THE JOURNAL OF CHEMICAL PHYSICS **125**, 064108 (2006)

High-accuracy extrapolated *ab initio* thermochemistry. II. Minor improvements to the protocol and a vital simplification

Yannick J. Bomble[a] and Juana Vázquez
Institute for Theoretical Chemistry, Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712

Mihály Kállay
Department of Physical Chemistry, Budapest University of Technology and Economics, P.O. Box 91, H-1521 Budapest, Hungary

Christine Michauk
Institut für Physikalische Chemie, Universität Mainz, D-55099 Mainz, Germany

Péter G. Szalay and Attila G. Császár
Institute of Chemistry, Eötvös Loránd University, P.O. Box 32, H-1518 Budapest, Hungary

Jürgen Gauss
Institut für Physikalische Chemie, Universität Mainz, D-55099 Mainz, Germany

John F. Stanton
Institute for Theoretical Chemistry, Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712

(Received 24 March 2006; accepted 27 April 2006; published online 10 August 2006)

The recently developed high-accuracy extrapolated *ab initio* thermochemistry method for theoretical thermochemistry, which is intimately related to other high-precision protocols such as the Weizmann-3 and focal-point approaches, is revisited. Some minor improvements in theoretical rigor are introduced which do not lead to any significant additional computational overhead, but are shown to have a negligible overall effect on the accuracy. In addition, the method is extended to completely treat electron correlation effects up to pentuple excitations. The use of an approximate treatment of quadruple and pentuple excitations is suggested; the former as a pragmatic approximation for standard cases and the latter when extremely high accuracy is required. For a test suite of molecules that have rather precisely known enthalpies of formation {as taken from the active thermochemical tables of Ruscic and co-workers [*Lecture Notes in Computer Science*, edited by M. Parashar (Springer, Berlin, 2002), Vol. 2536, pp. 25–38; J. Phys. Chem. A **108**, 9979 (2004)]}, the largest deviations between theory and experiment are 0.52, −0.70, and 0.51 kJ mol^{-1} for the latter three methods, respectively. Some perspective is provided on this level of accuracy, and sources of remaining systematic deficiencies in the approaches are discussed. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2206789]

I. INTRODUCTION

The field of high-accuracy thermochemistry is going through a renaissance. Owing to improvements in theoretical methods and the computational power that is needed to apply them, *ab initio* methods have entered a domain of accuracy that was previously unachievable.[1–5] Perhaps even more important, the precision of "experimental"[6] results for small molecules has been vastly improved in the past few years due to the use of the active table approach of Ruscic and co-workers.[7] The active thermochemical tables (ATcTs) have been instrumental in improving the accuracy and reducing the uncertainties of key species such as HO, HO_2, acetylene, and the carbon atom and have moved the standard of precision into what might be termed the spectroscopic range: a few tenths of a kJ mol^{-1} or, alternatively, tens of wave numbers.

It is perhaps not fully appreciated that the ATcT database provides the best set of benchmark data for calibrating the performance of various "theoretical model chemistries"[8] in predicting the thermodynamic stability of small molecules. Before ATcTs, it had been effectively impossible to rigorously evaluate methods for all but a very few molecules. This is due to errors and/or relatively large uncertainties found in the standard [CODATA,[9] Gurvich *et al.*[10] and JANAF (Ref. 11)] thermochemical databases, which comprised most of the experimental numbers used in comparisons. In addition to the fact that the molecular data are imperfect, there is another important source of error. Specifically, total atomization energies (TAEs) are used by most theoretical model chemistries, from which molecular enthalpies of formation can be calculated by subtracting the appropriate experimental atomic contributions from the TAEs. However, the canonical value of $\Delta_f H^o$ for the carbon atom (711.19±0.45 kJ mol^{-1} at 0 K (Ref. 9)] carries a fairly

---

[a] Electronic mail: ybomble@mail.utexas.edu

significant error bar; the correct value appears to be 711.5 kJ mol^{-1}.[12,13] Thus, a systematic error exists in all enthalpies of formation calculated for organic molecules using the TAE approach, the magnitude of which is proportional to the number of carbon atoms.[13]

Recently, a theoretical model chemistry known as high-accuracy extrapolated *ab initio* thermochemistry (HEAT) was defined and tested for a set of roughly 30 atoms and small molecules. HEAT is closely related to other approaches that are capable of very high accuracy—the focal-point approach of Császár *et al.*,[4] the W3 and W4 methods[1,2] from the Weizmann group, and various strategies based on estimates of the basis set limit that have been developed elsewhere.[3] In HEAT, no empirical scaling or correction factors are applied, and the method is best viewed as being designed for the estimation of absolute energies. Thermochemical quantities of interest, specifically enthalpies of formation, can then be calculated by a number of means; these include atomization energies, elemental formation reactions, or (when a suitable reaction can be designed that comprises only atoms and molecules in the HEAT database) reactions that are isodesmic[14] or nearly so.

While similar to other high-level methods, HEAT is undoubtedly the most costly model chemistry developed to date, a shortcoming that limited its application to just one molecule with more than two nonhydrogen atoms. The motivation for HEAT was to develop a method that represents the current state of the art for small molecules using generally available computational resources, rather than a more broadly applicable method such as W3. In this work, effects of a number of minor improvements to the original HEAT method are assessed, and a simplification is made which helps us to define a new model chemistry that is considerably less expensive to apply than the original HEAT approach but which retains essentially the same accuracy. Finally, we explore the realm of increased accuracy by extending the theoretical treatment of electron correlation beyond that used previously.

II. THEORY

A. Overview

According to the original HEAT protocol, total energies are calculated as a sum of eight terms,

$$E_{\mathrm{HEAT}} = E_{\mathrm{HF}}^{\infty} + \Delta E_{\mathrm{CCSD(T)}}^{\infty} + \Delta E_{\mathrm{CCSDT}} + \Delta E_{\mathrm{HLC}} + \Delta E_{\mathrm{rel}} + \Delta E_{\mathrm{DBOC}} + \Delta E_{\mathrm{SO}} + \Delta E_{\mathrm{ZPE}}. \tag{1}$$

The various contributions to E_{HEAT} are described below.

E_{HF}^{∞}. This is an estimate of the Hartree-Fock limit energy, approximated by extrapolating Hartree-Fock self-consistent-field (HF-SCF) energies calculated with the aug-cc-pCVXZ hierarchy[16] of basis sets (X=T, Q, and 5) using the formula suggested by Feller,[15]

$$E_{\mathrm{HF}}^{X} = E_{\mathrm{HF}}^{\infty} + a\exp(-bX), \tag{2}$$

where E_{HF}^{X} is the HF-SCF energy obtained with the aug-cc-pCVXZ basis set and a and b are parameter fits to the calculated energies.

$\Delta E_{\mathrm{CCSD(T)}}^{\infty}$. This is an estimate of the CCSD(T) (Ref. 17) correlation energy at the basis set limit, through extrapolation of energies obtained with the aug-cc-pCVQZ and aug-cc-pCV5Z basis sets using the approach advocated by Helgaker *et al.*,[18] which has a somewhat tenuous basis in the atomic partial wave expansion of the correlation energy,

$$\Delta E_{\mathrm{CCSD(T)}}^{X} = \Delta E_{\mathrm{CCSD(T)}}^{\infty} + aX^{-3}. \tag{3}$$

Here, $\Delta E_{\mathrm{CCSD(T)}}^{X}$ is the CCSD(T) correlation energy obtained with the aug-cc-pCVXZ basis sets. Note that through CCSD(T), core and valence correlation effects are not assumed to be additive; the CCSD(T) calculations above are done for all electrons, using (quite large) basis sets that are appropriate for the treatment of both core and valence correlation effects.

$\Delta E_{\mathrm{CCSDT}}$. This is an estimate of the difference between full CCSDT (Ref. 19) and CCSD(T) correlation energies at the basis set limit, obtained by extrapolating CCSDT and CCSD(T) energies with the cc-pVTZ and cc-pVQZ basis sets and Eq. (3). Core correlation effects are *not* included at this stage of the calculation; it is thus implicitly assumed that they are converged at the CCSD(T) level of theory.

ΔE_{HLC}. This "higher-level correlation" contribution attempts to account, approximately, for deficiencies in the CCSDT method and is the most expensive step involved in the HEAT protocol. Numerically, it is given by the difference in CCSDTQ (Ref. 20) and CCSDT correlation energies, both of these obtained in valence-only calculations using the cc-pVDZ basis set. This correction is one of the focuses of this research.

ΔE_{rel}. This is an estimate of the relativistic contribution to the total energy, in which the one-electron Darwin and mass-velocity terms[21] are calculated using perturbation theory at the CCSD(T)/aug-cc-pCVTZ level.

ΔE_{DBOC}. This is an estimate of the diagonal Born-Oppenheimer correction,[22,23] which is the expectation value of the nuclear kinetic energy operator, taken over the electronic (clamped-nucleus) wave function. This contribution is calculated at the HF-SCF level of theory using the aug-cc-pVTZ basis set.

ΔE_{SO}. This refers to the difference between the true ground state level for radicals in degenerate electronic states and the weighted average that corresponds to the results calculated with nonrelativistic electronic structure programs. In HEAT, ΔE_{SO} is calculated using a spin-orbit configuration interaction (CI) program and relativistic effective core potentials, using the cc-pVDZ basis set.

ΔE_{ZPE}. This is the zero-point vibrational contribution to the energy. This is calculated from the CCSD(T)/cc-pVQZ quartic force field using the formula[24]

$$\Delta E_{\mathrm{ZPE}} = \sum_i \frac{\omega_i}{2} + \frac{1}{4}\sum_{i<j} x_{ij}, \tag{4}$$

which includes a contribution from vibrational anharmonicity.

Before continuing, two additional things should be mentioned. First, all calculations are performed at the geometry obtained by minimizing the CCSD(T)/cc-pVQZ all-electron energy of the electronic ground state. Second, while straightforward restricted Hartree-Fock calculations are done for all closed-shell molecules, the electronic energy calculations done on radicals used unrestricted Hartree-Fock (UHF) wave functions up through CCSDT. However, program limitations at the time of publication prevented UHF reference functions from being used for the $\Delta E_{\mathrm{CCSDTQ}}$ correction.[5] UHF calculations were also used in the ΔE_{ZPE} calculations, apart from the NO, CF, CCH, and OF radicals,[25] where problems due to the onset of spin contamination[26] necessitated the use of restricted open-shell (ROHF) reference functions.

B. Modest improvements in protocol

We first investigate the effect of the following three improvements to the original HEAT protocol.

(1) The zero-point energy formula [Eq. (4)] used in the original HEAT paper is based on second-order vibrational perturbation theory (VPT2). This equation suffices to calculate energy differences and is therefore entirely sufficient for spectroscopic analyses, but it neglects a constant contribution to the energy levels. The correct equation is

$$\Delta E_{\mathrm{ZPE}} = G_0 + \sum_i \frac{\omega_i}{2} + \frac{1}{4}\sum_{i<j} x_{ij}. \tag{5}$$

Formulas for the G_0 term have been derived independently by three groups[5,27–29] in the past few years, and, apart from some apparent misprints,[30] all three agree. This contribution is included in this work and designated as Δ_{G_0}.

(2) At the time that the HEAT project was initiated, the higher-level correlation correction ΔE_{HLC} could not be calculated for UHF reference functions. As a result, there is a minor inconsistency in the original protocol for radicals, as all correlated calculations up to CCSDT were performed with UHF wave functions, and the correction for residual correlation embodied in ΔE_{HLC} was based on ROHF reference functions. The implicit assumption—that CCSDT energies obtained with UHF and ROHF wave functions are the same[31]—is expected to lead to very small errors (on the order of 0.1 kJ mol^{-1}). In this work, we address this question and report results from UHF-based CCSDTQ calculations. The differences between the original protocol and those obtained by the consistent use of UHF reference functions is denoted as Δ_{RU}.

(3) The relativistic corrections used in the original HEAT protocol are obtained via a perturbative treatment based on the one-electron mass-velocity and Darwin terms[21] using CCSD(T) with the aug-cc-pCVTZ basis. In the present work, this treatment is expanded to include the corresponding two-electron Darwin term referred to here as Δ_{D2}. For the corresponding computational expression see, for example, Ref. 32. In addition, some calculations have been performed to check on the basis set sensitivity. However, as those were found to be negligible (less than 0.1 kJ mol^{-1}), this work focuses on corrections obtained at the CCSD(T)/aug-cc-pCVTZ level.

The improvements to the original protocol described above are based on advances in theory and/or program capabilities; all of them are associated with a negligible amount of additional computation. Hence, it is felt that they should be included in the "standard" method. Despite a statement to the contrary in the literature about the wisdom of defining a class of similar methods with additional acronyms,[5] it is necessary to introduce a convention by which we might differentiate the model chemistry with and without these modest improvements. Accordingly, the original HEAT method will retain its name, and the methods introduced in this paper will be named in a different manner in which the treatment of higher-level correlation (defined as that beyond CCSDT) and the sequence of basis sets used in the extrapolations determine the name of the method. The HEAT method defined earlier and all of those of this paper are based on extrapolations using the X=T(3), 4, and 5 basis sets. However, one could use the larger sequence X=4, 5, and 6 (as we have done in some ongoing projects[33]), and it behooves us to introduce albeit cumbersome notation that distinguishes these methods. Hence, we will refer to the method comprising the minor improvements above as HEAT345-Q.

C. New model chemistries: HEAT345-QP and HEAT345-(Q)

In addition to the modest improvements outlined above, two other model chemistries are investigated here. The first will be known as HEAT345-QP, in which the higher-level correlation correction in HEAT345-Q is replaced by the difference between (valence-only) CCSDTQP and CCSDT energies, obtained with the cc-pVDZ basis set. This method is considerably more expensive than HEAT345-Q and consequently has a severely limited scope of application. However, it was possible to carry out HEAT345-QP calculations for all of the molecules used in our test suite. Nevertheless, in cases where electron correlation effects are extremely important, it is believed that HEAT345-QP will provide results superior to those obtained with the simpler HEAT345-Q approach.

Finally, we introduce a *simplified* method in which the higher-level correlation correction is based on the recently developed CCSDT(Q) model of Bomble *et al.*[34] This method treats the effects of quadruple excitations noniteratively, is analogous to the extremely popular and successful CCSD(T) treatment of triple excitations, and has already been used in high-accuracy thermochemical investigations of *trans*-butadiene,[35] as well as isomers of C_3H_3,[36] all of which present very demanding cases for HEAT345-Q.

In this work, enthalpies of formation (at 0 K) are calculated from TAEs, and thus correspond to the entries denoted by $\Delta_f H(I)$ in Table IV of Ref. 5. In this paper, enthalpies of formation for the carbon atom (711.58 kJ mol^{-1}) and fluorine (77.11 kJ mol^{-1}) from the ATcT database have been used in preference to the CODATA values, from which they differ by +0.39 and −0.16 kJ mol^{-1}, respectively.

It should be noted that we still advocate the HEAT family of methods as approaches to calculate total molecular and atomic energies and most judiciously used for thermochemi-

TABLE I. Enthalpies of formation (at 0 K, in kJ mol^{-1}, as calculated from atomization energies using the current ATcT value for the carbon and fluorine atoms, see text) according to the original and an improved HEAT protocol (columns 2 and 6) and contributions due to modest improvement discussed in the text.

Species	HEAT	Δ_{G0}	Δ_{RU}	Δ_{D2}	HEAT345-Q	ATcT value[a]
N_2	0.12	0.00	0.00	−0.07	0.05	0.00±0.00
H_2	−0.40	0.10	0.00	−0.01	−0.31	0.00±0.00
F_2	−0.22	0.00	0.00	−0.03	−0.26	0.00±0.00
O_2	0.29	0.00	0.33	−0.02	0.60	0.00±0.00
CO	−113.40	0.00	0.00	−0.04	−113.44	−113.82±0.03
C_2H_2	229.16	0.15	0.00	0.02	229.34	228.82±0.30
CCH	563.96	0.13	−0.01	0.04	564.12	563.94±0.31
CH_2	391.25	−0.13	0.00	0.02	391.15	390.96±0.27
CH	592.91	0.03	0.00	0.00	592.93	592.96±0.25
CH_3	149.74	0.32	0.00	0.01	150.07	149.97±0.10
CO_2	−392.84	0.05	0.00	−0.03	−392.83	−393.11±0.01
H_2O_2	−129.06	0.03	0.00	−0.03	−129.06	−129.78±0.07
H_2O	−239.36	0.05	0.00	−0.01	−239.32	−238.91±0.03
HCO	42.34	−0.26	0.03	−0.01	42.10	41.92±0.26
HF	−272.87	0.05	0.00	−0.01	−272.83	−272.73±0.24
HO_2	15.15	−0.13	0.02	−0.03	15.01	15.21±0.25
NO	91.22	0.00	0.09	−0.05	91.27	90.59±0.08
OH	36.73	0.04	0.00	−0.01	36.76	37.09±0.05

[a] Reference 43.

cal purposes in connection with isodesmic or nearly isodesmic reaction schemes. However, the purpose of this paper is simply to provide some measure of the relative accuracies of the various approaches, and the atomization energy approach has been adopted for this purpose. We believe that it lends itself to a more straightforward analysis, as the accuracy of all calculations ultimately rests on its ability to describe molecules relative to their constituent atoms.

III. RESULTS

A. HEAT345-Q versus original HEAT method

Table I documents the incremental differences in calculated enthalpies of formation due to the minor improvements outlined in Sec. II B. It can be seen that all three of these corrections amount to less than 0.2 kJ mol^{-1} (about 16.5 cm^{-1}) in almost all cases. The only exception is for O_2, where the consistent use of UHF reference functions in the correlation energy calculations results in a 0.33 kJ mol^{-1} difference. This difference with the value obtained with the original HEAT protocol actually results in a computed enthalpy of formation that is significantly worse—0.60 versus 0.29 kJ mol^{-1} (the correct value, of course, is zero). As will be seen in the Sec. III B, however, the effects of electron correlation are quite profound for O_2; the ΔE_{HLC} correction based on CCSDTQ is simply inadequate. In this troublesome case, pentuple excitations contribute nearly −0.5 kJ mol^{-1}, moving the result much closer to the correct value. Note that the two-electron relativistic contributions as well as the Δ_{G_0}[37] contributions to the zero-point energy are much smaller.

Overall, the net effect of these three modifications (see Table III) is essentially negligible. The statistical analysis of the data, based on mean absolute, mean signed, and root-mean-square errors (MAE, MSE, and rms, respectively) as well as the maximum absolute errors, shows that the performance of the HEAT345-Q method (that which includes the three contributions discussed in the paragraph above) is *very modestly* worse than the original HEAT approach, although this can mostly be attributed to the surprising behavior seen for O_2. However, the magnitude of the difference in performance is so small (especially when uncertainties in the ATcT values are taken into account) that it is impossible to say whether HEAT345-Q or the original HEAT method is superior. All that can be said is that the former is more complete in the theoretical sense and therefore preferred by us.

B. Effects of enhanced and simplified higher-level correlation treatments

In Table II, enthalpies of formation (at 0 K) are listed for the test suite of molecules, as calculated with HEAT345-Q, HEAT345-QP, and HEAT345-(Q) methods, while the statistical analyses of these results are summarized in Table III. A graphical summary of these results, along with those given by the original HEAT method, is given in Figs. 1 and 2. Inclusion of pentuple excitations in the higher-level correlation treatment of HEAT345-QP has a decidedly beneficial effect on the calculated enthalpies of formation, as the error measures all drop fairly significantly; the maximum error is now barely more than 0.5 kJ mol^{-1}. It is significant to note that HEAT345-QP improves upon HEAT345-Q for every molecule except N_2 that contains a multiple bond—O_2, C_2H_2, CCH, HCO, CO, CO_2, and NO—and uniformly leads to a smaller enthalpy of formation (a more stable molecule, as one would expect) in all multiply bonded systems. Also encouraging is that our results were calculated with the less rigorous, but less expensive, CCSDTQ(P) (Ref. 35) method which treats pentuple excitations in a noniterative manner. While still more expensive than CCSDTQ, it is a method that

TABLE II. Enthalpies of formation (0 K, in kJ mol^{-1}, as calculated from atomization energies using the current ATcT value for the carbon and fluorine atoms) using different variants of the HEAT procedure. Boldface values differ by more than 0.25 kJ mol^{-1} form the range spanned by the current ATcT estimates and corresponding uncertainties.

Species	HEAT345-Q	HEAT345-QP	HEAT345-(Q)	ATcT value[a]
N_2	0.05	**−0.42**	**−0.59**	0.00±0.00
H_2	**−0.31**	**−0.31**	**−0.31**	0.00±0.00
F_2	**−0.26**	**−0.42**	**−0.70**	0.00±0.00
O_2	**0.60**	0.17	0.08	0.00±0.00
CO	**−113.44**	−113.57	−113.85	−113.82±0.03
C_2H_2	229.34	229.01	229.01	228.82±0.30
CCH	564.12	563.81	563.95	563.94±0.31
CH_2	391.15	391.14	391.15	390.96±0.27
CH	592.93	592.93	592.95	592.96±0.25
CH_3	150.07	150.05	150.07	149.97±0.10
CO_2	**−392.83**	−392.99	−393.74	−393.11±0.01
H_2O_2	**−129.06**	**−129.26**	−129.46	−129.78±0.07
H_2O	**−239.32**	**−239.36**	**−239.42**	−238.91±0.03
HCO	42.10	41.94	41.57	41.92±0.26
HF	−272.83	−272.84	−272.89	−272.73±0.24
HO_2	15.01	14.80	14.54	15.21±0.25
NO	**91.27**	90.93	90.65	90.59±0.08
OH	**36.76**	**36.75**	**36.76**	37.09±0.05

[a]Reference 43.

can be used in some cases where a full-blown CCSDTQP calculation is not possible. As can be seen from Table III, this method provides an overall accuracy quite similar to that of CCSDTQP itself.

Let us turn to the results obtained with the computationally simpler HEAT345-(Q) method, which exploits the newly formulated CCSDT(Q) treatment of electron correlation. As mentioned earlier in this manuscript, calculations at the CCSDT(Q) level require considerably less time than those done at the CCSDTQ level. The computational requirements of the latter involve the iterative solution to a coupled system of nonlinear equations, the cost of which rises with the tenth power of the basis set dimension. On the other hand, CCSDT(Q) features an iterative step with an n^8 dependence and a single noniterative enegy evaluation with an n^9 scaling. The result, which is perhaps best seen in Figs. 1 and 2, is that the performance of HEAT345-(Q) is seen to be nearly as good as the original HEAT method.[38] It is remarkable that the CCSDT(Q) treatment of higher level correlation is, in general closer to that of the HEAT345-QP method than to HEAT345-Q. In fact, for *all* examples in which the pentuple correction (as measured by the difference between HEAT345-QP and HEAT345-Q) exceeds 0.1 kJ mol^{-1}, HEAT345-(Q) is closer to HEAT345-QP than to HEAT345-Q. In these cases, the CCSDT(Q) treatment of higher-level correlation tends to overshoot CCSDTQ in much the same way that CCSD(T) tends to overshoot CCSDT. And again, in the same sense that CCSD(T) often gives fortuitously better results than does the more theoretically complete CCSDT approach, so does CCSDT(Q) occasionally "outperform" CCSDTQ. In the overall statistical analyses of the various methods, however, CCSDT(Q) does less well than both HEAT345-Q and HEAT345-QP, owing to its tendency to overshoot fairly badly when the CCSDTQP/CCSDTQ difference is large, as seen for F_2, N_2, and HO_2. Despite this, it is truly remarkable that the great computational simplification resulting from the use of CCSDT(Q) does not cause a serious degradation in the accuracy. Indeed, there are still no examples in which HEAT-(Q) has an error larger than 0.75 kJ mol^{-1}.

TABLE III. Statistical analysis of errors obtained with the methods defined in this paper, using the atomization energy approach. Values are given in kJ mol^{-1}. MAE: mean absolute error; MSE: mean signed error; rms: root-mean-square error; MER: maximum error. The rows are ordered by the relative cost of the various methods.

Method	MAE	MSE	rms	MER
HEAT345-(Q)	0.29	−0.19	0.37	−0.70
HEAT	0.30	0.09	0.35	0.72
HEAT345-(Q)	0.31	0.12	0.37	0.72
HEAT345-Q(P)	0.27	−0.03	0.31	0.51
HEAT345-QP	0.25	−0.04	0.29	0.52

C. Discussion of remaining errors

It is interesting to note that in five of the six cases where the sophisticated HEAT-QP procedure differs from the corresponding ATcT estimate by more than 0.25 kJ mol^{-1} (see Table II), the calculation tends to give an enthalpy of formation that is *below* the ATcT estimate. That is, theory is overestimating the stability of these molecules—the underlying atomization energies are apparently a bit too large. While an investigation of the causes of this behavior on a case-by-case basis is beyond the scope of this paper, some general discussion of the possible sources is desirable. Perhaps the most obvious shortcoming is the use of the cc-pVDZ basis set in evaluating the higher-level correlation correction. In work on the W3 model, Boese *et al.*[1] advocated scaling the higher-level correction (calculated by CCSDTQ with the cc-pVDZ basis as well in their work) by 1.2532. However, as the higher-level correlation correction to atomization energies is uniformly positive (typically about 2 kJ mol^{-1} for multiply bonded or other "difficult" systems such as F_2 and <1 kJ mol^{-1} for hydrides) scaling this correction would tend towards increasing the predicted stability of molecules such as F_2 and N_2, thereby increasing the magnitude of the discrepancy. While the scale factor advocated in Ref. 1 is quite speculative, and based on a small amount of data, it is reasonable to assume that the use of the unscaled cc-pVDZ basis set results in our work does indeed tend towards artificially high molecular enthalpies of formation (smaller atomization energies). Indeed, viewed from this prospective the agreement between the results listed in Table II and the ATcT values is quite remarkable and suggests that any attempt to pinpoint a single systematic shortcoming of the HEAT-QP model that results in the "poor" performance noted for F_2, N_2, H_2O_2, HO, and H_2O is probably futile.

Based on a few calculations that we have carried out with HEAT456 approaches, a fairly systematic trend has

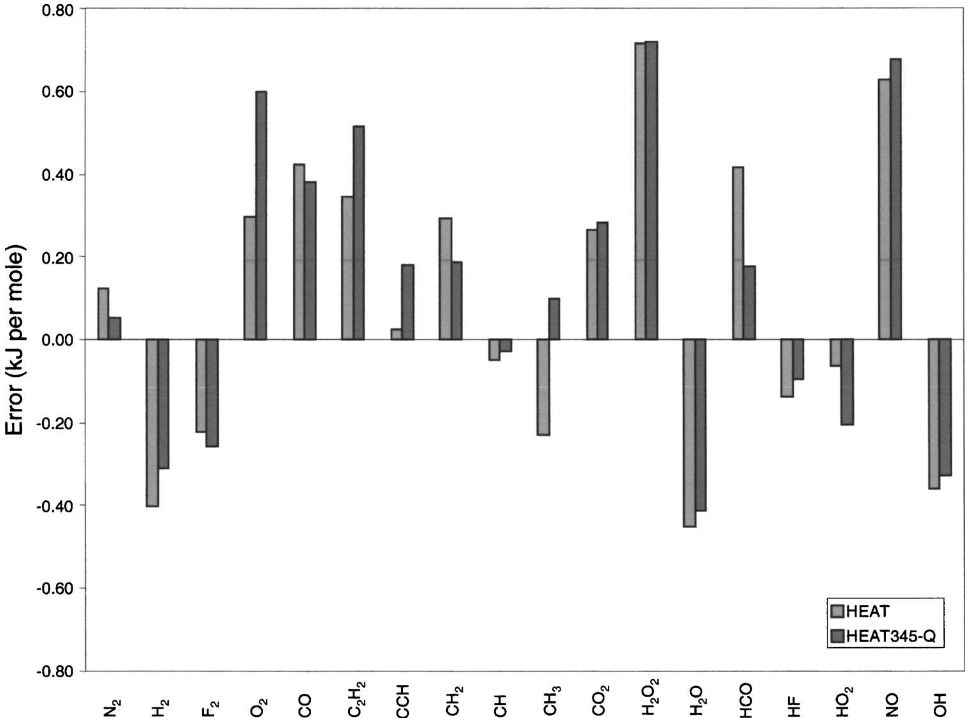

FIG. 1. Errors in the enthalpies of formation for HEAT and HEAT345-Q in kJ mol^{-1} (Table I).

emerged. Specifically, HF-SCF atomization energies obtained from the 456-based extrapolation tend to be slightly more positive than their 345 counterparts,[39] while the 456 extrapolations appear to always give larger (in magnitude, hence more negative) correlation contributions. From a limited number of examples, it appears that the overshooting of the correlation extrapolation based on the 345 series is larger than the undershooting (in terms of total atomization energy) of the HF-SCF contributions. Hence, the overall atomization energies calculated from the 456 sequence are smaller (more positive molecular enthalpies of formation) than those from the 345 sequence upon which the original HEAT method was based. In the two cases that have been studied, CO and CN, the differences in the HF-SCF and CCSD(T) atomization energy contributions are 730.16 versus 729.87 kJ mol^{-1} [HF-SCF] and 356.53 versus 356.91 kJ mol^{-1} [CCSD(T)] [CO] and 379.48 versus 379.38 kJ mol^{-1} [HF-SCF] and 371.03 versus 371.64 kJ mol^{-1} [CCSD(T)] [CN] for the 456 and 345 extrapolations, respectively. The corresponding total atomization energies [exclusive of all other effects apart from HF-SCF and CCSD(T) energy] at the CCSD(T) level are thus 1086.78 [345] and 1086.70 kJ mol^{-1} [456] for CO and

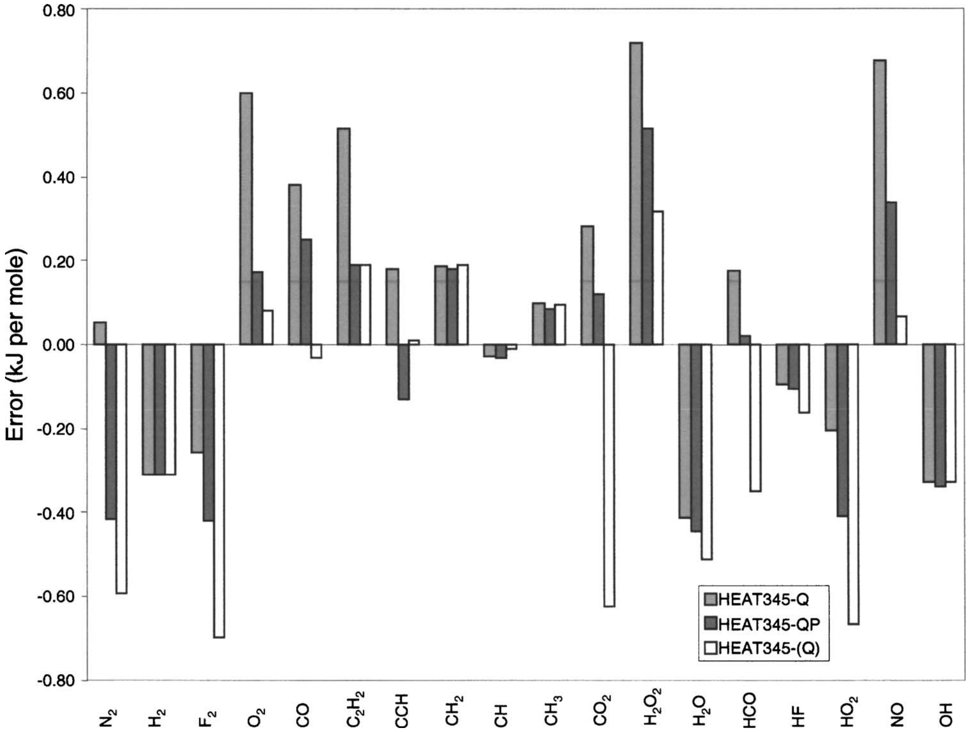

FIG. 2. Errors in the enthalpies of formation for HEAT345-Q, HEAT345-QP, and HEAT345-(Q) in kJ mol^{-1} (Table II).

751.02 [345] and 750.50 kJ mol^{-1} [456] for CN. In the course of this research, similar calculations have been carried out for H_2 and OH. The same trend is observed; total atomization energies with the 345- and 456-based extrapolations are 448.53 and 448.44 kJ mol^{-1} for OH and 458.22 and 458.15 kJ mol^{-1} for H_2.

It is interesting to note that a substantial part of the error for H_2 lies in the fact that the diagonal Born-Oppenheimer approximation is calculated at the HF-SCF level in HEAT. In the work of Valeev and Sherrill,[23] the correlation correction to the diagonal Born-Oppenheimer correction (DBOC) for H_2 (as calculated with configuration interaction in the singles and doubles approximation, which is exact for H_2 albeit subject to basis set limitations) is +13 cm^{-1} (0.16 kJ mol^{-1}). This value agrees to within 0.1 cm^{-1} with the essentially exact result of Kolos and Wolniewicz and can be regarded as the correct correlation contribution. Adding the DBOC correlation contribution to the HEAT-Q value of −0.30 kJ mol^{-1} gives a value of −0.14 kJ mol^{-1}. Most of the remaining errors arise from the extrapolation issues discussed in the preceding paragraph, and $\Delta_f H^o$ is reduced to just −0.07 kJ mol^{-1} when the 456-based extrapolations are used instead of those using the 345 sequence. Finally, the use of a zero-point correction calculated from the experimental spectroscopic constants[40] of 26.08 kJ mol^{-1} further improves the result to −0.05 kJ mol^{-1}, which is roughly −5 cm^{-1}. It is interesting to note that this result is virtually identical, and carries the same sign, as the best calculation by Kolos and Wolniewicz.[41] It would be foolish to further analyze this error.

At this point, it is appropriate to comment that the magnitude of the energy differences and "errors" being discussed in this section is—from the point of view of typical quantum chemistry—preposterously small. The HEAT345-QP method has a rms and maximum error of 0.29 and 0.52 kJ mol^{-1}. To put the smallness of these numbers in context (at least for those of a certain predilection), these results correspond to 24 and 43 *wave numbers*. Nevertheless, we attempt here to make some general comments about errors in the HEAT procedures. The principal shortcomings are inherent limitations of the extrapolation procedures, particularly with the 345 sequence; errors in the zero-point energy calculations, even for rigid molecules; and the approximation inherent in using cc-pVDZ basis sets in attempting to account for the higher-level (beyond CCSDT) correlation corrections. It appears that the 345-based extrapolations uniformly lead to an overestimation of atomization energies and the zero-point correction error tends to underestimate them.[42] The work of Boese *et al.*[1] as well as in our laboratories indicates that the cc-pVDZ basis set tends to underestimate correlation corrections. There are thus two principal sources of error that tend to overestimate the stability of molecules, while it is just the zero-point energy term that acts in the other direction. This is, quite probably, the reason why the mean signed error of HEAT345-QP is negative and the cause of the behavior discussed in the first sentence of this section.

IV. SUMMARY

This paper is concerned with theoretical enthalpies of formation, as based on atomization energy calculations. This strategy lends itself to a more straightforward analysis than reaction-based schemes, despite the fact that atomization energies are more difficult, in principle, to calculate. By extending the HEAT method to include quintuple excitations, perhaps the most accurate theoretical model chemistry to date has been developed; this accuracy is amply demonstrated by comparisons with the only "experimental" database of similar precison—the ATcT results from Ruscic and co-workers. In addition, it is shown that the use of an approximate treatments of high-level excitations does not seriously degrade the accuracy of the methods. In all methods investigated, including the original HEAT scheme, enthalpies of formation (at 0 K) remain below 1 kJ mol^{-1} for all of the molecules in our test suite.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation, the Robert A. Welch Foundation (Y.J.B., J.V., and J.F.S.), the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie (C.M. and J.G.), and Scientific Research Fund of Hungary (OTKA) [Grant Nos. D048583, T047182, and T047185 to three of the authors (M.K., P.G.S., and A.G.C., respectively)]. The research presented in this paper is part of a current and future work by a Task Group of the International Union of Pure and Applied Chemistry (2000-013-2-100) to determine structures, vibrational frequencies, and thermodynamic functions of free radicals of importance in atmospheric chemistry. We also thank Branko Ruscic (Argonne National Laboratory) for his enthusiastic support of our research.

[1] A. D. Boese, M. Oren, O. Atasolyu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

[2] J. M. L. Martin, (private communication).

[3] K. A. Peterson, D. E. Woon, and T. H. Dunning, J. Chem. Phys. **100**, 7410 (1994); D. Feller and D. A. Dixon, J. Phys. Chem. A **104**, 3048 (2000); J. Chem. Phys. **115**, 3484 (2001); D. A. Dixon, W. A. de Jong, K. A. Peterson, and J. S. Francisco, J. Phys. Chem. A **106**, 4725 (2002); D. Feller, K. A. Peterson, W. A. de Jong, and D. A. Dixon, J. Chem. Phys. **118**, 3510 (2003).

[4] A. G. Császár, W. D. Allen, and H. F. Schaefer III, J. Chem. Phys. **108**, 9751 (1998); A. G. Császár, M. L. Leininger, and V. Szalay, *ibid.* **118**, 10631 (2003).

[5] A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[6] As is well appreciated, "experimental" enthalpies of formation for most molecules do not come from a single experiment. Instead, the best values usually come from taking several experimental results into consideration in a critical data evaluation.

[7] G. von Laszewski, B. Ruscic, P. Wagstrom *et al.*, in *Lecture Notes in Computer Science*, edited by M. Parashar (Springer, Berlin, 2002), Vol. 2536, pp. 25–38; B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszevski, S. J. Bittner, S. G. Nijsure, K. A. Amin, M. Minkoff, and A. F. Wagner, J. Phys. Chem. A **108**, 9979 (2004).

[8] The term theoretical model chemistry was coined by People and is given to an approach in which all molecules are treated at a specific and consistent level of theory.

[9] J. D. Cox, D. D. Wagman, and V. A. Medvedev, *CODATA Key Values for Thermodynamics* (Hemisphere, New York, 1989), see also http://www.codata.org/codata/databases/key1.html

[10] L. V. Gurvich, I. V. Veyts, and C. B. Alcock, *Thermodynamic Properties of Individual Substances*, 4th ed. (Hemisphere, New York, 1989).

[11] Well-known tabulations of molecular enthalpies of formation include M. W. Chase, Jr., J. Phys. Chem. Ref. Data **6**, 27 (1998).

[12] B. Ruscic, M. Kállay, A. G. Császár, and J. F. Stanton (unpublished)

[13] G. Tasi, R. Izsák, G. Matisz, A. G. Császár, M. Kállay, B. Ruscic, and J. F. Stanton, Angew. Chem., Int. Ed. Ed. (submitted).

[14] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Molecular Orbital Theory* (Wiley, New York, 1986), p. 298.

[15] D. Feller, J. Chem. Phys. **96**, 6104 (1992).

[16] R. A. Kendall and T. H. Dunning, J. Chem. Phys. **96**, 6796 (1992).

[17] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[18] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[19] J. Noga and R. J. Bartlett, J. Chem. Phys. **86**, 1041 (1987); G. E. Scuseria and H. F. Schaefer, Chem. Phys. Lett. **132**, 382 (1988); J. D. Watts and R. J. Bartlett, J. Chem. Phys. **93**, 6104 (1993).

[20] S. A. Kucharski and R. J. Bartlett, Theor. Chim. Acta **80**, 387 (1991); J. Chem. Phys. **97**, 4282 (1992); N. Oliphant and L. Adamowicz, *ibid.* **94**, 1229 (1991).

[21] R. D. Cowan and M. Griffin, J. Opt. Soc. Am. **66**, 1010 (1976); R. L. Martin, J. Phys. Chem. **87**, 750 (1983).

[22] H. Sellers and P. Pulay, Chem. Phys. Lett. **103**, 463 (1984); N. C. Handy, Y. Yamaguchi, and H. F. Schaefer, J. Chem. Phys. **84**, 4481 (1986); N. C. Handy and A. M. Lee, Chem. Phys. Lett. **252**, 425 (1996); W. Kutzelnigg, Mol. Phys. **90**, 909 (1997).

[23] E. F. Valeev and C. D. Sherrill, J. Chem. Phys. **118**, 3921 (2003).

[24] I. M. Mills, in *Modern Spectroscopy: Modern Research*, edited by K. N. Rao and C. W. Matthews (Academic, New York, 1972), pp. 115–140.

[25] An error appears in note 62 of the original paper, in which it is stated that ROHF calculations were also used for the CN molecule. The numbers in the tables, however, are based on the ZPE calculated from UHF for CN. In this paper, the ROHF results are used.

[26] P. G. Szalay, J. Vázquez, C. Simmons, and J. F. Stanton, J. Chem. Phys. **121**, 7624 (2004).

[27] V. Barone, J. Chem. Phys. **120**, 3059 (2004).

[28] V. Barone and C. Minishino, J. Mol. Struct.: THEOCHEM **330**, 365 (1995).

[29] M. S. Schuurman, S. R. Muir, W. D. Allen, and H. F. Schaefer, J. Chem. Phys. **120**, 11586 (2004).

[30] Numerical results from all three implementations appear to agree with each other, and a sum-over-states direct summation program [J. Vázquez and J. F. Stanton, Mol. Phys. **104**, 377 (2006)]. The correct equation is

$$G_0 = -\frac{1}{4}\sum_{\alpha} B_e^{\alpha} + \frac{1}{64}\sum_{k} \phi_{kkkk} - \frac{7}{576}\sum_{k} \frac{\phi_{kkk}^2}{\omega_k} + \frac{3}{64}\sum_{k\neq l} \frac{\omega_l \phi_{kkl}^2}{4\omega_k^2 - \omega_l^2} - \frac{1}{4}\sum_{k<l<m} \frac{\phi_{klm}^2 \omega_k \omega_l \omega_m}{D_{klm}} - \frac{1}{8}\sum_{k\neq l}\sum_{\alpha} B_e^{\alpha}(\zeta_{kl}^{\alpha})^2.$$

See Ref. 5 for a definition of the symbols in this equation. In Ref. 5, the fifth and sixth terms were in error. Furthermore, the kinetic energy elements in Ref. 28 were in error, which was addressed to some degree in Ref. 27.

[31] This is true, of course, for full configuration interaction wave functions and tends to be an excellent approximation for coupled-cluster methods of all types that include single excitations. Nevertheless, this assumption remains as an undesirable approximation of our original work and is rectified here.

[32] W. Klopper, J. Comput. Chem. **18**, 20 (1997).

[33] J. Vázquez, J. F. Stanton, and J. M. L. Martin (unpublished)..

[34] Y. J. Bomble, J. F. Stanton, M. Kállay, and J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

[35] M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

[36] S. E. Wheeler, K. A. Robertson, W. D. Allen, H. F. Schaefer, Y. J. Bomble, and J. F. Stanton (unpublished).

[37] A strong Fermi resonance exists in CO2 between $\nu_1(\Sigma_g^+)$ and $2[\nu_2(\Pi_u)]$, which results in a large value of G_0 (-0.86 kJ/mol) when calculated according to the formula in Ref. 30. However, one may alternatively skip the corresponding resonance denominator, which affects the ZPE calculated from the term level expression without G_0 as well as G_0 itself (the total ZPE, of course, is not affected by resonances between excited vibrational levels). In either case, the total zero-point energies are the same. The G_0 contribution for CO2 in Table I is calculated with the corresponding resonance denominator omitted, as was the ZPE given in Ref. 5.

[38] Alternatives to CCSDT(Q) for the approximate treatment of quadruple excitations include the CCSDT[Q], CCSDT$(\mathrm{Q})_\Lambda$, CCSDTQ-1a, CCSDTQ-1b, CC4, and CCSDTQ-3 schemes (see Ref. 35), of which the first two are, in fact, slightly cheaper to apply than the others. Numerical tests indicate that for total energies, the CCSDT$(\mathrm{Q})_\Lambda$, CCSDTQ-1b, and CC4 approaches perform better than CCSDT(Q), while performance of the CCSDTQ-3 method is perhaps comparable (see Ref. 35). The CCSDT[Q] and the CCSDTQ-1a methods, on the other hand, are less accurate than any of these approaches. rms errors of the calculated heats of formation for the dataset were also evaluated using all of these methods and found to be (in kJ mol^{-1}) 1.17 (CCSDT[Q]), 0.31 (CCSDT$(\mathrm{Q})_\Lambda$), 1.12 (CCSDTQ-1a), 0.32 (CCSDTQ-1b), 0.33 (CC4), and 0.69 (CCSDTQ-3). The conclusions concerning the performance of the approximate quadruples methods for heats of formation are similar to those for total energies; however, the difference between CCSDT(Q) and the more expensive methods is, ultimately, negligible (their use is consequently not justified). We note in passing that heats of formation were also computed using CCSDTQ as well as the aforementioned approximate methods with the cc-pVTZ basis set. However, no significant improvement has been achieved with respect to the experimental values, meaning that again the extra computational cost is not justified.

[39] It is interesting to note that both the HF-SCF and CCSD(T) energies extrapolated with the 345 sequence are more negative than their 456 counterparts, although the former extrapolation tends to give smaller atomization energies in one case (HF-SCF) and larger in the other (CCSD(T)). The latter is rather obvious and sensible: correlation energy always tends to increase atomization energies, and a method that tends to overestimate correlation energies would tend to overestimate binding energies, if the extrapolation error were somewhat systematic. However, for the HF-SCF cases, the extrapolation error (as measured by the difference between 345- and 456-based extrapolations) is *larger* for free atoms than those in molecules in the cases we have investigated, which is in turn responsible for the underestimated atomization energies. Differences between 345- and 456-based extrapolations(in μH) are 269 (oxygen atom), 111 (nitrogen atom), 35 (carbon atom), 22 (hydrogen atom) [atoms], 239 (OH), 109 (CN), and 190 (CO). It is interesting, indeed a bit odd, that the error for the oxygen atom is the largest.

[40] Experimental ZPE was calculated using Eq. (5) and experimental values for ω_e, $\omega_e x_e$, the equilibrium rotational constant, B_e and the rotation-vibration interaction constant, α_e [M. W. Chase, Jr., J. Phys. Chem. Ref. Data **6**, 27 (1998)]..

[41] W. Kolos and L. Wolniewicz, J. Chem. Phys. **49**, 404 (1968).

[42] In many cases, vibrational frequencies are slightly overestimated in cc-pVQZ calculations, with a consequent overestimation of the zero-point energy.

[43] B. Ruscic (private communication); unpublished results from Active Thermochemical Tables ver. 1.25 using the Core (Argonne) Thermochemical Network ver. 1.048..