Extrapolation of High-Order Correlation Energies: The WMS Model

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-08-2018-004973.R1
Article Type:	Paper
Date Submitted by the Author:	10-Oct-2018
Complete List of Authors:	Zhao, Yan ; State Key Laboratory of Advanced Technology for Materials Synthesis and Processing, International School of Materials Science and Engineering, Wuhan University of Technology Xia, Lixue; Wuhan University of Technology, International School of Materials Science and Engineering Liao, Xiaobin; Wuhan University of Technology, State Key Laboratory of Advanced Technology for Materials Synthesis and Processing He, Qiu; Wuhan University of Technology, State Key Laboratory of Advanced Technology for Materials Synthesis and Processing Zhao, Maria; Lynbrook High School Truhlar, Donald; University of Minnesota System, Chemistry

Extrapolation of High-Order Correlation Energies: The WMS Model†

Yan Zhao,[a] Lixue Xia,[a] Xiaobin Liao,[a] Qiu He,[a] Maria X. Zhao,[b] and Donald G. Truhlar[c]

Abstract. We have developed a new composite model chemistry method called WMS (Wuhan-Minnesota scaling method) with three characteristics: (1) a composite scheme to approximate the complete configuration interaction valence energy with the affordability condition of requiring no calculation more expensive than CCSD(T)/jul-cc-pV(T+d)Z, (2) low-cost methods for the inner-shell correlation contribution and scalar relativistic correction, (3) accuracy comparable to methods with post-CCSD(T) components. The new method is shown to be accurate for the W4-17 database of 200 atomization energies with an average mean unsigned error (averaged with equal weight over strongly correlated and weakly correlated subsets of the data) of 0.45 kcal/mol, and the performance/cost ratio of these results compares very favorably to previously available methods. We also assess the WMS method against the DBH24-W4 database of diverse barrier heights and the energetics of the reactions of three strongly correlated Criegee intermediates with water. These results demonstrate that higher-order correlation contributions necessary to obtain high accuracy for molecular thermochemistry may be successfully extrapolated from the lower-order components of CCSD(T) calculations, and chemical accuracy can now be obtained for larger and more complex molecules and reactions.

---

[a]*State Key Laboratory of Silicate Materials for Architectures,*
International School of Materials Science and Engineering
Wuhan University of Technology, Wuhan 430070, People's Republic of China.
Email: yan2000@whut.edu.cn

[b]*Lynbrook High School, 1280 Johnson Avenue, San Jose, California 95129*

[c]*Department of Chemistry and Supercomputing Institute, University of Minnesota, 207*
Pleasant Street S.E. Minneapolis, MN 55455-0431
Email: truhlar@umn.edu

---

† Electronic supplementary information (ESI) available: Calculated barrier heights for the DBH24-W4 database; RMSEs of the basis-set-extrapolated Hartree Fock, valence CCSD, valence (T), and HOVC energies, the scalar relativistic contributions, the core-valence correction energies, and the composite methods for the W4-17 database; the calculated Born-Oppenheimer atomization energies for the W4-17 database; examples of input files; and perl scripts.

1. Introduction

The first prerequisite for accurate simulation and modeling of chemical processes is to get the energetics right, and for many species involved in modeling catalysis, atmospheric and environmental chemistry, or combustion, one can obtain more accurate energetics from quantum mechanical calculations than from experiment. Density functional theory is very useful for large and complex systems, but wave function theory (WFT) is often more accurate for smaller systems.[1] However, the cost of the most accurate WFT methods increases as a high power of system size, so the development of lower-cost, but still highly accurate WFT methods is an important goal for extending this capability to larger systems for new practical applications.

Here we propose to extend this capability by using the coupled cluster (CC) method with higher-order contributions obtained by extrapolation. Coupled cluster calculations are based on excitations from a reference function and must be converged with respect to the excitation level (double, triple, quadruple… excitations from a reference function) and with respect to the one-electron basis set. Including all excitation levels for a given one-electron basis is called full configuration interaction (FCI), and carrying out FCI for a complete basis set (CBS) is called complete configuration interaction (CCI). For systems with more than a very few electrons, brute-force FCI and CCI are unaffordable, and one must extrapolate to reach CCI. Many methods are available to try to extrapolate to the CCSD(T)/CBS limit, where CCSD(T) denotes CC with single and double excitations and a quasiperturbative treatment of connected triple excitations, and CBS denotes a complete one-electron basis. Although extrapolation (either approximately or accurately) to a complete one-electron basis is now standard, extrapolation to include higher excitation levels is much less developed. Here we propose a scheme for the latter.

Extrapolation methods are based on a set of calculations at various levels and/or with various basis sets, and for that reason they are often called composite methods. An important component of many composite methods is second-order Møller-Plesset perturbation theory[2] (MP2) because of its low cost and the fact that it involves only double excitations, but one must also include higher-order contributions for good accuracy. Extrapolation methods based on single-configuration reference functions were pioneered by Petersson[3] and composite methods based on single-configuration reference functions were brought to a Nobel-Prize level of quality by Pople and coworkers.[4] Pople[4] also emphasized that an approximate procedure should be precisely formulated, and a precisely formulated procedure is called a theoretical model chemistry or – for short – a model chemistry. Model chemistries have the advantage that they can be unambiguously validated.

For the last two decades, many composite WFT model chemistries have been developed to approximate the CCSD(T)/CBS limit, including the complete basis set (CBS-*n*) model chemistries of Petersson, Radom, and coworkers,[5-8] the G*n* composite methods of Pople, Curtiss, Raghavachari, and coworkers,[9-13] the multicoefficient correlation methods (MCCMs),[14-16] the lower-order Weizmann methods (W1 and W2[17-19] and W1-F12 and W2-F12[20]) of Martin and coworkers, and the correlation-consistent composite approach (ccCA) methods of Wilson and coworkers.[28] The CCSD(T)/CBS limit is accurate to about 0.4 kcal/mol for reaction energies and barrier heights when a single-configuration reference function is adequate.[21]

Some composite WFT methods (sometimes called post-CCSD(T) methods) also include higher-order-than-CCSD(T) calculations either to give higher accuracy in general or to give better descriptions of multireference systems. Multireference systems, sometimes called strongly correlated systems, are systems with near-degeneracy correlation effects such that the nearly degenerate configurations should be included in the zero-order description; in contrast systems for which a single configuration state function can serve as a good reference function are called single-reference systems or weakly correlated systems. Simple examples of multi-reference systems are B_2, BN, C_2, O_3, FOOF, and S_4. The CC theory with a single-configuration reference function is not adequate for such systems unless one includes very high-level excitations, typically full triple excitations and at least quasiperturbative quadruple excitations. Examples of post-CCSD(T) composite methods are the W3/W4 Weizmann methods,[22, 23] the High Accuracy Extrapolated ab initio Thermochemistry (HEAT) method of Stanton and coworkers,[24] and the Feller-Dixon-Peterson (FPD) composite approach.[25] An important aspect of very recent work is trying to reduce the cost of post-CCSD(T) composite methods as in the W3X-L method of Radom and coworkers[26] and the diet-HEAT-F12 method of Csontos and coworkers.[27] The present article continues in this vein, i.e., the development of less expensive but still accurate post-CCSD(T) methods.

The problem of the slow convergence of the conventional correlation methods with respect to basis set size is due to the poor description of the cusp in the many-electron wave functions when two electrons come together; products of the usual one-electron Gaussian basis functions cannot reproduce this cusp condition easily. One needs to use basis functions with very high angular momentum (for example, those in a basis set larger than correlation-consistent quadruple zeta) to converge the correlation energy by conventional methods. Unfortunately even MP2 calculations with the necessary one-electron basis sets have been shown to be a bottleneck of the ccCA model for molecules of moderate size.[28] The early work of Hylleraas on the helium atom[29] showed that the use of explicitly correlated wave functions, i.e., wave functions containing terms that depend explicitly on the interelectronic distances r_{ij} between electrons i and j, can achieve rapid convergence of correlated WFT calculations with respect to the size of the basis set. Building on advances in resolution of the identity (RI) methods[30, 31] and robust density fitting (DF) approximations for evaluating integrals[32, 33] and on the proposal of correlation factor (F12) methods,[34-37] the explicitly correlated CCSD(T)-F12-type models[38, 39] have been shown to provide rapid basis-set convergence of the CCSD(T) method, which has opened the possibility that one can extrapolate to the CBS limit based on calculations with smaller basis sets than is possible without explicit correlation.

The objective of the present work is to develop a new composite model chemistry method, to be called WMS, that has three features: (1) a separate extrapolation scheme to obtain the CCSD(T)/CBS valence correlation energy using the CCSD(T)-F12b method with the constraint of using only double zeta and triple zeta basis sets – not quadruple zeta or larger (see Table 1 for the notation and references of the basis sets employed in the new methods); (2) parametrization to implicitly extrapolate the higher-order valence correlation energy by using the MP2/CBS, CCSD/CBS, and CCSD(T)/CBS correlation energies; and (3) validated low-cost methods for the inner-shell correlation contribution and scalar relativistic corrections. Because the method is optimized against ab initio calculations including high-order correlation components at the nearly complete basis set limit, the optimized parameters implicitly include both one-electron CBS extrapolation and many-electron higher-order-correlation extrapolation.

This paper is organized as follows: Section 2 presents the details of the data that are employed in this study and references for the methods to which we compare, and Section 3 describes key computational details. The new methods are presented in Section 4. Section 5 gives the results and discussion. Section 6 concludes the paper.

2. Tests

2.1 Test Data. The present paper is concerned with Born-Oppenheimer energies E in the ground electronic state; note that E is the electronic energy plus nuclear repulsion at a fixed geometry, and it corresponds to the potential energy function for internuclear motion. It includes relativistic effects, but it does not include zero point energy or thermal energy. We consider three kinds of comparison to experimental observables: equilibrium atomization energies (labeled D_e), classical barrier heights (labeled $E_f^\ddagger$ for the forward reaction and $E_r^\ddagger$ for the reverse reaction), and classical energies of reaction (labeled ΔE). The dissociation energy D_e is the sum of the energies of the dissociated atoms minus the energy of the molecule at its equilibrium geometry. The classical barrier height is the energy of the transition structure (i.e., the saddle point along the lowest-energy reaction path for a chemical reaction) minus the energy of the reactant (for a unimolecular reaction) or minus the sum of the energies of the reactants (for a bimolecular reaction). The energy of a reaction is sum of the D_e values of reactants minus the sum of the D_e values of products.

For atomization energies we obtained best estimates needed to test the new method by subtracting Born-Oppenheimer corrections from the W4-17 database[40] of 200 non-Born-Oppenheimer D_e values. W4-17 is an extension of the earlier W4-11 dataset;[41] it includes molecules and radicals composed of atoms H through Cl with up to eight heavy atoms (a heavy atom is defined here as an atom heavier than H). Most of the D_e values in the W4-17 database were obtained by using the layered CCSDTQ5/CBS or CCSDTQ56/CBS level of theory. The W4-17 dataset contains two subsets (i) a single-reference subset of 183 systems denoted as SR183 in the present paper and (ii) a multireference subset of 17 systems denoted as MR17 in the present paper.

We have also tested the new methods against the barrier heights and reaction energies in the DBH24-W4 database and the energetics of the reaction of three strongly correlated Criegee intermediates with water. The reactions and geometries in DBH24-W4 database are the same as in our previous DBH24/08 barrier height database,[42] but with all the energetic data updated to new reference values taken from the W4/W3.2 calculations of Karton et al.[43] For the best estimates of the reactions of the Criegee intermediates,[44] we used the W3X-L calculations (which are post-CCSD(T) estimates) of Long et al.[44]

2.2 Methods Tested. We test our new method (which is described in Section 4), against several published composite methods, in particular four MCCMs (BMC-CCSD,[45] MCG3-MPW,[46] MCG3/3,[15] MCQCISD-MPW[46]), five G*n* methods (G4,[9] G3SX(MP3),[47] ROG4(MP2)-6X,[48] G3,[11] and G2[10]), five Weizmann methods (W3X-L,[26] W3X,[49] W2-F12,[20] W2X,[26] and W1-F12 [20]), a composite method of Wilson and coworkers (ccCA-PS3[50]), a composite method of Petersson and coworkers (ROCBS-QB3[51]), and a dual-level method studied in a recent paper of Papajak and one of the authors[21] based on MP2-F12[31] and CCSD(T)-F12a[38, 39] and defined by:

$$E(\text{DL-jun}) = E(\text{MP2-F12/jun-T}) + [E(\text{CCSD(T)-F12a}) - E(\text{MP2-F12})]/\text{jun-DZ}$$

where basis set abbreviations are explained in Table 1.

3. Computational Methods

3.1. Geometries. The geometries for all molecules and radicals in the W4-17 paper are taken from Supporting Information of the W4-17 paper, where they were optimized at the CCSD(T)/V(Q+d)Z level of theory. The geometries for the DBH24-W4 database are from previous studies,[42] and they have been optimized at the QCISD/MG3 level of theory. The geometries for the reaction of the Criegee intermediate were taken from a previous paper,[44] and they have been optimized at the CCSD(T)-F12a/T-F12 or QCISD/T level of theory.

3.2 Basis Sets. The basis sets that we have employed in the WMS method are shown in Table 1, with the corresponding references, in order of basis set size. This table also shows the method of basis set abbreviation that we use.

3.3 Software and Reference Functions. All calculations in this article were performed with version 2015.1.13 of the *Molpro* software[52] and version 3.0 of *MLGauss*.[53] *MLGauss* employs *Gaussian 09*[54] for the calculations of the energy components in some of the sMCCMs.

All calculations and timing for the BMC-CCSD,[45] MCG3-MPW,[46] MCG3/3,[15] MCQCISD-MPW,[46] G2, G3, G4, ROG4(MP2)-6X, ROCBS-QB3, and G3SX(MP3) methods were carried out with the *Gaussian 09* software, and *Molpro* is used for the computations of all other methods. For all the methods using the *Gaussian 09* software except ROG4(MP2)-6X and ROCBS-QB3, the references for the open-shell systems are unrestricted Hartree-Fock (UHF) wave function. For all the methods using *Molpro* and for the ROG4(MP2)-6X and ROCBS-QB3 methods, the references are restricted or restricted-open-shell HF (RHF or ROHF) wave functions. Note that the CCSD/CCSD-F12 open-shell calculations in *Molpro* employed the Knowles-Hampel-Werner definition[55] of the RHF-UCCSD scheme. By default, *Gaussian09* employs UHF-CCSD, but *Gaussian09* can do ROHF-CCSD calculations using the Watts-

Gauss-Bartlett[56] scheme, which has a subtle difference from the Knowles-Hampel-Werner RHF-UCCSD scheme in *Molpro*.

3.4 Computational Cost Estimates. It is useful to compare the computational costs of composite methods, and we have used the computational time for thiophene as an indicator of the computational cost. Thiophene has four hydrogens, four 2p atoms (carbon), and one 3p element (sulfur). All the timing estimates have been performed on an INSPUR supercomputer using 8 cores of the Intel Xeon E5-2630 v4 CPU, with the same memory limit (1300 MW). We have normalized all compute times by dividing by the timing of the *Molpro* MP2/jul-D calculation (MP2/jul-D is a good choice for normalization because the timing for this method is similar for the two software packages that we used).

4. The WMS Methods

The new composite model chemistry method is built on the following formula for the Born-Oppenheimer energy:

$$\begin{aligned} E(\text{WMS}) = {} & E(\text{CCSD(T)-F12b/jul-D}) + c_{\text{HF}}[\Delta E(\text{HF})] + c_{\text{CABS}}[\Delta E(\text{CABS})] \\ & + c_{\text{MP2}}[\Delta E(\text{MP2-cor})] + c_{\text{F12}}[\Delta E(\text{F12})] + c_{\text{CCSD}}[\Delta E(\text{CCSD-HO})] \\ & + c_{\text{(T)}}[\Delta E(\text{CCSD(T)-F12})] + \text{E}_{\text{CV}} + \text{E}_{\text{SRel}} + \text{E}_{\text{SO}} \end{aligned} \quad (1)$$

where

$$\Delta E(\text{X}) = \text{E(X/jul-T)} - \text{E(X/jul-D)} \quad (2)$$

The various methods X are explained below, the jul-D and jul-T basis sets are explained in Table 1, E_{CV} is the core-valence correlation energy, E_{SRel} is the scalar relativistic contribution, E_{SO} is vector relativistic contribution (which is labeled in the usual way as the spin-orbit (SO) term), and c_{HF}, c_{CABS}, c_{MP2}, c_{F12}, c_{CCSD}, $c_{\text{(T)}}$ and the parameters in E_{CV} have been optimized against the W4-17 database by minimizing the RMSE of the atomization energies in W4-17.

Details of how each of the terms in eq. (1) is calculated are given in the following subsections.

4.1 CBS Extrapolation. Two of the terms (explained below) in eq 1 involve extrapolating to the CBS limit by using the two-point power-law formula:[57]

$$E(n) = E^{CBS} + A/n^{\alpha} \qquad (3)$$

where n is 2 for double zeta and 3 for triple zeta, and α is a parameter.

4.2 Hartree-Fock Components. In the WMS scheme, the HF components include the complementary auxiliary basis singles (CABS)[58] correction as adopted in the Wn-F12[20, 59] and WnX[26, 60] models, and the canonical HF energy and the CABS energy are treated as separate terms (X = HF and X = CABS) in eq. (1).

4.3 Explicitly Correlated Calculations. MP2 and CCSD calculations with F12 and CCSD(T) calculations with F12b suffixes are explicitly correlated methods; in these calculations the configuration state functions contain an explicit correlating factor

$$F = \exp(-\beta r_{12}) \qquad (4)$$

For all explicitly correlated calculations for valence correlation energies, we have used $\beta = 0.9$ a. u. for the jul-D basis set and $\beta = 1.0$ a. u. for jul-T basis set, based on recommendations by Peterson et al.[61] and Hill et al.[62] The default value for β in *Molpro* is 1.0. The supporting information gives input examples for setting β to the nondefault values of the WMS calculations.

In all of the explicitly correlated coupled cluster calculations, we have employed the fixed-amplitude 3C(FIX) ansatz[63-65] for the CCSD(T)-F12b method of Knizia et al.[38, 65] The 3C(FIX) ansatz is the default in *Molpro*; the F12b method is expected to slightly underestimate the full F12 energy.

4.4 CCSD Valence Correlation Energy. In the WMS model, the CCSD-F12 correlation energy (which equals the CCSD energy minus the HF energy) is decomposed into three contributions:

$$\mathrm{E(CCSD\text{-}cor) = E(MP2\text{-}cor) + E(F12) + E(CCSD\text{-}HO)} \tag{5}$$

where

$$E(\mathrm{MP2\text{-}cor}) = E(\mathrm{MP2}) - E(\mathrm{HF}) \tag{6}$$

$$E(\mathrm{F12}) = E(\mathrm{MP2\text{-}F12}) - E(\mathrm{MP2}) \tag{7}$$

and

$$E(\mathrm{CCSD\text{-}HO}) = E(\mathrm{CCSD\text{-}F12}) - E(\mathrm{MP2\text{-}F12}) \tag{8}$$

4.5 The Scalar Relativistic Component. The scalar relativistic contribution (in the second-order Douglas–Kroll–Hess approximation)[66, 67] is extrapolated as the difference between non-relativistic MP2/jul-*n* (where (*n* = D or T) and relativistic MP2/jul-*n*-DK calculations using Eq. (3) with α = 2.0.

4.6 The Core-Valence Correlation Component. The calculation of the core-valence correlation component can be a bottleneck of the computational cost of a composite method because of the demanding full-electron correlation calculations. We employed separate extrapolation of the MP2 and CCSD-minus-MP2 contributions to obtain the CCSD core-valence correlation. The total core-valence correlation contribution for WMS is defined as:

$$\Delta\mathrm{E}_{\mathrm{CV}}^{\mathrm{WMS}} = \Delta\mathrm{E}_{\mathrm{MP2}}^{\mathrm{CV-CBS}} + (\Delta\mathrm{E}_{\mathrm{CCSD}}^{\mathrm{CV}} - \Delta\mathrm{E}_{\mathrm{MP2}}^{\mathrm{CV}} + c\Delta\mathrm{E}_{\mathrm{(T)}}^{\mathrm{CV}})_{\mathrm{wCVDZ}} \tag{8}$$

where $\Delta E_{\mathrm{MP2}}^{\mathrm{CV-CBS}}$ is extrapolated with the wCV{D,T} basis sets using Eq. (3) and the optimized α is 3.55. The perturbative triple excitation contribution of Eq. (8) is evaluated with the wCVDZ basis set, and scaled by an optimized factor *c* = 3.8. As described in Section 4, the parameter α and c for the CV calculations are optimized globally (along with the parameters in valence correlations) against the TAEs of W4-17.

We tried using the CCSD(T)-F12b method to calculate the core-valence correlation, but we found that it is inferior to CCSD(T) from the standpoint of cost-to-performance. We also found that the wCVDZ basis set gives better performance than the CVDZ basis sets when combined with CCSD(T).

4.7 Spin-Orbit Coupling. To first order, E_{SO} is zero by symmetry for closed-shell molecules, for linear molecules in Σ states, and for singlet and doublet molecules in A or B states. For cases with nonzero E_{SO}, it can be obtained from experiment for monatomic species (from data in Moore's tables,[68] which are reproduced, with only slight updates, on the NIST website) and by calculations for molecules (e.g., by state-averaged complete active space self-consistent field calculations).

In the present study, E_{SO} values for calculating atomization energies have been taken from the W4-17 paper.[40] (they could also be computed on the back of an envelope for any monatomic species by using Moore's tables). For DBH24, the nonzero E_{SO}values for F, Cl, and OH are taken from a previous paper,[69] and E_{SO} for SH is taken from our BMC-CCSD paper.[45]

For the reactions of Criegee intermediates, all the E_{SO}values are zero.

4.8 Optimized parameters. The optimized parameters for the WMS valence correlation energies are in Table 2. All parameters are positive and greater than unity, as expected for a physical extrapolation.

5. The Performance of the WMS Composite Method

In the limit of large *N*, scaling of the computational effort of the CCSD(T) method is N^7 where *N* is the number of atoms in the molecule, and the computational costs of the direct calculation of high-order valence correlation (HOVC) using CCSDTQ and CCSDTQ5 are respectively N^{10} and N^{12}; thus these latter methods are unaffordable even for systems of moderate size with (for example, molecules with 10 heavy atoms). Although extrapolation has been widely used for basis sets, it is not used for excitation levels in the most popular methods. The key new aspect of the present work is that we optimize the parameters to effectively include HOVC by extrapolation. By fitting to high-level calculations including high-order correlation, the goal of the present method is accuracy greater than CCSD(T)/CBS, but at cost no higher than CCSD(T)-F12b/jul-T. Both scaling of the correlation energy and extrapolation to an infinite basis set involving taking linear combinations of differences of energy components. If desired, this can be viewed as replacing the several “high level correction” parameters employed in most of the G*n* methods. (The G*n* methods based on scaling do not have such parameters; however, the G*n* methods without scaling are not size extensive.[70, 71]) In order to make the present method size extensive, we chose a functional form based on scaling[13, 15, 72-74] so that the final results retain the size extensivity of the underlying ab initio methods.

In the text we will discuss mean unsigned errors (MUEs); root-mean-square errors, which are less robust, are given in the Electronic Supplementary Information (ESI). Tables with a prefix S are in the ESI.

5.1 The Scalar Relativistic Component. Tables 3 and S2 show that the extrapolated scalar relativistic contributions give an MUE of 0.04 kcal/mol for the MR17 database and 0.02 kcal/mol for the SR183 database.

5.2 Atomization Energies. In Table 4, we tabulate the mean signed and unsigned errors (MSEs and MUEs) for several methods on the MR17 and SR183 data, and we also show AMUE, which is the average of the MUEs for the MR17 and SR183 databases. Of the methods in Table 4, W3X-L performs best for the MR17 database, and it is the second best performer for the SR183 database; it has an AMUE of 0.37 kcal/mol. Note, for perspective, that the average number of bonds for the molecules in the data set is 2.24 for MR17 and 4.14 for SR183.

If we divided by the numbers of bonds to obtain the mean errors on a per bond basis, the mean errors would be smaller by a factor of 2 to 4. The excellent performance of W3X-L, especially its good performance for MR17, demonstrates that it is an effective lower-cost variant of the W4 method.

WMS is the second best performer for MR17, and it is the best performer for SR183, with an AMUE of 0.45 kcal/mol. The computational cost of WMS is two orders of magnitude lower than that of W3X-L for thiophene and it has better scaling for going to larger molecules, as indicated in the final column of the table. Encouragingly, the performance of WMS is better than that of W3X, which includes actual calculations of post-CCSD(T) contributions. A graphical comparison of the performances and timings is given in Figure 1.

We note, as a key result of the present paper, the importance of treating HOVC contributions to the energy, especially for the MR data. For the SR183 data, the work of Karton et al. shows that average absolute magnitude of the quadruple, pentuple, and sextuple contributions to the atomization energies are respectively 0.97, 0.05, and 0.00 kcal/mol, and for the MR17 data, they are respectively 2.23, 0.14, and 0.01 kcal/mol.

If we use AMUEs to gauge the performance of the methods in Table 4, only W3X-L WM-*n*, W3X, W*n*-F12, W2X, ccCA-PS3, and G4 methods give AMUEs with accuracy better then 1 kcal/mol, which is commonly called the borderline of “chemical accuracy”.

We have already mentioned the last column of Table 4, which gives the scaling of the steepest-scaling component in each method. The best N^6 method in Table 4 is MCQCISD-MPW, which is the only N^6 method that has a density functional component.

The G2 row of Table 4 is particularly interesting in illustrating the progress in the field. The G2 method is the method that Pople presented in his Nobel Prize lecture 20 years ago; the

AMUE is 2.33 kcal/mol. It is now possible to do chemical accuracy much more reliably than when Pople's work was justifiably recognized as a breakthrough for quantum chemistry.

The relatively large error of the straightforward DL-jun method might be surprising since DL-jun was shown[21] to be accurate to about a half kcal/mol for a set of reaction energies and barrier heights. It is however, well appreciated that many sources of error largely cancel out in reaction energies and to a slightly lesser extent in barrier heights, and these sources of error do not cancel as much for the more drastic changes involved in atomization where all bonds are completely broken. This shows the difficulty of the present test of theory and makes the good performance of the better performing methods even more impressive. The DL-jun scheme systematically underestimates the atomization energy of hydrocarbons; the signed errors (calculated minus best estimate) are -2.60, -4.64, -8.68, an -10.57 kcal/mol for CH_4, C_2H_6, C_4H_{10}, and C_5H_{12}, which can be explained by the CCSD(T)-F12a calculations overestimating the correlation energy of the carbon atom relative to a carbon atom in a molecule.

Unexpectedly, the DL-jun and ROG4(MP2)-6X methods have smaller errors for strongly correlated MR17 than for the weakly correlated ST183; all other methods have in Table 4 have MUE(MR17)/MUE(SR183) > 1, with values in the range 1.2 to 3.0.

5.3 Diverse Barrier Heights and Reaction Energies. Table 5 presents the performance of the WMS method for the DBH24-W4 barrier heights database. The overall MUE for the DBH24/08 database is 0.16 kcal/mol for WMS. We remind the reader that no barrier height data was used in parameterization.

In a test[42] of 64 N^7 methods, including some very expensive ones, against the DBH24 barrier heights in 2009, the MUEs ranged from 0.46 to 7.79 kcal/ mol, with an average MUE of 2.71 kcal/mol. In this light, the results in Table 5 represent a remarkable step forward.

5.4 Criegee Intermediate Reactions. Table 6 presents the performance of WMS for the energetics in a class of challenging systems, namely the reactions of three Criegee intermediates with water. Criegee intermediates are carbonyl oxides, which have strong diradical character (sometimes described as zwitterionic character), and understanding their atmospheric chemistry is important for modeling climate change.[75] Previously, Long, Bao, and one of us[43] have shown that W2X predictions differ largely from W3X-L predictions, with MUEs range 0.65-0.80 kcal/mol for the three reactions in Table 6. The average mean unsigned deviation of WMS from the expensive W3X-L results is only 0.21 kcal/mol; this finding confirms that our linear extrapolation scheme does capture most of the higher-order valence correlation energies for a set of difficult practical cases.

5.5 Timings. The relative timings of several composite methods are given in Table 4. Figure 1 is a plot of the performance (MUEs) and the computational cost (relative timing, blue balls in Figure 1). Discussion of timings is sometimes controversial, because timings depends on the molecule, the software, the computer, and the computer load, but we give sample timings for a molecule small enough to run even the more expensive methods, just to give a rough idea of the cost savings achievable by using the new WMS method. The timing comparison in Table 4 and Figure 1 is very dramatic.

6. Concluding Remarks

By combining the scaling approach with carefully crafted computation schemes for the inner-shell and the scalar relativistic contributions, we have proposed a composite model chemistry methods, WMS, with low computational costs and N^7 cost scaling. The new method has been shown to be accurate for the 200 atomization energies in the W4-17 database with an average mean unsigned error (AMUE) of 0.45 kcal/mol. The WMS method was also tested against the DBH24 database of barrier heights for diverse set of reactions, and the reaction energies and barrier heights for reactions of three Criegee intermediates with water. These tests confirm the robustness of the linear extrapolation scheme.

The extrapolation was developed with the goal of designing a method that can be used to estimate the higher-order valence correlation energies for large systems, even when they have multireference character, because it is not feasible to perform higher-order correlation methods such as CCSD(TQ) even with small double-zeta basis sets for large systems. Although no post-CCSD(T) calculations are used in the new composite method, the results are more accurate than the W3X method, which does include such calculations and which is much more expensive.

The present method was developed based on the atomization energies of main-group molecules (the W4-17 database contains molecules with atoms no heavier than Cl) at the equilibrium geometry, and its validity was verified for chemical reaction barrier heights. The MR17 subset of W4-17 contain elements B, C, N, F, G, S, and Cl; the SR183 subset contains these elements plus H, Al, Si, and P. There are only five data containing a metallic element (five compounds containing Al). All data are for equilibrium geometries. Further study will be required to test the method for the transition metal compounds or stretched main-group molecules, and in fact the method may need further development to be accurate for the kinds of strong correlation involved in such systems. For example, one might want to use different components to obtain a method valid for stretched bonds with bond length much longer than those involved in transition states. Therefore the new methods might be improvable upon further study. Other composite methods, such as the G*n* and W*n* methods, have been improved in a sequence of iterations, and maybe this is possible in the present case as well, but we believe the present work is important in already showing the possibilities of greatly improved accuracy for a wide class of strongly correlated systems at low cost. The present investigation used a human-guided strategy for extrapolation, but it is possible that machine learning could be used to develop even more powerful excitation-level extrapolation schemes.

Author Information

ORCID

Yan Zhao: 0000-0002-1234-4455
Lixue Xia: 0000-0001-6006-4055
Xiaobin Liao: 0000-0002-2455-832X
Qiu He: 0000-0002-5782-4962
Maria X. Zhao: 0000-0002-8662-0176
Donald G. Truhlar: 0000-0002-7742-7294

Acknowledgments

This work was supported in part by the Thousand Innovative Talents Plan of the Chinese Government and by the U.S. National Science Foundation.

References

1. A. Karton, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2016, **6**, 292-310.
2. C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618-622.
3. G. A. Petersson, A. K. Yee and A. Bennett, *J. Chem. Phys.*, 1985, **83**, 5105-5128.
4. J. A. Pople, *Rev. Mod. Phys.*, 1999, **71**, 1267-1274.
5. J. A. Montgomery, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.*, 1994, **101**, 5900-5909.
6. J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.*, 2000, 6532-6542.
7. J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.*, 1999, **110**, 2822-2827.
8. P. M. Mayer, C. J. Parkinson, D. M. Smith and L. Radom, *J. Chem. Phys.*, 1998, **108**, 604-615.
9. J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari and L. A. Curtiss, *J. Chem. Phys.*, 1989, **90**, 5622-5629.
10. L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, *J. Chem. Phys.*, 1991, **94**, 7221-7230.
11. L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov and J. A. Pople, *J. Chem. Phys.*, 1998, **109**, 7764-7776.
12. L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 2007, **126**, 084108.
13. L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, *J. Chem. Phys.*, 2000, **112**, 1125-1132.
14. P. L. Fast and D. G. Truhlar, *J. Phys. Chem. A*, 2000, **104**, 6111-6116.
15. B. J. Lynch and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 3898-3906.
16. Y. Zhao, B. J. Lynch and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **7**, 43-52.
17. J. M. L. Martin and G. d. Oliveira, *J. Chem. Phys.*, 1999, **111**, 1843-1856.
18. S. Parthiban and J. M. L. Martin, *J. Chem. Phys.*, 2001, **114**, 6014-6029.
19. A. Karton, N. Sylvetsky and J. M. L. Martin, *Journal of Computational Chemistry*, 2017, **38**, 2063-2075.
20. A. Karton and J. M. L. Martin, *J. Chem. Phys.*, 2012, **136**, 124114.
21. E. Papajak and D. G. Truhlar, *J. Chem. Phys.*, 2012, **137**, 064110.
22. A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay and J. Gauss, *J. Chem. Phys.*, 2004, **120**, 4129-4141.
23. A. Karton, E. Rabinovich, J. M. L. Martin and B. Ruscic, *J. Chem. Phys.*, 2006, **125**, 144108.
24. A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez and J. F. Stanton, *J. Chem. Phys.*, 2004, **121**, 11599.
25. D. Feller, K. A. Peterson and D. A. Dixon, *J. Chem. Phys.*, 2008, **129**, 204105.
26. B. Chan and L. Radom, *J. Chem. Theory Comput.*, 2015, **11**, 2109-2119.
27. Á. Ganyecz, M. Kállay and J. Csontos, *J. Chem. Theory Comput.*, 2017, **13**, 4193-4204.
28. A. Mahler and A. K. Wilson, *J. Chem. Theory Comput.*, 2013, **9**, 1402-1407.
29. E. A. Hylleraas, *Z. Phys*, 1928, **48**, 469-494.
30. W. Kutzelnigg, *Theor. Chim. Acta*, 1985, **68**, 445-469.
31. W. Klopper and C. C. M. Samson, *J. Chem. Phys.*, 2002, **116**, 6397-6410.
32. S. Ten-no and F. R. Manby, *J. Chem. Phys.*, 2003, **119**, 5358-5363.
33. F. R. Manby, *J. Chem. Phys.*, 2003, **119**, 4607-4613.
34. D. P. Tew and W. Klopper, *J. Chem. Phys.*, 2005, **123**, 074101.
35. S. Ten-no, *Chem. Phys. Lett.*, 2004, **396**, 56-61.
36. A. J. May, E. F. Valeev, R. Polly and F. R. Manby, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2710-2713.
37. T. Helgaker, W. Klopper and D. P. Tew, *Mol. Phys.*, 2008, **106**, 2107-2143.
38. T. B. Adler, G. Knizia and H.-J. Werner, *J. Chem. Phys.*, 2007, **127**, 221106.

39. G. Knizia, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2009, **130**, 054104.
40. A. Karton, N. Sylvetsky and J. M. L. Martin, *J. Comp. Chem.*, 2017, **38**, 2063-2075.
41. A. Karton, S. Daon and J. M. L. Martin, *Chem. Phys. Lett.*, 2011, **510**, 165-178.
42. J. J. Zheng, Y. Zhao and D. G. Truhlar, *Journal of Chemical Theory and Computation*, 2009, **5**, 808-821.
43. A. Karton, A. Tarnopolsky, J.-F. Lamere, G. C. Schatz and J. M. L. Martin, *J. Phys. Chem. A*, 2008, **112**, 12868.
44. B. Long, J. L. Bao and D. G. Truhlar, *J. Am. Chem. Soc.*, 2016, **138**, 14409-14422.
45. B. J. Lynch, Y. Zhao and D. G. Truhlar, *Journal of Physical Chemistry A*, 2005, **109**, 1643-1649.
46. Y. Zhao and D. G. Truhlar, *Journal of Physical Chemistry A*, 2005, **109**, 4209-4212.
47. L. A. Curtiss, P. C. Redfern, K. Raghavachari and J. A. Pople, *Chem. Phys. Lett.*, 2002, **359**, 390.
48. B. Chan, A. Karton, K. Raghavachari and L. Radom, *J. Phys. Chem. A*, 2016, **120**, 9299-9304.
49. B. Chan and L. Radom, *J. Chem. Theory Comput.*, 2013, **9**, 4769-4778.
50. N. J. DeYonker, T. R. Cundari and A. K. Wilson, *J. Chem. Phys.*, 2006, **124**, 114104.
51. G. P. F. Wood, L. Radom, G. A. Petersson, E. C. Barnes, M. J. Frisch and J. A. Montgomery, *J. Chem. Phys.*, 2006, **125**, 094106.
52. H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby and M. Schtz, *WIREs Comput Mol Sci*, 2012, **2**, 242--253.
53. Y. Zhao, R. Meana-Pañeda and D. G. Truhlar, *MLGAUSS-version 3.0*, University of Minnesota, Minneapolis, 2012.
54. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian09*, Gaussian, Inc., Pittsburgh PA, Revision D.01 edn., 2009.
55. P. J. Knowles, C. Hampel and H.-J. Werner, *J. Chem. Phys.*, 1993, **99**, 5219-5227.
56. J. D. Watts, J. Gauss and R. J. Bartlett, *J. Chem. Phys.*, 1993, **98**, 8718-8733.
57. D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45-48.
58. G. Knizia and H.-J. Werner, *J. Chem. Phys.*, 2008, **128**, 154103.
59. N. Sylvetsky, K. A. Peterson, A. Karton and J. M. L. Martin, 2016, **144**, 214101.
60. B. Chan and L. Radom, *J. Chem. Theory Comput.*, 2012, **8**, 4259-4269.
61. K. A. Peterson, *J. Chem. Phys.*, 2008, **128**, 084102.
62. J. G. Hill, S. Mazumder and K. A. Peterson, *J. Chem. Phys.*, 2010, 054108.
63. S. Ten-no and J. Noga, *WIREs Comput Mol Sci*, 2012, **2**, 114-125.
64. H. J. Werner, T. B. Adler and F. R. Manby, *J. Chem. Phys.*, 2007, **126**, 164102.
65. G. Knizia and H.-J. Werner, *J. Chem. Phys.*, 2009, **128**, 154103.
66. M. Douglas and N. M. Kroll, *Ann. Phys.*, 1974, **82**, 89-155.
67. B. A. Hess, *Phys. Rev. A*, **33**, 3742-3748.
68. C. E. Moore, *Atomic Energy Levels*, National Bureau of Standards, Washington, 1948.
69. P. L. Fast, J. Corchado, M. L. Sanchez and D. G. Truhlar, *J. Phys. Chem. A*, 1999, **103**, 3139.
70. P. L. Fast, M. L. Sanchez, J. C. Corchado and D. G. Truhlar, *J. Chem. Phys.*, 1999, **110**, 11679-11681.

71. P. L. Fast, M. L. Sanchez and D. G. Truhlar, *Chem. Phys. Lett.*, 1999, **306**, 407-410.
72. F. B. Brown and D. G. Truhlar, *Chem. Phys. Lett.*, 1985, **117**, 307-313.
73. M. S. Gordon and D. G. Truhlar, *J. Am. Chem. Soc.*, 1986, 5412-5419.
74. Y. Zhao, B. J. Lynch and D. G. Truhlar, *Physical Chemistry Chemical Physics*, 2005, **7**, 43-52.
75. C. J. Percival, O. Welz, A. J. Eskola, J. D. Savee, D. L. Osborn, D. O. Topping, D. Lowe, S. R. Utembe, A. Bacak, G. McFiggans, M. C. Cooke, P. Xiao, A. T. Archibald, M. E. Jenkin, R. G. Derwent, I. Riipinen, D. W. K. Mok, E. P. F. Lee, J. M. Dyke, C. A. Taatjes and D. E. Shallcross, *Faraday Discuss.*, 2013, **165**, 45-73.
76. R. A. Kendall, T. H. Dunning Jr. and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796-6806.
77. D. E. Woon and T. H. Dunning Jr., *J. Chem. Phys.*, 1993, **99**, 3730-3737.
78. T. H. Dunning Jr., K. A. Peterson and A. K. Wilson, *J. Chem. Phys.*, 2001, **114**, 9244-9253.
79. E. Papajak and D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 10-18.
80. K. A. Peterson and T. H. Dunning, *J. Chem. Phys.*, 2002, **117**, 10548-10560.
81. J. T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007-1023.
82. D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1993, **98**, 1358-1371.
83. W. A. de Jong, R. J. Harrison and D. A. Dixon, *J. Chem. Phys.*, 2001, **114**, 48-53.

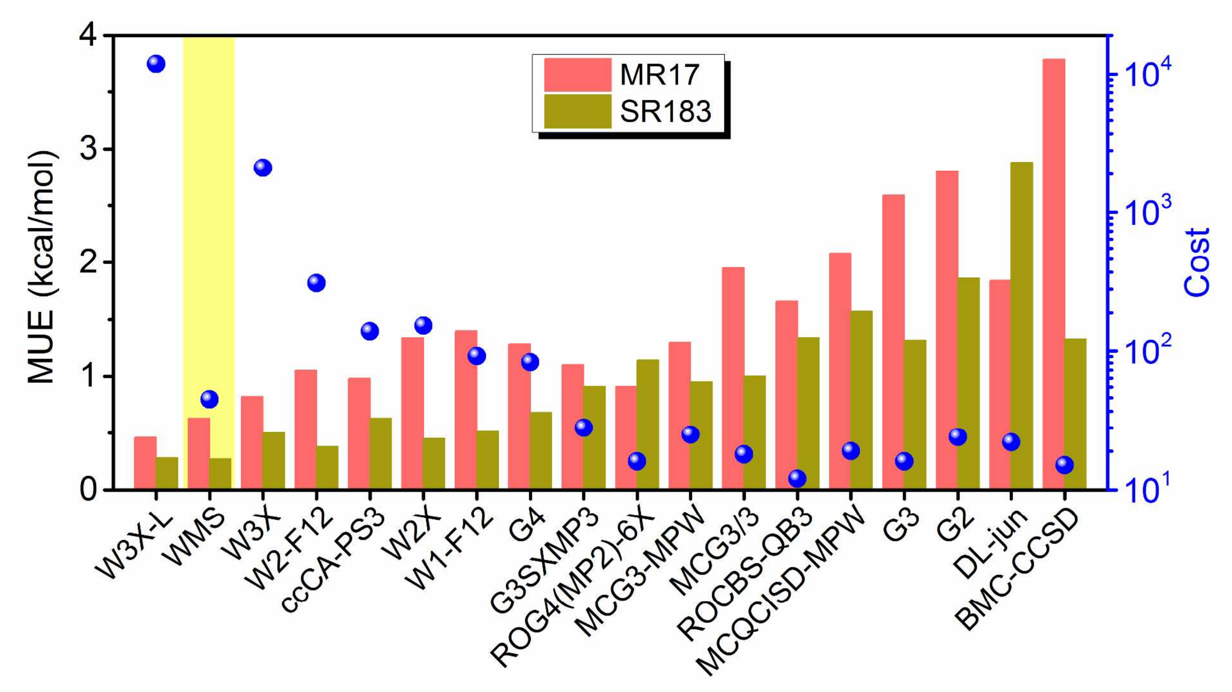

Figure 1. Performance (MUE denotes mean unsigned error) and computational cost

Table 1. Basis sets

Short name	Description	N^a	Ref.[b]
jun-D	jun-cc-pV(D+d)Z	73	76-79
wCVDZ	cc-pwCVDZ	78	80
jul-D	jul-cc-pV(D+d)Z[c]	88	76-79
jul-D-DK	cc-pVDZ-DK for hydrogen, and aug-cc-pVDZ-DK for heavy atoms	88	79, 81-83
T	cc-pVTZ	136	76-78
jun-T	jun-cc-pV(T+d)Z	163	76-79
jul-T	jul-cc-pV(T+d)Z[d]	184	76-79
jul-T-DK	cc-pVTZ-DK for hydrogen, and aug-cc-pVTZ-DK for heavy atoms	184	79, 81-83
wCVTZ	cc-pwCVTZ	187	80
T-F12	cc-pVTZ-F12	222	61

[a] N is the number of contracted basis functions for vinyl chloride C_2H_3Cl

[b] references for the basis set

[c] equivalent to cc-pVDZ for hydrogen, and aug-cc-pV(D+d)Z for heavy atoms

[d] equivalent to cc-pVTZ for hydrogen, and aug-cc-pV(T+d)Z for heavy atoms

Table 2. Optimized parameters for WMS

Coefficients	Value
c_{HF}	2.178
c_{CABS}	2.309
c_{MP2}	1.018
c_{F12}	1.126
c_{CCSD}	1.569
$c_{(T)}$	2.175

Table 3. Statistical analysis of the scalar relativistic contribution in WM (kcal/mol) [a]

Basis sets	α	MR17		SR183	
		MSE	MUE	MSE	MUE
jul-V{D,T}Z-DK	2.0	0.04	0.04	0.01	0.02

[a] The reference scalar relativistic contribution were taken from the W4-17 paper,[40] and they are obtained from the difference between nonrelativistic CCSD(T)/aug'-V(Q+d)Z and relativistic CCSD(T)/aug'-V(Q+d)Z-DK calculations. MSE = mean signed error = mean signed deviation from referecne value; MUE = mean unsigned error = mean absolute deviation from reference value.

Table 4. Performance of the composite methods for the W4-17 database (kcal/mol)[a,b]

Method	Ref. [c]	MR17		SR183		AMUE	Time [d]	Post-CCSD(T) [e]	Scaling n[f]
		MSE	MUE	MSE	MUE				
W3X-L	This work	0.15	0.46	0.16	0.28	0.37	11862	Yes	9
WMS	This work	0.01	0.63	-0.06	0.27	0.45	45	Extrapolation	7
W3X	This work	0.25	0.82	0.13	0.50	0.66	2096	Yes	9
W2-F12	[40]	-0.99	1.05	-0.07	0.38	0.72	308	No	7
ccCA-PS3	[40]	-0.26	0.98	0.25	0.63	0.81	140	No	7
W2X	[40]	-1.17	1.34	–0.01	0.45	0.90	154	No	7
W1-F12	[40]	-1.68	1.40	-0.45	0.51	0.96	92	No	7
G4	[40]	-0.67	1.28	-0.06	0.68	0.98	83	HLC	7
G3SXMP3	This work	-0.09	1.10	-0.01	0.91	1.01	28	No	7
ROG4(MP2)-6X	[40]	0.71	0.91	0.90	1.14	1.03	16	HLC	7
MCG3-MPW	This work	-0.12	1.30	-0.38	0.95	1.12	25	No	7
MCG3/3	This work	-0.45	1.95	-0.02	1.00	1.47	18	No	7
ROCBS-QB3	[40]	-0.61	1.66	0.32	1.34	1.50	12	HLC	7
MCQCISD-MPW	This work	1.73	2.08	-0.50	1.57	1.82	19	No	6
G3	[40]	-2.20	2.59	-0.68	1.32	1.96	16	HLC	7
G2	This work	-1.66	2.80	-0.64	1.86	2.33	24	HLC	7
DL-jun	This work	-0.34	1.84	-2.63	2.88	2.36	22	No	7
BMC-CCSD	This work	3.14	3.79	-0.23	1.33	2.56	15	No	6

[a] The reference TAEs were taken from the W4-17 paper,[40] and they are obtained by using the W4 protocol and its variants.

[b] MSE = mean signed error = mean signed deviation (MSD); MUE = mean unsigned error = mean absolute deviation (MAD). AMUE is the average of the MUEs of the MR17 and SR183 databases.

[c] The reference given is for the performance data; see section 2.2 of the text for references for the methods.

[d] Relative computational time for thiophene as normalized by the time of the MP2/jul-VDZ calculation.

[e] Some of the methods include empirical high-level correction (HLC) parameters to account of higher-order effects.

[f] In the limit of large *N*, the cost of the most expensive component scales as N^n where *N* is the number of atoms.

Table 5. Performance of the WM methods for the DBH24-W4 database (kcal/mol)[a]

Method	Heavy atom transfer		S_N2		Unimolecular or Association		Hydrogen atom transfer		DBH24-W4	
	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
WM-S	-0.17	0.26	-0.15	0.16	0.05	0.14	-0.03	0.07	-0.07	0.16

[a] For best estimates, we employed the W3.2 and W4 results in a paper by Karton et al.[43]

Table 6. Performance of the WM methods for the forward energy barriers ($\Delta E_f^{\neq}$), reverse energy barriers ($\Delta E_r^{\neq}$), and reaction energies (ΔE) of the reactions of Criegee intermediates with H_2O (kcal/mol)[a]

Method	$\Delta E_{\mathrm{f}}^{\neq}$		$\Delta E_{\mathrm{r}}^{\neq}$		ΔE		MUD[b]
			$CH_2OO + H_2O \rightarrow CH_2(OH)OOH$				
	B1a-TS1	B1a-TS2	B1a-TS1	B1a-TS2	B1a-P1	B1a-P2	
W3X-L//CCSD(T)-F12a/VTZ-F12	0.49	1.45	46.91	47.66	-46.42	-46.19	0.00
WMS//CCSD(T)-F12a/VTZ-F12	0.65	1.53	47.23	47.90	-46.57	-46.37	0.19
W2X//CCSD(T)-F12a/ VTZ-F12	-0.04	0.91	47.44	48.16	-47.48	-47.25	0.70
			syn-$CH_3CHOO + H_2O \rightarrow HC(OH)CH_3OOH$				
	B2a-TS1	B2a-TS2	B2a-TS1	B2a-TS2	B2a-P1	B2a-P2	
W3X-L//QCISD/VTZ	5.15	6.51	45.02	45.94	-39.87	-39.43	0.00
WMS//QCISD/VTZ	5.25	6.91	45.29	46.21	-40.05	-39.30	0.22
W2X//QCISD/ VTZ	4.66	6.01	45.51	46.42	-40.85	-40.41	0.65
			anti-$CH_3CHOO + H_2O \rightarrow HC(OH)CH_3OOH$				
	B3a-TS1	B3a-TS2	B3a-TS1	B3a-TS2	B3a-P1	B3a-P2	
W3X-L//QCISD/ VTZ	-1.41	-0.67	43.16	43.9	-44.58	-44.57	0.00
WMS//QCISD/ VTZ	-1.50	-0.76	43.39	44.15	-44.88	-44.90	0.21
W2X//QCISD/ VTZ	-2.04	-1.3	43.64	44.37	-45.68	-45.67	0.80
			Overall average over 18 data				
W3X-L//QCISD/ VTZ							0.00
WMS//QCISD/ VTZ							0.21
W2X//QCISD/ VTZ							0.72

[a] The results for W2X and W3X-L were taken from a previous paper.[44] We have employed the W3X-L data as the reference. We have used the same name convention to label the transition states and products for different reaction pathways as in Ref. [44].

[b] Mean unsigned deviation averaged over the six quantities in the previous columns. The deviation is with respect to the W3X-L results.