ANNUAL REVIEWS

Annual Review of Physical Chemistry Straightening the Hierarchical Staircase for Basis Set Extrapolations: A Low-Cost Approach to High-Accuracy Computational Chemistry

António J.C. Varandas

Coimbra Chemistry Center and Department of Chemistry, University of Coimbra, Coimbra 3004-535, Portugal; email: varandas@uc.pt

Annu. Rev. Phys. Chem. 2018. 69:177-203

First published as a Review in Advance on February 2, 2018

The Annual Review of Physical Chemistry is online at physchem.annualreviews.org

https://doi.org/10.1146/annurev-physchem-050317-021148

Copyright © 2018 by Annual Reviews. All rights reserved

ANNUAL Further REVIEWS Further Click here to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
 Explore related articles
- Search keywords

Keywords

electronic structure calculations, Hartree–Fock energy, correlation energy, basis set extrapolation, potential energy surfaces

Abstract

Because the one-electron basis set limit is difficult to reach in correlated post-Hartree-Fock ab initio calculations, the low-cost route of using methods that extrapolate to the estimated basis set limit attracts immediate interest. The situation is somewhat more satisfactory at the Hartree-Fock level because numerical calculation of the energy is often affordable at nearly converged basis set levels. Still, extrapolation schemes for the Hartree-Fock energy are addressed here, although the focus is on the more slowly convergent and computationally demanding correlation energy. Because they are frequently based on the gold-standard coupled-cluster theory with single, double, and perturbative triple excitations [CCSD(T)], correlated calculations are often affordable only with the smallest basis sets, and hence single-level extrapolations from one raw energy could attain maximum usefulness. This possibility is examined. Whenever possible, this review uses raw data from second-order Møller-Plesset perturbation theory, as well as CCSD, CCSD(T), and multireference configuration interaction methods. Inescapably, the emphasis is on work done by the author's research group. Certain issues in need of further research or review are pinpointed.

1. INTRODUCTION

The solution of the time-independent electronic Schrödinger equation (SE) has long been a challenge of the utmost importance in computational molecular science (1, 2). This is not surprising, since accurate energies and eigenfunctions are crucial information for spectroscopy, thermochemistry, and kinetics, to note only a few fields where it plays a key role.

The oldest and simplest method for solving the SE is the Hartree-Fock (HF) or mean-field approach, in which the correlated motion of the electrons is ignored. Because the error arising from this disregard is significant, more sophisticated single-reference (SR) methods have been proposed. These include variational methods such as configuration interaction (CI) and the perturbative second-order Møller–Plesset (MP2) and coupled-cluster (CC) methods, which are perhaps the most popular. Of relevance is coupled-cluster theory with single, double, and perturbative triple excitations [CCSD(T)], which is considered the gold standard by quantum chemists because of its accuracy and affordability even for moderate-sized systems: It often yields results closer to more correlated CC methods (such as CCSDTQ, which has explicit triple and quadruple excitations) and full configuration interaction (FCI) than the less approximate CCSDT method (3). Still, SR methods leave out the static (nondynamical) correlation. To recover it, multireference (MR) methods are needed, such as the one based on the complete active space self-consistent field (CASSCF or CAS) wave function, with MRCI accounting for the dynamical correlation by inclusion of single and double electronic excitations (MRCISD). Since integration of the SE with accuracy at demand is unreachable, extrapolation of the raw energies to target converged values is key and the focal point of this work.

The outcome of the pointwise solution of the SE for an N-atom species is the potential energy surface (PES) (4), a function defined in a configuration space of dimensionality $\tau = 3N - 6$. Scaling as n_u^4 per point (where n_u is the number of unoccupied orbitals), the cost of the PES easily becomes unaffordable when combined with that of generating the PES itself, adding up to $\propto n_u^4 \bar{m}^{t}$ (where \bar{m} is a typical number of calculations required per dimension). Fortunately, the focus is often on relative energies, with the important chemical changes occurring in the valence orbitals. Despite this, conventional post-HF methods have an agonizingly slow rate of convergence with basis set size. Moreover, basis set superposition error (BSSE), which is inextricably tangled with the inability to exactly solve the SE, plagues all ab initio calculations. Extrapolation to the oneelectron complete basis set (CBS) limit, or just scaling of the correlation energy (if not the total energy itself; see 5), can thus be valuable. In fact, besides allowing one to avoid correcting for BSSE, the utility of such methods is enhanced in combination with fragment-based methods, in which a large molecule is made tractable by explicitly considering all parts into which it can be fragmented as a many-body expansion (MBE) (4). Note that if it follows a first split of the SE eigenenergies into HF- and correlation-type contributions as in modern ab initio theory, this approach assumes the double many-body expansion (DMBE) form (6, 7). To this end, CBS extrapolation schemes are here surveyed. For convenience and following common practice in the literature, energies are given in hartree ($E_{\rm h}$), kcal mol⁻¹, or kJ mol⁻¹; 1 $E_{\rm h} = 627.5095$ kcal mol⁻¹ = 2,625.4996 kJ mol⁻¹.

2. GENERAL CONCEPTS

2.1. Explicit Correlation Versus Complete Basis Set Extrapolation

The need to go beyond HF arises from the correlated motions of the electrons dictated by the electron–electron repulsion operator. Because this shows a singularity when any two electrons coalesce ($r_{12} = 0$), the exact wave function has a cusp (discontinuous derivative), as implied by the famous Kato (8) cusp condition, $\langle \partial \Psi / \partial r_{12} \rangle_{r_{12}=0} = \frac{1}{2} \Psi(r_{12} = 0)$, which implies that the exact wave function must be linear in r_{12} at small interelectronic distances. It turns out that conventional

correlation methods expand the wave function in terms of Slater determinants built from orbitals, thus failing to obey this condition; this explains their extremely slow convergence. Although the purist's approach to solving the problem is to make the many-electron wave function depend explicitly on \mathbf{r}_{12} (2, 9), the alternative consists of systematizing the error made by conventional methods and seeking an extrapolation scheme that allows one to predict the inherent error, which is the approach highlighted here.

For two-electron atoms, where the first-order wave function of a 1/Z expansion in partial waves [Legendre polynomials of the cosine of the angle between the electronic vectors \mathbf{r}_1 and \mathbf{r}_2 , $P_l(\cos\theta)$ behaves, for small r_{12} , as $\psi = (1 + \kappa r_{12}) \Phi + \mathcal{O}(r_{12}^2)$, with $\kappa = \frac{1}{2}$ (or $\frac{1}{4}$) for singlet (or triplet) states (8, 10) and with Φ being the unperturbed wave function, Kutzelnigg & Morgan (11) established that: (a) For natural-parity singlet states [a state's parity is natural if it equals $(-1)^{L}$ and unnatural if it equals $(-1)^{(L+1)}$, the leading contribution to the energy at the second order of perturbation theory is proportional to $(\ell + \frac{1}{2})^{-4}$ with no odd terms proportional either to $(\ell + \frac{1}{2})^{-5}$ or $(\ell + \frac{1}{2})^{-7}$ and with the term $(\ell + \frac{1}{2})^{-6}$ universally $-\frac{5}{4}$ of $(\ell + \frac{1}{2})^{-4}$; (b) for triplet states, the leading term is proportional to $(\ell + \frac{1}{2})^{-6}$; (c) for unnatural-parity singlet states, the coefficient is proportional to $(\ell + \frac{1}{2})^{-8}$. They also found the results to remain essentially unaltered when Φ is the HF wave function, although the ratio (11) of the $(\ell + \frac{1}{2})^{-6}$ and $(\ell + \frac{1}{2})^{-4}$ coefficients for natural-parity singlet states is no longer $-\frac{5}{4}$. Moreover, the results were found to generalize to the MP2 energy of atoms with any number of electrons (11, 12). Such findings followed pioneering work using perturbation theory by Schwartz (13), who analyzed the partial-wave expansion of the second-order expression and obtained only the coefficients of powers -4 and -6, since the $(l+\frac{1}{2})^{-5}$ term vanishes. Thus, if $\Delta E_l \propto \sum_{m=4} a_m (l+\frac{1}{2})^{-m}$ is assumed to be the convergence error when all terms with $l \ge L$ are omitted, one has, for large enough L values,

$$\Delta E = \sum_{l=1}^{\infty} \Delta E_{l} \sim \sum_{m=4}^{\infty} a_{m} \int_{L+\frac{1}{2}}^{\infty} \left(l + \frac{1}{2}\right)^{-m} \mathrm{d}l = \sum_{m=4}^{\infty} A_{m-1} (L+1)^{-m+1}, \qquad 1.$$

where $A_3(L+1)^{-3}$ is the leading asymptotic term and $A_5(L+1)^{-5}$ the next one. Use of a one-term expression may then limit accuracy, since the subset of natural singlet pairs in an MP2 calculation for the Zn²⁺ ion contributes only 54.7% of the total correlation energy (12). Schwenke (14) dissected this issue through MP2 calculations on He and Ne as a function of the maximum angular momentum using basis sets with converged radial extent and values of *l* up to 14. While the form $(L + 1)^{-3}$ was reached for Ne, higher-order terms were normally important in the range of *l* values used in molecular calculations. Additionally, a similar analysis for an open-shell atom and a small molecule has shown that the actual form may gain even further intricacy (14).

Largely motivated by the above, modern basis sets (15–21) are commonly built according to a principal expansion (thus increasing by inclusion of all $\sim n^2$ functions in the next shell), with each shell contributing an amount of energy $\propto n^{-4}$. The popular correlation-consistent basis sets (cc-pVXZ or VXZ; aug-cc-pVXZ or AVXZ and d-aug-cc-pVXZ or dAVXZ when augmented with extra diffuse functions; cc-pCVXZ or CVXZ when functions for core-correlation are included) of the Dunning family (15, 16, 22) are an example. Because the cardinal number (X = 2 : D, 3 : T, 4 : Q, ...) can be identified with L + 1, the common X^{-3} power-law error (23, 24) arises. Its slow convergence compounds with the time required for MP2, CCSD, and CCSD(T) calculations, which scale as $N^5N_b^4$, $N^6N_b^4$, and $N^7N_b^4$ (25); N is the number of atoms and N_b the number of basis functions per atom, so that $N_b \simeq X^3$ for a VXZ basis (24, 26). Calculations beyond QZ are thus unaffordable for many interesting systems, making the output raw energies far from safe for extrapolation with many CBS schemes. This contrasts with the HF energy, which is found numerically and analytically to converge exponentially (27), possibly $\propto \exp(-\sqrt{L})$. There are then two ways to obtain accurate energies: (*a*) solution of the SE after explicit introduction of correlation in the wave function (2) via \mathbf{r}_{ij} coordinates and (*b*) exploitation of the convergence of hierarchized correlation-consistent basis sets toward the CBS limit. Although this may conceivably become a matter of merely historical interest (28), with standard CCSD(T) calculations giving way to their explicit correlated counterparts, any analysis of the current literature or direct comparison of standard versus explicitly correlated methods (generally denoted by R12) shows such a scenario to be far away, since R12 methods are relatively new, with even analytical derivatives being largely unavailable (29). Actually, despite a fast convergence ($\propto X^{-7}$), R12 methods perform inefficiently with small basis sets (2). Besides, both conventional and R12 methods converge to the same asymptotic energy, with CBS extrapolations doing well and often outperforming R12 methods in overcoming noncompleteness of the one-electron basis set, a merit reflected (3) in the number of CBS schemes vying with the more rigorous R12 techniques. Indeed, to facilitate the CBS extraction process, an extrapolation calculator has recently been developed (30) for some of the schemes discussed here.

2.2. Correlation Scaling Versus Complete Basis Set Extrapolation

The problem of generating an accurate ab initio PES has been examined by Varandas & Piecuch (31; see also 32), who suggested predicting it for a larger target basis set from results of smaller basis set calculations by scaling the correlation energy. To understand this, see **Figure 1**, which shows the relation between the basis set and intrinsic, adiabatic, and apparent error includes corrections due to relativistic and nonadiabatic effects. Varandas & Piecuch (31) have shown that by calculating the PES with bases at hierarchies (m - 1) and m and one point at hierarchy (m + 1), one should be able to predict the entire PES at the (m + 1) target level. For this, the required universal scaling factor should be (a) determined at the chosen (reference) nuclear geometry over a series of systematically constructed basis sets and (b) applied to the electron correlation energies of the remaining geometries defining the PES, which have been calculated with smaller basis sets in order to predict the PES corresponding to the larger target basis set. The idea



Figure 1

Basis set versus intrinsic versus adiabatic versus apparent errors in electronic structure theory (1). Hierarchical numbers are $m - 1, m, \ldots$; the experimental value of the calculated property is indicated and the Born–Oppenheimer plane is shown in gray.

in this correlation scaling (CS) approach (31) is then similar to that in other approaches offering large savings in computer effort: scaling of all (total) correlation (SAC) (33) in SR calculations and scaling of the external correlation (SEC) (34). It should be noted that if more than one channel exists, a compromise scaling parameter must be sought in SAC or SEC (33, 34), although such a requirement can be eliminated by first expanding the total energy as a DMBE (thus allowing for one scaling parameter per distinct *n*-body fragment, $n \leq N$ in a procedure known as DMBE-SEC (35) that has been much used in modeling global PESs (see 5 and references therein). The major difference in CS (31) then lies in not using (31, 32) any empirical data. The CS scheme has also been extended to an arbitrary number of reference geometries (32, 36) and used with success for small molecules (31, 32, 36) and even molecules as large as butadiene (37). Although extendable to large bases, CS is impractical for reaching the CBS limit. Yet their joint use can conceivably be useful and has actually proven reliable (36, 38). For example, there are regions of configuration space where the cheaper CCSD(T) method may break down while MRCISD meets no obstacle. Thus, by CBS extrapolating the two best sets of energies and seeking via CS their smooth merging at the breakup region, one may achieve a global high-level CCSD(T)-quality PES. This might well be called a cuckoo-like approach, putting us in mind of the bird that does not bother to build its own nest but instead mimics the eggs of other birds and takes over their nests.

2.3. Intramolecular and Intermolecular Energies

Perhaps not surprisingly, noncovalent intermolecular interactions form the core of a promising area of research in which the use of CBS extrapolation schemes can be most relevant (39). Because such interactions are typically much weaker than intramolecular ones, the ability to apply highly accurate methods is key. Examples include advanced quantum chemical studies of base stacking and base pairing in the chemistry of RNA and DNA (40, 41) and the accurate description of the activation barrier in the thioldisulfide exchange reaction, since disulfide bonds are abundant in proteins (42). Since the systems of interest are generally of larger size than when studying a simple molecule, the use of high levels of correlation or even large basis sets is prohibited, with CBS being the only route to accuracy. Another important issue in this context is size consistency, which must be satisfied particularly with weak van der Waals (vdW) interactions: While MP and CC methods are size-consistent, MRCI is not (1). There is also BSSE, which manifests for finite basis sets as a spurious attraction even in the absence of interaction and plagues all ab initio molecular orbital (MO) calculations, seriously contaminating the interaction energy. While the popular counterpoise (CP) (43) technique (or variants thereof), which consists of calculating the monomer energies with the full dimer basis, is commonly employed to remedy the problem, full elimination of BSSE is elusive in the case of many-fragment systems (44-46). In contrast, BSSE must vanish in the limit of a complete basis, and hence CBS extrapolation should in principle render CP unnecessary (47, 48). This was, to the best of our knowledge, first demonstrated by the author via highly accurate calculations (47, 48) for He₂, and has been much utilized to calculate PESs ever since (42, 49–57). Similar findings were later reported (58–63), in particular by Mentel & Baerends (61), who noted that CP-corrected bond energies deviate more from the basis set limit numbers than do uncorrected ones. With this applying at HF, CASSCF, and, even more strongly, correlated levels including FCI, the answer to the question, "can the CP correction for BSSE be justified?" is simply no (61).

2.4. Composite Schemes

The extrapolation schemes here discussed differ from other popular ones such as Gn, CBS-Q, and their variants (1). These have various prescriptions and often assume basis set additivity or

use mixed methodologies, while adding empirical corrections of various types such as uniform scaling of force constants and energy corrections based on the number of electrons. Whether they can be considered to be ab initio is then a subtle question. Conversely, the CBS procedures here discussed perform explicit extrapolations of ab initio energies without empiricism. Unless indicated otherwise, they use also a single set of extrapolation parameters for all systems, and hence are called global extrapolation schemes to contrast with system-dependent ones that use a different set of parameters for each system (64). Thus, in principle, global schemes allow CBS extrapolations from arbitrary steps of the hierarchical staircase. Note that the so-called Wn chemistries (65) rely on the extrapolation of the HF and correlation energies (and even include relativistic corrections), thus sharing techniques similar to the ones here surveyed. They have been quite successful, making predictions of an accuracy often higher than that of experimental counterparts. So, not surprisingly, more recent schemes use R12 methods to accelerate basis set convergence (66). Note that even CCSD calculations with a *TZ* basis are often rather expensive for interactions with large molecules. Methods akin to the focal-point approach (67) [used in G2 theory (68) and in our own work (48, 50)] may thus be very useful, although they are not discussed here.

3. SOME GUIDELINES AND PROCEDURES FOR COMPLETE BASIS SET EXTRAPOLATION

CBS extrapolation has been much discussed in papers (14, 23, 24, 26, 27, 29, 69–93), review articles (2, 94), and books (1) and is not recapitulated here except where relevant for the present review. Due to distinct asymptotic dependences, we focus on schemes that split the total energy into HF and correlation components, which are then separately extrapolated:

$$E_X(\mathbf{R}) = E_X^{\rm HF}(\mathbf{R}) + E_X^{\rm cor}(\mathbf{R}), \qquad 2.$$

where cor stands for correlation, **R** stands for the molecular geometry (omitted hereinafter), and all other symbols have the meaning assigned above. Since the only long-range HF contributions arise from electrostatic and induction interactions, the partition in Equation 2 is traceable to the HF-dispersion model (95), which is in turn homologous to the first-order exchange-perturbation plus damped dispersion model (96) and even to DMBE theory (6, 7).

3.1. Hartree–Fock-Type Energy

To treat the HF energy, one of two approaches is usually followed. The first, utilized in early days (97) to extrapolate the energy from correlated wave functions, uses the protocol

$$E_X^{\rm HF} = E_{\infty}^{\rm HF} + A \exp(-\tilde{\beta}X), \qquad 3.$$

where E_{∞}^{HF} , A, and $\tilde{\beta}$ are parameters to be determined from raw energies calculated for at least three cardinal numbers. If these numbers are sequential (X - 1, X, and X + 1), one gets (87)

$$E_{\infty}^{\rm HF} = E_{X-1}^{\rm HF} + \frac{\left(E_{X-1}^{\rm HF} - E_X^{\rm HF}\right)^2}{E_{X-1}^{\rm HF} - 2E_X^{\rm HF} + E_{X+1}^{\rm HF}}.$$
4.

Halkier et al. (72) observed that $\tilde{\beta}$ in Equation 3 has an approximately constant value for their test set of main group species, with $\tilde{\beta} = 1.63$ performing well in dual-level extrapolations with small correlation-consistent bases. This not only reduced the number of calculations to two but also gave deviations from the numerical limits smaller than those obtained from three-point extrapolations. Rather than an algebraic average, Williams et al. (87) suggested that $\tilde{\beta}$ be chosen so as to minimize the maximum absolute deviation (MAD) for a training set with analytical solutions from duallevel extrapolations. Other exponentially decaying functions that were employed to extrapolate the total energy may also be useful for extrapolating the HF part alone, in particular a mixed Gaussian/exponential form (70). With three linear parameters, this widely used form has been found to perform more accurately than Equation 3 for the $H + H_2$ reaction (70) and for transition metal diatomics (87).

Following work by Jensen (80), Karton & Martin (81) proposed the dual-level scheme:

$$E_X^{\rm HF} = E_\infty^{\rm HF} + A(X+1)\exp(-b\sqrt{X}), \qquad 5.$$

where b = 9. Equation 5 shows a root-mean-square deviation (RMSD) of only 0.1 m E_h with the AVQZ and AV5Z basis sets, and 10 μE_h with the (AV5Z, AV6Z) pair. Unfortunately, it does not perform well even with (*T*, *Q*). This led to testing of alternative expressions (81), including Equation 6 below, which is commonly used for the correlation energy with β optimized from a set of test systems. This gives RMSDs of 3.54 (72), 5.34, 8.74, 9.43, and 8.18, respectively, for cardinal number pairs (*D*, *T*), (*T*, *Q*), (*Q*, 5), (5, 6), and (6, 7).

Equation 3 also performs reliably for the CASSCF energy (82). In fact, being system specific and accounting for only the nondynamical correlation, the CASSCF energy is shown in Section 8 to share similar protocols with HF. Often, both are affordable with large basis sets, with the values obtained being assumed to represent the converged result without significant error (76). It turns out that, when used to extrapolate the CASSCF energy from the three highest affordable bases, Equation 3 tends to predict energies that are too low when the exact results are known (82). Given the problematic nature of a fit of three points with three parameters, a test with the four largest bases was done (82). While this was also mostly unsuccessful, it was found (82) that an average of the highest-*X* raw energy with the triple-level CBS one gave striking agreement with predictions from Equation 5.

3.2. Correlation Energy

As noted above, the use of CBS extrapolations of the correlation energy is supported by the energy's dependence on the partial wave quantum number in two-electron atoms and on second-order pair energies in many-electron atoms (see 11–13, 98, 99, to mention only some pioneering work). Thus,

$$E_X^{\rm cor} = E_\infty^{\rm cor} + \frac{A_{\rm cor}}{X^\beta},\tag{6}$$

with all parameters having the usual meaning (24). Theoretically, one expects $\beta = 3$, leading to the so-called X3 protocol. However, recent work suggests that the exponent may vary with the system and the different contributions of the energy (14). Aiming at a rule of potential interest for moderate-sized systems, Truhlar (71) directed attention to extrapolations from small bases (X = D, T) and valence-only limits. By minimizing the RMSD in fits to the estimated basis set limits (24) for Ne, HF, and H₂O, optimal exponents of $\beta_{MP2} = 2.2$ and $\beta_{CCSD} = \beta_{CCSD(T)} = 2.4$ were recommended (33) for MP2, CCSD, and CCSD(T) energies. These were later reevaluated with a larger test set (73), and the protocol was applied (39) to binding energies in noncovalent interactions. Most recently, the idea of allowing $\beta \neq 3$ has been reinforced (91). This topic is revisited in Section 4.3.

Bakowies (84) suggested the use of variable-exponent inverse power laws and used β to quantify deviations from the asymptotic convergence behavior. Both analytical examples (for other asymptotic expansions, see 23, 26, 77–79) and numerical extrapolations to either large but finite or CBS targets were also analyzed (84). He noted that the deviations from asymptotic convergence are most significant for small bases but are still noticeable even when the bases are as large as V5Z or V6Z. By defining exponents $\beta_{opt}(X, X + 1, X + N)$ that are optimal for a set of molecules, Bakowies (84) suggested picking β so as to minimize the RMSD between

extrapolated and explicitly correlated E_{X+N} energies. Instead of a single exponent, the use of specific exponents for specific extrapolations is advocated, a procedure related to the one used by Schwenke (79) but with analytical rather than numerical scaling.

The most popular asymptotic dual-level law is (23, 26, 75, 82, 85, 100, 101)

$$E_X^{\rm cor} = E_\infty^{\rm cor} + \frac{A_3}{(X+\alpha)^3},$$
7.

where the offset angular momentum is determined from some additional criterion (82, 85). On the basis of Equation 1, Klopper (77) suggested the use of distinct one-term expansions for the singlet and triplet pairs, thus accounting for the expected X^{-3} and X^{-5} dependences. Although use of multiple monomial terms in inverse powers of X may then come to mind, the procedure gets cumbersome and not necessarily more accurate. Of course, the parameterization of a multiplemonomial expansion requires as many calculations of a given raw energy using bases of distinct cardinal numbers as there are expansion terms. To alleviate the burden, we have suggested a uniform singlet- and triplet-pair extrapolation (USTE) (82) scheme in which the total correlation is not decomposed, for three reasons. First, the decomposition is unnecessary for accurate results (82). Second, it cannot be implemented for open-shell CCSD calculations (79) (the wave function is not a spin eigenfunction in practical implementations of CCSD theory), nor is it commonly available in CC codes. Third, such a decomposition is extraneous to MRCI: A single excitation out of the reference space can be counted as either a singlet pair or a triplet pair depending on the spin coupling of the N - 1 part of the determinant. These reasons may legitimize reading of the USTE acronym as interchangeable with unified singlet- and triplet-pair extrapolation.

Equations 6 and 7 are both limiting cases of the following rational fraction:

$$E_X^{\rm cor} = \frac{p_0 + p_M X^{-M} + p_{M+1} X^{-(M+1)} + \dots}{1 + q_N X^{-N}},$$
8.

where p and q are coefficients, $M \ge N$, and X may also be thought of as $X + \alpha$. This equation assumes the form of Equations 6 and 7 when $M = \infty$, since for large X values, a Taylor expansion in X^{-1} yields $\lim_{X^{-1}\to 0} [1/(1+q_N X^{-N})] \sim 1-q_N X^{-N}$. Thus, if N = 3, $E_{\infty}^{cor} = p_0$ and $A_3 = -p_0q_3$. Another possibility (see above) is to limit the numerator to odd powers and M = 5. Such a continued fraction has proven reliable for representing correlation energies (86): With $M = \infty$ and the two parameters fixed from energies for Ne with basis sets of large cardinal number (X = 9 and 10) (3), it yielded $E_{\infty}^{cor,S} = -210.64 \text{ m}E_{h}$ and $E_{\infty}^{cor,T} = -104.88 \text{ m}E_{h}$ for the singletpair and triplet-pair CCSD/VXZ energies, versus $-210.61 \text{ m}E_{h}$ and $-104.87 \text{ m}E_{h}$ from CCSD-R12 calculations. It may then be ideal to fit unusual dependences as in R12 calculations (102).

4. COMPLETE BASIS SET EXTRAPOLATION SCHEMES FOR CORRELATION ENERGY

4.1. A Dual-Level Scheme: Uniform Singlet- and Triplet-Pair Extrapolation

The USTE (82) scheme has its origin in a three-parameter rule implicit in Equation 8:

$$E_X^{\rm cor} = E_\infty^{\rm cor} + \frac{A_3}{(X+\alpha)^3} + \frac{A_5}{(X+\alpha)^5},$$
 9.

where E_{∞}^{cor} , A_3 , and A_5 are parameters and the offset factor α effectively accounts for the terms beyond the fifth power in Equation 1. Chosen with the requirement that Equation 7 must reproduce the exact correlation energy for He from FCI/CBS(5, 6) calculations, the value $\alpha = -\frac{3}{8}$ is assumed hereinafter except where otherwise specified (82). By then defining the auxiliary relation $A_5 = A_5^\circ + c A_3^m$ (82), where $A_5^\circ = A_5$ ($A_3 = 0$) and where *c* and *m* are optimized from ab initio energies for a variety of systems in a training set, Equation 9 is transformed into the dual-level USTE rule. **Supplemental Table 1** gathers all parameters fixed from MRCI(Q)/AVXZ energies for 24 systems (82) as well as MP2/VXZ, CCD/VXZ, and CCSD/VXZ energies (77). Naturally, they vary with the method and basis set used, but not significantly for methods and bases belonging to related families. In fact, correlation energies obtained by SR methods and the dynamical part from MRCI(Q) calculations (36), and even correlation energies obtained via intrinsic scaling (103), are accurately extrapolated with USTE even for systems out of the calibrating set (36, 38). For single-reference MP and CC energies, USTE assumes the analytic form (88)

$$E_{\infty}^{\text{cor}} = E_{X_2}^{\text{cor}} - A_5^{\circ}(X_2 + \alpha)^{-5} + \frac{E_{X_1}^{\text{cor}} - E_{X_2}^{\text{cor}} + A_5^{\circ}[(X_2 + \alpha)^{-5} - (X_1 + \alpha)^{-5}]}{c[(X_2 + \alpha)^{-5} - (X_1 + \alpha)^{-5}] + (X_2 + \alpha)^{-3} - (X_1 + \alpha)^{-3}}[(X_2 + \alpha)^{-3} + c(X_2 + \alpha)^{-5}],$$
10.

where X_1 and X_2 are cardinal numbers. Although Equation 10 looks superficially more complicated, it is as easy to solve as the traditional inverse cubic power law for X3.

The USTE scheme may also assume the form

$$E_X^{\rm cor} = E_\infty^{\rm cor} + \frac{A_3}{(X+\alpha)^3} \left[1 + \frac{\tau_{53}}{(X+\alpha)^2} \right],$$
 11.

where $\tau_{53} = A_5/A_3 = (A_5^\circ/A_3) + cA_3^{m-1}$. Showing the proper asymptotic behavior, Equations 9 and 11 therefore allow an estimate of the ratio τ_{53} to be extracted from ab initio correlation energies. Values have been reported at the MP2, CCD, and CCSD levels of theory with VXZ basis sets, as well as at the MRCI(Q)/AVXZ level (86). Note that τ_{53} assumes a value of -0.75 for natural-parity singlet states in the 1/Z expansion for various states of the He isoelectronic series [it corresponds to the ratio of the $(\ell + \frac{1}{2})^{-6}$ and $(\ell + \frac{1}{2})^{-4}$ terms, namely $-\frac{5}{4}$; see 11] (82), although its value may vary if the HF wave function is used (11), as shown in MP2 studies of symmetry-adapted pairs (12). In fact, τ_{53} accounts both for the $(\ell + \frac{1}{2})^{-6}$ term arising from natural parity singlet states and for the leading term with the same dependence in triplet states, and thus for all electron pairs (for details, see 86).

4.2. Dual-Level Generalized Uniform Singlet- and Triplet-Pair Extrapolation

One might wonder at this point how τ_{53} varies over configuration space: If not significantly, can it be constrained at some best estimate (86) leading to a generalized USTE (GUSTE)? This could identify systems with similar A_3 values but distinct A_5 values, such as He-like systems (11), where some triplet states have different values of A_5 for $A_3 = 0$. The generalization consists of writing $A_5 = \eta A_5^{\alpha} + c A_3^m$ (86), where the scaling factor η moves the curve up and down parallel to the original law $[A_5(0) = A_5^{\alpha}, \text{ with } \eta = 1]$ to mimic a desired result. Obviously, one has again the three-parameter rule in Equation 9 unless a criterion is used to fix η . The simplest way is to select η from a calculation with X = Q; this method is called GUSTE/DTQ. Thus, $\eta_{DTQ} = A_3 \left(\tau_{53}^{DTQ} - c A_3^{m-1}\right) / A_5^{\alpha}$, where DTQ implies that only the three smallest cardinal numbers are employed. Naturally, A_3 varies with the cardinal numbers used for extrapolation, and so does η (86). If higher cardinals are affordable (86), one has τ_{53}^{Q56} and GUSTE/Q56. GUSTE commonly employs the average ratio $\overline{\tau}_{53} = \left(\tau_{53}^{DTQ} + \tau_{53}^{Q56}\right)/2$ (86).

www.annualreviews.org • Straightening the Hierarchical Staircase 185



4.3. Other Dual-Level Schemes and Beyond

Because triple-level or superior schemes require raw calculations with ≥ 3 distinct bases, they can be cumbersome and computationally demanding, and hence are only briefly addressed.

If X is large enough, high-order terms can be neglected and the offset parameter α set to 0, with the popular X3 dual-level protocol assuming the form (24, 74, 75, 104–107)

$$E_{\infty}^{\rm cor} = E_X^{\rm cor} + c_X \left(E_X^{\rm cor} - E_{X-1}^{\rm cor} \right),$$
 12.

where $c_X = (X-1)^3/[X^3-(X-1)^3]$; a corresponding expression is obtained if a power of $X \neq 3$ is used (71). This is denoted by X3(X-1,X) or by (X-1,X) if no confusion arises; other notations such as [X-1,X] are occasionally used (91). Much utilized in extrapolations of both intra- and intermolecular energies (39), this method is inaccurate when extrapolating from small cardinals.

An accurate numerical dual-level scheme is due to Schwenke (79) and denoted by $SCH(F_X)$:

$$E_{\infty}^{\rm cor} = E_{X-1}^{\rm cor} + F_X \left[E_X^{\rm cor} - E_{X-1}^{\rm cor} \right],$$
 13.

where $F_X = X^p / [X^p - (X - 1)^p]$ are coefficients (identical to c_X above if the power is p = 3), which were determined (79) by using CCSD-R12 energies (77). Thus, they are dependent both on the basis set used and the correlated method.

Recently, Okoshi, Atsumi & Nakai (91) suggested two schemes for CBS(D, T) extrapolation of MP2, CCSD, and CCSD(T) energies, namely OAN(C)[2,3] and OAN(P)[2,3]. The first uses Equation 6 but with the power optimized using the condition that it must fit the CBS-extrapolated correlation energies calculated with the X3(3, 4) protocol using VXZ and AVXZ basis sets. As test data, the G3X set was utilized for MP2 and CCSD, whereas G2 was employed for CCSD(T). For MP2/VXZ, the authors found a sharp peak for 187 molecules at s = 2.15-2.20 (except for Li₂, for which the peak appears at 2.385); a value of 2.174 was thus chosen. The OAN(P)[2,3] variant represents the prefactor as $A_3(3, 4) = \overline{\alpha}A_3(2, 3) + \overline{\beta}$, with $\overline{\alpha}$ and $\overline{\beta}$ chosen for each combination of method and basis set from the G2 and G3X test sets. Both variants perform accurately versus CBS(T, Q) provides reliable references; second, OAN(C)[2,3] does not follow the s = 3 asymptotic law. A similar comment applies to OAN(P)[2,3], and a stringent test would require high-level explicitly correlated energies.

Multilevel schemes have occasionally been used for benchmarking purposes (29). Using primarily large bases at the frozen-core [fc, as opposed to all-electron (ae)] level, fc-CCSD(T)/AVXZ, Feller (29) tested modifications of five CBS dual- and triple-level extrapolation formulas to reproduce a diverse set of 183 atomization energies. The most effective appears to be a generalized version of Schwenke's form with nine adjustable parameters (versus three in the original protocol), which seems capable of yielding results comparable to the raw values requiring basis sets three steps higher in the hierarchical staircase. A weighted average of such formulas was also tested (29), which turned out to be better behaved than the components in the sense that some approach the asymptote from below while others do it from above.

4.4. A Dual-Level Scheme with Basis Set Rehierarchization: USTE(x - 1, x)

Achieving an accuracy as high as one possibly can get implies obtaining (E_{∞}^{cor}, A_3) from raw energies with the highest affordable rank. Although the (5, 6) cardinal pair is affordable for few-electron systems with most ab initio methods, a tractable CBS/(D, T) is key for larger ones. This is a goal of GUSTE (86), with τ_{53} kept invariant over configuration space once it is determined from a single calculation with a larger X value. This quickly gets out of reach for medium-sized molecules, and we turn to another scheme (48) aiming at the same goal.

The approach begins by recognizing that extrapolations are best discriminated if their form is linear. Equation 9 is then rewritten as $E_X^{\text{cor}} = E_{\infty}^{\text{cor}} + A_3 \tilde{X}^{-3}$ (48), which is equivalent to defining a new hierarchical number (generally a real positive number):

$$\tilde{X} = (X + \alpha) \left[1 + \frac{A_5/A_3}{(X + \alpha)^2} \right]^{-1/3},$$
14.

where A_3 and A_5 have the same meaning as in the original USTE scheme (82). For the reassignment, various approaches were tested while obtaining a highly accurate potential for the weakest interaction in He₂ (48). Subsequently, MP2, CCSD, and CCSD(T) calculations were done for 18 systems at the fc level (108), and nearly all of these were used for USTE (26, 82, 109). The values of \tilde{X} were then determined from the requirement that they fall onto the straight line defined by E_X^{cor} versus \tilde{X}^{-3} , a procedure applied to the whole test set using the largest affordable cardinal numbers (48). The newly defined $x = \tilde{X}$ values are shown in **Supplemental Table 2** using the hybrid Latin/Greek labeling x = d, t, q, p, b. Tested on 106 neutral and closed-shell species (84) including H₂, the method (108) keeps the original acronym but is accompanied by the hierarchical numbers used for the extrapolation: USTE(x_1, x_2).

Shown in **Supplemental Figure 1** are the average deviations between the novel hierarchical and cardinal numbers, $\Delta \chi$. Notably, the cardinal is always larger than the hierarchical number for CC methods, but smaller for MP2 in the case of the *DZ* basis. The actual basis is then statistically better than *DZ* for MP2, while others are slightly poorer than the ones they represent. Also visible from the same plot is the fact that the hierarchical number tends asymptotically to a value $\simeq 0.26$ smaller than the cardinal, both for CC and for MP2. Such a pattern is well described by

$$\Delta \chi = p_0 + \frac{1}{2} \tanh\left[\frac{(X - X_0)}{p_1}\right],$$
15

where $\Delta \chi = X - x$, $X_0 = 1.75$ (the crossing point of the MP2 and CC curves) is a reference, and p_i are parameters: $p_0 = -0.2110$ and $p_1 = 1.7211$ for MP2; $p_0 = -0.2004$ and $p_1 = 0.3770$ for CC. Good one-parameter fits are still obtained with p_0 fixed at $p_0 = -0.2057$, yielding $p_1 = 1.7816$ and $p_1 = 0.3678$ for MP2 and CC, respectively. Clearly, X - x attains a nearly constant value for CC from TZ onward. In fact, the utilized basis sets tend to converge asymptotically, in both the CC and MP methods, to hierarchical numbers $\simeq 0.3$ smaller than the corresponding cardinals. The only anomaly is DZ, which partly explains why CBS extrapolations from basis sets with the two lowest cardinal numbers are problematic with the X3 (23, 24) protocol. Indeed, USTE(d, t) significantly outperforms the original USTE (82) scheme. The universality of the novel hierarchical numbers across other hierarchized bases has been verified (90), and the results have even been enhanced for AVXZ basis sets. In general, USTE(x - 1, x) (108) outperforms USTE(X - 1, X) (82) and outperforms X3 (23, 24) even more. In turn, it shows excellent agreement with the results from Schwenke's (79) numerical protocol for CC energies and polarizabilities, which are critical for accurate long-range interactions (108, 110).

4.5. More on Dual-Level Schemes with Basis Set Rehierarchization

Consider the inverse X power law in Equation 7 with an exponent β . One has (64)

$$E_{\infty}^{\rm cor} = E_{X-1}^{\rm cor} + \left(E_X^{\rm cor} - E_{X-1}^{\rm cor}\right) \left[\left(\frac{X+\alpha}{X+\alpha-1}\right)^{\beta} - 1 \right]^{-1},$$
 16.



Figure 2

Unnormalized frequency distributions of the optimal extrapolation exponents (ideal β obtained from Equation 17) for a 106-species test data set as fitted to Gaussian distributions. Abbreviations: CCSD(T), coupled-cluster theory with single, double, and perturbative triple excitations; MP2, second-order Møller–Plesset theory.

from which it follows that

$$\beta = \ln\left(\frac{E_X^{\text{cor}} - E_{X-1}^{\text{cor}}}{E_{\infty}^{\text{cor}} - E_X^{\text{cor}}} + 1\right) / \ln\left(\frac{X+\alpha}{X+\alpha-1}\right).$$
 17.

Two questions arise. The first question is whether the optimal shifted angular momenta $(X + \alpha)$ obtained by matching the hierarchical numbers in USTE(x - 1, x) predict the ideal asymptotic β value. The second one is whether, by varying α and β to minimize the residuals, one might enhance the performance of USTE(x - 1, x). These questions are particularly pertinent for (d, t) and (t, q), since CBS extrapolation from larger pairs is expected to mimic the asymptotic behavior. Because the second question demands investigation, only the first one, concerning whether $X + \alpha$ can be approximated by $X - \Delta \chi$ in Equation 15, is considered, using the 106-species test set and USTE(5, 6) as a reference.

Figure 2 shows histograms of the optimum β values for (d, t), (t, q), and (q, p), all broadly depicting a Gaussian-type shape, except the (d, t) distribution, which shows a somewhat asymmetrical form with an outlier point at $\beta \sim 4$. They were fitted to a three-parameter Gaussian form, not necessarily normalized, with the fits mimicking the histograms, except in the (d, t) case, which has an asymmetric bell shape and is unpopulated for values $x \leq 2.8$ with an outlier close to 3.5 (the value is for H₂ and is thus of no concern for extrapolation purposes; see 108). As shown, all distributions peak close to 3, with the peak positions and average values being, in the order given above, (2.95, 2.98), (2.98, 2.97), and (2.96, 2.94) for MP2 and (2.91, 2.94), (3.00, 2.99), and (2.98, 2.95) for CCSD. This supports a faster MP2 versus CCSD convergence to the CBS limit: All β values get closer to 3 for MP2 than for CCSD, from (d, t) onward. The results seem also to corroborate the scheme employed in the USTE(x - 1, x) calibration. Although they use the hierachical numbers in **Supplemental Table 2** (48), they should be similar to the results obtained by combining Equations 15 and 17. Indeed, a fully analytical USTE $(\tilde{x} - 1, \tilde{x})$ may be $E_X^{cor} = E_{\infty}^{cor} + A_3 \tilde{x}^{-3}$, where $\tilde{x} = X - \Delta \chi$ and where $\Delta \chi$ is given by Equation 15; analysis of this topic is outside the scope of the present review.

With various dual-level CBS schemes in the literature for extrapolating the correlation energy, one might think of recasting them into a unifying form. Seino & Nakai (93) have shown such a

form to be analogous to Equation 12, but with the constant redefined as $c_X = (X - 1 + \eta)^{\beta} / [(X + \gamma)^{\beta} - (X - 1 + \eta)^{\beta}]$, where η and γ are parameters; they validated it for various dual-level models (23, 48, 84, 91, 93, 108, 111) using *DZ* and *TZ* basis sets.

4.6. A Single-Level Scheme: USPE(*x*)

Until recently, the only single-level protocol was $E_X^{\text{cor}}/E_{\infty}^{\text{cor}} = 1 - 2.4X^{-3}$ (with energy in E_h), a golden rule (26) hereafter denoted by V2000, where the prepower constant was chosen from 33 distinct data sets referring to 20 molecules and one atom (24, 109). Although E_{∞}^{cor} lies closer to the CBS limit than E_X^{cor} , it performs modestly compared with dual-level protocols, which is not surprising because of the lack of parameters to characterize even the method.

For a single-level scheme, A needs to be determined from some other criterion. The simplest would be to consider its average value for a test set, but this is not expected to yield any significant improvement over V2000, since A varies widely with the system. The alternative is to relate A to a byproduct of the correlation calculation, for example (89)

$$E_{\infty}^{\rm cor} = E_X^{\rm cor} + \frac{\mathcal{A}(E_X^{\rm tot})}{x^3},$$
18.

with $A(E_X^{\text{tot}})$ chosen by imposing that the extrapolations mimic some reference values. Using the USTE(5, 6) predictions (82) and the 18-system data set (108), this yields $A(E_X^{\text{tot}}) = b + a E_X^{\text{tot}}$, with good fits reported for the MP2 and CCSD methods as well as the VXZ and AVXZ bases (89). This may be explained by the fact that A represents an energy that grows approximately with the number of electrons. However, the dependences on the atomic number (Z) of E_X^{tot} and E_X^{cor} are distinct (112, 113), and hence Equation 19 fails for atoms in the stringent criterion of consistency of dimensionality. In fact, the total energy (112) varies as $-AZ^{7/3} + BZ^2 - CZ^{5/3} + \cdots$, while the leading dependence of the correlation energy was found numerically (113) to be $-a Z^{4/3} - b Z^{2/3} - c Z^{1/3}$. Yet the above is strictly valid for atoms, whereas the ideal formula must also apply to molecules, although thus far only molecules made of atoms from H to Ne have been considered. Since no oscillations are expected, and primarily because Z is not defined for a molecule, Equation 18 may effectively warrant such a requirement. Thus,

$$E_{\infty}^{\rm cor} = E_X^{\rm cor} + \frac{af_x E_X^{\rm tot}}{x^3} = E_X^{\rm cor} + \frac{a[1+\kappa(X)]E_X^{\rm tot}}{x^3},$$
19

where *a* characterizes the method and basis set and where $\kappa(x)$ [or $\kappa(X)$] is an adaptative auxiliary form (48, 108). Clearly, Equation 19 is applicable only to systems with $E_X^{\text{cor}} \neq 0$; **Supplemental Table 2** gathers the average values of *a* in this unified single-parameter extrapolation [USPE(*x*)] scheme. Despite its limitation to first-row atoms, it often outperforms dual-level schemes.

When aiming at applicability to atoms and molecules as well as distinct rows of the periodic table, the above lack of dimensionality may require some solution. One such solution consists of relating correlation energies calculated at two distinct levels of theory, one much cheaper than the target level. Two such methods are density functional theory (DFT) and MP2. Since the former is part of a plethora of methods not considered in this work, we focus on MP2. **Supplemental Figure 3***a* shows that a linear relationship holds between the CCSD(T) and MP2 correlation energies, with the latter being costless since it is part of the CCSD(T) output. The calculations (all employing MOLPRO; see 114) include the 106-species test set (89) plus 20 systems involving elements of higher periods that are part of a training set used (115) to calibrate a restricted-openshell CBS model chemistry. Although only the 18-system set is used in **Supplemental Figure 3***b*, a linear relationship holds between $A_{cor}^{MP2}(X)$ correlation and MP2 correlation, thus setting a promising scenario for predictions as consistent as those using USPE(x). This is a fair expectation,

Supplemental Material

since $A(E_{\rm HF})$ should be well approximated linearly for both MP2 and CC, which also suggests an approximately linear relationship between $E_{\rm cor}^{\rm CC}$ versus $E_{\rm cor}^{\rm MP2}$. Since this is shown only for atoms of the first period, it requires validation for other atoms. Yet even if distinct linear relations apply, the ratio is expected, on the basis of **Supplemental Figure 3**, to be linear. Note that a scaling relation for CCSD(T) and MP2 correlation interactions of a noncovalent complex was also reported (116).

Assessed with the 106-species test set, USPE(*t*) outperforms (d, t), while (X > T)-predicted correlation energies show deviations similar to (X - 1, X) or (x - 1, x) (89). In turn, USPE(*q*) outperforms computationally demanding CCSD(F12)/def2-QZVPP results (117). Regarding CCSD-F12b/3C(FIX) (118), the predicted values tend to fall slightly below. Notably, USPE(*d*) outperforms USTE(*D*, *T*) and X3(*D*, *T*) for MP2, and X3(*D*, *T*) for CCSD.

Regarding Equation 19, MP2(F12)/AV6Z (119) predicts slightly larger correlation energies than most extrapolations (89), a trend seen from a statistical analysis of the error/valence electron in **Supplemental Figure 4**. Also visible is the fact that explicitly correlated methods tend to perform inefficiently with small bases but remarkably well with large ones. Note that both conventional and R12-type methods must perform similarly in convergence when approaching the CBS limit, and hence any convergence delay between the two must reach a maximum at some intermediate cardinal number. This is the case with MP2 and MP2-F12, with $\kappa(X)$ in Equation 19 being the accumulated delay in CBS convergence up to a given X. For the 18-system set at MP2(F12)/AV6Z (119) and MP2 (VXZ and AVXZ) levels, one gets (89) $f_d = 1.00$, $f_t = 1.02$, $f_q = 1.04$, $f_p = 1.08$, and $f_b = 1.13$, which are well mimicked by

$$\kappa^{\text{MP2}}(X) = \eta + \frac{1}{2} \tanh\left(\frac{X - X_m}{5}\right),$$
20.

where $\eta = 0.46$ for both the newly hierarchized and conventional bases (see **Supplemental Figure 4**). In turn, $X_m = 10$ and $\delta = X_m - x_m = 0.3$, with the denominator of the tanh argument fixed at 5 to make it close to -1 at x, X : d, D. Note that $\eta \neq 0.5$ reflects the fact that the smallest hierarchical number is D: 2 (d), where $\kappa^{MP2} = 0$, while in the high-X limit, $\kappa^{MP2} \sim 1$. This suggests that the accumulated delay of MP2 attains $\sim 50\%$ of its maximum at $X \sim 10$, a nonverifiable prediction until bases with cardinal numbers as large as $X \gtrsim 10$ become available. Although a novel scheme (USPE_{fx}) may then be proposed, its rate of delay $\propto \operatorname{sech}^2[(X - X_m)/5]$ may require validation from ab initio calculations. Regarding CCSD-F12, $f_{x,X}^{CCSD} \sim 1$, as USTE and USPE predict slightly more negative CBS energies, which may be explained by the slower basis set convergence (118) of CCSD-F12.

Note that a Taylor series expansion may suggest an effective convergence power law that is somewhat smaller than inverse cubic. Interestingly, this may explain some power laws (71, 84, 91) available in the literature, although the current protocol shows the well-established asymptotic power. For CCSD, some previous work (71, 84, 91) predicts a negative power law larger than for MP2, also in line with the above analysis.

5. CAN COMPLETE BASIS SET EXTRAPOLATION AVOID COUNTERPOISE?

A challenging undertaking in modern electronic structure theory is the quantitative treatment of intramolecular and intermolecular interactions. The traditional approach, which is the focal point of the present analysis, combines advances in methods for treating electron correlation with a hierarchical expansion of the basis, which is then systematically expanded in both its radial and angular parts to reach the CBS limit. The question then arises whether a quantitative account of the BSSE is required prior to CBS (74). From an accurate study (47, 48) on He₂, the author

has shown that CBS extrapolation of both CP-corrected and CP-uncorrected energies leads to essentially the same value (within 10^{-2} K or so) of the vdW well depth, with the sole benefit of CP being a smoother convergence pattern. Although the suggestion (47, 48) that CP could be eschewed gathered support (58–63), a debate arose. The topic is here elaborated with the help of a few case studies.

Being a simple concept, BSSE is highly accepted, and so is CP as a way to approximately correct it (43). However, CP has drastic effects, besides making the calculations expensive and not necessarily more accurate in absence of CBS (32, 38, 47, 48). For example, the CBS(D, T) binding energy of the formic acid dimer gets unusually large when including BSSE with augmented DZ and TZ bases, and it does not decrease with basis set size. In line with other findings (60, 120), this is attributed to two effects. One effect is the BSSE itself. The resulting error should then decrease with basis set size, which implies that CP is superfluous at the CBS limit. The other effect is ascribed (60) to an intramolecular BSSE, supposedly due to an artificial improvement of the electron correlation. Imputed to an unbalanced description of the electron correlation, it would significantly increase the intermolecular BSSE, since both effects seem related by a power law (120). Due to such undesirable intramonomer effects, both BSSE corrections were questioned, with Sheng et al. (60) refraining from including any corrections at all for their CCSD(T) results, in agreement with the above (47, 48).

Another elucidative study uses the barely bound helium dimer and a FCI treatment. Two errors were assigned from basis set incompleteness. One is BSSE, while the other, called basis set convergence error (BSCE), is the exact interaction energy minus the CP-corrected FCI energy in the same basis (60). They were found from FCI/AVXZ (X = D-5) calculations to have opposite contributions. BSCE (negative) is larger at short range and remains larger than (but closer to) BSSE around the vdW minimum. Only at long range does BSSE become larger. Because they largely cancel each other out over the vdW well, it was recommended (60) that BSSE not be included in calculations from short distances until well beyond the vdW minimum. In other words, CP is required only if accurate tail behavior is at stake. Since caution must be exercised there anyway (even after CBS) if the basis misses sufficient diffuse functions, one can refrain from using CP altogether for most purposes. Tests were also done for other systems where basis set free benchmark numbers are available for monomers and dimers, with Be₂ being another example. A positive answer to the question has in this case too been verified (61). Indeed, the conclusions seem to hold both at HF and, even more strongly, at correlated [CCSD(T), FCI] levels (61).

A final remark should be made on results suggesting that BSSE should account for substantial geometry changes of the monomers due to formation of strong hydrogen bonds in the dimer (121). Considered key (122) for obtaining accurate binding energies of the formic acid dimer (since the BSSE displays the expected decrease with basis set size), the problem was later attributed (123) to undue use of a valence basis set jointly with correlation of all electrons, an effect akin to that reported (55) in a study of the HS₂ reaction course. In summary, even if CP-corrected and CP-uncorrected energies yield identical results upon CBS extrapolation, the latter are easier and cheaper to get (32, 38, 47, 48). It thus seems legitimate to obviate CP altogether when extrapolating for basis set incompleteness.

6. COMPLETE BASIS SET EXTRAPOLATION OF MOLECULAR PROPERTIES

6.1. Tensorial Properties

So far, only extrapolations of energies have been considered, which may lead to more accurate energy differences. Since other properties can be calculated by electronic structure methods (1),

one might ask how much they benefit from CBS extrapolation. To answer this, we first recognize that the envisaged property may be defined as the response of a wave function, an energy, or an expectation value of an operator to a perturbation, this being any operator not present in the Hamiltonian, for example an external field. Relativistic terms are here ignored, and so are properties, such as transition moments, that depend on matrix elements between two distinct wave functions. The analysis is further restricted to static properties and electronic contributions to properties of a single molecule at a fixed geometry.

Consider the energy expanded as a Taylor series in the perturbation strength λ ,

$$E = E(0) + \frac{\partial E}{\partial \lambda} \lambda + \frac{1}{2!} \frac{\partial^2 E}{\partial \lambda^2} \lambda^2 + \frac{1}{3!} \frac{\partial^3 E}{\partial \lambda^3} \lambda^3 + \cdots, \qquad 21.$$

where the perturbation is usually a vector, the first derivative is also a vector, the second derivative is a matrix, the third derivative is a (third-order) tensor, etc., and $\partial^n E/\partial \lambda^n$ is an *n*th-order property. For example, if the perturbation is due to moving a nucleus, Equation 21 becomes

$$E(\mathbf{R}) = E(\mathbf{R}_0) + \frac{\partial E}{\partial \mathbf{R}} (\mathbf{R} - \mathbf{R}^0) + \frac{1}{2!} \frac{\partial^2 E}{\partial \mathbf{R}^2} (\mathbf{R} - \mathbf{R}^0)^2 + \cdots$$

= $E(0) + \sum_{i=1}^3 G_i (R_i - R_i^0) + \frac{1}{2!} \sum_{i,j=1}^3 H_{ij} (R_i - R_i^0) (R_j - R_j^0) + \cdots, \qquad 22.$

where the expansion has been truncated at the second order: The first derivative is the gradient **G**, the second derivative is the force constant (Hessian) **H**, etc. Thus, $\mathbf{G} = \mathbf{0}$ if \mathbf{R}_0 is a stationary point, and the force constant matrix may be used to evaluate the harmonic vibrational frequencies and normal coordinates **Q**.

First consider the case where the perturbation is an external electric field:

$$E(\mathbf{F}) = E(0) + \sum_{i=1}^{3} \mu_i F_i + \frac{1}{2!} \sum_{i,j=1}^{3} \alpha_{ij} F_i F_j + \cdots, \qquad 23.$$

where F_i are components of the applied field and E(0) is the total energy in its absence. If **F** is uniform and applied along the axes i = x, y, or z, then $\mu = \mu_i$ is the dipole moment and $\alpha = \alpha_{ii}$ the dipole polarizability, respectively the negative of the first and second derivatives of the energy relative to the field (124). As these are fundamental properties of matter that are difficult to experimentally determine (the inherent uncertainties are often undesirably large; see 125), ab initio methods offer a valuable route for their accurate determination.

We now write

$$E(F_i) = E^{\rm HF}(F_i) + E^{\rm cor}(F_i), \qquad 24.$$

with CBS extrapolations done as described in previous sections. If Equation 23 is now truncated at the fourth order and E(0) and $E(F_i)$ are extrapolated at four field values ($F_i = \pm 0.005, \pm 0.01$ a.u.), a system of linear equations can be solved and the properties calculated (83). If the procedure is then carried out for the x, y, and z field components, average polarizabilities can be obtained in good agreement with experiment (83). Dual-level (triple-level for HF) CBS extrapolations of polarizabilities were also reported for tetravalent actinide ions (126) with ae-relativistic core-correlation bases, and at the triple level with AVXZ and dAVXZ (X = Q, 5, 6). Moreover, calculations of static dipole polarizabilities were performed for Ne, CO, N_2 , F_2 , HF, H₂O, HCN, and C₂H₂ at the CBS level by focal-point analysis (127).

Because the above procedure requires four correlated calculations with applied electric fields and one without for two cardinal numbers (83), the method is expensive, and hence we next consider a simpler variant. For this, let the properties in question be written as (110)

$$\zeta = \zeta^{\rm HF} + \zeta^{\rm cor}, \qquad 25$$

where $\zeta = -\partial^{\sigma} E(F_i)/\partial F_i^{\sigma}$, with $\zeta = \mu_i, \alpha_{ii}$, etc. Thus, $\zeta^{\text{HF}} = -\partial^{\sigma} E^{\text{HF}}(F_i)/\partial F_i^{\sigma}$ and $\zeta^{\text{cor}} = -\partial^{\sigma} E^{\text{cor}}(F_i)/\partial F_i^{\sigma}$. If the HF and correlation energies in the presence of an external field are next assumed to converge to the CBS limit at a similar rate as in its absence, one has

$$\zeta_{\text{CBS}}^{\text{HF}} = \zeta_{X_2}^{\text{HF}} + \frac{X_2^{-\kappa}}{X_1^{-\kappa} - X_2^{-\kappa}} \Big(\zeta_{X_2}^{\text{HF}} - \zeta_{X_1}^{\text{HF}} \Big),$$
 26

where $X_1 = X_2 - 1$ and κ depends on the cardinal pair (X_1, X_2) used for extrapolation. Similarly, if USTE(x - 1, x) or the X3 protocol is used for the correlation energy,

$$\zeta_{\rm CBS}^{\rm cor} = \zeta_{X_2}^{\rm cor} + \frac{x_2^{-3}}{x_1^{-3} - x_2^{-3}} \Big(\zeta_{X_2}^{\rm cor} - \zeta_{X_1}^{\rm cor} \Big),$$
 27

where $x_i = d, t, ...$ (108) for USTE(x - 1, x) and $X_i = D, T, ...$ for X3 (23). Thus,

$$\zeta^{\text{CBS}} = \zeta_{X_2} + \frac{x_2^{-3}}{x_1^{-3} - x_2^{-3}} \left(\zeta_{X_2} - \zeta_{X_1} \right) + \left(\frac{X_2^{-\kappa}}{X_1^{-\kappa} - X_2^{-\kappa}} - \frac{x_2^{-3}}{x_1^{-3} - x_2^{-3}} \right) \left(\zeta_{X_2}^{\text{HF}} - \zeta_{X_1}^{\text{HF}} \right), \qquad 28.$$

where ζ_{X_1} and ζ_{X_2} are properties at the full correlated level, and $\zeta_{X_1}^{\text{HF}}$ and $\zeta_{X_2}^{\text{HF}}$ at the HF level. Alternatively, with USTE(X - 1, X) (82), this assumes the form

$$\zeta^{\text{CBS}} = \zeta_{X_2} + B(X_1, X_2) \left(\zeta_{X_2} - \zeta_{X_1} \right) + \left[\frac{X_2^{-\kappa}}{X_1^{-\kappa} - X_2^{-\kappa}} - B(X_1, X_2) \right] \left(\zeta_{X_2}^{\text{HF}} - \zeta_{X_1}^{\text{HF}} \right), \qquad 29$$

where

approximation is plausible.

$$B = \frac{-\left[(X_2 + \alpha)^{-3} + c(X_2 + \alpha)^{-5}\right]}{c\left[(X_2 + \alpha)^{-5} - (X_1 + \alpha)^{-5}\right] + (X_2 + \alpha)^{-3} - (X_1 + \alpha)^{-3}};$$
 30.

unlike Equations 10 and 29, this does not depend on A_{5}° . In turn, with USPE(x), one has

$$\zeta^{\text{CBS}} = \left(1 + \frac{a}{x_2^3}\right)\zeta_{X_2} + \frac{X_2^{-\kappa}}{X_1^{-\kappa} - X_2^{-\kappa}} \left(\zeta_{X_2}^{\text{HF}} - \zeta_{X_1}^{\text{HF}}\right),$$
31

where *a* is defined in **Supplemental Table 3** and other symbols have the meanings given above. If a finite difference method is used, $\alpha_{ii} = -[E(F_i^{(1)}) - 2E(F_i^{(2)}) + E(F_i^{(3)})]_0/b^2$, where $F_i^{(1)} = -b$, $F_i^{(2)} = 0$, and $F_i^{(3)} = b$ are the applied electric fields along i = x, y, z. Clearly, *b* must be small to warrant converged results, but large enough (0.005) to facilitate differentiation (125). With USPE, the average polarizability will look like Equation 31, but with $\bar{\alpha}$ replacing ζ , which is manifestly cost-effective, since only one correlated calculation is required. The good performance of USTE (83, 110) has been demonstrated: Even at the (d, t) level, the predicted polarizabilities show errors (<1%) similar to those of the most accurate MP2 and CC raw values. Note that the direct CBS extrapolation of raw polarizabilities is erroneous, albeit with small errors (110). For large hierarchical numbers, the difference between $\zeta_{ii(X_2)}^{\text{HF}}$ and $\zeta_{ii(X_1)}^{\text{HF}}$ is small, and such an

Consider next the gradients required to integrate the equations of motion in direct dynamics or the frequencies that imply second derivatives. The solution stems from Equations 28–30, since for a continuous function f(x) in the interval [x - b, x + b], the intermediate value theorem warrants that there is a point $\xi \in (x - b, x + b)$ such that $f^{(n)}(\xi) = [f^{(n)}(\xi_1) + f^{(n)}(\xi_2)]/2$, with $\xi_1 \in (x - b, x)$ and $\xi_2 \in (x, x + b)$. Gradients and Hessians can then be extrapolated when such raw data are available. Two observations are in order. First, the second term of Equation 28 is nonnegligible [for (d, t): -0.235 in MP2 and -1.118 in CCSD], which may lead to nonmonotonic

behavior, depending on the ζ -differences at HF and correlated levels. This has been verified from calculations done here for 8 diatomics, all at optimum ground-state geometries: H₂, N₂, HF (hydrogen fluoride), C₂, O₂, CO, HCl, and S₂. Second, the equilibrium geometry varies with X, which may hamper correct extrapolation. Suffice it to add that convergence of the raw ζ^{HF} and ζ^{MP2} properties is itself often nonmonotonic, although (d, t) predictions from Equation 28 mimic the MP2 and CC values at the V6Z level to within about 2%, a deviation similar to that often obtained in optimizing the equilibrium geometry R_{ϵ} . Expectedly, enhanced convergence can be envisaged by numerical differentiation or analytically from ζ -CBS extrapolations with high hierarchical pairs.

6.2. Atomization Energies

An essential component of ab initio calculations is their capability to predict or rationalize experimental outcomes. The atomization energy of a molecule is an example closely related to the heat of formation, since it differs only by the reference state and neglect of vibrational effects: For the former, the reference state is the isolated atoms, and for the latter, the elements in their most stable form. Because the dissociation energy for the reference molecules can be measured, the atomization energy is an experimental observable.

Accurate thermochemical calculations are feasible only for small systems formed from firstand second-row atoms (128, 129), with CCSD(T) being the state-of-the-art method once effects are added due to convergence to the one-particle CBS limit, core/valence correlation, scalar relativistic correction, and sources (see **Figure 1**). Zero-point vibrational energy (ZPE) corrections are also important for heats of formation and introduce their own computational difficulties. Still, a correction to the dissociation energy that accounts for the intrinsic error in CCSD(T) must be added. Ideally, a FCI wave function and a basis capable of capturing the magnitude of the effect should be employed. Use of FCI and a *TZ* valence basis jointly with the above composite scheme could then be a strategic choice, since it can differentiate between heats of formation that differ by ≤ 0.5 kcal mol⁻¹ (128). Alternatively, inclusion of fully iterative triple excitations (CCSDT) may be adopted (117, 129), as the involved $\sim n^8$ scaling may still be affordable.

The performance of a CBS extrapolation scheme can therefore be plausibly addressed from calculations of atomization energies once some additivity scheme is adopted. One due to Klopper et al. (117) adds to fc-CCSD/CVQZ energies the following: core-valence correction, perturbative corrections for connected quadruple excitations [CCSDT(Q)], scalar relativistic and spin-orbit effects, anharmonic ZPE vibrational effects, and HF corrections. To improve the correlation at the fc-CCSD/CVQZ level, fc-CCSD(F12)/def2-QZVPP calculations are advocated (17), with an ΔE_{F12} correction believed to give subchemical accuracy: few kJ mol⁻¹ versus chemical (<1 kcal mol⁻¹) versus spectroscopic (<1 cm⁻¹). Thus (117),

$$E_{\text{total}} = E_{\text{fc-CCSD}(T)/CVQZ} + \Delta E_{CV} + \Delta E_{ZPE} + \Delta E_{\text{anh}} + \Delta E_{F12} + \Delta E_{MVD} + \Delta E_{SO} + \Delta E_{T} + \Delta E_{(Q)}, \qquad 32.$$

with all but ΔE_{anh} computed (130) at the ae-CCSD(T)/CVTZ equilibrium geometry: ΔE_{CV} is a CVQZ core-valence correlation correction obtained as the difference between ae-CCSD(T) and fc-CCSD(T) energies, ΔE_{ZPE} is the ae-CCSD(T)/CVTZ harmonic ZPE, ΔE_{anh} is the correction to ZPE due to anharmonic effects at the fc-MP2/VDZ level, and ΔE_{MVD} is the ae-CCSD(T)/CVTZ correction for scalar relativistic effects (130). Spin-orbit corrections (131) are also considered: $\Delta E_{SO} = 0.35399$, 0.93278, and 1.61153 kJ mol⁻¹ for C, O, and F, respectively. Also added is the full correction for connected triples and a perturbative correction for connected quadruple excitations (132):

$$\Delta E_T = E_{\text{CCSDT/VTZ}} - E_{\text{CCSD(T)/VTZ}},$$
33

$$\Delta E_{(Q)} = E_{\text{CCSDT}(Q)/\text{VDZ}} - E_{\text{CCSDT}/\text{VDZ}}.$$
34.

Finally, a correction for basis set truncation error at the MP2-F12 level of theory is included, ΔE_{F12} , which (for enhanced results) was subsequently estimated at the CCSD(F12) level of theory,

$$\Delta E_{\text{F12}} = \Delta E_{\text{fc-CCSD(F12)/def2-}QZVPP} - \Delta E_{\text{fc-CCSD/CV}QZ},$$
35.

as well as a HF correction

$$\Delta_{\rm HF} = E_{\rm HF/CV5Z} - E_{\rm HF/CVQZ} \qquad 36.$$

and, for the triple excitations [via fc-CCSD(T) with CVQZ and CV5Z bases followed by X3 extrapolation], a correction to the CCSD/CVQZ energies (117).

Recently, we employed (108) the full scheme of Klopper et al. (117) but replaced ΔE_{F12} by CBS-extrapolated correlation energies, denoted by $\Delta(\mathcal{X} - 1, \mathcal{X})$:

$$\Delta(\mathcal{X} - 1, \mathcal{X}) = \left(E_{(\mathcal{X} - 1, \mathcal{X})}^{\text{cor(atom)}} - E_{\text{CCSD/CV}QZ}^{\text{cor(atom)}} \right) - \left(E_{(\mathcal{X} - 1, \mathcal{X})}^{\text{cor(mol)}} - E_{\text{CCSD/CV}QZ}^{\text{cor(mol)}} \right).$$
37

Also, geometries were optimized at the ae-CCSD(T)/CVTZ level (thus ensuring calculations at minima), and atomic or molecular correlation energies used HF or restricted open-shell HF references. A comparison of RMSDs for calculated atomization energies of 73 systems from the Active Thermochemical Tables (133) shows a significant improvement for CBS extrapolations beyond (d, t) or F_Q relative to CCSD/CVQZ values (108). Generally, USTE(x - 1, x) and SCH (F_X) perform best, and USTE(q, p) and SCH (F_Q) show the smallest errors. However, it is unwise to extract definite conclusions, since some of the steps involved in the process of calculating molecular atomization energies may involve somewhat uncertain approximations.

7. ASSESSMENT OF LOW-COST COMPLETE BASIS SET EXTRAPOLATION SCHEMES

After examining various schemes for CBS extrapolation of the correlation energy, one might seek those schemes that are least expensive and potentially most useful in practice. Included in this discussion are the USTE(x - 1, x), USPE(x), SCH (F_X) , and OAN(X - 1, X) schemes, whose performances are assessed from the RMSD for the 106-species test set (84, 108, 117, 134) using MP2, CCSD, and CCSD(T) with VXZ (15, 16, 22), XZP (18), and X-ZaP (19, 20) basis sets. The results (90) are shown in Supplemental Figure 5, with MP2(F12)/AV6Z (119) and USTE(5, 6)/VXZ taken as reference energies. The first observation concerns the VXZ and X-ZaP bases, with USPE(t) outperforming USTE(d, t) at a lower cost. The performance of USTE(d, t) is, however, enhanced with XZP, with improvements greater than 1 kcal mol⁻¹ in MP2 and in CCSD versus the VXZ homolog. The largest USTE(x - 1, x) deviations correspond to the X-ZaP basis, with USTE(D, T) outperforming USTE(d, t) for CCSD, a comparison that may vary slightly when basis-specific x values are used. Still, USTE(d, t) and USPE(t) outperform OAN(C)(D, T)(135) for MP2 with any basis, but the opposite is valid for CCSD even if by ≤ 0.4 kcal mol⁻¹. In turn, SCH(F_T)/CCSD outperforms USTE(d, t) with VXZ and X-ZaP bases, but USPE(t) outperforms SCH(F_T) in all cases. Notably, USPE(x) can be competitive with USTE(x - 1, x), particularly for $x \equiv t$. In fact, USPE(t) outperforms USTE(d, t) with VXZ and X-ZaP bases, while USPE(t)/X-ZaP generally outperforms $SCH(F_{(T)})$ and OAN(C)(D, T). If TZ calculations are affordable, USPE(t) may be the best choice if only H and first-row atoms are involved. Often, all schemes show performances better than R12/TZ-type calculations.

8. COMPLETE BASIS SET EXTRAPOLATION OF HARTREE-FOCK-LIKE ENERGIES: AN UPDATE

Recently (92) we developed a dual-level scheme for the CBS extrapolation of the HF energy from arbitrary hierarchical numbers, not necessarily sequential. The new protocol generally outperforms the best available ones and is also valid for CASSCF-type energies.

Consider Equation 3, where, for a fixed universal parameter $\tilde{\beta}$,

$$E_{\infty}^{\text{CAS}} = \frac{E_{X_i} e^{\beta X_i} - E_{X_j} e^{\beta X_j}}{e^{\beta X_i} - e^{\beta X_j}}.$$
38.

Aiming at straightening the hierarchical staircase, assume E_{∞}^{CAS} , β , and one hierarchical number $x_k = X_k$ as known. The others $(i = 2, 3, ... \text{ and } i \neq k)$ are determined from (92)

$$x_i = \frac{1}{\tilde{\beta}} \ln \left[\frac{e^{\tilde{\beta} x_k} (E_{\infty}^{\text{CAS}} - E_{x_k}^{\text{CAS}})}{E_{\infty}^{\text{CAS}} - E_{x_i}} \right],$$
39.

where $x_2 = d$, $x_3 = t$, $x_4 = q$, $x_5 = p$, $x_6 = b$, and $x_7 = s$ (48, 108), with the method being denoted by CAS-E(x_i, x_j) or CAS-E when used for CASSCF energies (HF-E for HF, or SCF-E if used for both). Regarding E_{∞}^{CAS} , values from Equation 3 with the (6, 7) cardinal pair (15, 136, 137) and $\tilde{\beta} = 1.26$ (81) were used. Specific *x*-numbers were then obtained as averages of homologs so calculated for the 18-system set (89, 90, 108). Because they should be close to the associated cardinals (90, 108), x_k was constrained to minimize the RMSD between the two, with the search carried out (92) near $x_k = 7$ for the grid $1.2 \le \tilde{\beta} \le 2.0$ and $6.50 \le x_k \le 7.50$. The results are in **Supplemental Table 4**. Notably, the SCF-E schemes are more reliable than Karton & Martin's (81) protocols even with the smallest basis sets x = d, t, q.

Supplemental Material

9. CONCLUDING REMARKS

It is impractical to saturate a basis in ab initio calculations, especially for large molecules. With direct dynamics pointing toward the future in the simulation of chemical reactions (138), a double partition of the energy followed by CBS extrapolation from low hierachical bases (plus derivatives) may offer a route to on-the-fly integration of the equations of nuclear motion or may even help to discover chemistry with an ab initio nanoreactor (139).

Naturally, CBS extrapolation can hardly correct any major handicaps of the methods themselves. For example, unbalanced basis sets originate artifacts that cannot conceivably be eliminated, although they may likely be minimized. The same holds for properties, such as polarizabilities, that depend mainly on the wave function tail and are thus energetically unimportant. Such cautionary remarks extend also to systems requiring diffuse functions, which are to be considered whenever loosely bound electrons are present (anions and excited states). Indeed, USTE(d, t), USPE(t), and often even USPE(d) show enhanced results that may explain the optimistic view that CBS extrapolation helps to minimize basis set unbalance. Nothing can be said, though, about the wave function, with ab initio methods standing per se, especially of the R12 type. Showing significantly faster convergence, extrapolations of the energy are still doable for the latter, and Equation 8 or CS may help in the endeavor.

Finally, there is the question of whether CBS extrapolation can have an impact on large molecules. The consensus view is that (often heavily) parameterized Kohn–Sham DFT is the way, and hence any attempt to point to a wave function–based alternative may look dated. With most functionals calibrated so that calculated energies lie as close to known targets as possible, this

does not imply that they yield electron densities close to the exact one (140). In turn, a comparison with experimental data requires at least CBS extrapolation of any raw energies. Taken together, MP2/CBS(d, t) is likely a scheme capable of competing with Kohn–Sham DFT in terms of both time and accuracy (141). This is an issue worthy of further exploration.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am grateful for the contributions of all my collaborators whose names appear as coauthors of work cited here. The financial support over the years from Fundação para a Ciência e a Tecnologia, Portugal, is appreciated, as is the support of the Coimbra Chemistry Centre. Also acknowledged is support from COST Action CM1405.

LITERATURE CITED

- 1. Helgaker T, Jørgensen P, Olsen J. 2000. Molecular Electronic-Structure Theory. Chichester, UK: Wiley
- Hättig C, Klopper W, Köhn A, Tew DP. 2012. Explicitly correlated electrons in molecules. *Chem. Rev.* 112:4–74
- Feller D, Peterson KA, Crawford TD. 2006. Sources of error in electronic structure calculations on small chemical systems. *J. Chem. Phys.* 124:054107
- 4. Murrell JN, Carter S, Farantos SC, Huxley P, Varandas AJC. 1984. *Molecular Potential Energy Functions*. Chichester, UK: Wiley
- Varandas AJC, Galvão BLR. 2014. Exploring the utility of many-body expansions: a consistent set of accurate potentials for the lowest quartet and doublet states of the azide radical with revisited dynamics. *J. Phys. Chem. A* 118:10127–33
- 6. Varandas AJC. 1988. Intermolecular and intramolecular potentials: topographical aspects, calculation, and functional representation via a DMBE expansion method. *Adv. Chem. Phys.* 74:255–338
- Varandas AJC. 2013. Putting together the pieces: a global description of valence and long-range forces via combined hyperbolic inverse power representation of the potential energy surface. In *Reaction Rate Constant Computations: Theories and Applications*, ed. KL Han, TS Chu, pp. 408–45. London: R. Soc. Chem.
- 8. Kato T. 1957. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* 10:151–77
- 9. Kutzelnigg W, Klopper W. 1991. Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory. *7. Chem. Phys.* 94:1985–2001
- 10. Pack RT, Brown WB. 1966. Cusp conditions for molecular wavefunctions. J. Chem. Phys. 45:556-59
- Kutzelnigg W, Morgan JD. 1992. Rates of convergence of the partial-wave expansions of atomic correlation energies. *J. Chem. Phys.* 96:4484–508
- Flores JR, Slupski R, Jankowski K. 2006. Toward benchmark second-order correlation energies for large atoms. II. Angular extrapolation problems. *J. Chem. Phys.* 124:104107
- Schwartz C. 1962. Importance of angular correlations between atomic electrons. *Phys. Rev.* 126:1015– 19
- 14. Schwenke DW. 2012. On one-electron basis set extrapolation of atomic and molecular correlation energies. *Mol. Phys.* 110:2557–67
- 15. Dunning TH Jr. 1989. Gaussian basis sets for use in correlated molecular calculations. The atoms boron through neon and hydrogen. *J. Chem. Phys.* 90:1007–23

- Dunning TH Jr., Peterson KA, Woon DE. 1998. Basis sets: correlation consistent sets. In *Encyclopedia of Computational Chemistry*, ed. PvR Schleyer, NL Allinger, T Clark, J Gasteiger, PA Kolman, et al., pp. 88–115. Chichester, UK: Wiley
- Weigend F, Ahlrichs R. 2005. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7:3297–305
- Neto AC, Muniz EP, Centoducatte R, Jorge FE. 2005. Gaussian basis sets for correlated wave functions. Hydrogen, helium, first- and second-row atoms. *J. Mol. Struct.* 718:219–24
- Zhong S, Barnes EC, Petersson GA. 2008. Uniformly convergent n-tuple-ζ augmented polarized (nZaP) basis sets for complete basis set extrapolations. I. Self-consistent field energies. J. Chem. Phys. 129:184116
- Ranasinghe DS, Petersson GA. 2013. CCSD(T)/CBS atomic and molecular benchmarks for H through Ar. J. Chem. Phys. 138:144104
- 21. Jensen F. 2012. Atomic orbital basis sets. WIREs Comput. Mol. Sci. 3:273-95
- Dunning TH Jr. 2000. A road map for the calculation of molecular binding energies. J. Phys. Chem. A 104:9062–80
- Helgaker T, Klopper W, Koch H, Noga J. 1997. Basis-set convergence of correlated calculations on water. J. Chem. Phys. 106:9639–46
- Halkier A, Helgaker T, Jørgensen P, Klopper W, Koch H, et al. 1998. Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* 286:243–52
- Raghavachari K, Anderson JB. 1996. Electron correlation effects in molecules. J. Phys. Chem. 100:12960– 73
- 26. Varandas AJC. 2000. Basis-set extrapolation of the correlation energy. J. Chem. Phys. 113:8880-87
- Kutzelnigg W. 1994. Theory of the expansion of wave functions in a Gaussian basis. Int. J. Quantum Chem. 51:447–63
- Papajak E, Truhlar DG. 2012. What are the most efficient basis set strategies for correlated wave function calculations of reaction energies and barrier heights? J. Chem. Phys. 137:064110
- Feller D. 2013. Benchmarks of improved complete basis set extrapolation schemes designed for standard CCSD(T) atomization energies. *J. Chem. Phys.* 138:074103
- Vasilyev V. 2017. Online complete basis set limit extrapolation calculator. Comput. Theor. Chem. 1115:1– 3
- Varandas AJC, Piecuch P. 2006. Extrapolating potential energy surfaces by scaling electron correlation at a single geometry. *Chem. Phys. Lett.* 430:448–53
- Varandas AJC. 2007. Accurate global ab initio potentials at low-cost by correlation scaling and extrapolation to the one-electron basis set limit. *Chem. Phys. Lett.* 443:398–407
- Gordon MS, Nguyen KA, Truhlar DG. 1989. Parameters for scaling the correlation-energy of the bonds Si–H, P–H, S–H, and Cl–H and application to the reaction of silyl radical with silane. *J. Phys. Chem.* 93:7356–58
- Brown FB, Truhlar DG. 1985. A new semi-empirical method of correcting large-scale configuration interaction calculations for incomplete dynamic correlation of electrons. *Chem. Phys. Lett.* 117:307– 13
- Varandas AJC. 1989. A semiempirical method for correcting configuration interaction potential energy surfaces. *J. Chem. Phys.* 90:4379–91
- Varandas AJC. 2007. Accurate ab initio potentials at low cost via correlation scaling and extrapolation: application to CO(A¹Π). *J. Chem. Phys.* 127:114316
- Lutz JJ, Piecuch P. 2008. Extrapolating potential energy surfaces by scaling electron correlation: isomerization of bicyclobutane to butadiene. *J. Chem. Phys.* 128:154116
- Varandas AJC. 2007. Accurate ab initio-based molecular potentials: from extrapolation methods to global modelling. *Phys. Scr.* 76:C28–35
- Scheiner S. 2012. Extrapolation to the complete basis set limit for binding energies of noncovalent interactions. *Comput. Theor. Chem.* 998:9–13
- Šponer J, Jurečka P, Hobza P. 2006. Base stacking and base pairing. In *Computational Studies of RNA and DNA*, ed. J Šponer, F Lankaš, pp. 343–88. Dordrecht, Neth.: Springer

- Haldar S, Gnanasekarana R, Hobza P. 2015. A comparison of ab initio quantum-mechanical and experimental D₀ binding energies of eleven H-bonded and eleven dispersion-bound complexes. *Phys. Chem. Chem. Phys.* 17:26645–52
- Neves RPP, Fernandes PA, Varandas AJC, Ramos MJ. 2014. Benchmarking of density functionals for the accurate description of thioldisulfide exchange. *J. Chem. Theory Comput.* 10:4842–56
- Boys F, Bernardi F. 1970. The calculation of small molecular interactions by the differences of separate total energy. Some procedures with reduced error. *Mol. Phys.* 19:553–66
- 44. Wells BH, Wilson S. 1983. Van der Waals interaction potentials—many-body basis set supperposition errors. *Chem. Phys. Lett.* 101:429–34
- 45. White JC, Davidson ER. 1990. An analysis of the hydrogen-bond in ice. J. Chem. Phys. 93:8029-35
- Valiron P, Mayer I. 1997. Hierarchy of counterpoise corrections for N-body clusters: generalization of the Boys–Bernardi scheme. Chem. Phys. Lett. 275:46–55
- 47. Varandas AJC. 2008. Can extrapolation to the basis set limit be an alternative to the counterpoise correction? A study on the helium dimer. *Theor. Chem. Acc.* 119:511–21
- Varandas AJC. 2010. Extrapolation to the complete basis set limit without counterpoise. The pair potential of helium revisited. *J. Phys. Chem. A* 114:8505–16
- Varandas AJC. 2010. Spin-component-scaling second-order Møller–Plesset theory and its variants for economical correlation energies: unified theoretical interpretation and use for quartet N₃. *J. Chem. Phys.* 133:064104
- Varandas AJC. 2010. Helium-fullerene pair interactions: an *ab initio* study by perturbation-theory and coupled-cluster methods. *Int. J. Quantum Chem.* 111:416–29
- Song YZ, Varandas AJC. 2011. Accurate double many-body expansion potential energy surface for ground-state HS₂ based on ab initio data extrapolated to the complete basis set limit. *J. Phys. Chem. A* 115:5274–83
- Varandas AJC. 2012. Ab initio treatment of bond-breaking reactions: accurate course of HO₃ dissociation and revisit to isomerization. *J. Chem. Theory Comput.* 8:428–41
- 53. Srivastava S, Sathyamurthy N, Varandas AJC. 2012. An accurate ab initio potential energy curve and the vibrational bound states of $X^2 \Sigma^+_u$ state of H_2^- . *Chem. Phys.* 398:160–67
- Galvão BRL, Varandas AJC. 2013. Accurate study of the two lowest singlet states of HN₃: stationary structures and energetics at the MRCI complete basis set limit. *J. Phys. Chem. A* 117:4044–50
- 55. Varandas AJC. 2013. Accurate determination of the reaction course in HY₂ ⇒ Y + YH(Y = O, S): a detailed analysis of the covalent- to hydrogen-bonding transition. *J. Phys. Chem. A* 117:7393–407
- Varandas AJC. 2014. On carbon dioxide capture: an accurate *ab initio* study of the Li₃N + CO₂ insertion reaction. *Comput. Theor. Chem.* 1036:61–71
- 57. Varandas AJC. 2014. Is HO₃⁻ multiple-minimum and floppy? Covalent to van der Waals isomerization and bond rupture of a peculiar anion. *Phys. Chem. Chem. Phys.* 16:16997–7007
- Bytautas L, Ruedenberg K. 2008. Correlation energy and dispersion interaction in the ab initio potential energy curve of the neon dimer. *J. Chem. Phys.* 128:214308
- Alvarez-Idaboy JR, Galano A. 2010. Counterpoise corrected interaction energies are not systematically better than uncorrected ones: comparison with CCSD(T) CBS extrapolated values. *Theor. Chem. Acc.* 126:75–85
- Sheng XW, Mentel L, Gritsenko OV, Baerends EJ. 2011. Counterpoise correction is not useful for short and van der Waals distances but may be useful at long range. *J. Comput. Chem.* 32:2896–901
- Mentel LM, Baerends EJ. 2014. Can the counterpoise correction for basis set superposition effect be justified? J. Chem. Theory Comput. 10:252–67
- 62. Sheng X, Hu F, Qian S. 2016. On the complete basis set extrapolation procedures for the interaction energies. *Comput. Theor. Chem.* 1102:1–4
- Deb DK, Sarkar B. 2017. Theoretical investigation of gas-phase molecular complex formation between 2-hydroxy thiophenol and a water molecule. *Phys. Chem. Chem. Phys.* 19:2466–78
- Spackman PR, Karton A. 2015. Estimating the CCSD basis-set limit energy from small basis sets: basisset extrapolations versus additivity schemes. *AIP Adv.* 5:057148
- 65. Boese AD, Oren M, Atasoylu O, Martin JML, Kállay M, Gauss J. 2004. W3 theory: robust computational thermochemistry in the kJ/mol accuracy range. *J. Chem. Phys.* 120:4129–41

- Karton A, Martin JML. 2012. Explicitly correlated Wn theory: W1-F12 and W2-F12. J. Chem. Phys 136:124114
- Tarczay GT, Császár AG, Klopper W, Szalay V, Allen WD, Schaefer HF III. 1999. The barrier to linearity of H₂O. *J. Chem. Phys.* 110:11971–81
- Curtiss LA, Raghavachari K, Trucks GW, Pople JA. 1991. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* 94:7221–30
- Klopper W, Kutzelnigg W. 1986. Gaussian basis sets and the nuclear cusp problem. J. Mol. Struct. 135:339-56
- 70. Peterson KA, Woon DE, Dunning TH Jr. 1994. Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the $H + H_2 \rightarrow H_2 + H$ reaction. J. Chem. Phys. 100:7410–15
- 71. Truhlar DG. 1998. Basis-set extrapolation. Chem. Phys. Lett. 294:45-48
- Halkier A, Helgaker T, Jorgensen P, Klopper W, Olsen J. 1999. Basis-set convergence of the energy in molecular Hartree–Fock calculations. *Chem. Phys. Lett.* 302:437–46
- Fast PL, Sanchez ML, Truhlar DG. 1999. Infinite basis limits in electronic structure theory. J. Chem. Phys. 111:2921–26
- Halkier A, Klopper W, Helgaker T, Jørgensen P, Taylor PR. 1999. Basis set convergence of the interaction energy of hydrogen-bonded complexes. *J. Chem. Phys.* 111:9157–67
- Lee JS, Park SY. 2000. Basis set convergence of correlated calculations on He, H₂, and He₂. *J. Chem. Phys.* 112:10746–53
- 76. Klopper W. 2001. A critical note on extrapolating helium pair potentials. J. Chem. Phys. 115:761-65
- Klopper W. 2001. Highly accurate coupled-cluster singlet and triplet pair energies from explicitly correlated calculations in comparison with extrapolation techniques. *Mol. Phys.* 99:481–507
- Huh SB, Lee JS. 2003. Basis set and correlation dependent extrapolation of correlation energy. *J. Chem. Phys.* 118:3035–42
- Schwenke DW. 2005. The extrapolation of one-electron basis sets in electronic structure calculations: how it should work and how it can be made to work. J. Chem. Phys. 122:014107
- Jensen F. 2005. Estimating the Hartree–Fock limit from finite basis set calculations. *Theor. Chem. Acc.* 113:267–73
- Karton A, Martin JML. 2006. Comment on "Estimating the Hartree–Fock limit from finite basis set calculations". *Theor. Chem. Acc.* 115:330–33
- Varandas AJC. 2007. Extrapolation to the one-electron basis-set in electronic structure calculations. *J. Chem. Phys.* 126:244105
- Junqueira GMA, Varandas AJC. 2008. Extrapolating to the one-electron basis set limit in polarizability calculations. *J. Phys. Chem. A* 112:10413–19
- Bakowies D. 2007. Extrapolation of electron correlation energies to finite and complete basis set targets. *J. Chem. Phys.* 127:084105
- Bakowies D. 2007. Accurate extrapolation of electron correlation energies from small basis sets. J. Chem. Phys. 127:164109
- Varandas AJC. 2008. Generalized uniform singlet- and triplet-pair extrapolation of the correlation energy to the one electron basis set limit. *J. Phys. Chem. A* 112:1841–50
- Williams TG, DeYonker NJ, Wilson AK. 2008. Hartree–Fock complete basis set limit properties for transition metal diatomics. *J. Chem. Phys.* 128:044101
- Varandas AJC. 2009. Møller–Plesset perturbation energies and distances for HeC₂₀ extrapolated to the complete basis set limit. *J. Comp. Chem.* 30:379–88
- Pansini FNN, Varandas AJC. 2015. Toward a unified single-parameter extrapolation scheme for the correlation energy: systems formed by atoms of hydrogen through neon. *Chem. Phys. Lett.* 631–632:70– 77
- Pansini FNN, Neto AC, Varandas AJC. 2015. On the performance of various hierarchized bases in extrapolating the correlation energy to the complete basis set limit. *Chem. Phys. Lett.* 641:90–96
- Okoshi M, Atsumi T, Nakai H. 2015. Revisiting the extrapolation of correlation energies to complete basis set limit. *J. Comput. Chem.* 36:1075–82

- 92. Pansini FNN, Neto AC, Varandas AJC. 2016. Extrapolation of Hartree–Fock and multiconfiguration self-consistent-field energies to the complete basis set limit. *Theor. Chem. Acc.* 135:261
- Seino J, Nakai H. 2016. Informatics-based energy fitting scheme for correlation energy at complete basis set limit. *J. Comput. Chem.* 37:2304–15
- 94. Peterson KA. 2007. Gaussian basis sets exhibiting systematic convergence to the complete basis set limit. In Annual Reports in Computational Chemistry, Vol. 3, ed. DC Spellmeyer, R Wheeler, pp. 195–206. Amsterdam: Elsevier
- Hepburn J, Scoles G, Penco R. 1975. A simple but reliable method for the prediction of intermolecular potentials. *Chem. Phys. Lett.* 36:451–56
- Murrell JN, Varandas AJC. 1975. Perturbation calculations of rare-gas potentials near the van der Waals minimum. *Mol. Phys.* 30:223–36
- Feller D. 1992. Application of systematic sequences of wave functions to the water dimer. J. Chem. Phys. 96:6104–14
- Bunge CF. 1970. Electronic wave functions for atoms. 2. Some aspects of the convergence of the configuration interaction expansion for ground states of He isoelectronic series. *Theor. Chim. Acta* 16:126– 44
- Hill RN. 1985. Rates of convergence and error estimation formulas for the Rayleigh–Ritz variational method. *J. Chem. Phys.* 83:1173–96
- Martin JML, Taylor PR. 1997. Benchmark quality total atomization energies of small polyatomic molecules. *J. Chem. Phys.* 106:8620–23
- 101. Valeev EF, Allen WD, Hernandez R, Sherrill CD, Schaefer HF. 2003. On the accuracy limits of orbital expansion methods: explicit effects of k-functions on atomic and molecular energies. J. Chem. Phys. 118:8594
- Klopper W, Samson CCM. 2002. Explicitly correlated second-order Møller–Plesset methods with auxiliary basis sets. J. Chem. Phys. 116:6397–410
- 103. Bytautas L, Ruedenberg K. 2005. Correlation energy extrapolation by intrinsic scaling. IV. Accurate binding energies of the homonuclear diatomic molecules carbon, nitrogen, oxygen and fluorine. *J. Chem. Phys.* 122:154110
- 104. Bak KL, Halkier A, Jorgensen P, Olsen J, Helgaker T, Klopper W. 2001. Chemical accuracy from 'Coulomb hole' extrapolated molecular quantum-mechanical calculations. J. Mol. Struct. 567:375– 84
- Parthiban S, Martin JML. 2001. Fully ab initio atomization energy of benzene via Weizmann-2 theory. J. Chem. Phys. 114:2051–54
- Klopper W, Noga J. 2003. Accurate quantum-chemical prediction of enthalpies of formation of small molecules in the gas phase. *ChemPhysChem* 4:32–48
- Helgaker T, Ruden TA, Jorgensen P, Olsen J, Klopper W. 2004. A priori calculation of molecular properties to chemical accuracy. *J. Phys. Org. Chem.* 17:913–33
- Varandas AJC, Pansini FNN. 2014. Narrowing the error in electron correlation calculations by basis set re-hierarchization and use of the unified singlet and triplet electron-pair extrapolation scheme: application to a test set of 106 systems. *J. Chem. Phys.* 141:224113
- Klopper W, Bak KL, Jørgensen P, Olsen J, Helgaker T. 1999. Highly accurate calculations of molecular electronic structure. *J. Phys. B* 32:R103
- Pansini FNN, Neto AC, Varandas AJC. 2015. Application of the unified singlet and triplet electron-pair extrapolation scheme with basis set rehierarchization to tensorial properties. *J. Phys. Chem. A* 119:1208– 17
- Huh SB, Lee JS. 2003. Basis set and correlation dependent extrapolation of correlation energy. *J. Chem. Phys.* 118:3035–42
- 112. Englert BG, Schwinger J. 1985. Atomic binding-energy oscillations. Phys. Rev. A 32:47-63
- 113. Kais S, Sung S, Herschbach DR. 1994. Large-Z and -N dependence of atomic energies from renormalization of the large-dimension limit. Int. J. Quant. Chem. 49:657–74
- 114. Werner HJ, Knowles PJ, Lindh R, Manby FR, Schütz M, et al. 2010. *Molpro v2010.1*. Comput. Softw. http://www.molpro.net

- Wood GPF, Radom L, Petersson GA, Barnes EC, Frisch MJ, Montgomery JA Jr. 2006. A restrictedopen-shell complete-basis-set model chemistry. *J. Chem. Phys.* 125:094106
- Fabiano E, Della Sala F, Grabowski I. 2015. Accurate non-covalent interaction energies via an efficient MP2 scaling procedure. *Chem. Phys. Lett.* 635:262–67
- Klopper W, Rafal AB, Hättig C, Tew DP. 2010. Accurate computational thermochemistry from explicitly correlated coupled-cluster theory. *Theor. Chem. Acc.* 126:289–304
- Hill JG, Peterson KA, Knizia G, Werner HJ. 2009. Assessment of basis sets for F12 explicitly-correlated molecular electronic-structure methods. *J. Chem. Phys.* 131:194105
- Bischoff FA, Wolfsegger S, Tew DP, Klopper W. 2009. Assessment of basis sets for F12 explicitlycorrelated molecular electronic-structure methods. *Mol. Phys.* 107:963–75
- 120. Balabin RM. 2010. Intramolecular basis set superposition error as a measure of basis set incompleteness: Can one reach the basis set limit without extrapolation? *J. Chem. Phys.* 132:211103
- 121. Xantheas SS. 1996. On the importance of the fragment relaxation energy terms in the estimation of the basis set superposition error correction to the intermolecular interaction energy. *J. Chem. Phys.* 104:8821–23
- Kalescky R, Kraka E, Cremer D. 2014. Accurate determination of the binding energy of the formic acid dimer: the importance of geometry relaxation. *J. Chem. Phys.* 140:084315
- 123. Miliordos E, Xantheas SS. 2015. On the validity of the basis set superposition error and complete basis set limit extrapolations for the binding energy of the formic acid dimer. J. Chem. Phys. 142:094311
- Buckingham AD. 1967. Permanent and induced molecular moments and long-range intermolecular forces. Adv. Chem. Phys. 12:107–42
- 125. Kobus J, Moncrieff D, Wilson S. 2007. Comparison of the polarizabilities and hyperpolarizabilities obtained from finite basis set and finite difference Hartree–Fock calculations for diatomic molecules: III. The ground states of N₂, CO and BF. *J. Phys. B* 40:877–96
- 126. Ral F, Vallet V, Clavagura C, Dognon JP. 2008. In silico prediction of atomic static electric-dipole polarizabilities of the early tetravalent actinide ions: Th⁴⁺ (5f⁰), Pa⁴⁺ (5f¹), and U⁴⁺ (5f²). *Phys. Rev. A* 78:052502
- Montena R, Hajgatóa B, Deleuze MS. 2011. Many-body calculations of molecular electric polarizabilities in asymptotically complete basis sets. *Mol. Phys.* 109:2317–39
- 128. Ruscic B, Wagner AF, Harding LB, Asher RL, Feller D, et al. 2002. On the enthalpy of formation of hydroxyl radical and gas-phase bond dissociation energies of water and hydroxyl. *J. Phys. Chem. A* 106:2727–47
- Peterson KA, Feller D, Dixon DA. 2012. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. *Theor. Chem. Acc.* 131:1079
- Coriani S, Helgaker T, Jørgensen P, Klopper W. 2004. A closed-shell coupled-cluster treatment of the Breit–Pauli first-order relativistic energy correction. *J. Chem. Phys.* 121:6591–98
- Sansonetti JE, Martin WC. 2005. Handbook of basic atomic spectroscopic data. J. Phys. Chem. Ref. Data 34:1559–2259
- Klopper W, Ruscic B, Tew DP, Bischoff FA, Wolfsegger S. 2009. Atomization energies from coupledcluster calculations augmented with explicitly-correlated perturbation theory. *Chem. Phys.* 356:14–24
- 133. Ruscic B, Pinzon RE, Morton ML, Srinivasan NK, Su MC, et al. 2006. Active thermochemical tables: accurate enthalpy of formation of hydroperoxyl radical, HO₂. *J. Phys. Chem. A* 110:6592–601
- 134. Pansini FNN, Varandas AJC. 2015. Toward a unified single-parameter extrapolation scheme for the correlation energy: systems formed by first- and second-row atoms. *Chem. Phys. Lett.* 631–632:70–77
- Okoshi M, Atsumi T, Nakai H. 2015. Revisiting the extrapolation of correlation energies to complete basis set limit. *J. Comput. Chem.* 36:1075–82
- Feller D, Peterson KA. 1999. Re-examination of atomization energies for the Gaussian-2 set of molecules. *J. Chem. Phys.* 110:8384–96
- Feller D, Sordo JA. 2000. Performance of CCSDT for diatomic dissociation energies. J. Chem. Phys. 113:485–93
- Paranjothy M, Sun R, Zhuang Y, Hase WL. 2013. Direct chemical dynamics simulations: coupling of classical and quasiclassical trajectories with electronic structure theory. WIREs Comput. Mol. Sci. 3:296– 316

- 139. Wang LP, Titov A, McGibbon R, Liu P, Panda VS, Martínez TJ. 2014. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* 6:1044–48
- 140. Medvedev MG, Bushmarinov IS, Sun J, Perdew JP, Lyssenko KA. 2017. Density functional theory is straying from the path toward the exact functional. *Science* 355:49–52
- 141. Varandas AJC, González MM, Montero-Cabrera LA, de la Vega JMG. 2017. Assessing how correlated molecular orbital calculations can perform versus Kohn–Sham DFT: barrier heights/isomerizations. *Chem. Eur. J.* 23:9122–9129