A Critical Study of Basis Set Effects and the Use of Approximate Natural Orbitals in SCF-CI Calculations of Molecular Geometries and Heats of Reaction

Gunnar Karlström and Bo Jönsson

Division of Physical Chemistry 2, Chemical Center, P.O.B. 740, S-220 07 Lund, Sweden

Björn O. Roos and Per E. M. Siegbahn

Institute of Theoretical Physics, University of Stockholm, Vanadisvägen 9, S-113 46 Stockholm, Sweden

SCF-CI calculations have been performed on a number of chemical reactions between closed shell molecules in order to determine the heats of reaction. Contracted Gaussian type atomic basis sets of three different qualities were used and the CI calculations were performed in a truncated approximate natural orbital space. The conclusions to be drawn from these calculations are rather pessimistic. For heats of reaction, errors up to 6 kcal/mole are obtained on the SCF-level with a double zeta plus polarization atomic basis. A further improvement is only possible if extended basis sets are used. Correlation effects on heats of reaction are of the same size and CI calculations are therefore only meaningful with large atomic basis sets.

For the CI calculations a one-electron space of approximate natural orbitals, obtained from second order RS perturbation theory, was used. Different truncations, using the occupation number as criterion, were tested. The general conclusion is that errors in energy differences obtained with a truncated basis set are of the same magnitude as the error in the total correlation energy. In practice this means that not more than 20–30% of the approximate natural orbitals can be deleted if the error is to be kept less than a few kcal/mole.

Finally the truncation error in calculations of bond distances was tested for a few cases. Errors of around 10% of the total change due to correlation were found when 30% of the lowest occupied natural orbitals were deleted.

Key words: Basis set effects – Approximate natural orbitals

1. Introduction

A proper choice of the atomic basis set and the molecular one-electron space is of crucial importance in SCF-CI calculations of energy surfaces for chemical reactions. It is necessary to know to what extent a given calculation can be expected

to be of predictive value. At the same time, it is desirable to keep the number of basis functions as small as possible both in the SCF and in the CI part of the calculation. Smaller basis sets mean that larger systems can be studied and also lead to a reduction of the computational effort.

One quantity which is of interest in studies of chemical reactions is the heat of reaction, that is, the energy difference between the products and the reactants. This quantity is often well known experimentally, and is therefore suitable for testing purposes. In the literature one also finds a number of reports on calculations of heats of reaction (especially hydrogenation energies) which have been made in order to investigate the reliability of theoretical predictions [1–4]. These investigations have been made on the Hartree–Fock (HF) level of approximation with the primary aim of testing the basis set effects on such predictions. It is then implicitly assumed that energies of reactions between closed shell systems can to a good approximation be calculated within the HF approximation [1]. The results of these calculations indicate that large basis sets including polarization functions are in general necessary, if energy differences close to the HF limiting values are to be obtained. However, they also show that errors as large as 10–20 kcal/mole may remain at the HF limit. Since such errors are too large to be acceptable in accurate studies of chemical reactions, it is also of interest to investigate the correlation effects on heats of reaction in some detail.

A number of results are available, e.g. [4–12], which shows the correlation effects on energy differences in reactions between closed shell systems, but so far no systematic study of the basis set effects, and the error due to the approximations involved in the calculation of the correlation energy, seems to be available.

The most common procedure to calculate the correlation energy is to use the method of configuration interaction (CI), where the wave function is expanded in a given set of configurations built from a chosen set of one-particle functions. Usually in studies of closed-shell systems the CI expansion is limited to single and double replacements out of a given reference state, which is normally taken to be the HF wave function. One way of truncating this expansion is to delete some part of the virtual one-particle space. This is most easily done if the one-particle space is taken to be approximate natural orbitals (ANO's) or by using the Pseudo-Natural-Orbital Configuration Interaction Scheme [13]. If these ANO's are good approximations to the true natural orbitals of the systems, there exists a close relation between the occupation number of a given ANO and its importance in the CI wave function. It is therefore of interest to investigate the possibility of deleting ANO's with small occupation numbers from the one-particle space. The convergence properties of the total correlation energy with respect to this type of truncation of the CI expansion have earlier been investigated to some extent [11, 14, 15, 17]. However, from these studies no general conclusions can be made about the efficiency of the ANO's in calculations of equilibrium geometries, heats of reaction or potential energy surfaces in general for closed-shell systems. The usefulness of the ANO's may also be basis set dependent.

In the present work we report an investigation of some hydrogenation reactions,

where effects of the atomic basis set choice and truncation of the ANO space have been studied. SCF-CI calculations have been performed using the direct CI technique [16]. The details of the calculations are presented in Sect. 2. In Sect. 3 we present the results for the hydrogenation reactions, and in Sect. 4 the results from some studies of equilibrium geometries.

2. Computational Methods and Details

All calculations were performed with the program system MOLECULE-CI [16–18] on the UNIVAC 1108 at the University of Lund. Three different atomic basis sets were chosen extending from "double zeta" to a large basis set including two sets of polarization functions on the heavy centres and one set of p-type functions on hydrogen. Some details of these basis sets are given in Table 1.

Table 1. Atomic basis sets

Atom	Basis	Uncontracted basis	Contracted basis	Contraction scheme	Polarization functions	Ref.
C, O	I	$10s, 6p$	$6s, 4p$	$(5, 1, 1, 1, 1, 1; 3, 1, 1, 1)$	$2d(\zeta=1.2$ and $0.3)$	[19]
H	I	$5s$	$3s$	$(3, 1, 1)$	$1p(\zeta=0.8)$	,, [a]
C, O	II	$9s, 5p$	$4s, 2p$	$(6, 1, 1, 1; 4, 1)$	$1d(\zeta=0.8)$	,,
H	II	$4s$	$2s$	$(3,1)$	$1p(\zeta=0.8)$	,,
Cl	III	$12s, 9p$	$6s, 4p$	$(6, 2, 1, 1, 1, 1; 6, 1, 1, 1)$	—	[20]
C, O	III	$7s, 3p$	$4s, 2p$	$(4, 1, 1, 1; 2, 1)$	—	[21][a]
H	III	$4s$	$2s$	$(3,1)$	—	[17][a]

[a] The hydrogen s-exponents were multiplied with the factor 1.34.

The actual calculation of the CI wave function involves several steps and can be briefly described as follows:

1) Select M atomic basis functions. They are automatically transformed into symmetry adapted basis functions by MOLECULE. Compute the two-electron supermatrix in this basis, and
2) obtain the LCAO-MO-SCF wave function with m occupied orbitals.
3) Calculate the two-electron integrals over the symmetry adapted basis.
4) Construct molecular two-electron integrals of the type (ai/bj) where i and j are occupied valence orbitals (the core electrons are left uncorrelated) and a and b virtual orbitals. This is a much faster procedure than the full transformation, since in general m is much smaller than M.
5) Construct ANO's by a second order RS perturbation calculation using the transformed two-electron integrals. The density matrix is directly obtained from the list of integrals which is a very efficient procedure [17]. The occupied SCF orbitals are not changed in this method.
6) Select the N ANO's having the largest occupation numbers. Transform the two-electron integrals over the symmetry adapted basis to two-electron integrals over a basis consisting of the valence part of the occupied orbitals and the ANO's.

7) Perform a CI calculation including all single and double replacements relative to the SCF ground state in this one-electron basis. In order to minimize the computational efforts we saved the transformed two-electron integrals over the full ANO space and sorted out the integrals needed for the truncated CI calculations, which is a very fast procedure.

Using the methods and technique described above we have performed a series of *ab initio* MO-LCAO-SCF-CI calculations on a number of compounds using the three different basis sets and a varying number of ANO's in the CI wave function. The different results of these calculations will be discussed in their context in the following paragraphs.

3. Heats of Reaction

When a heat of a chemical reaction is studied experimentally two different values are frequently reported, the ΔH-value and the ΔU-value, depending on whether the measurement is made under constant pressure or constant volume. At zero temperature where there is no translational motion these two values coincide. Computationally the heat of a reaction at zero temperature is calculated as the

Table 2. SCF atomization energies using different basis sets (energy values in kcal/mole)

Reaction	Estimated HF value[a]	STO extended	GTO extended	Basis set, this work		
				I	II	III
$H_2O \rightarrow 2H + O$	162 ± 1	159.4^b	161.0^c	159.1	138.4	116.8
$CO \rightarrow C + O$	184 ± 1	182.9^d	—	180.4	148.7	131.7
$CH_4 \rightarrow C + 4H$	332 ± 1	—	330.6^e	330.0	325.9	312.5

[a] HF limit values for the free atoms are taken from Ref. [23]. The HF energy for H_2O is assumed to be -76.067 ± 0.002 a.u. [24]; for CO -112.791 ± 0.001 a.u. [25] and for CH_4 -40.220 ± 0.002 a.u. [26].
[b] $5s, 4p, 1d$ from Ref. [23].
[c] $13s, 8p, 3d, 1f$ for oxygen, contracted to $8s, 5p, 3d, 1f$ and $6s, 2p, 1d$ for hydrogen, contracted to $4s, 2p, 1d$ [24].
[d] $5s, 4p, 1d$, and $1f$, from Ref. [27].
[e] $12s, 6p, 3d, 1f$ for carbon, contracted to $10s, 6p, 3d, 1f$, and $6s, 1p$ for hydrogen contracted to $4s, 1p$. One s-function is added for each bond [26].

total energy difference between the products and the reactants with a correction for zero point vibration. There are two main reasons for performing calculations of heats of reaction. First, reaction involving radicals or excited states are hard to study experimentally and theoretical predictions of these heats of reaction may therefore be important. Second, since many heats of reaction are known to a high accuracy experimentally, they can be used to calibrate the calculations of the corresponding full potential energy surfaces. In our investigation, which is a pure methodological one, we have concentrated on three aspects of the calculation of

Table 3. SCF hydrogenation energies calculated with different atomic basis sets (energies in kcal/mole)

Reaction	Estimated HF value[a]	I	II	III	GTO[b] SB	GTO[b] HP
		Basis sets, this work				
1. $CO + 3H_2 \rightarrow CH_4 + H_2O$	-60 ± 3	-58.2	-65.8	-56.0	-81.5	-58.4
2. $CO + H_2 \rightarrow H_2CO$	0 ± 4	$+1.7^c$	$+0.4$	-3.6	-11.4	$+0.9$
3. $C_2H_2 + 3H_2 \rightarrow 2CH_4$	-113 ± 4	-112.4	-118.1	-115.8	-120.9	-117.9
4. $C_2H_2 + H_2 \rightarrow C_2H_4$	-54 ± 7	-53.8	-54.4	-50.8	-54.4	-53.2
5. $CO + 2H_2 \rightarrow CH_3OH$	-43 ± 8	—	-34.6	-27.8	-53.8	-28.4

[a] The HF-limit values for H_2O, CH_4 and CO are given in footnotes to Table 3. The HF energy for H_2CO is assumed to be -113.925 a.u. [28], for C_2H_2 -76.860 a.u. [29], for C_2H_4 -78.080 a.u. [2], for CH_3OH -115.127 a.u. [30] and for H_2 -1.134 a.u. [31].
[b] For comparison we have added the results of Snyder and Basch (SP) [3] and Hariharan and Pople (HP) [4]. SB used a double zeta basis constructed from $10s$ and $5p$ primitive Gaussians and HP used a 6-31G** basis with polarization functions of approximately the same quality as our basis set II.
[c] Estimated from [28] where a very similar basis set is used.

heats of reaction. In the first section we discuss the choice of atomic basis set on the one configurational SCF level. The next two sections deal with the calculation of the correlation effects on heats of reaction by means of the configuration interaction method. Again the atomic basis set choice is discussed. In order to make CI expansions shorter, and thereby more efficient, truncations based on a natural orbital transformation are often performed. We have investigated the convergence of the heat of the reaction as a function of the number of approximate natural orbitals

Table 4. Correlation contributions to the hydrogenation energies with different basis sets (energies in kcal/mole)

Reaction	$\Delta E_E^a - \Delta E_{HF}$	ΔE_{corr} Basis I	Basis II	Basis III
1. $CO + 3H_2 \rightarrow CH_4 + H_2O$	-4 ± 3	-4.5 (-3.9)[b]	$+1.2$ ($+1.5$)	$+17.1$ ($+21.9$)
2. $CO + H_2 \rightarrow H_2CO$	-7 ± 4	—	-3.2 (-2.8)	$+2.1$ ($+1.9$)
3. $C_2H_2 + 3H_2 \rightarrow 2CH_4$	$+7 \pm 4$	-1.6 ($+1.1$)	$+0.7$ ($+3.9$)	$+14.8$ ($+19.5$)
4. $C_2H_2 + H_2 \rightarrow C_2H_4$	$+6 \pm 6$	—	$+0.1$ ($+1.0$)	$+5.6$ ($+7.0$)
5. $CO + 2H_2 \rightarrow CH_3OH$	$+10 \pm 8$	—	—	$+15.5$ ($+16.3$)

[a] The experimental ΔE values, corrected for contribution from the zero-point vibrational energy, have been taken from Ref. [3].
[b] Values within parentheses include the unlinked cluster contribution to the correlation energy estimated according to $E_{ulc} = (1 - C_0^2)E_{corr}$, where C_0 is the coefficient of the HF reference configuration in the CI wave function.

kept in the expansion. The approximate natural orbitals were constructed by the use of second order Rayleigh-Schrödinger perturbation theory as outlined in Sect. 2. Five different reactions involving the eight molecules H_2, H_2O, CH_4, C_2H_2, C_2H_4, CO, H_2CO and CH_3OH were investigated. The reactions are given in Tables 3 and 4.

3.1. SCF Energy Differences and Their Dependence on the Atomic Basis Set Choice

In all calculations reported here contracted Gaussian type orbitals were used. In order to get an idea of the dependence of the heat of reaction on the basis set choice

Table 5. SCF energies and correlation energies as function of occupation number limit for deleted ANO's [a]

Compound	Basis	Number of ANO's	Approximative occupation number limit	$-E_{SCF}$ a.u.	$-E_{corr}$ a.u.	C_0
C_2H_4	III		no truncation	77.946811	0.19428	0.956061
C_2H_2	III		,,	76.739791	0.18344	0.956345
CO	III	11	no truncation	112.56579	0.19740	0.963674
		10	10^{-3}		0.19441	0.965059
		7	$3 \cdot 10^{-3}$		0.18070	0.966810
		6	$6 \cdot 10^{-3}$		0.16541	0.972310
		5	$9 \cdot 10^{-3}$		0.15983	0.978145
		4	10^{-2}		0.13982	0.973101
H_2CO	III	14	no truncation	113.69188	0.21511	0.960263
		13	10^{-3}		0.21212	0.970897
		8	$6 \cdot 10^{-3}$		0.19807	0.973219
		5	$9 \cdot 10^{-3}$		0.14984	0.978115
CH_3OH	III	17	no truncation	114.86396	0.21817	0.964791
		14	10^{-3}		0.21438	0.965495
		11	$3 \cdot 10^{-3}$		0.20343	0.966985
		6	$9 \cdot 10^{-3}$		0.14418	0.973394
H_2	III	3	no truncation	1.12599	0.02384	0.993350
		2	10^{-3}		0.02352	0.993702
		1	$7 \cdot 10^{-3}$		0.01828	0.994866
CH_4	III	12	no truncation	40.15113	0.11414	0.975622
		10	10^{-3}		0.11088	0.976450
		8	$3 \cdot 10^{-3}$		0.10862	0.977431
		5	$7 \cdot 10^{-3}$		0.08463	0.981447
H_2O	III	8	no truncation	75.88473	0.12616	0.982499
		6	10^{-3}		0.12493	0.983333
		5	$3 \cdot 10^{-3}$		0.11760	0.984321
		4	$7 \cdot 10^{-3}$		0.11138	0.988365
CH_4	II	30	no truncation	40.20171	0.17610	0.969090
		19	10^{-3}		0.16639	0.969898
		9	$3 \cdot 10^{-3}$		0.12906	0.973460
		4	$7 \cdot 10^{-3}$		0.07316	0.981269
CH_3OH	II		—	115.04014	—	—
H_2O	II	20	no truncation	76.01945	0.19333	0.974162
		10	10^{-3}		0.17718	0.975647
		9	$3 \cdot 10^{-3}$		0.17368	0.975909
		4	$7 \cdot 10^{-7}$		0.11014	0.982161

Compound	Basis	Number of ANO's	Approximative occupation number limit	$-E_{SCF}$ a.u.	$-E_{corr}$ a.u.	C_0
C_2H_2	II	33	no truncation	76.82154	0.24938	0.952298
		26	10^{-3}		0.24238	0.952760
		12	$3 \cdot 10^{-3}$		0.19375	0.956313
		5	$7 \cdot 10^{-3}$		0.12761	0.964230
C_2H_4	II	42	no truncation	78.03945	0.27807	0.950286
		36	10^{-3}		0.26968	0.951136
		14	$3 \cdot 10^{-3}$		0.21143	0.955498
		6	$7 \cdot 10^{-3}$		0.12522	0.964885
H_2	II	9	no truncation	1.13127	0.03542	0.991785
		4	10^{-3}		0.03368	0.991952
		2	$3 \cdot 10^{-3}$		0.02491	0.993034
		1	$7 \cdot 10^{-3}$		0.01808	0.994484
CO	II	23	no truncation	112.72241	0.26466	0.957048
		17	10^{-3}		0.25848	0.957589
		12	$3 \cdot 10^{-3}$		0.23516	0.959881
		4	$7 \cdot 10^{-3}$		0.13653	0.973301
H_2CO	II	32	no truncation	113.85938	0.29867	0.954013
		21	10^{-3}		0.28651	0.955053
		14	$3 \cdot 10^{-3}$		0.25096	0.957760
		5	$7 \cdot 10^{-3}$		0.14703	0.966093
CO	I	51	no truncation	112.78113	0.30938	0.955812
		43	$2 \cdot 10^{-5}$		0.30920	0.955775
		35	10^{-4}		0.30636	0.955901
		23	$2 \cdot 10^{-4}$		0.29482	0.956256
CH_4	I	48	no truncation	40.21245	0.19458	0.966808
		42	$2 \cdot 10^{-5}$		0.19438	0.966830
		34	10^{-4}		0.19180	0.966927
		25	$3 \cdot 10^{-4}$		0.18508	0.967192
C_2H_2	I	63	no truncation	76.84787	0.28000	0.949678
		57	$4 \cdot 10^{-6}$		0.27995	0.949688
		50	$3 \cdot 10^{-5}$		0.27945	0.949700
		32	$2 \cdot 10^{-4}$		0.27086	0.949760
C_2H_4	I		—	78.06226	—	—
H_2O	I	36	no truncation	76.05950	0.23054	0.972102
		31	$2 \cdot 10^{-5}$		0.23042	0.972156
		25	10^{-4}		0.22726	0.972575
		20	$3 \cdot 10^{-4}$		0.22217	0.972889
H_2	I	11	no truncation	1.13262	0.03573	0.991650
		10	$2 \cdot 10^{-5}$		0.03572	0.991650
		9	$3 \cdot 10^{-5}$		0.03566	0.991672
		8	$8 \cdot 10^{-5}$		0.03547	0.991728
		6	$2 \cdot 10^{-4}$		0.03481	0.991639
		5	$3 \cdot 10^{-4}$		0.03432	0.991815
		4	$3 \cdot 10^{-4}$		0.03372	0.991858
		2	$2 \cdot 10^{-3}$		0.02510	0.992907
		1	$4 \cdot 10^{-3}$		0.01818	0.994336

[a] All geometries are experimental. See: Snyder, L. C., Basch, H.: Molecular wave functions and properties: Tabulated from SCF calculations in a Gaussian basis set. New York: John Wiley & Sons 1972 for further references.

three different sizes of basis sets were investigated (cf. Table 1). The smallest basis set labelled III is of double zeta type, comparable in quality to the frequently used STO 4-31G basis [22]. In basis set II, *d*-type polarization functions are added to the first row atoms and *p*-type functions on hydrogen. There is also an increased number of uncontracted Gaussian *s*, *p*-functions. Finally basis set I is of better than triple-zeta quality and has two *d*-type functions on the first row atoms.

The first and perhaps most severe test of the adequacy of the basis set to describe heats of reaction is shown in Table 2, where the molecule is dissociated into its component atoms. Only basis set I gives reasonable dissociation energies, around 3 kcal/mole lower than the estimated HF values. The results are lower than the correct values probably because atom-optimized basis sets have been used. The errors with basis set II are quite different for the three molecules, which may be understood in the following way. CH_4 will have a comparatively high density of basis functions with only a small fraction of the charge centered far away from where the basis functions are located, and the error is therefore only 6 kcal/mole. On H_2O the error has increased to 24 kcal/mole which probably has to do with the difficulty of describing the oxygen lone-pair with basis functions centered on oxygen. For CO the error is as large as 35 kcal/mole, illustrating the well known fact that large basis sets are needed to describe the complicated binding situation between two first-row atoms. Finally the double zeta quality basis set III is completely inadequate for calculating these types of ΔE-values with errors of up to 52 kcal/mole for CO, which is 28% of the total Hartree–Fock binding energy.

The changes in basis set requirements between free atoms and molecules may however be expected to be larger than between different closed-shell molecules since the basis sets are optimized for free atoms, and it is therefore usually assumed that it is easier to reproduce ΔE_{HF} values for ordinary closed-shell chemical reactions. Hurley [2] concludes that the consistent use of a good double zeta plus polarization functions basis set would probably reproduce the ΔE_{HF} values for ordinary chemical reactions well. In Table 3 the calculated ΔE-values for five hydrogenation reactions using the three different basis sets are presented, together with the estimated Hartree–Fock limit values. The difficulty in estimating the HF values makes a comparison of accuracy between Tables 2 and 3 harder, but in general it seems that the calculated results in Table 3 are more accurate than those in Table 2, which supports the above statement concerning the accuracy for ordinary molecular reactions. We note however that there is an extreme sensitivity of the results to details of the basis set. This is particularly clear from the comparison between the two basis sets of double zeta quality, our basis set III and the basis set used by Snyder and Basch [3]. Differences of up to 25 kcal/mole are obtained for the first and last reaction. It is obvious already from the results of Table 2 that a double zeta basis set does not give a good description of the binding in the molecules studied. A good result for a molecular reaction therefore has to rely on a cancellation of errors which originates in an equally bad description of reactants and products. Such a cancellation of errors apparently takes place for reactions 3 and 4 but not for 1 and 5. Many reasons may be found to explain these results for basis set III, such as the capability of the basis set to describe

Table 6. Correlation contribution to the hydrogenation energies for different truncations of the ANO space (energies in kcal/mole)

Reaction	Basis	No truncation		Truncation 1			Truncation 2			Truncation 3		
		ΔE_{corr}	Number ANO's	ΔE_{corr}	Number ANO's	OCC num. limit	ΔE_{corr}	Number ANO's	OCC num. limit	ΔE_{corr}	Number ANO's	OCC num. limit
$CO+3H_2 \rightarrow$ CH_4+H_2O	I	−3.9	84	−3.8	73	$3 \cdot 10^{-4}$	−2.4	59	10^{-4}	−2.7	45	$3 \cdot 10^{-4}$
	II	+1.5	50	+12.1	29	$8 \cdot 10^{-4}$	+6.9	18	$3 \cdot 10^{-3}$	—	—	—
	III	+21.9	20	+22.4	16	10^{-3}	+19.1	13	$3 \cdot 10^{-3}$	—	—	—
$CO+H_2 \rightarrow H_2CO$	II	−2.8	32	+0.3	21	$8 \cdot 10^{-4}$	+3.7	14	$3 \cdot 10^{-3}$	+3.0	5	$7 \cdot 10^{-3}$
	III	+1.9	14	+2.2	13	10^{-3}	+6.0	8	$3 \cdot 10^{-3}$	+6.9	5	$9 \cdot 10^{-3}$
$CO+2H_2 \rightarrow CH_3OH$	III	+16.3	17	+16.5	14	10^{-3}	+15.1	11	$3 \cdot 10^{-3}$	+22.6	6	$7 \cdot 10^{-3}$
$C_2H_2+3H_2 \rightarrow 2CH_4$	I	+1.1	96	+1.2	84	$3 \cdot 10^{-5}$	+2.6	68	10^{-4}	+5.7	50	$3 \cdot 10^{-4}$
	II	+3.9	60	+8.8	38	$8 \cdot 10^{-4}$	+8.6	18	$3 \cdot 10^{-3}$	+24.3	8	$9 \cdot 10^{-3}$
$C^2H^2+H^2 \rightarrow C^2H^4$	II	+1.0	42	+2.0	30	$8 \cdot 10^{-4}$	+3.4	14	$3 \cdot 10^{-3}$	+12.8	6	$9 \cdot 10^{-3}$

σ-π reorganization and the incapability to describe charge transfer. Such arguments are not too useful however, since the number of explanations quickly becomes larger than the number of reactions studied. The double zeta basis set therefore has to be used either purely empirically, or maybe with more confidence after extensive basis set optimizations on the separate molecules. For basis set II which is essentially equivalent to the basis set of Hariharan and Pople, the errors have decreased somewhat but many of the arguments given above are still applicable for this basis set. The large effect of adding polarization functions noticed in Table 2 is diminished in Table 3 because they usually give a similar contribution on both sides of the reaction. In spite of these cancellation effects, errors of around 7 kcal/mole are common with basis set II when compared with basis set I, and again there does not seem to be any consistent way of correcting the results without further optimization of exponents or enlargement of the basis set. Finally the results with basis set I are all close to the estimated HF limit, as could be expected from the results of Table 2. Even with this basis set some cancellation of errors is found to occur, since the difference in the HF-value in Table 2 is around 3 kcal/mole and in Table 3 only around 1 kcal/mole.

In conclusion basis set I gives very accurate heat of reactions, close to the HF limit, whereas the double zeta basis sets (Basis set III) are unpredictable, yielding large errors in some cases. For basis set II the situation is less clear; the basis set may be used if errors of up to around 6 kcal/mole are tolerated. Finally, an accurate estimate of the quality of a basis set and a prediction of the size of the errors can be obtained by studying the deviation from the HF limit for the energy of dissociation into component atoms, as was done in Table 2.

3.2. CI Energy Differences and their Dependence on the Atomic Basis Set Choice

Since the one-configurational Hartree–Fock approximation is known to describe closed-shell molecules in their equilibrium configurations rather accurately, the correlation effects on the heats of reaction between such molecules might be expected to be small. This is not generally true however, as can be seen in Table 4, where the correlation contributions to the heat of reaction have been estimated for the five reactions discussed in the preceding section. The correlation contribution for reaction 5 is for example 10 kcal/mole which is evidently not negligible if a quantitative prediction of a heat of reaction is to be made. There are two different aspects on the choice of atomic basis set that should be considered when the correlation effects on the heats of reaction are calculated. First, it is well known that the correlation energy shows a much slower convergence behaviour with respect to the size of the atomic basis set than the Hartree–Fock energy. On the other hand, since the correlation effect is generally smaller than the value obtained on the HF level, we may be satisfied with a larger relative error on the CI-level than on the SCF-level. Since about the same absolute error is wanted on the SCF-level as on the CI-level, the question is thus whether a smaller or larger basis set is needed to describe the correlation effects.

The correlation energy was calculated using the direct CI method [16, 17], and

Table 7. Some equilibrium geometries and force constants obtained with different ANO truncations using basis set III[a]

System	SCF r_e	SCF k	No truncation r_e	k	No. of ANO's	OCC. no. limit	Truncation 1 r_e	k	No. of ANO's	OCC. no. limit	Truncation 2 r_e	k	No. of ANO's	OCC. no. limit	Truncation 3 r_e	k	No. of ANO's	OCC. no. limit
$H_3O_2^-$ [b]	0.223	0.0400	0.155	0.061	14	0.0	0.611	0.060	9	$2 \cdot 10^{-3}$	0.131	0.064	6	$1 \cdot 10^{-2}$	0.087	0.070	3	$2 \cdot 10^{-2}$
HCl [c]	1.305	0.338	1.329	0.331	11	"	1.327	0.325	6	$5 \cdot 10^{-4}$	1.330	0.325	4	$3 \cdot 10^{-3}$	1.327	0.363	3	$7 \cdot 10^{-3}$
H_2 [d]	0.735	0.394	0.757	0.388	3	"	0.752	0.389	2	10^{-3}	0.757	0.386	1	$7 \cdot 10^{-3}$	—	—	1	—

[a] Distances are in Ångström units and force constants in atomic units.

[b] r_e is here the deviation from a symmetric position of the central hydrogen atom, with an assumed OO distance of 2.65 Å

[c] HF-limit values are $r_e = 0.736$ Å and $k = 0.386$ calculated from Ref. [37], and experimental values $r_e = 0.742$ Å and $k = 0.368$ calculated from Ref. [38].

[d] HF-limit values are $r_e = 1.264$ Å and $k = 0.375$ from Ref. [39], and the experimental values $r_{e^-} = 1.275$ Å and $k = 0.331$ from Ref. [38].

the wave function comprised all single and double replacement from the HF reference configuration. It is well known that the correlation energies obtained for the separate systems with such a non-complete configuration expansion are not additive. This is due to the unlinked cluster contribution to the correlation energy which is partly included when the energies are added, but are not included in a parallel way if the number of electron pairs in the subsystems change during the reaction. In order to avoid calculations on the supermolecule a correlation has therefore been added to the correlation energy, which corrects for the non-additivity [32].

As the correlation contribution to the hydrogenation energies studied varies between 4 and 10 kcal/mole, the SCF value must be calculated with a basis set of at least the same size as basis set I in order to keep the error on the SCF level smaller than the total correlation contribution. Even with the fairly large basis set II the error is sometimes larger than the total correlation effect.

3.3. The Effects of Using Truncated ANO Space

It was shown in the last section that very large basis sets are needed in order to make useful predictions of correlation effects on energy differences in chemical reactions. In fact larger basis sets are needed to describe the correlation effects than to obtain an SCF result of a given accuracy, at least when the error is small. The situation is very troublesome since CI calculations require about five times as much computation time as the corresponding SCF calculation. A normal CI calculation consists of three parts, integral evaluation, integral transformation and the solution of the secular problem. The integral programs available today are very efficient and the integral evaluation therefore takes a relatively small portion of the total computing time. The transformation time is proportional to $M \cdot N^4$ (see Sect. 2 for definition of M and N). Consequently it is highly desirable to minimize the number of orbitals used to build the CI wave function. The time spent in finding the lowest root and eigen vector to the CI matrix is proportional to $m^2(N-m)^2 N^2$ [16]. One way to speed up this part of the calculation is to delete configurations which give very small contributions to the correlation energy. The drawback of this method is that it only speeds up the matrix diagonalization, while the integral transformation, which is equally time consuming when large basis sets are used, is not affected. It would be more effective to be able to delete orbitals from the virtual one-particle space. Unfortunately all virtual canonical HF orbitals give approximately equal contributions to the correlation energy [2, 15, 17]. Consequently it is not possible just to exclude virtual canonical orbitals from the one-particle space.

Natural spin-orbitals (NSO) and for closed-shell systems natural orbitals (NO) have been shown to have optimal convergence properties relative to the wave function [33–35]. The properties of the NO's have been further discussed by Davidson [36]. Knowledge of the natural orbitals implies however that the wave function is known and no calculation is then necessary. Several methods have been proposed to construct approximate natural orbitals [36] and they all repro-

duce the NO's surprisingly well. In the present work we have constructed ANO's using Rayleigh-Schrödinger perturbation theory [17], but the conclusions drawn should be valid for all types of ANO's. Unfortunately very little is known about the relative convergence of NO's or ANO's for different systems containing the same atoms. The only study of this type has been made by Hay [14], who determined the dissociation energies for H_2O and O_2 using different number of ANO's. Hay's study shows for example that using only 14 ANO's gives an error of 2.5 kcal/mole for the dissociation energy of H_2O compared to an untruncated calculation with 26 orbitals. Using 18 ANO's out of total space of 30 gives an error of 8.8 kcal/mole for the dissociation energy of O_2. The reactions studied by Hay are however not of the closed-shell type and little could be said about the effect of other basis sets.

In the present work we have studied the effect of truncating the ANO space on the hydrogenation energies for the five reactions given in Table 3. The calculated SCF energies, and the correlation energies as function of the occupation number limit for the deleted ANO's, are presented in Table 5. This table also gives the coefficient C_0 for the HF reference configuration in the CI wave function. From these results we have obtained ΔE values for the hydrogenation reactions, which are presented in Table 6.

Several interesting features of the results presented in Table 5 could be noticed:

1) A CI calculation on H_2O with 4 ANO's constructed from a double zeta basis gives a larger correlation energy than a CI calculation with 4 ANO's constructed from a double zeta plus polarization functions basis. This indicates that the ANO's are not optimized to include maximum correlation energy, but maximum correction to the wave function.
2) When a large but equivalent number of ANO's are used, a large atomic basis is always superior to a smaller.
3) The contribution to the correlation energy from the 20% of the ANO's, which have the lowest occupation numbers, amounts to only a few kcal/mole.
4) The effect of the least populated ANO's diminishes when larger basis sets are used.
5) It is not possible to assign an occupation number limit for a given error in the correlation energy. This limit has to be decreased when the basis set is increased in order to keep the error constant.

Analysis of Table 6 together with the results of Table 5 shows that ΔE-values calculated with a truncated ANO basis reproduces those of a full calculation well, only as long as the total correlation energies obtained with a full basis are well reproduced by those calculated with the truncated basis. It should be noted that the occupation number limits have been kept constant for all reactions within a given basis, but different limits have been used for the different basis sets. The inequivalent convergence properties of the ANO expansions may be due to the fact that the coefficients in the CI-expansion to first order have the form $H_{0\mu}/\Delta E_\mu$, where $H_{0\mu}$ is a matrix element between a doubly excited configuration and the HF state and ΔE_μ the corresponding energy difference. The energy contribution

is on the other hand $H_{0\mu}^2/\Delta E_\mu$. Since $H_{0\mu}$ in general is a small number, configurations corresponding to small energy differences will become too important. Consequently, as could be seen from Table 5, the energy contribution from an orbital in the H_2 molecule, where the ΔE_μ's are smaller, is less than from an orbital with the same occupation number in e.g. CO where the energy differences are larger. Also the number of ANO's which falls between two occupation numbers may be very different. Compare for example H_2O and CH_4 with basis set II. Only 36.7% of the ANO's in CH_4 fall between the occupation numbers 0.0 and 10^{-3}, but 50% of the ANO's in H_2O. It is even worse between the two limits 10^{-3} and $3 \cdot 10^{-3}$ where the corresponding figures are 33.3 and 5%, respectively. Thus it can be concluded that ANO expansions for different systems are not parallel, and that good estimates of correlation energy between different systems require near convergence in the calculated correlation energy for the separated systems. Nevertheless ANO expansions may be a powerful tool for shortening CI expansions since a fairly large number of ANO's have a very small effect on the calculated correlation energy, a number which increases with basis set size. For example, only 70% of the virtual space of basis I is needed to obtain results with an accuracy of approximately 1 kcal/mole for the ΔE values of the reactions studied here.

4. Basis Set Effect on Equilibrium Geometries

In the previous paragraphs we have studied the quality of different basis sets with respect to large changes in the internal geometry of the system, i.e. undergoing a chemical reaction. It might also be of some interest to investigate whether the conclusions drawn from these studies hold also in the case of small changes in the geometry, i.e. when calculating properties like bond distances, force constants etc. Table 7 presents the results of some calculations of bond distances and force constants with different truncations in the ANO space. These calculations were performed with the small basis set III and the conclusions which can be drawn from them therefore have a rather limited value. It seems, however, as if the general conclusions made in the previous sections hold also here. A reduction of the ANO space with 20–30% of the least occupied orbitals leads to only minor changes in the results (less than 10% of the total correlation effect is lost). Greater truncations still give good results in some of the studied cases, but are of less predictive value.

5. Concluding Remarks

We may conclude from the present work that it is a difficult task to attempt to calculate heats of reaction between closed-shell molecules with an accuracy better than about 10 kcal/mole. To obtain results with this accuracy on the HF level one is compelled to use basis sets of better than "double zeta plus polarization" quality. The correlation effects on heat of reactions are of this order of magnitude and therefore have to be considered. Extended basis sets are needed also here.

Calculations of the correlations effects, using the CI method with a one-particle

space consisting of approximate natural orbitals, give errors in relative energies which are of the same size as the error in the total correlation energy. It is therefore only possible to delete those natural orbitals which have a negligible effect on the correlation energy. The present calculations indicate that it is possible to delete around 20% of the least occupied natural orbitals without loosing more than a few per cent in accuracy. This seems to hold both for heats of reaction and equilibrium properties of the molecules.

References

1. Snyder, L. C.: J. Chem. Phys. **46**, 3602 (1967)
2. Hurley, A. C.: Advan. Quantum Chem. **7**, (1973)
3. Snyder, L. C., Basch, H.: J. Am. Chem. Soc. **91**, 2189 (1969)
4. Hariharan, P. C., Pople, J. A.: Theoret. Chim. Acta (Berl.) **28**, 213 (1973)
5. Karlström, G., Jönsson, B., Roos, B., Wennerström, H.: J. Am. Chem. Soc. **98**, 6851 (1976)
6. Støgård, A., Strich, A., Almlöf, J., Roos, B.: Chem. Phys. **8**, 405 (1975)
7. Diercksen, G., Kraemer, N., Roos, B.: Theoret. Chim. Acta (Berl.) **36**, 249 (1975); **42**, 77 (1976)
8. Meyer, W., Jakubetz, W., Schuster, P.: Chem. Phys. Letters **21**, 97 (1973)
9. Jönsson, B., Karlström, G., Roos, B., Wennerström, H.: Chem. Phys. Letters **41**, 317 (1976)
10. Jönsson, B., Karlström, G., Wennerström, H., Forsén, S., Roos, B., Almlöf, J.: J. Am. Chem. Soc. (in press)
11. Weber, J., Yoshimine, M., McLean, A. D.: J. Chem. Phys. **64**, 4159 (1976)
12. Keil, F., Ahlrichs, R.: J. Am. Chem. Soc. **98**, 4787 (1976)
13. Meyer, W.: J. Chem. Phys **58**, 1017 (1973)
14. Hay, P. J.: J. Chem. Phys. **59**, 2468 (1973)
15. Siu, A. K. Q., Hayes, E. F.: J. Chem. Phys. **61**, 37 (1974)
16. Roos, B.: Chem. Phys. Letters **15**, 153 (1972)
17. Roos, B., Siegbahn, P.: The direct configuration interaction method from molecular integrals, in: Modern theoretical chemistry, 3. Methods of electronic structure theory, Schaefer III, F. H., ed. New York: Plenum Publishing Corp.
18. Almlöf, J.: MOLECULE – a program system for non-empirical calculations of the electronic structures of molecules. USIP Report 74-29, University of Stockholm 1974
19. Huzinaga, S.: J. Chem. Phys. **42**, 1293 (1965)
20. Veillard, A.: Theoret. Chim. Acta (Berl.) **12**, 405 (1968)
21. Roos, B., Siegbahn, P.: Theoret. Chim. Acta **17**, 209 (1970)
22. Hehre, W. J., Ditchfield, R., Radom, L., Pople, J. A.: J. Am. Chem. Soc. **92**, 4796 (1970)
23. Schaefer III, H. F.: The electronic structure of atoms and molecules. A survey of rigorous quantum mechanical results. Reading, Massachusetts: Addison-Wesley Publishing Corp. 1972
24. Clementi, E., Popkie, H.: J. Chem. Phys. **57**, 1077 (1972)
25. Huo, W. M.: J. Chem. Phys. **43**, 624 (1965)
26. Meyer, W.: J. Chem. Phys. **55**, 1017 (1973)
27. McLean, A. D., Yoshimine, M.: Intern. J. Quantum Chem. **1S**, 313 (1967)
28. Garrison, B. J., Schaefer III, H. F., Lester Jr., W. A.: J. Chem. Phys. **61**, 3039 (1974)
29. McLean, A. D., Yoshimine, M.: Tables of linear molecule wave functions. Research Report, IBM Corporation 1967
30. Tel, L. M., Wolfe, S., Csizmadia, I. G.: J. Chem. Phys. **59**, 4047 (1968)
31. Kolos, W., Roothan, C. C. J.: Rev. Mod. Phys. **32**, 214 (1960)
32. Karlström, G.: Theoret. Chim. Acta (Berl.) **44**, 165 (1977)
33. Löwdin, P. O.: Phys. Rev. **97**, 1474 (1955)
34. Löwdin, P. O., Shull, H.: Phys. Rev. **101**, 1730 (1956)
35. Hurley, A. C.: Rev. Mod. Phys. **32**, 400 (1960)

36. Davidson, E. R.: Advan. Quantum Chem. **6**, 235 (1972)
37. Kowalewski, J., Roos, B., Siegbahn, P., Vestin, R.: Chem. Phys. **9**, 29 (1975)
38. Herzberg, G.: Spectra of diatomic molecules. Princeton, N.J. U.S.A.: Van Nostrand 1950
39. Cade, P. E., Huo, W. M.: J. Chem. Phys. **47**, 649 (1967)