

Basis-set convergence of the energy in molecular Hartree–Fock calculations

Asger Halkier^{a,*}, Trygve Helgaker^b, Poul Jørgensen^a, Wim Klopper^b,
Jeppe Olsen^a

^a Department of Chemistry, University of Aarhus, DK-8000 Århus, C, Denmark

^b Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway

Received 18 December 1998; in final form 5 February 1999

Abstract

The basis-set convergence towards the numerical limit of the Hartree–Fock total energy and binding energy is investigated for the correlation-consistent cc-pVXZ basis sets. For both energies, solid improvements are obtained with each increment in X . The basis-set errors for the total energy (ΔE) fit an exponential form better than a power form, and the total energy is better fitted than the binding energy. It is difficult to find generally reliable extrapolation schemes for the total energy. In most cases, the most successful scheme gives results extrapolated beyond a given X that are comparable to the cc-pV(X + 1)Z results, but occasionally it fails dramatically for large X . Indeed, explicit calculation of the energy in a larger basis set, especially the cc-pV6Z set for which $\Delta E \leq 0.1$ mE_h, gives the most reliable estimate of the basis-set limit. © 1999 Elsevier Science B.V. All rights reserved.

1. Introduction

Basis-set convergence and extrapolations are among the most important issues of contemporary molecular ab initio theory; see, e.g., Refs. [1–11] and references therein. For total energies, it is well known that the convergence of the correlation part is significantly slower than that of the Hartree–Fock (H-F) part. First, this suggests that, when studying the basis-set convergence of the energy, one should treat the H-F and correlation parts separately. Second, it shows that convergence acceleration is most important for the correlation part. Consequently, recent

work on basis-set convergence and extrapolations has focused primarily on the correlation energy [1,5–8,10].

When examining basis-set convergence, it is mandatory to have a hierarchical sequence of basis sets with systematic improvements from level to level. The correlation-consistent polarized valence basis sets cc-pVXZ of Dunning et al. [12,13] constitute such a hierarchy and have been extensively used in our previous work on the basis-set convergence of the correlation energy [8,10]. In these studies, we proposed the expression

$$E_X^{\text{corr}} = E_{\text{lim}}^{\text{corr}} + AX^{-3}, \quad (1)$$

for the convergence of the correlation energy E_X^{corr} obtained with the basis set of cardinal number X ($D:2, T:3, \dots$) towards the basis-set limit $E_{\text{lim}}^{\text{corr}}$.

* Corresponding author. Fax: +45 86 19 6199; e-mail: asger@kemi.aau.dk

From this simple expression, reliable estimates of the basis-set limit of the correlation energy may be obtained. The main purpose of the present study is to investigate whether it is possible to establish a similarly simple formula for the H-F energy. Such a formula would be of practical importance in calculations of accurate total electronic energies at the H-F level itself and also at the correlated level (to which the H-F energy contributes). Clearly, if a suitable technique can be developed for the extrapolation of the H-F energy, the cost of many calculations aimed at the accurate prediction of reaction enthalpies of chemical reactions between molecules, and other thermochemical data, may be reduced. In addition, the convergence of the H-F energy constitutes in itself an interesting theoretical problem, on which the numerical data generated in the present investigation may throw some light.

As shown in Ref. [8], the expression (1) exhibits the correct asymptotic behaviour of the correlation energy for large X compared with the theoretical analysis of the partial-wave expansion of the helium atom [14]. The extrapolation for the correlation energy is therefore, in some sense, *theoretically* motivated. For the H-F energy, on the other hand, no such theoretical foundation appears to exist and the question of basis-set convergence must therefore be approached with caution so as to avoid any unsystematic parameter fitting. In this study, we shall examine two of the most popular extrapolation forms – namely, the *exponential form*

$$E_X^{\text{H-F}} = E_{\text{lim}}^{\text{H-F}} + B \exp(-\alpha X), \quad (2)$$

which has been extensively used by Feller et al. for extrapolations of various properties (see, e.g., Refs. [3,4]) and the *power form*

$$E_X^{\text{H-F}} = E_{\text{lim}}^{\text{H-F}} + BX^{-\alpha}, \quad (3)$$

which forms the basis for the CBS model of Petersson et al. (see Refs. [1,2] and references therein). Our purpose is to establish whether one particular form is preferable to the other for H-F energies, and therefore also reaction enthalpies as mentioned above, and how reliable and accurate this form is.

As the convergence of the correlation energy is governed by the description of the electronic cusp, the expression (1) applies equally well to atoms and molecules. Therefore, the convergence of the correlation part of the binding energy is the same as the

convergence of the total (molecular or atomic) correlation energy. However, there is no electronic cusp in Hartree–Fock to dictate the convergence of the energy to a given form irrespective of the nature of the system. On the contrary, as the partial wave expansion at the H-F level is finite for atoms (with contributions from only s and p functions for C, for example) but infinite for molecules (with contributions from all angular-momentum functions), the convergence of the molecular total and binding energies may well be different. We shall therefore examine binding energies at the H-F level in some detail as well.

To investigate the convergence towards the true basis-set limit, one needs to establish this limit from a source that is independent of the basis sets under study. For diatomics, H-F basis-set limits may be determined in numerical orbital calculations and several such studies have appeared in the literature [15–17]. We therefore consider primarily diatomics in the present study, investigating the following nine systems: H_2 , C_2 , N_2 , F_2 , BH, HF, BF, CO, and NO^+ . These systems, which contain from 2 to 18 electrons, cover a reasonably broad spectrum of diatomics ranging from non-polar to polar, from singly bonded to triply bonded, and with different weights of the H-F determinant in the full configuration interaction wave function. Our strategy is to calculate the H-F energy for these diatomics using the cc-pVXZ sets with cardinal numbers $X \leq 6$ and to compare these energies, and the corresponding extrapolated energies, with the numerical basis-set limits.

Finally, we investigate H_2O and CH_2O to see whether the conclusions for the diatomics are valid for polyatomic molecules as well. Because of the large number of systems investigated and since we compare the results with the true basis-set limit, the present study appears to be the most extensive systematic investigation of H-F energy convergence carried out to date.

2. Results and discussion

2.1. Computational details

For the diatomics, the cc-pVXZ H-F calculations with $X \leq 6$ were carried out at the following fixed

Table 1
Calculated H-F energies (in E_h) for the nine diatomics

	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pV5Z	cc-pV6Z	Numerical
H ₂	-1.128709	-1.132961	-1.133459	-1.133608	-1.133625	-1.133630 ^a
C ₂	-75.387018	-75.401381	-75.405700	-75.406426	-75.406539	-75.406565 ^b
N ₂	-108.95455	-108.98409	-108.99174	-108.99342	-108.99374	-108.99381 ^c
F ₂	-198.685679	-198.752055	-198.768284	-198.772750	-198.773300	-198.773323 ^b
BH	-25.125332	-25.129933	-25.131295	-25.131556	-25.131622	-25.131639 ^d
HF	-100.01941	-100.05801	-100.06769	-100.07043	-100.07074	-100.07082 ^c
BF	-124.10747	-124.15586	-124.16599	-124.16846	-124.16873	-124.16885 ^c
CO	-112.74929	-112.78036	-112.78886	-112.79062	-112.79085	-112.79095 ^c
NO ⁺	-128.92696	-128.96512	-128.97519	-128.97742	-128.97768	-128.97778 ^c

^aRef. [15].

^bRef. [18].

^cRef. [16].

^dRef. [23].

geometries (which correspond to the bond distances employed in the numerical calculations): $R_{\text{H}_2} = 1.400 a_0$, $R_{\text{C}_2} = 2.358 a_0$, $R_{\text{N}_2} = 2.068 a_0$, $R_{\text{F}_2} = 2.668 a_0$, $R_{\text{BH}} = 2.3289 a_0$, $R_{\text{HF}} = 1.7328 a_0$, $R_{\text{BF}} = 2.386 a_0$, $R_{\text{CO}} = 2.132 a_0$, and $R_{\text{NO}^+} = 2.0092 a_0$. The accuracy of the energies of C₂ and F₂ given in Ref. [17] is not high enough for our purpose, but more accurate numbers were kindly supplied to us by Kobus [18]. The H-F calculations in the cc-pVXZ basis sets with $X \leq 6$ and in the aug-cc-pV6Z basis set¹ for H₂O and CH₂O were carried out at the experimental equilibrium geometries of $R_{\text{O-H}} = 0.9572 \text{ \AA}$, $\angle_{\text{H-O-H}} = 104.52^\circ$ (Ref. [19]) and $R_{\text{C-O}} = 1.203 \text{ \AA}$, $R_{\text{C-H}} = 1.100 \text{ \AA}$, $\angle_{\text{H-C-H}} = 116.3^\circ$ (Ref. [20]), respectively. All the molecular calculations were performed with the Dalton program package [21], whereas the atomic calculations were carried out with the LUCAS program [22].

2.2. Convergence towards the basis-set limit

The cc-pVXZ and numerical results for the diatomics are listed in Table 1 and the corresponding

¹ Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, version 1.0, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352, USA, and funded by the US Department of Energy. The Pacific Northwest Laboratory is a multi-program laboratory operated by Battelle Memorial Institute for the US Department of Energy under Contract DE-AC06-76RLO 1830. Contact David Feller, Karen Schuchardt, or Don Jones for further information.

atomic results are given in Table 2. The error in the total molecular energy calculated in the cc-pVXZ basis

$$\Delta E_X^{\text{mol}} = E_X^{\text{mol}} - E_{\text{lim}}^{\text{mol}}, \quad (4)$$

is given in the upper half of Table 3. The error in the total molecular energy may be divided into an atomic part and a bonding part. The basis-set error in the H-F binding energy is given by

$$\Delta E_X^{\text{bnd}} = \Delta E_X^{\text{mol}} - \sum \Delta E_X^{\text{atom}}, \quad (5)$$

where ΔE_X^{atom} is defined analogously to ΔE_X^{mol} . ΔE_X^{bnd} is listed in the lower half of Table 3.

For the total energy, there are solid improvements from level to level in the series. Typically, the error is reduced by a factor of 4–5 from double to triple zeta and again from triple to quadruple zeta. From quadruple to quintuple zeta and from quintuple to sextuple zeta, the convergence is less systematic but the error is still reduced by at least a factor of 3 with each increment in X . The convergence for F₂ is different from that of the other systems, becoming very rapid for large X , with a reduction in the error from quintuple to sextuple zeta by more than a factor of 20.

For the binding energy, solid improvements from level to level are also observed. However, the variation from system to system is larger than for the total energy. Moreover, the error in the binding energy is not a constant fraction of the total error – whereas, for small basis sets, the binding error is typically less than one half of the total error, for cc-pV6Z, the binding error is typically larger than one half of the

Table 2
Calculated atomic H-F energies (in E_h)

	H	B	C	N	O	F
cc-pVDZ	−0.499278	−24.526564	−37.682391	−54.388414	−74.786188	−99.371080
cc-pVTZ	−0.499810	−24.528098	−37.686662	−54.397358	−74.803078	−99.399194
cc-pVQZ	−0.499946	−24.528898	−37.688234	−54.400176	−74.807975	−99.406980
cc-pV5Z	−0.499995	−24.529036	−37.688572	−54.400852	−74.809258	−99.409128
cc-pV6Z	−0.499999	−24.529057	−37.688612	−54.400923	−74.809381	−99.409324
Numerical ^a	−0.500000	−24.529061	−37.688619	−54.400934	−74.809398	−99.409349

^aNumerical results from Ref. [24].

total error. We thus have the first indication that, in H-F theory, the binding energy does not converge in the same manner as the total energy. Again F_2 behaves differently from the other systems, the atomic error being larger than the total error for the cc-pV6Z basis, leading to a non-monotonic convergence of the binding energy as seen in Table 3.

It is gratifying to see that the cc-pV5Z total energies are safely within 1 mE_h of the basis-set limit and that the cc-pV6Z error is about 0.1 mE_h . Concerning the binding energies, the cc-pVQZ results are already within 1 mE_h of the basis-set limit and the cc-pV6Z results within 0.1 mE_h . For most purposes, therefore, the H-F total energy and binding energy may be considered to be converged to the basis-set limit at the cc-pV6Z level, especially compared with the accuracy with which we are able to determine the correlation energy. Thus, at the cc-pV5Z and cc-pV6Z levels, the errors in the H-F energy are at least an order of magnitude smaller than the errors in the correlation energy [10].

Table 3
Errors in the total molecular energy, ΔE_X^{mol} , (upper half) and the binding energy, ΔE_X^{bnd} , (lower half) (in mE_h)

	H ₂	C ₂	N ₂	F ₂	BH	HF	BF	CO	NO ⁺
cc-pVDZ	4.92	19.55	39.26	87.64	6.31	51.41	61.38	41.66	50.82
cc-pVTZ	0.67	5.18	9.72	21.27	1.71	12.81	12.99	10.59	12.66
cc-pVQZ	0.17	0.87	2.07	5.04	0.34	3.13	2.86	2.09	2.59
cc-pV5Z	0.02	0.14	0.39	0.57	0.08	0.39	0.39	0.33	0.36
cc-pV6Z	0.005	0.03	0.07	0.02	0.02	0.08	0.12	0.10	0.10
cc-pVDZ	3.48	7.09	14.22	11.11	3.09	12.42	20.61	12.22	15.09
cc-pVTZ	0.29	1.27	2.57	0.96	0.55	2.47	1.87	2.31	2.76
cc-pVQZ	0.06	0.10	0.55	0.30	0.13	0.71	0.33	0.28	0.41
cc-pV5Z	0.01	0.05	0.23	0.13	0.05	0.16	0.14	0.14	0.14
cc-pV6Z	0.003	0.01	0.05	−0.03	0.01	0.05	0.09	0.08	0.07

2.3. Exponential and power fits to the total energy

Turning our attention to the exponential and power fits, we note that (2) and (3) may be brought to a linear form by taking the logarithm:

$$\ln(\Delta E_X^{\text{mol}}) = \ln(E_X^{\text{mol}} - E_{\text{lim}}^{\text{mol}}) = \ln(B) - \alpha X, \quad (6)$$

$$\ln(\Delta E_X^{\text{mol}}) = \ln(E_X^{\text{mol}} - E_{\text{lim}}^{\text{mol}}) = \ln(B) - \alpha \ln(X). \quad (7)$$

Provided these forms are suitable for extrapolations, the logarithm of the error should give straight lines when plotted against X and $\ln(X)$ for the exponential and power forms, respectively. We shall here investigate their suitability by performing linear regression on each of the linear forms (6) and (7) for the total energy. The results of these fits are given in Table 4.

The exponential form fits the data well – the correlation coefficient is good and α appears to be rather insensitive to what data that are included in the fit. By contrast, for the power form, α changes systematically with variations in the data, invariably increasing from the 2 – 5 fit to the 2 – 6 fit and again from the 2 – 6 fit to the 3 – 6 fit. Moreover, the correlation coefficients for the power fits are poorer than those for the exponential fits. Thus, for fits and extrapolations of the H-F energy, the exponential form is preferable to the power form. In the following, therefore, we shall be concerned mainly with extrapolations based on the exponential form.

2.4. Exponential and power fits to the binding energy

For the binding energies, we have performed an analysis similar to that carried out for the total energies. Replacing ΔE_X^{mol} with ΔE_X^{bnd} in (6) and

Table 4

The parameters α and B extracted from linear regressions using data with $X \in [X_{\min}, X_{\max}]$ for the two different fitting forms for the total molecular energy

	X_{\min}	X_{\max}	Exponential form			Power form		
			α	B	Corr.	α	B	Corr.
H ₂	2	6	1.7198	0.1400	−0.9983	6.2071	0.5076	−0.9884
	2	5	1.7595	0.1576	−0.9972	5.6921	0.3034	−0.9885
	3	6	1.6740	0.1113	−0.9971	7.1768	2.2985	−0.9895
C ₂	2	6	1.6864	0.6756	−0.9986	6.0260	2.2069	−0.9790
	2	5	1.6629	0.6296	−0.9974	5.3266	1.0969	−0.9790
	3	6	1.7714	1.0335	−0.9998	7.6230	26.550	−0.9960
N ₂	2	6	1.5875	1.0580	−0.9991	5.6740	3.2304	−0.9797
	2	5	1.5382	0.9126	−0.9992	4.9478	1.5631	−0.9849
	3	6	1.6469	1.4244	−0.9997	7.0606	28.004	−0.9921
F ₂	2	6	2.0105	8.1500	−0.9826	7.0350	27.465	−0.9432
	2	5	1.6530	2.7888	−0.9939	5.2763	4.7349	−0.9721
	3	6	2.2662	29.273	−0.9849	9.5660	1415.4	−0.9623
BH	2	6	1.4855	0.1331	−0.9995	5.3254	0.3862	−0.9830
	2	5	1.4593	0.1230	−0.9993	4.7055	0.2078	−0.9874
	3	6	1.5248	0.1620	−0.9997	6.5558	2.6254	−0.9950
HF	2	6	1.6423	1.6392	−0.9971	5.8536	5.0912	−0.9750
	2	5	1.6054	1.4672	−0.9946	5.1280	2.4650	−0.9735
	3	6	1.7311	2.5547	−0.9975	7.4168	58.105	−0.9892
BF	2	6	1.5980	1.5197	−0.9978	5.7564	4.9571	−0.9860
	2	5	1.6689	1.8798	−0.9978	5.3575	3.3271	−0.9815
	3	6	1.6046	1.5701	−0.9957	6.9229	30.499	−0.9944
CO	2	6	1.5533	0.9885	−0.9981	5.5866	3.0846	−0.9848
	2	5	1.6137	1.1851	−0.9978	5.1723	2.0387	−0.9800
	3	6	1.5833	1.1488	−0.9967	6.8427	21.820	−0.9970
NO ⁺	2	6	1.6022	1.3769	−0.9980	5.7485	4.3723	−0.9823
	2	5	1.6437	1.5593	−0.9968	5.2599	2.6828	−0.9775
	3	6	1.6496	1.7456	−0.9970	7.1146	36.711	−0.9953

‘Corr.’ is the statistical correlation coefficient for the fit.

(7) and performing linear regression, we investigate whether the convergence of the binding energy fits an exponential form or a power form. The results of these fits are given in Table 5.

Comparing Tables 4 and 5, we see that the situation is less clear-cut for the binding energy than for the total energy. First, the power fit is now better than the exponential fit, but not in all cases (only in about 70% of these). Second, the fits are now in general significantly poorer than for the total energies. Thus, whereas, for the total energy, the poorest correlation coefficient for the best form is -0.9946 , it is down to -0.9760 for the binding energy and about 60% of the correlation coefficients for the power form are poorer than -0.9946 . At the H-F level, therefore, the convergence of the binding energy is rather different from that of the total energy,

and the latter is somewhat easier to work with. In the following, we shall therefore consider only the total energy.

The reason for the special results of F₂ is partly understood. From Table 3, we note that, relative to the total error, the binding error is particularly small for this molecule. For F₂, therefore, the molecular error is dominated by the atomic error to a much higher degree than for the other diatomics. As the molecular and atomic energies do not converge similarly, this may explain the results obtained for F₂. However, no matter what the explanation is, the special results for F₂ will diminish the generality of any extrapolation scheme. In the discussion of extrapolations below, we have chosen to exclude F₂ from the statistical analysis and instead treat it as a special worst case.

Table 5

The parameters α and B extracted from linear regressions using data with $X \in [X_{\min}, X_{\max}]$ for the two different fitting forms for the binding energies

	X_{\min}	X_{\max}	Exponential form			Power form		
			α	B	Corr.	α	B	Corr.
H ₂	2	6	1.7292	75.08	−0.9936	6.3303	308.41	−0.9979
	2	5	1.8530	108.85	−0.9933	6.0793	239.97	−0.9986
	3	6	1.5362	28.60	−0.9995	6.6133	479.25	−0.9960
C ₂	2	6	1.6103	135.02	−0.9834	5.9166	517.69	−0.9912
	2	5	1.7773	222.77	−0.9813	5.8265	473.12	−0.9858
	3	6	1.4733	68.04	−0.9694	6.4671	1220.27	−0.9849
N ₂	2	6	1.3813	185.26	−0.9953	5.0243	548.92	−0.9932
	2	5	1.3959	193.59	−0.9911	4.5908	355.90	−0.9987
	3	6	1.2836	113.66	−0.9950	5.5203	1188.52	−0.9905
F ₂ ^a	2	5	1.4478	127.75	−0.9681	4.8446	265.41	−0.9927
BH	2	6	1.3446	36.62	−0.9948	4.8978	106.38	−0.9941
	2	5	1.3666	39.12	−0.9902	4.4997	71.45	−0.9991
	3	6	1.2365	21.33	−0.9953	5.3179	204.67	−0.9908
HF	2	6	1.3586	164.69	−0.9983	4.9283	470.81	−0.9934
	2	5	1.4230	199.80	−0.9989	4.6204	346.09	−0.9938
	3	6	1.2924	118.28	−0.9988	5.5652	1269.69	−0.9955
BF	2	6	1.3411	149.15	−0.9567	5.0607	544.47	−0.9905
	2	5	1.6633	392.21	−0.9782	5.5306	870.92	−0.9966
	3	6	0.9895	25.71	−0.9591	4.3809	188.86	−0.9829
CO	2	6	1.2944	108.68	−0.9702	4.8200	348.36	−0.9911
	2	5	1.5449	230.41	−0.9830	5.0811	452.25	−0.9907
	3	6	1.0926	39.62	−0.9506	4.8462	362.86	−0.9760
NO ⁺	2	6	1.3687	167.33	−0.9823	5.0591	545.70	−0.9961
	2	5	1.5994	334.31	−0.9943	5.2259	644.67	−0.9955
	3	6	1.2030	73.05	−0.9732	5.2954	788.47	−0.9916

^aBecause of the negative error in the binding energy at the cc-pV6Z level we have only performed the fit with $X \in [2,5]$.

‘Corr.’ is the statistical correlation coefficient for the fit.

2.5. Three-point non-linear exponential extrapolations

For the correlation energy, an extrapolation based on formula (1) requires only two sets of results (X , E_X^{corr}). For the H-F energy, we would likewise like to keep the number of sets of results (X , $E_X^{\text{H-F}}$) needed for the extrapolation as small as possible. Moreover, since the extrapolations for the correlation energy are most accurate for two consecutive X [10], we require the X values in the H-F extrapolations to be consecutive as well. Since the exponential form constitutes a three-parameter model, we begin with extrapolations based on formula (2) with results for three consecutive X . This gives three equations in three unknowns, which may be solved analytically for α , B , and $E_{\text{lim}}^{\text{H-F}}$. The values of $E_{\text{lim}}^{\text{H-F}}$ obtained in this way for the eight diatomics are compared with

the numerical values (from Table 1) in Table 6, where we give the mean error, the standard deviation, the maximum absolute error, and the mean absolute error for the extrapolated limits relative to the numerical numbers.

As can be seen from the table, the extrapolations based on the $D - Q$ energies are a good deal better than the cc-pVQZ energies, the mean and maximum errors being about a factor of 3 smaller. On the other hand, the extrapolations overshoot the limit by about 30% (compared with the cc-pVQZ error) and the extrapolated values are not as good as the cc-pV5Z results.

The $T - 5$ and $Q - 6$ extrapolations are worse. Thus, the $T - 5$ extrapolations overshoot the limit with about 80% compared with the cc-pV5Z error and the maximum absolute error of the extrapolated values is about twice that of the cc-pV5Z basis. In

Table 6

Mean error ($\bar{\Delta}$), standard deviation (Δ_{std}), maximum absolute error (Δ_{max}), and mean absolute error ($\bar{\Delta}_{\text{abs}}$) of the extrapolated basis-set limits for the sample of eight diatomics (F_2 being excluded) (in mE_h)

X	cc-pVXZ			Exponential three-point extrapolation				Simple linear two-point extrapolation			
	$\bar{\Delta}$	Δ_{std}	Δ_{max}	$\bar{\Delta}$	Δ_{std}	Δ_{max}	Δ_{abs}	$\bar{\Delta}$	Δ_{std}	Δ_{max}	$\bar{\Delta}_{\text{abs}}$
2	34.41	21.52	61.38								
3	8.29	5.07	12.99					1.93	1.39	3.40	2.02
4	1.77	1.15	3.13	-0.47	0.53	1.11	0.54	0.18	0.29	0.77	0.22
5	0.26	0.15	0.39	-0.20	0.25	0.69	0.21	-0.10	0.11	0.28	0.11
6	0.07	0.04	0.12	0.02	0.05	0.09	0.04	0.02	0.02	0.05	0.02

For a given X the exponential three-point extrapolation is based on results with $X' \in [X - 2, X]$, and the simple linear two-point extrapolation on results with $X' \in [X - 1, X]$. For comparison, the corresponding statistical information about the cc-pVXZ results is given.

this case, therefore, the extrapolation does not even represent an improvement on the cc-pV5Z basis. For the $Q - 6$ extrapolations, the mean error is a good deal better than that of the cc-pV6Z basis, but the maximum error is only reduced by 25% by the extrapolation.

The performance of this type of extrapolation is obviously rather poor. The reason for this behaviour is that, in order to extrapolate beyond a given cardinal number X , we need to use the result for $X - 2$. The error in the latter energy is about a factor of 20 larger than that of the former energy, which from the point of view of extrapolation is not ideal.

2.6. Two-point linear exponential extrapolations with a global α

The discussion in Section 2.5 indicates that we might improve our scheme by reducing the number of parameters from three to two, thereby rendering the $X - 2$ result superfluous for the extrapolation. Reviewing the data in Table 4, we note that, except for F_2 , α is practically the same for all fits and for all diatomics. Indeed, for the $D - 6$ fits, the mean value of α is 1.609 with a standard deviation of 0.074. For the $D - 5$ fits and $T - 6$ fits, the corresponding numbers are 1.619 ± 0.090 and 1.648 ± 0.079 , respectively. This observation suggests that we could try to fix α in the fits. An obvious way of choosing α is as the mean value from all the 24 fits. This strategy gives $\alpha = 1.63$, which corresponds to a reduction of the error by a factor of 5.1 with each increment in X . A noteworthy benefit of this approach is that the three-parameter scheme is reduced

not only to a two-parameter scheme but to a *linear* two-parameter scheme.

The performance of the extrapolated limits obtained from solving the two equations for pairs of consecutive X values in the two parameters B and $E_{\text{lim}}^{\text{H-F}}$ with fixed $\alpha = 1.63$ is shown in Table 6. We immediately note that this simple two-point extrapolation gives results that are significantly better than those obtained with the exponential three-point extrapolation. For the lower cardinal numbers, the accuracy of the extrapolated energies is now comparable with that of the energies in the cc-pV($X + 1$)Z basis set – compare the $D - T$ extrapolated energies with the cc-pVQZ energies and also the $T - Q$ extrapolated energies with the cc-pV5Z energies. The $Q - 5$ extrapolation is not quite as good – it overshoots the limit by about 40% compared with the cc-pV5Z error and the extrapolated energies are not as good as the cc-pV6Z energies. Finally, the $5 - 6$ extrapolation represents a reasonable improvement upon the cc-pV6Z basis, the maximum absolute and mean errors being reduced by more than a factor of 2 and 3, respectively.

For the eight ‘regular’ diatomics, the performance of this simple two-point extrapolation is thus rather satisfactory so at this point we should investigate how it performs for F_2 . For this molecule, the errors of the two-point extrapolations are: 5.09 mE_h ($D - T$), 1.08 mE_h ($T - Q$), -0.52 mE_h ($Q - 5$), and -0.11 mE_h ($5 - 6$). Some similarities with the other diatomics are immediately seen. The $D - T$ energy is close to the cc-pVQZ energy and the extrapolation becomes poorer for higher cardinal numbers. Thus, the $Q - 5$ and in particular the $5 - 6$ energies are

poor compared with the cc-pV5Z and cc-pV6Z energies, respectively.

Altogether, we may summarize the performance of the two-point extrapolation as follows:

- For $X \leq 4$, the extrapolated energies are comparable to the cc-pV($X + 1$)Z energies.
- The $T - Q$ extrapolated energies are within 1 mE_h of the basis-set limit.
- For $X > 4$, the extrapolation usually represents an improvement on the cc-pVXZ energy, but the improvements are smaller than for $X \leq 4$. In difficult cases such as F₂, the extrapolated energy may even be poorer than the cc-pVXZ energy.

Somewhat reluctantly, we conclude that it is difficult to find a simple global extrapolation scheme that works well for all X . Indeed, in sharp contrast to the correlation energy, the most reliable approximation to the basis-set limit of the H-F energy is obtained by calculating the energy in the next basis in the series. If this is not feasible, one can obtain results of comparable quality using the two-point extrapolation as long as $X < 5$ for the best basis; for $X \geq 5$, however, one should be careful in employing the extrapolation. However, if one is performing both H-F and highly correlated calculations, the additional larger H-F calculation does not represent any new computational bottleneck, being less expensive than the highly correlated calculation in the original largest basis. For the cc-pV6Z basis, the problem of course arises that there exists no larger basis. However, as the errors at this level are only about 0.1 mE_h, the cc-pV6Z energies are for most purposes converged to the basis-set limit anyway.

2.7. Two-point linear exponential extrapolations with an X -dependent α

Although the linear two-parameter scheme considered so far is not fully satisfactory, it still represents an improvement on the three-parameter scheme, but so far we have only considered a scheme where the same α is used for all X . We may, however, take the idea one step further and work with different α for different X pairs. The value of α to be used for a particular X pair (e.g., 2 and 3) is obtained by minimizing the root-mean-square (RMS) error of the extrapolated limits over the sample of diatomics. The

optimal α are listed in Table 7, where we have given also the corresponding mean deviations, standard deviations, maximum errors, and mean absolute errors.

The average error for these extrapolations is very good, which is not surprising as the extrapolation is designed to give a small mean error. Comparing the standard deviations and maximum errors with those of Table 6, we see that the $D - T$ numbers are a little better than those of the cc-pVQZ basis, the $T - Q$ numbers are a little worse than those of the cc-pV5Z basis, the $Q - 5$ numbers are comparable to those of the cc-pV6Z basis, and the 5 - 6 numbers are a good deal better than those of the cc-pV6Z basis. For the eight well-behaved diatomics, this type of extrapolation (with $X' \in [X - 1, X]$) thus seems capable of giving results that are of an accuracy comparable to that of the cc-pV($X + 1$)Z basis and it appears to be better than the simple 'global' $\alpha = 1.63$ extrapolation from above.

However, one should realize that, since the scheme with varying exponents by construction is biased towards the molecules with the largest errors, there is a risk of failure for molecules with small errors. Indeed, the $Q - 5$ extrapolation for BH recovers only about 50% of the error at the cc-pV5Z level. More seriously, the $D - T$ extrapolation for H₂ overshoots the limit with 100% compared with the error at the cc-pVTZ level and thus does not represent any improvement at all for this molecule. Finally, as for the simple 'global' extrapolation, this scheme performs well for F₂ for small X , but the error of the 5 - 6 extrapolated value is more than a factor of 6 larger than the cc-pV6Z basis-set error.

Table 7

The optimized α , mean error ($\bar{\Delta}$), standard deviation (Δ_{std}), maximum absolute error (Δ_{max}), and mean absolute error ($\bar{\Delta}_{\text{abs}}$) of the extrapolated basis-set limits for the sample of eight diatomics (F₂ being excluded)

X	α	$\bar{\Delta}$	Δ_{std}	Δ_{max}	$\bar{\Delta}_{\text{abs}}$
3	1.43	0.07	1.04	2.23	0.80
4	1.54	-0.02	0.25	0.49	0.17
5	1.94	0.01	0.05	0.11	0.04
6	1.41	< 0.01	0.02	0.03	0.02

The latter four are all given in units of mE_h, and for a given X the two-point extrapolation is based on results with $X' \in [X - 1, X]$.

The overall conclusion thus remains that the safest and most reliable estimate of the basis-set limit is obtained by calculating the energy in the next basis in the series. Moreover, it appears difficult to find a reliable extrapolation scheme for the H-F energy with the cc-pVXZ basis sets. The underlying reason for this difficulty may be that the cc-pVXZ sets are constructed to give systematically improved correlation energies rather than improved H-F energies.

2.8. Two-point linear power extrapolations with an X -dependent α

An extrapolation based on the idea of an optimal exponent for each X pair was recently proposed by Truhlar [11], who minimized the RMS error of extrapolated limits based on the power form (Eq. (3)) with $X \in [2,3]$ for three molecules, obtaining an optimal value of $\alpha = 3.4$. For our broader sample of eight ‘regular’ diatomics, the corresponding value is $\alpha = 3.54$ and the performance of extrapolations of the form (3) with this value of α is summarized by: $\bar{\Delta} = 0.13 \text{ mE}_h$, $\Delta_{\text{std}} = 1.04 \text{ mE}_h$, $\Delta_{\text{max}} = 2.22 \text{ mE}_h$, and $\bar{\Delta}_{\text{abs}} = 0.83 \text{ mE}_h$.

Although the extrapolated results are not as good as the cc-pV6Z results (as originally stated [11]), the performance is quite good – that is, between cc-pVQZ and cc-pV5Z and comparable with the corresponding exponential scheme. As this extrapolation is constructed in much the same way as the exponential one, some similarities are noted. The average error is quite good and the scheme is biased towards the molecules with large errors. Thus, this scheme overshoots the H-F limit for H_2 by 99% compared with the error at the cc-pVTZ level.

2.9. Polyatomic molecules

The results for H_2O and CH_2O obtained with the cc-pVXZ basis sets, the aug-cc-pV6Z basis set, and with the simple ‘global’ extrapolation are given in Table 8. Although we do not have the exact limits for these molecules, a good deal of information can be extracted from the data in Table 8.

First, according to the findings for the diatomics, the errors at the cc-pV6Z and aug-cc-pV6Z levels should be smaller than a few tenths of one mE_h – which for H_2O is in agreement with the near-basis set limit energy of 76.067488 E_h obtained by Mon-

Table 8

H-F energies (in E_h) for the H_2O and CH_2O molecules obtained with the correlation-consistent basis sets and with the simple ‘global’ two-point extrapolation based on $X' \in [X-1, X]$

X	H_2O		CH_2O	
	cc-pVXZ	Extrapolated	cc-pVXZ	Extrapolated
2	-76.026799		-113.876435	
3	-76.057168	-76.064569	-113.911996	-113.920661
4	-76.064835	-76.066703	-113.921007	-113.923203
5	-76.067091	-76.067640	-113.923227	-113.923768
6	-76.067407	-76.067484	-113.923499	-113.923565
a ^a	-76.067462		-113.923516	

^aaug-cc-pV6Z result.

crieff and Wilson, using a large atom-centered basis augmented with many bond functions [25]. The differences between the cc-pVXZ energy for small X and the aug-cc-pV6Z energy is consequently very close to the true basis-set errors. These differences are 40.66 mE_h (D), 10.29 mE_h (T), 2.63 mE_h (Q) for H_2O and 47.08 mE_h (D), 11.52 mE_h (T), 2.51 mE_h (Q) for CH_2O . These errors are similar to the average errors for the diatomics in Table 6. Furthermore, the $D - T$ extrapolations are very close to the cc-pVQZ energies, the $T - Q$ extrapolations are comparable to the cc-pV5Z energies, and the $Q - 5$ extrapolations appear to overshoot the basis-set limit slightly. All these findings agree with the results for the diatomics. Finally, for H_2O and CH_2O , the 5 – 6 extrapolations are respectively 0.02 and 0.05 mE_h below the aug-cc-pV6Z energy and appear to be close to the basis-set limit, as is further substantiated by the excellent agreement between the 5 – 6 extrapolated energy and the near basis-set limit of Moncrieff and Wilson for H_2O . Our conclusions for the diatomics thus appear to be in agreement with the results for H_2O and CH_2O .

3. Conclusions

From comparison with results obtained in numerical orbital studies, the convergence towards the basis-set limit of the H-F total and binding energies has been investigated for the cc-pVXZ basis sets. Solid improvements are obtained with each increment in the cardinal number. At the cc-pV6Z level, the errors

in the total and binding energies are about 0.1 mE_h . For most purposes, therefore, we may regard the cc-pV6Z H-F total and bonding energies as converged to the basis-set limit.

An exponential form (2) fits the basis-set errors of the cc-pVXZ sets for the total molecular energy well and better than does a power form (3). Furthermore, the total energies are in general better fitted than the binding energies. However, it is very difficult to develop a reliable scheme for the extrapolation of the total molecular H-F energy which works well for all X . A three-point exponential scheme is not very successful since it requires the use of a basis with an error 20 times larger than that of the largest basis used in the extrapolation. A simple 'global' two-point scheme with a fixed $\alpha = 1.63$ works better: For $X < 5$, the extrapolated energies are comparable with the energies obtained with the cc-pV(X + 1)Z basis but for larger X the scheme is less reliable and may give poor results. A two-point extrapolation with different α for different X (with α chosen to minimize the RMS error of the two-point extrapolated limits over the sample) gives extrapolated energies that are usually comparable with the energies obtained in the cc-pV(X + 1)Z basis. Unfortunately, this scheme performs poorly in some cases, even for small X . The overall conclusion is therefore that, for a given X , the most reliable estimate of the basis-set limit is obtained by calculating the energy in the larger cc-pV(X + 1)Z basis. Eventually, at the cc-pV6Z level, this scheme yields an energy within 0.1 mE_h of the basis-set limit. If this is not feasible, one should be careful with any extrapolation scheme. Still, for applications where the size of the system prevents one from using basis sets with $X > 4$, the 'global' two-point scheme with $\alpha = 1.63$ (the most reliable scheme for small X) should give an accuracy comparable with that of the cc-pV(X + 1)Z basis and thus improve the accuracy and usefulness of the calculations.

Acknowledgements

We wish to thank Jacek Kobus for running the new numerical calculations on C_2 and F_2 for us. AH

wishes to thank Sonia Coriani for helpful discussions. This work has been supported by the Danish Research Council (Grant No. 9600856).

References

- [1] M.R. Nyden, G.A. Petersson, *J. Chem. Phys.* 75 (1981) 1843.
- [2] G.A. Petersson, M. Braunstein, *J. Chem. Phys.* 83 (1985) 5129.
- [3] D. Feller, *J. Chem. Phys.* 96 (1992) 6104.
- [4] D. Feller, *J. Chem. Phys.* 98 (1993) 7059.
- [5] W. Klopper, *J. Chem. Phys.* 102 (1995) 6168.
- [6] J.M.L. Martin, *Chem. Phys. Lett.* 259 (1996) 669.
- [7] J.M.L. Martin, P.R. Taylor, *J. Chem. Phys.* 106 (1997) 8620.
- [8] T. Helgaker, W. Klopper, H. Koch, J. Noga, *J. Chem. Phys.* 106 (1997) 9639.
- [9] A. Halkier, H. Koch, O. Christiansen, P. Jørgensen, T. Helgaker, *J. Chem. Phys.* 107 (1997) 849.
- [10] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A.K. Wilson, *Chem. Phys. Lett.* 286 (1998) 243.
- [11] D. Truhlar, *Chem. Phys. Lett.* 294 (1998) 45.
- [12] T.H. Dunning, *J. Chem. Phys.* 90 (1989) 1007.
- [13] A.K. Wilson, T. van Mourik, T.H. Dunning, *J. Mol. Struct. (THEOCHEM)* 388 (1996) 339.
- [14] R.N. Hill, *J. Chem. Phys.* 83 (1985) 1173.
- [15] L. Laaksonen, P. Pyykkö, D. Sundholm, *Int. J. Quantum Chem.* 23 (1983) 319.
- [16] D. Sundholm, P. Pyykkö, L. Laaksonen, *Mol. Phys.* 56 (1985) 141.
- [17] P. Pyykkö, D. Sundholm, L. Laaksonen, *Mol. Phys.* 60 (1987) 597.
- [18] J. Kobus, private communication.
- [19] A.R. Hoy, I.M. Mills, G. Strey, *Mol. Phys.* 24 (1972) 1265.
- [20] J.L. Duncan, *Mol. Phys.* 28 (1974) 1177.
- [21] T. Helgaker, H.J.A. Jensen, P. Jørgensen, J. Olsen, K. Ruud, H. Ågren, T. Andersen, K.L. Bak, V. Bakken, O. Christiansen, P. Dahle, E.K. Dalskov, T. Enevoldsen, B. Fernandez, H. Heiberg, H. Hettema, D. Jonsson, S. Kirpekar, R. Kobayashi, H. Koch, K.V. Mikkelsen, P. Norman, M.J. Packer, T. Saue, P.R. Taylor, O. Vahtras, DALTON, an ab initio electronic structure program, release 1.0, 1997. See <http://www.kjemi.uio.no/software/dalton/dalton.html>.
- [22] LUCAS, an atomic program written by J. Olsen, D. Sundholm, Lund.
- [23] P. Jørgensen, T. Helgaker, J. Olsen, Lecture Notes from the Fifth Sostrup Summer School on Quantum Chemistry and Molecular Properties, 1998.
- [24] C.F. Fischer, *The Hartree-Fock Method for Atoms. A Numerical Approach*, Wiley, New York, 1977.
- [25] D. Moncrieff, S. Wilson, *J. Phys. B* 29 (1996) 6009.