13 September 1996

ELSEVIER

Chemical Physics Letters 259 (1996) 669–678

CHEMICAL PHYSICS LETTERS

Ab initio total atomization energies of small molecules – towards the basis set limit

Jan M.L. Martin [1]

The Institute of Chemistry, The Hebrew University of Jerusalem, Givat-Ram, Jerusalem 91904, Israel

Received 27 June 1996; in final form 31 July 1996

Abstract

Total atomization energies (TAEs) of about a dozen small polyatomic molecules have been calculated at the CCSD(T) level using correlation consistent basis sets of up to [7s6p5d4f3g2h] quality, and including core correlation. Single bond energies are close to convergence with such basis sets: for multiple bonds, extrapolation remains mandatory. 'Augmented' basis sets specifically designed for anions yield improved atomization energies for highly polar molecules. An extrapolation of the form $A + B/(l + 1/2)^4 + C/(l + 1/2)^6$, with l the maximum angular momentum in the basis set, is found to yield TAEs accurate, on average, to 0.5 kcal/mol using at most [sdpfg] basis sets. Using [spdfgh] basis sets and a small additivity correction for triple bonds, this can be reduced to 0.2 kcal/mol.

1. Introduction

The calculation of total atomization energies (TAEs; $\sum D_e$ values) – and thermochemical data in general – to chemical accuracy (commonly defined as 1 kcal/mol) has been a long-standing goal of quantum chemistry. Obtaining these directly has been a losing battle due to the well-known slow convergence behavior of the electron–electron interaction, which asymptotically [1] converges as $(l+1/2)^{-4}$ (see also Refs. [2,3] for detailed treatments). For example, in a pioneering basis set saturation study [4] on N_2 it was found that even a multireference configuration interaction (MRCI) calculation in a [6s5p4d3f2g1h1i] basis set, after corrections for core correlation and basis set superposition error, does not come closer than 2.4 kcal/mol to the experimental value.

One solution for this problem has been the use of wavefunctions that explicitly contain interelectronic distances, such as the MP2-R12 method of Kutzelnigg and Klopper [5] and Gaussian-type geminals, which have mainly found application in small systems, but which have been applied, e.g., to Ne [6] and H_2O [7] by Szalewicz and co-workers, and recently by Persson and Taylor [8].

Another has been to introduce various empirical correction schemes. The most popular such scheme is G2 (Gaussian-2) theory [9], in which a total energy is obtained from basis sets of spdf quality, additivity approximations involving fourth-order perturbation theory and approximate coupled cluster energies, and an empirical correction formula of the form

$$\Delta E = An_{\text{unpaired}} + (A + B)n_{\text{paired}}, \qquad (1)$$

which depends on the numbers of unpaired and paired

[1] On sabbatical from: Limburgs Universitair Centrum, Department SBG, Universitaire Campus, 3590 Diepenbeek, Belgium and University of Antwerp (UIA), Institute for Materials Science, Department of Chemistry, Universiteitsplein 1, 2610 Wilrijk, Belgium.

PII S0009-2614(96)00898-6

electrons. Determining A and B from a large sample of experimental atomization energies, a mean absolute error of about 1.1 kcal/mol could be achieved.

A less popular scheme has been Martin's 3-parameter additivity correction [10,11]

$$\Delta E_{\text{correction}} = a_\sigma n_\sigma + b_\pi n_\pi + (n_\sigma + n_\pi + n_{\text{lone pair}})c_{\text{pair}}, \quad (2)$$

in which $n_\sigma, n_\pi, n_{\text{lone pair}}$ represent the numbers of σ bonds, π bonds, and lone pairs, respectively, and the coefficients $a_\sigma, b_\pi, c_{\text{pair}}$ are specific for the basis set, electron correlation treatment, and (level of theory used for the) reference geometry. (They are determined by least-squares fitting to a fairly small sample of accurately known TAEs.) For coupled cluster calculations using spdf basis sets, performance of the scheme is somewhat better than G2 theory; for spdfg basis sets, mean absolute errors as low as 0.46 kcal/mol can be reached. Experimentation with even larger basis sets [12] appears to indicate that the point of diminishing returns has been reached.

Both schemes have their empirical nature in common; an additional disadvantage of Eq. (2) – which does allow us to go beyond 'mere' chemical accuracy – is that the molecule must exhibit clearly defined directional bonding, which severely restricts applications to, e.g., boranes and transition metal complexes.

A third alternative, which is neither empirical nor requires calculations with explicitly correlated wavefunctions, and does not require connectivity information, is extrapolation from a systematic sequence of energies using progressively larger basis sets. The most widely used such formula was proposed by Feller [13] and extensively used by the Dunning group (e.g. Refs. [14,15]):

$$E(l) = E_\infty + A\exp(-Bl), \quad (3)$$

where l is the maximum angular momentum quantum number in the basis set.

In the present paper, we investigate the convergence behavior of computed TAEs for about a dozen molecules where the experimental value is precisely known. The basis sets will include up to h functions, and, in the case of N_2, up to i functions. Secondly, we assess the performance of Feller's extrapolation method. Thirdly, a new extrapolation method will be proposed which is shown to perform markedly better than Feller's. Fourthly, it will be shown that, in combination with 'augmented' basis sets that include up to h functions, a target accuracy of better than 1 kJ/mol (0.23 kcal/mol) is within reach.

2. Methods

All calculations were carried out using the MOLPRO 96[2] ab initio package running on a DEC Alpha TurboLaser 8200 5/300 at the Hebrew University, and, for the most demanding calculations in terms of disk space, on the Cray C90 at San Diego Supercomputer Center.

The CCSD(T) electron correlation method [16], as implemented for open-shell systems by Hampel et al. [17], has been used throughout. The acronym stands for coupled cluster with all single and double substitutions [18] augmented by a quasiperturbative account for triple excitations [16]. From extensive studies (see Ref. [19] for a review) this method is know to yield total energies close to the exact basis set correlation energy. Specifically for the N_2 molecule, the difference in the total atomization energy with an extensive CASSCF/CI calculation in the cc-pVQZ basis set was only 0.3 kcal/mol (CCSD(T) being the lowest).

Only the valence electrons were correlated in the various calculations except when indicated otherwise. The effect of core correlation was taken as the difference between calculations with valence electrons only correlated, and all electrons correlated, using the Martin–Taylor [20] core correlation basis set. This was previously found [21] to recover essentially the entire core correlation effect for those molecules where more accurate information was available.

All other basis sets are from the 'correlation consistent' family of Dunning and co-workers [22,23]. The following have been considered:

– the cc-pVnZ basis sets (correlation consistent polarized valence n-tuple zeta), with n = D,T,Q,5,6 (double, triple, quadruple, quintuple, hextuple). These basis sets consist of [3s2p1d/2s1p], [4s3p2d1f/3s2p1d], [5s4p3d2f1g/4s3p2d1f], [6s-

[2] MOLPRO 96 is an ab initio MO package by H.J. Werner and P.J. Knowles, with contributions from J. Almlöf, R.D. Amos, M.J.O. Deegan, S.T. Elbert, C. Hampel, W. Meyer, K.A. Peterson, R.M. Pitzer, A.J. Stone, P.R. Taylor and R. Lindh.

5p4d3f2g1h/5s4p3d2f1g], and [7s6p5d4f3g2h1i/6s5p4d3f2g1h] contractions, respectively, where the part behind the slash indicates the basis set for hydrogen and that before the basis set for first-row atoms

– the aug-cc-pVnZ basis sets, in which one 'diffuse' or 'soft' (i.e. low-exponent) function of each angular momentum is added. The exponents were optimized for the anions of the corresponding atoms [23]

– the aug′-cc-pVnZ basis sets (acronym coined by Del Bene [24]), in which the soft functions are only added to the heavy atoms and omitted on the hydrogen atoms

It was previously established [11] that different families of large basis sets with similar contracted sizes (e.g. atomic natural orbital [25]) yielded results for the total atomization energy of the same quality as the corresponding correlation consistent basis sets. Of the correlation-consistent core correlation basis sets [26], the cc-pCVQZ basis set should yield results of comparable quality as the Martin–Taylor one [21].

For the cc-pVnZ basis sets (n = D,T,Q), energies were calculated at the true CCSD(T)/cc-pVnZ equilibrium geometry: for n = 5 and up, and for the aug-cc-pVnZ and aug′-cc-pVnZ (n = D, T, Q, 5) basis sets, CCSD(T)/cc-pVQZ reference geometries were used. Full geometry optimization was performed in the calculations with the Martin–Taylor basis set.

The anharmonic zero-point energies (ZPEs) were obtained from experimental anharmonicity constants (where full accurate sets are available) or accurate ab initio anharmonic force field studies (see Table 1 for detailed references).

The present calculations are all nonrelativistic. For the present purpose, the main consequence is that the spin–orbit components of B(^{2}P), C(^{3}P), O(^{3}P), and F(^{2}P) are all degenerate, which of course is not the case in Nature. In order to permit direct comparison, the effect of spin–orbit splitting in the atoms should therefore be extracted out of the experimental TAEs. For example, for every oxygen atom present, TAE should be increased by $[E(^3P_0)+3E(^3P_1)+5E(^3P_2)]/9-E(^3P_0)$ (see e.g. Ref. [27]). The largest two effects of atomic spin–orbit splitting in the present work are 0.8 kcal/mol (for F_2) and 0.6 kcal/mol (for CO_2), clearly on the order of magnitude of the accuracy we are trying to achieve.

3. Results and discussion

CCSD(T) total energies with the different basis sets are given in the top half of Table 2. After using the core correlation contributions (Table 1) as an additive correction, this leaves the residual errors given in the bottom half of Table 2. While the mean absolute error for the CCSD(T)/cc-pV5Z+core results does drop to 2.42 kcal/mol (maximum 4.64 kcal/mol), and still somewhat further to 2.14 (maximum 4.04) kcal/mol for CCSD(T)/aug′-cc-pV5Z+core, this is clearly outside the definition of 'chemical accuracy', even though this kind of accuracy on 'raw' (uncorrected) results would have been unimaginable five years ago due to computational power limitations.

More detailed investigation does reveal that single bonds are, on average, calculated only about 0.5 kcal/mol too weak, which suggests approach to convergence. (With the cc-pVQZ basis set, this effect is less systematic.) Multiple bonds are however still calculated 2–3 kcal/mol short of experiment, suggesting that they are somewhat further from convergence even though, again, the effects appear to be somewhat more systematic than with the cc-pVQZ basis set. As expected, the difference between cc-pVnZ and aug′-cc-pVnZ errors diminishes as n increases, with the aug′ basis sets being obviously superior.

We have repeated part of the calculations with the aug-cc-pVnZ basis sets. While there are significant differences with aug′-cc-pVnZ for n = D, these dwindle to $\sim$ 0.1 kcal/mol or less for n = 5. This suggests that the contribution of diffuse functions on hydrogen in the aug-cc-pVDZ basis is largely due to basis set superposition error.

All in all, the results suggest that some form of empirical correction and/or extrapolation will remain indispensible for some time.

The first possibility is the exponential extrapolation, eq. (3). A minimum of three different values of l is required to enable the use of the formula. In the present work, we will denote with Feller(lmn) the E_∞ obtained from CCSD(T)/cc-pVlZ, CCSD(T)/cc-pVmZ, and CCSD(T)/cc-pVnZ results, and with aug′-Feller(lmn) the corresponding value obtained with aug′ basis sets.

The extrapolated values are shown in Table 3. As it turns out, Feller(DTQ), Feller(TQ5), aug′-Feller(DTQ), and aug′-Feller(TQ5) results, which

Table 1
Experimental $\sum D_0$ values, zero point energies, $\sum D_0$ values and computed core correlation contributions. All units are kcal/mol. For detailed references, see Ref.[12]

	$\sum D_0$	ZPE	$\sum D_e$	w/o spin-orbit	Core corr.
C_2H_2	388.90(24)	16.46	405.36(24)	405.53(24)	2.44
CH_4	392.51(14)	27.6	420.15(14)	420.23(14)	1.25
CO	256.16(12)	3.11	259.27(12)	259.58(12)	0.96
CO_2	381.91(6)	7.24	389.15(6)	389.68(6)	1.78
H_2	103.27(0)	6.21	109.48(0)	109.48(0)	0.00
H_2O	219.35(2)	13.25	232.60(2)	232.83(2)	0.38
HF	135.33(17)	5.85	141.18(17)	141.57(17)	0.18
NH_3	276.73(10)	21.33	298.06(10)	298.06(10)	0.66
N_2	225.06(3)	3.36	228.42(3)	228.42(3)	0.85
H_2CO	357.25(16)(12)	16.53	373.78(16)	374.09(16)	1.32
F_2	36.94(10)	1.30	38.24(10)	39.01(10)	−0.07
HNO	196.85(6)	8.56	205.41(6)	205.64(6)	0.48
N_2O	263.61(10)	6.77	270.38(10)	270.60(10)	1.26

Table 2
Experimental $\sum D_e$ values and errors for directly computed CCSD(T) energies. All values in kcal/mol

	Exp.	Regular				Aug′			
		VDZ	VTZ	VQZ	V5Z	AVDZ	AVTZ	AVQZ	AV5Z
C_2H_2	405.53(24)	34.25	12.27	6.10	4.22	35.19	11.26	5.59	4.03
CH_4	402.23(14)	24.78	7.21	3.16	2.03	24.11	6.70	3.03	1.99
CO	259.58(12)	18.12	7.90	3.53	2.25	19.16	7.58	3.27	2.06
CO_2	389.68(6)	33.13	13.87	6.11	3.75	32.47	12.46	5.30	3.34
H_2	109.48(0)	5.87	1.10	0.36	0.14	6.02	1.10	0.36	0.14
H_2O	232.83(2)	23.94	7.70	2.87	1.22	14.77	5.09	1.77	0.93
HF	141.57(17)	15.12	4.52	1.51	0.52	7.79	2.49	0.69	0.31
NH_3	298.06(10)	30.37	9.86	4.01	2.04	23.00	7.36	3.00	1.79
N_2	228.42(3)	27.79	11.97	5.55	3.30	27.31	10.43	4.62	2.83
H_2CO	374.09(16)	28.78	10.29	4.34	2.47	26.52	8.85	3.66	2.20
F_2	39.01(10)	11.82	4.18	2.29	1.44	9.52	2.56	1.33	0.94
HNO	205.64(6)	27.46	10.91	4.74	2.54	23.09	8.69	3.54	2.06
N_2O	270.60(10)	40.08	16.87	7.70	4.64	38.20	14.48	6.35	4.04
mean abs. error		24.73	9.13	4.02	2.35	22.09	7.62	3.27	2.05
max. abs. error		40.08	16.87	7.70	4.64	38.20	14.48	6.35	4.04

should ideally all be nearly equal, may differ quite appreciably, particularly for polar compounds. For CO_2, for example, aug′-Feller(DTQ) and aug′-Feller(TQ5) differ by no less than 1.21 kcal/mol, one value too low, the other being too high. The mean absolute error for aug′-Feller(TQ5) is 0.71 kcal/mol, and the maximum (for N_2O) 1.89 kcal/mol. While this does fall within the goal of chemical accuracy, the fact that one cannot do any better than this with a [7s6p5d4f3g2h/5s4p3d2f1g] basis set is hardly encouraging. Actually, Martin's empirical correction scheme [10,11] achieves better results with much smaller basis sets.

The main problem with Eq. (3) is that it is purely empirical – derived from the observed convergence behavior [22] of the contributions of progressively higher angular momenta to total energies, as well as of those of extra functions of one particular angular

Table 3
Experimental $\sum D_e$ values (kcal/mol) and errors (kcal/mol) in extrapolated CCSD(T) calculated values after correction for core correlation

≡ correction	Exp.	Feller(DTQ)	Feller(TQ5)	Schwartz4(TQ)	Schwartzα(TQ5)	√	Schwartz4(Q5)	√	Schwartz6(TQ5)	√
Regular cc-pV*n*Z basis sets										
C_2H_2	405.53(24)	1.24	0.97	0.09	0.37	−0.03	0.26	−0.04	0.32	0.02
CH_4	402.23(14)	0.74	0.35	−0.43	0.04	0.04	−0.13	−0.13	−0.02	−0.02
CO	259.58(12)	−0.16	0.77	0.06	0.35	−0.05	0.24	−0.06	0.29	−0.01
CO_2	389.68(6)	−0.34	0.96	−0.15	0.18	−0.22	0.05	−0.25	0.13	−0.17
H_2	109.48(0)	0.23	0.06	−0.07	0.00	0.00	−0.03	−0.03	−0.01	−0.01
H_2O	232.83(2)	0.44	−0.01	−0.31	−0.64	−0.64	−0.49	−0.49	−0.56	−0.56
HF	141.57(17)	0.18	−0.15	−0.41	−0.52	−0.52	−0.47	−0.47	−0.50	−0.50
NH_3	298.06(10)	1.02	0.37	−0.03	−0.39	−0.39	−0.23	−0.23	−0.31	−0.31
N_2	228.42(3)	0.94	1.26	0.99	0.31	−0.09	0.62	0.32	0.48	0.18
H_2CO	374.09(16)	0.23	0.31	−0.41	−0.34	−0.34	−0.37	−0.37	−0.36	−0.36
F_2	39.01(10)	1.86	0.91	1.27	0.24	0.24	0.83	0.83	0.67	0.67
HNO	205.64(6)	0.89	0.86	0.71	−0.11	−0.11	0.27	0.27	0.12	0.12
N_2O	270.60(10)	1.15	1.87	1.14	0.72	0.32	0.90	0.60	0.81	0.51
mean abs.err.		0.72	0.68	0.47	0.32	0.23	0.38	0.28	0.35	0.26
max.abs.err.		1.86	1.87	1.27	0.72	0.64	0.90	0.83	0.81	0.67
aug′-cc-pV*n*Z basis sets										
C_2H_2	405.53(24)	1.49	0.99	−0.12	0.56	0.16	0.31	0.01	0.47	0.17
CH_4	402.23(14)	0.92	0.32	−0.34	0.02	0.02	−0.11	−0.11	−0.09	−0.09
CO	259.58(12)	−0.10	0.66	−0.18	0.28	−0.12	0.12	−0.18	0.22	−0.08
CO_2	389.68(6)	−0.32	0.89	−0.61	0.30	−0.10	−0.03	−0.33	0.18	−0.12
H_2	109.48(0)	0.24	0.06	−0.07	0.00	0.00	−0.03	−0.03	−0.01	−0.01
H_2O	232.83(2)	−0.31	0.31	−0.53	0.05	0.05	−0.14	−0.14	−0.01	−0.01
HF	141.57(17)	−0.42	0.12	−0.54	−0.04	−0.04	−0.18	−0.18	−0.04	−0.04
NH_3	298.06(10)	0.66	0.66	−0.18	0.31	0.31	0.14	0.14	0.22	0.22
N_2	228.42(3)	0.83	1.17	0.41	0.58	0.18	0.51	0.21	0.55	0.25
H_2CO	374.09(16)	0.20	0.33	−0.65	−0.12	−0.12	−0.31	−0.31	−0.20	−0.20
F_2	39.01(10)	1.50	0.83	0.69	0.68	0.68	0.69	0.69	0.68	0.68
HNO	205.64(6)	0.34	1.01	0.09	0.55	0.55	0.38	0.38	0.48	0.48
N_2O	270.60(10)	0.99	1.89	0.40	1.18	0.78	0.90	0.60	0.94	0.64
mean abs.err.		0.64	0.71	0.37	0.36	0.24	0.30	0.22	0.32	0.23
max. abs.err.		1.50	1.89	0.69	1.18	0.78	0.90	0.69	0.94	0.68

momentum, to the total energy. Beyond that, it has no physical basis.

Schwartz [1], on the other hand, considered the convergence of the second-order energy of helium-like atoms, and showed analytically (see also Kutzelnigg [2]) that the contributions to the energy for a given angular momentum l converge as

$$\Delta E(l) = A/(l+1/2)^4 + B/(l+1/2)^6 + O(l^{-8}). \quad (4)$$

(Kutzelnigg [2] in addition showed that for a wavefunction that includes a linear term in the interelectronic distance, A and B vanish, and the leading term becomes $O(l^{-8})$. This explains the great success of methods that explicitly include interelectronic distances.)

Kutzelnigg and Morgan [3] showed that third-order contributions have a leading term $\propto 1/(l+1/2)^5$, while Hill [28] showed that for a general variational calculation

$$\Delta E(l) = A/(l+1/2)^4 + B/(l+1/2)^5 + O(l^{-6}). \quad (5)$$

The error for a calculation with functions of at most angular momentum L is then given by

$$E_\infty - E(L) = \sum_{l=L+1}^{\infty} \left(\frac{A}{(l+1/2)^4} + \frac{B}{(l+1/2)^5} + \frac{C}{(l+1/2)^6} + \ldots \right) \quad (6)$$

$$= A\psi^{(3)}(L+3/2)/6 + B\psi^{(4)}(L+3/2)/24 + C\psi^{(5)}(L+3/2)/120 + \ldots, \quad (7)$$

where $\psi^{(n)}(x)$ represents the polygamma function [29] of order n, which has an asymptotic expansion with leading terms

$$\psi^{(n)}(x) = (-1)^{n-1}\left(\frac{(n-1)!}{x^n} + \frac{n!}{2x^{n+1}} + O(x^{-n-2}) \right) = (-1)^{n-1}\frac{(n-1)!}{(x-1/2)^n} + O(x^{-n-2}), \quad (8)$$

which explains the asymptotic $(L+1)^{-3}$ dependence of the total energy discussed elsewhere [3,28,30].

However, if we define an 'effective' exponent α from the equation

$$\psi^{(n)}(l+1/2) = (n-1)!/l^\alpha \quad (9)$$

that is

$$\alpha = \ln[(n-1)!/\psi^{(n)}(l+1/2)]/\ln l \quad (10)$$

some numerical experimentation reveals that α differs quite appreciably from n for practical values (2–6) of l. Moreover, in a compound made up of both hydrogen and first-row atoms, a calculation in a cc-pVlZ basis set would require a compromise between L and $L+1$. We therefore propose an extrapolation based on the function

$$E = A + B/(l+1/2)^\alpha. \quad (11)$$

Like Eq. (3), Eq. (11) is monotonic and has a clearly defined limit for infinite l: however, its asymptotic convergence behavoir is much slower than that of $\exp(-Bl)$.

For the species in Table 3, the average values of α are 3.9 for cc-pVnZ, and 4.5 for aug′-cc-pVnZ calculations. Evidently, we are still somewhat removed from the regime where $(L+1)^{-3}$ behavior dominates convergence.

In order to provide an alternative for the somewhat clumsy nonlinear fitting, we introduce two approximations with integer exponents:

$$E = A + B/(l+1/2)^4 + C/(l+1/2)^6, \quad (12)$$

$$E = A + B/(l+1/2)^4, \quad (13)$$

which coincidentally have the same form as the expression for the second-order *increments*. We here introduce the notation Schwartzα(lmn) for the extrapolated energy obtained from CCSD(T)/cc-pVlZ, CCSD(T)/cc-pVmZ, and CCSD(T)/cc-pVnZ energies and Eq. (11), Schwartz6(lmn) for ditto using Eq. (12), and Schwartz4(mn) for E_∞ obtained from extrapolation using Eq. (13) from CCSD(T)/cc-pVmZ and CCSD(T)/cc-pVnZ energies. aug′-Schwartz4(mn) then stands for the extrapolation using Eq. (13) from CCSD(T)/aug′-cc-pVmZ and CCSD(T)/aug′-cc-pVnZ energies, and similarly for aug′-Schwartz6(lmn) and aug′-Schwartzα(lmn).

In order to use Eqs. (11)–(13) as extrapolation formulas for the CCSD(T) total energy of a molecule, we have to assume that, in the region of interpolation or fit, (a) the molecular dynamical correlation is dominated by atomic correlation effects; (b) the convergence behavior is dominated by the electron–electron cusp. It would appear that the cc-pVDZ energy is too incomplete to be dominated by the interelectron cusp: therefore we shall only consider results starting at cc-pVTZ.

As it turns out, even Schwartz4(TQ) performs better than Feller(TQ5) or aug′-Feller(TQ5), with a mean absolute error of 0.47 kcal/mol and a maximum error of 1.27 kcal/mol. This is approximately the accuracy obtained using the Martin correction [11,10], with the added advantage that no information is required on the connectivity and bond orders. For aug′-Schwartz4(TQ), the mean absolute error is somewhat improved to 0.37 kcal/mol: more significantly, the largest remaining error drops to 0.69 kcal/mol. Schwartz4(Q5) delivers a similar accuracy, while for aug′-Schwartz4(Q5), the mean absolute error drops to a respectable 0.30 kcal/mol, which is

near the 1 kJ/mol target accuracy. Schwartz6(TQ5), on the other hand, delivers a mean absolute error of 0.35 kcal/mol, while for aug′-Schwartz6(TQ5), this drops to 0.32 kcal/mol. Encouragingly, aug′-Schwartz4(Q5) and aug′-Schwartz6(TQ5) yield similar results on the individual energies, as do Schwartz4(Q5) and Schwartz6(TQ5).

Detailed inspection of the errors for aug′-Schwartz6(TQ5) or aug′-Schwartz4(Q5) reveals a number of things. First, the errors for single AH bonds using aug′-Schwartz6(TQ5) are particularly small: per bond errors in HF, H_2O, NH_3, and CH_4 amount to −0.04, −0.01, 0.07, and −0.02 kcal/mol, respectively! (For aug′-Schwartz4(Q5), the corresponding numbers are −0.18, −0.07, +0.03, and −0.03 kcal/mol.) There does remain a sizeable error of 0.68 kcal/mol for F_2, which could be partly due to residual imperfections in the electron correlation treatment.

Secondly, with the exception of F_2, the largest errors remain in cases with triple bonds (C_2H_2, N_2) or a bond approaching triple-bond character (NNO). (Interestingly, the error for CO appears to be quite reasonable.) Errors for double bonds, as far as can be seen, appear to be fairly small: the mean absolute error for species including at most double bonds is 0.2 kcal/mol. This suggests that the convergence behavior for a triple bond may not quite be dominated yet by the cusp condition over the (TQ5) range. While cc-pV6Z calculations proved to be beyond the available computational facilities, MRCI/cc-pVnZ (n = T,Q,5,6) results for N_2 are available from the work of Bauschlicher and Partridge [31], and for CN and C_2 from the work of Pradhan et al. [32].

The results are summarized in Table 4. After subtracting the core correlation contribution, the 'valence only' experimental $\sum D_e$ amounts to 227.6 kcal/mol. Schwartzα(Q56), Schwartz6(Q56), and Schwartz4(56) reach values of 227.7, 227.5, and 227.4 kcal/mol, respectively, in excellent agreement with experiment. Schwartz6(TQ5) and Schwartz4(Q5) are 0.3 kcal/mol lower than the corresponding data from the higher angular momenta; Schwartzα(TQ5) represents a drop by 0.4 kcal/mol. This would imply that at least 0.3 kcal/mol (0.4 kcal/mol for the Schwartzα values) of the remaining error in Table 3 is due to non-'cusp' basis set incompleteness, and similar observations can probably be made for the other triply bonded species.

Table 4
Computed[31,32] and extrapolated CASSCF/CI total atomization energies of N_2 (kcal/mol), CN and C_2 (eV)

	N_2 kcal/mol	CN eV[a]	CO eV[a]
cc-pVTZ	216.6	7.299	5.899
cc-pVQZ	223.1	7.521	6.061
cc-pV5Z	225.3	7.591	6.110
cc-pV6Z	226.3	7.618	6.129
aug-cc-pVTZ	218.2	7.353	5.928
aug-cc-pVQZ	224.0	7.546	6.077
aug-cc-pV5Z	225.8	7.602	6.118
core corr.	+0.80	0.051	0.067
exp.	228.42(3)	7.738(20)[b]	6.297(20)[c]
w/o core	227.62(3)	7.687(20)	6.23(2)
	Feller		
(TQ5)	226.4	7.623	6.131
aug(TQ5)	226.6	7.625	6.134
(Q56)	227.1	7.635	6.141
	Schwartz6		
(TQ5)	227.2	7.647	6.148
aug(TQ5)	227.3	7.645	6.148
(Q56)	227.5	7.645	6.149
	Schwartz4		
(TQ)	226.7	7.649	6.155
aug(TQ)	227.4	7.657	6.163
(Q5)	227.1	7.648	6.150
aug(Q5)	227.3	7.647	6.151
(56)	227.4	7.646	6.149
	Schwartzα		
(Q5)	227.3	7.647	6.147
aug(Q5)	227.2	7.642	6.145
(56)	227.7	7.646	6.147

[a] Units of Ref. [32] retained on purpose.
[b] Ref. [34]. [c] Ref. [35].

No such effects are seen for C_2 and for the CN radical, where Schwartz6(TQ5) and Schwartz6(Q56) are in nearly total agreement, as are Schwartz4(Q5) and Schwartz4(56), or Schwartzα(TQ5) and Schwartzα(Q56). It should be noted that Feller(TQ5) and Feller(Q56) differ by 0.7 kcal/mol for N_2, with the Feller(Q56) value still 0.5 kcal/mol short

Table 5
Fitted and extrapolated atomic energies from CCSD(T)/aug′-cc-pVnZ calculations (hartree)

	Fit of $A + BC^{-l}$			
	R^2	A	B	C
H	0.99993	−0.50000(1)	0.0102(7)	3.74(14)
B	0.99995	−24.60191(9)	0.135(7)	3.28(10)
C	1.00000	−37.78878(3)	0.349(3)	3.50(2)
N	0.99999	−54.52882(17)	0.62(2)	3.51(5)
O	0.99999	−75.00264(33)	0.96(3)	3.21(5)
F	0.99999	−99.66521(47)	1.29(3)	3.06(4)

	Fit of $A + B/(l+1/2)^4 + C/(l+1/2)^{-6}$			
	R^2	A	B	C
H	0.99999	−0.50004	0.0381(9)	−0.053(5)
B	0.99978	−24.60240(17)	0.78(8)	−1.7(4)
C	0.99994	−37.78988(21)	1.605(91)	−2.8(5)
N	0.99997	−54.53076(26)	2.83(11)	−4.9(7)
O	0.99999	−75.00611(29)	5.95(13)	−13.7(7)
F	1.00000	−99.67013(27)	9.45(12)	−24.2(7)

	Feller(DTQ)	Feller(TQ5)	Schwartzα(TQ5)	Schwartz6(DTQ)	Schwartz6(TQ5)	Schwartz4(TQ)	Schwartz4(Q5)
B	−24.60207	−24.60178	−24.60196	−24.60265	−24.60201	−24.60239	−24.6021
C	−37.78873	−37.78883	−37.78934	−37.79019	−37.78941	−37.78977	−37.7895
N	−54.52855	−54.52910	−54.53008	−54.53114	−54.53018	−54.53045	−54.53025
O	−75.00209	−75.00323	−75.00575	−75.00653	−75.00546	−75.00474	−75.00527
F	−99.66441	−99.66606	−99.67021	−99.67054	−99.66954	−99.66745	−99.66898

of experiment. Such differences are only 0.2 to 0.3 kcal/mol for CN and C_2, reflecting that Eq. (3) will do better with faster overall convergence.

Any remaining error in N_2 is probably mostly due to residual imperfections in the electron correlation treatment: Bauschlicher and Partridge [31] found in their core correlation basis that a multireference averaged coupled pair functional (ACPF) calculation from a CASSCF (complete active space SCF) reference comes out 0.3 kcal/mol lower than the CCSD(T) result.

Returning to Table 3, we can compare performance for the regular and augmented basis sets. As expected, the performance gap dwindles as the underlying basis sets increase, yet the augmented basis sets are clearly superior for highly polar molecules like NH_3, H_2O, and HF.

Summarizing, we have established that using CCSD(T)/aug′-cc-pV5Z calculations and a simple extrapolation formula based on the known convergence rate of the electron cusp, total atomization energies accurate to, on average, 0.2 kcal/mol can be obtained if no multiple bonds are present. If a 0.3 kcal/mol correction term is applied for the triple bonds, similar accuracy can be obtained even for triply bonded species. The principal source of the remaining error, besides the limitations of the extrapolation method, is probably residual imperfection of the electron correlation treatment.

If we make a small compromise with our goal of a 'nonempirical' extrapolation and simply add in 0.3 kcal/mol for each triple bond or cumulenic double bond pair to the Schwartz4(Q5) and Schwartz6(TQ5) numbers (0.4 kcal/mol in the case of Schwartzα(TQ5), we obtain the best numbers overall in Table 3. The mean absolute error for Schwartz6(TQ5) drops to 0.26 kcal/mol for the nonaugmented, and 0.23 kcal/mol for the augmented,

basis sets. For Schwartz4(Q5), the corresponding numbers are 0.28 and 0.22 kcal/mol, respectively; for Schwartzα(TQ5), 0.23 and 0.24 kcal/mol. For the augmented basis sets, almost half the error is contributed by two molecules: F_2 and N_2O. When these are omitted from consideration, the mean absolute error drops to 0.15 kcal/mol with Schwartz6(TQ5), 0.13 kcal/mol with Schwartzα(TQ5), and 0.12 kcal/mol with Schwartz4(Q5)! This error is of the order of magnitude of the experimental error bar for most species in Table 3.

As a final note, let us look at the convergence behavior of the atomic energies. (The use of the l^{-4} dependence for estimating atomic energies at the infinite basis set limit has been discussed by Petersson et al. [33].) Table 5 reveals that the quality of a 4-point Schwartz6(DTQ5) fit actually improves with increasing Z, and that the $O(l^{-6})$ term assumes an increasing importance as Z increases. Furthermore, the Schwartz4 and Schwartz6 extrapolated total energies are quite substantially lower (5 mE_h for F, 3.5 mE_h for O, 2.1 mE_h for N, 1.1 mE_h for C) than the corresponding Feller values. Also, the Feller values for F differ by about 2 mE_h depending on the choice of points, while the difference between Schwartz6(TQ5) and Schwartz4(Q5) is only 0.6 mE_h. Given the above observations on the success of the Schwartz-type extrapolation for molecules, we are inclined to consider the Schwartz-extrapolated atomic total energies to be better estimates for the valence correlation-only total energy than the Feller-extrapolated ones.

4. Conclusions

In this Letter, we have established that, using a sequence of CCSD(T)/aug′-cc-pVnZ (n = T, Q, 5) calculations and a Schwartz-type extrapolation formula, as well as separate calculations to determine the core correlation contribution, we can determine total atomization energies of small polyatomic molecules with a mean absolute error of 0.30 kcal/mol *without any empirical corrections*. If a small correction term of 0.3 kcal/mol is added for every triple bond present, this is improved to 0.20–0.23 kcal/mol. If the problematic molecules F_2 and N_2O are removed from the sample, this goes down to 0.12–0.15 kcal/mol.

Use of nonaugmented basis sets results in a significant loss of accuracy for highly polar compounds. Schwartz4(TQ) yields results of comparable quality to the Martin correction, at comparable cost, and without requiring connectivity information. Feller's exponential-type extrapolation formula is found to be distinctly inferior in performance to the Schwartz-type formulas.

As a general conclusion, the 'state of the art' for accurate theoretical thermochemistry of small molecules is no longer at the 1 kcal/mol, but at the 1 kJ/mol or better level.

Acknowledgement

The author is a Senior Research Associate ('Onderzoeksleider') of the National Science Foundation of Belgium (NFWO/FNRS) and acknowledges a Golda Meir Fellowship from the Hebrew University. He thanks Dr. Thom. H. Dunning Jr. (Pacific Northwest Laboratories), Prof. Trygve Helgaker (University of Oslo, Norway), Dr. Timothy J. Lee (NASA Ames Research Center), and Dr. Peter R. Taylor (San Diego Supercomputer Center) for helpful discussions, Prof. Chava Lifshitz for her hospitality, and Dr. Peter R. Taylor for helpful comments on a draft of the manuscript and technical assistance with one of the calculations that required an inordinate amount of disk space (HNO CCSD(T)/aug′-cc-pV5Z). Generous allowances of computer time were provided by the Institute of Chemistry (Hebrew University) and by the San Diego Supercomputer Center.

References

[1] C. Schwartz, in: Methods in computational physics, ed. B.J. Alder (Academic Press, New York, 1963).
[2] W. Kutzelnigg, Theor. Chim. Acta 68 (1985) 445.
[3] W. Kutzelnigg and J.D. Morgan III, J. Chem. Phys. 96 (1992) 4484; 97 (1992) 8821 (E).
[4] J. Almlöf, B.J. DeLeeuw, P.R. Taylor, C.W. Bauschlicher Jr. and P.E.M. Siegbahn, Intern. J. Quantum Chem. Symp. 23 (1989) 345.
[5] W. Kutzelnigg and W. Klopper, J. Chem. Phys. 94 (1991) 1985.
[6] K.B. Wenzel, J.G. Zabolitzky, K. Szalewicz, B. Jeziorski and H.J. Monkhorst, J. Chem. Phys. 85 (1986) 3964.
[7] R. Bukowski, B. Jeziorski, S. Rybak and K. Szalewicz, J. Chem. Phys. 102 (1995) 888.
[8] B.J. Persson and P.R. Taylor, J. Chem. Phys., in press.

[9] L.A. Curtiss, K. Raghavachari, G.W. Trucks and J.A. Pople, J. Chem. Phys. 94 (1991) 7221.
[10] J.M.L. Martin, J. Chem. Phys. 100 (1994) 8186.
[11] J.M.L. Martin, J. Chem. Phys. 97 (1992) 5012.
[12] J.M.L. Martin, J. Mol. Struct. (THEOCHEM), to be published (WATOC '96 conference issue).
[13] D. Feller, J. Chem. Phys. 96 (1992) 6104.
[14] K.A. Peterson and T.H. Dunning Jr., J. Phys. Chem. 99 (1995) 3898.
[15] K.A. Peterson, R.A. Kendall and T.H. Dunning Jr., J. Chem. Phys. 99 (1993) 9790.
[16] K. Raghavachari, G.W. Trucks, J.A. Pople and M. Head-Gordon, Chem. Phys. Lett. 157 (1989) 479;
[17] C. Hampel, K.A. Peterson and H.J. Werner, Chem. Phys. Lett. 190 (1992) 1
[18] G.D. Purvis III and R.J. Bartlett, J. Chem. Phys. 76 (1982) 1910.
[19] T.J. Lee and G.E. Scuseria, in: Quantum mechanical electronic structure calculations with chemical accuracy, ed. S.R. Langhoff (Kluwer, Dordrecht, 1995).
[20] J.M.L. Martin and P.R. Taylor, Chem. Phys. Lett. 225 (1994) 473.
[21] J.M.L. Martin, Chem. Phys. Lett. 232 (1995) 343.
[22] T.H. Dunning Jr., J. Chem. Phys. 90 (1989) 1007.
[23] R.A. Kendall, T.H. Dunning Jr and R.J. Harrison, J. Chem. Phys. 96 (1992) 6796.
[24] J.E. Del Bene, J. Phys. Chem. 97 (1993) 107.
[25] J. Almlöf and P.R. Taylor, J. Chem. Phys. 86 (1987) 4070.
[26] D.E. Woon and T.H. Dunning Jr., J. Chem. Phys. 103 (1995) 4572.
[27] T.J. Lee, J.M.L. Martin, C.E. Dateo and P.R. Taylor, J. Phys. Chem. 99 (1995) 15858.
[28] R.N. Hill, J. Chem. Phys. 83 (1985) 1173.
[29] M. Abramowitz and I.A. Stegun, Handbook of mathematical functions (Dover, New York, 1972).
[30] T. Helgaker, H. Koch, W. Klopper and J. Noga, "Benchmark coupled-cluster calculations on the H_2O molecule", to be published.
[31] C.W. Bauschlicher Jr. and H. Partridge, J. Chem. Phys. 100 (1994) 4329.
[32] A. Pradhan, H. Partridge and C.W. Bauschlicher Jr., J. Chem. Phys. 101 (1994) 3857.
[33] G.A. Petersson, A.K. Yee and A. Bennett, J. Chem. Phys. 83 (1985) 5105.
[34] Y. Huang, S.A. Barts and J.B. Halpern, J. Phys. Chem. 96 (1992) 425.
[35] R.S. Urdahl, Y. Bao and W.M. Jackson, Chem. Phys. Lett. 178 (1991) 425.