Theoretical Chemistry Accounts \circledcirc Springer-Verlag 1997

Coupling between the convergence behavior of basis set and electron correlation: a quantitative study

Jan M.L. Martin

Department of Organic Chemistry, Kimmelman Building, Weizmann Institute of Science, IL-76100 Rehovot, Israel e-mail: comartin@wicc.weizmann.ac.il

Received: 7 February 1997 / Accepted: 6 May 1997

Abstract. The coupling between improvement of the basis set and the valence electron correlation method has been studied quantitatively for the total atomization energies (TAEs) of a number of small molecules, using basis sets of up to $[7s6p5d4f3g2h/5s4p3d2f1g]$ quality. Very significant coupling is found to exist. Using a scaled basis set extrapolation beyond $[6s5p4d3f2g/$ 4s3p2d1f] at the MP2 or CCSD level, mean absolute errors of 0.18 and 0.15 kcal/mol, respectively, can be obtained for the TAEs of a number of small polyatomic molecules, compared to 0.12 kcal/mol using CCSD(T) throughout.

Key words: Convergence $-\text{Coupling} - \text{Basis set} -$ Electron correlation $-$ Atomization energy

1 Introduction

One of the most vexing problems in carrying out highaccuracy ab initio calculations has traditionally been the very slow basis set convergence of the correlation energy. As has been known for some time $[1-3]$, the increments to the correlation energy of a two-electron atom as a function of angular momentum l converge asymptotically as $A/(l + 1/2)^4 + B/(l + 1/2)^5 + \cdots$, where l is the highest angular momentum present in the basis set. This means that the truncation error for a basis with maximum angular momentum L would be given as

$$
E(L) = E_{\infty} - \frac{A(L+1)^{-3}}{3} + \frac{B(L+1)^{-4}}{4} + O(L^{-5}) \quad , \quad (1)
$$

which is clearly a rather slow convergence. For example, Almlöf et al. [4] showed in a landmark paper that even a multireference configuration interaction (MRCI) calculation in a $[6s5p4d3f2g1h1i]$ basis set, after corrections for core correlation and basis set superposition error, still falls 2.4 kcal/mol short of the experimental dissociation energy of N_2 (see also Ref. [5]).

By contrast, the electron correlation problem appears to be largely solved, in that the $CCSD(T)$ method [6, 7] (i.e., coupled cluster with all single and double excitations [8] and a quasiperturbative treatment of triple excitations [6]) yields, for molecules where the SCF wave function is a good zero-order reference, correlation energies very close to the exact solution with the given basis set [9].

Very recently, Martin and Taylor [10] reported a study in which total atomization energies (TAEs, ΣD_e values) of a number of small polyatomics were obtained with a mean absolute error of 0.12 kcal/mol, which is on a par with the error bars on the experimental values. These authors performed CCSD(T) calculations with basis sets as large as $[7s6p5d4f3g2h/5s4p3d2f1g]$, and subsequently carried out separate extrapolations of the SCF and valence correlation components of TAE to the one-particle basis limit. (Core correlation contributions were considered to be additive and taken from previous work [11].) For the SCF component TAE_{SCF} , a twopoint extrapolation of the form $A + B/(l + 1/2)^5$ was carried out, paralleling the similar technique in the CBS method [12]; for the valence correlation component TAEcorr, a three-point extrapolation of the form $A + B/(l + 1/2)^C$ was used following our earlier arguments $[13]$. The final total atomization energy is then obtained as

$$
TAE = TAE_{SCF}(l \to \infty) + TAE_{corr}(l \to \infty) + TAE_{core} ,
$$
\n(2)

where the contribution of the inner-shell correlation TAE_{core} is evaluated as the difference between calculated binding energies with and without inner-shell electrons correlated, using special basis sets [14, 15] that contain a large complement of high-exponent (hard) polarization functions. Since these latter functions are very different from those required to accommodate valence shell correlation, the core correlation contribution can fairly safely be considered additive.

The main drawback of this method is its formidable computational expense, owing to the steep dependence of CPU time in CCSD(T) on the number of electrons n and the size of the basis set N ; as is well known, each CCSD iteration scales roughly as n^2N^4 , while the final quasiperturbative triples step scales as n^3N^4 . As a result, ethylene was the largest molecule that could be treated

by Martin and Taylor [10] with the available computational hardware.

The question therefore arises of whether the computational expense of this procedure cannot be reduced by using more approximate correlation treatments for the one-particle calibration (i.e., for gauging the effect of further basis set expansion beyond some point of reference). That is, by using an additivity approximation of the type

$$
E[CCSD(T)/(l \to \infty)] \approx E[CCSD(T)/(l = L_a)]
$$

+
$$
(E[M/(l \to \infty)] - E[M/(l = L_a)])
$$
 (3)

for some approximate electron correlation method M and some reference basis set with comparatively low angular momentum L_a .

Approximations similar to the above are common in such procedures as G2 theory [16]. They presuppose that one-particle basis/*n*-particle space coupling is negligible beyond the $M/(l = L_a)$ level of theory. While this assumption was shown not to affect the performance of $G2$ theory appreciably [17], it is not at all clear that this is still the case for high-accuracy work. A quantitative assessment of the importance of one-particular basis/ n -particle space coupling effects near the basis set limit is therefore the subject of the present study. As a by-product, some approximate methods will be derived that permit accurate calculation of TAEs at significantly reduced computational expense.

2 Methods

All closed-shell coupled cluster and MP2 calculations were carried out using the MOLPRO 96¹ ab initio package running on a DEC Alpha TurboLaser 8200 5/300 at the Hebrew University, on a DEC Alpha 500/ 500 MHz workstation at the Weizmann Institute and, for the most demanding calculations in terms of disk space, on the Cray C90 at San Diego Supercomputer Center. The open-shell coupled cluster [18] and MP2 [19] calculations required for the atoms were carried out using the $ACES$ II package² run on the DEC Alpha at the Weizmann Institute.

Dunning's correlation consistent basis sets [20, 21] have been used throughout. These basis sets are designed in such a way that they give a balanced treatment of radial and angular atomic correlation, and therefore constitute a good starting point for any extrapolation based on the maximum angular momentum in the basis set. It was shown repeatedly (e.g., [13, 14]) that the availability of a full complement of diffuse (low-exponent) functions in the basis set is essential for calculations of the highest accuracy as opposed to the common practice [22] of adding only diffuse s and p functions to the basis set, which only recovers about half the effect [14] because no angular correlation in the tail region can thus be accommodated. Therefore, the particular variant of the correlation consistent basis sets used are the aug¢ $cc-pVnZ$ ones, where the prime refers to the suggestion by Del Bene [23] that the diffuse functions be omitted from hydrogen atoms unless they carry significant negative partial charges. (We previously found [13] that this will affect results by less than 0.1 kcal/mol for sufficiently large basis sets.) The contracted sizes of these basis sets are:

aug'-cc-pVDZ $[4s3p2d/2s1p]$, aug'-cc-pVTZ $[5s4p3d2f/3s2p1d]$, aug'-cc-pVQZ $[6s5p4d3f2g/4s3p2d1f]$, aug'-cc-p-V5Z $[7s6p5d4f3g2h/5s4p3d2f1g]$.

For convenience, we will use the abbreviated notation A'VnZ for aug'-cc-pVnZ.

All calculations were carried out from the CCSD(T)/ cc-pVQZ optimized geometries [24], and only valence electron correlation was included explicitly. Core correlation contributions were taken from earlier work [11, 25].

3 Results and discussion

The basis set increments $-$ that is, the difference between aug'-cc-pVnZ and aug'-cc-pV(n-1)Z results – for the SCF and valence correlation part of the total atomization energies of 14 small polyatomic molecules are given in Table 1, including the final increments added on by the extrapolation step.

It is obvious that SCF convergence beyond A^{$V5Z$} is not an issue: in all cases, the extrapolation suggests that the SCF/A'V5Z values are removed from the Hartree-Fock limit by less than 0.1 kcal/mol. The increments for the valence correlation energy are rather more sizable and, as expected, do not decay quite so rapidly as the SCF ones. Even A'V5Z appears to fall short by a substantial amount for multiple bonds. The amounts included through the extrapolation basically make the difference between "mere" chemical accuracy (about 1 kcal/mol) and calibration accuracy which, given the prevalent use of kJ/mol units in such thermochemical reference compendia as the JANAF tables [26], will be arbitrarily defined here as 1 kJ/mol (i.e., 0.24 kcal/mol) or less.

If a particular approximate electron correlation method would not be severely affected by one-particle $basis/n$ -particle space coupling, then either these increments would be reproduced faithfully (in the ideal case), or at least a fairly constant percentage of them would be recovered. Table 2 depicts the increments obtained at the MP2 level, as well as the percentages of the corresponding CCSD(T) increments that are being recovered.

228

¹ MOLPRO 96 is an ab initio MO package written by H.J. Werner and P.J. Knowles, with contributions from J.E. 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

² ACES II is an ab initio program system written by J.F. Stanton, J. Gauss, J.D. Watts, W. Lauderdale, and R.J. Bartlett, incorporating the MOLECULE molecular integral program by J.E. Almlöf and P.R. Taylor and a modified version of the ABACUS integral derivative package by T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, J. Olsen, and P.R. Taylor

Table 1. Basis set increments (kcal/mol) at the SCF and CCSD(T) levels

Table 2. Performance for computed basis set increments

at the MP2 level

HCN 11.49 4.84 1.84 2.02 88.0 100.8 119.0 157.0 HF 3.89 1.70 0.49 0.34 103.0 105.1 131.2 178.0 NH₃ 11.63 4.33 1.54 1.48 97.2 110.7 136.8 190.1

This table immediately reveals that very substantial coupling does indeed exist. On average, the percentage recovered evolves from a mild underestimate for the A'VTZ-A'VDZ increment to a severe overestimate for the $A'V5Z-A'VQZ$ increment. Not only that, but the percentage of that latter increment recovered is quite

erratic, ranging from 107% for $CO₂$ to 194% for $H₂$. As a consequence, the extrapolation term $A'V\infty Z$ $A'V5Z$ is overshot even more severely, with the 153% of N_2 and 188% of CH₄ being quite typical.³

The fact that the onset of this behavior has apparently not yet occured for basis sets of *spdf* quality may explain why one-particle basis set calibration at the MP2 level was found to perform reasonably well in the context of G2 theory, which strives for mere chemical accuracy at an affordable computational cost. Since, however, "calibration" accuracy mandates the use of much larger basis sets, MP2 would appear clearly unacceptable for our purposes.

The same analysis is made in Table 3 for the CCSD method. The behavior here is seen to be completely different: with the exception of F_2 (which has a large contribution of triple excitations to its binding energy), CCSD recovers a fairly constant percentage of the $CCSD(T)$ basis set increments starting at the $A'VQZ$ A¢VTZ one. It would therefore appear natural to suggest the following approximate extrapolation:

$$
\begin{aligned} &\text{TAE}_{\text{corr}}[\text{CCSD(T)}/\text{A}'\text{V}\infty\text{Z}] \approx \text{TAE}_{\text{corr}}[\text{CCSD(T)}/\text{A}'\text{VQZ}] \\& + \left(\frac{\text{TAE}_{\text{corr}}[\text{CCSD(T)}/\text{A}'\text{VQZ}] - \text{TAE}_{\text{corr}}[\text{CCSD(T)}/\text{A}'\text{VTZ}]}{\text{TAE}_{\text{corr}}[\text{CCSD}/\text{A}'\text{VQZ}] - \text{TAE}_{\text{corr}}[\text{CCSD}/\text{A}'\text{VTZ}]} \right) \;, \\ & \times \left(\text{TAE}_{\text{corr}}[\text{CCSD}/\text{A}'\text{V}\infty\text{Z}] - \text{TAE}_{\text{corr}}[\text{CCSD}/\text{A}'\text{VQZ}]\right) \end{aligned} \eqno{(4)}
$$

where it is recalled that $TAE_{corr}[CCSD/A'V\infty Z]$ is obtained by $A + B/(l + 1/2)^C$ extrapolation from CCSD/A'VTZ, CCSD/A'VQZ, and CCSD/A'V5Z results. Such an extrapolation would at least obviate the need for the (T) step in the largest basis set, which will result in significant time savings for large molecules. Moreover, a recent direct implementation has made CCSD calculations with 500 or more basis functions or more, and with dozens of electrons, a realistic possibility even on workstation computers.

³ Similar behavior is seen in unpublished work by Bennett [27], who considered the use of Feller's geometric extrapolation [28] in conjunction with MP2/cc-pVDZ, MP2/cc-pVTZ, and MP2/ cc-pVQZ calculations

	CCSD increments (kcal/mol)				Percentage recovered			
	$AYTZ-$ A'VDZ	$A'VQZ-$ A'VTZ	$A'V5Z-$ A'VQZ	$A'V\infty Z-$ A'V5Z	$AYTZ-$ A'VDZ	$A'VOZ-$ A'VTZ	$A'V5Z-$ A'VQZ	$A'V\infty Z-$ A'V5Z
HNO	9.11	3.99	1.28	1.06	80.4	90.2	88.9	86.8
CO ₂	9.76	5.31	1.70	1.39	77.3	90.7	89.4	87.4
CO	5.27	3.03	1.02	0.90	76.0	90.5	89.0	86.6
F ₂	1.38	0.88	0.24	0.15	42.6	70.5	63.5	54.5
$\rm N_2$	9.60	4.43	1.56	1.49	82.9	92.4	91.3	89.5
N_2O	13.32	6.18	1.97	1.60	79.4	90.4	89.0	86.9
C_2H_2	13.73	4.49	1.30	0.91	88.7	92.7	91.9	90.8
CH ₄	11.26	3.09	0.87	0.58	93.6	94.9	95.1	95.5
H ₂ CO	10.32	4.08	1.24	0.95	84.7	91.7	90.6	88.9
H_2O	6.59	2.70	0.73	0.46	84.1	92.2	91.2	89.9
H ₂	2.94	0.60	0.18	0.13	100.0	100.0	100.0	100.0
HCN	11.19	4.44	1.42	1.15	85.8	92.4	91.5	90.0
ΗF	2.87	1.47	0.33	0.16	76.1	90.9	88.2	84.8
NH ₃	10.87	3.69	1.06	0.73	90.8	94.4	94.2	93.8

Table 4. Performance for several approximate basis set extrapolation methods for total atomization energies (TAE, ΣD_e). All quantities are in kcal/mol

^a Including the correction of $0.126 \times$ (bond order) kcal/mol suggested by Martin and Taylor [10] for each nitrogen-containing bond b Compiled in Ref. [34]. Anharmonic zero-point energies and atomic spin-orbit splitting ΣD_0 values to obtain spin-orbit averaged ΣD_e values that can be directly compared with the present calculations. See Ref. [34] for details and specific references

c Estimated

The performance of this extrapolation is shown in Table 4. It can be seen that the CCSD-based extrapolation, Eq. (4), performs very well, producing results within 0.05 kcal/mol (on average), of the CCSD(T) based extrapolation. The largest deviations, 0.10 kcal/ mol, occur for F_2 and N₂O. After applying the additional correction of $0.126 \times$ (bond order) kcal/mol per nitrogen-involving bond suggested by Martin and Taylor [10], the mean absolute deviation of the total atomization energies themselves from the experimental values comes to 0.20 kcal/mol or, after eliminating F_2 , to 0.15 kcal/mol. (The F_2 molecule has a correlation contribution of about 180% to TAE, being unbound at the SCF level. This probably accounts for it being an outlier.) Such agreement with the experimental results requires no further explanation.

It should be noted that the results using the full $CCSD(T)$ extrapolated TAE_{corr} given in Table 4 differ by a few hundredths of a kcal/mol from the results of Martin and Taylor [10] owing to the different definitions of open-shell CCSD(T) being used for the atomic energies. In [10], the definition as programmed in MOLPRO [7] was employed, while in the present study, the use of the definition in ACES II [18] was required for consistency with the open-shell MP2 definition in use. The accuracy of the results is not affected: as in $[10]$, the

mean absolute error, excluding the pathological case of F_2 , is 0.12 kcal/mol, which is on a par with the average of the experimental error bars, 0.11 kcal/mol.

Now how would the analog of Eq. (4) based on MP2 fare? That is, the expression

$$
TAE_{corr}[CCSD(T)/A'V\infty Z] \approx TAE_{corr}[CCSD(T)/A'VQZ]
$$

+
$$
\left(\frac{TAE_{corr}[CCSD(T)/A'VQZ] - TAE_{corr}[CCSD(T)/A'VTZ]}{TAE_{corr}[MP2/A'VQZ] - TAE_{corr}[MP2/A'VTZ]}\right) \times (TAE_{corr}[MP2/A'V\infty Z] - TAE_{corr}[MP2/A'VQZ]) .
$$
 (5)

Again these results are given in Table 4. And, surprisingly, their performance is quite creditable, their average error of 0.22 kcal/mol (with respect to experiment) is on a par with Eq. (4) despite a somewhat less systematic error pattern. (The largest deviation between Eq. (5) and using CCSD(T) calculations throughout is 0.27 kcal/mol for CO_2 ; the average discrepancy is 0.09 kcal/mol.) After applying the same correction for bonds involving nitrogen as in the CCSD case, the error drops to 0.18 kcal/mol. Since MP2 can be carried out for quite large systems using direct methods $[31-33]$, this opens the perspective of theoretical thermochemistry in the 0.24 kcal/mol (i.e., 1 kJ/mol) accuracy range for medium-sized molecules.

In situations where both direct MP2 and direct CCSD are feasible for the largest basis set, direct CCSD may still be more worthwhile, particularly for molecules where some nondynamical correlation is present.

4 Conclusions

Based on the present study, we can assert that substantial coupling exists between a one-particle basis set and n-particle treatment for total atomization energies. At the MP2 level, successive basis set increments are progressively more severely over-estimated, and at the CCSD level, a fairly constant percentage of the basis set increments is obtained. By scaling the MP2 basis set increments beyond A'VQZ by the ratio between the $CCSD(T)$ and $MP2$ results for the A $'VQZ$ -A $'VTZ$ increment, or similarly using CCSD instead of MP2, very good approximations of the full CCSD(T) extrapolated values are obtained. Using MP2 or CCSD for the basis set extrapolations beyond A'VOZ together with Eqs. (4) and (5) yields total atomization energies with mean absolute errors of 0.17 and 0.15 kcal/mol, respectively, compared with 0.12 kcal/mol when CCSD(T) values are used throughout.

Acknowledgements. The author is a Yigal Allon Fellow and an Honorary Research Associate ("Onderzoeksleider in eremandaat") of the National Science Foundation of Belgium (NFWO/FNRS). He thanks Prof. Peter R. Taylor (San Diego Supercomputer Center and University of California at San Diego) for helpful discussions.

The DEC Alpha workstation at the Weizmann Institute was purchased with USAID (United States Agency for International Development) funds.

References

- 1. Schwartz C (1963) In: Alder BJ (ed) Methods in computational physics 2. Academic Press, New York
- 2. Hill RN (1985) J Chem Phys 83:1173
- 3. Kutzelnigg W, Morgan III JD (1992) J Chem Phys 96:4484; erratum (1992) 97:8821
- 4. Almlöf JE, DeLeeuw BJ, Taylor PR, Bauschlicher Jr. CW, Siegbahn PEM (1989) Int J Quantum Chem Symp 23:345
- 5. Bauschlicher Jr. CW, Partridge H (1994) J Chem Phys 100:4329 and references therein
- 6. Raghavachari K, Trucks GW, Pople JA, Head-Gordon M (1989) Chem Phys Lett 157:479
- 7. Hampel C, Peterson KA, Werner HJ (1992) Chem Phys Lett $190 \cdot 1$
- 8. Purvis III GD, Bartlett RJ (1982) J Chem Phys 76:1910
- 9. Lee TJ, Scuseria GE (1995) In: Langhoff SR (ed) Quantum mechanical electronic structure calculations with chemical accuracy. Kluwer, Dordrecht
- 10. Martin JML, Taylor PR (1997) J Chem Phys 106:8620
- 11. Martin JML (1995) Chem Phys Lett 242:343
- 12. Montgomery Jr. JA, Ochterski JW, Petersson GA (1994) J Chem Phys 101:5900 and references therein
- 13. Martin JML (1996) Chem Phys Lett 259:669
- 14. Martin JML, Taylor PR (1994) Chem Phys Lett 225:473
- 15. Woon DE, Dunning TH (1995) J Chem Phys 103:4572
- 16. Curtiss LA, Raghavachari K, Trucks GW, Pople JA (1991) J Chem Phys 94:7221
- 17. Curtiss LA, Carpenter JE, Raghavachari K, Pople JA (1992) J Chem Phys 96:9030
- 18. Watts JD, Gauss J, Bartlett RJ (1993) J Chem Phys 98:8718
- 19. Lauderdale WJ, Stanton JF, Gauss J, Watts JD, Bartlett RJ (1992) J Chem Phys 97:6606
- 20. Dunning Jr. TH (1989) J Chem Phys 90:1007
- 21. Kendall RA, Dunning Jr. TH, Harrison RJ (1992) J Chem Phys 96:6796
- 22. Clark T, Chandrasekar J, Spitznagel GW, von Ragué Schleyer P (1983) J Comp Chem 4:294
- 23. Del Bene JE (1993) J Phys Chem 97:107
- 24. Martin JML (1994) J Chem Phys 100:8186
- 25. Martin JML, Taylor PR (1996) Chem Phys Lett 248:336
- 26. Chase Jr. MW, Davies CA, Downey Jr. JR, Frurip DJ, McDonald RA, Syverud AN (1985) JANAF thermochemical tables, 3rd edn. J Phys Chem Ref Data 14, supplement 1
- 27. Bennett FW (personal communication)
- 28. Feller D (1992) J Chem Phys 96:6104
- 29. Koch H, Christiansen O, Kobayashi R, Jørgensen P, Helgaker T (1994) Chem Phys Lett 228:233
- 30. Koch H, Sanchez de Meras A, Helgaker T, Christiansen O (1996) J Chem Phys 104:4157
- 31. Frisch MJ, Head-Gordon M, Pople JA (1990) Chem Phys Lett 166:281
- 32. Almlöf JE (1991) Chem Phys Lett 181:319
- 33. Häser M, Almlö JE (1992) J Chem Phys 96:489
- 34. Martin JML J Mol Struct (Theochem), (in press; WATOC96 issue)