

# Factorized Quadruples and a Predictor of Higher-Level Correlation in Thermochemistry

James H. Thorpe,<sup>†</sup> Zachary W. Windom,<sup>‡</sup> Rodney J. Bartlett,<sup>\*,‡</sup> and Devin A.

Matthews<sup>\*,†</sup>

<sup>†</sup>*Department of Chemistry, Southern Methodist University, Dallas, TX 75275, USA*

<sup>‡</sup>*Quantum Theory Project, University of Florida, Gainesville, Florida 32611, USA*

E-mail: bartlett@qtp.ufl.edu; damatthews@smu.edu

## Abstract

Coupled cluster theory has had a momentous impact on the ab initio prediction of molecular properties, and remains a staple ingratiate in high-accuracy thermochemical model chemistries. However, these methods require inclusion of at least some connected quadruple excitations, which generally scale at best as  $\mathcal{O}(N^9)$  with the number of basis functions. It is very difficult to predict, a priori, the effect correlation past CCSD(T) has on a given reaction energy. The purpose of this work is to examine cost-effective quadruple corrections based on the factorization theorem of many-body perturbation theory that may address these challenges. We show that the  $\mathcal{O}(N^7)$ , factorized CCSD(TQ<sub>f</sub>) method introduces minimal error to predicted correlation and reaction energies as compared to the  $\mathcal{O}(N^9)$  CCSD(TQ). Further, we examine the performance of Goodson’s continued fraction method in the estimation of CCSDT(Q)<sub>Λ</sub> contributions to reaction energies, as well as a “new” method related to %TAE[(T)] that we refer to as a scaled perturbation estimator. We find that the scaled perturbation estimator based upon CCSD(TQ<sub>f</sub>)/cc-pVDZ is capable of predicting CCSDT(Q)<sub>Λ</sub>/cc-pVDZ contributions to

reaction energies with an average error of  $0.07 \text{ kcal mol}^{-1}$  and an  $L_2D$  of  $0.52 \text{ kcal mol}^{-1}$  when applied to a test-suite of nearly 3000 reactions. This offers a means by which to reliably “ballpark” how important post-CCSD(T) contributions are to reaction energies while incurring no more than CCSD(T) formal cost and a little mental math.

# Introduction

Coupled cluster (CC) theory provides a systematically improvable route toward capturing the instantaneous electron correlation of small to medium-sized molecules. By modulating the maximal rank of the cluster operator according to calculation affordability and/or the desired accuracy, a hierarchy of CC methods are available albeit using algorithms that scale as increasingly high-order polynomials. This feature of CC theory has made it the method of choice for so-called composite model chemistries, which describe procedures by which to predict experimental measurements of molecular properties such as bond energies or vibrational frequencies to a given degree of accuracy/precision/trueness via computed, or easily obtained, values.<sup>1</sup> The exact details of which coupled-cluster contributions are included with which basis sets depends on the goals of the model chemistry in question. High-accuracy model chemistries, which seek to determine molecular enthalpies of formation to within 1 kJ mol<sup>-1</sup>, such as HEAT,<sup>2-5</sup>  $W_n$ ,<sup>6-11</sup> FPA,<sup>12-16</sup> FPD,<sup>17-23</sup> and others (see Ref. 1), generally include not just CCSD(T)<sup>24,25</sup> with a sizeable basis set, but also some amount of instantaneous four-electron correlation via CCSDT(Q),<sup>26-28</sup> CCSDTQ,<sup>29</sup> or, more recently, CCSDT(Q)<sub>Λ</sub>,<sup>28,30</sup> typically in a cc-pVDZ basis set. On the other hand, model chemistries that seek chemical accuracy (one kcal mol<sup>-1</sup> within experimental measurements), such as G- $n$ <sup>31-35</sup> and CBS,<sup>36-41</sup> among others,<sup>1</sup> generally stop at CCSD(T) with a relatively small basis set, as they target systems for which CCSD(T) itself is already an expensive calculation. While it is recognized that the neglect of post-(T) corrections, often referred to as higher-level correlation (HLC) corrections, in multireference systems is dangerous, it is also true that the absence of these terms in a model chemistry aiming for chemical accuracy introduces errors that, even for well behaved species, may put the goal of sub-kcal mol<sup>-1</sup> accuracy in danger when combined with other approximations that must be made to reduce cost. Further, it is difficult to tell, quantitatively, when this will be the case a priori. The purpose of this work is to investigate some routes towards estimating the size of this post-CCSD(T), non-relativistic electron correlation, without incurring any increase in formal cost

beyond  $\mathcal{O}(N^7)$  already expended in CCSD(T).

This effort naturally begins with perturbation theory. Initial attempts to increase the accuracy of CCSD<sup>24</sup> centered on accounting for the effect of the  $T_3$  triple electron excitation operator, either perturbatively<sup>24,25,42–45</sup> or iteratively.<sup>46–48</sup> These efforts eventually culminated in the  $\mathcal{O}(N^8)$  CCSDT method.<sup>49</sup> Pursuit of increasingly accurate calculations lead to similar attempts at incorporating quadruple electron excitations into the underlying CC framework.<sup>26–28,50,51</sup> The groundwork for perturbative treatment of quadruple excitation operators like  $[Q]$ ,  $(Q)$ , and  $(Q)_\Lambda$  was originally postulated by Kucharski and Bartlett,<sup>27,30</sup> which naturally lead to the  $\mathcal{O}(N^{10})$  CCSDTQ method.<sup>29,52</sup> A later derivation by Bomble and Stanton used Löwdin’s partitioning on a CCSDT reference to derive the versions of CCSDT(Q) and CCSDT(Q) $_\Lambda$  that have been employed in high-accuracy thermochemistry, and resolved ordering issues similar to those that plague the definitions and performance of +T, [T], and (T).<sup>3,28</sup> Of particular note is that perturbative triple corrections like [T] or (T), and perturbative quadruple corrections like [Q] or (Q), are  $\mathcal{O}(N)$  cheaper than their counterparts that are complete through a given cluster operator rank. An important consequence of this reduced scaling is the capability not only to address larger molecular systems, but also to employ larger atomic basis sets in higher-level correlation calculations of small systems when exceedingly high accuracy is called for.<sup>53,54</sup>

However, in spite of the order  $\mathcal{O}(N)$  in savings afforded by using methods like CCSDT(Q) $_\Lambda$  (which scales as  $\mathcal{O}(N^9)$ ) as compared to full CCSDTQ, this is a far cry from the  $\mathcal{O}(N^7)$  cost of CCSD(T). The first step to reduce this cost is to remove the need to perform the iterative CCSDT component of these methods (an  $\mathcal{O}(N^8)$  procedure), typically by replacing the converged  $T_3$  CC amplitudes with a denominator-weighted contraction of  $T_2$  with the Hamiltonian, as is done in CCSD(T). The remaining cost then arises from contractions with the Fock denominator,  $D_4$ —in this case, a 8-index tensor—used to construct an approximation to the  $T_4$  operator. These methods are typically referred to as CCSD(TQ).<sup>30</sup> A reasonable way to reduce the algorithmic scaling, and the central point of this manuscript, focuses on

eliminating the 8-index,  $T_4$  denominator by invoking the factorization theorem of MBPT.<sup>55,56</sup> This allows for the elimination of the  $D_4$  denominator in exchange for a product of two, 4-index denominators (e.g.  $D_2^A D_2^B$ ), which is the idea used to define the  $Q_f$  correction<sup>57–59</sup> that facilitates an estimate of connected quadruples excitations at  $\mathcal{O}(N^7)$  cost. Initial studies of these factorized methods focused on a small selection of molecules, with the general conclusion being that negligible error was introduced into the calculation by adopting the  $Q_f$  method over analogous, but more expensive quadruple corrections.<sup>30,42,60–63</sup> Later benchmarks studied the performance of a variety of methods in the prediction of ground state correlation energies.<sup>64,65</sup> However, a robust analysis of these factorized approaches centered in the realm of accurate thermochemistry remains largely unexplored.

After establishing the accuracy and characteristics of this factorization approximation, this work will investigate how a low-order approximation of the importance of the  $T_4$  operator might be correlated with the size of post-CCSD(T) contributions of the more complete CCSDT(Q) <sub>$\Lambda$</sub>  method,<sup>27,28,30</sup> which has recently been demonstrated as a very accurate higher-level correlation approximation.<sup>11,54,66–69</sup> Although several indices have been proposed to indicate a strong deficiency in the zeroth-order approximation to the wavefunction<sup>70–84</sup>, which usually coincides with large HLC corrections, there are limited means by which to correlate these metrics with a quantitative estimate of post-CCSD(T) contributions in thermochemistry. This is further complicated by the fact that many multireference indices do not display a strong statistical correlation with the size of the higher-level correlation (HLC) terms in model chemistries, see Ref. 82 and references therein. A notable exception to this statement are several “energy-based” diagnostics such as Martin’s %TAE[(T)]<sup>8,66</sup> and DFT-based diagnostics.<sup>82,83</sup> Additionally, approaches based on Goodson’s continued fraction formalism have seen some success in the prediction of post-CCSDTQ correlation.<sup>85–87</sup>

The remainder of this manuscript is organized as follows. First, we briefly introduce the theory behind the CCSD(T)-scaling methods that are the focal point of this work. We then describe how these methods are benchmarked against a subset of the W4.17 dataset,<sup>88</sup> and

how this dataset is expanded to include several thousand chemical reactions. We then use these test-suites to analyze the errors incurred by factorizing the quadruple excitation contributions of these CCSD(T)-scaling methods, along with the size of the correlation missing between these methods and CCSDT(Q)<sub>Λ</sub>, in the context of both raw correlation energies and reaction energies. Finally, we present a novel predictor of post-CCSD(T) higher-level correlation based upon these factorized quadruples methods that serves as a reliable guide as to when HLC effects need to be accounted for in a model chemistry.

# Theory

## Coupled Cluster Theory and Perturbative Quadruples Corrections

Coupled cluster theory is an exponential parametrization of the wavefunction

$$|\Psi\rangle = e^T |0\rangle \quad (1)$$

for some single reference Slater determinant  $|0\rangle$ . The cluster operator,  $T$ , is expressed in terms of a sum of individual operators responsible for single, double,  $\dots$ , up to  $n$ -fold electron excitations

$$T = T_1 + T_2 + \dots + T_n \quad (2)$$

In the limit where the cluster operator is left untruncated, the FCI wavefunction is rigorously recovered. Otherwise, a truncation point is chosen to include up to a maximal rank cluster operator,  $T_k$  defined as

$$T_k = (k!)^{-2} \sum t_{ij\dots}^{ab\dots} a^\dagger b^\dagger \dots ji \quad (3)$$

The definition for the normal-ordered Hamiltonian is

$$H_N = \sum_{pq} f_{pq} \{p^\dagger q\} + \frac{1}{4} \sum_{pqrs} \langle pq || rs \rangle \{p^\dagger q^\dagger sr\} \quad (4)$$

$$= f_N + W_N \quad (5)$$

which is expressed in terms of one and two body integrals. In this work we assume a canonical (Hartree–Fock) reference determinant such that  $f_{pq} = \delta_{pq}\epsilon_p$ , with  $\epsilon_p$  being the orbital energies.

As there is some ambiguity in the literature, we wish to define the pertinent quadruple

corrections that are the focal point of this study in the interest of clarity. We adopt the nomenclature [Q] to mean the energy correction associated with quadruple excitations correct through fifth-order in MBPT,<sup>42</sup> given by

$$\Delta E_{[Q]}^{[5]} = \langle 0 | T_2^\dagger W_N T_4^{[3]} | 0 \rangle \quad (6)$$

where the approximation to  $T_4$  is correct through third-order in MBPT and only includes connected (C) diagrams

$$T_4^{[3]} = R_4 \left( W_N \frac{T_2^2}{2} + W_N T_3 \right)_C \quad (7)$$

The resolvent operator,  $R_n$ , is defined as

$$R_n(X) = (n!)^{-2} \sum \frac{\langle \Phi_{ij\dots}^{ab\dots} | X | 0 \rangle}{\epsilon_i + \epsilon_j + \dots - \epsilon_b - \epsilon_a} a^\dagger b^\dagger \dots ji \quad (8)$$

and is used to enforce the projection of a pure excitation operator onto the proper subspace. For example, with the double excitation cluster operator  $T_2$  represented as a graphical fragment with external indices  $A = \{a, b, i, j\}$ , then  $R_2(T_2) = \frac{1}{4} \sum_{abij} \frac{t_{ij}^{ab}}{\epsilon_i + \epsilon_j - \epsilon_b - \epsilon_a} a^\dagger b^\dagger ji$ , which is the same operator with amplitudes divided by denominators  $D_2^A = \epsilon_i + \epsilon_j - \epsilon_b - \epsilon_a$ . Informally, we can also write the action of the resolvent as, for example,  $R_2(T_2) = D_2^{-1} T_2$ , and similarly, with a modified definition, the resolvent can be applied to de-excitation operators such as  $R_2(T_2)^\dagger = T_2^\dagger D_2^{-1}$ .

At this point, it is prudent to briefly review the tenets of the factorization theorem in MBPT. We first note that the first-order approximation of  $T_2$  is

$$T_2 \approx T_2^{(1)} = R_2(W_N) = D_2^{-1} W_N \quad (9)$$



This allows (6) to be rewritten through fourth order as

$$\langle 0 | W_N D_2^{-1} W_N T_4^{[3]} | 0 \rangle \quad (10)$$

The two  $W_N$  operators must be disconnected since they only contain double de-excitation components. Thus, the external indices spanned by the quadruple excitation fragment  $T_4^{(3)}$  can be split into two doubly-excited parts, as can the associated orbital energy denominator,  $D_4 = \epsilon_i + \epsilon_j + \epsilon_k + \epsilon_l - \epsilon_a - \epsilon_b - \epsilon_c - \epsilon_d = (\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b) + (\epsilon_k + \epsilon_l - \epsilon_c - \epsilon_d) = D_2^A + D_2^B$ . The index groups  $A$  and  $B$  are associated with the two  $W_N$  fragments. Due to the symmetry of  $T_4$ , however, we can associate the index groups to the fragments in either order. Finally, we introduce the identity as  $D_4^{-1} D_4$ . This leads directly to the factorization

$$\langle 0 | T_2^{(1)\dagger} W_N T_4^{[3]} | 0 \rangle = \frac{1}{2} \langle 0 | \left( W_N^A (D_2^A)^{-1} W_N^B D_4^{-1} D_4 T_4^{[3]} + \right. \quad (11)$$

$$\left. W_N^B (D_2^B)^{-1} W_N^A D_4^{-1} D_4 T_4^{[3]} \right)_C | 0 \rangle \quad (12)$$

$$= \frac{1}{2} \langle 0 | \left( W_N^A (D_2^A)^{-1} W_N^B (D_2^B)^{-1} D_4 T_4^{[3]} \right)_C | 0 \rangle \quad (13)$$

$$= \frac{1}{2} \langle 0 | T_2^{(1)\dagger} T_2^{(1)\dagger} \left( W_N \frac{T_2^2}{2} + W_N T_3 \right)_C | 0 \rangle \quad (14)$$

where the factor of  $1/2$  arises because of the two equivalent orderings, and we have used the fact that  $D_4 = D_2^A + D_2^B$ . From (7), the application of  $D_4$  to  $T_4^{(3)}$  “undoes” the four-electron resolvent and removes the denominator coupling of the eight external indices. Kucharski and Bartlett used this transformation as a basis to propose what they referred to as the  $(Q_f)$  correction,<sup>57,58</sup> which is given to be

$$E_{(Q_f)}^{[5]} = \frac{1}{2} \langle 0 | T_2^\dagger T_2^{(1)\dagger} \left( W_N \frac{T_2^2}{2} + W_N T_3 \right)_C | 0 \rangle. \quad (15)$$

where the reintroduction of the full  $T_2^\dagger$  provides additional higher-order terms. It should be noted that despite the notation  $(Q_f)$  is a factorized version of the  $[Q]$  (“bracket Q”) correction

in (6), rather than the (Q) (“parentheses Q”) correction described below. An overview of the factorization theorem as it pertains to coupled-cluster theory is discussed elsewhere.<sup>89</sup>

What we refer to as the (Q) correction includes an additional term to those of (6). This is shown to be<sup>27,28,30</sup>

$$\Delta E_{(Q)} = \Delta E_{[Q]} + \langle 0 | T_3^\dagger W_N T_4^{[3]} | 0 \rangle \quad (16)$$

with  $T_4^{[3]}$  given by (7). The distinction between these methods is shown diagrammatically in Fig. 1. Note that this extra term would show up as a sixth-order correction to the energy if taking an expectation value coupled cluster (XCC) viewpoint.<sup>50,90,91</sup> Additionally, this term is not factorizable—meaning there is no way to improve the  $\mathcal{O}(N^9)$  scaling of the second term in (16) which is shown diagrammatically in Fig. 1C. A final non-iterative quadruples correction is given by<sup>28</sup>

$$\Delta E_{(Q)\Lambda} = \langle 0 | \Lambda_2 W_N T_4^{[3]} | 0 \rangle + \langle 0 | \Lambda_3 W_N T_4^{[3]} | 0 \rangle \quad (17)$$

where  $\Lambda$  are the left-hand wavefunction components of the coupled cluster ground state.

The complete coupled cluster methods and total correlation energies defined by these various approximation, as used in this work, are CCSD(TQ) ( $E_{\text{CCSD(TQ)}} = E_{\text{CCSD(T)}} + \Delta E_{[Q]}$ ), CCSD(TQ<sub>f</sub>) ( $E_{\text{CCSD(TQ}_f\text{)}} = E_{\text{CCSD(T)}} + \Delta E_{(Q_f)}$ ), CCSDT[Q] ( $E_{\text{CCSDT[Q]}} = E_{\text{CCSDT}} + \Delta E_{[Q]}$ ), CCSDT(Q) ( $E_{\text{CCSDT(Q)}} = E_{\text{CCSDT}} + \Delta E_{(Q)}$ ), and CCSDT(Q)<sub>Λ</sub> ( $E_{\text{CCSDT(Q)}_\Lambda} = E_{\text{CCSDT}} + \Delta E_{(Q)\Lambda}$ ). All of the coupled cluster methods used here have been previously published, and we urge readers to review the original work for the full working equations and other details.<sup>26–28,30,42,50,57–63</sup>

## Higher-Level Correlation Predictors

This work attempts to provide a quantitative means by which to estimate the correlation energy missing at the conclusion of a CCSD(T) calculation. As recent work has demonstrated

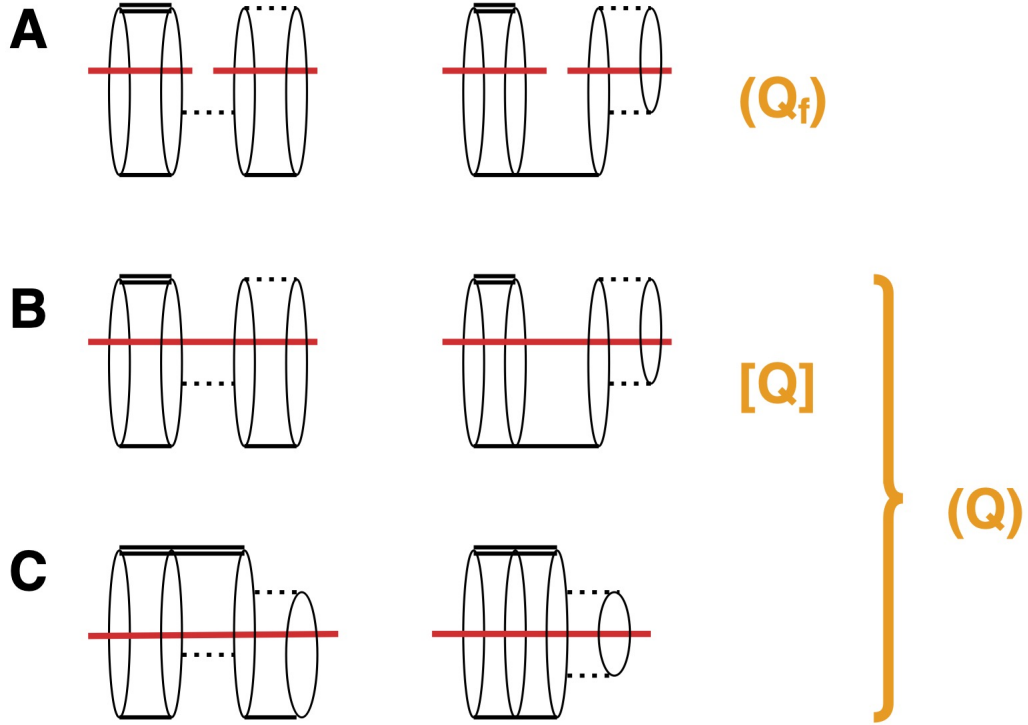


Figure 1: Comparison of the pertinent diagrams defining various, perturbative approximations targeting quadruple excitations. Red, horizontal lines indicate denominators. A) The two diagrams defining the  $(Q_f)$  approximation. B) The two diagrams defining the  $[Q]$  approximation. C) These two diagrams in addition to those in B) define the  $(Q)$  approximation. In the  $(Q)_\Lambda$  approximation the top  $T^\dagger$  vertices in  $(Q)$  are replaced by  $\Lambda$ .

the capabilities of  $\text{CCSDT}(\text{Q})_\Lambda$  to act as a second Pauling point in the CC expansion (the “platinum standard” in comparison to  $\text{CCSD}(\text{T})$ ’s “gold standard”), we will define the residual higher-level correlation (HLC) as  $\Delta E_{\text{HLC}} = E_{\text{CCSDT}(\text{Q})_\Lambda} - E_{\text{CCSD}(\text{T})}$ .<sup>11,53,54,68,69</sup> Previous diagnostics of the multi-reference nature of a given molecule are not generally correlated with a quantifiable prediction of the influence of HLC effects on, say, a bond energy. An exception to this, however, is the work of Goodson in 2002,<sup>85</sup> which was later extended by Schröder and coworkers in 2015<sup>92</sup> and made use of in some FPD model chemistries.<sup>87,93,94</sup>

Goodson used continued fractions and truncated Padé approximants to derive three formulas for predicting the post- $\text{CCSD}(\text{T})$  electron correlation energy  $\Delta E_{\text{HLC}}$ :

$$\Delta E_{cf} = \frac{E_{\text{CCSD}}^2}{E_0} + \frac{\Delta E_{(\text{T})}^2}{E_{\text{CCSD}}} + 2 \frac{E_{\text{CCSD}} \Delta E_{(\text{T})}}{E_0} \quad (18)$$

$$\Delta E_R = \frac{\Delta E_{(\text{T})}^2}{E_{\text{CCSD}}} + \frac{\Delta E_{(\text{T})}^3}{E_{\text{CCSD}}^2} \quad (19)$$

$$\Delta E_Q = 2 \frac{\Delta E_{(\text{T})}^2}{E_{\text{CCSD}}} + 5 \frac{\Delta E_{(\text{T})}^3}{E_{\text{CCSD}}^2} \quad (20)$$

where  $E_0$  is the SCF reference energy,  $E_{\text{CCSD}}$  is the CCSD correlation energy, and  $\Delta E_{(\text{T})}$  is the perturbative triples contribution from  $\text{CCSD}(\text{T})$ . We will refer to these methods, in order, as CF, R, and Q. In this work, we will also apply these approximations to  $\text{CCSD}(\text{TQ}_f)$ , in which case we define  $\Delta E_{(\text{TQ}_f)} \equiv E_{\text{CCSD}(\text{TQ}_f)} - E_{\text{CCSD}} = \Delta E_{(\text{T})} + \Delta E_{(\text{Q}_f)}$  and use that in place of  $\Delta E_{(\text{T})}$ .

It should be noted that Goodson’s continued fraction formula (18) could be extended to take the  $(\text{Q}_f)$  correction as a separate contribution in a similar manner to Schröder et al.,<sup>92</sup> although we have not done so here.

A better-known alternative to Goodson’s continued fraction approach is the %TAE[(T)] metric of Martin.<sup>8,66</sup> Specifically devised in the context of total atomization energies—which, at the time, were the primary means of determining ab initio predictions of enthalpies of formation of small molecules—the %TAE[(T)] is determined by calculating the contribution

of the perturbative triples portion of CCSD(T) to an atomization energy of some species  $X$ ,

$$\%TAE[(T)](X) \equiv \frac{N_C \Delta E_{(T)}(C) + N_N \Delta E_{(T)}(N) + N_O \Delta E_{(T)}(O) + \cdots - \Delta E_{(T)}(X)}{N_H E_{SCF}(H) + N_C E_{CC}(C) + N_N E_{CC}(N) + \cdots - E_{CC}(X)}, \quad (21)$$

where  $N_H, N_C, N_N, N_O$  are the number of hydrogen, carbon, nitrogen, and oxygen atoms in the species  $X$ ,  $\Delta E_{(T)}(X)$  is the perturbative triples contribution to the energy of a given species (as above), and  $E_{CC}(X)$  is the full CCSD(T) energy (both SCF and correlation) of a given species. This method has been demonstrated to be a good predictor of when electron correlation beyond CCSD(T) is important in model chemistries based upon total atomization energies,<sup>8</sup> but is not usually extended to general chemical reactions, a point that will be addressed later in this manuscript.

# Methods

## Chemical Test-Suite

In order to benchmark the errors associated with the factorization of the  $D_4$  denominator via CCSD(TQ<sub>f</sub>), we have constructed two test-suites drawn from the singlet species of the W4.17 dataset<sup>88</sup> containing the atoms H, B, C, N, O, and F. These species span a range of chemical properties and bonding—from those well described by single reference methods, such as water, to those of decidedly multi-reference character, such as BN and C<sub>2</sub>—and total 93 individual molecules. The first test-suite is composed of the absolute correlation energies of these compounds. This metric contains information about the trueness<sup>95</sup> of the quantum chemical methods applied to correlation energies without fortuitous or designed cancellation of errors, but is applied to properties of molecules that cannot be determined experimentally.

The second test-suite, the “reaction” suite, is constructed by extending the ANL- $n$  scheme for determining enthalpies of formations of molecules.<sup>96</sup> In that work, the enthalpy of formation of a molecule containing only H, C, N, and O atoms was obtained by balancing a chemical reaction where the reactants (LHS) contains the species of interest, the products (RHS) contains an appropriate number of CH<sub>4</sub>, NH<sub>3</sub>, and H<sub>2</sub>O as “reference species” to balance the atom count of the LHS, and additional H<sub>2</sub> molecules are added to the reactants or products in order to balance the remaining H atoms. This work extends this scheme in two ways. First, we add B<sub>2</sub>H<sub>6</sub> and HF as reference species for the B and F atoms. Second, we extend the scheme to include non-reference molecules in not just the reactants, but also in the products of the chemical reaction, which are then balanced with the species of the reference set (H<sub>2</sub>, B<sub>2</sub>H<sub>6</sub>, CH<sub>4</sub>, NH<sub>3</sub>, H<sub>2</sub>O, and HF). By considering all unique pairs of non-reference molecules as reactants/products, this test-suite is expanded from a comparatively small number of species to a much more significant number of chemical reactions (93 species and 2859 reactions in the case of the cc-pVDZ basis set,<sup>97</sup> and 73 species and 1600 reactions in the case of the cc-pVTZ basis set<sup>97</sup>). The reactions contained within span a considerable

range of chemical space, from simple decompositions such as  $\text{H}_2\text{CCH}_2 \rightarrow \text{HCCH} + \text{H}_2$  to complicated reactions such as  $\text{BF}_3 + \text{NH}_3 \rightarrow \text{BN} + 3 \text{HF}$ , which preserve little-to-no bonding from the products to reactants. The species, reactions, and raw energetic data for all results presented below may be generated from the data and scripts in the Electronic Supplementary Information.

It should be noted that great care must be applied to the training or fitting of any model to the information contained within this reaction energy test-suite. The reaction energies contained within will be significantly, though subtly, influenced by the set of reference species chosen to balance the left and the right hand side of the chemical reactions. That is, a great majority of the reactions in this reaction energies suite will be connected to a subset of reference molecules containing only six species, and any model trained to minimize errors in the chemical behavior may fall dangerously close to overfitting any information that correlates with features of these six molecules. In the language of machine learning, we would term the absolute energy test-suite as the “training” set, and the reaction test-suite as the “validation” set.<sup>98</sup> It is perhaps informative that the former in our case is far smaller than the latter; indicative of our desire to remain as *ab initio* as possible in our approach and to avoid (or at least minimize) the aforementioned dangers of fitting the specific reaction scheme utilized.

## Software

The closed-shell data generated in this work was obtained via the NCC module of the CFOUR quantum chemistry package,<sup>99</sup> using the current development version. All calculations were sufficiently converged such that numerical noise is several orders of magnitude below the statistics presented here. The  $(Q_f)$  contributions for the open-shell atoms were determined using a Python package designed for rapid prototyping of coupled cluster methods,<sup>100</sup> which in turn relies on an interface to PYSCF for one and two-body integrals.<sup>101</sup> The open-shell CCSDT(Q)<sub>A</sub> calculations were performed using the general CC code, MRCC, of

Kállay and coworkers.<sup>102</sup> All post-SCF calculations include only valence correlation.

## Basis Sets

The basis sets tested include Dunning’s cc-pVDZ and cc-pVTZ,<sup>97</sup> with the majority of the results obtained using the small cc-pVDZ basis set. This is a result of three independent considerations. First, the effects of quadruple excitations in the majority of high-accuracy model chemistries are determined using this basis set, even if the increment from CCSD(T) to CCSDT is often treated with a cc-pVTZ or larger basis. Second, the purpose of this work is not to replace the HLC terms of these model chemistries, but rather to demonstrate that the magnitude and sign of these corrections can be estimated using perturbation theory, and cc-pVDZ is sufficient for this purpose. Third, it has been previously reported that CCSDT[Q] displays basis-set dependence discordant with that of CCSDT(Q) and CCSDT(Q)<sub>Λ</sub>, accumulating undesirable errors with basis sets larger than cc-pVDZ.<sup>28</sup> This may need to be accounted for if one were to attempt to include the fifth-order (Q<sub>f</sub>) and [Q] corrections in a model chemistry, but as our purpose is simply to provide guidelines for estimating HLC contributions, we will put this issue aside.

## Analysis

The statistical analysis of the absolute and reaction energy test-suites discussed above was performed with the R language,<sup>103</sup> with scripts available in the SI. Each statistical measurement reported below was further characterized by BCa (bias-corrected and accelerated)<sup>104,105</sup> analysis of ten-thousand bootstrap replicas<sup>106</sup> in order to establish 95% confidence intervals, all performed using the `boot` package<sup>105,107</sup> in R version 4.3.3 (2024-02-29). As we are most interested in the “trueness”<sup>95</sup> of the methods investigated here, we define a new error measure



as the (bias corrected)  $l_2$  norm of the deviations  $x_i$  (denoted  $L_2D$ ),

$$L_2D = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N-1}}, \quad (22)$$

which is also identical to the usual standard deviation formula with the average measurement  $\bar{x}$  taken as zero.

# Results and Discussion

## Factorization Error: CCSD(TQ) vs CCSD(TQ<sub>f</sub>)

To characterize errors that might arise by factorizing the  $D_4$  denominator as in the CCSD(TQ<sub>f</sub>) method, our initial results focus on the raw correlation energy differences between the fifth-order [Q] and [Q<sub>f</sub>] corrections. The top panel of Fig. 2 displays a histogram of the difference between the fifth-order [Q] and (Q<sub>f</sub>) correlation energies applied to the absolute energy test-suite calculated with the cc-pVDZ basis set. As shown by the histogram itself, the CCSD(TQ<sub>f</sub>) and CCSD(TQ) energies coincide to a remarkable degree. This analysis is corroborated in Table 1, which records an average error of 4.2  $\mu\text{E}_\text{H}$ , with an asymmetric 95% confidence interval that skews toward (Q<sub>f</sub>) slightly under-predicting [Q]. The reported L<sub>2</sub>D value of 24.9  $\mu\text{E}_\text{H}$  is exceedingly small, and the only outliers are BN and C<sub>2</sub>, which are particularly nasty species in this test-set. These results corroborate previous findings in the literature, which demonstrated small errors incurred by this factorization in smaller datasets.<sup>42,58,62</sup>

This trend continues when analyzing the impact of these quadruples’ corrections to reaction energies of the reaction energy test-suite calculated with the cc-pVDZ basis, as illustrated in the bottom panel of Fig. 2. Here, we again find a tight error distribution centered around a mean value -1.5 cal mol<sup>-1</sup>, with outliers corresponding to reactions containing C<sub>2</sub> and BN. As shown in Table 1, the 95% confidence interval is very tight, with an L<sub>2</sub>D of 28.1 cal mol<sup>-1</sup>—entirely negligible on the scale of chemical accuracy (1 kcal mol<sup>-1</sup>).

Results for the smaller cc-pVTZ test-suites for molecular correlation energies and reaction energies track with those found at the cc-pVDZ level. We find that (Q<sub>f</sub>)/cc-pVTZ tends to underestimate both raw correlation and reaction energies. Table 1 documents the mean error in the prediction of raw correlation energies to be 7.95  $\mu\text{E}_\text{H}$  using the larger cc-pVTZ basis set, with a L<sub>2</sub>D of 22.9  $\mu\text{E}_\text{H}$ . Reaction energies determined using the cc-pVTZ basis set tend to slightly increase the mean error to -2.6 cal mol<sup>-1</sup> and decrease the L<sub>2</sub>D to 22.4

cal mol<sup>-1</sup>, compared to cc-pVDZ. There is virtually no impact in L<sub>2</sub>D between the cc-pVDZ and cc-pVTZ basis sets, which seems to indicate that the factorization of [Q] into (Q<sub>f</sub>) is not a highly basis-set dependent approximation, even if the [Q] correction itself is.

In summary, we find that the errors arising from the factorization of  $D_4$  in CCSD(TQ) to be manifestly negligible in both the description of correlation energy and in correlation energy contributions to reaction energies.

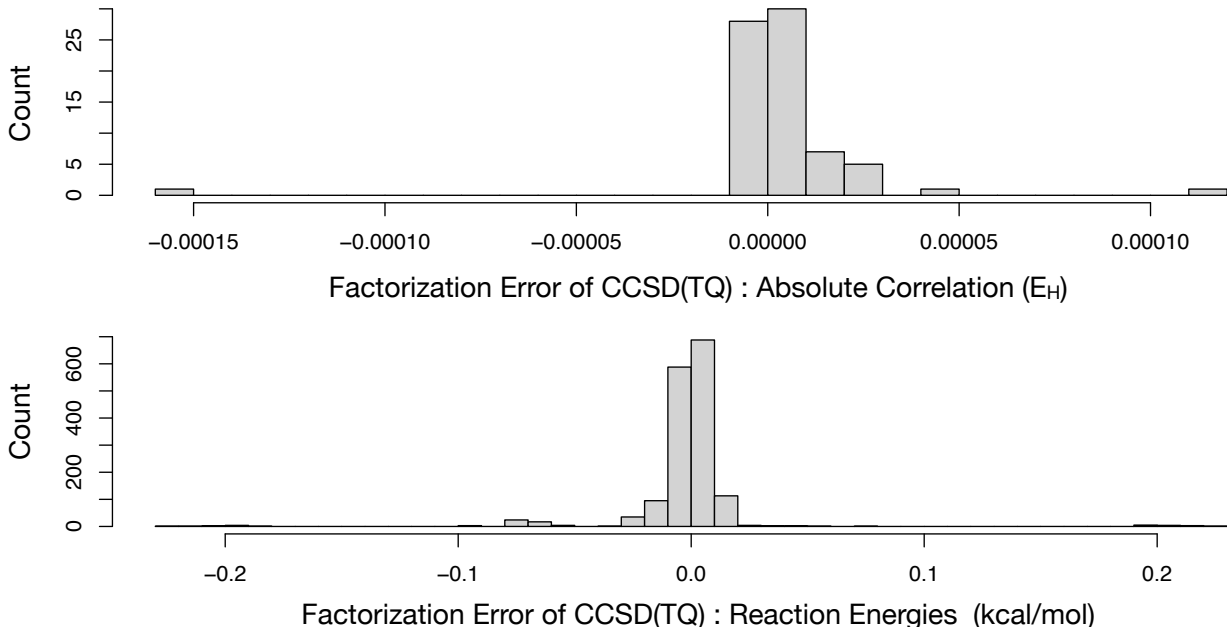


Figure 2: Histogram of the error incurred by factorizing the  $D_4$  denominator in CCSD(TQ)/cc-pVDZ correlation energies (top) and contributions to reaction energies (bottom). Outliers on the left and right hand side of the x-axis are BN and C<sub>2</sub> (top) and reactions involving these species (bottom).

## Scaled-Perturbation Estimators of Higher-Level Correlation

Having determined that the factorization approximations in CCSD(TQ<sub>f</sub>) introduce no significant errors on the scale of kcal mol<sup>-1</sup> accuracy, we may begin to consider what role these methods might play in model chemistries. To this end, Fig. 3 displays histograms of the difference between CCSDT(Q)<sub>Λ</sub> (our “platinum standard” method for post-CCSD(T) calculations) and CCSD(T)/CCSD(TQ<sub>f</sub>) in the context of cc-pVDZ contributions to reaction

Table 1: Mean signed error (MSE),  $L_2D$ , and max absolute error (MAXE) values incurred by the factorization of the  $D_4$  denominator in CCSD(TQ<sub>f</sub>) for both absolute correlation energies and the test-suite of reaction energies. Subscripts and superscripts correspond to 95% confidence intervals determined via BCa analysis of 10,000 bootstrap samples.

Absolute Energy ( $\mu E_H$ )		
	cc-pVDZ	cc-pVTZ
MSE	$4.2^{+4.5}_{-7.0}$	$8.0^{+5.4}_{-4.5}$
$L_2D$	$24.9^{+19.2}_{-14.4}$	$22.9^{+16.3}_{-9.1}$
MAXE	154.2	129.9
Reaction Energy (cal mol <sup>-1</sup> )		
	cc-pVDZ	cc-pVTZ
MSE	$-1.5^{+1.4}_{-1.3}$	$-2.6^{+1.1}_{-1.1}$
$L_2D$	$28.1^{+4.6}_{-3.9}$	$22.4^{+2.7}_{-2.5}$
MAXE	227.5	168.9

energies. The mean signed error,  $L_2D$ , and max absolute errors are also tabulated in Table 3 (under the “None” column, indicating that no additional correction or extrapolation is applied). In general, we find that CCSD(T) displays smaller mean HLC contributions than CCSD(TQ<sub>f</sub>) (0.05 vs 0.10 kcal mol<sup>-1</sup>), along with reduced  $L_2D$  (0.66 vs 0.77 kcal mol<sup>-1</sup>) of the missing HLC, but with larger maximum errors (5.53 vs 4.00 kcal mol<sup>-1</sup>). While CCSD(TQ) has been shown to improve the treatment of total correlation energies<sup>42,58,62</sup> (a result we have also observed in the process of performing this study), the inclusion of fifth-order quadruples corrections at the CCSD level of theory does not improve the prediction of reaction energies, and in fact harms it.

However, a more prescient application of CCSD(TQ<sub>f</sub>) is not as a direct substitute for more complete theories such as CCSDT(Q)<sub>A</sub>, but rather as a predictor of them. For instance, we may ask the following question: “given just a CCSD(T)-cost calculation, might it be possible to estimate the size of the missing HLC terms to, say, one or two kcal mol<sup>-1</sup>”? Tacit within our attempt to address this question are three critical approximations: that the cc-pVDZ basis set is sufficient to represent post-CCSD(T) correlation (model chemistries aiming for kcal mol<sup>-1</sup> accuracy may find this cumbersome and model chemistries aiming for

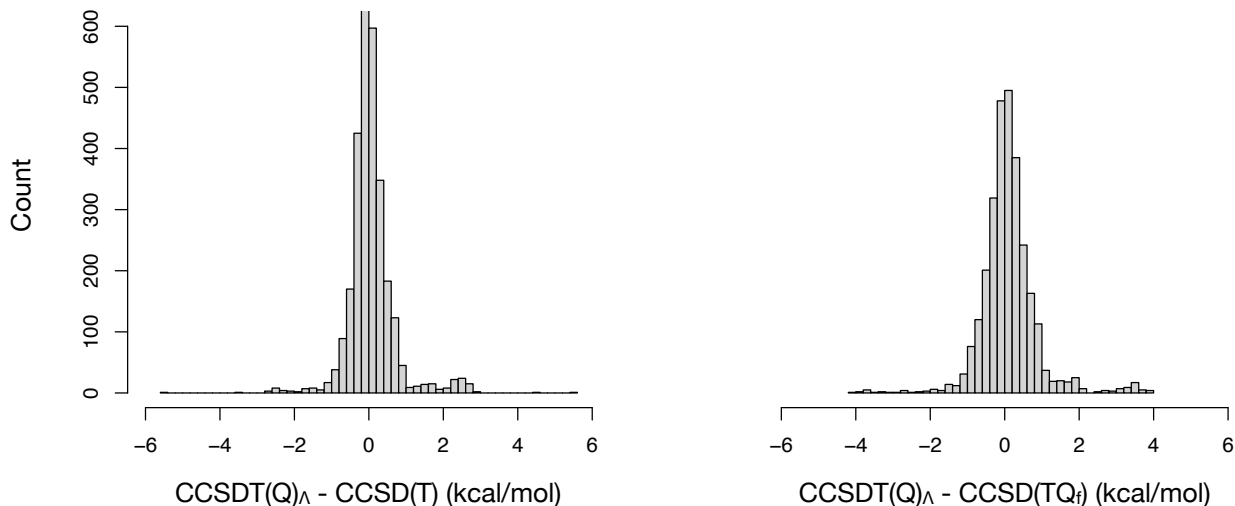


Figure 3: Histogram of differences between  $\text{CCSDT(Q)}_{\Lambda}$  and  $\text{CCSD(T)}/\text{cc-pVDZ}$  (left),  $\text{CCSD(TQ}_f)/\text{cc-pVDZ}$  (right) for reaction energies.

$\text{kJ mol}^{-1}$  accuracy may find this insufficient), that  $\text{CCSDT(Q)}_{\Lambda}$  is a sufficient representation of correlation effects after  $\text{CCSD(T)}$  (evidence indicates that this is true for well-behaved species, but it may not be sufficient for multireference molecules such as  $\text{C}_2$  and  $\text{BN}$ , see Refs. 11,53,54,66), and that core-valence correlation beyond  $\text{CCSD(T)}$  may be safely neglected for accuracy on the scale of one to two  $\text{kcal mol}^{-1}$  (which is certainly true for molecules containing atoms  $Z \leq 10$ , but not so obviously true for heavier atoms).

Related versions of this question have been asked before, specifically in the pursuit of so-called multi-reference indices (MRI).<sup>70–84</sup> But, while these indices are useful guides to the degree of complexity in the electronic wavefunction, it is not always obvious how to translate this into quantitative prediction of HLC effects. This is demonstrated in Fig. 4, where the often quoted MRI, the max  $\text{CCSD } T_2$ , is plotted on the x-axis and the missing HLC of  $\text{CCSD(TQ}_f)/\text{cc-pVDZ}$  is plotted on the y-axis. Despite the fact that this index certainly contains some information about the suitability of the  $\text{CCSD}$  wavefunction to address these molecular correlation energies, this index does not track in a particularly quantitative manner with extent of correlation missing in  $\text{CCSD(T)}$ . Quantitative correlation is certainly improved by using Martin’s  $\% \text{TAE}[(\text{T})]$ , displayed in the central panel of Fig. 4, which show a generally

linear relationship between the fraction of the TAE that is accounted for by the perturbative triples correction and the extent of the the HLC corrections. The agreement is not perfect—the trend appears to break down for particularly large %TAE[(T)]—and it is not generalized to arbitrary molecular reactions, but it is certainly an improvement. We can, however, go further. The right-most panel of this figure employs a different type of metric that is intimately related to the %TAE[(T)] of Martin. However, instead of taking a fraction of an atomization energy (a quantity which blends features of the absolute and reaction energy measures), we display the relationship between post-CCSD(TQ<sub>f</sub>) correlation energy and simply the difference between CCSD(TQ<sub>f</sub>) and CCSD in a cc-pVDZ basis set. Said otherwise, rather than examining the fraction of an atomization energy that is accounted for by perturbative triples, we examine the total of a molecule’s correlation energy that is accounted for by perturbative triples and quadruples. This approach reveals a linear relationship between the HLC correction to CCSD(TQ<sub>f</sub>) and the (TQ<sub>f</sub>) correction to CCSD, and may thus be modeled as by a simple equation of the form  $y \approx b * x$ , where  $b$  is the slope obtained by least squares fit to the data in the rightmost panel of Fig. 4.

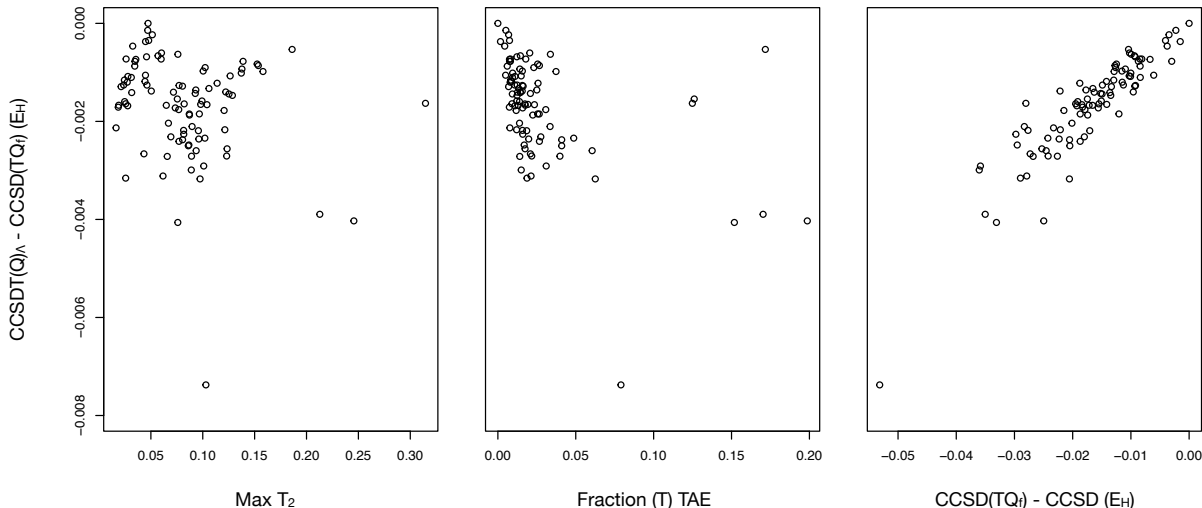


Figure 4: Max  $T_2$  (left) and %TAE[(T)] (middle, fractional form) multireference diagnostics as predictors of higher-level correlation, compared against the (TQ<sub>f</sub>) perturbative corrections (right) suggested in this work.

This has a particularly intuitive interpretation. Let us begin by writing out an expression

for the approximation of  $\text{CCSDT}(\text{Q})_\Lambda$  using this linear fit (assuming all terms are evaluated with the same basis set).

$$E_{\text{CCSDT}(\text{Q})_\Lambda} \approx E_{\text{CCSD}(\text{TQ}_f)} + b(E_{\text{CCSD}(\text{TQ}_f)} - E_{\text{CCSD}}) \quad (23)$$

$$\approx E_{\text{CCSD}} + (1 + b)\Delta E_{(\text{TQ}_f)} \quad (24)$$

The above equation claims that the difference between CCSD and  $\text{CCSDT}(\text{Q})_\Lambda$  may be captured by the leading-order perturbative treatment of  $T_3$  and  $T_4$  from  $\text{CCSD}(\text{TQ}_f)$  multiplied by a scaling factor (hopefully) close to unity. We thus term this approach as a “scaled perturbation estimator”, or SPE for short, and will examine its application in the remainder of this manuscript. Importantly, the SPE scaling factor is not fit to data from the reaction energies, but to the correlation energies of individual molecules.

Fig. 5 displays two variations of this SPE using either  $\text{CCSD}(\text{T})$  or  $\text{CCSD}(\text{TQ}_f)$  as the “base” method. It also displays the relationship of the HLC of  $\text{CCSD}(\text{T})$  and  $\text{CCSD}(\text{TQ}_f)$  with Goodson’s CF approximation, which we find also scales roughly linearly with the size of higher-level correlation contributions to molecular energies. It should be noted that we also tested Goodson’s R and Q metrics, but found them inferior to Goodson’s CF method. The point of these plots is that the location of each black point on the x-axis can be determined at  $\mathcal{O}(N^7)$  cost, while the y-value requires  $\mathcal{O}(N^9)$ . Thus, a linear fit of the form  $y \approx b * x$  to any of these plots could potentially be used as a HLC estimator at a cost no greater than  $\text{CCSD}(\text{T})$ . The coefficients of the linear fits of these trends are displayed in Table 2, and are portrayed by the red lines in Figure 5.

We then examine the performance of six of these models in the prediction of cc-pVDZ HLC in the reaction energies test-suite:

- HLC of  $\text{CCSD}(\text{T})$  predicted by linear fit of (T),
- HLC of  $\text{CCSD}(\text{T})$  predicted by Goodson CF of  $\text{CCSD}(\text{T})$ ,

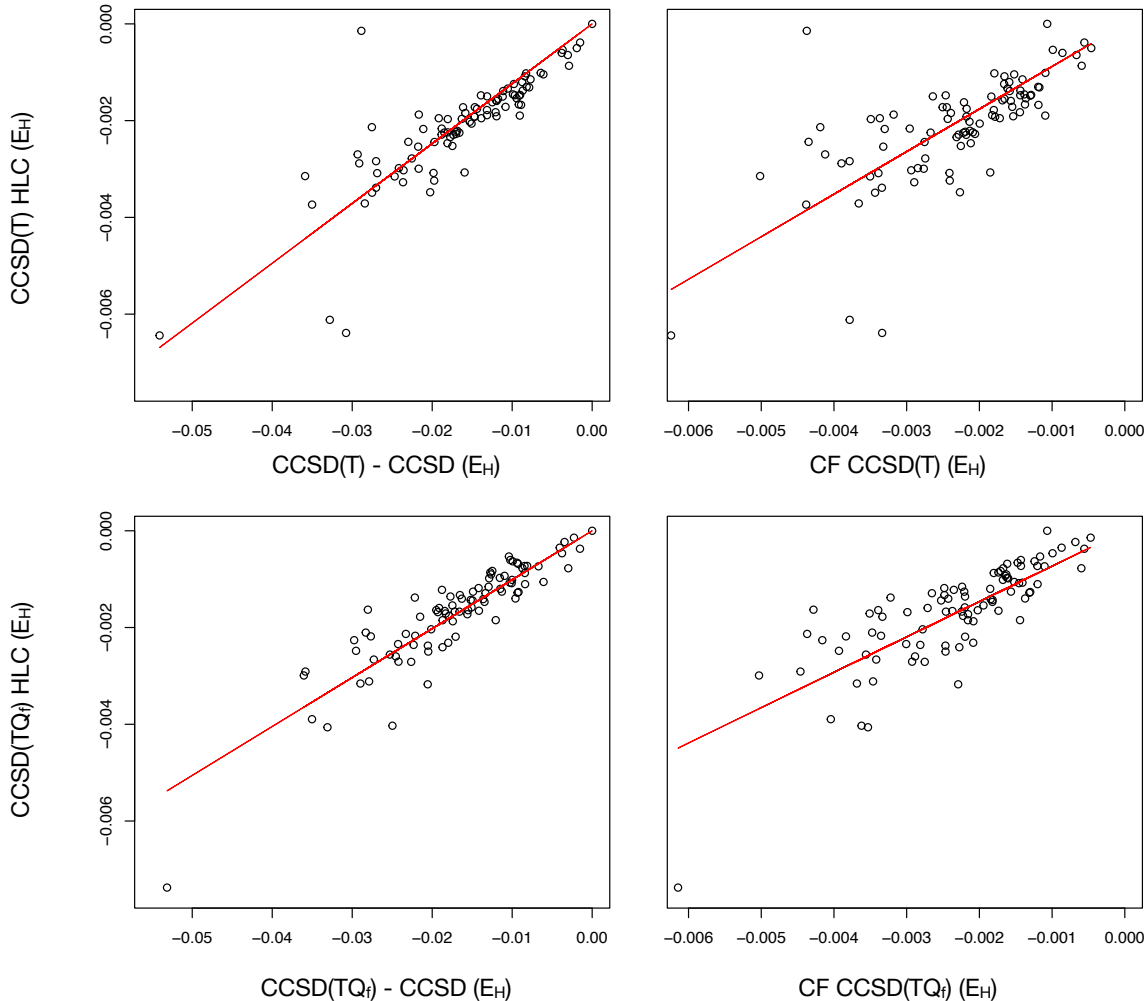


Figure 5: Trends of CCSD(T)-scaling methods vs their HLC corrections in a cc-pVDZ basis set applied to molecular correlation energies. The top row uses CCSD(T) as the base method, and the bottom row uses CCSD(TQ<sub>f</sub>) as the base method. The red line indicates a linear fit with zero intercept.

Table 2: Parameters of linear fits ( $y \approx b*x$ ) to CCSD(T) and CCSD(TQ<sub>f</sub>) HLC corrections in a cc-pVDZ basis set, determined from molecular correlation energies using either the “scaled perturbation estimator” (SPE) or Goodson’s continued fraction (CF) method.<sup>85</sup>

Base Method		SPE	Goodson CF
		b	b
CCSD(T)	Best Fit	0.1235	0.8786
	Std. Err.	0.0035	0.0336
CCSD(TQ <sub>f</sub> )	Best Fit	0.1012	0.7309
	Std. Err.	0.0025	0.0247



- HLC of CCSD(T) predicted by linear fit of Goodson CF of CCSD(T),
- HLC of CCSD(TQ<sub>f</sub>) predicted by linear fit of (TQ<sub>f</sub>),
- HLC of CCSD(TQ<sub>f</sub>) predicted by Goodson CF of CCSD(TQ<sub>f</sub>),
- HLC of CCSD(TQ<sub>f</sub>) predicted by linear fit of Goodson CF of CCSD(TQ<sub>f</sub>).

These results are summarized in Fig. 6 and Table 3, the latter of which also includes the HLC errors accrued in model chemistries that ignore higher-level correlation altogether in the “None” column. It is immediately obvious that the fairly reasonable fits in molecular correlation energies do not necessarily translate to even qualitatively accurate predictions of HLC contributions to reaction energies. Take for example Goodson’s continued fraction method applied to CCSD(T) (top right panel of Fig. 6). Not only does the method itself show no trend with the actual size of the HLC contributions to reaction energies (reflected in a large L<sub>2</sub>D of 1.79 kcal mol<sup>-1</sup> and max absolute error of 8.70 kcal mol<sup>-1</sup>), the linear fit of this metric to the molecular correlation energies does not even qualitatively track the true data in the reaction energies. This is an unfortunately common theme in computational thermochemistry : approximate methods that are reasonable for raw correlation energies are not guaranteed to be reasonable for reaction energies. The SPE approaches based off CCSD(T) and CCSD(TQ<sub>f</sub>) perform markedly better, with L<sub>2</sub>D of 0.61 and 0.52 kcal mol<sup>-1</sup>, respectively. We also observe some trends that indicate that correct physics is being accounted for in these approaches: the sign of the perturbative correction to CCSD generally corresponds to the sign of the missing HLC (hence the diagonal slant to this data). Perhaps most surprising is that CCSD(TQ<sub>f</sub>), which is a less accurate treatment of correlation in the determination of reaction energies than CCSD(T), performs significantly better than any of the other methods considered here when used to inform a simple linear regression. This is particularly reflected in the maximum error exhibited by the CCSD(TQ<sub>f</sub>)-based SPE, which is nearly 2 kcal mol<sup>-1</sup> smaller than the nearest competitor. Visually, we also see that the inclusion of these fifth-order quadruples corrections in CCSD(TQ<sub>f</sub>) tightens (makes more

diagonal) the relationship between the perturbative treatments of  $T_3$  and  $T_4$  and the higher-level correlation correction, even if the spread of said HLC contributions is larger. That the factorized, fifth-order quadruple corrections in CCSD(TQ<sub>f</sub>) play such an important role in these predictions has, to the best of our knowledge, not been observed in the extant literature. It also suggests that the claims of equation (23) are not unfounded, and that the scaled perturbation of CCSD by (TQ<sub>f</sub>) may be made into a reasonable estimate for the more rigorous treatment of  $T_3$  and  $T_4$  in CCSDT(Q)<sub>Λ</sub>.

That this simple, empirical scaling works so well is a bit surprising, but nonetheless provides a means by which to reasonably estimate the size and sign of HLC contributions in chemical reactions with nothing more than values that may be routinely calculated at CCSD(T)-cost and a little bit of mental math. The answer to the above question: “Is it possible to estimate higher-level correlation effects to one or two kcal mol<sup>-1</sup> using just CCSD(T)-cost methods?” seems to have been answered in the affirmative (even accounting for the factor of two increase in the L<sub>2</sub>D that would better align these values with a 95% confidence interval), at least for the reactions within the current test-suite.

It should be reiterated that these models are not necessarily intended to be used as replacements for higher-level correlation in thermochemical recipes, especially not in anything that aims for accuracy on the order of a kcal mol<sup>-1</sup>. Such an application would need to be carefully studied in the context of the individual model chemistry under examination, and would require further developments to reduce the L<sub>2</sub>D of the SPE predictions of higher level correlation. While the (TQ<sub>f</sub>)-based SPE developed in this work is the best predictor of HLC that we know of, a prediction of a HLC contribution to a reaction energy on the order of 2 kcal mol<sup>-1</sup> should probably be taken as no more than an indication that a more rigorous treatment of post-CCSD(T) correlation is in order if kcal mol<sup>-1</sup> accuracy is desired.

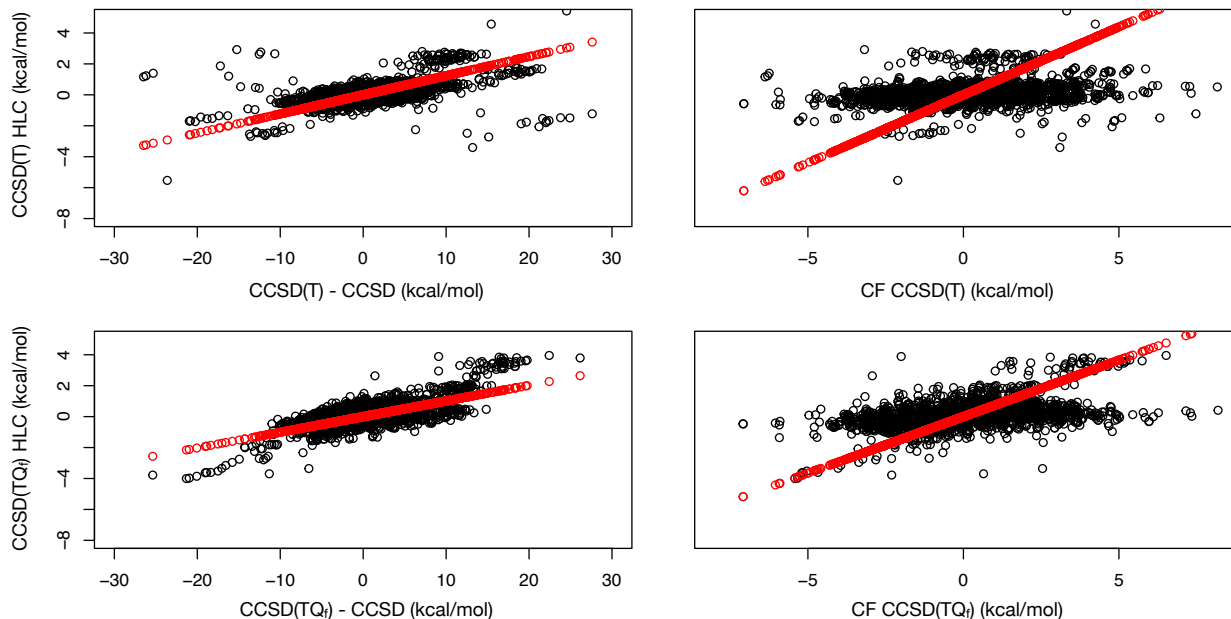


Figure 6: Relationship of CCSD(T) and CCSD(TQ<sub>f</sub>) vs their HLC corrections with a cc-pVDZ basis set applied to reaction energies. The top row uses CCSD(T) as the base method, and the bottom row uses CCSD(TQ<sub>f</sub>) as the base method. Red dots indicate the values obtained using the linear fits derived from molecular correlation energies, obtained at CCSD(T) cost.

Table 3: Mean signed error (MSE), L<sub>2</sub>D, and max absolute errors (MAXE) of various CCSD(T)-cost predictors of higher-level correlation contributions to reaction energies, as described in the main text. Values in subscripts and superscripts correspond to 95% confidence intervals determined via BCa analysis of the reported predictor applied to 10,000 bootstrap samples of the reaction test-suite, respectively. All values are in kcal mol<sup>-1</sup>

Base Method		None	SPE	Linear CF	Goodson CF
CCSD(T)	MSE	0.05 <sup>+0.02</sup> <sub>-0.02</sub>	0.06 <sup>+0.02</sup> <sub>-0.02</sub>	0.02 <sup>+0.06</sup> <sub>-0.06</sub>	0.02 <sup>+0.06</sup> <sub>-0.07</sub>
	L <sub>2</sub> D	0.66 <sup>+0.04</sup> <sub>-0.04</sub>	0.61 <sup>+0.06</sup> <sub>-0.05</sub>	1.58 <sup>+0.06</sup> <sub>-0.05</sub>	1.79 <sup>+0.06</sup> <sub>-0.06</sub>
	MAXE	5.53	5.04	7.80	8.70
CCSD(TQ <sub>f</sub> )	MSE	0.10 <sup>+0.03</sup> <sub>-0.03</sub>	0.07 <sup>+0.02</sup> <sub>-0.02</sub>	0.08 <sup>+0.04</sup> <sub>-0.05</sub>	0.06 <sup>+0.06</sup> <sub>-0.06</sub>
	L <sub>2</sub> D	0.77 <sup>+0.04</sup> <sub>-0.04</sub>	0.52 <sup>+0.02</sup> <sub>-0.02</sub>	1.21 <sup>+0.04</sup> <sub>-0.04</sub>	1.64 <sup>+0.05</sup> <sub>-0.05</sub>
	MAXE	4.00	2.97	5.57	7.78

## Connection to Martin’s %TAE

Finally, it is worth commenting on the connection between these scaled perturbation estimators and the percent-(T) TAE based methods that are commonly employed. In Karton

et al., 2006,<sup>8</sup> it was found that the percentage of connected  $T_4$  and  $T_5$  contribution to total atomization energies post-CCSD(T) was reasonably correlated with contribution from (T):  $\%TAE[T_4 + T_5] \approx 0.126 \times \%TAE[(T)]$ . Note that the factor of  $\sim 0.126$  has been obtained by fitting to a set of atomization energies, but is rather similar to the  $\sim 0.123$  we obtain for the fit of the SPE fit of HLC based on CCSD(T).

Now, this may be simplified to better compare to the methods presented in this work. If we insert the definition of  $\%TAE[X]$  (21), the denominators may be cancelled such that

$$\begin{aligned} N_C \Delta E_{\text{HLC}}(C) + N_N \Delta E_{\text{HLC}}(N) + \cdots - \Delta E_{\text{HLC}}(X) \\ \approx 0.126 \times (N_C \Delta E_{(\text{T})}(C) + N_N \Delta E_{(\text{T})}(N) + \cdots - \Delta E_{(\text{T})}(X)), \end{aligned} \quad (25)$$

where  $\Delta E_{\text{HLC}}$  indicates the high-level correlation correction from  $T_4$  and  $T_5$ . Then we may determine an estimate of the HLC contributions for a given reaction, for example, the isomerization of HCN to HNC,

$$\begin{aligned} \Delta E_{\text{HLC}}^{\text{rxn}}(\text{HCN} \rightarrow \text{HNC}) &= \Delta E_{\text{HLC}}^{\text{TAE}}(\text{HCN}) - \Delta E_{\text{HLC}}^{\text{TAE}}(\text{HNC}) \\ &\approx 0.126 (E_{\text{atoms}} - \Delta E_{(\text{T})}(\text{HCN})) \\ &\quad - 0.126 (E_{\text{atoms}} - \Delta E_{(\text{T})}(\text{HNC})) \\ &= 0.126 (\Delta E_{(\text{T})}(\text{HNC}) - \Delta E_{(\text{T})}(\text{HCN})), \end{aligned} \quad (26)$$

where  $E_{\text{HLC}}^{\text{rxn}}(\text{HCN} \rightarrow \text{HNC})$  is the contribution of HLC to the reaction energy and  $E_{\text{atoms}}$  is a shorthand for sum of the  $\Delta E_{(\text{T})}$  for the H, C, and N atoms. Thus, the  $\%TAE[(\text{T})]$  metric may actually be used in a nearly identical manner to the scaled perturbation estimators developed in this work, though the fit of the scaling parameters in that context are determined from percentages of atomization energies rather than absolute correlation energies, which can have significant consequences for the accuracy of the resulting model. The main contributions of this work, then, are the theoretical justification of performance of these scaled perturbation

estimators and/or %TAE[(T)] metrics, the determination and benchmarking of these scaling factors for a set of 93 molecules and 2859 reactions containing atoms H, B, C, N, O and F in a cc-pVDZ basis set, and the demonstration that the factorization of the CCSD(TQ<sub>f</sub>) method both incurs negligible error for the vast majority of species/reactions contained in these test-suites and noticeably improves the performance of SPE methods for estimating higher-level correlation.

## Conclusions

The development of new tools and methodologies which circumvent the sometimes unaffordable computational cost associated with high-level coupled cluster theory is particularly important in the domain of thermochemistry targeting chemical accuracy. However, there are currently only limited means in which to evaluate the importance of post-CCSD(T) corrections to correlation energies of molecules or reactions *a priori*. Furthermore, the viability of cheaper quadruples corrections like ( $Q_f$ ) has not been holistically assessed in the domain of thermochemistry. The current work takes steps to address these two gaps in the literature. We find that factorizing the fifth-order [Q] diagrams of CCSD(TQ), leading to the ( $Q_f$ ) correction and CCSD(T $Q_f$ ) method, introduces minimal error and a  $\mathcal{O}(N^2)$  reduction of cost in applications to thermochemistry, even when applied to significantly multireference species such as  $C_2$  and BN.

We then examined the use of these methods in the context of higher-level correlation corrections in chemical reaction thermodynamics. While we find that CCSD(T $Q_f$ ) on its own is not sufficient to account for HLC effects in reaction energies, we demonstrate that these factorized quadruples methods can play an important role in the estimation of post-CCSD(T) correlation energies such as CCSDT(Q) $_{\Lambda}$ . In particular, we investigate two potential HLC estimators; Goodson-style continued fraction methods and a method we refer to as scaled perturbation estimators. While we did not find the first to be a reasonable route to estimating HLC effects, we do find that a simple linear regression of the leading-order perturbative corrections to CCSD in CCSD(T $Q_f$ ) reliably predicts the size of post-CCSD(T $Q_f$ ) correlation in a cc-pVDZ basis set, with a mean error of 0.07 kcal mol $^{-1}$ , a  $L_2D$  of 0.52 kcal mol $^{-1}$ , and a maximum error of 2.97 kcal mol $^{-1}$ . These results improves upon the CF method applied to this same dataset by a factor of more than three. Importantly, we also find that the maximum absolute errors of these predictors are dramatically reduced by the inclusion of the factorized quadruples terms. We conclude by showing the connection between these methods and the %TAE[(T)] that is commonly used as a multireference index. Given their low cost, these

scaled perturbation estimates may be a viable means by which to “ballpark” the importance of post-CCSD(T) contributions to reactions energies in other studies without expending more than CCSD(T)/cc-pVDZ effort. This provides a useful guide for electronic-structure practitioners who are uncertain if these higher-level effects are of particular relevance in a given reaction.

## Acknowledgement

D.A.M. and J.H.T. acknowledge support by the US National Science Foundation under grant CHE-2143725. J.H.T. acknowledges support from the SMU Moody School of Graduate and Advanced Studies. R.J.B and Z.W.W acknowledge support from the Air Force Office of Scientific Research under AFOSR Award No. FA9550-23-1-0118. Z.W.W. thanks the National Science Foundation and the Molecular Sciences Software Institute for financial support under Grant No. CHE-2136142.

## Supporting Information Available

Additional supplemental information is available as a .zip file, containing all of the molecular structures, raw energetic data, and analysis scripts (in R) used to obtain the data presented in the manuscript.

## References

- (1) Karton, A. In *Chapter Three - Quantum mechanical thermochemical predictions 100years after the Schrödinger equation*; Dixon, D. A., Ed.; Annual Reports in Computational Chemistry; Elsevier, 2022; Vol. 18; pp 123–166.
- (2) Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated *ab initio* thermochemistry. *The Journal of Chemical Physics* **2004**, *121*, 11599–11613.
- (3) Bomble, Y. J.; Vázquez, J.; Kállay, M.; Michauk, C.; Szalay, P. G.; Császár, A. G.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated *ab initio* thermochemistry. II. Minor improvements to the protocol and a vital simplification. *The Journal of Chemical Physics* **2006**, *125*, 064108.



- (4) Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated *ab initio* thermochemistry. III. Additional improvements and overview. *The Journal of Chemical Physics* **2008**, *128*, 114111.
- (5) Thorpe, J. H.; Lopez, C. A.; Nguyen, T. L.; Baraban, J. H.; Bross, D. H.; Ruscic, B.; Stanton, J. F. High-accuracy extrapolated *ab initio* thermochemistry. IV. A modified recipe for computational efficiency. *The Journal of Chemical Physics* **2019**, *150*, 224102.
- (6) Martin, J. M. L.; de Oliveira, G. Towards standard methods for benchmark quality *ab initio* thermochemistry—W1 and W2 theory. *The Journal of Chemical Physics* **1999**, *111*, 1843–1856.
- (7) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. W3 theory: Robust computational thermochemistry in the kJ/mol accuracy range. *The Journal of Chemical Physics* **2004**, *120*, 4129–4141.
- (8) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *The Journal of Chemical Physics* **2006**, *125*, 144108.
- (9) Karton, A.; Martin, J. M. L. Explicitly correlated Wn theory: W1-F12 and W2-F12. *The Journal of Chemical Physics* **2012**, *136*, 124114.
- (10) Sylvetsky, N.; Peterson, K. A.; Karton, A.; Martin, J. M. L. Toward a W4-F12 approach: Can explicitly correlated and orbital-based *ab initio* CCSD(T) limits be reconciled? *The Journal of Chemical Physics* **2016**, *144*, 214101.
- (11) Semidalas, E.; Karton, A.; Martin, J. M. L. W4 $\Lambda$ : Leveraging  $\Lambda$  Coupled-Cluster for Accurate Computational Thermochemistry Approaches. *The Journal of Physical Chemistry A* **2024**, *128*, 1715–1724.

- (12) Császár, A. G.; Allen, W. D.; Schaefer, H. F. In pursuit of the *ab initio* limit for conformational energy prototypes. *The Journal of Chemical Physics* **1998**, *108*, 9751–9764.
- (13) East, A. L. L.; Allen, W. D. The heat of formation of NCO. *The Journal of Chemical Physics* **1993**, *99*, 4638–4650.
- (14) Kenny, J. P.; Allen, W. D.; Schaefer, H. F. Complete basis set limit studies of conventional and R12 correlation methods: The silicon dicarbide (SiC–2) barrier to linearity. *The Journal of Chemical Physics* **2003**, *118*, 7353–7365.
- (15) Schuurman, M. S.; Muir, S. R.; Allen, W. D.; Schaefer, H. F. Toward subchemical accuracy in computational thermochemistry: Focal point analysis of the heat of formation of NCO and [H,N,C,O] isomers. *The Journal of Chemical Physics* **2004**, *120*, 11586–11599.
- (16) Jaeger, H. M.; Schaefer, H. F.; Demaison, J.; Császár, A. G.; Allen, W. D. Lowest-Lying Conformers of Alanine: Pushing Theory to Ascertain Precise Energetics and Semiexperimental Re Structures. *Journal of Chemical Theory and Computation* **2010**, *6*, 3066–3078.
- (17) Ruscic, B.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Harding, L. B.; Asher, R. L.; Wagner, A. F. Evidence for a Lower Enthalpy of Formation of Hydroxyl Radical and a Lower Gas-Phase Bond Dissociation Energy of Water. *The Journal of Physical Chemistry A* **2000**, *105*, 1–4.
- (18) Ruscic, B.; Wagner, A. F.; Harding, L. B.; Asher, R. L.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Song, Y.; Qian, X.; Ng, C.-Y.; et al. On the Enthalpy of Formation of Hydroxyl Radical and Gas-Phase Bond Dissociation Energies of Water and Hydroxyl. *The Journal of Physical Chemistry A* **2002**, *106*, 2727–2747.

- (19) Feller, D.; Peterson, K. A.; Dixon, D. A. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *The Journal of Chemical Physics* **2008**, *129*, 204105.
- (20) Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical accuracy in *ab initio* thermochemistry and spectroscopy: current strategies and future challenges. *Theoretical Chemistry Accounts* **2012**, *131*, 1079.
- (21) Feller, D.; Peterson, K. A.; Dixon, D. A. Further benchmarks of a composite, convergent, statistically calibrated coupled-cluster-based approach for thermochemical and spectroscopic studies. *Molecular Physics* **2012**, *110*, 2381–2399.
- (22) Dixon, D. A.; Feller, D.; Peterson, K. A. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Ed.; Elsevier, 2012; Vol. 8; Chapter 1, pp 1–28.
- (23) Feller, D.; Peterson, K. A.; Ruscic, B. Improved accuracy benchmarks of small molecules using correlation consistent basis sets. *Theoretical Chemistry Accounts* **2013**, *133*, 1407.
- (24) Purvis, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics* **1982**, *76*, 1910–1918.
- (25) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters* **1989**, *157*, 479–483.
- (26) Kucharski, S. A.; Bartlett, R. J. Coupled-cluster methods that include connected quadruple excitations, T4: CCSDTQ-1 and Q (CCSDT). *Chemical Physics Letters* **1989**, *158*, 550–555.

- (27) Kucharski, S. A.; Bartlett, R. J. Coupled-cluster methods correct through sixth order. *Chemical physics letters* **1993**, *206*, 574–583.
- (28) Bomble, Y. J.; Stanton, J. F.; Kállay, M.; Gauss, J. Coupled-cluster methods including noniterative corrections for quadruple excitations. *The Journal of Chemical Physics* **2005**, *123*, 054101.
- (29) Kucharski, S. A.; Bartlett, R. J. The coupled-cluster single, double, triple, and quadruple excitation method. *The Journal of chemical physics* **1992**, *97*, 4282–4288.
- (30) Kucharski, S. A.; Bartlett, R. J. Sixth-order energy corrections with converged coupled cluster singles and doubles amplitudes. *The Journal of chemical physics* **1998**, *108*, 5255–5264.
- (31) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *The Journal of Chemical Physics* **1989**, *90*, 5622–5629.
- (32) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. Gaussian-1 Theory of Molecular Energies for Second-Row Compounds. *The Journal of Chemical Physics* **1990**, *93*, 2537–2545.
- (33) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 Theory for Molecular Energies of First- and Second-Row Compounds. *The Journal of Chemical Physics* **1991**, *94*, 7221–7230.
- (34) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *The Journal of Chemical Physics* **1998**, *109*, 7764–7776.
- (35) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *The Journal of Chemical Physics* **2007**, *126*, 084108.

- (36) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *The Journal of Chemical Physics* **1988**, *89*, 2193–2218.
- (37) Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *The Journal of Chemical Physics* **1991**, *94*, 6081–6090.
- (38) Petersson, G. A.; Tensfeldt, T. G.; Montgomery, J., J. A. A complete basis set model chemistry. III. The complete basis set-quadratic configuration interaction family of methods. *The Journal of Chemical Physics* **1991**, *94*, 6091–6101.
- (39) Montgomery, J., J. A.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. IV. An improved atomic pair natural orbital method. *The Journal of Chemical Physics* **1994**, *101*, 5900–5909.
- (40) Ochterski, J. W.; Petersson, G. A.; Montgomery, J., J. A. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *The Journal of Chemical Physics* **1996**, *104*, 2598–2619.
- (41) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *The Journal of Chemical Physics* **1999**, *110*, 2822–2827.
- (42) Kucharski, S. A.; Bartlett, R. J. Noniterative energy corrections through fifth-order to the coupled cluster singles and doubles method. *The Journal of chemical physics* **1998**, *108*, 5243–5254.
- (43) Crawford, T. D.; Stanton, J. F. Investigation of an asymmetric triple-excitation correction for coupled-cluster energies. *International journal of quantum chemistry* **1998**, *70*, 601–611.

- (44) Stanton, J. F. Why CCSD(T) works: a different perspective. *Chemical Physics Letters* **1997**, *281*, 130–134.
- (45) Taube, A. G.; Bartlett, R. J. Improving upon CCSD(T): ACCSD(T). I. Potential energy surfaces. *The Journal of Chemical Physics* **2008**, *128*, 044110.
- (46) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. Towards a full CCSDT model for electron correlation. *The Journal of Chemical Physics* **1985**, *83*, 4041–4046.
- (47) Noga, J.; Bartlett, R. J.; Urban, M. Towards a full CCSDT model for electron correlation. CCSDT-n models. *Chemical physics letters* **1987**, *134*, 126–132.
- (48) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. A coupled cluster approach with triple excitations. *The Journal of chemical physics* **1984**, *81*, 5906–5912.
- (49) Noga, J.; Bartlett, R. J. The full CCSDT model for molecular electronic structure. *The Journal of chemical physics* **1987**, *86*, 7041–7050.
- (50) Noga, J.; Kucharski, S. A.; Bartlett, R. J. A coupled-cluster method that includes connected quadruple excitations. *The Journal of chemical physics* **1989**, *90*, 3399–3400.
- (51) Eriksen, J. J.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P.; Gauss, J. A Lagrangian framework for deriving triples and quadruples corrections to the CCSD energy. *The Journal of Chemical Physics* **2014**, *140*, 064108.
- (52) Kucharski, S. A.; Bartlett, R. J. Recursive intermediate factorization and complete computational linearization of the coupled-cluster single, double, triple, and quadruple excitation equations. *Theoretica chimica acta* **1991**, *80*, 387–405.
- (53) Thorpe, J. H.; Kilburn, J. L.; Feller, D.; Changala, P. B.; Bross, D. H.; Ruscic, B.; Stanton, J. F. Elaborated thermochemical treatment of HF, CO, N<sub>2</sub>, and H<sub>2</sub>O: Insight into HEAT and its extensions. *The Journal of Chemical Physics* **2021**, *155*, 184109.

- (54) Thorpe, J. H.; Feller, D.; Bross, D. H.; Ruscic, B.; Stanton, J. F. Sub 20 cm<sup>-1</sup> computational prediction of the CH bond energy – a case of systematic error in computational thermochemistry. *Phys. Chem. Chem. Phys.* **2023**, *25*, 21162–21172.
- (55) Frantz, L. M.; Mills, R. L. Many-body basis for the optical model. *Nuclear Physics* **1960**, *15*, 16–32.
- (56) Bartlett, R. J.; Silver, D. M. Some aspects of diagrammatic perturbation theory. *International Journal of Quantum Chemistry* **1975**, *9*, 183–198.
- (57) Kucharski, S. A.; Bartlett, R. J. In *Fifth-Order Many-Body Perturbation Theory and Its Relationship to Various Coupled-Cluster Approaches*; Löwdin, P.-O., Ed.; Advances in Quantum Chemistry; Academic Press, 1986; Vol. 18; pp 281–344.
- (58) Kucharski, S. A.; Bartlett, R. J. An efficient way to include connected quadruple contributions into the coupled cluster method. *The Journal of chemical physics* **1998**, *108*, 9221–9226.
- (59) Kucharski, S. A.; Musiał, M. Connected quadruple excitations in the coupled-cluster theory. *Molecular Physics* **2010**, *108*, 2975–2985.
- (60) Kucharski, S. A.; Bartlett, R. J. Connected Quadruples for the Frequencies of O<sub>3</sub>. *The Journal of chemical physics* **1999**, *110*, 8233–8235.
- (61) Kucharski, S. A.; Kolaski, M.; Bartlett, R. J. Toward the limits of predictive electronic structure theory: Connected quadruple excitations for large basis set calculations. *The Journal of Chemical Physics* **2001**, *114*, 692–700.
- (62) Musiał, M.; Bartlett, R. J. Critical comparison of various connected quadruple excitation approximations in the coupled-cluster treatment of bond breaking. *The Journal of Chemical Physics* **2005**, *122*, 224102.

- (63) Musial, M.; Bartlett, R. J. Improving upon CCSD(TQ<sub>f</sub>) for potential energy surfaces:  $\Lambda$ CCSD(TQ<sub>f</sub>) models. *The Journal of Chemical Physics* **2010**, *133*, 104102.
- (64) Eriksen, J. J.; Matthews, D. A.; Jørgensen, P.; Gauss, J. Assessment of the accuracy of coupled cluster perturbation theory for open-shell systems. I. Triples expansions. *The Journal of Chemical Physics* **2016**, *144*, 194102.
- (65) Eriksen, J. J.; Matthews, D. A.; Jørgensen, P.; Gauss, J. Assessment of the accuracy of coupled cluster perturbation theory for open-shell systems. II. Quadruples expansions. *The Journal of Chemical Physics* **2016**, *144*, 194103.
- (66) Martin, J. M. The eight-valence-electron systems re-examined: convergence of the coupled-cluster series and performance of quasiperturbative methods for quadruple excitations. *Molecular Physics* **2014**, *112*, 785–793.
- (67) Thorpe, J. H. Applications of the Coupled Cluster Ansatz to the Electronic and Vibrational Structure of Molecules. Ph.D. thesis, University of Florida, 2022.
- (68) Franke, P. R.; Stanton, J. F. Rotamers of Methanediol: Composite Ab Initio Predictions of Structures, Frequencies, and Rovibrational Constants. *The Journal of Physical Chemistry A* **2023**, *127*, 924–937.
- (69) Spiegel, M.; Semidalas, E.; Martin, J. M. L.; Bentley, M. R.; Stanton, J. F. Post-CCSD(T) corrections to bond distances and vibrational frequencies: the power of  $\Lambda$ . *Molecular Physics* **2024**, *122*, e2252114.
- (70) Lee, T. J.; Rice, J. E.; Scuseria, G. E.; Schaefer, H. F. Theoretical investigations of molecules composed only of fluorine, oxygen and nitrogen: determination of the equilibrium structures of FOOF, (NO)<sub>2</sub> and FN<sub>2</sub> and the transition state structure for FN<sub>2</sub> cis-trans isomerization. *Theoretica chimica acta* **1989**, *75*, 81–98.



- (71) Lee, T. J.; Taylor, P. R. A diagnostic for determining the quality of single-reference electron correlation methods. *International Journal of Quantum Chemistry* **1989**, *36*, 199–207.
- (72) Ivanov, V. V.; Lyakh, D. I.; Adamowicz\*, L. New indices for describing the multi-configurational nature of the coupled cluster wave function. *Molecular Physics* **2005**, *103*, 2131–2139.
- (73) Bartlett, R. J.; Park, Y. C.; Bauman, N. P.; Melnichuk, A.; Ranasinghe, D.; Ravi, M.; Perera, A. Index of multi-determinantal and multi-reference character in coupled-cluster theory. *The Journal of Chemical Physics* **2020**, *153*, 234103.
- (74) Faulstich, F. M.; Kristiansen, H. E.; Csirik, M. A.; Kvaal, S.; Pedersen, T. B.; Laestadius, A. S-Diagnostic—An a Posteriori Error Assessment for Single-Reference Coupled-Cluster Methods. *The Journal of Physical Chemistry A* **2023**, *127*, 9106–9120.
- (75) Janssen, C. L.; Nielsen, I. M. New diagnostics for coupled-cluster and Møller–Plesset perturbation theory. *Chemical Physics Letters* **1998**, *290*, 423–430.
- (76) Leininger, M. L.; Nielsen, I. M.; Crawford, T.; Janssen, C. L. A new diagnostic for open-shell coupled-cluster theory. *Chemical Physics Letters* **2000**, *328*, 431–436.
- (77) Lee, T. J. Comparison of the T1 and D1 diagnostics for electronic structure theory: a new definition for the open-shell D1 diagnostic. *Chemical Physics Letters* **2003**, *372*, 362–367.
- (78) Tishchenko, O.; Zheng, J.; Truhlar, D. G. Multireference Model Chemistries for Thermochemical Kinetics. *Journal of Chemical Theory and Computation* **2008**, *4*, 1208–1219.
- (79) Jiang, W.; DeYonker, N. J.; Wilson, A. K. Multireference Character for 3d Transition-

- Metal-Containing Molecules. *Journal of Chemical Theory and Computation* **2012**, *8*, 460–468.
- (80) Mok, K.; Neumann, R.; Handy, N. Dynamical and nondynamical correlation. *Journal of Physical Chemistry* **1996**, *100*, 6225–6230.
- (81) Krylov, A. I.; Sherrill, C. D.; Byrd, E. F. C.; Head-Gordon, M. Size-consistent wave functions for nondynamical correlation energy: The valence active space optimized orbital coupled-cluster doubles model. *The Journal of Chemical Physics* **1998**, *109*, 10669–10678.
- (82) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. A simple DFT-based diagnostic for nondynamical correlation. *Theoretical Chemistry Accounts* **2012**, *132*, 1291.
- (83) Martin, J. M. L.; Santra, G.; Semidalas, E. An exchange-based diagnostic for static correlation. *AIP Conference Proceedings* **2022**, *2611*, 020014.
- (84) Xu, X.; Soriano-Agueda, L.; López, X.; Ramos-Cordoba, E.; Matito, E. All-Purpose Measure of Electron Correlation for Multireference Diagnostics. *Journal of Chemical Theory and Computation* **2024**, *20*, 721–727.
- (85) Goodson, D. Z. Extrapolating the coupled-cluster sequence toward the full configuration-interaction limit. *The Journal of Chemical Physics* **2002**, *116*, 6948–6956.
- (86) Karton, A.; Taylor, P. R.; Martin, J. M. L. Basis set convergence of post-CCSD contributions to molecular atomization energies. *The Journal of Chemical Physics* **2007**, *127*, 064104.
- (87) Feller, D.; Peterson, K. A. Probing the limits of accuracy in electronic structure calculations: Is theory capable of results uniformly better than “chemical accuracy”? *The Journal of Chemical Physics* **2007**, *126*, 114105.

- (88) Karton, A.; Sylvetsky, N.; Martin, J. M. L. W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *Journal of Computational Chemistry* **2017**, *38*, 2063–2075.
- (89) Bartlett, R. J. Perspective on Coupled-cluster Theory. The evolution toward simplicity in quantum chemistry. *Physical Chemistry Chemical Physics* **2024**, *26*, 8013–8037.
- (90) Bartlett, R. J.; Noga, J. The expectation value coupled-cluster method and analytical energy derivatives. *Chemical physics letters* **1988**, *150*, 29–36.
- (91) Bartlett, R. J.; Kucharski, S. A.; Noga, J. Alternative coupled-cluster ansätze II. The unitary coupled-cluster method. *Chemical physics letters* **1989**, *155*, 133–140.
- (92) Schröder, B.; Sebald, P.; Stein, C.; Weser, O.; Botschwina, P. Challenging High-Level ab initio Rovibrational Spectroscopy: The Nitrous Oxide Molecule. *Zeitschrift für Physikalische Chemie* **2015**, *229*, 1663–1690.
- (93) Feller, D.; Peterson, K. A.; de Jong, W. A.; Dixon, D. A. Performance of coupled cluster theory in thermochemical calculations of small halogenated compounds. *The Journal of Chemical Physics* **2003**, *118*, 3510–3522.
- (94) Feller, D. Estimating the intrinsic limit of the Feller-Peterson-Dixon composite approach when applied to adiabatic ionization potentials in atoms and small molecules. *The Journal of Chemical Physics* **2017**, *147*, 034103.
- (95) Ruscic, B. Uncertainty Quantification in Thermochemistry, Benchmarking Electronic Structure Computations, and Active Thermochemical Tables. *International Journal of Quantum Chemistry* **2014**, *114*, 1097–1101.
- (96) Klippenstein, S. J.; Harding, L. B.; Ruscic, B. Ab Initio Computations and Active Thermochemical Tables Hand in Hand: Heats of Formation of Core Combustion Species. *The Journal of Physical Chemistry A* **2017**, *121*, 6580–6602.

- (97) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.
- (98) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep learning*; MIT press, 2016.
- (99) Matthews, D. A.; Cheng, L.; Harding, M. E.; Lipparini, F.; Stopkowitz, S.; Jagau, T.-C.; Szalay, P. G.; Gauss, J.; Stanton, J. F. Coupled-cluster techniques for computational chemistry: The CFOUR program package. *The Journal of Chemical Physics* **2020**, *152*, 214108.
- (100) Windom, Z. W. UT2: A python-based suite of coupled-cluster methods designed to converge towards an “ultimate” T2 method. <https://github.com/zzw-4855/ut2>, Visted on 07/03/2024.
- (101) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; others PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1340.
- (102) Kállay, M. *et al.* The MRCC program system: Accurate quantum chemistry from water to proteins. *The Journal of Chemical Physics* **2020**, *152*, 074107.
- (103) R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2024.
- (104) Efron, B. Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association* **1987**, *82*, 171–185.
- (105) A. C. Davison; D. V. Hinkley *Bootstrap Methods and Their Applications*; Cambridge University Press: Cambridge, 1997; ISBN 0-521-57391-2.

- (106) Efron, B.; Tibshirani, R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* **1986**, *1*, 54 – 75.
- (107) Angelo Canty; B. D. Ripley boot: Bootstrap R (S-Plus) Functions. 2024; R package version 1.3-29.

# TOC Graphic

