Theor Chem Acc (2012) 131:1079
DOI 10.1007/s00214-011-1079-5

REGULAR ARTICLE

Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges

Kirk A. Peterson · David Feller · David A. Dixon

Received: 4 July 2011 / Accepted: 25 August 2011 / Published online: 11 January 2012

Abstract The current state of the art in wavefunction-based electronic structure methods is illustrated via discussions of the most important effects incorporated into a selection of high-accuracy methods chosen from the chemical literature. If one starts with a high-quality correlation treatment, such as provided by the CCSD(T) coupled cluster method, the leading effects include convergence of the results with respect to the 1-particle basis set, (outer) core/valence correlation, scalar relativistic effects and a number of smaller effects. For thermochemical properties such as the heat of formation, the zero-point vibrational energy also becomes important, introducing its own set of difficulties to the computational approach. Changes in the various components as the chemical systems incorporate heavier elements and as the size of the systems grows are also considered. Finally, challenges arising from the desire to extend existing methods to transition metal and heavier elements are considered.

Keywords Ab initio thermochemistry · Ab initio spectroscopy · Chemical accuracy · Composite methods · Coupled cluster

Published as part of the special collection of articles celebrating the 50th anniversary of Theoretical Chemistry Accounts/Theoretica Chimica Acta.

K. A. Peterson (✉) · D. Feller
Department of Chemistry, Washington State University, Pullman, WA 99164-4630, USA
e-mail: kipeters@wsu.edu

D. A. Dixon
Department of Chemistry, The University of Alabama, Shelby Hall, Box 870336, Tuscaloosa, AL 35487-0336, USA

1 Introduction

The last 20 years or so have been marked by phenomenal progress in the area of quantitatively accurate quantum chemistry. Arguably, the greatest impact due to these advances has been in the area of ab initio thermochemistry and to a lesser extent ab initio spectroscopy. The steep increase in computing power over these years has certainly been a major driver for these advances, but these gains have also been matched by significant methodological developments in solving the electronic Schrödinger equation. Simultaneously with both advances, researchers have benefited from the implementation of these methods in easy-to-use software packages that perform well on high-performance computers. In the current paper, the emphasis is not on density functional theory but necessarily on wavefunction-based methods, e.g., the hierarchy of coupled cluster techniques, since they represent a scheme that can systematically approach the exact solution of the Schrödinger equation in a given basis set. The other main approximation that strongly influences the accuracy of a quantum chemistry calculation is the choice of 1-particle basis set used to describe the underlying molecular orbitals. The introduction of the systematically convergent (to the complete basis set (CBS) limit) correlation consistent basis sets by Dunning [1] in 1989, cc-pVnZ where n = D, T, Q, 5, etc., has led to a paradigm shift in strategies for accurate solutions of the electronic Schrödinger equation. Using sequences of these basis sets with a chosen approximation method, e.g., coupled cluster singles and doubles with perturbative triples [2], CCSD(T), the CBS limit can be accurately estimated. This capability effectively eliminates the basis set incompleteness error from the calculation. In addition, it can minimize issues with basis set superposition error (BSSE). This strategy is now a common thread in all

high-accuracy ab initio thermochemistry and spectroscopy methodologies whether coupled cluster or multireference configuration interaction (MRCI) methods are used to solve the electronic Schrödinger equation.

Our goal is to design methods that provide at least chemical accuracy in terms of thermochemistry (and spectroscopy). So what does one mean by "chemical accuracy"? In the thermochemistry literature, this is almost universally interpreted as 1 kcal/mol or about 4 kJ/mol (1 kcal/mol = 4.184 kJ/mol). In the computational chemistry literature, a mean absolute deviation (MAD), or analogously a mean unsigned deviation (MUD), of 1 kcal/mol has been used to establish the criterion of chemical accuracy for a given thermochemistry methodology. As has been previously noted in the literature, however, an MUD of 1 kcal/mol does not actually represent chemical accuracy analogous to experiment since for experimentalists, the latter is generally based on a 95% confidence level or about 2 standard deviations. Hence, a more stringent MUD of about 0.5 kcal/mol is actually required to achieve an experimentalist's notion of chemical accuracy in a quantum chemistry methodology. The situation is a bit more ambiguous in terms of spectroscopic properties, such as equilibrium structures and vibrational frequencies. The term "spectroscopic accuracy" in the quantum chemistry literature generally refers to vibrational frequencies accurate to within ± 1 cm^{-1}. While not typically defined for equilibrium bond lengths (which in any event are not directly observable quantities), this high level of accuracy for vibrational frequencies can be loosely equated to bond lengths accurate to better than ± 0.0005 Å. This represents a serious challenge for computational quantum chemistry, and from a thermochemistry perspective, such error bars are more in line with kJ/mol or sub-kJ/mol accuracy. For the purposes of the present contribution, which is focused on *chemical* accuracy, the rather arbitrary definition of ± 0.005 Å and ± 15 cm^{-1} for structures and frequencies, respectively, will be used.

Attempting to achieve reliable chemical accuracy for a general polyatomic molecule from a single quantum chemistry calculation is a daunting if not impossible task. As will be demonstrated in detail below, results of this quality require the use of methods capable of recovering significant amounts of electron correlation together with 1-particle basis sets near the CBS limit. For example, the CCSD(T) method, which often has an inherent accuracy (defined as the limiting accuracy at the CBS limit) equivalent to chemical accuracy, at least for thermochemical properties, has a computational expense that scales roughly as n^3N^4, where n is the number of occupied orbitals and N the number of virtual orbitals (this is loosely given as M^7 where M is the number of basis functions, although for a fixed molecule, the cost actually scales as just N^4). To directly achieve chemical accuracy with CCSD(T) without extrapolation would require correlation of both the valence and outer-core electrons with a basis set of at least aug-cc-pCV6Z or aug-cc-pCV7Z quality [3]. For very small molecules, this is not especially difficult in terms of the computational cost, but for molecules of any size, this strategy becomes impractical, especially for the determination of the equilibrium geometry and the vibrational frequencies. Furthermore, if the goal is to push beyond chemical accuracy toward spectroscopic accuracy, this approach becomes completely intractable.

The solution to this problem is the use of so-called composite approaches, which are based on additivity arguments. The dominant sources of error in the calculation are treated as accurately as possible, and the remaining energy contributions are tackled with lower levels of approximation. For molecular systems with wavefunctions dominated by a single Slater determinant, one of the best foundations for such a composite scheme is based on the frozen-core CCSD(T) method. In order to mitigate the steep basis set requirements of these calculations, the systematic convergence characteristics of the correlation consistent family of basis sets of Dunning and co-workers can be very effectively exploited via simple extrapolation to the CBS limit. As detailed below, extrapolation of the results of two or more basis sets of increasing size can in many cases reduce the computational burden from requiring a very large sextuple-ζ or septuple-ζ calculation (often including augmenting diffuse functions) to one of only quadruple-ζ. As also discussed below, this can be alternatively treated by using explicitly correlated methods whereby functions with explicit r_{12} dependence are included in the basis set.

In the present work, Sect. 2 briefly reviews some of the more common composite thermochemistry methodologies in use to date in the quantum chemistry literature, while Sect. 3 follows with a detailed discussion of the strategies involved in effectively calculating the major components of these schemes. In Sect. 4, any modifications of these thermochemical strategies in order to calculate "chemically accurate" spectroscopic properties will be addressed with future challenges and a short summary following thereafter in Sects. 5 and 6, respectively.

2 Theoretical thermochemical procedures from the literature

Electronic structure methods designed to achieve "high accuracy" in predictions of thermochemical properties have proliferated over the past several decades. In the following section, we briefly discuss a select subset of methods taken from the chemical literature. This list is

representative of commonly used methods and not meant to be exhaustive.

2.1 Gaussian-n

In 1989, Pople and co-workers introduced a composite procedure, a so-called model chemistry, named "Gaussian-1 theory" (G1) that was intended to predict properties that depended upon energy differences, such as total atomization energies, ionization potentials, as well as proton and electron affinities [4]. G1 attempted to approximate a time-consuming, high-level calculation [QCISD(T)/6-311+G(2df,p)] at a greatly reduced computational cost. It consisted of a fixed collection of well-defined steps that were easily implemented in the popular Gaussian program (it was readily invoked in the Gaussian program with the command line option G1) [5]. G1 blended high-level methods, as judged by the standards of the time, with a correction (based on the H_2 molecule and H atom) to achieve a target accuracy of ± 2 kcal/mol in atomization energies for most of a 31-molecule test set containing elements taken from the first row of the Periodic Table. Errors for seven molecules exceeded 2 kcal/mol, with a maximum error of 4.0 kcal/mol. Subsequent extension to second-row elements yielded an accuracy of approximately ± 3 kcal/mol [6]. It is important to note that the need to approach the complete basis set limit to deal with issues in the 1-particle space was not well understood when G1 was designed since the correlation consistent basis sets had just been developed [1].

Due to the "black box" nature of G1, it could be readily adopted by researchers who preferred to focus on chemical questions using an off-the-shelf, calibrated procedure. The more time-consuming alternative, which was common practice at the time, was to begin most research projects by investing resources in deciding which theoretical method (or combination of methods) was best suited for that project. In this regard, G1 represented a departure from the norm for high-accuracy methods in the 1980s but was in line with Pople's concepts about model chemistries that he had been developing for more than 30 years [7]. Because techniques were seldom applied to large numbers of atoms or molecules, drawing conclusions about the robustness and generality of the method were difficult to reach. The term "G1" served as a convenient abbreviation for a complex, 7-step procedure. However, what began as a widely used communication shorthand was heavily eroded by the explosion of model chemistries in the following decades. It became nearly impossible for the casual user to assimilate the myriad acronyms and oftentimes complex sequence of steps they represented.

Due to G1's focus on energy differences, the developers felt justified in performing most of the individual computational steps at reference geometries obtained from a much lower level of theory (MP2(full)/6-31G*, where "full" means all electrons are correlated). Although the 6-31G* basis set is inherently not capable of describing core/valence correlation effects, the choice of MP2(full) was predicated on the availability of analytic first derivatives for MP2 in the version of Gaussian available at that time. The harmonic frequencies that are needed for computing zero-point vibrational energies were acquired from scaled HF/6-31G* values. The practice of using reference geometries obtained at a much lower level of theory than the ultimate desired level, as well as frequencies (often scaled) from a lower-level calculation, is a general characteristic shared by nearly all model chemistries that were to follow.

The desire for improved accuracy led to the introduction of Gaussian-2 theory (G2) by Pople and co-workers in 1991 [8]. It started with the G1 energy and added a two-part MP2 correction designed to account for the error arising from an additivity assumption in G1 and a basis set correction involving a third *d* function on non-hydrogen atoms and a second set of *p* functions on hydrogen. Finally, an empirical, "higher-level correction" (HLC) based on the number of electron pairs was added. This term was determined from a best fit to the experimental atomization energies of 55 molecules for which the experimental values were well established. An important concept that arose at the time of the G2 method was the need to develop an experimental data set of energetic properties that are known to ± 1 kcal/mol for testing their approach, as well as for the necessary parameterization. For a set of 39 small first- and second-row compounds, the MUD was reduced from 1.4 kcal/mol (G1) to 0.9 kcal/mol (G2). Differences exceeding 5 kcal/mol were still observed in some atomization energies. Modifications to one or more of the steps in G2 intended to reduce the computing requirements led to a number of related methods, including: G2(MP2) and G2(MP3) [9], G2(B3LYP/MP2/CC) [10], G2(MP2,SVP) [11] and G2(MP2,SV) [12].

The next major modification, Gaussian-3 theory (G3), appeared in 1998 [13]. It replaced several steps in the G2 procedure and incorporated a new higher-level correction (still empirical) as well as added corrections for spin–orbit effects in atoms and a core/valence correlation correction. For a test set of 148 experimental enthalpies of formation at 298 K, G3 reduced the MUD from 1.56 (G2) to 0.94 kcal/mol. Nonetheless, numerous cases were found where the unsigned deviations were much larger, e.g., $C_2F_4 = 4.9$, $PF_3 = 4.8$, $Na_2 = 4.0$ and $SO_2 = 3.8$ kcal/mol. As with G2, there were a number of derivative model chemistries based on a range of smaller modifications to the procedure, such as G3(MP2) [14], G3//B3LYP, G3(MP2)//B3LYP

[15], G3(CCSD), G3(MP2,CCSD) [16], G3X, G3X(MP2), G3X(MP3), G3SX [17], G3-RAD and G3X-RAD [18].

In 2007, the development of the Gaussian-4 (G4) procedure was reported by Curtiss et al. [19]. It incorporated five modifications to G3: (1) the Hartree–Fock (HF) component was extrapolated to the basis set limit, (2) an additional set of *d* functions was added to the basis set, (3) CCSD(T) replaced QCISD(T), (4) B3LYP was used for determining the geometries and zero-point vibrational energies and (5) two new higher-level corrections replaced the former ones. However, the correction for scalar relativistic effects was removed and is now included in the HLCs. On the G3/05 test set of 454 experimental energy differences, G4 yielded a MUD of 0.83 kcal/mol versus 1.13 kcal/mol for G3. It should be noted that the experimental energy test set included many ionization potentials, electron affinities and proton affinities, which are in principle easier to calculate than heats of formation from total atomization energies. The root mean square (RMS) deviations were 1.19 kcal/mol (G4) versus 1.67 kcal/mol (G3). The number of molecules where the error exceeded 2 kcal/mol was 35, and the maximum error was 8.9 kcal/mol (ionization potential of B_2F_4). A follow-up paper that same year reported variants G4(MP2) and G4(MP3) whose performance on the G3/05 test set was only slightly worse than G4, e.g., ε_{MAD}[G4(MP2)] = 1.03 kcal/mol [20]. Thus, we see that the original G1 procedure eventually spawned a family of 22 different model chemistries based on four primary forms, each with their own strengths and weaknesses. This pattern of proliferating models will be seen to repeat itself with the other approaches discussed below.

2.2 Petersson-style CBS models

Petersson and co-workers have produced a large body of work focused on procedures for estimating the correlation energy at the complete basis set limit. Their publications cover more than three decades [21]. One of the earliest model chemistries to incorporate the core ideas was the CBS$^{(\infty,3)}$(Full)/DZ+P atomic pair natural orbital model, which was limited to closed-shell systems [22]. An RMS accuracy of ±0.0014 E_h (0.9 kcal/mol) in total energies was claimed with a very modest double-zeta plus polarization quality basis set. A subsequent unrestricted Hartree–Fock-quadratic-CI-based method, CBS-CI(full)/(14s9p4d 2f,6s3pld)/[6s6p3d2f,4s2p1d] was reported to yield total energies accurate to ±0.0012 E_h (0.8 kcal/mol) [23].

Further development work led to several new procedures that were applied to energy differences. The CBS2(FC)/6-311+G**-QCI/6-311+G†† model was found to produce dissociation energies, electron affinities and ionization energies that were only slightly worse than G1 values while being 10 times faster [24]. These conclusions were based on a small set of first-row atoms and diatomic molecules. A further factor of ten reduction in computer time was found with the CBS2(FC)/6-311+G†† model, but the RMS error in dissociation energies more than doubled (1.91 → 4.23 kcal/mol, and the error in electron affinities increased by a factor of 9 (0.18 → 1.58 eV).

Following the release of the previously discussed Gaussian-2 method, further refinements in the approach of Petersson and co-workers led to the CBS-QCI/APNO model [25]. It included a higher-order correction based on QCISD(T)/6-311++G(2df,p) calculations and a size-consistent empirical correction. Tests on a collection of 64 first-row compounds (the method was only defined for first-row elements) showed a MUD of 0.53 kcal/mol for various energy differences (dissociation energies, ionization potentials and electron affinities). Although maximum errors were not explicitly given, several errors exceeding 1 kcal/mol were evident in the included Tables. Additional tests on 20 hydrocarbon bond energies showed an RMS error of 0.9 kcal/mol versus 2.2 kcal/mol for G2 and a maximum error of 2.6 kcal/mol.

A year later, three new methods (CBS-4, CBS-q and CBS-Q) were introduced in an attempt to increase the size of the systems that could be handled to six non-hydrogen atoms by significantly lowering the computational cost [26]. The mean unsigned deviations measured with respect to the 125-member G2 test set were 2.0, 1.7 and 1.0 kcal/mol, respectively, for the three methods. Maximum errors of 5.4, 4.6 and 2.3 kcal/mol were obtained for atomization energies. These three methods and the CBS-QCI/APNO method all exploit the asymptotic convergence of the second-order Møller-Plesset (MP2) pair energies. CBS-Q was later modified to use B3LYP geometries and frequencies, leading to a method labeled CBS-QB3 [27]. That modification produced a reduction in the maximum error for the entire G2 test set from 3.9 to 2.8 kcal/mol. More recently, a modification was made to the CBS-QB3 method to allow the spin correction to be removed. In terms of heats of formation, the resulting ROCBS-QB3 method did not perform quite as well as the method it was meant to replace, ε_{MUD} = 0.91 kcal/mol versus 0.69 kcal/mol for CBS-QB3, based on the G2/97 test set [28].

Radom and co-workers proposed a modification to the CBS-Q method in which the geometries and zero-point vibrational energies were obtained from QCISD/6-31G(d) calculations and the highest level of theory involved CCSD(T) rather than QCI [29]. This method, which was intended for free radical thermochemistry, was called CBS-RAD. It included an empirical correction of 0.30 kcal/mol per triple or double bond.

Like the Gaussian-n methods, the collection of Petersson-style CBS model chemistries has undergone a continual evolution that has resulted in many different

implementations. The CBS chemistries are also readily invoked by keywords in various versions of the Gaussian program system. We are unaware of any direct comparisons between the most recent Gaussian-n and Petersson-style methods across a large body of experimental data.

2.3 Weizmann-n

A decade following the appearance of the Gaussian-1 method, Martin and co-workers introduced the Weizmann-1 (W1) and Weizmann-2 (W2) "black box" model chemistries, adopting a naming convention analogous to the Gn sequence [30]. It was intended that the Wn family of models include all terms that contribute ~0.24 kcal/mol (1 kJ/mol) to atomization energies and that all of the components individually be demonstrated to have converged with respect to the 1-particle basis set to approximately this same level of accuracy. Experimentally derived empirical parameters were not to be used. Some of the earliest Wn implementations did not conform to these admittedly high standards due to the computational cost of certain terms.

Although atomic electron affinities were used to help guide the selection of the best CBS extrapolation formula, the Wn models were primarily intended for the prediction of heats of formation. W1 used B3LYP/cc-pVTZ+1 geometries and zero-point energies, with the latter being scaled by 0.985. The "+1" notation signified the addition of a tight d function for second-row elements. Separate CBS extrapolations were performed on the SCF, CCSD and (T) energy components. The most expensive single-point CCSD(T) calculation involved an aug-cc-pVTZ+2d1f basis set, while the lower-level CCSD calculation used a larger aug-cc-pVQZ+2d1f basis set.

The more expensive W2 included up through h-type basis functions. W1 included a molecule-independent, empirical parameter not present in W2. The former was reported to yield a mean unsigned deviation of 0.30 kcal/mol for atomization energies, while the latter method reduced that to 0.23 kcal/mol (or 0.18 kcal/mol for strongly single configuration dominant species). The maximum errors were 1.01 (W1) and 0.64 kcal/mol (W2). These statistical metrics were based on a collection of 28 small, well-behaved, first- and second-row molecules. For comparison purposes, the corresponding CBS-QB3 and G3 mean unsigned deviations were 0.91 and 0.86 kcal/mol. Using a different test set of 30 molecules, including the more difficult-to-describe O_3, W2 had a MUD of 0.40 kcal/mol and a maximum error of 3.0 kcal/mol [31]. It becomes readily apparent that the error metrics can be quite different from one small collection of molecules to the next. A number of W1 and W2 variants were subsequently introduced, including W1U, W1USc, W1BD and W1RO [32] and W1w, W2w and W2.2 [33]. Mintz et al. discussed additional modifications labeled W2C, W2C-CAS-ACPF and W2C-CAS-AQCC methods [34].

W3, the next major Weizmann-n revision, appeared 5 years later [31]. For W3, the contribution of triple excitations was determined from CCSDT calculations with the cc-pVDZ and cc-pVTZ basis sets. The effects of quadruple excitations were based on CCSDTQ/cc-pVDZ calculations scaled by 1.2532, a value derived from a small training set. Results included atomization energies, ionization potentials and electron affinities. For the same set of 30 small molecules used with the W2 method, W3 provided a smaller mean unsigned deviation (0.22 vs. 0.40 kcal/mol) and reduced the maximum error from 3.00 to −0.78 kcal/mol. A number of variations have also been discussed, including W3a, W4a and W4b.

Karton et al. introduced the Weizmann-4 (W4) model for first- and second-row compounds 2 years later [33]. In this implementation, reference geometries were based on CCSD(T)(FC)/cc-pVQZ calculations. The vibrational zero-point energies were taken from the best available sources, which often translated to a combination of experimental fundamentals and theoretical anharmonic corrections. For a test set of 26 small molecules for which active thermochemical table (ATcT) [35] atomization energies (which are in principle highly accurate) were available, W4 yielded a MUD of 0.15 kcal/mol with an RMS of 0.09 kcal/mol. When the test set was expanded to 30 molecules, both error metrics became 0.15 kcal/mol. Individual cases that exceeded 0.24 kcal/mol (1 kJ/mol) included ClCN (−0.63), PH_3 (+0.31) and O_3 (−0.24). Table I in the Karton et al.'s [33]. paper included a total of 36 molecules, but some of these were excluded from the final statistical analysis. Among these was C_2 ($^1\Sigma_g^+$) which has an even larger higher-order correction than O_3 and for which an ATcT value is available [35]. As with the earlier Wn models, a number of related models were subsequently proposed, including W4lite, W4.2, W4.3 [36] W3.2 [33] and W4.4 [36]. Thus, we see that while the Weizmann-n model chemistries are much younger than either the Gaussian-n or the Petersson-style CBS methods, they nonetheless have spawned many variations as well.

Karton and Martin have also discussed using the W4 electronic energies evaluated at fixed geometries, i.e., not the reference CCSD(T)(FC)/cc-pVQZ geometries used in the standard application of W4, to map out portions of the potential energy surface in order to predict high-accuracy spectroscopic properties such as bond lengths, harmonic frequencies and anharmonicities [37].

2.4 High-accuracy extrapolated ab initio thermochemistry

In 2004, Tajti et al. introduced the high-accuracy extrapolated ab initio thermochemistry (HEAT) model which was

intended to "achieve high accuracy for enthalpies of formation of atoms and small molecules" without resorting to empirical scale factors [38]. Reference geometries were determined at the CCSD(T)(CV)/cc-pVQZ level of theory, and open-shell systems were described with unrestricted Hartree–Fock (UHF) wavefunctions. The CCSD(T)(CV) basis set limit was estimated by a $1/l_{max}^3$ extrapolation [39] of aug-cc-pCVQZ and aug-cc-pCV5Z energies. A frozen-core higher-order correction was determined by combining a $1/l_{max}^3$ extrapolation of CCSDT/cc-pVTZ and CCSDT/cc-pVQZ energies with a CCSDTQ/cc-pVDZ energy. Anharmonic zero-point vibrational energies were evaluated at the CCSD(T)(CV)/cc-pVQZ level of theory. For a test set of 26 small molecules involving 5 atoms (H, C, N, O and F), HEAT produced a mean unsigned deviation of 0.09 kcal/mol relative to well-established experimental ΔH_f(0 K) values. The maximum observed error was 0.33 kcal/mol (C_2H_2).

Two years later, Bomble et al. discussed a number of improvements that lead to the HEAT345-Q, HEAT345-(Q), HEAT345-Q(P) and HEAT345-Q(P) methods [40]. When applied to a set of 18 small molecules, these variants yielded mean unsigned deviations that were marginally better than the original HEAT method. The most accurate was HEAT345-QP which had a MUD that was 0.01 kcal/mol smaller than the original HEAT. The maximum error declined from 0.17 kcal/mol (HEAT) to 0.12 kcal/mol (HEAT345-QP).

In 2008, Harding et al. explored the impact of extending the basis sets used in the HEAT protocol up through aug-cc-pCV6Z and including a diagonal Born–Oppenheimer correction. This led to the HEAT456-Q and HEAT-456QP methods [41]. A statistical comparison based on a collection of 18 small molecules for which ATcT values were available showed small *increases* in the RMS errors for the larger basis set variants, e.g., 0.09 (HEAT345-Q) versus 0.14 (HEAT456-Q) kcal/mol. The maximum errors were also found to increase slightly, e.g., 0.17 (HEAT345-Q) versus 0.29 (HEAT456-Q) kcal/mol. This suggests that the earlier implementations may have benefited from a fortuitous cancellation of errors, as have many of the other methods described in the current paper. The authors commented that HEAT was not intended strictly for the purpose of computing atomization energies, but "rather for estimating the total energies of molecules".

2.5 Correlation consistent composite approach

The MP2-based correlation consistent composite approach (ccCA) of Cundari, Wilson and co-workers was proposed considerably later than the first four families of model chemistries as an alternative to the Gaussian-n methods [42]. The design goal was to achieve a mean unsigned deviation of ±1 kcal/mol for the same energetic quantities targeted by Gn (enthalpies of formation, electron affinities, ionization potentials and proton affinities) but without reliance on empirical corrections. Compared to the Wn models, ccCA is computationally faster and therefore can be applied to much larger systems. There have already been many variations of this basic approach. The first paper in this series discussed five different implementations (ccCA-DZ, ccCA-TZ, ccCA-aTZ, ccCA CBS-1 and ccCA CBS-2). All of them used reference geometries and scaled harmonic frequencies obtained at the B3LYP/6-31(d) level of theory. The first three of these begin with an MP4(FC)/cc-pVnZ (n = D,T) or MP4(FC)/aug-cc-pVTZ calculation. The final two variations of ccCA involve an extrapolation to the MP2(FC) complete basis set limit using aug-cc-pVnZ (n = D,T,Q) basis sets with either an exponential [43] or mixed Gaussian/exponential formula [44]. Finally, a QCISD(T)/cc-pVTZ correction was added to better account for correlation effects. On a collection of 28 molecules taken from the G2-2 and G3/99 test sets, the ccCA-CBS2 method yielded a mean unsigned deviation of 1.16 kcal/mol after incorporating additional corrections for atomic spin–orbit effects and scalar relativistic effects based on first-order stationary direct perturbation theory values reported by Kedziora et al. [45]. In several cases, the error in ΔH_f exceeded 3 kcal/mol, e.g., $AlCl_3 = -3.8$, $SF_6 = -3.8$ and $C_2F_4 = -3.5$ kcal/mol.

Subsequent modifications of ccCA (ccCA-F, ccCA-P, ccCA-S4, ccCA-S3 and ccCA-WD) explored the use of different MP2/CBS extrapolation formulas [46]. Other changes included the following: (1) replacing QCISD(T) with CCSD(T), (2) using MP2 Douglas–Kroll calculations for computing scalar relativistic corrections and (3) using cc-pV(n + d)Z basis sets for second-row elements. The MUD for the complete G3/99 test set (223 enthalpies of formation, 88 ionization potentials, 58 electron affinities and 9 proton affinities) was 0.96 and 0.97 kcal/mol with the mixed Gaussian/exponential and $1/(l_{max} + 1/2)^4$ formulas, respectively. This compares to 0.95 kcal/mol for G3X. In cases where the ccCA errors for ΔH_f were unusually large, e.g., $B_2F_4^+ = 7.7$, $C_{10}H_8$ (azulene) = −5.3 kcal/mol, the authors recommended a re-examination of the experimental data.

A year after the ccCA method was announced, further changes were introduced in order to handle transition metal compounds, leading to the ccCA-TM variant [47, 48]. A much looser target of "transition metal chemical accuracy", corresponding to ± 3 kcal/mol, was adopted. Further modifications intended to allow the method to describe potential energy surfaces were introduced in the multireference ccCA (MR-ccCA) method [49]. By replacing the standard self-consistent field and MP2 steps with the corresponding resolution of the identity (RI) counterparts,

Prascher et al. reported improved computational efficiency with the development of the RI-ccCA and RI-ccCA+L methods [50]. The +L notation indicates that the density fitting, local CCSD(T) method of Schütz and Werner was used [51, 52]. The average savings in computer time for the RI-ccCA method was 72%. With the +L alternative, the computational savings increased slightly to 76%. However, as might be expected, both approximations introduced additional error relative to ccCA. For RI-ccCA, the average unsigned error relative to ccCA is 0.27 kcal/mol, compared to 2.63 kcal/mol for RI-ccCA+L on a test set of 120 molecules. The maximum positive and negative errors are as follows: RI-ccCA = 3.81 and −3.56 kcal/mol, RI-ccCA+L = 9.20 and −1.77 kcal/mol.

Five more modifications to ccCA were proposed in 2009, including changes to the basis set used in the B3LYP calculations, separate extrapolations of the SCF and MP2 energies and the use of new vibrational scale factors [53]. When applied to the G3/05 test set (454 comparisons), the latest version (ccCA-PS3) produced a mean unsigned deviation of 1.01 kcal/mol. The corresponding G4(MP2) value was 1.04 kcal/mol. However, maximum errors were not reported.

2.6 Focal-point analysis

All of the approaches considered to this point fall under the category of model chemistries. Like the original G1 method, they consist of a multi-step, fixed recipe implicitly defined for a subset of the periodic table. In many cases, this meant hydrogen and the main group elements from the first two rows. The focal-point approach (FPA) of Allen, Császár and co-workers differs in that it is more of a flexible strategy that can be tailored for particular research interests [54, 55]. In the initial 1993 report, the method was applied to the determination of the heats of formation of NCO ($^2\Pi$) (cyanato radical) and HCNO (isocyanic acid) using a mixture of experimental and theoretical reference geometries. For NCO, a CISD/DZ(d,p) optimized structure was adopted, although the CO bond length differed substantially from the experimental r_0 value. Five independent reactions, involving a mixture of experimental and theoretical data, were used to determine the NCO ΔH_f(0 K). A variety of basis sets, the largest of which was an [13s,8p,6d,4f/8s,6p,4d] contraction, were combined with various levels of perturbation theory and coupled cluster theory, all of which utilized the frozen-core approximation. Essentially, CCSD(T)/(2d,1f/2p,1d) reaction energies were adjusted by MP2 energies evaluated with the largest basis set, allowing the very expensive CCSD(T) calculations in the large basis set to be avoided. The analysis was carried out using two sequences of methods: MP2 → MP3 → MP4(SDTQ) → MP5(SDTQ) and MP3 → CCSD → CCSD(T).

Calibration calculations on some prototypical bonds, e.g., $CO_2 \rightarrow O + CO$ and $N_2 \rightarrow N + N$, showed that the FPA level of theory underestimated bond strengths by as much as 6.4 kcal/mol (N_2). Consequently, bond additivity corrections were adopted for the four (out of five) reactions involving homolytic bond cleavage.

A 1998 focal-point study of conformational energies (e.g., the torsional barrier in ethane) replaced the basis sets used in earlier studies with members of the correlation consistent basis set family up through cc-pV6Z, aug-cc-pV5Z and cc-pCVQZ [56]. In some cases, it was necessary to use a hybrid cc-pVTZ(C,N,O)/cc-pVDZ(H) basis set. Correlation recovery was accomplished through fifth-order perturbation theory or through coupled cluster calculations at the CCSD, CCSD(T) and CCSDT levels. Geometries were optimized with CCSD(T)(CV)/cc-pVTZ or CCSD(T)(CV)/aug-cc-pVTZ calculations. The largest basis set used in a CCSD(T) calculation varied from system to system. In the case of ethane, the largest basis set was aug-cc-pVTZ. Relativistic corrections were determined with first-order perturbation theory using the mass velocity and one-electron Darwin terms [57]. The diagonal Born–Oppenheimer correction was computed at the Hartree–Fock level of theory [58]. Once again, there were two sequences of correlation recovery procedures, one based on perturbation theory through fifth order and another based on coupled cluster theory through CCSDT. A Padé approximant was used to extrapolate the perturbation theory sequence to infinite order. The authors described this as a two-dimensional extrapolation grid in the 1-particle and n-particle expansions.

There have been a substantial number of studies involving the focal-point method, but a comprehensive survey is beyond the scope of the present work. As an illustration of a more up-to-date application of the method, we consider a 2004 study which re-examined the heats of formation of NCO and HNCO by Allen and co-workers [59]. An early focal-point study had reported ΔH_f(0 K) values of 35.3 kcal/mol (NCO) and −26.1 kcal/mol (HNCO) [60]. Another study that same year by the same authors offered revised values of 31.4 ± 0.5 kcal/mol (NCO) and −27.5 ± 0.5 kcal/mol (HNCO) [55]. In the 2004 investigation, the heat of formation of NCO was established via two formation reactions. The corresponding heat of formation for HNCO involved seven additional reactions. After accurately determining the reaction energies, reference heats of formation for component species were taken from experiment or in a few cases from theory. This approach seeks to avoid computing total atomization energies in favor of the use of isogyric reactions. The MPn (n = 2–5) scheme used in previous FPA studies was abandoned in favor of the HF → MP2 → CCSD → CCSD(T) → CCSDT sequence of methods. The largest

CCSD(T) calculation employed the cc-pV5Z basis set. Separate extrapolations were performed on the Hartree–Fock and correlation energy components. Core/valence correlation effects were treated with up to CCSD(T)(CV)/cc-pCVQZ and MP2(CV)/cc-pCV5Z calculations, followed by extrapolation to the CBS limit.

For closed-shell systems, the MP2/CBS limit was estimated with explicitly correlated MP2-R12/A calculations performed with an uncontracted cc-pV5Z basis set augmented with additional tight and diffuse functions. Open-shell systems were described with conventional UHF-MP2(FC)/cc-pVnZ (n = D–5) followed by extrapolation to the CBS limit. In their previous studies, the highest level of correlation recovery was obtained from CCSDT computations. For the newer work, the impact of quadruple excitations was explored via a series of homolytic bond breaking reactions, three of which involved a single bond and four of which involved a multiple bond. For this purpose, the authors chose the Brueckner orbital coupled cluster theory, BD(TQ) and CCSD(2) methods. Although the impact of quadruple excitations was sometimes found to be "not negligible", the two approaches to including quadruples gave "disconcerting inconsistencies" and an "anomalously large correction" in certain cases. Therefore, they recommended not attempting to correct for quadruple excitations until full-blown CCSDTQ calculations were affordable for all of the species in their formation reactions. The final recommended 0 K heats of formation were 30.5 ± 0.2 kcal/mol (NCO) and −27.6 ± 0.2 kcal/mol (HNCO). The latter is in essentially exact agreement with the 1993 value of −27.5 ± 0.5 kcal/mol, and the former falls slightly outside the error bars of the earlier value.

To the best of our knowledge, there have been no surveys comparing the performance of the focal-point approach to other approaches for high-accuracy thermochemistry across a broad range of molecules.

2.7 Feller-Peterson-Dixon procedure

The general strategy of this approach developed by the current authors for the studies of thermochemical or spectroscopic properties such as molecular structures and vibrational frequencies is similar in spirit to the focal-point analysis technique of Allen, Császár and co-workers in that it is a flexible, multi-step approach intended to address all of the major sources of error, as opposed to a fixed recipe model chemistry. For the sake of this discussion, the approach will be referred to as the Feller–Peterson–Dixon (FPD) procedure that has been under development for almost 15 years. The overall goal is to include all physically significant effects that contribute to the property of interest at a level that will guarantee results within the target accuracy. For example, if the goal is computing the dissociation energy of a heavy element diatomic like HI ($^1\Sigma^+$) to an accuracy of ≤0.5 kcal/mol, a second-order molecular spin–orbit correction would be included even though for lighter elements the contribution is negligible [61, 62]. Consequently, the basic approach can be easily modified to include effects that contribute significantly in some cases and not in others. This characteristic enables it to address many properties across a large portion of the periodic table. An important feature of the approach is that, whenever possible, an attempt is made to gauge the uncertainty in each of the (up to) eight components required to achieve a final value. Even a rough knowledge of the uncertainty in the individual components improves our ability to avoid combining pieces with markedly different accuracies, so that computational resources can be directed to the most problematic areas where they can be used most effectively. While this approach differs in detail from FPA and the other high-level approaches already discussed, it also shares many common features.

For thermochemical studies, the initial step in the procedure involves a series of frozen-core CCSD(T) geometry optimizations using the diffuse function–augmented correlation consistent sets, denoted aug-cc-pVnZ and aug-cc-pV(n + d)Z, n = D, T, Q, 5,…,10. The wide variation in available basis set sizes allows the treatment of fairly large systems, such as C_8H_{18}, [63] with up through quadruple-ζ quality sets or the ability to employ sets of 10-ζ quality for atoms [64, 65]. The largest basis set possible for the system of interest is used in order to reduce the reliance on basis set extrapolation procedures, e.g., use of aug-cc-pV6Z through aug-cc-V8Z whenever they are computationally tractable [3, 62, 66–68]. In the FPD approach, a combination of up to five CBS extrapolation formulas is utilized, taking the average as the best estimate and half the spread in the values as a crude measure of the associated uncertainty. The application of these formulas have been discussed in detail elsewhere [62]. A very recent study of the effectiveness of these formulas across a collection of 141 small-to-medium size chemical systems for which reliable estimates of the CBS limit were available supports the accuracy of this approach (see below) [3]. Open-shell calculations are based on the R/UCCSD(T) method, which is based on restricted open-shell Hartree–Fock (ROHF) orbitals but allows a small amount of spin contamination in the solution of the CCSD equations for open-shell systems. Calculations on isolated atoms imposed full atomic symmetry on the orbitals. As in the HEAT or Wn methods, higher-order (HO) correlation effects are handled with the CCSDT and CCSDTQ or CCSDT(Q) sequence of methods. Combinations such as CCSDT/cc-pVQZ+CCSDTQ/cc-pVTZ or CCSDT/cc-pVTZ+CCSDTQ/cc-pVDZ are typically used to provide a balanced treatment. When both combinations are affordable, the difference between them

provides a measure of the uncertainty in the HO correction. In a limited number of cases, CCSDTQ5 or even FCI calculations have been performed. In most other instances, the FCI result has been estimated by using a continued fraction (cf) approximant that involves the CCSD, CCSDT and CCSDTQ sequence of energies [62, 69].

Corrections for outer core/valence (CV) correlation effects are treated with the weighted core/valence basis sets [70], cc-pwCV*n*Z, n = D, T, Q and 5 with extrapolation to the CBS limit using the same procedure as used for the frozen-core energies. It is also possible to compute the CV correction using higher-order correlation methods, but the calculations are extremely time-consuming. For molecules with 4–6 "heavy" atoms, it seems that generally the HO CV correction typically falls into the 0.1–0.2 kcal/mol range. Thus, they are only required in studies aiming for the highest possible accuracy. Scalar relativistic (SR) effects are described with second-order Douglas-Kroll-Hess (DKH) CCSD(T)(FC) calculations [71, 72] using the cc-pVnZ-DK basis sets [73]. Spin–orbit coupling effects in the atoms are taken from the experimental zero-field splittings, while, when applicable, molecular SO corrections are taken either from experiment or calculated with the state-interacting approach using CI wavefunctions.

Anharmonic zero-point vibrational energies (ZPVEs) for diatomic molecules are obtained from sixth-degree Dunham fits of the potential energy curves [74]. For triatomic systems fits to potential energy surfaces are often used with ZPVEs calculated using second-order vibrational perturbation theory. Finally, anharmonic ZPEs for general polyatomic species are determined by combining CCSD(T) harmonic frequencies with anharmonic corrections obtained from second-order Møller-Plesset perturbation theory (MP2). For example, CCSD(T)(FC)/aug-cc-pVQZ harmonic frequencies might be combined with MP2(FC)/aug-cc-pVTZ anharmonic corrections. Whenever possible, the degree of convergence in the ZPE is tracked as the underlying basis sets are improved.

Another factor considered in studies aiming for very high accuracy, especially when the molecule contains hydrogen atoms, is the effect of the diagonal Born–Oppenheimer correction (see, e.g., Ref. [75]). Typically, the FPD approach uses the aug-cc-pVTZ basis set at either the CISD or CCSD levels of theory to evaluate this component.

As an indication of the performance of the FPD procedure for atomization energies, a mean signed deviation of −0.04 kcal/mol, RMS = 0.28 kcal/mol and MAD = 0.17 kcal/mol is found when comparing against 121 molecules whose experimental uncertainties are ±1 kcal/mol or less [62]. If the comparison is restricted to molecules with an experimental uncertainty of ±0.3 kcal/mol or less, the MAD falls to 0.10 kcal/mol (61 comparisons).

Molecular structures are also accurately predicted with this approach. Mean unsigned deviations for CCSD(T)(FC) bond lengths between non-hydrogen atoms drop from 0.029 Å (281 comparisons) with the aug-cc-pVDZ basis set to 0.012 Å with aug-cc-pVTZ and 0.007 Å with aug-cc-pVQZ [62]. At the estimated basis set limit and including all of the minor corrections discussed above except the DBOC, a MAD value of 0.0006 Å is found (103 comparisons). Specific comparisons with accurate semi-experimental bond lengths for polyatomic molecules have been discussed elsewhere [76].

3 Components of an accurate composite methodology

Among the composite methods reviewed above, many include a common collection of similar contributions. In this section, the most important are discussed with an emphasis on the calculation of molecular atomization energies (needed for heats of formation) since these are among the most demanding properties to predict. In general, the conclusions based on this particular property also carry over to equilibrium geometries and vibrational frequencies.

3.1 Estimating the frozen-core, CCSD(T) complete basis set limit

Notwithstanding molecules containing alkali and alkaline earth metals, the basis set incompleteness error in the correlation of the *valence* electrons tends to be one of the largest sources of error in the calculation of an atomization energy, particularly when the accurate CCSD(T) method is used. A detailed discussion of the error introduced separating the frozen-core and core correlation contributions is given in the next section. Figure 1 shows the basis set convergence of the total energies and dissociation energies toward the CBS limit for the diatomic series N_2 through Bi_2 using a sequence of correlation consistent basis sets. For these results the aug-cc-pV*n*Z sets [1, 77] were used for N_2, aug-cc-pV(n + *d*)Z for P_2, [78] and aug-cc-pV*n*Z-PP was used [79] for As_2 through Bi_2. The -PP (pseudopotential) basis sets are accompanied by a Stuttgart-Köln, energy consistent, small-core effective core potential that replaces 10 electrons in As, 28 electrons in Sb and 60 electrons in Bi. The CBS limits were obtained by extrapolation of the total energies (n = 4 and 5 for QZ and 5Z, respectively) by the formula (see below)

$$E(n) = E(CBS) + A\left(n + \frac{1}{2}\right)^{-4}. \quad (1)$$

Whereas there is considerable spread in the basis set error at the double-zeta level, smooth regular convergence is

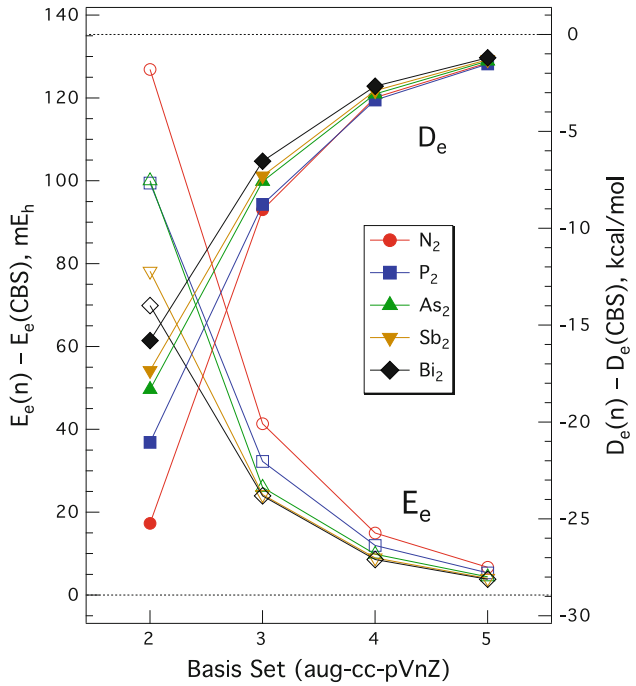

Fig. 1 Convergence of the frozen-core CCSD(T) total and dissociation energies with respect to the CBS limit of N_2 through Bi_2 as a function of the correlation consistent basis set size (aug-cc-pVnZ, aug-cc-pV(n + d)Z, or aug-cc-pVnZ-PP)

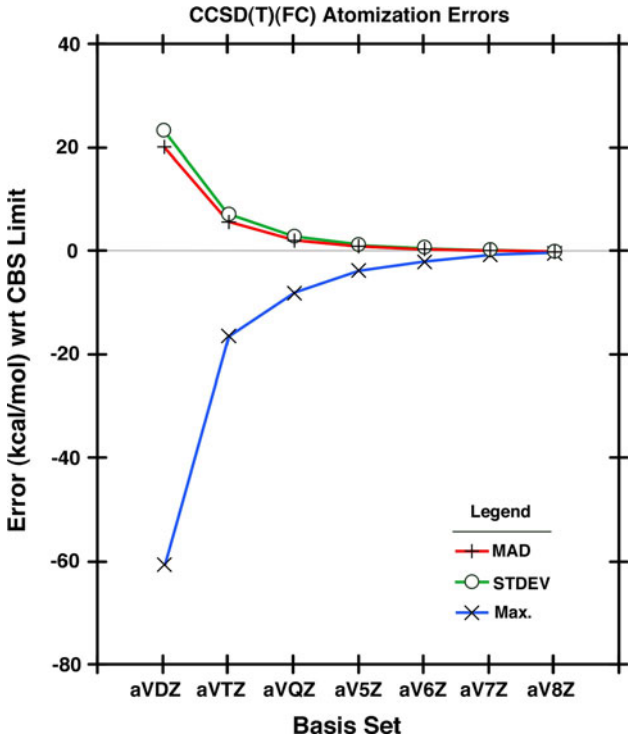

Fig. 2 Frozen-core CCSD(T) atomization energy errors (kcal/mol) with respect to the estimated complete basis set limit as a function of the correlation consistent basis set size (aVnZ = aug-cc-pVnZ and/or aug-cc-pV(n + d)Z) for 141 molecules

observed. Curiously, at least for these simple non-polar molecules, the convergence rate is the slowest for N_2 and fastest for Bi_2. Presumably this is related to the decrease in the correlation energy between these species. To the extent one can generalize from the data in this figure, it seems likely that to obtain an accuracy of 1 kcal/mol in the atomization energy from a calculation with a single finite basis set, one would need a set of at least aug-cc-pV6Z quality. This conjecture is strengthened by Fig. 2, which demonstrates the slow convergence of atomization energies for 141 molecules of various sizes, mostly consisting of elements of the first and second rows. Beginning with the aug-cc-pVTZ basis set, each increment in basis set index cuts the error metrics by approximately a factor of two. Depending on which error metric one chooses to employ, achieving an accuracy of ±0.5 kcal/mol or less in this component without the use of extrapolation techniques to the CBS limit would require the use of basis sets on the order of aug-cc-pV7Z. Such calculations are currently only possible for very small molecules. This analysis was based on data contained in the Computational Results Database (CRDB) [80]. The basis set errors can also be strongly influenced by the molecular size. Figure 3 shows the basis set errors in the CCSD(T) atomization energies for a series of simple alkanes from C_2H_6 to C_6H_{14} [68]. While the results for the smallest member of this series mimics the behavior seen in Figs. 1 and 2, there is a strong size dependence capped by the nearly 10 kcal/mol error for C_6H_{14} at the aug-cc-pVQZ level. Whereas it is clear that for a size-extensive method like CCSD(T), the errors of the individual energies should scale with system size, it is perhaps a bit disconcerting that the relative energy should also display such a strong size dependence. This would appear to indicate that as the molecules increase in size not only is the correlation treatment more computationally intensive due to the growth in the number of electrons, but a larger basis set must also be used to maintain the same level of accuracy in the atomization energy. As discussed in detail below, this can be mitigated by effective basis set extrapolation strategies.

In lieu of basis set extrapolation, there is a more fundamental technique to improve the basis set convergence rate at the correlated level, namely the use of explicitly correlated methods. The slow convergence of the correlation energy with basis set size is due to the poor description of the electron coalescence cusp by products of one-electron functions when the interelectronic distance r_{12} becomes small. Recently developed F12 methods [81, 82] utilize a Slater-type correlation factor, $F_{12} = e^{-\gamma r_{12}}$, and exhibit much faster convergence toward the CBS limit. Figure 4 compares the convergence of conventional CCSD(T) and CCSD(T)-F12b atomization energies toward

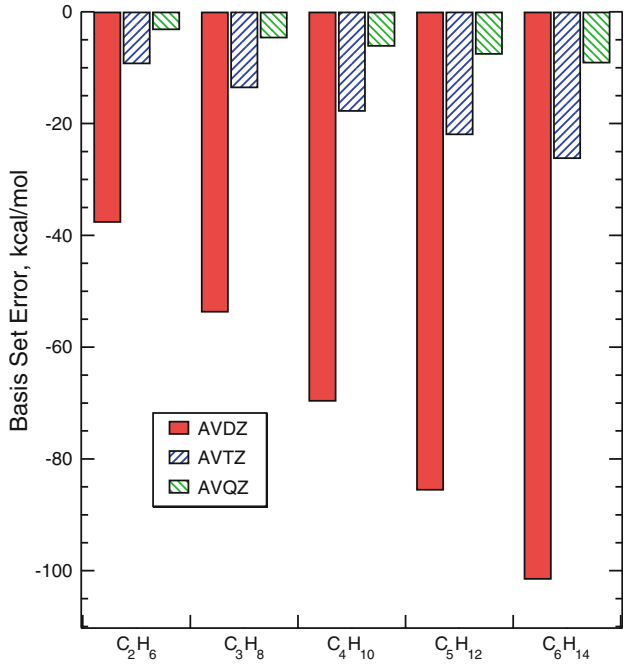

Fig. 3 Convergence toward the CBS limit of the frozen-core CCSD(T) atomization energies for a series of simple alkanes (AVnZ = aug-cc-pVnZ, n = D, T, Q)

the CBS limit for a set of 51 small molecules [3]. The F12b calculations [83] have used the cc-pVnZ-F12 sequence of basis sets [84], which are similar in size to the standard aug-cc-pVnZ [aug-cc-pV(n + d)Z for the second row) sets. From Fig. 4, it is clear that the use of F12 methods improves the basis set convergence rate by about two basis set index (n) levels, e.g., the cc-pVDZ-F12 set has errors similar to conventional calculations utilizing aug-cc-pVQZ basis sets and the cc-pVQZ-F12 results are very similar to those obtained from computationally expensive aug-cc-pV6Z calculations. Some of the benefits of F12 calculations exhibited in Fig. 4 are accomplished by a reduction in the HF basis set error by the effective use of the auxiliary basis set required in the explicitly correlated treatment [85]. The evaluation of additional multi electron integrals does result in an increased computational cost for F12 calculations compared to the analogous conventional cases, but, with the CCSD(T)-F12b method, the conventional (T) correction begins to again dominate the calculation already at the cc-pVTZ-F12 level. From the results shown in Fig. 4, chemical accuracy in the atomization energies of these small molecules is obtained on average with just the cc-pVTZ-F12 basis set and in the worst case with cc-pVQZ-F12. As shown in Fig. 5 for the same saturated hydrocarbon series as in Fig. 3, while a similar increase in basis set error with system size is evident (although perhaps now approaching a maximum at C_5H_{12}), even the C_6H_{14} molecule is nearly converged to within ~1 kcal/mol at the CCSD(T)-F12b/cc-pVQZ-F12 level of theory.

Given the regular convergence patterns clearly observed in Figs. 1, 2, 3, 4, and 5, it is not surprising that there have been a tremendous number of publications proposing and testing various basis set extrapolation procedures, nearly exclusively using correlation consistent basis sets. It is outside the scope of the present contribution to review this literature, but the reader is referred to the recent comprehensive study of Ref. [3]. A total of 5 commonly used CBS extrapolation formulas were studied, as well as an average of the results obtained when these formulas are each used on the total energies. The latter approach has generally

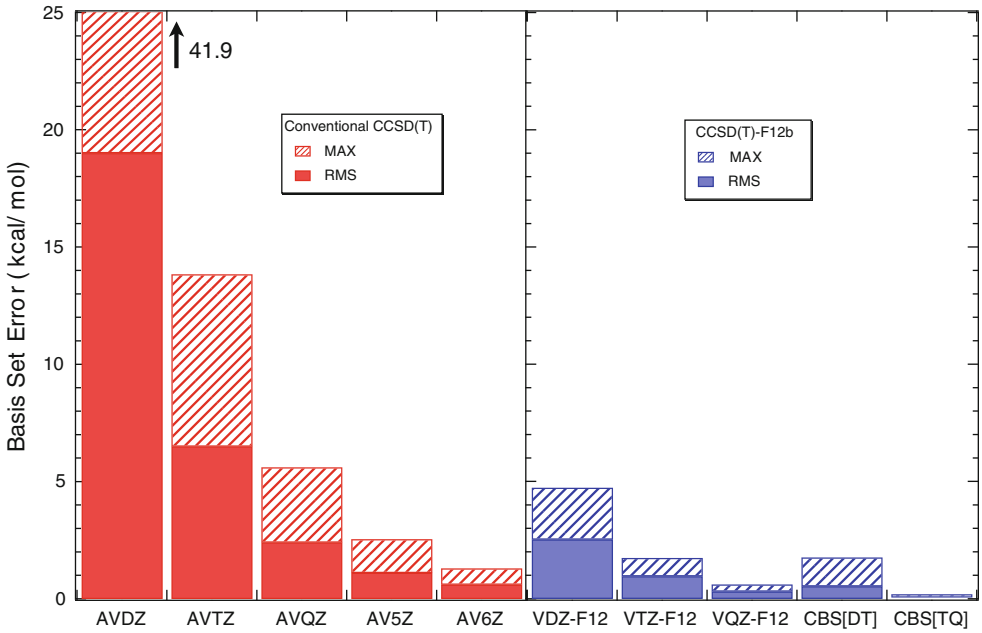

Fig. 4 Comparison of the convergence toward the CBS limit of frozen-core conventional CCSD(T) and explicitly correlated CCSD(T)-F12b calculations of the atomization energies of 51 small molecules. Also shown are the F12b CBS extrapolated values, CBS[DT] and CBS[TQ]. See the text. (AVnZ = aug-cc-pVnZ or aug-cc-pV(n + d)Z and VnZ-F12 = cc-pVnZ-F12, n = D, T, Q)

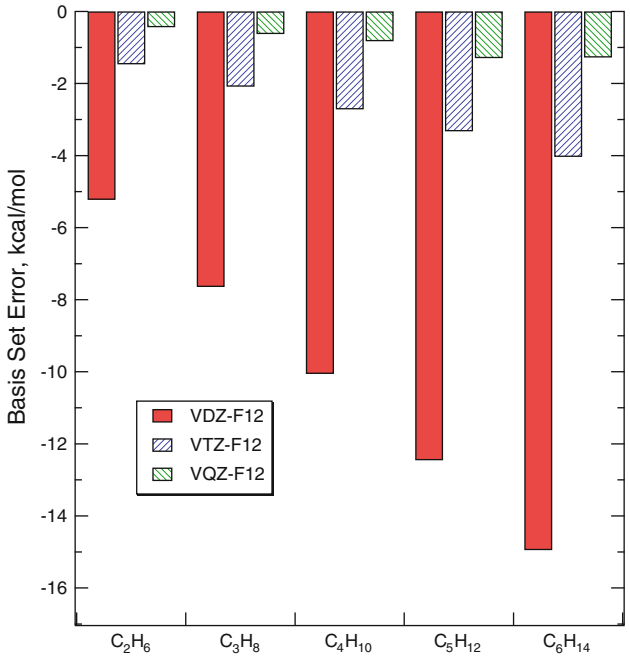

Fig. 5 Convergence toward the CBS limit of the frozen-core CCSD(T)-F12b atomization energies for a series of simple alkanes

been used in the FPD composite method outlined above. With a reference set of 141 relatively small molecules, Eq. 1 applied to the total energies and the 5-formula average exhibited the best statistical results on the whole. Both the MUD and RMS across the entire 141 molecule dataset are shown in Fig. 6 for four of the approaches described in Ref. [3]. In Fig. 6, the results for the popular $l_{\max}^{-3}$ formula [39] involved a separate extrapolation of the HF energy using the Karton and Martin relation [86] followed by an extrapolation of the correlation energy by $E(n) = E(CBS) + An^{-3}$. Whereas the RMS error is very similar to that of Eq. 1, the maximum error is noticeably larger and was generally positive in sign. Compared to the basis set errors of Fig. 2, all of the formulas shown in Fig. 6 are very successful at reducing the basis set incompleteness error; both the RMS and mean unsigned errors decrease by factors of 3–4. Given the smooth convergence with basis set shown in Fig. 4 for the CCSD(T)-F12b method, it is not surprising that CBS extrapolations can be beneficial in these calculations as well. Also shown in Fig. 4 are the results of extrapolating both DZ/TZ and TZ/QZ basis set pairs using the Schwenke-style approach [87] of Hill et al. [88]. For the small DZ/TZ pair, the RMS error is decreased by a factor of two compared to the raw cc-pVTZ-F12 result, but no improvement is observed for the maximum error. On the other hand, extrapolation of the TZ/QZ pair yields results on par with the estimated accuracy of the reference data, i.e., about $\sim$0.1 kcal/mol. Both conventional and F12 extrapolations were used previously on the series of alkanes in Figs. 3 and 5. In these cases, the DZ/TZ extrapolations from the F12 calculations were in excellent agreement with the more computationally intensive F12 TZ/QZ results, as well as the conventional TZ/QZ extrapolations, although the latter had relatively large estimated uncertainties.

In summary, in order to reliably converge the frozen-core CCSD(T) atomization energies to better than 1 kcal/mol one has at least three choices:

1. perform a conventional CCSD(T) calculation with basis sets of aug-cc-pV6Z to aug-cc-pV7Z quality,

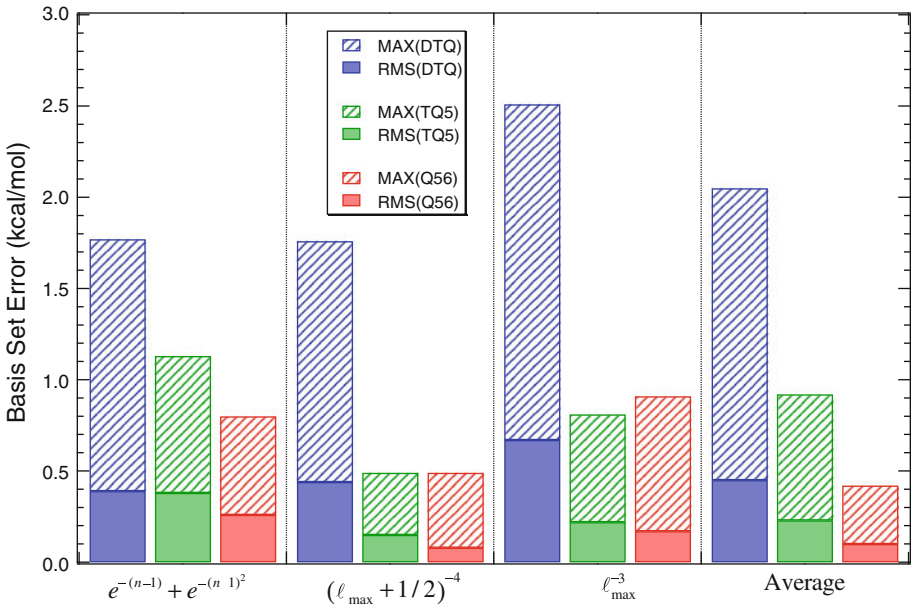

Fig. 6 A comparison of the effectiveness of various extrapolation formulas. Only the $\ell_{\max}^{-3}$ results employed separate extrapolations of the HF and correlation energy components. Except for the mixed exponential/gaussian formula, only the largest two basis sets in each group were used in the extrapolations. See the text

2. use conventional basis sets of at least aug-cc-pVQZ and aug-cc-pV5Z quality to be used with a CBS extrapolation with a formula such as Eq. 1, or
3. perform a CCSD(T)-F12b/cc-pVQZ-F12 calculation.

In the above summary, it is assumed that the outcome of any given calculation might result in the maximum possible errors that are shown in Figs. 2, 4 and 6. To obtain accuracies in the 1–2 kJ/mol range, it seems clear that either a basis set extrapolation using conventional CCSD(T) with sets of at least aug-cc-pV6Z quality must be used or a TZ/QZ extrapolation at the CCSD(T)-F12b level of theory using cc-pV*n*Z-F12 basis sets.

3.2 Contributions due to correlation of outer-core electrons

Since the estimation of the frozen-core CBS limit is presumed to be the dominant source of error in this CCSD(T)-based procedure, it is important that this and all subsequent contributions be converged with respect to basis set and level of theory to much better than 1 kcal/mol so that the final composite result will achieve chemical accuracy. For the recovery of core correlation effects, only the outer-core shell of electrons is usually treated, i.e., the principal shell of electrons lying just below the valence electrons (1*s* for 1st row, 2*s*2*p* for 2nd row, etc.). Basis sets specially designed for correlating these types of electrons, e.g., cc-pwCV*n*Z or cc-pCV*n*Z [70, 89], must be used in order to minimize spurious basis set superposition errors. In principle, the combination of core and valence correlation could be combined into a single contribution by using a sequence of core/valence correlation consistent basis sets. Based on the discussion in the previous section, large sets of at least aug-cc-pwCV5Z sizes would be required for chemical accuracy (with CBS extrapolations), while very large cc-pwCV6Z or larger basis sets would be needed for more accurate work. As mentioned above, this is the procedure utilized in the HEAT method. All other composite schemes choose to separate the two contributions, and the rationale relies on the fact that the effect of core correlation on properties such as atomization energies generally converges much more rapidly with basis set than the frozen-core contribution.

As a simple example, consider again the diatomic series N_2 through Bi_2. In these cases, the definition of outer-core correlation ranges from including just the (4) 1s electrons in N_2 to the (20) 5*d* electrons in Bi_2 (it should be noted that generally the (*n*-1)*sp* electrons are included in the outer-core definition for the post-*d* elements but these relatively small contributions have been neglected in this discussion). Figure 7 shows the convergence with basis set of the CCSD(T) core correlation effects on the dissociation energies, i.e., $\Delta D_e = D_e(\text{valence} + \text{core}) - D_e(\text{valence})$ using aug-cc-pwCV*n*Z basis sets [70] (aug-cc-pwCV*n*Z-PP for As, Sb, and Bi) [90] on *both* the frozen-core (valence) and core correlated (val + core) calculations. The CBS limits were obtained via extrapolation with Eq. 1. The use of core/valence basis sets for both calculations (FC and core correlated) is essential since the tight functions that are included in the core/valence basis sets do make small contributions to the frozen-core correlation energy. Using a basis set such as cc-pwCV*n*Z for valence + core correlation but using only cc-pV*n*Z in the frozen-core calculation will result in an overestimation of the core correlation contribution. Even if the results calculated in this way are extrapolated to the CBS limit, the problem can persist since the accuracy of the frozen-core extrapolation will depend on whether valence or core/valence basis sets are used. Returning to Fig. 7 and comparing these results to those of Fig. 1 clearly show that the core correlation contributions to D_e are much more rapidly convergent with basis set than the frozen-core dissociation energies. In the former case, convergence to within 1 kcal/mol is achieved with core/valence basis sets two levels smaller in the basis set index than what was required for the FC dissociation energy. It should be stressed, however, that the cc-pwCV*n*Z series of core/valence basis sets converge this contribution

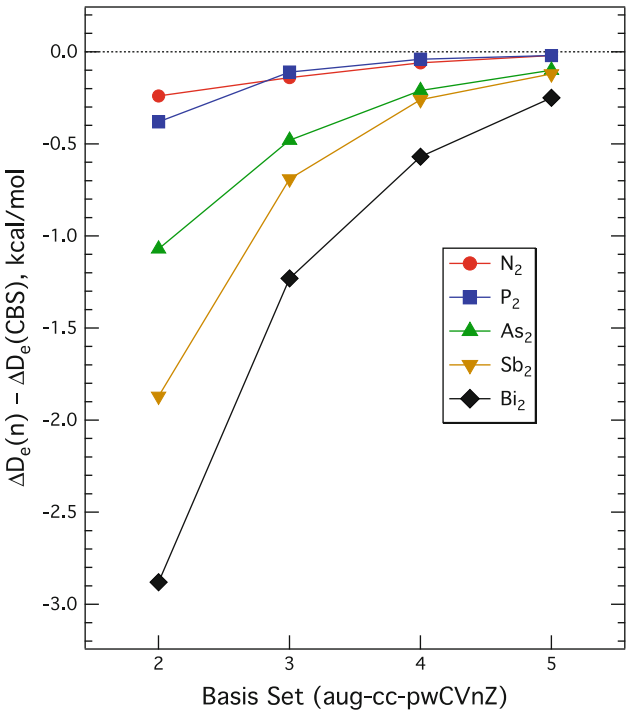

Fig. 7 Convergence toward the CBS limit of the CCSD(T) core correlation contributions to the dissociation energies of N_2 through Bi_2 as a function of basis set (aug-cc-pwCV*n*Z or aug-cc-pwCV*n*Z-PP)

considerably faster than the cc-pCVnZ basis sets, so, if the latter are used, the comparison may not be nearly as dramatic [41]. Except for the Bi_2 case, even basis sets of aug-cc-pwCVTZ(-PP) level are sufficient for chemical accuracy. It is obvious, however, that accurate recovery of core correlation for molecules containing post-d elements is much more challenging than in the case of molecules containing just first- and second-row elements. The core correlation effects are much larger as well; at the CBS limit, the core correlation contributions to D_e for these 5 diatomics are 0.80 (N_2), 0.77 (P_2), 2.39 (As_2), 3.16 (Sb_2) and 4.34 (Bi_2) kcal/mol. It should be noted that for these cases, the "core d" orbitals are not active in the valence space and that compounds such as AsF_3, SbF_3 and BiF_3 may require these d orbitals to be treated in the valence space since the F 2s orbitals are of comparable energy to the "core d" orbitals. Of course the core correlation contribution also grows with the molecular size, e.g., the CCSD(T)/CBS limit values for the core correlation contributions to the atomization energies of C_2H_6, C_3H_6 and C_4H_6 have been reported to be (in kcal/mol) 2.42, 3.72, and 4.78, respectively [62]. For a system as large as C_8H_{18} (n-octane), it is has been calculated to be at least 8.60 kcal/mol [63].

Given the smooth convergence with basis set for the core/valence effects shown in Fig. 7, it is not surprising that extrapolation is also effective in this case as well. Using Eq. 1 with $n = 3$ and 4 with the aug-cc-pwCVnZ series of basis sets (TZ and QZ, respectively) yields results nearly identical to the CBS limits obtained with $n = 4$ and 5, except for Bi_2, where the [TQ] extrapolation underestimated the more accurate [Q5] CBS limit by 0.18 kcal/mol. It should be noted that a CCSD(T) [TQ] extrapolation of aug-cc-pwCVnZ basis sets is the same level of theory used in the W4 method [33] to accurately account for core correlation effects. For the series of molecules of Fig. 7, even a [DT] extrapolation is reasonably effective, yielding a factor of 2 or better improvement compared to the raw aug-cc-pwCVTZ results.

Given the dramatic improvement in the basis set convergence rate when explicitly correlated F12 methods are used, it might be expected that similar behavior will occur for core correlation effects. This is indeed the case and CCSD(T)-F12b results analogous to those shown for conventional CCSD(T) in Fig. 7 for N_2 through Bi_2 are shown in Fig. 8. These results used the cc-pCVnZ-F12 basis sets of Hill et al. [91]. for N_2 and P_2 and the newly developed cc-pVnZ-F12-PP sets of Peterson et al. [92] for the heavier elements. In the F12 cases, convergence to within about 0.2–0.5 kcal/mol is observed already at the DZ level, while at least 0.2 kcal/mol accuracy is obtained in all cases at just the TZ level. Some small non-monotonic behavior for Bi_2 is observed, but it is only on the order of 0.2 kcal/mol. The

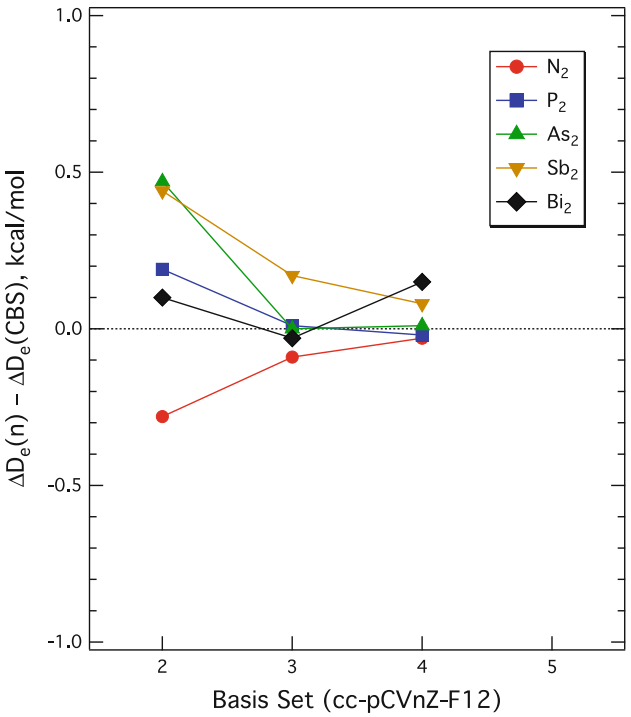

Fig. 8 Convergence toward the CBS limit of the core correlation contributions to the dissociation energies of N_2 through Bi_2 calculated with CCSD(T)-F12b as a function of basis set (cc-pCVnZ-F12 or cc-pVnZ-F12-PP)

latter can probably be attributed to the cc-pVnZ-F12-PP basis sets, which were only minimally augmented for outer-core correlation. In any event, it is obvious that F12 methods have the potential of providing a very efficient means to accurately account for core correlation effects on atomization energies.

3.3 Accounting for relativistic effects

3.3.1 Scalar relativity

For molecules containing elements from H through Ar, the effects of scalar relativity on atomization energies is generally small, on the order of a few to just several tenths of a kcal/mol for small molecules, and this will increase with molecular size. There are exceptions, of course, such as the sulfur oxides SO_2 and SO_3, where the scalar relativistic contributions have been calculated to be −0.78 and −1.76 kcal/mol, respectively [62], as well as a contribution of −2.3 kcal/mol for C_8H_{18} [63]. The first-row transition metals already exhibit strong effects due to scalar relativity, for example the contribution to the dissociation energy of CuH amounts to 2.5 kcal/mol [93]. For molecules containing post-3d elements, the effect can also be very large. For example, the scalar relativistic effect on the atomization energy of $GaCl_3$ has been calculated to be 6 kcal/mol

[94]. Most composite methodologies have focused on three techniques for recovery of scalar relativistic effects: the 1-electron mass velocity and Darwin operators from the Breit-Pauli Hamiltonian in first-order perturbation theory (MVD) [57, 95], the 2nd-order (or sometimes 3rd order) Douglas-Kroll-Hess (DKH) Hamiltonian [71, 96], and in a few cases stationary direct perturbation theory (DPT) [97, 98]. The MVD approach is fairly accurate for light elements but can lead to small errors compared to more accurate treatments even for molecules containing second-row elements like SO_2 [99]. In addition, the MVD approach is much more sensitive to the underlying basis set than other methods, and generally requires uncontracting at least the *s* functions of the basis set for the most accurate results [100, 101]. The workhouse for ab initio thermochemistry and spectroscopy, however, is the 1-electron DKH method, typically truncated at 2nd-order (DKH2). After recontracting the underlying basis set for DKH, e.g., cc-pVnZ-DK [73], the scalar relativistic correction is simply obtained as the difference in two energy evaluations, one with and one without the DKH hamiltonian, e.g., DK-CCSD(T)/cc-pVTZ-DK–CCSD(T)/cc-pVTZ. For relative energies, DKH3 does not seem to be required until one encounters elements as heavy as Z = 54 (Xe) [102]. The basis set dependence of the scalar relativistic contributions evaluated with DKH2 is not overly strong. Convergence to within ~0.1 kcal/mol is often reached at the TZ level and is converged to within a few hundredths of a kcal/mol with aug-cc-pVQZ-DK [99].

For molecules containing heavy elements, scalar relativistic effects should be included from the beginning (see Sec. III A and B above), either in all-electron calculations with the DKH Hamiltonian or using accurate relativistic effective core potentials (ECPs) or pseudopotentials (PPs). The family of correlation consistent basis sets has been extended to much of the periodic table in conjunction with newly optimized Stuttgart-style PPs, e.g., cc-pV*n*Z-PP [79, 103–106]. When the latter approach is taken, scalar relativistic effects are accurately taken into account for the heavy elements, but scalar relativity on any remaining light elements is not accounted for in addition to the small errors due to using the pseudopotential approximation. To this end, a DKH correction has been used, generally using TZ-level correlation consistent basis sets, i.e., ΔDK = DK-CCSD(T)/cc-pVTZ-DK–CCSD(T)/cc-pVTZ-PP, where in the 2nd term standard cc-pVTZ basis sets are used on any light elements that do not employ PPs. In a series of diatomic molecules containing 5*d* metal atoms, this "PP correction" ranged from 4.9 kcal/mol for HfO to just −0.07 kcal/mol for IrN [107]. In the case of the halogen oxides BrO and IO, this contribution amounted to at most 0.17 kcal/mol [108]. It should be noted at this point that the PPs used in the cc-pV*n*Z-PP basis sets have been adjusted to data that contain contributions not included at the DKH level of theory, namely the Breit interaction, which can become important for heavy elements, e.g., 5*d* and post-5*d* elements. The above DKH correction effectively removes these contributions.

3.3.2 Spin–orbit effects

For molecules involving only light elements, spin–orbit (SO) effects on atomization energies are generally limited to the spin–orbit fine structure of the atoms, which can be easily extracted from experiment [109] for lighter atoms by simply *J*-averaging over the levels of the ground-state term:

$$\Delta E_{SO} = -\frac{\sum_J (2J+1)E_J}{\sum_J (2J+1)} \quad (2)$$

For heavier atoms, it may not be possible to use this simple expression due to mixing of different spin/spatial states with the ground state. For diatomics or linear molecules with non-zero spin and orbital angular momentum, e.g., the $X^2\Pi$ state of NO, the spin–orbit contribution is given as one-half the spin–orbit constant obtained from experiment. In either case, these quantities can also be accurately calculated at the configuration interaction (CI) level using 1st-order perturbation theory and the Breit-Pauli Hamiltonian [110]. For heavier elements, even starting with the post-3*d* elements, 2nd-order spin–orbit molecular effects begin to have non-negligible contributions. For example, molecular SO effects on the closed-shell BrF and Br_2 molecules contribute 0.3 and 0.4 kcal/mol, respectively, to their dissociation energies [61]. In the case of I_2, this effect increases to 1.6–2.0 kcal/mol [61, 102]. These contributions can be recovered using a state-interacting approach using either the all-electron Breit-Pauli Hamiltonian or spin–orbit potentials from relativistic effective core potentials (ECPs), but the large number of electronic states that need to be included to accurately recover these 2nd-order effects can be cumbersome [111]. In the case of small molecules, this can be accurately overcome by carrying out a double-group relativistic CI calculation with ECPs [112]. Perhaps the most promising avenue is the use of 2-component Kramers-restricted CCSD using SO ECPs. This approach has seen only very limited use, however, in composite schemes due in part to its lack of availability in commonly used software [113].

3.4 Electron correlation beyond CCSD(T)

For molecules whose wavefunctions are dominated by the HF determinant, the contributions discussed to this point are usually sufficient to achieve chemical accuracy for an atomization energy at the equilibrium geometry. For cases

with multireference character or when more accurate results (<1 kcal/mol) are required, extension of the correlation treatment beyond CCSD(T) is required. Estimating the remaining contributions between full CI (FCI) and CCSD(T) is generally carried out by performing calculations with higher-order coupled cluster methods (CCSDT, CCSDT(Q), CCSDTQ, etc.). In the majority of cases, these higher-order correlation corrections are carried out at the frozen-core level of theory with relatively small basis sets since the computational expense scales very strongly with both the number of correlated electrons and basis functions. Studies have also shown that smaller basis sets are justified with higher-order correlated methods because the corrections due to each step up in the excitation level converges much more rapidly in terms of the basis set size than the previous excitation level [64]. It also has been generally observed in the literature that CCSD(T) atomization energies (and often bond lengths) are fortuitously closer to the FCI limit than the CCSDT values. Hence, any coupled-cluster-based correction for higher-order correlation should also include some estimate of connected quadruple excitations. To make matters somewhat more complicated, in the majority of cases the CCSDT correction, defined as CCSDT–CCSD(T), and the quadruples correction, e.g., CCSDTQ–CCSDT, enter with opposite signs and often with comparable magnitudes. So extra care must be taken to balance the convergence with respect to basis set of the two contributions, e.g., a CBS extrapolation of the CCSDT contribution combined with a CCSDTQ or CCSDT(Q) correction using just a cc-pVDZ basis set is not generally recommended [62]. A balanced approach might involve CBS extrapolation of both components, e.g., using TZ/QZ for CCSDT and DZ/TZ for CCSDTQ or CCSDT(Q), or by simply utilizing cc-pV(n + 1)Z for the CCSDT correction and cc-pVnZ for the quadruples since CCSDT has a stronger observed basis set dependence than CCSDTQ [33, 36, 114].

Going beyond corrections for connected quadruple excitations is often necessary for high accuracy. However, extending the procedure to pentuples, i.e., CCSDTQ5–CCSDTQ, is applicable to only the smallest of molecules. Another avenue that has proven fruitful is to carry out a continued fraction (cf) extrapolation using the sequence CCSD, CCSDT and CCSDTQ [62]. Although the results from this procedure can be a bit erratic, we reported that in 42 out of 49 comparisons with FCI, it improved atomization energies based on CCSDTQ [62]. In cases where large cf corrections are predicted, the results should be viewed with caution.

3.5 Miscellaneous contributions

For molecules composed of light elements, especially hydrogen, the commonly employed Born–Oppenheimer approximation, which separates electronic and nuclear motions, introduces a potentially significant error into the calculation of molecular properties when judged by the goals of high-accuracy studies. The diagonal correction to the BO approximation (DBOC) has been shown to contribute up to a few tenths of a kcal/mol to the atomization energies of hydrogen-containing molecules [115]. Calculating this contribution at the HF level of theory can lead to overestimations compared to more accurate CCSD calculations with cc-pCVTZ basis sets with outer-core correlation [116]. Even the use of cc-pCVDZ basis sets in the latter case yields nearly converged results. Including DBOC corrections for molecules that do not include hydrogen does not seem to be warranted at this time for the accuracies discussed above.

Related to the scalar relativistic correction discussed above is the leading quantum electrodynamic (QED) effect, the Lamb shift. While generally neglected in most composite methodologies, it has been shown previously by Dyall et al. [117]. that the contribution of the Lamb shift to the atomization energy can approach 3–5% of the scalar relativistic contribution. In cases where the scalar relativistic correction is large, for example when the valence s orbital occupation strongly changes, the Lamb shift can contribute $\sim$0.1 to 0.2 kcal/mol to the atomization energy for even relatively light systems like AlF_3 and GaF_3. This latter work calculated the contributions using perturbation theory. In the work of Shepler et al. [102]. involving small molecules containing iodine and mercury, the Lamb shift was calculated using the model potential approach of Pyykkö and Zhao [118] in DKH calculations. For the various mercury-containing species reported in Ref. [102], the Lamb shift contributed 0.3–0.7 kcal/mol to the atomization energies. Molecules such as I_2 or IBr exhibited contributions of only about 0.1 kcal/mol.

3.6 Zero-point vibrational corrections

The situation with zero-point vibrational energy (ZPVE) contributions in many ways parallels the situation with the atomization energies themselves. The largest contribution comes from estimating the harmonic value at something approaching the CCSD(T)(FC)/CBS limit, followed by a number of corrections of differing signs that may largely cancel but nevertheless need to be included for very high-accuracy studies. Anharmonic contributions to the ZPVE seem to rarely exceed 0.5 kcal/mol for small molecules [62, 119], but can increase strongly with the size of the molecule and number of bonds to hydrogen. Several of the model chemistries outlined in Sect. 2 utilize scaled harmonic frequencies with scale factors empirically chosen for a given method and basis set combination. Others use quartic force fields together with 2nd-order vibrational

perturbation theory. Whereas both the HEAT and Wn approaches calculate these at the CCSD(T) level of theory, the FPD approach currently exploits the fact that the anharmonic contribution can be accurately determined by lower levels of theory, so accurate CCSD(T) harmonic frequencies (at least with cc-pVQZ) are combined with anharmonic effects calculated at the MP2/aug-cc-pVDZ level of theory. It is recommended that core correlation contributions be neglected in the calculation of the CCSD(T) harmonic frequencies for ZPVEs if higher-order correlation effects are neglected (see below). Otherwise, the resulting ZPVEs tend to be too large. Certainly the error associated with the computed ZPVE is often overlooked in many composite thermochemistry methodologies, and its error can easily grow to the 1 kcal/mol range for small molecules and can be much larger for large molecules, especially if they contain many hydrogen atoms [62, 120]. For example, the ZVPE is estimated to be 151.8 kcal/mol for C_8H_{18} of which the anharmonic correction could be on the order of at least 10 kcal/mol [63].

4 Additional considerations for ab initio spectroscopy

In general, the application of an accurate composite methodology originally designed for thermochemistry applications needs little modification for use in the accurate prediction of spectroscopic properties such as rotational constants and vibrational band origins. Compared to calculating an atomization energy, the biggest challenge is simply one of dimensionality—all of the contributions from Sects. 3.1–3.6 must now be applied (in principle) on many near-equilibrium geometries of the 3N-6 dimensional potential energy surface (PES) instead of just at the minimum. For diatomic and even triatomic molecules it is relatively straightforward to explicitly sample the near-equilibrium part of the PES, to obtain an analytical representation and to evaluate spectroscopic constants from the derivatives of the PES using perturbation theory or variational methods. Each point sampled on the PES is constructed from the several contributions of the composite approach just as in thermochemical applications. From such small molecule benchmarks, it is clear that the dominant contribution to the overall accuracy is still the estimation of the frozen-core CCSD(T)/CBS values. For molecules containing only 1st- and 2nd-row elements, the resulting FC-CCSD(T)/CBS equilibrium bond lengths exhibit typical average errors with respect to experiment of about +0.003 Å (too long) and harmonic frequencies too large by an average of about 6 cm^{-1} [37, 98, 121]. The addition of outer-core correlation reduces the error in the equilibrium structures by about a factor of two but leaves the distances uniformly too short on average. The errors in the harmonic frequencies exhibit a further increase, tending to be even further away from experiment. Both of these deficiencies are unfortunately not ameliorated until the contributions of connected quadruples are included into the methodology (the inclusion of full iterative triples also tends to worsen the agreement with experiment compared to CCSD(T) alone). So, it is clear that the success of the FC-CCSD(T)/CBS level of theory is to some extent due to a fortuitous cancellation of errors. For equilibrium structures, it is clear that correlation of the outer-core electrons is desirable, particularly as shown recently for post-*d* elements where *d*-electron correlation can lead to substantial bond shortening [90]. On the other hand for harmonic frequencies, if core correlation contributions are included, then some treatment of relativistic and higher-order correlation effects must also be addressed.

For polyatomic molecules, a common strategy is to assume that the various contributions of the composite methodology are additive for the structural parameters and harmonic frequencies [62]. The CCSD(T)(FC)/CBS limits can be estimated by either large basis set conventional calculations, explicitly correlated CCSD(T), or, in the case of bond lengths, by direct CBS extrapolation [62, 122]. Of course it is also possible to incorporate the CBS extrapolation scheme into a numerical or analytical derivative calculation [121, 123]. Overall, the conclusions reached by the small molecule studies have been shown to be applicable to larger polyatomics as well [62]. In particular for harmonic frequencies, however, the importance of quadruple excitations makes the accurate calculation (better than 10–15 cm^{-1}) of harmonic frequencies (and subsequent prediction of anharmonic band origins) very computationally challenging for polyatomic molecules [123].

Of course neither equilibrium geometries nor harmonic frequencies are directly comparable to experiment; hence, some calculation of the contributions of vibrational anharmonicity must be considered in order to be truly predictive. Fortunately, the anharmonic part of the PES is not nearly as sensitive to the electron correlation method or basis set as compared to the harmonic portion. For polyatomic molecules, this can be exploited by a multilayer approach to the calculation of vibrational frequencies, whereby the harmonic frequencies are calculated as accurately as possible, perhaps also with the diagonal anharmonicities, and then the remaining anharmonic contributions are reliably calculated at some lower level of theory, e.g., MP2 or DFT [62, 124, 125]. Generally, the anharmonic contributions are calculated from quartic force fields and 2nd-order vibrational perturbation theory [126, 127] or by explicit sampling of the PES with multimode normal coordinate expansions and subsequent vibrational SCF or vibrational CI calculations (see, e.g., Ref [128]). For the prediction of ground-state rotational constants, vibrationally averaged structures

can be obtained from these same calculations from the cubic force field, which defines the vibration–rotation interaction constants [129]. These latter ab initio values can also be combined with experimental ground-state rotational constants to produce accurate semi-experimental equilibrium geometries [76, 130–132], which are more directly comparable to bottom of the well values obtained from theory [133].

5 Future challenges

For molecules consisting of relatively light elements from the first few rows of the periodic table, sophisticated composite ab initio methodologies have demonstrated a potential for very high accuracy in both thermochemistry and spectroscopy. Given the important role of the FC-CCSD(T)/CBS limit in these schemes, further development and implementation of explicitly correlated methods, e.g., CCSD(T)-F12x, will bring immediate dividends to these methods, especially in terms of the size of molecules that then can be accurately treated. Particularly for the accurate calculation of vibrational frequencies beyond the FC-CCSD(T) level of theory, more efficient methods for recovering higher-order correlation corrections are essential in order to carry out more reliable calculations on polyatomic molecules. Finally, the biggest challenge is the extension of these ideas to molecules containing transition metals and other heavy elements. For molecules that are dominated by the HF determinant, the contributions outlined above in Sec. IIIA-E are perhaps sufficient [47, 134–137], but, in particular for the transition metals, the existence of multireference character is common, which puts the current composite schemes that are based on the single-reference CCSD(T) method in jeopardy. Recently attempts have been made to include multireference components, for example the MR-ccCA variant [138], which includes CASPT2 and MRCI contributions. It remains to be seen, however, whether this approach will provide sufficient accuracy or can be made sufficiently general for larger systems. Also as accurate composite calculations are pushed to heavier systems, relativistic effects begin to be extremely important. For the most part, scalar relativistic effects appear to be straightforward to include, but accurate recovery of spin–orbit effects remains a challenge. This, in part, is due to difficulties in calculating the correct energies even for the atoms, which can have very complicated spin–orbit fine structure [109].

6 Summary

Current strategies for the calculation of molecular thermochemical and spectroscopic properties have been reviewed with particular attention to composite schemes designed to obtain chemical accuracy and beyond. The most important consideration in these methodologies is the accurate estimation of the frozen-core complete basis set limit, either at the equilibrium geometry in the case of thermochemistry applications or as a function of the 3N-6 internal coordinates for the calculation of spectroscopic properties. Practical solutions to this problem include effective basis set extrapolation strategies using sequences of correlation consistent basis sets or newly developed F12 explicitly correlated methods. The effects of correlating the outer-core electrons are also discussed. These latter correlation corrections are essential for accurate thermochemistry, e.g., atomization energies or heats of formation, as well as decreasing the errors in calculated equilibrium bond lengths. In the case of vibrational frequencies determined at the CCSD(T) level, correlating the outer-core electrons tends to increase the errors with respect to experiment and should generally be accompanied by higher-order electron correlation contributions in order to improve upon the accuracy already attained with just the valence electrons correlated. Relativistic effects can also make important contributions to both thermochemical and spectroscopic properties, even in molecules containing only elements of the first few rows of the periodic table, and scalar relativistic effects need to be included from the outset in heavier systems. The treatment of chemical systems with significant multireference character, as well as the efficient and accurate recovery of spin–orbit coupling effects for molecules containing heavy elements, remains challenge for the area of chemically accurate thermochemistry.

Acknowledgments K. A. P. would like to acknowledge the funding of the National Science Foundation (CHE-0723997). D. A. D. was supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department of Energy (DOE) under the catalysis center program. D. A. D. thanks the Robert Ramsay Fund of the University of Alabama and Argonne National Laboratory for partial support.

References

1. Dunning TH Jr (1989) J Chem Phys 90:1007
2. Raghavachari K, Trucks GW, Pople JA, Head-Gordon M (1989) Chem Phys Lett 157:479
3. Feller D, Peterson KA, Hill JG (2011) J Chem Phys 135:044102
4. Pople JA, Head-Gordon M, Fox DJ, Raghavachari K, Curtiss LA (1989) J Chem Phys 90:5622
5. Frisch MJ, Trucks GW, Head-Gordon M, Gill PMW, Wong MW, Foresman JB et al (1992) Gaussian 92. Gaussian, Inc., Pittsburg
6. Curtiss LA, Jones C, Trucks GW, Raghavachari K, Pople JA (1990) J Chem Phys 93:2537
7. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) Ab initio molecular orbital theory. Wiley-Interscience, New York
8. Curtiss LA, Raghavachari K, Trucks GW, Pople JA (1991) J Chem Phys 94:7221

9. Curtiss LA, Raghavachari K, Pople JA (1993) J Chem Phys 98:1293
10. Bauschlicher CW Jr, Partridge H (1995) J Chem Phys 103:1788
11. Smith BJ, Radon L (1995) J Phys Chem 99:6468
12. Curtiss LA, Redfern PC, Smith BJ, Radom L (1996) J Chem Phys 104:5148
13. Curtiss LA, Raghavachari K, Redfern PC, Pople JA (1998) J Chem Phys 109:7764
14. Curtiss LA, Redfern PC, Raghavachari K, Rassolov V, Pople JA (1999) J Chem Phys 110:4703
15. Baboul AG, Curtiss LA, Redfern PC, Raghavachari K (1999) J Chem Phys 110:7650
16. Curtiss LA, Raghavachari K, Redfern PC, Baboul AG, Pople JA (1999) Chem Phys Lett 314:101
17. Curtiss LA, Redfern PC, Raghavachari K, Pople JA (2001) J Chem Phys 114:108
18. Henry DJ, Sullivan MB, Radom L (2003) J Chem Phys 118:4849
19. Curtiss LA, Redfern PC, Raghavachari K (2007) J Chem Phys 126:084108
20. Curtiss LA, Redfern PC, Raghavachari K (2007) J Chem Phys 127:124105
21. Nyden MR, Petersson GA (1981) J Chem Phys 75:1843
22. Petersson GA, Bennett A, Tensfeldt TG, Al-Laham MA, Shirley WA, Mantzaris J (1988) J Chem Phys 89:2193
23. Petersson GA, Al-Laham MA (1991) J Chem Phys 94:6081
24. Petersson GA, Tensfeldt TG, Montgomery JA Jr (1991) J Chem Phys 94:6091
25. Montgomery JA Jr, Ochterski JW, Petersson GA (1994) J Chem Phys 101:5900
26. Ochterski JW, Petersson GA, Montgomery JA Jr (1996) J Chem Phys 104:2598
27. Montgomery JA Jr, Frisch MJ, Ochterski JW, Petersson GA (1999) J Chem Phys 110:2822
28. Wood GPF, Radom L, Petersson GA, Barnes EC, Frisch MJ, Montgomery JA Jr (2006) J Chem Phys 125:094106
29. Mayer PM, Parkinson CJ, Smith DM, Radom L (1998) J Chem Phys 108:604
30. Martin JML, de Oliveira G (1999) J Chem Phys 111:1843
31. Boese AD, Oren M, Atasoylu O, Martin JML, Kallay M, Gauss J (2004) J Chem Phys 120:4129
32. Barnes EC, Petersson GA, Montgomery JJA, Frisch MJ, Martin JML (2009) J Chem Theory Comput 5:2687
33. Karton A, Rabinovich E, Martin JML, Ruscic B (2006) J Chem Phys 125:144108
34. Mintz B, Chan B, Sullivan MB, Buesgen T, Scott AP, Kass SR, Radom L, Wilson AK (2009) J Phys Chem A 113:9501
35. Ruscic B (unpublished) Results obtained from the active thermochemical tables, ATcT ver. 1.25 and the Core (Argonne) thermochemical network ver. 1.056 2006
36. Karton A, Taylor PR, Martin JML (2007) J Chem Phys 127:064104
37. Karton A, Martin JML (2010) J Chem Phys 133:144102
38. Tajti A, Szalay PG, Császár AG, Kállay M, Gauss J, Valeev EF, Flowers BA, Vázquez J, Stanton JF (2004) J Chem Phys 121:11599
39. Helgaker T, Klopper W, Koch H, Noga J (1997) J Chem Phys 106:9639
40. Bomble YJ, Vázquez J, Kállay M, Michauk C, Szalay PG, Császár AG, Gauss J, Stanton JF (2006) J Chem Phys 125:064108
41. Harding ME, Vazquez J, Ruscic B, Wilson AK, Gauss J, Stanton JF (2008) J Chem Phys 128:114111
42. DeYonker NJ, Cundari TR, Wilson AK (2006) J Chem Phys 124:114104
43. Feller D (1992) J Chem Phys 96:6104
44. Peterson KA, Woon DE, Dunning TH Jr (1994) J Chem Phys 100:7410
45. Kedziora GS, Pople JA, Ratner MA, Redfern PC, Curtiss LA (2001) J Chem Phys 115:718
46. DeYonker NJ, Grimes T, Yokel S, Dinescu A, Mintz B, Cundari TR, Wilson AK (2006) J Chem Phys 125:104111
47. DeYonker NJ, Peterson KA, Steyl G, Wilson AK, Cundari TR (2007) J Phys Chem A 111:11269
48. DeYonker NJ, Williams TG, Imel AE, Cundari TR, Wilson AK (2009) J Chem Phys 131:024106
49. Mintz B, Williams TG, Howard L, Wilson AK (2009) J Chem Phys 130:234104
50. Prascher BP, Lai JD, Wilson AK (2009) J Chem Phys 131:044130
51. Schütz M, Chem J (2000) Phys 113:9986
52. Schütz M, Werner H-J (2000) Chem Phys Lett 318:370
53. DeYonker NJ, Wilson BR, Pierpont AW, Cundari TR, Wilson AK (2009) Mol Phys 107:1107
54. East ALL, Allen WD, Császár AG (1993) In: Laane J, van der Veken B, Overhammer H (eds) Structure and conformations of non-rigid molecules. Kluwer, Dordrecht, p 343
55. East ALL, Allen WD (1993) J Chem Phys 99:4638
56. Császár AG, Allen WD, Schaefer HF, II I (1998) J Chem Phys 108:9751
57. Cowan RD, Griffin M (1976) J Opt Soc Am 66:1010
58. Handy NC, Yamaguchi Y, Schaefer HF (1986) J Chem Phys 84:4481
59. Schuurman MS, Muir SR, Allen WD, Schaefer HF, II I (2004) J Chem Phys 120:11586
60. East ALL, Johnson CS, Allen WD (1993) J Chem Phys 98:1299
61. Feller D, Peterson KA, de Jong WA, Dixon DA (2003) J Chem Phys 118:3510
62. Feller D, Peterson KA, Dixon DA (2008) J Chem Phys 129:204105
63. Pollack L, Windus TL, de Jong WA, Dixon DA (2005) J Phys Chem A 109:6934
64. Feller D, Peterson KA, Crawford TD (2006) J Chem Phys 124:054107
65. Barnes EC, Petersson GA, Feller D, Peterson KA (2008) J Chem Phys 129:194115
66. Feller D, Peterson KA, Dixon DA (2010) J Phys Chem A 114:61
67. Feller D, Peterson KA, Dixon DA (2011) J Phys Chem A 115:1440
68. Feller D, Peterson KA, Hill JG (2010) J Chem Phys 133:184102
69. Feller D, Peterson KA (2007) J Chem Phys 126:114105
70. Peterson KA, Dunning TH Jr (2002) J Chem Phys 117:10548
71. Douglas M, Kroll NM (1974) Ann Phys NY 82:89
72. Jansen G, Hess BA (1989) Phys Rev A 39:6016
73. de Jong WA, Harrison RJ, Dixon DA (2001) J Chem Phys 114:48
74. Dunham JL (1932) Phys Rev 41:713
75. Kutzelnigg W (1997) Mol Phys 90:909
76. Feller D, Craig NC, Matlin AR (2008) J Phys Chem A 112:2131
77. Kendall RA, Dunning Jr TH, Harrison RJ (1992) J Chem Phys 96:6796
78. Woon DE, Dunning TH Jr (1993) J Chem Phys 98:1358
79. Peterson KA (2003) J Chem Phys 119:11099
80. Feller D (1996) J Comp Chem 17:1571
81. Werner H-J, Knizia G, Adler TB, Marchetti O (2010) Z Phys Chem 224:493
82. Klopper W, Manby FR, Ten-no S, Valeev EF (2006) Int Rev Phys Chem 25:427
83. Knizia G, Adler TB, Werner H-J (2009) J Chem Phys 130:054104
84. Peterson KA, Adler TB, Werner H-J (2008) J Chem Phys 128:084102

85. Knizia G, Werner H-J (2008) J Chem Phys 128:154103
86. Karton A, Martin JML (2006) Theor Chem Acc 115:330
87. Schwenke DW (2005) J Chem Phys 122:014107
88. Hill JG, Peterson KA, Knizia G, Werner H-J (2009) J Chem Phys 131:194105
89. Woon DE, Dunning TH Jr (1995) J Chem Phys 103:4572
90. Peterson KA, Yousaf KE (2010) J Chem Phys 133:174116
91. Hill JG, Mazumder S, Peterson KA (2010) J Chem Phys 132:054108
92. Peterson KA, Hill JG (2011) (in preparation)
93. Balabanov NB, Peterson KA (2005) J Chem Phys 123:064107
94. Bauschlicher Jr CW (1998) J Phys Chem A 102:10424
95. Martin RL (1983) J Phys Chem 87:750
96. Hess BA (1986) Phys Rev A 33:3742
97. Kutzelnigg W, Ottschofski E, Franke R (1995) J Chem Phys 102:1740
98. Tew DP, Klopper W, Heckert M, Gauss J (2007) J Phys Chem A 111:11242
99. Boese AD, Oren M, Atasoylu O, Martin JML, Kallay M, Gauss J (2004) J Chem Phys 120:4129
100. Bauschlicher CW Jr, Martin JML, Taylor PR (1999) J Phys Chem A 103:7715
101. Michauk C, Gauss J (2007) J Chem Phys 127:044106
102. Shepler BC, Balabanov NB, Peterson KA (2005) J Phys Chem A 109:10363
103. Figgen D, Peterson KA, Dolg M, Stoll H (2009) J Chem Phys 130:164108
104. Peterson KA, Figgen D, Dolg M, Stoll H (2007) J Chem Phys 126:124101
105. Peterson KA, Puzzarini C (2005) Theor Chem Acc 114:283
106. Peterson KA, Figgen D, Goll E, Stoll H, Dolg M (2003) J Chem Phys 119:11113
107. Spohn B, Goll E, Stoll H, Figgen D, Peterson KA (2009) J Phys Chem A 113:12478
108. Peterson KA, Shepler BC, Figgen D, Stoll H (2006) J Phys Chem A 110:13877
109. Moore CE, (1971) Vol Natl Stand Ref Data Ser, National Bureau of Standards (US), Washington 35
110. Hess BA, Marian CM, Peyerimhoff SD (1995) In: Yarkony DR (ed) Modern electronic structure theory, part I, vol 2. World Scientific, Singapore, p 152
111. Shepler BC, Peterson KA (2003) J Phys Chem A 107:1783
112. Yabushita S, Zhang Z, Pitzer RM (1999) J Phys Chem A 103:5791
113. Cho WK, Choi YJ, Lee YS (2005) Mol Phys 103:925
114. Ruden TA, Helgaker T, Jorgensen P, Olsen J (2003) Chem Phys Lett 371:62
115. Ruscic B, Wagner AF, Harding LB, Asher RL, Feller D, Dixon DA, Peterson KA, Song Y, Qian XM, Ng CY, Liu JB, Chen WW (2002) J Phys Chem A 106:2727
116. Gauss J, Tajti A, Kallay M, Stanton JF, Szalay PG (2006) J Chem Phys 125:4111
117. Dyall KG, Bauschlicher Jr CW, Schwenke DW, Pyykkö P (2001) Chem Phys Lett 348:497
118. Pyykkö P, Zhao U (2003) J Phys B 36:1469
119. Barone V (2004) J Chem Phys 120:3059
120. Karton A, Ruscic B, Martin JML (2007) J Mol Struct 811:345
121. Heckert M, Kallay M, Tew DP, Klopper W, Gauss J (2006) J Chem Phys 125:4108
122. Puzzarini C (2009) J Phys Chem A 113:14530
123. Peterson KA, Francisco JS (2011) J Chem Phys 134:084308
124. Rauhut G, Knizia G, Werner H-J (2009) J Chem Phys 130:054105
125. Puzzarini C, Barone V (2008) Phys Chem Chem Phys 2008:6991
126. Barone V (2005) J Chem Phys 122:014108
127. Clabo DA Jr, Schaefer HF III (1987) Int J Quantum Chem 31:429
128. Rauhut G (2004) J Chem Phys 121:9313
129. Puzzarini C, Heckert M, Gauss J (2008) J Chem Phys 128:194108
130. Craig NC, Feller D, Groner P, Hsin HY, McKean DC, Nemchick DJ (2007) J Phys Chem A 111:2498
131. Feller D, Craig NC (2009) J Phys Chem A 113:1601
132. Feller D, Craig NC, Groner P, McKean DC (2011) J Phys Chem A 115:94
133. Puzzarini C, Cazzoli G, Gambi A, Gauss J (2006) J Chem Phys 125:054307
134. Craciun R, Picone D, Long RT, Li S, Dixon DA, Peterson KA, Christe KO (2010) Inorg Chem 49:1056
135. Li S, Hennigan JM, Dixon DA, Peterson KA (2009) J Phys Chem A 113:7861
136. Puzzarini C, Peterson KA (2005) Chem Phys 311:177
137. Pahl E, Figgen D, Borschevsky A, Peterson KA, Schwerdtfeger P (2011) Theor Chem Acc 129:651
138. Jiang W, Wilson AK (2011) J Chem Phys 134:034101