_5$Cl		−807.2	−807.3	−808.3	−815.6	−810.1	−812.28
dodecane	C$_{12}$H$_{26}$	−288.8	−289.0	−291.4			−289.70
hexadecane	C$_{16}$H$_{34}$	−368.9	−369.2	−372.2			−374.80
adamantane[c]	C$_{10}$H$_{16}$	−141.4	−141.6	−143.9	−139.7	−138.0	−137.90[b]
diadamantane	C$_{14}$H$_{20}$	−178.5	−178.7	−182.3			−145.90
pyrene	C$_{16}$H$_{10}$	206.2	206.0	215.8	221.5	221.4	225.68
fluoranthene	C$_{16}$H$_{10}$	264.5	264.3	273.4			288.99
anthracene	C$_{14}$H$_{10}$	213.3	213.1	221.7	227.3	226.8	230.91
phenazine	C$_{12}$H$_8$N$_2$	321.4	321.2	328.1	330.7	330.1	343.80[c]
trans-azobenzene	C$_{12}$H$_{10}$N$_2$	397.0	396.8	402.6	408.0	406.8	404.09
benzophenone	C$_{13}$H$_{10}$O	44.4	44.2	49.2	54.6		54.89
dibenzothiophene	C$_{12}$H$_8$S	188.6	188.4	197.3	208.6	207.7	205.10
3,6-diphenyl-1,2-dithiin[c]	C$_{16}$H$_{12}$S$_2$	355.3	355.1	369.6			422.42

[a]Experimental values obtained from ref 35 unless otherwise noted. [b]Reference 38. [c]See text for discussion.

and W1X-2 values for the smaller molecules within the set. We note that the experimental values all have quoted uncertainties of <1 kcal (4.2 kJ mol^{-1}). We can see that the G4(MP2)-type procedures sometimes yield heats of formation that differ significantly from experimental results. The largest discrepancies are found for 3,6-diphenyl-1,2-dithiin, with G4(MP2)[32] and G4(MP2)-5H[8a] giving deviations of ∼−67 kJ mol^{-1}, and the G4(MP2)-6X[8b] value differing from experimental values by −52.9 kJ mol^{-1}. Among the three G4(MP2)-type procedures, we find that, in general, G4(MP2)-6X provides the best agreement with the experimental ΔH$_f$ values.

While it was not feasible to do W1-type calculations for diadamantane and 3,6-diphenyl-1,2-dithiin, for which the deviations of G4(MP2)-type ΔH$_f$'s from experimental values are the largest (Table 8), we have obtained ΔH$_f$ values using the higher-level G4[39] and ccCA[2c] procedures. We find that the G4 ΔH$_f$'s for both diadamantane (−178.9 kJ mol^{-1}) and 3,6-diphenyl-1,2-dithiin (358.0 kJ mol^{-1}) are similar to those obtained with the G4(MP2)-type methods, but the ccCA method yields values (−155.0 and 389.1 kJ mol^{-1}, respectively, for diadamantane and 3,6-diphenyl-1,2-dithiin) that are closer to experimental values (−145.9 and 422.4 kJ mol^{-1}, respectively). For diadamantane, we have further obtained a ΔH$_f$ value using the ccCA energy for the isodesmic reaction: diadamantane + cyclohexane → 2 adamantane, together with experimental ΔH$_f$'s for other species.[40] This yields our best-estimated theoretical value of −168.0 kJ mol^{-1}. This value lies between the ccCA ΔH$_f$ and the values obtained with G4 and G4(MP2)-type procedures, but it deviates substantially from the experimental ΔH$_f$ of −145.90 kJ mol^{-1}. Thus, while there may be deficiencies within the G4 and G4(MP2)-type methods that lead to the large deviations from experimental values, the uncertainty associated with the experimental value of diadamantane may also be larger than the one quoted. For 3,6-diphenyl-1,2-dithiin, we note that there is also a large difference between the ccCA ΔH$_f$ (389.1 kJ mol^{-1}) and the experimental value (422.42 kJ mol^{-1}).

The two W1-type procedures generally give ΔH$_f$ values that are in close agreement with one another, and we find that W1X-1 usually agrees with experimental values somewhat more closely than does W1X-2. The largest disagreements between these two W1-type procedures are seen for C$_6$F$_6$ and C$_6$F$_5$Cl, for which the W1X-2 values are more negative than the corresponding W1X-1 values by 6.4 and 5.5 kJ mol^{-1}, respectively.

This may reflect a tendency for W1X-2 to overstabilize molecules containing multiple C−F bonds.[41] It is intriguing that this overstabilization of highly fluorinated molecules is consistent with the observation that W1X-2 also tends to overstabilize HF clusters and proton-exchange transition structures (Table 6). Finally, we note that the W1X-1 and W1X-2 ΔH$_f$'s (330.1 and 330.7 kJ mol^{-1}, respectively) for phenazine agree well with one another but deviate from the experimental value by ∼13 kJ mol^{-1}. Examination of the NIST Chemistry Webbook[42] shows that several experimental values are provided, and we find that our theoretical values are closer to the value obtained by Leitao et al. (328.8 ± 2.9 kJ mol^{-1}).[43]

■ DESCRIPTION OF THE W1X-1 AND W1X-2 PROTOCOLS

In the previous sections, we have demonstrated that the W1X-1 and W1X-2 procedures can provide a cost-effective means for obtaining accurate thermochemistry. For convenience, we now briefly summarize the procedures for carrying out W1X-1 and W1X-2 calculations.

- For both W1X-1 and W1X-2, geometries and vibrational frequencies are obtained at the B3-LYP/cc-pVTZ+d level. A scale factor of 0.985 is applied in the evaluation of ZPVEs and ΔH$_{298}$.
- For W1X-1, the valence CCSD(T)/CBS component is obtained as follows:
  ○ The HF/CBS energy is obtained by extrapolating the HF(CABS)/cc-pVDZ-F12 and HF(CABS)/cc-pVTZ-F12 energies,[44] where HF(CABS) is the HF energy with a complementary auxiliary basis singles correction. The extrapolation is carried out using the formula $E_L = E_{CBS} + AL^{-5}$.
  ○ The ΔCCSD/CBS component is obtained by extrapolating the CCSD-F12b/cc-pVDZ-F12 and CCSD-F12b/cc-pVTZ-F12 correlation energies using the formula $E_L = E_{CBS} + AL^{-3.67}$.[15,44]
  ○ The Δ(T) portion of the CCSD(T) correlation energy, i.e., $E_{CCSD(T)} − E_{CCSD}$, is obtained with the aug′-cc-pV(D+d)Z and aug′-cc-pV(T+d)Z basis sets and extrapolated using the formula $E_L = E_{CBS} + AL^{-2.04}$. The aug′ description denotes the inclusion of diffuse functions only for p-block elements.[14,15]

- For W1X-2, the valence CCSD(T)/CBS term is the sum of the following components:
  - An HF/CBS energy is obtained by extrapolating the HF(CABS)/aug′-cc-pVDZ and HF(CABS)/aug′-cc-pVTZ energies using the formula $E_L = E_{CBS} + AL^{-5}$.
  - A ΔCCSD/CBS component is obtained with CCSD-F12b using the aug′-cc-pVDZ and aug′-cc-pVTZ basis sets and extrapolated using the formula $E_L = E_{CBS} + AL^{-4.74}$.
  - A Δ(T)/CBS component is calculated using the (T) correlation energies of CCSD(T)-F12b/aug′-cc-pVDZ and CCSD(T)-F12b/aug′-cc-pVTZ and extrapolated using the formula $E_L = E_{CBS} + AL^{-2.09}$.
- For both W1X-1 and W1X-2, the combined core-correlation and scalar-relativistic component $[\Delta(C+R)]$ is taken as the difference between the frozen-core and all-electron-DKH MP2/cc-pCVTZ energies.

CONCLUDING REMARKS

In the present study, we have examined a number of approaches for reducing the computational requirements for the W1w and the recently proposed,[7] already more economical, W1-F12 procedures, while maintaining the accuracy. A number of key results emerge from our investigation:

1. The evaluation of core-correlation effects represents the most time-consuming step for both W1w and W1-F12. In terms of memory and disk requirements, the CCSD/cc-pVQZ calculation is the most demanding for W1w, while for W1-F12 it is the CCSD-F12b/cc-pVTZ-F12 calculation.

2. We have examined a number of alternative approaches for estimating the core-correlation energy. We find that MP2/cc-pCVTZ provides reliable results at a cost that is negligible in the context of W1-type procedures.

3. For the evaluation of the valence CCSD(T)/CBS energy, we find that truncation of the set of diffuse functions leads to a notable savings in CPU time, with only a minor deterioration in the performance. However, it is the CCSD/QZ step that represents an arguably unnecessary bottleneck, and the truncations are therefore not implemented in the W1X procedures.

4. The previously proposed W1-F12 protocol makes substantial progress in reducing the requirements for memory and disk, as well as computer time, by using the CCSD-F12b method to eliminate the use of a QZ-type basis set. We have adopted this approach in the W1X-1 procedure proposed in the present study. Furthermore, the W1X-1 method utilizes the cost-effective MP2/cc-pCVTZ procedure for the evaluation of core-correlation energy and is thus significantly more economical than both W1w and W1-F12.

5. The W1X-1 method, like W1-F12, employs four coupled-cluster calculations for the evaluation of the CCSD(T)/CBS energy. We have formulated the even more economical W1X-2 procedure by basing the CCSD(T)/CBS energy on only two CCSD(T)-F12b calculations.

6. We have evaluated the performance of the W1X-1 and W1X-2 procedures for a diverse range of thermochemical quantities. These include the E2 set of 526 relative energies, the BDE261 set of 261 bond dissociation energies, com-

plexation energies of HF clusters and the corresponding proton-exchange barriers, barriers for pericyclic reactions, and heats of formation of medium-sized molecules. We find that W1X-1 performs well and comparably to the W1w and W1-F12 procedures in virtually all these cases. The W1X-2 method gives equivalently good results for most properties with the notable exception of those for highly fluorinated molecules and hydrogen fluoride clusters.

ASSOCIATED CONTENT

S **Supporting Information**

Instructions for carrying out W1X-1 and W1X-2 calculations, example Molpro input and output files, incorporated into a single zip archive (w1x-n.zip); reactions in the DBH24, RR49, and PR10 sets (Table S1); zero-point vibrational energies (ZPVEs), enthalpy corrections at 298 K (ΔH_{298}), and W1w, W1-F12, W1X-1, and W1X-2 total electronic energies (Table S2); deviations from benchmark values (Table S3); and MADs (kJ mol^{-1}) for W1X-1 and W1X-2 using different basis sets for the evaluation of the $\Delta(C+R)$ term (Table S4). This material is available free of charge via the Internet at http://pubs.acs.org

AUTHOR INFORMATION

Corresponding Author
*E-mail: chan_b@chem.usyd.edu.au; radom@chem.usyd.edu.au.
Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) (a) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910. (b) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(2) For recent reviews, see for example: (a) Petersson, G. A. In *Quantum-Mechanical Prediction of Thermochemical Data*; Cioslowski, J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; pp 99−130. (b) Martin, J. M. L. *Ann. Rep. Comput. Chem.* **2005**, *1*, 31. (c) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. *Prog. Theor. Chem. Phys.* **2009**, *19*, 197. (d) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *WIREs: Comput. Mol. Sci.* **2011**, *1*, 810. (e) Peterson, K. A.; Feller, D.; Dixon, D. A. *Theor. Chem. Acc.* **2012**, *131*, 1079.

(3) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843.

(4) Chan, B.; Morris, M.; Radom, L. *Aust. J. Chem.* **2011**, *64*, 394.

(5) Martin, J. M. L. *THEOCHEM* **2006**, *771*, 19.

(6) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kallay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129.

(7) Karton, A.; Martin, J. M. L. *J. Chem. Phys.* **2012**, *136*, 124114.

(8) (a) Chan, B.; Coote, M. L.; Radom, L. *J. Chem. Theory Comput.* **2010**, *6*, 2647. (b) Chan, B.; Deng, J.; Radom, L. *J. Chem. Theory Comput.* **2011**, *7*, 112. (c) Chan, B.; Radom, L. *J. Chem. Theory Comput.* **2011**, *7*, 2852.

(9) See for example: (a) Hehre, W. J.; Radom, L.; Schleyer, P. v. P.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986. (b) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley: New York, 2001. (c) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; Wiley: Chichester, U. K., 2007.

(10) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.

P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.

(11) Werner, H.-J.; Knowles, P. J.; Manby, F. R.; Schütz, M.; Celani, P.; Knizia, G.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklaß, A.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO 2010.1*; University College Cardiff Consultants Limited: Cardiff, U. K., 2010.

(12) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

(13) (a) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063. (b) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.

(14) Standard cc-pVnZ basis sets are used for H, He, Li, and Be. For Na and Mg, cc-pCVnZ basis sets are utilized, as the inclusion of inner-valence functions in the basis sets, together with correlating inner-valence orbitals, can lead to significantly improved accuracy in post-Hartree−Fock treatments. See: Sullivan, M. B.; Iron, M. A.; Redfern, P. C.; Martin, J. M. L.; Curtiss, L. A.; Radom, L. *J. Phys. Chem. A* **2003**, *107*, 5617.

(15) The use of core-correlation basis sets, e.g., cc-pCVnZ, for Na and Mg implies the inclusion of inner-valence electrons in the valence-correlation treatment for these two elements, whereas the use of standard cc-pVnZ or cc-pVnZ-F12 basis sets for Na and Mg implies the use of the standard chemical core in the evaluation of valence-correlation energy. The associated core-correlation treatment in the composite procedure is then adjusted accordingly.

(16) Martin, J. M. L.; Sundermann, A.; Fast, P. L.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *113*, 1348.

(17) Martin, J. M. L.; Parthiban, S. In *Quantum-Mechanical Prediction of Thermochemical Data*; Cioslowski, J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; pp 31−66.

(18) (a) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 1293. (b) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703. (c) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105.

(19) Papajak, E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 10.

(20) For recent reviews, see for example: (a) Klopper, W.; Manby, F. R.; Ten-No, S.; Valeev, E. F. *Int. Rev. Phys. Chem.* **2006**, *25*, 427. (b) Shiozaki, T.; Valeev, E. F.; Hirata, S. *Ann. Rep. Comput. Chem.* **2009**, *5*, 131. (c) Werner, H.-J.; Knizia, G.; Adler, T. B.; Marchetti, O. *Z. Phys. Chem.* **2010**, *224*, 493. (d) Kong, L.; Bischoff, F. A.; Valeev, E. F. *Chem. Rev.* **2012**, *112*, 75.

(21) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.

(22) We have briefly investigated the use of several basis sets of triple-ζ quality, namely G3Large, cc-pCVTZ, cc-pWCVTZ, and aug′-cc-pCVTZ, for the evaluation of Δ(C+R) (Supporting Information, Table S4). We find that, with the exception of G3Large, which leads to noticeably larger deviations from the benchmark, all three Dunning-type basis sets yield very similar statistics in the context of W1-type procedures for the overall G2/97 set. The performance of the three Dunning-type basis sets improves, very slightly (with improvements in the MADs < 0.1 kJ mol^{-1}), in the order cc-pWCVTZ → cc-pCVTZ → aug′-cc-pCVTZ. We find that, between the two nonaugmented basis

sets, cc-pCVTZ gives a somewhat more balanced performance across the four subsets of G2/97 than does cc-pWCVTZ. We also find that these two basis sets perform comparably well for both first-row-only systems and molecules that contain one or more second-row atoms. As the Δ(C+R) term is used to account for scalar-relativistic as well as core-correlation effects, we have also examined the use of the cc-pCVTZ-DK basis set. We find that the use of this basis offers no apparent improvement in performance for the G2/97 test set. Taking into account all of the above considerations, we have adopted cc-pCVTZ in preference to the somewhat better performing aug′-cc-pCVTZ basis set because of its substantially less-demanding requirement on scratch disk space.

(23) See for example: (a) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175. (b) Hättig, C. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59. (c) Hellweg, A.; Hättig, C.; Höfener, S.; Klopper, W. *Theor. Chem. Acc.* **2007**, *117*, 587. (d) Katouda, M.; Nagase, S. *Int. J. Quantum Chem.* **2009**, *109*, 2121.

(24) Mintz, B.; Lennox, K. P.; Wilson, A. K. *J. Chem. Phys.* **2004**, *121*, 5629.

(25) (a) Adler, T. B.; Knizia, G.; Werner, H.-J. *J. Chem. Phys.* **2007**, *127*, 221106. (b) Knizia, G.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.

(26) Marchetti, O.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3400.

(27) Peterson, K. A.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2008**, *128*, 084102.

(28) The inclusion of tight-d functions in the basis set has been shown to be often important for highly polar bonds that involve second-row p-block elements. Thus, the exclusion of tight-d functions in the W1X-2 procedure can potentially lead to a deterioration in performance in these cases. A classic example of such a basis set effect is represented by SO_2,[29] which is also contained in our E0 test set. In this case, the deviations from the benchmark SO_2 atomization energy for the various W1-type procedures are −6.5 (W1w), −7.2 (W1-F12), −7.1 (W1X-1), and −6.2 (W1X-2) kJ mol^{-1}. Likewise, for SO_3, the deviations are −7.4 (W1w), −9.1 (W1-F12), −10.5 (W1X-1), and −9.7 (W1X-2) kJ mol^{-1}; and for ClO, the deviations are −6.2 (W1w), −5.8 (W1-F12), −3.9 (W1X-1); and −5.2 (W1X-2) kJ mol^{-1}.

(29) Martin, J. M. L. *J. Chem. Phys.* **1998**, *108*, 2791.

(30) Hill, J. G.; Peterson, K. A.; Knizia, G.; Werner, H.-J. *J. Chem. Phys.* **2009**, *131*, 194105.

(31) Chan, B.; Radom, L. *J. Phys. Chem. A* **2012**, *116*, 4975.

(32) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105.

(33) Karton, A.; O'Reilly, R. J.; Chan, B.; Radom, L. *J. Chem. Theory Comput.* **2012**, *8*, 3128.

(34) (a) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 11445. (b) Ess, D. H.; Houk, K. N. *J. Phys. Chem. A* **2005**, *109*, 9542.

(35) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.

(36) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.

(37) For adamantane in the MS14 set, the ΔH_f value employed in ref 35 is −134.4 ± 2.3 kJ mol^{-1}. In the present study, we utilize the experimental ΔH_f of adamantane to obtain our best theoretical estimate for the ΔH_f of diadamantane. Therefore, we have adopted an alternative value of −137.9 ± 0.79 kJ mol^{-1} of ref 38 for adamantane, owing to its quoted sub-kJ·mol^{-1} uncertainty.

(38) Mansson, M.; Rapport, N.; Westrum, E. F., Jr. *J. Am. Chem. Soc.* **1970**, *92*, 7296.

(39) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 084108.

(40) For this reaction, we have evaluated the sensitivity of the calculated reaction enthalpy to the level of theory used to obtain it. The values calculated at the various levels are 12.2 [HF/6-31G(d)], 12.3 [B3-LYP/6-31G(d)], 16.5 [MP2/aub-cc-pVQZ], 17.5 [G4(MP2)], 17.6 [G4(MP2)-6X], 17.9 [G4], and 15.3 [ccCA] kJ mol^{-1}. Both supplementary species involved in the calculation of the diadamantane ΔH_f using this reaction have experimental ΔH_f's with

sub-kJ mol^{-1} quoted uncertainties (cyclohexane, -123.1 ± 0.79, and adamantane, -137.9 ± 0.79 kJ mol^{-1}).

(41) We have further probed the relative behavior of W1X-1 and W1X-2 for the calculation of the ΔH_f's for several other fluorocarbons, namely CF_4, C_2F_6, and C_3F_8. We find that the W1X-2 values for CF_4 (-941.0 kJ mol^{-1}), C_2F_6 (-1354.9 kJ mol^{-1}), and C_3F_8 (-1770.6 kJ mol^{-1}) are consistently more negative than the respective W1X-1 values of -937.0, -1347.6, and -1759.3 kJ mol^{-1}. We also note that the difference between W1X-1 and W1X-2 grows with the size of the fluorocarbon molecules. The experimental values[42] are -930 ± 20 (CF_4), -1343.9 (C_2F_6), and -1784.7 ± 8.8 (C_3F_8) kJ mol^{-1}. We also note that the original W1 protocol also exhibits certain difficulties with molecules that contain highly polar bonds, which were attributed to the slow convergence in the AVnZ basis sets.[21]

(42) *NIST Chemistry WebBook*; Linstrom, P. J.; Mallard, W. G., Eds.; NIST Standard Reference Database Number 69; National Institute of Standards and Technology, Gaithersburg, MD, 2011. http://webbook.nist.gov (accessed September 2012).

(43) Leitao, M. L.; Pilcher, G.; Acree, W. E., Jr.; Zvaigzne, A. I.; Tucker, S. A.; Ribeiro Da Silva, M. D. M. C. *J. Chem. Thermodyn.* **1990**, *22*, 923.

(44) The cc-pCVnZ-F12 basis sets are used for Na and Mg. These basis sets are designed to recover core-correlation effects together with the F12-type procedures. See: Hill, J. G.; Mazumder, S.; Peterson, K. A. *J. Chem. Phys.* **2010**, *132*, 054108.