Thermochemical Data from *Ab Initio* Calculations. I. Estimation of SCF Energies for Augmented Basis Sets and Hartree–Fock Limit Energies

Dieter Cremer

Lehrstuhl für Theoretische Chemie, Universität zu Köln, D-5000 Köln 41, West Germany
Received 27 April 1981; accepted 15 July 1981

A procedure is outlined which allows an estimation of molecular energies both for a finite basis set including polarization functions and for the Hartree–Fock limit. It is shown that the orbital error of a given minimal basis is covered to a certain relatively constant percentage by an augmented basis set calculation. Thus an improvement factor Q_{av} can be determined by analyzing the corresponding results of small molecules where reasonable estimates of HF limit energies can be taken from the literature. For a combination of Pople's STO-3G and 6-31G* basis sets Q_{av} turns out to be 0.955.

I. INTRODUCTION

This is the first of two articles, the objective of which is to obtain estimates of characteristic molecular energies which in turn can be used for calculating thermochemical data of molecular systems. It is well-known that *ab initio* calculations lead to energy differences which are valid only for isolated molecules in the gas phase at 0 K and at a hypothetical state of rest.[1] In order to compare computed values with observed energies, zero-point vibrations and temperature effects have to be considered. Even if these corrections are available, it is generally not known to what extent computed energy differences approximate the true theoretical difference

$$\Delta E = E(\text{THEO})-E'(\text{THEO}) \approx E(\text{S}) - E'(\text{S}) \quad (1)$$

where $E(\text{S})$s have been obtained by solving the nonrelativistic electronic Schrödinger equation and $E(\text{THEO})$s contain in addition all relativistic effects (see Fig. 1).

If theoretical calculations are based on the Hartree–Fock (HF) approach, the molecular energy will depend on the constraints imposed on the wavefunctions, e.g., setting MOs $\psi_i^\alpha = \psi_i^\beta$ in the restricted HF (RHF) method.[2] For closed shell molecules one gets a self-consistent field energy $E(\text{SCF})$ which primarily reflects the size of the basis set X. The molecular energy $E(\text{SCF}/\text{X})$ can be improved by optimizing the molecular geometry $[\to E(\text{SCF}/\text{X};\text{G})]$ and in addition the orbital exponents $\zeta[\to E(\text{SCF}/\text{X};\text{G};\zeta)]$.

In finite basis set calculations $E(\text{SCF})$ will always stay above the HF limit energy $E(\text{HF})$ that would be obtained with an infinite number of basis functions.[1] The difference $\Delta E(\text{ORB}/\text{X}) = E(\text{SCF}/\text{X}) - E(\text{HF})$ computed at a reference geometry G with standard ζs is called the orbital or basis set error of X.[1] Although $\Delta E(\text{ORB}/\text{X})$ changes from molecule to molecule, it can be used to characterize X as will be shown below. Both $E(\text{HF})$ and $\Delta E(\text{ORB})$ are generally not known.

In their pioneering work on the nitrogen molecule Cade, Sales, and Wahl[3] have examined the convergence of $E(\text{SCF}/\text{X}) = E(\text{X})$ energies towards $E(\text{HF})$. For a stepwise improvement of the basis these authors found a continuous decrease of the total energy. By extrapolation a reasonable estimate of $E(\text{HF})$ for N_2 could be obtained.

Although such a procedure seems to be straightforward, it suffers from two major drawbacks. First of all, it implies extensive and systematic calculations with large augmented basis sets which are too costly for many-atom molecules. Secondly, the uncertainty of the HF limit energy thus obtained may be at best 2 mhartree. It results mainly from the fact that extrapolation of SCF energies $E(\text{X})$ to $E(\text{HF})$ has to be done in a completely arbitrary way. There seems to be no simple parameter which describes the quality of the basis set in a quantitative way. Even if a quantity is found that changes continuously toward an HF limiting value for a steady enlargement of the basis, it is in general not predictable whether these
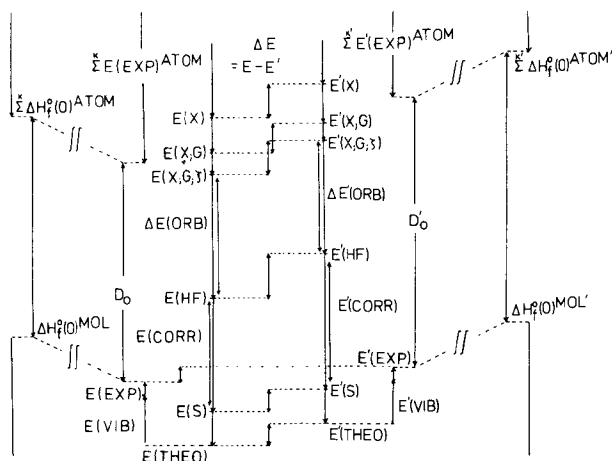
Figure 1. Dependence of calculated energy differences $\Delta E = E - E'$ on the level of theory. (For an explanation of energy terms, see text of this and following article.[25] Note that E and ΔH are given for different scales.)

changes are parallel to the lowering of the molecular energy.

In spite of these difficulties HF limit energies $E(\mathrm{HF})$ have been estimated for a number of small molecules by various authors.[4-7] Noteworthy are the compilations of $E(\mathrm{HF})$ values published by Hurley[5] and by Pople and Binkley.[6] These authors have taken available estimates or "near HF" results from the literature and derived therefrom the most likely HF energies.

As shown in Figure 1, the true $E(\mathrm{HF})$ value—even if known—is not sufficient to obtain accurate energy differences which can be compared directly with experimental ΔE values. This can only be done if the difference $E(\mathrm{S}) - E(\mathrm{HF})$, namely the correlation energy $E(\mathrm{CORR})$, is known for each of the molecular systems under consideration. Although there are several methods available which lead to correlation corrected energies,[8] a direct calculation of $E(\mathrm{CORR})$ is as impossible as is the determination of $\Delta E(\mathrm{ORB})$.

In order to get anywhere near theoretical energies that can be compared with experimental quantities, one has to establish a generally applicable procedure of estimating both $\Delta E(\mathrm{ORB})$ and $E(\mathrm{CORR})$. Such a procedure should fulfill two important requirements. First, it should be cheap and secondly, it should lead to molecular energies which are of value to the experimentalist.

The first requirement excludes large basis set calculations and optimization of nonlinear parameters. If we distinguish between five types of basis sets, namely minimal (MB), double zeta (DZ), heavy-atom augmented DZ (DZ + D), fully augmented DZ (DZ + D + P), and fully augmented

extended basis sets (EX + D + P), only MB and DZ calculations are economically feasible for large molecules. Repeated calculations with extended basis sets as carried out for the N_2 molecule[3] are far too costly. On the other hand, it is quite obvious that one should get, at least, energies from augmented DZ sets, in order to get anywhere near the HF limit energy $E(\mathrm{HF})$.

Hence, the first objective of this investigation is to find a way to estimate DZ + D or DZ + D + P energies at low cost and sufficient accuracy. Once these energies have been obtained, one has to develop a strategy of getting to HF limit energies from DZ + $D(+P)$ values. For this purpose it is expedient to utilize the $E(\mathrm{HF})$ energies found for small molecules. Thus, the second objective of this work is to analyze the orbital errors $\Delta E(\mathrm{ORB/X})$ of molecules for which reasonable estimates of $E(\mathrm{HF})$ are known. This analysis will lead to a general procedure of obtaining $E(\mathrm{HF})$ from $E(\mathrm{SCF/X})$ where X is a basis set of moderate size.

II. QUANTUM CHEMICAL METHODS

In this and the following article restricted HF(RHF) theory for closed-shell molecules[2] and unrestricted HF(UHF) theory for open-shell cases are used.[9] In order to avoid any ambiguity with regard to geometrical parameters, standard geometries are chosen throughout.[10] As suitable basis sets Pople's STO-3G (basis A), 4-31G (B), 6-31G* (C), and 6-31G** (D) basis sets have been chosen.[11-13] Although B, C, and D are not true examples of a DZ, DZ + D, and DZ + D + P basis set—the DZ quality holds only for the valence shell—they lead to a reasonable compromise between computational cost and accuracy. In addition, they have been used in many systematic studies on small- and medium-sized molecules.[13-15] Therefore, most of the molecular energies $E(\mathrm{SCF/X})$ used in this investigation can be taken from the literature at no additional cost.

In some cases it is necessary to employ a more elaborate basis set. For this purpose basis set E is used which corresponds to Dunning's [$4s3p1d/2s1p$] contraction of a ($9s5p1d/4s1p$) set of primitive Gaussian-type functions (GTFs).[16]

In order to obtain reasonable estimates of correlation corrections, second-order Rayleigh–Schrödinger–Møller–Plesset (RSMP) perturbation theory[17] has been employed. In this method the RHF (UHF) wave function and energy are taken as $\Phi^{(0)}$ and $E^{(0)}$. Then the RSMP total energy for a given basis set X is determined by

$$E(\text{RSMP}/X) = E^{(0)}(\text{SCF}/X) + E^{(2)}(X) \quad (2)$$

where $E^{(2)}$ represents the second-order correlation correction.

III. HF AND RSMP ENERGIES FOR AUGMENTED BASIS SETS

Generally, one would like to know to what extent small basis set SCF results change when both the method and the basis set are improved, e.g., to a level beyond the HF theory and to DZ + D + P basis set quality. If this could be done for a quantity like the molecular energy at low computational cost, then the analysis of SCF energies as well as the prediction of molecular quantities depending on energy differences could be facilitated considerably.

In Figure 2 the direction of quantum chemical improvements is symbolized by an arrow in a method/basis set diagram. For the molecular energy the diagonal arrow marks a line of decreasing energy values that leads directly to the Schrödinger energy $E(\text{S})$. Energies $E(\text{B})$ and $E(\text{RSMP}/\text{C})$ may be considered as being two points on this line. The question is how to obtain the value of $E(\text{RSMP}/\text{C})$ from easily available $E(\text{B})$ energies without any additional quantum chemical calculations.

As illustrated in Figure 2 this can be done by first estimating the energy lowering due to a basis set enlargement from B to C, thus obtaining an approximated value of $E(\text{C})$. In the next step $E(\text{RSMP}/\text{C})$ is determined by finding an appropriate estimate of the correlation energy $E^{(2)}(\text{C})$. The alternative way, i.e., estimation of $E^{(2)}(\text{B})$, and subsequently the effect of polarization functions at the RSMP level, should lead to the same results. In other words, if the effect of adding polarization functions to basis B can be estimated at the HF level, this should also be possible at the RSMP level (step 1). In addition, an empirical evaluation of correlation energies $E^{(2)}$—if possible—should work for both basis sets B and C (step 2). In the following we concentrate first on accomplishing step 1. The empirical implementation of step 2 is investigated in Part II.

When going from B to C two extensions of the 4-31G basis are introduced.[13] First, the inner-shell 1s orbital is described by six rather than four contracted GTFs. Secondly, six second-order GTFs corresponding to five 3d and a single 3s GTF are added. The effects of these basis set enlargements can be separated by comparing energies $E(\text{C})$ with both $E(\text{B})$ values and energies obtained with a
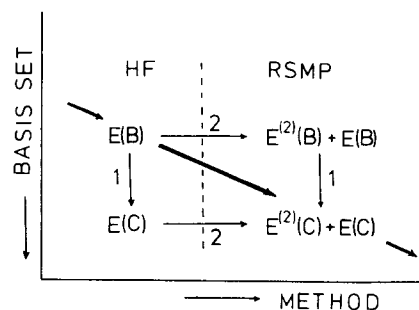

Figure 2. Dependence of *ab initio* energies on method and basis set. (Arrows indicate energy lowering due to method or basis set improvements.)

6-31G basis.[12] This has been done by Hariharan and Pople[13] for small molecules. From computed $E(\text{C})$ and $E(6\text{-}31\text{G})$ values these authors compiled a list of energy lowerings per heavy atom which shows that energy lowerings depend on the nature and the chemical environment of the atom in question. These d lowerings, supplemented by some additional values, are listed in the third column of Table I.

From known 6-31G and 4-31G results[13,15] the 1s part of the energy difference $E(\text{B}) - E(\text{C})$ can be determined easily. It is given in the second column of Table I.

In exactly the same way the 1s and d parts can be evaluated for the difference $E^{(2)}(\text{B}) - E^{(2)}(\text{C})$ at the RSMP level. We have calculated for the molecules of Table I three sets of energies $E^{(2)}$ using the 4-31G, 6-31G, and 6-31G* basis sets (see Appendix). The energy lowerings per atom derived therefrom are also listed in Table I.

From an inspection of Table I the following points become obvious.

(1) For each of the four heavy atoms C, N, O, and F a 1s increment $\Delta E(\text{SCF})^{1s}$ can be defined, which is almost independent of the chemical environment of the heavy atom. The value of the 1s increment increases with the electronegativity of the heavy atom and that always by ca. 11 kcal/mol per atomic number starting at C with 26 and ending at F with 59 kcal/mol.

(2) The d-part $\Delta E(\text{SCF})^{3d}$ varies from 8 to 36 kcal/mol indicating its dependence on nature and bonding state of the heavy atom. Polarity and multiplicity of the heavy atom bond(s), the number of XH bonds, and their geometrical arrangement influence the magnitude of $\Delta E(\text{SCF})^{3d}$.

(3) All 1s increments $\Delta E^{(2)1s}$ are negative. By employing the 6-31G basis the second-order correlation corrections decrease by 0.5 kcal/mol per heavy atom where this value does not depend on

Table I. Lowering of the molecular energy due to an enlargement of basis B to C. The importance of d-type polarization functions in the heavy atom basis. (Energies ΔE are given in kcal/mol per heavy atom.)[a]

| Atom | Molecule | Lowering of $E(\text{SCF/B})$[b] | | Lowering of $E^{(2)}(\text{B})$ | | Total lowering |
		ΔE^{1s}	ΔE^{3d}	ΔE^{1s}	ΔE^{3d}	$\Delta E(\text{RSMP})^t$
$\rangle$C$\langle$	CH_4	25.6	9.2	−0.6	26.4	60.6
	C_2H_6	25.6	9.8	−0.5	26.6	61.5
$\rangle$C$=$	C_2H_4	26.0	8.5	−0.7	25.4	59.1
$—$C$=$	C_2H_2	25.5	7.8	−0.7	23.1	55.7
$\searrow$N$\nearrow$	NH_3,C_{3v}	36.8	12.9	−0.4	35.5	84.8
	D_{3h}	37.0	5.9	−0.3	36.3	78.9
	N_2H_4	36.6	15.3	−0.4	35.1	86.6
	CH_3NH_2	36.7	16.2	−0.6	35.3	87.6
$\nearrow$N$=$	N_2H_2	36.6	19.9	−0.7	31.5	87.3
	CH_2NH	36.6	20.2	−0.6	30.8	87.1
N$\equiv$	N_2	35.9	23.4	−0.7	24.3	82.9
	HCN	35.2	20.8	−0.6	25.7	81.1
$\underset{/\backslash}{\overset{O}{}}$	H_2O	48.1	15.6	−0.4	37.5	100.8
	H_2O_2	47.1	14.5	−0.4	36.4	97.7
	CH_3OH	47.6	20.0	−0.7	36.9	103.8
O$=$	O_2	47.5	22.1	−1.1	30.8	99.2
	CH_2O	46.8	26.5	−0.7	31.2	103.8
O$\equiv$	CO	46.6	36.1	−0.4	25.6	107.9
F$—$	HF	60.3	12.2	−0.5	32.8	104.8
	F_2	58.9	8.4	−0.6	31.0	97.7
	CH_3F	59.5	17.1	−0.8	32.2	107.9

[a] Increments ΔE have been calculated from the energies listed in the Appendix and in refs. 12 and 13. For molecules of type H_pCYH_q the average ΔE_{CY} value is used in the calculation of ΔE_Y.
[b] Reference 13.

either electronegativity or bonding state of the atom in question.

(4) The d-part $\Delta E^{(2)3d}$ is with one exception (O in CO) always greater than the corresponding $\Delta E(\text{SCF})^{3d}$ value. It shows similar dependence on the chemical environment as the corresponding $\Delta E(\text{HF})^{3d}$ value.

Observations (1) and (3) verify that inner-shell electrons do not take part in bonding. The negative $\Delta E^{(2)1s}$ reveal a deficiency of the larger basis set. The common understanding that a basis developed for HF calculations is also a good basis for correlation-corrected calculations is certainly not true.[18] The RSMP calculation, of course, leads to a different electron distribution. For example, the electron density close to the nucleus decreases significantly as compared with the HF density. This requires a contraction of the basis functions describing the region close to the nuclei. The HF basis C does not fulfill this requirement thus leading to a deterioration of the absolute value of $E^{(2)}$. In view of the consistency of $\Delta E^{(2)1s}$ this seems to have no consequence for the bond description.

Observations (2) and (4) underline the necessity of polarization functions especially for the description of anisotropic electron distributions.

Their importance increases in correlation-corrected calculations where in the case of basis C the high $\Delta E^{(2)3d}$ values may indicate that the defi-

Table II. Lowering of the molecular energy due to an enlargement of basis C to D. The importance of p-type polarization functions in the H basis. (Energies are given in kcal/mol per Y—H bond.)[a]

Bond	Molecule	Lowering of $E(\text{SCF}/C)$	Lowering of $E^{(2)}(C)$	Total lowering $\Delta E(\text{RSMP})^t$
H—H	H_2	1.4	2.8	4.2
C—H	CH_4	1.0	4.1	5.1
	C_2H_6	1.0	4.2	5.2
	C_2H_4	1.1	4.1	5.2
	C_2H_2	1.3	3.5	4.8
	NCH	1.2	3.4	4.7
	OCH_2	1.1	4.0	5.1
	FCH_3	1.1	4.1	5.2
N—H	NH_3,C_{3v}	2.4	3.8	6.2
	D_{3h}	2.6	3.5	6.2
	N_2H_4	2.2	3.9	6.1
	N_2H_2	2.0	4.0	6.0
	CH_3NH_2	2.3	3.9	6.2
	CH_2NH	2.3	3.9	6.2
O—H	H_2O	4.0	3.3	7.3
	H_2O_2	4.0	3.2	7.2
	CH_3OH	4.0	3.4	7.4
F—H	FH	5.3	2.7	8.0

[a] Increments ΔE have been calculated from the energies listed in the Appendix and in ref. 13.

Table III. Estimation of basis C energies from basis B energies. (Molecular energies in hartree, errors in kcal/mol.)[a]

Formula	Skeleton	Molecule, conformation	$E(\text{SCF/C})$ estimated	Error[b,c]	$E(\text{RSMP/C})$ estimated	Error[b,d]
C_3H_8	C—C—C	Propane	−118.2599	1.3	−118.6696	2.1
C_3H_6	C—C=C	Propene	−117.0679	0.5	−117.4653	1.2
C_3H_4	C—C≡C	Propyne	−115.8627	0.6	−116.2513	2.5
	C=C=C	Allene	−115.8625	−1.4	−116.2488	−0.5
C_2NH_7	C—C—N	Ethylamine, gauche	−134.2452	0.5	−134.6840	1.3
C_2NH_7	C—N—C	Dimethylamine	−134.2317	2.8	−134.6707	3.2
C_2NH_5	C=C—N	Vinylamine	−133.0644	−6.4	−133.4903	−3.4
	C—C=N	Acetaldimene, HCNH cis	−133.0662	0.3	−133.4984	0.9
	C=N—C	N-Methylformaldimine	−133.0455	2.5	−133.4820	2.6
C_2NH_3	C—C≡N	Acetonitrile	−131.9254	0.2	−132.3467	2.4
	C≡C—N	Ethynylamine	−131.8514	−6.0	−132.2665	−2.4
	C=C=N	Ketenimine	−131.8566	−1.3	−132.2835	−1.2
C_2OH_6	C—C—O	Ethanol, CCOH trans	−154.0737	0.4	−154.5259	1.1
	C—O—C	Dimethyl ether	−154.0553	3.9	−154.5082	3.7
C_2OH_4	C—C=O	Acetaldehyde, HCCO cis	−152.9125	0.2	−153.3550	1.3
	C=C—O	Vinyl alcohol, COOH cis	−152.8819	0.8	−153.3238	2.2
C_2OH_2	C=C=O	Ketene	−151.7213	1.7	−152.1572	1.3
	C≡C—O	Ethynol	−151.6552	1.6	−152.0932	2.5
CO_2H_4	O—C—O	Methane diol, HOCO gauche	−189.8964	−0.5	−190.3848	5.6
	C—O—O	Methyl hydroperoxide, COOH trans	−189.7791	3.0	−190.2838	3.5
CO_2H_2	O—C=O	Formic acid, HOCO cis	−188.7501	2.5	−189.2423	2.4
CO_2	O=C=O	Carbon dioxide	−187.6166	10.2	−188.1101	4.3
C_2FH_5	C—C—F	Fluoroethane	−178.0755	0.4	−178.5166	1.3
C_2FH_3	C=C—F	Fluoroethylene	−176.8780	−0.1	−177.3110	0.9
C_2FH	C≡C—F	Fluoroacetylene	−175.6377	2.0	−176.0690	2.1
CF_2H_2	F—C—F	Difluoromethane	−237.8919	1.8	−238.3703	1.7

[a] Increments ΔE of compounds H_pXYH_q [X = C; Y = C,N,O,F, and X = Y =O (for CH_3OOH)] have been used.
[b] Errors E(estimated) − E(computed).
[c] Average error $\bar{f}$ = 2.0 kcal/mol; standard deviation σ = 3.1 kcal/mol.
[d] $\bar{f}$ = 2.2 kcal/mol; σ = 2.6 kcal/mol.

ciency of the inner-shell functions are partially compensated by the $3d$ GTFs.

A similar analysis has been carried out for the augmentation of basis C with hydrogen polarization functions thus leading to basis D. The corresponding energy improvements are conveniently given per XH bond.[13] In Table II they are listed for HF and RSMP calculations where the values $\Delta E(\text{SCF})^{2p}$ have been taken from Hariharan and Pople.[13] While the SCF values increase from 1 to 5 kcal/mol in going from the CH to the FH bond, most of the $\Delta E^{(2)2p}$ values are close to 4 kcal/mol. The slight decrease of the latter values when increasing the electronegativity of X in XH may also result from the fact that basis D has not been reoptimized for RSMP calculations. Obviously the bonding status of X is only of minor consequence for the $2p(H)$ increments.

From Tables I and II suitable energy increments can be derived in order to estimate $E(C)$ [$E(D)$] energies from the corresponding basis B (C) values. From the tabulated data it is obvious that such a procedure is the more accurate the larger the

starting basis is. Basis B can be considered as one of the smallest possible starting basis sets.

For a set of three-heavy-atom molecules $E(\text{SCF/C})$ and $E(\text{RSMP/C})$ values are estimated using increments ΔE^{1s} and ΔE^{3d} taken from two-heavy-atom molecules with similar heavy atom bonds. Thus the value $E(\text{RSMP/C})$ for formic acid is estimated from the $\Delta E(\text{RSMP})^t$ increments of C ($H_2C=CH_2$) and O (CH_2O and CH_3OH):

$$E(\text{RSMP/C}) \approx E(\text{RSMP/B}) - [\Delta E(\text{RSMP})^t_{C=}$$
$$+ \Delta E(\text{RSMP})^t_{O=} + E(\text{RSMP})^t_{O-}]/627.52$$

$$\approx -188.8173 - (59.1 + 103.8 + 103.8)/627.52$$

$$\approx -189.2423 \text{ hartree}$$

The estimated energies thus obtained are listed in Table III. As can be seen from the error analysis [standard deviations σ are 3.1 (SCF) and 2.3 kcal/mol (RSMP)] approximation of the calculated basis C energies by estimated values is satisfying. The latter are mostly somewhat too low indicating that nonbonded interactions, not accounted for in the increments, contribute to the total energy. Since

Table IV. Orbital errors of basis sets A, B, C, and D calculated for small molecules from HF limit energies.[a]

Molecule	Orbital error of A		Improvement by[b]			$E(\text{HF})^c$
	absolute	relative	B	C	D	(hartree)
H_2	0.01684	1.48	59.4	59.4	86.3	−1.1336
CH_4	0.492	1.22	83.9	95.1	96.5	−40.219
C_2H_2	1.007	1.31	85.2	95.8	96.2	−76.860
C_2H_4	1.009	1.29	84.2	95.1	95.8	−78.080
C_2H_6	0.965	1.22	83.9	95.6	96.6	−79.270
NH_3	0.774	1.37	84.3	94.5	96.0	−56.226
N_2	1.501	1.38	83.8	96.3	96.3	−108.997
N_2H_2	1.508	1.37	83.9	95.9	96.3	−110.053[e]
N_2H_4	1.500	1.35	83.9	94.9	95.9	−111.243[e]
OH_2	1.109	1.46	85.4	94.6	95.7	−76.070
$O_2{}^d$	2.036	1.36	86.4	97.3	97.3	−149.670
O_2H_2	2.117	1.40	85.5	94.8	95.4	−150.860[e]
FH	1.500	1.50	87.8	95.5	96.0	−100.071
F_2	2.820	1.42	88.4	96.0	96.0	−198.785
CH_3NH_2	1.244	1.31	83.4	94.7	95.7	−95.274[e]
CH_2NH	1.271	1.35	83.4	94.9	95.5	−94.086[e]
HCN	1.245	1.34	84.8	96.2	96.4	−92.920
CH_3OH	1.561	1.36	84.9	95.4	96.1	−115.106
CH_2O	1.580	1.39	84.8	95.7	95.9	−113.932
CO	1.566	1.39	84.8	96.6	96.6	−112.791
CH_3F	1.942	1.40	86.9	96.1	96.4	−139.110[e]

[a] Absolute orbital error in hartree, relative orbital error in percent. For the SCF energies see ref. 14 [$E(A)$], ref. 15 [$E(B)$], and ref. 13 [$E(C)$ and $E(D)$].

[b] Improvement of $\Delta E(\text{ORB/A})$ in percent.

[c] HF limit energies from refs. 5–7.

[d] From UHF energies of lowest triplet state.

[e] This work, uncertainty: 0.002 hartree.

the number of these interactions increases with the size of the molecule, deviations between estimated and computed energies will also increase for larger molecules. On the other hand, the set of increments can be improved by using available SCF and RSMP energies of three-, four-, and five-heavy-atom molecules.

It can be concluded that step 1 of diagram 2 can be done in a successful manner. Based on an analysis of energy lowerings due to basis set enlargement of a standard basis, helpful estimates of the energies of larger molecules can be made, provided that these molecules already have been calculated with the standard basis set.

IV. THE ORBITAL ERROR

In the last column of Table IV the HF limit energies of 21 small molecules are collected. Most of these data have been taken from Hurley's work,[5] and some of them slightly corrected by more recent estimates.[6,7] HF limit energies of N_2H_2, N_2H_4, H_2O_2, CH_3NH_2, CH_2NH, and CH_3F are based on our own estimates. They have been obtained as is briefly outlined for the H_2O_2 molecule (see Fig. 3).

Combing our own results on H_2O_2 (ref. 19) with

those published earlier[20-23] we extrapolate to an HF limit energy which is slightly lower than that of Hurley[5] but considerably higher than the estimate of Guidotti, Lamanna, Maestro, and Moccia (GLMM)[23] (see Fig. 3). At first sight the chosen energies seem to gather closely to a smooth curve $E(N)$ (N = number of basis functions) which leads directly to the GLMM estimate of $E(\text{HF})$. Closer examination, however, reveals that there are two
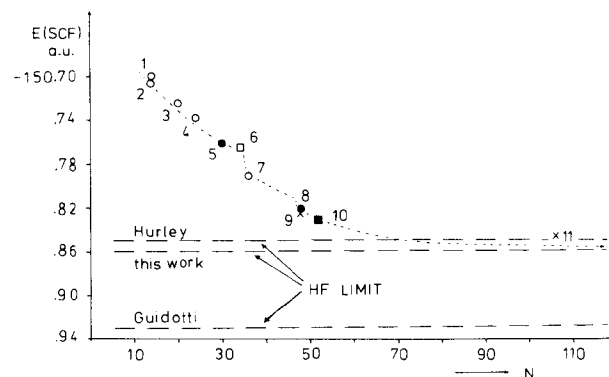

Figure 3. Determination of $E(\text{HF})$ for H_2O_2. [Points 1–4, 7: MB, MB + P, DZ, DZ + D calculations (O), ref. 20; 5, 8: DZ, DZ + D (●), ref. 22; 6: triple zeta (TZ) (□), ref. 21; 10: TZ + D + P (■), ref. 23; 9, 11: DZ + D + P, extended + $2D$ + $2P$ (×), ref. 19.]

drops indicating augmentation of the O basis by $3d$ and the H basis by $2p$ GTFs. Actually only the energy points 9, 10, and 11 can be used for the extrapolation procedure, where special consideration has to be given to the fact that points 9 and 11 have been calculated with a basis of GTFs (ref. 19) and point 10 with a basis of STFs.[23] In addition, the calculation of point 11 was done with fixed rather than optimized exponents of the polarization functions, which may cause a change of 5–8 mhartree in the molecular energy. In view of these considerations an $E(\text{HF})$ of $-150.860(2)$ hartree is predicted for H_2O_2. This value coincides with an independent estimate obtained by Ermler and Kern.[24]

Absolute and relative orbital errors of basis sets A–D are listed in Table IV. They have been derived by comparing the HF energies of the last column with published $E(\text{A})$,[14] $E(\text{B})$,[15] $E(\text{C})$, and $E(\text{D})$ (ref. 13) molecular energies. The absolute values $\Delta E(\text{ORB/A})$ illustrate the well-known fact that MB calculations lead to orbital errors of the order of 1–2 hartree even for small molecules. In most quantum chemical studies, however, the relative rather than the absolute orbital error matters, i.e., the consistency of the values $\Delta E(\text{ORB/X})$ is of importance.* If, for example, the relative orbital error of the MB basis A remains constant, a basis A calculation would suffice for most purposes.

As can be seen from Table IV, basis A does not provide consistent molecular energies. The relative orbital error is fairly low for hydrocarbons, but increases with the electronegativity of the heavy atoms(s). In addition, it depends on the number of H atoms per heavy atom (compare NH_3 and N_2H_4, etc.), and on the multiplicity of the heavy atom bond (compare C_2H_6, C_2H_4, C_2H_2). Since all these features influence the anisotropy of the electron distribution around the nuclei, it is safe to say that with an increasing anisotropy the relative orbital error of A becomes larger.

Bases B, C, and D compensate for some of the deficiencies of A. This can be seen from the corresponding data of Table IV. They give the improvement brought about by basis B, C, or D in percent of $\Delta E(\text{ORB/A})$.

Excluding the two exceptional cases H_2 and O_2[†] two observations can be made. With basis B, ca. 85% of the orbital error of A is covered. Application of C and D leads to an improvement of this value to ca. 95.5 and 96%, respectively. Secondly, the improvements become more consistent in going from B to D. Thus the range of relative orbital errors narrows from 3.5% (basis B) to 2.1 (C) and 1.2% (D).

V. ESTIMATION OF HF LIMIT ENERGIES

In view of this consistent improvement of basis A energies, it should be possible to obtain crude estimates of $E(\text{HF})$ for larger molecules by just two calculations, namely, one with basis A and one with basis C. The energy difference $E(\text{A}) - E(\text{C})$ leads immediately to an estimate of $E(\text{HF})$:

$$E(\text{HF}) \approx E(\text{A}) - [\text{E(A)} - E(\text{C})]/Q_{\text{av}} \qquad (2)$$

where $Q_{\text{av}} = 0.955$ is the improvement factor found for basis C.

Such a procedure can be improved by:

(1) increasing the set of HF limit energies given in Table IV by reliable values for other molecules;

(2) calculating for this set the orbital errors of a larger basis set than A, e.g., B;

(3) employing augmented basis sets E, F, ..., X in order to evaluate from $E(\text{X})$ an improvement factor $Q(\text{X}) = [\text{E(B)} - E(\text{X})]/\Delta E(\text{ORB/B})$ which is closer to 1;

(4) applying an average factor $Q_{\text{av}}(\text{X})$ when estimating the unknown HF limit energy of a large molecule from its energies $E(\text{B})$ and $E(\text{X})$.

This procedure is used to get estimates of $E(\text{HF})$ for 26 three-heavy-atom compounds. These data are presented in Part II.[25] In addition, we have estimated HF limit energies in several practical problems,[26] one of which will be discussed below. Before doing so, it is appropriate to compare the present method of estimating $E(\text{HF})$ with a similar procedure suggested by Ermler and Kern (EK).[24]

---

* A basis X leading to a small orbital error is certainly a valid basis for describing chemical bonding. On the other hand, a large orbital error does not necessarily indicate that chemical bonding is insufficiently described by the corresponding basis. Neglect of inner-shell functions increases $\Delta E(\text{ORB})$ but is of minor consequence for the description of the bond region.

† In the case of H_2, the relative orbital errors indicate that the H basis sets should be improved. Since, however, the absolute orbital errors are rather small, the severe deficiencies of the H sets hardly show up in the orbital errors of larger molecules. As for O_2, one has to consider that its triplet ground-state energy has been calculated with the UHF method which leads to a lower $E(\text{HF})$ value than the RHF method.

Table V. HF limit energies of molecules formed or destroyed in the ozonolysis of olefins.

Formula	Name	E(HF) (hartree)		
		Method I[a]	Method II[b]	EK Method[c]
O_3	Ozone	−224.401	−224.391	−224.390
C_2H_4	Ethylene	−78.077	−78.075	−78.076
$C_2O_3H_4$	1,2,3-Trioxolane	−302.602	−302.588	−302.590
CO_2H_2	Carbonyl oxide	−188.658	−188.649	−188.648
CO_2H_2	Formic acid	−188.882	−188.874	−188.872
COH_2	Formaldehyde	−113.937	−113.932	−113.931
$C_2O_3H_4$	1,2,4-Trioxolane	−302.699	−302.680	−302.681
OH_2	Water	−76.060	−76.071	−76.070
O_2H_2	Hydrogen peroxide	−150.859	−150.866	−150.865
O_3H_2	Hydrogen trioxide	−225.673	−225.678	−225.677

[a] Calculated from E(A) and E(C) energies with Q_{av} = 0.955.[28]
[b] Calculated from E(A) and E(E) energies with Q_{av} = 0.980.[28]
[c] Determined with the Ermler–Kern procedure[24]; f_E = 1.00027.

VI. COMPARISON WITH THE ERMLER-KERN PROCEDURE[24]

The EK procedure for estimating HF limit energies is based on the assumption that a basis set X of DZ + D + P quality is stable in the sense that SCF energies do change consistently upon enlargement of X. If for a suitable reference molecule E(HF) is known with sufficient accuracy, then a proportionality factor f can be defined from the computed energy E(X):

$$f(X) = E(\text{HF})/E(X) \qquad (3)$$

Once f(X) has been determined, the HF limit energy of any other molecule is estimated by multiplying the corresponding E(X) value by f(X). EK demonstrated that with basis E (ref. 16) and f(E) = 1.00027, reasonable estimates of E(HF) are obtained.[24] These values agree within 2–3 mhartree with the corresponding limit energies given in Table IV (see also Table V).

Although the procedure presented in this work and the one suggested by EK have been developed in the same spirit, there are three major differences between them.

First of all, the factor f is based on just one molecule while Q is derived as an average using HF and SCF energies of a series of small molecules. This feature, however, can be incorporated easily into the EK procedure.[27]

Secondly, the EK procedure requires a fairly large basis in order to secure stability of the computed energy values. The Q method, on the other hand, can be carried out with relatively inexpensive basis sets since it uses the ratio of basis set errors rather than the ratio of absolute energies. For example, RHF/C calculations are about four times less costly than RHF/E calculations. The

necessity of doing two SCF calculations for the Q method does not mean a disadvantage since the computational cost for carrying out RHF/A (RHF/MB) calculations is negligible.

The third difference between EK and Q procedure becomes apparent when assuming for a moment that the same basis set X is used in both procedures. The data of Table IV indicate that with an increase of the anisotropy of the electron distribution the relative orbital error x of basis X becomes larger. Inconsistencies of x will also result when using for the same number of electrons different numbers of basis functions or vice versa. In Figure 4 the orbital errors of two molecules 1 and 2 are depicted as a function of the anisotropy of the distribution ρ and the size of the basis set (Y > X). Analyzing the data of Table IV, it can be shown that the basic requirement of the Q method, namely,
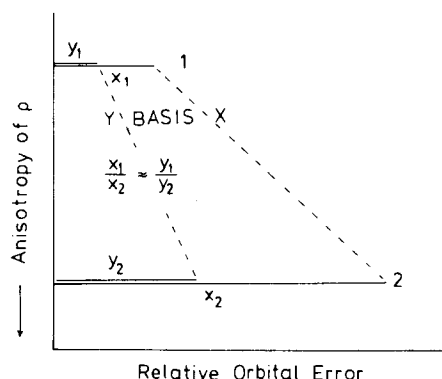
$$\frac{x_1}{x_2} \approx \frac{y_1}{y_2} \qquad (4)$$

Figure 4. Dependence of the orbital error $x(y)$ on the anisotropy of the electron distribution ρ and the size of the basis X (Y).

is generally fulfilled. From eq. (4) it follows that

$$\frac{x_1}{x_2} \approx \frac{x_1 - y_1}{x_2 - y_2} \tag{5}$$

and

$$Q(Y) = \frac{x_1 - y_1}{x_1} \approx \frac{x_2 - y_2}{x_2} \tag{6}$$

With the definition of the relative orbital error y (for a not to small basis Y)

$$y = \frac{E(Y) - E(HF)}{|E(HF)|} \approx \frac{E(Y) - E(HF)}{|E(Y)|} \tag{7}$$

the factor $f(Y)$ is given by

$$f(Y) = \frac{E(HF)}{E(Y)} \approx 1 + y \tag{8}$$

Since $y_2 > y_1$, $f(Y)_2$ and $f(Y)_1$ will not be equal even for a fairly large basis Y. If, however, weighted differences $f(Y) - f(X)$ rather than absolute values f are considered, then eq. (9),

$$\frac{f(X)_i - f(Y)_i}{x_i} = Q(Y) \tag{9}$$

holds for all molecules i.

When using an average factor $f_{av} = (f_1 + f_2)/2$, the value $f_{av}E(Y)$ differs by $0.5E(Y)(y_2 - y_1)$ from $E(HF)$ obtained with Q. For example, the HF limit energy of the C_2H_6 molecule computed with the average factor $f(C) = 1.00063$ is 8 mhartree lower than the value given in Table IV.

Therefore, it is concluded that for large basis sets both the Q and the EK procedures lead to the same HF limit energies within an uncertainty of 2–3 mhartree. Since the Q method compensates for inconsistencies of $E(X)$ values, it is favorably used in those cases where calculations with stable basis sets cannot be carried out because of high computational cost.

VII. APPLICATION

In order to demonstrate the usefulness of the procedures developed in Secs. III and V we refer to a recent investigation of the reaction,

$$O_3 + C_2H_4 \xrightarrow{a} \overline{CH_2OOOCH_2} \xrightarrow{b} CH_2O$$

$$+ CH_2OO \xrightarrow{c} \overline{OCH_2OCH_2O}$$
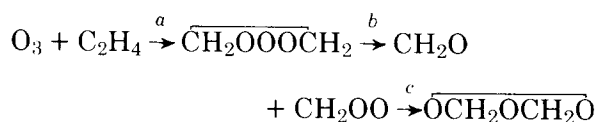
which is considered as the dominant path of the solution-phase ozonolysis.[28]

The three reaction energies were first calculated at the RHF/B level. As these values differed con-

siderably from thermochemically estimated reaction enthalpies, it was asked whether inclusion of polarization functions into the basis would alter RHF/B results significantly. While the calculation of O_3, C_2H_4, CH_2O, and CH_2O_2 was carried out easily at the RHF/C level, calculation of the two trioxolanes with basis C (83 basis functions) seemed to be problematic since it involved the evaluation of about 5 million two-electron integrals. Thus, it was first checked whether this computational effort was necessary. This was done by estimating the energy difference $E(B) - E(C)$ from the data of Table IV. $E(C)$ energies of -302.424 and -302.503 hartree were predicted for the 1,2,3- and the 1,2,4-trioxolane. Utilizing these values together with the computed basis C energies, changes of basis B reaction energies of $+16$ (a), -10 (b), and $+9$ kcal/mol (c) were predicted. These changes were considered to be significant, thus justifying explicit basis C calculations of the two trioxolanes. This lead to RHF/C reaction energies which differed from the corresponding basis B values by 17.3, -10.5, and 11.2 kcal/mol, in reasonable agreement with the estimated changes. Later the same procedure was repeated at the RSMP/B level.

In Table V the $E(HF)$ values of molecules possibly formed or destroyed in the course of the solution-phase ozonolysis of olefins are listed. They have been obtained in three different ways. First, basis A and basis C energies were calculated and the factor $Q_{av}(C) = 0.955$ was applied. Then, the procedure was repeated with a larger basis set E for which a factor $Q_{av}(E) = 0.980$ had been determined. The $E(HF)$ values estimated with this factor (Table V) deviate on the average by 6 kcal/mol, mostly turning out somewhat less negative than those obtained with $Q_{av}(C)$.

Some of the molecules used with their SCF energies for standard geometries to determine $Q_{av}(C)$ and $Q_{av}(E)$ are also included in Table V. Since in ref. 30 extensive geometry optimization had to be carried out, $E(HF)$ values of these molecules indicate to what extent the lower $E(X,G)$ energies lead to a change of HF limit energies. If only for one basis geometry optimization has been carried out, $E(HF)$ will be either lower (optimization with C) or higher (optimization with basis A) than the value obtained at standard geometries. If for both basis sets optimal energies are used, the change of $E(HF)$ will not be predictable (see Table V). In any case, changes of $E(HF)$ are rather small as can be seen from inspection of Table V.

HF limit energies computed with $Q_{av}(E)$ coin-

Table VI. RSMP energies (hartree) of one- and two-heavy-atom molecules at standard geometries.

Molecule	Basis A	Basis B 4-31G	Basis B' 6-31G	Basis C	Basis D
H_2	−1.12990	−1.14414	−1.4414	−1.14414	−1.15763
CH_4	−39.78326	−40.24045	−40.28033	−40.33704	−40.36977
$HC\equiv CH$	−76.00324	−76.89801	−76.97726	−77.07566	−77.09090
$H_2C=CH_2$	−77.19445	−78.10521	−78.18554	−78.29362	−78.32669
H_3C-CH_3	−78.41216	−79.30714	−79.38711	−79.50314	−79.55285
NH_3	−55.49872	−56.22163	−56.27974	−56.35679	−56.38630
$N\equiv N$	−107.65143	−108.99503	−109.10706	−109.25938	−109.25938
$HN=NH$	−108.66484	−110.03853	−110.15314	−110.31695	−110.33623
H_2N-NH_2	−109.83185	−111.22701	−111.34230	−111.50289	−111.54192
OH_2	−74.99577	−76.03783	−76.11384	−76.19840	−76.22168
$O=O^a$		−149.63650	−149.78573	−149.95277	−149.95277
$HO-OH$	−148.81939	−150.81022	−150.95910	−151.12156	−151.14431
FH	−98.58858	−100.01686	−100.11228	−100.18396	−100.19671
$F-F$	−196.01857	−198.72735	−198.91325	−199.03882	−199.03882
H_3C-NH_2	−94.12731	−95.27588	−95.37322	−95.51280	−95.55749
$H_2C=NH$	−92.94079	−94.08624	−94.18384	−94.31925	−94.34570
$HC\equiv N$	−91.83052	−92.94820	−93.04297	−93.16635	−93.17380
H_3C-OH	−113.63245	−115.08920	−115.20394	−115.35203	−115.38863
$H_2C=O$	−112.46884	−113.91415	−114.02778	−114.17375	−114.19001
$C\equiv O$	−111.35374	−112.76662	−112.87977	−113.02728	−113.02728
H_3C-F	−137.23540	−139.07270	−139.20607	−139.34204	−139.36690

[a] UHF/RSMP energy of the lowest triplet state.

cide within 2 mhartree with estimates of $E(HF)$ which have been obtained according to the EK procedure.[24] This confirms the discussion given in Sec. VI.

Appendix

RSMP energies have been calculated with the four basis sets considered in this work. They are depicted in Table VI for one- and two-heavy-atom molecules. The corresponding second-order correlation corrections $E^{(2)}$ are listed in Table VII. Energies obtained with the 6-31G basis (B') are added.

Note added in proof. Recently, the applicability of the EK procedure has been investigated by George, Trachtman, and Bock.[29] These authors demonstrated that the factor f depends on the choice of the reference molecule and, therefore, should only be used for molecules with structural features similar to those of the reference mole-

Table VII. Second-order RSMP correlation corrections $E^{(2)}$ (hartree) of one- and two-heavy-atom molecules at standard geometries.

Molecule	Basis A	Basis B 4-31G	Basis B' 6-31G	Basis C	Basis D
H_2	−0.01314	−0.01738	−0.01738	−0.01738	−0.02634
CH_4	−0.05657	−0.10090	−0.09994	−0.14198	−0.16818
$HC\equiv CH$	−0.14985	−0.18696	−0.18465	−0.25833	−0.26953
$H_2C=CH_2$	−0.12324	−0.18470	−0.18236	−0.26325	−0.28915
H_3C-CH_3	−0.10720	−0.19230	−0.19060	−0.27539	−0.31561
NH_3	−0.04617	−0.11713	−0.11654	−0.17304	−0.19131
$N\equiv N$	−0.15493	−0.24158	−0.23944	−0.31704	−0.31704
$HN=NH$	−0.11988	−0.22751	−0.22521	−0.32572	−0.33855
H_2N-NH_2	−0.08873	−0.22509	−0.22379	−0.33557	−0.36045
OH_2	−0.03505	−0.12941	−0.12876	−0.18853	−0.19913
$O=O^a$		−0.24382	−0.24027	−0.33837	−0.33837
$HO-OH$	−0.07799	−0.25739	−0.25597	−0.37208	−0.38216
FH	−0.01748	−0.12958	−0.12885	−0.18115	−0.18549
$F-F$	−0.05332	−0.26897	−0.26720	−0.36592	−0.36592
H_3C-NH_2	−0.09726	−0.20782	−0.20605	−0.30461	−0.33729
$H_2C=NH$	−0.12620	−0.21063	−0.20857	−0.29820	−0.31746
$HC\equiv N$	−0.15556	−0.21739	−0.21534	−0.29318	−0.29865
H_3C-OH	−0.08695	−0.21900	−0.21712	−0.31817	−0.34362
$H_2C=O$	−0.11631	−0.22219	−0.21989	−0.31005	−0.32291
$C\equiv O$	−0.12899	−0.21427	−0.21255	−0.29011	−0.29011
H_3C-F	−0.06752	−0.21622	−0.21407	−0.30756	−0.32729

[a] UHF/$E^{(2)}$ energy of lowest triplet state.

cule. This observation is in line with the discussion presented in Sec. VI.

References

1. See, e.g., A. C. Hurley, *Introduction to the Electron Theory of Small Molecules*, Academic, New York, 1976.

2. (a) C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951); (b) G. G. Hall, *Proc. R. Soc. London, Ser. A*, **A205**, 541 (1951).

3. P. E. Cade, K. D. Sales, and A. C. Wahl, *J. Chem. Phys.*, **44**, 1973, (1966).

4. See, e.g., C. Hollister and O. Sinanoğlu, *J. Am. Chem. Soc.*, **88**, 13 (1966); C. D. Ritchie and H. F. King, *J. Chem. Phys.*, **47**, 564 (1967); A. C. Hopkinson, N. K. Holbrook, K. Yates, and I. G. Csizmadia, *J. Chem. Phys.*, **49**, 3596 (1968); L. C. Snyder, Robert A. Welch Foundation Bulletin 29 (1971); W. C. Ermler and C. W. Kern, *J. Chem. Phys.*, **58**, 3458 (1973).

5. A. C. Hurley, *Adv. Quantum Chem.*, **7**, 315 (1973).

6. J. A. Pople and J. S. Binkley, *Mol. Phys.*, **29**, 599 (1975).

7. P. A. Christiansen and E. A. McCullough, Jr., *J. Chem. Phys.*, **67**, 1877 (1977).

8. See, e.g., *Modern Theoretical Chemistry, Vol. 3, Methods of Electronic Structure Theory*, H. F. Schaefer III, Ed., Plenum, New York, 1977.

9. J. A. Pople and R. K. Nesbet, *J. Chem. Phys.*, **22**, 571 (1954).

10. J. A. Pople and M. S. Gordon, *J. Am. Chem. Soc.*, **89**, 4253 (1967).

11. W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).

12. R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.*, **54**, 724 (1971).

13. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, **28**, 213 (1973).

14. L. Radom, W. J. Hehre, and J. A. Pople, *J. Chem. Soc. Sec. A*, 2299 (1971).

15. L. Radom, W. J. Hehre, and J. A. Pople, *J. Am. Chem. Soc.*, **93**, 289 (1971).

16. T. H. Dunning, Jr., *J. Chem. Phys.*, **53**, 2823 (1970); **55**, 3958 (1971). See also D. Cremer, *J. Chem. Phys.*, **70**, 1898 (1979).

17. C. Møller and M. S. Plesset, *Phys. Rev.*, **46**, 618 (1934).

18. D. Cremer, "Is a good HF basis also a good basis for calculation of correlation corrections?," 15th Symposium für Theoretische Chemie, Berlin, 1979.

19. D. Cremer, *J. Chem. Phys.*, **69**, 4440 (1978).

20. R. B. Davidson and L. C. Allen, *J. Chem. Phys.*, **55**, 519 (1971).

21. A. Veillard, *Theor. Chim. Acta*, **18**, 21 (1970).

22. T. H. Dunning, Jr. and M. W. Winter, *J. Chem. Phys.*, **63**, 1847 (1975).

23. C. Guidotti, U. Lamanna, M. Maestro, and R. Moccia, *Theor. Chim. Acta*, **27**, 55 (1972).

24. W. C. Ermler and C. W. Kern, *J. Chem. Phys.*, **61**, 3860 (1974).

25. D. Cremer, *J. Comput. Chem.*, **3**, 165 (1982), following article in this issue.

26. D. Cremer, unpublished result.

27. R. Ahlrichs, F. Driessler, H. Lischka, and V. Staemmler, *J. Chem. Phys.*, **62**, 1235 (1975), in effect used the EK procedure with an average value of f based on data for BH, HF, and Ne to estimate $E(HF)$ for CH_4 and H_2O.

28. D. Cremer, *J. Am. Chem. Soc.*, **103**, 3619, 3627, 3633 (1981).

29. P. George, M. Trachtman, and C. W. Bock, *J. Comput. Chem.*, **2**, 334 (1981).