

Polarization consistent basis sets: Principles

Frank Jensen

Department of Chemistry, SDU, Odense University, DK-5230 Odense M., Denmark

(Received 10 April 2001; accepted 5 September 2001)

The basis set convergence of Hartree–Fock energies for the H_2 , H_3^+ , C_2 , N_2 , N_4 , O_2 , O_3 , F_2 , HF, and CH_4 molecules is analyzed using optimized basis functions. Based on these analysis a sequence of polarization consistent basis sets are proposed which should be suitable for systematically improving Hartree–Fock and density functional energies. Analogous to the correlation consistent basis sets designed for correlation energies, higher angular momentum functions are included based on their energetical importance. In contrast to the correlation consistent basis sets, however, the importance of higher angular momentum functions decreases approximately geometric, rather than arithmetic. It is shown that it is possible to design a systematic sequence of basis sets for which results converge monotonic to the Hartree–Fock limit. The primitive basis sets can be contracted by a general contraction scheme. It is found that polarization consistent basis sets provide a faster convergence than the correlation consistent basis sets. Results obtained with polarization consistent basis sets can be further improved by extrapolation. © 2001 American Institute of Physics. [DOI: 10.1063/1.1413524]

I. INTRODUCTION

In recent years there has been considerable interest in designing efficient methods for obtaining accurate results, with the primary focus being on energetics. These models combine various standard *ab initio* methods for estimating the infinite basis, infinite correlation result. The Gaussian-1, -2, and -3 models assume simple additivity,^{1–3} while the CBS family of models employ an extrapolation scheme for the most important contribution to the correlation energy.^{4–6} Both these families of models use a small number of empirically adjusted parameters to improve the performance.

The major problem in obtaining accurate results is the slow convergence of the correlation energy as a function of basis set size.⁷ Theoretically the correlation energy is known to converge as an inverse power series as a function of the highest angular momentum included in the basis set.^{8–11} Extrapolation of results obtained with the correlation consistent basis sets cc-pVXZ, (X=D, T, Q, 5, 6) developed by Dunning and co-workers^{12,13} has proved an efficient route for obtaining very accurate results, as exemplified by the recently proposed W1 and W2 models.¹⁴

In recent work we have shown that the basis set convergence of Hartree–Fock (HF) and density functional (DF) energies is exponential,^{15–17} implying a significantly faster convergence than for the correlation energy. Given the impressive accuracy of many DF methods, it would seem desirable to have a systematic way of establishing the basis set limit, and thus the limiting accuracy of a given functional. Common applications^{18,19} and developments of density functionals²⁰ typically use basis sets of double or triple zeta quality, and the inherent error in such models is distributed between the functional and the basis set. With the anticipation of development of more accurate functionals, it would be desirable to be able to quantify the basis set error.

While there is little doubt that results with existing basis

sets, like cc-pVXZ (Refs. 21–23) or even-tempered basis sets,^{24–26} will converge toward the basis set limit, it is not clear that they provide the fastest or smoothest convergence. In the present work we analyze the relative importance of polarization functions and examine the possibility of generating a hierarchical sequence of basis sets for extrapolation to the HF or DF limits. Given the plethora of different DF methods, and the lack of fully numerical reference data, we will in the present paper focus on HF results, but the similarity of HF and DF convergences¹⁷ suggests that the results will be valid for DF methods as well.

II. INITIAL CONSIDERATIONS

Basis sets have traditionally been designed by optimizing the exponents of a suitable number of basis functions with angular momenta required for describing the isolated atom, which for first row elements translate into *s*- and *p*-functions.^{27,28} Normally some of these basis functions are subsequently contracted for improving the computational efficiency. Higher angular momentum functions are added to describe charge polarization at the HF level, and electron correlation at correlated levels. Traditionally these have been called polarization functions, although some prefer the term correlation functions. In correlated calculations they describe both effects, and the correlation effect normally dominates. As the atomic HF energy does not depend on polarization functions, the polarization exponents must be determined either from molecular HF calculations or from correlated calculations on atoms. In older work typically only a single optimum exponent for a given type of polarization function was determined, and multiple polarization functions were generated by symmetrically splitting around the single optimized value.²⁹ More recent work have used explicitly optimized exponents.^{12,13}

A major step towards a systematic way of improving basis set for describing the correlation energy was the atomic natural orbital analysis by Almlöf and co-workers,³⁰ which lead Dunning and co-workers to propose the correlation consistent basis sets of double, triple, quadruple, etc. quality.^{12,13} The important concept in designing these basis sets is that functions which contribute similar amounts of correlation energy are included at the same stage. As shown by detailed analysis, this leads to (contracted) basis sets with compositions $ns(n-1)p(n-2)d(n-3)f$, etc. The original cc-pVXZ basis sets concentrated on recovering the valence correlation energy, but they have later been extended with diffuse functions³¹ (for describing properties) and more tight functions (for describing core and core/valence correlation).³²

As the convergence of HF and DF energies is significantly faster than the correlation energy, it implies that the optimum composition in terms of basis functions will be shifted towards lower values of angular momentum functions than for the correlation consistent basis sets. Furthermore, basis set exponents which are optimum for describing the correlation energy, will not necessarily be optimum for describing charge polarization. In the present paper we analyze the relative importance of polarization functions at the HF level of theory, and propose a new hierarchy of basis sets for approaching the basis set limit. In analogy with the procedure for the development of the correlation consistent basis set, we will employ an energy criterium, i.e., functions which contribute similar amounts of energy are included at the same stage.

III. ENERGY ANALYSIS

Basis set exponents have been optimized by a pseudo-Newton–Raphson approach with $\ln(\zeta)$ as the variables and gradients generated by central finite differences.^{15–17,33} As shown previously,^{15,16} this is capable of achieving nanohartree accuracy in absolute energies compared to numerical HF values. Molecular geometries have either been taken from experimental work or MP2/cc-pVTZ optimized, if experimental geometries were not readily available. For open shell species we have employed the ROHF formalism. Atomic energies have been calculated with spherical averaged densities.

In contrast to the development of the correlation consistent basis sets, it is not possible to employ atoms for analyzing the importance of polarization functions. By using a molecule as the optimization target, there is a risk of biasing the results in the direction of describing a specific molecule better than the general case. In our analysis we have concentrated on using symmetric homonuclear molecules, since the energy partition here is unambiguous. In a heteronuclear molecule like FH the basis functions on hydrogen will to some extent compensate for deficiencies in the fluorine basis (and vice versa). Since we employ fully optimized basis functions, there will be a tendency for the hydrogen functions to become diffuse in order to improve the description of the wave function near the fluorine atom, since this is the energetically important region. Furthermore, in a molecule like FH the relative importance of the hydrogen and fluorine basis functions will be different than in comparable mol-

ecules, like for example CH₄. In the present case we examine the HF energy for the H₂, H₃⁺ (cyclic), C₂, N₂, N₄ (tetrahedral), O₂, O₃ (cyclic), and F₂ and molecules in detail. The diatomic systems sample differences in bonding and nuclear charge, while H₃⁺, N₄, and O₃ test the sensitivity of varying the bonding pattern. In order to obtain more data for especially hydrogen we have also performed analysis of the FH and CH₄ systems.

In order to reduce the dependence of the results on the chosen target molecule, optimized *s*- or *sp*-basis sets were initially generated for the isolated atoms. For the subsequent optimization of polarization functions, we selected a 19*s* atomic basis for hydrogen and a 26*s*17*p* basis for C–F. These basis sets are capable of reproducing the numerical limit for the atomic energies to within a few nanohartree.³⁴ As a good fraction of the molecular binding typically is described by the atomic *s*- and *sp*-orbitals, only the higher angular momentum functions will be biased toward the chosen optimization target. By leaving the polarization functions uncontracted, the bias will only be in the specific values for the exponents, and the importance of the exact values will diminish as multiple functions are included.

Given that the atomic HF error is essentially eliminated with the above atomic basis sets, the higher angular momentum functions describe the charge polarization due to molecular bond formation. Given a fixed number of polarization functions, the optimum composition [number of (*p*-) *d*-, *f*- etc. functions] and exponents are determined, keeping the *s*- or *sp*-exponents at their atomic values. Examining the results for increasingly larger polarization spaces allows an analysis of the relative importance of each type of polarization functions, and thereby selection of a consistent set of polarization functions. Subsequently, a proper number of *s*- or *sp*-functions from the corresponding atomic optimization is selected, based on the criteria that the energy error due to incomplete *s*- or *sp*-function space should be comparable to the error due to incomplete polarization space. Finally, the optimum polarization exponents are determined for a representative set of molecules, and suitable average values are chosen in order to minimize the molecular bias.

The general criteria is that errors from each type of function should be balanced, and the overall error is consequently determined by the highest angular momentum function included. We will coin the name polarization consistent basis sets, with the acronym pc. Increasingly larger basis sets will be denoted by a number indicating the level of polarization beyond the isolated atom, i.e., a pc-1 basis will have a function with an angular momentum one higher than required for the isolated atom, a pc-2 basis will have a function with an angular momentum two higher than required for the isolated atom, etc. In the development stage the pc basis sets are uncontracted, and the question of contraction for improving the computational efficiency is addressed subsequently.

A. Results for N₂

The optimum composition of polarization functions and corresponding energies for N₂ at an internuclear distance of 2.068 a.u. (1.094 Å) using the 26*s*17*p* atomic functions is shown in Table I. The logarithm of the energy lowering per

TABLE I. Hartree–Fock energy convergence for N_2 as a function of number of polarization functions N_{pol} with $26s17p$ -functions fixed at their atomic values.

N_{pol}	Composition	E_{HF}	ΔE (per atom)
0		-108.911 247 020	
1	1d	-108.983 771 370	-0.036 262 175
2	2d	-108.989 601 734	-0.002 915 182
3	2d 1f	-108.992 631 153	-0.001 514 710
4	3d 1f	-108.993 288 156	-0.000 334 002
5	3d 1f 1g	-108.993 538 103	-0.000 119 474
6	4d 1f 1g	-108.993 675 042	-0.000 068 470
7	4d 2f 1g	-108.993 774 793	-0.000 049 756
8	5d 2f 1g	-108.993 791 609	-0.000 008 408
9	5d 2f 1g 1h	-108.993 804 474	-0.000 006 433
10	5d 3f 1g 1h	-108.993 811 932	-0.000 003 729
11	6d 3f 1g 1h	-108.993 816 931	-0.000 002 500
12	6d 3f 2g 1h	-108.993 820 770	-0.000 001 920
13	7d 3f 2g 1h	-108.993 822 002	-0.000 000 616
14	7d 4f 2g 1h	-108.993 823 424	-0.000 000 711
15	7d 4f 2g 1h 1i	-108.993 824 095	-0.000 000 336
HF limit		-108.993 826 ^a	

^aReference 41.

atom by each polarization function is shown in Fig. 1, along with the corresponding atomic s - and p -functions results.

It is clear that the relative importance of the polarization functions will depend on the molecular geometry (at infinite separation polarization functions have no energetic importance), and we have therefore performed a similar analysis at a distance of 2.68 a.u. (1.42 Å) with the results shown in Fig. 1. Compared to the results at $R=2.068$ a.u., the polarization functions are seen to be slightly more important, but the geometry effect is very minor.

At either distance the first d -function has an energy contribution which requires seven s -functions and four p -functions for a comparable error. The kink in the s -function energy contribution for six and seven functions is due to the fact that the sixth function primarily improves the $1s$ -orbital, while the seventh s -function primarily improves the $2s$ -orbital.

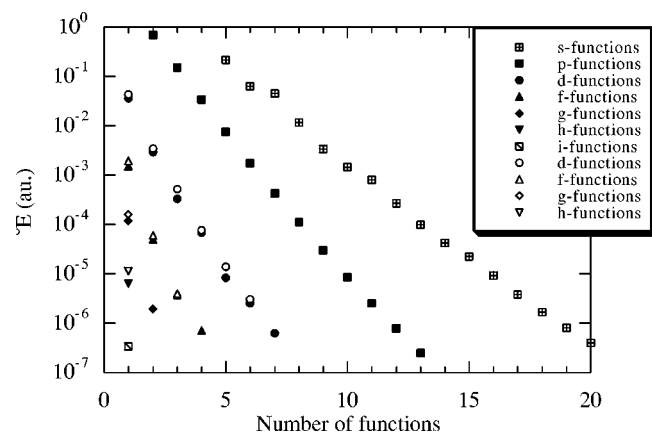


FIG. 1. Energy contribution from each polarization function for N_2 (Table I), and corresponding s - and p -function atomic contributions. Filled symbols are for an internuclear distance of 2.068 a.u., while open symbols are for a distance of 2.68 a.u.

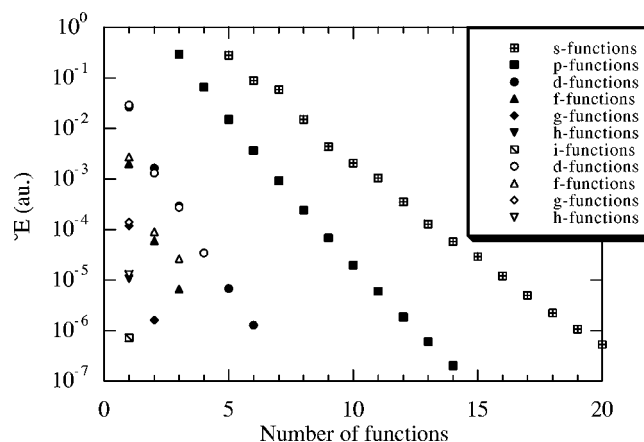


FIG. 2. Energy contribution from each polarization function for O_2 and O_3 (cyclic form) and corresponding s - and p -function atomic contributions. Filled symbols are for O_2 while open symbols are for O_3 .

The second consistent polarization level is $2d1f$, which should be combined with ten s - and six p -functions, i.e., a pc-2 basis set is $10s6p2d1f$ in composition. The pc-3 basis set includes the first g -function, which should be combined with two f - and four d -functions. Depending on the internuclear distance, either 8 or 9 p -functions and either 13 or 14 s -functions should be included. Finally, a pc-4 basis set involving an h -function is $18s11p6d3f2g1h$ in composition.

B. Results for C_2 , O_2 , F_2 , O_3 , and N_4

Plots corresponding to Fig. 1 for C_2 , O_2 , and F_2 are given in Fig. 2 (O_2) and as supplementary material.³⁵ The consistent choices for pc- n basis sets for these molecules are given in Table II, where the notation 7/8 indicates that the consistent choice is intermediate between seven and eight functions. Given that the C_2 – F_2 series includes variation both in bonding and internuclear distance, there is a good agreement between the consistent choices.

Although the homonuclear diatomic molecules sample a variation in orbital occupation and nuclear charge, they all have significant π -bonding. This is unlike the majority of applications which have molecules dominated by σ -bonding. We thus also included O_3 in a cyclic triangular form and the tetrahedral N_4 molecule in the analysis. The latter has a bonding pattern resembling the typical single bonding in many molecule. The results for O_3 are shown in Fig. 2 together with the O_2 data, while the N_4 results are provided as supplementary material.³⁵ The associated choices for the pc-1, pc-2, and pc-3 basis sets are given in Table II. Given that these two molecules have very different bonding than O_2 and N_2 , the good agreement indicates that the analysis on the homonuclear diatomics are representative for systems in general.

C. Results for H_2 , H_3^+ , FH , and CH_4

A logarithmic plot of the energy contributions for each polarization function for H_2 at an internuclear distance of 1.40 a.u. (0.74 Å) using the 19 atomic s -functions is shown in Fig. 3. The equilibrium bond distance in H_2 is significantly shorter than typical bonding distances between hydrogen and

TABLE II. Consistent choices for polarization functions.

	pc-1			pc-2				pc-3				pc-4						
	<i>s</i>	<i>p</i>	<i>d</i>	<i>s</i>	<i>p</i>	<i>d</i>	<i>f</i>	<i>s</i>	<i>p</i>	<i>d</i>	<i>f</i>	<i>g</i>	<i>s</i>	<i>p</i>	<i>d</i>	<i>f</i>	<i>g</i>	<i>h</i>
C ₂	7/8	4	1	10	6	2	1	13/14	8	3/4	2	1	17/18	10/11	5/6	3	2	1
N ₂ (<i>R</i> =2.068)	7	4	1	10	6	2	1	13/14	8/9	4	2	1	18	11	6	3	2	1
N ₂ (<i>R</i> =2.68)	7	4	1	10	6	2	1	14	9	4	2	1						
O ₂	7/8	4/5	1	10	6/7	2	1	13/14	8/9	3/4	2	1	17	11	5	3	1/2	1
F ₂	8/9	5/6	1	12	8	2	1	14	9/10	3	2	1	17	11	4/5	3	1/2	1
N ₄	7	4	1	9/10	6	2	1											
O ₃	7/8	4/5	1	10/11	6/7	2	1	13/14	8/9	3/4	2	1						
H ₂ (<i>R</i> =1.40)	3/4	1		7	3	1		11	5/6	2/3	1		15	8	4/5	2	1	
H ₂ (<i>R</i> =2.80)	4	1		7	2/3	1		10	4/5	2	1		13/14	6/7	4	2	1	
H ₃ ⁺	3/4	1		6/7	2/3	1		11	5	2/3	1		15	7/8	4/5	2	1	
H(CH ₄)	3/4	1		6	2	1		8	3	2	1		10	4/5	3	2	1	
H (FH)	3/4	1		5/6	2/3	1		7	3/4	1/2	1							

other elements. We have therefore also performed the analysis at an internuclear distance of 2.80 a.u. (1.48 Å) with the results shown Fig. 3. Compared to the results at $R = 1.40$ a.u., the p -functions are less important, the d -function contribution change only slightly, while the f - and g -functions become more important. The corresponding results for H₃⁺ (regular triangle with an internuclear distance of 1.65 a.u.) are shown as supplementary material.³⁵

In all three cases, the first p -function is the most important polarization function, and either three or four s -functions should be selected to give a balanced description. For H₂ at $R = 1.40$ a.u. the contribution from the first d -function is comparable to that of the third p -function, while for H₂ at $R = 2.80$ a.u. and for H₃⁺ the contribution from the first d -function is intermediate between the second and third p -function. Either six or seven s -function are required to give a comparable atomic error.

The energy contribution from the first f -function is intermediate between the second and third d -function, the fifth p -function, and 10 or 11 s -functions. Similar the energy contribution from the first g -function is comparable to that from

the second f -function, the fourth or fifth d -function and the seventh or eight p -function. 14 or 15 s -functions are required to give a balanced description.

Hydrogen is special since it is the only chemically important element which only has valence electrons. Given the relatively large sensitivity of the polarization functions to the internuclear distance (Fig. 3), we have also performed analysis on the FH and CH₄ molecules. Although the energy partitioning is not unique in these systems, they provide information about the relative importance of the hydrogen polarization functions. The conclusions for the hydrogen polarization functions are given in Table II.

IV. POLARIZATION CONSISTENT BASIS SETS

A. Composition

The consistent choices for polarization functions based on energetical analysis for the above molecules are shown in Table II. In practical calculations it is more important to have a good balance between basis sets for a range of atoms, than to have a good absolute energy for each individual atom. We therefore wish to have the same basis set composition for C, N, O, and F, and we propose a set of polarization consistent basis sets with the compositions shown in Table III. We have also included an unpolarized pc-0 basis set with the composition $5s3p$. Although this is not expected to give useful accuracy, it provides a reference point for extrapolations, as discussed below.

For hydrogen the choice is less obvious, since the results dependent somewhat on the molecular system used for analysis. Based primarily on the results for FH, CH₄, and H₂ at $2R_e$, we propose that a polarization consistent basis set for hydrogen at the first polarization level, pc-1, is $4s1p$ in composition. At the second level, pc-2, the consistent choice is $6s2p1d$. Similar the consistent choices for pc-3 and pc-4 are $9s4p2d1f$ and $11s6p4d2f1g$. We note that these choices follow the composition of polarization functions for the C–F elements. Analogously we have defined a pc-0 basis set as $3s$ for hydrogen.

Also shown in Table III are compositions for two other widely used basis sets, the correlation consistent and Pople style STO-3G and k - $lmnG$ basis sets. Of these only the cor-

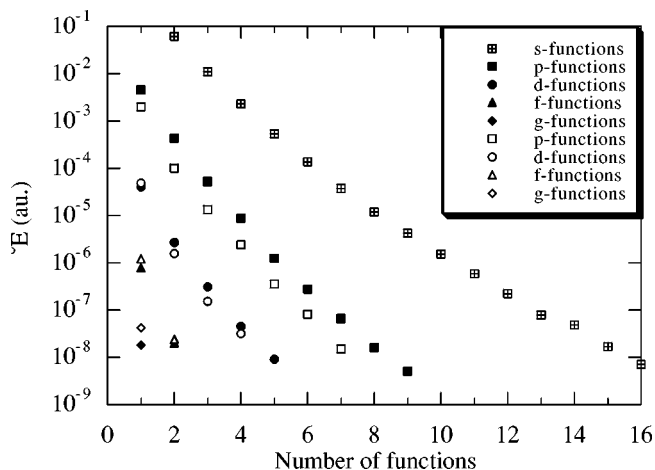


FIG. 3. Energy contribution from each polarization function for H₂, and corresponding s -function atomic contributions. Filled symbols are for an internuclear distance of 1.40 a.u., while open symbols are for a distance of 2.80 a.u.

TABLE III. Basis set compositions (uncontracted).

C-F					
pc-0	5s 3p			STO-3G	6s 3p
pc-1	7s 4p 1d	cc-pVDZ	9s 4p 1d	6-31G(<i>d,p</i>)	11s 4p 1d
pc-2	10s 6p 2d 1f	cc-pVTZ	10s 5p 2d 1f	6-311G(2 <i>df</i> ,2 <i>pd</i>)	11s 5p 2d 1f
pc-3	14s 9p 4d 2f 1g	cc-pVQZ	12s 6p 3d 2f 1g		
pc-4	18s 11p 6d 3f 2g 1h	cc-pV5Z	14s 8p 4d 3f 2g 1h		
		cc-pV6Z	16s 10p 5d 4f 3g 2h 1i		
H					
pc-0	3s			STO-3G	3s
pc-1	4s 1p	cc-pVDZ	4s 1p	6-31G(<i>d,p</i>)	4s 1p
pc-2	6s 2p 1d	cc-pVTZ	5s 2p 1d	6-311G(2 <i>df</i> ,2 <i>pd</i>)	5s 2p 1d
pc-3	9s 4p 2d 1f	cc-pVQZ	6s 3p 2d 1f		
pc-4	11s 6p 3d 2f 1g	cc-pV5Z	8s 4p 3d 2f 1g		
		cc-pV6Z	10s 5p 4d 3f 2g 1h		

relation consistent basis sets provide a systematic hierarchy for approaching the basis set limit. While the correlation consistent basis sets have the importance of the angular momentum functions decreasing arithmetically, the polarization consistent basis sets have the importance decreasing roughly geometrically, i.e., balanced pc basis sets have the approximate composition $ns(n/2)p(n/4)d(n/8)f$, etc.

The difference in composition between the cc- and pc-basis sets can be rationalized as follows. With only a single polarization function, the combination of charge polarization and electron correlation makes the *d*-function energy contribution larger for the cc-pVDZ basis set than for the pc-1 basis set, and the cc-basis set consequently includes more *s*-functions than the pc-1 basis set. For the larger basis sets, the exponential convergence of the HF energy makes the lower angular momentum functions more important than for the correlation energy. The pc-3 basis set thus have two *s*-functions, three *p*-functions, and one *d*-function more than the cc-pVQZ basis. The pc-2/cc-pVTZ basis sets constitute the point where these two effects roughly balance out, and they are consequently almost identical in composition.

B. Exponents

The energy analysis above are based on atomic basis sets which are effectively saturated and with optimum polarization exponents. For consistency reasons we argue that the polarization exponents should be reoptimized using the reduced *s*- or *sp*-basis, as given in Table III. Furthermore, since the optimum exponents are quite sensitive to the geometry (internuclear distances), we wanted to investigate the optimum exponents for a selection of typical molecules, and using the large 26*s*17*p* atomic basis set makes this cumbersome. Initial explorations with reoptimization of the polarization exponents using the reduced *sp*-basis sets showed that this in some cases produced very diffuse functions. The problem is that the reduced *sp*-function space does not contain sufficient diffuse functions (which are present in the 26*s*17*p* basis) to prevent polarization function from drifting outward to describe the wave function tail. This is particularly troublesome for polar systems or with large basis sets, like pc-3 and pc-4. A heuristic fix was made by including one additional diffuse *s*- and *p*-function, with an expo-

TABLE IV. Optimum polarization exponents as a function of the *sp*-basis.^a

Molecule	<i>sp</i>	<i>d</i> ₁	<i>d</i> ₁	<i>d</i> ₂	<i>f</i> ₁	<i>d</i> ₁	<i>d</i> ₂	<i>d</i> ₃	<i>d</i> ₄	<i>f</i> ₁	<i>f</i> ₂	<i>g</i> ₁	
F ₂	26 <i>s</i> 17 <i>p</i>	1.03	3.03	0.83	1.31	13.00	3.48	1.11	0.48	3.52	1.05	1.46	
	7 <i>s</i> 4 <i>p</i>	0.96											
	7 <i>s</i> 4 <i>p</i> + <i>sp</i>	1.01											
	10 <i>s</i> 6 <i>p</i>		2.64	0.73	1.27								
	10 <i>s</i> 6 <i>p</i> + <i>sp</i>		3.03	0.80	1.29								
	14 <i>s</i> 9 <i>p</i>						5.17	1.39	0.60	0.16	3.54	1.06	1.55
	14 <i>s</i> 9 <i>p</i> + <i>sp</i>						12.91	3.47	1.09	0.46	3.56	1.05	1.46
N ₂	26 <i>s</i> 17 <i>p</i>	0.97	1.66	0.50	1.41	6.13	1.69	0.68	0.25	1.82	0.74	1.91	
	7 <i>s</i> 4 <i>p</i>	0.93											
	7 <i>s</i> 4 <i>p</i> + <i>sp</i>	0.93											
	10 <i>s</i> 6 <i>p</i>		1.59	0.48	1.37								
	10 <i>s</i> 6 <i>p</i> + <i>sp</i>		1.61	0.49	1.36								
	14 <i>s</i> 9 <i>p</i>						4.33	1.27	0.44	0.11	1.76	0.73	1.90
	14 <i>s</i> 9 <i>p</i> + <i>sp</i>						5.95	1.64	0.65	0.24	1.79	0.72	1.90

^aNotation 7*s*4*p* + *sp* indicates that the 7*s*4*p* atomic basis has been augmented by an additional set of diffuse *s*- and *p*-functions.

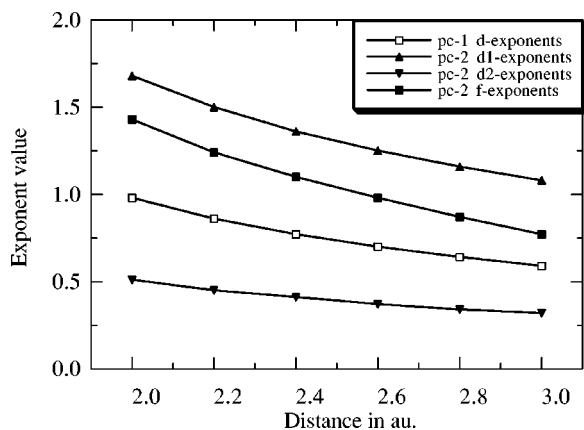


FIG. 4. Optimum polarization exponent for the augmented pc-1 and pc-2 basis sets for N_2 as a function of internuclear distance.

ment 2.5 times smaller than the outermost function. Table IV shows that this essentially removes the problem, and leads to optimum exponents close to those for the large $26s19p$ atomic basis.

The optimum polarization exponents for the pc-1 and pc-2 (augmented with diffuse s - and p -functions) for N_2 as a function of internuclear distance are shown in Fig. 4. It is clear that there is a quite strong dependence of the exponents on the molecular geometry. Table V shows the optimum d -function exponents for the augmented pc-1 basis for all the diatomic molecules corresponding to combinations of H, C, N, O, and F. For the HX systems a common internuclear distance of 1.90 a.u. (1.01 Å) was used, while for the XY molecules a common distance of 2.30 a.u. (1.22 Å) was employed. Table V thus samples the sensitivity of the exponent to the introduction of molecular asymmetry at a constant geometry. Figure 4 and Table V together show that the optimum exponents depend both on the molecular geometry and the bonding situation.

In order to select suitable values for the polarization exponents, we have optimized the exponents for a selection of small molecules. The results for the augmented pc-1 and pc-2 basis sets are shown in Tables VI–X, and results for the augmented pc-3 basis set for a smaller sample of molecules

TABLE V. Optimum polarization exponents for the augmented pc-1 basis set. $R_{HX}=1.90$ a.u., $R_{XY}=2.30$ a.u.

Molecule	H	C	N	O	F
CH	0.79	0.79			
NH	0.86		0.93		
OH	0.72			1.16	
FH	0.87				1.13
C ₂		0.74			
CN		0.64	0.91		
CO		0.74		0.95	
CF		0.87			1.09
N ₂			0.81		
NO			0.89	0.93	
NF			1.03		1.09
O ₂				1.01	
OF				1.15	1.12
F ₂					1.22

TABLE VI. Optimum hydrogen polarization exponents for the augmented pc-1 and pc-2 basis sets.

Molecule	pc-1		pc-2	
	$1p$	$1p$	$2p$	$1d$
CH ₄	0.93	1.63	0.53	1.29
C ₂ H ₆	0.97	1.54	0.47	1.21
C ₃ H ₈	1.02	1.52	0.47	1.22
C ₄ H ₁₀	1.03	1.50	0.46	1.21
C ₂ H ₄	0.99	1.70	0.53	1.28
C ₆ H ₆	1.04	1.77	0.55	1.30
C ₂ H ₂	0.95	2.12	0.56	1.07
HCN	0.92	1.81	0.39	0.43
NH ₃	0.70	1.35	0.28	1.58
N ₂ H ₄	0.80	1.38	0.32	1.49
H ₂ O	0.69	1.49	0.29	1.75
H ₂ O ₂	0.79	1.52	0.33	1.66
HF	0.86	1.66	0.36	1.92
CH ₃ NH ₂	0.84	1.33	0.34	1.40
CH ₃ OH	0.84	1.35	0.33	1.52
CH ₃ F	0.87	1.88	0.61	1.27
(CH ₃) ₃ N	0.97	1.73	0.55	1.19
(CH ₃) ₂ O	0.95	2.03	0.65	1.23
H ₂ CO	0.90	2.28	0.71	1.21
(CH ₃) ₂ CO	0.86	1.44	0.41	1.23
CH ₃ CONH ₂	0.95	1.66	0.36	1.92

are given in Table XI. It is clear that the optimum exponents for a given element in a molecule depend both on the nature of the neighbor element and the bond distance (*vide supra*). Compare for example the optimum pc-1 d -exponent for oxygen in H₂O, CH₃OH, and (CH₃)₂O, 1.25, 1.00, 0.88 (Table IX). In H₂O both the bonds are short, favoring a large exponent value, while the longer bonds in (CH₃)₂O favor a smaller value. The CH₃OH molecule is intermediate between these two. The difference between the optimum exponent value for (CH₃)₂O (0.88) and (CH₃)₂CO (1.04) is primarily due to the shorter bond distance between carbon and oxygen in the latter. It should be noted that small molecules like

TABLE VII. Optimum carbon polarization exponents for the augmented pc-1 and pc-2 basis sets.

Molecule	pc-1		pc-2	
	$1d$	$1p$	$2d$	$1f$
CH ₄	0.92	1.45	0.41	1.84
C ₂ H ₆	0.83	1.53	0.47	0.80
C ₃ H ₈	0.82	1.57	0.49	0.79
C ₄ H ₁₀	0.81	1.58	0.49	0.79
C ₂ H ₄	0.83	1.31	0.41	1.00
CH ₂ NH	0.78	1.13	0.27	1.22
C ₆ H ₆	0.83	1.49	0.50	0.83
C ₂ H ₂	0.82	1.58	0.50	1.12
HCN	0.73	1.26	0.31	1.20
CH ₃ NH ₂	0.79	1.36	0.41	1.00
CH ₃ OH	0.78	1.24	0.35	1.07
CH ₃ F	0.81	1.23	0.35	1.06
CF ₄	0.80	1.26	0.39	1.00
(CH ₃) ₃ N	0.80	1.33	0.41	0.98
(CH ₃) ₂ O	0.79	1.25	0.36	1.11
H ₂ CO	0.77	1.26	0.33	1.35
(CH ₃) ₂ CO	0.78	1.43	0.46	0.92
CH ₃ CONH ₂	0.76	1.27	0.37	1.20

TABLE VIII. Optimum nitrogen polarization exponents for the augmented pc-1 and pc-2 basis sets.

Molecule	pc-1		pc-2	
	1 <i>d</i>	1 <i>d</i>	2 <i>d</i>	1 <i>f</i>
NH ₃	1.09	1.92	0.52	0.71
N ₂ H ₄	0.93	1.91	0.58	0.97
N ₂ H ₂	0.86	1.49	0.45	1.20
N ₂	0.94	1.61	0.49	1.36
HCN	0.97	2.05	0.66	1.20
H ₂ CNH	0.85	1.78	0.54	0.96
CH ₃ NH ₂	0.96	1.98	0.58	0.82
(CH ₃) ₃ N	0.87	2.10	0.64	0.87
CH ₃ CONH ₂	1.18	2.09	0.66	0.88

CH₄, NH₃, H₂O, and HF, despite their popularity for testing purposes, are not representative for the general case, since they only have short bonds.

Based on the results in Tables VI–XI we selected an initial set of representative polarization exponents for the pc-1, -2, and -3 basis sets. These exponents were analyzed and adjusted to conform to common principles, e.g., exponents should increase with the nuclear charge and with the angular momentum. As seen from the results in Table IX and X, the optimum *d*-exponent for fluorine tend to be smaller than for oxygen in the molecules sampled. This is mainly due to the fact that oxygen participates in more varied bonding environments than fluorine. In many fluorine containing systems of common interest the fluorine is bonded to carbon, and the three lone pairs on fluorine tend to favor a relative small value for the *d*-exponent. Oxygen, on the other hand, is typically found in carbonyl, alcohol or ether functional groups. In both carbonyl and alcohol groups the short bond distance favors a large oxygen exponent value. The bonding in ethers is similar to that in carbon–fluorine, and here a slightly smaller optimum exponent is indeed found.

As the basis set becomes larger, the energy dependence on each exponent becomes a softer function, leading to more variation in the optimum exponents with respect to molecular environment. Polarization exponents for the pc-4 basis were thus estimated from extrapolation from the pc-1, -2, and -3 basis sets, and from results with explicit optimized 26s17p6d3f2g1h basis sets for the C₂, N₂, O₂, and F₂ systems (19s6p3d2f1g for H₂). All basis set exponents for the pc-*n* basis sets employed in this work are given as supple-

TABLE IX. Optimum oxygen polarization exponents for the augmented pc-1 and pc-2 basis sets.

Molecule	pc-1		pc-2	
	1 <i>d</i>	1 <i>d</i>	2 <i>d</i>	1 <i>f</i>
H ₂ O	1.25	2.23	0.60	1.00
H ₂ O ₂	0.99	2.29	0.66	1.10
F ₂ O	1.00	1.88	0.58	1.07
F ₂ O ₂	1.02	1.87	0.60	1.26
CH ₃ OH	1.00	2.37	0.67	1.00
(CH ₃) ₂ O	0.88	2.50	0.71	1.00
H ₂ CO	1.05	2.16	0.67	1.14
(CH ₃) ₂ CO	1.04	1.84	0.58	1.25
CH ₃ CONH ₂	1.04	2.22	0.70	1.23

TABLE X. Optimum fluorine polarization exponents for the augmented pc-1 and pc-2 basis sets.

Molecule	pc-1		pc-2	
	1 <i>d</i>	1 <i>d</i>	2 <i>d</i>	1 <i>f</i>
HF	1.33	2.51	0.69	1.33
CH ₃ F	0.91	3.18	0.83	1.18
CH ₂ F ₂	0.92	3.07	0.83	1.20
CF ₄	0.95	2.91	0.81	1.21
F ₂ O	0.93	3.21	0.84	1.29
F ₂ O ₂	0.70	2.51	0.64	1.06

mentary material.^{35,36} Compared to the cc-pVXZ basis sets, the polarization exponents vary less with the nuclear charge. The cc-pVDZ/pc-1 and cc-pVTZ/pc-2 basis sets are quite similar in composition, but the pc-1/2 basis sets have different polarization exponents. For H, C, and N the pc-exponents are larger than those for the corresponding cc-basis sets, while the opposite is true for O and F. The pc-3 and pc-4 basis sets have significantly more low angular momentum functions than the corresponding cc-pVQZ and cc-pV5Z basis sets, and it is clear that the compositions will make the cc-pVXZ basis sets less than optimum for HF and DF calculations.

We note that the present analysis is based on neutral molecules, and there may be systematic deficiencies for other types of molecules and properties. For systems and properties depending on the wave function tail we expect that an explicit addition of diffuse functions will improve the performance, analogous to the situation for the augmented correlation consistent basis sets.³¹ Such extensions will be considered at a later stage.

TABLE XI. Optimum polarization exponents for the augmented pc-3 basis set.

Molecule	1 <i>d</i>	2 <i>d</i>	3 <i>d</i>	4 <i>d</i>	1 <i>f</i>	2 <i>f</i>	1 <i>g</i>
	H						
H ₂ O ₂	6.63	1.47	0.48	0.18	1.77	0.40	2.29
N ₂ H ₄	6.53	1.46	0.48	0.17	1.71	0.44	2.24
C ₂ H ₆	6.76	1.53	0.53	0.20	1.88	0.63	2.12
CH ₃ NH ₂	6.57	1.49	0.50	0.15	1.83	0.58	2.20
CH ₃ OH	6.68	1.52	0.50	0.16	1.89	0.57	2.18
CH ₃ F	6.75	1.57	0.54	0.20	2.01	0.67	2.17
H ₂ CO	6.98	1.64	0.52	0.21	2.22	0.61	0.82
C							
C ₂ H ₆	4.74	1.32	0.52	0.20	1.68	0.72	1.22
CH ₃ NH ₂	4.45	1.23	0.48	0.18	1.25	0.42	1.37
CH ₃ OH	4.30	1.18	0.46	0.18	1.25	0.42	1.38
CH ₃ F	4.50	1.23	0.48	0.17	1.36	0.55	1.35
H ₂ CO	4.46	1.19	0.45	0.16	1.34	0.43	1.58
N							
N ₂ H ₄	5.99	1.73	0.67	0.26	2.13	0.80	1.18
CH ₃ NH ₂	5.73	1.66	0.63	0.23	2.09	0.77	1.13
O							
H ₂ O ₂	7.49	2.17	0.80	0.32	2.66	0.87	1.25
CH ₃ OH	6.59	1.94	0.73	0.25	2.59	0.89	1.28
H ₂ CO	7.02	1.97	0.77	0.28	2.35	0.91	1.56
F							
CH ₃ F	10.14	3.01	1.00	0.41	3.48	1.08	1.47

TABLE XII. Errors in the total atomization energy, equilibrium distance, and dipole moment relative to the HF limit for the NH molecule with uncontracted basis sets.

Basis set	M	ΔE_{atom} (a.u.)	ΔR_e (Å)	$\Delta\mu$ (D)
HF-limit ^a		-0.077 651	1.017 313	1.6133
STO-3G	6	0.006 080	0.061 898	-0.3048
6-31G(d,p)	35	0.003 038	0.005 899	0.0944
6-311G(2 $df,2pd$)	59	0.001 416	0.000 619	0.0012
cc-pVDZ	33	0.004 375	0.006 669	0.0605
cc-pVTZ	58	0.001 085	0.001 076	0.0236
cc-pVQZ	100	0.000 345	0.000 218	0.0139
cc-pV5Z	166	0.000 095	0.000 038	0.0120
cc-pV6Z	256	0.000 024	0.000 009	0.0062
pc-0	17	-0.016 581	0.027 843	0.2479
pc-1	31	0.004 503	0.005 210	0.0844
pc-2	62	0.000 866	0.000 546	0.0636
pc-3	122	0.000 061	0.000 007	0.0082
pc-4	198	0.000 005	0.000 000	0.0006

^aReference 41. Error is defined as value-reference value. ΔE_{atom} and $\Delta\mu$ are for a fixed internuclear distance of 1.95 a.u.

V. COMPARISON WITH NUMERICAL HARTREE-FOCK RESULTS

Atoms and diatomic molecules can be subjected to numerical HF calculations³⁷ which provide the reference point for finite basis set calculations. We have selected the NH, FH, CO, and NF molecules for testing the performance of the pc- n basis sets, and compared the results with those from the cc-pVXZ basis sets.^{12,13} In order to facilitate a direct comparison, these have been used in the uncontracted form. We furthermore compare the results with those from the popular STO-3G,³⁸ 6-31G(d,p),³⁹ and 6-311G(2 $df,2pd$) (Refs. 29,40) basis sets, with the latter two in their uncontracted forms. We note that the cc-pVXZ basis sets have not been designed for HF or DF calculations, but they are currently the only other hierarchical basis sets available for comparison.

We will focus on three properties: total atomization energy, equilibrium distance, and dipole moment. The first two of these are especially relevant for many typical applications of DF methods. The dipole moment is known to be sensitive to the wave function tail, and therefore a property which is likely to be improved by addition of diffuse functions. Only the results for NH is discussed in detail; corresponding results for CO, FH, and NF are provided as supplementary material.³⁵

The results for the four diatomic molecules are shown in Tables XII–XV in addition to the results from numerical HF calculations,⁴¹ and the logarithmic errors relative to the limiting values for the NH molecule are shown in Figs. 5–7. The abscissa axis is the total number of functions since the cc and pc basis sets have different compositions.

For NH there is little difference in performance of the three DZP type basis sets, pc-1, cc-pVDZ, and 6-31G(d,p) for the total atomization energy, Fig. 5 and Table XII. The pc-3 result, however, is better than the result with the cc-pV5Z basis, despite the smaller number of functions. Furthermore, the results from the pc- n sequence of basis sets clearly provides a faster convergence than those from the

TABLE XIII. Errors in the total atomization energy, equilibrium distance, and dipole moment relative to the HF limit for the CO molecule with uncontracted basis sets.

Basis set	M	ΔE_{atom} (a.u.)	ΔR_e (Å)	$\Delta\mu$ (D)
HF-limit ^a		-0.292 890	1.101 780	0.2650
STO-3G	10	0.070 953	0.043 701	-0.4332
6-31G(d,p)	56	0.009 165	0.007 459	0.0487
6-311G(2 $df,2pd$)	86	-0.000 055	0.000 586	-0.0040
cc-pVDZ	52	0.010 141	0.004 026	-0.0253
cc-pVTZ	84	0.001 589	0.001 840	-0.0039
cc-pVQZ	136	0.000 185	0.000 204	0.0002
cc-pV5Z	216	0.000 078	0.000 040	0.0015
cc-pV6Z	322	-0.000 024	0.000 013	0.0004
pc-0	28	0.075 587	0.044 224	0.4093
pc-1	48	0.007 511	0.004 089	0.0333
pc-2	90	0.001 358	0.000 464	0.0244
pc-3	168	0.000 096	0.000 067	0.0021
pc-4	262	0.000 001	0.000 014	0.0003

^aReference 41. Error is defined as value-reference value. ΔE_{atom} and $\Delta\mu$ are for a fixed internuclear distance of 2.132 a.u.

cc-pVXZ basis sets. The results for the equilibrium distance follows the same trend, Fig. 6 and Table XII. Note that the pc-4 result is missing in Fig. 6, since it reproduces the limiting value to within the numerical accuracy.

The electric dipole moment (for a fixed geometry) also converge monotonic, Fig. 7 and Table XII, but less smoothly than the atomization energy and equilibrium distance. The convergence of the pc- n basis set results has a pronounced kink between pc-1 and pc-2, and the pc-2 result is inferior to the corresponding cc-pVTZ results. This is a general results, and is due to the fact that the outer $d(p)$ -function for C, N, and H in the pc-2 basis set is larger than for the corresponding cc-pVTZ. The dipole moment is a first order property and therefore sensitive to the presence of sufficiently diffuse $d(p)$ -functions. The tighter $d(p)$ -functions for the pc-2 basis set improves the atomization energies and equilibrium distances, but deteriorates the dipole moment. It is expected that addition of diffuse functions will remedy this problem, and

TABLE XIV. Errors in the total atomization energy, equilibrium distance, and dipole moment relative to the HF limit for the FH molecule with uncontracted basis sets.

Basis set	M	ΔE_{atom} (a.u.)	ΔR_e (Å)	$\Delta\mu$ (D)
HF-limit ^a		-0.161 453	0.897 005	1.9218
STO-3G	6	0.043 765	0.058 458	-0.6333
6-31G(d,p)	35	0.010 674	0.003 776	0.0425
6-311G(2 $df,2pd$)	59	0.002 761	-0.000 822	-0.0070
cc-pVDZ	33	0.010 107	-0.000 903	0.0659
cc-pVTZ	58	0.002 365	0.000 990	0.0194
cc-pVQZ	100	0.000 672	-0.000 106	0.0117
cc-pV5Z	166	0.000 138	-0.000 056	0.0103
cc-pV6Z	256	0.000 034	-0.000 020	0.0052
pc-0	17	0.024 859	0.023 477	0.2789
pc-1	31	0.009 668	0.001 258	0.0792
pc-2	62	0.001 404	0.000 075	0.0606
pc-3	122	0.000 064	0.000 001	0.0052
pc-4	198	0.000 007	0.000 001	0.0005

^aReference 41. Error is defined as value-reference value. ΔE_{atom} and $\Delta\mu$ are for a fixed internuclear distance of 1.7328 a.u.

TABLE XV. Errors in the total atomization energy, equilibrium distance, and dipole moment relative to the HF limit for the NF molecule for uncontracted basis sets.

Basis set	M	ΔE_{atom} (a.u.)	ΔR_e (Å)	$\Delta\mu$ (D)
HF-limit ^a		-0.032 134	1.293 090	0.3668
STO-3G	10	0.008 389	0.044 089	-0.1380
6-31G(<i>d,p</i>)	56	0.006 882	0.013 144	-0.0480
6-311G(2 <i>df</i> ,2 <i>pd</i>)	86	0.001 173	-0.001 398	-0.0006
cc-pVDZ	52	0.009 300	0.005 164	0.0892
cc-pVTZ	84	0.001 374	0.001 134	0.0116
cc-pVQZ	136	0.000 429	0.000 160	0.0045
cc-pV5Z	216	0.000 127	0.000 043	0.0059
cc-pV6Z	322	0.000 029	-0.000 006	0.0030
pc-0	28	0.015 570	0.086 322	0.2348
pc-1	48	0.007 039	0.006 621	0.0652
pc-2	90	0.001 353	0.001 146	0.0355
pc-3	168	0.000 134	0.000 011	0.0057
pc-4	262	0.000 013	0.000 000	0.0006

^aReference 41. Error is defined as value-reference value. ΔE_{atom} and $\Delta\mu$ are for a fixed internuclear distance of 2.49 a.u.

such extensions will be considered later. The error in the dipole moment for the STO-3G basis set is of the opposite sign of the others, but the 6-311G(2*df*,2*pd*) result is actually very close to the limiting result. This appears to be a general result, and is presumably due to the presence of slightly more diffuse polarization functions in the 6-311G(2*df*,2*pd*) basis set.

The corresponding results for CO, FH, and NF are shown in Tables XIII–XV, with similar trends as for NH. We note in passing that the total energy for CO with the pc-4 basis set, which contains 262 functions, is very close to that of an even-tempered basis set containing 623 functions.²⁶ Another notable point is the dipole moment of CO, which is a difficult property for quantum chemical methods. For the cc-pVXZ and STO-3G, 6-31G(*d,p*), 6-311G(2*df*,2*pd*) basis sets, the convergence is erratic, with the cc-pVQZ result accidentally close to the limiting value. The pc-*n* results, on the other hand, converges smoothly to the limiting value from above, with the previous mentioned kink between the pc-1 and pc-2 results.

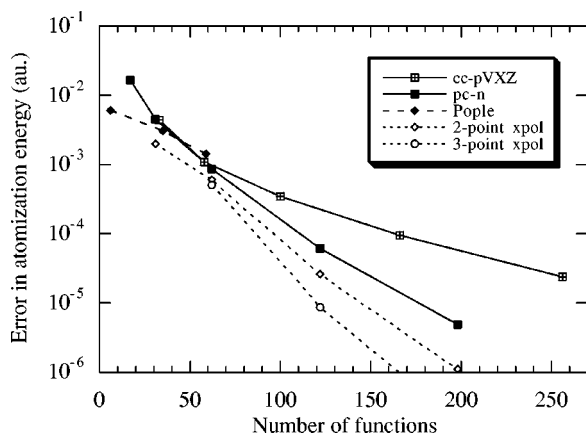


FIG. 5. Errors in atomization energy (a.u.) relative to the HF limit for NH for different uncontracted basis sets (Table XII). Open symbols corresponds to either 2- or 3-point extrapolations based on the pc-*n* results and Eq. (7) in the text.

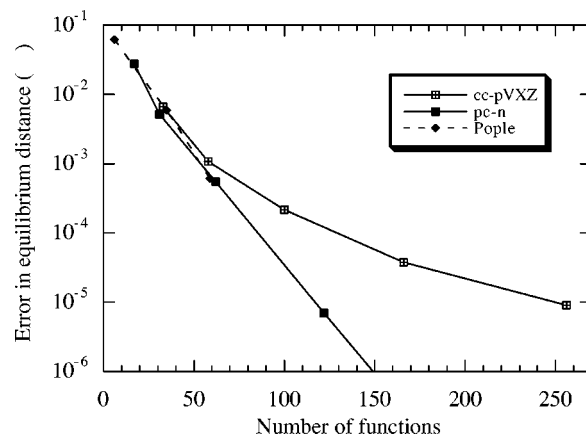


FIG. 6. Errors in equilibrium distance (Å) relative to the HF limit for NH for different uncontracted basis sets (Table XII). The pc-4 result is not shown as it reproduces the limiting value to within the numerical accuracy.

For all of our present 9 testcases ($\text{H}_2, \text{C}_2, \text{N}_2, \text{O}_2, \text{F}_2, \text{NH}, \text{FH}, \text{CO}, \text{NF}$) the total atomization energy, the equilibrium distance and the dipole moment calculated with the pc-*n* basis sets converge monotonic towards the limiting value. The pc-4 results are in excellent agreement with numerical HF data,⁴¹ and the pc-3 results are already close to the HF limit. The results from the cc-pVXZ basis sets, and the STO-3G, 6-31G(*d,p*), 6-311G(2*df*,2*pd*) sequence, converges less systematic and in some cases overshoots the limiting value, as seen from Tables XII–XV.

VI. EXTRAPOLATIONS BASED ON pc-*n* BASIS SETS

The correlation energy is known from theory to converge as an inverse power series in the maximum angular momentum included in the basis set, with the leading term being L^{-3} .^{8–11} As this is fairly slowly convergent, extrapolation is necessary for achieving a high accuracy. The correlation consistent basis sets have been shown to form a good basis for such extrapolations.^{14,22,23,42–45} It has been found that generalized inverse power formulas like $E_L = E_\infty + AL^{-B}$ do not provide improvement over a fixed form with $B=3$, i.e., E_L

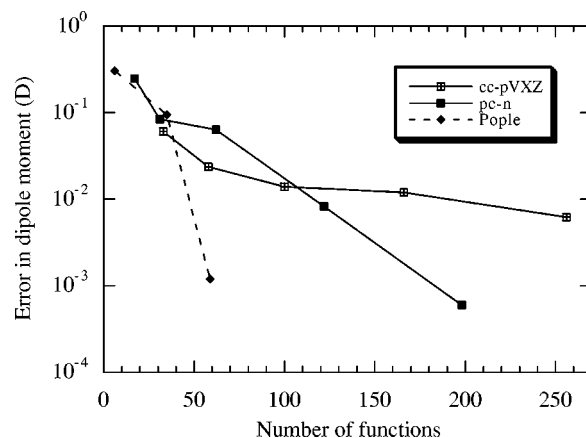


FIG. 7. Errors in dipole moment (Debye) relative to the HF limit for NH for different uncontracted basis sets (Table XII). The point corresponding to the STO-3G basis set (Pople, $M=6$) has an error of different sign that the other points.

$=E_\infty + AL^{-3}$.^{22,42-45} This is due to the fact that the former is a three point extrapolation, compared to the two point extrapolation with $B=3$. A three point extrapolation scheme requires that one additional point must be included, and for a given maximum L value, the additional $(L-2)$ point included is quite far from the limiting value. The additional flexibility in the fitting function by allowing a variable exponent is thus offset by the necessity of including an additional fitting point.

We have considered the possibility of extrapolating the results with the pc- n basis sets for estimating the basis set limits. For the HF energy, the convergence is exponential,^{15,16} which in the absence of theoretical or computational guidelines for choosing the parameters gives three point extrapolations schemes. Analogous to the situation for the correlation consistent basis set, we would expect that including low order basis sets in the extrapolation would lead to results of lower accuracy. Extrapolations based on pc-0, -1, and -2 basis sets are thus likely to be less successful than based on results from pc-1, -2, and -3, or pc-2, -3, and -4, basis sets. However, results from pc-3 calculations are already close to the limiting value due to the fast convergence of the HF energy. This illustrates that there is inherently less room for improving HF and DF results by extrapolation, than for correlation energies.

For the correlation energy, theory indicates that the maximum angular momentum L is a natural extrapolation parameter,⁸⁻¹¹ but no corresponding theoretical analysis is available for the HF energy. A straightforward extrapolation formula of the type shown in Eq. (1),

$$E_L = E_\infty + A e^{-BL}, \quad (1)$$

was in all cases found to give substantially better extrapolated results than a power from $(E_L = E_\infty + AL^{-B})$,⁴⁶ in agreement with the established exponential convergence.^{15,16} For the total energy, however, Eq. (1) tends to predict an E_∞ value which is below the numerical HF limit, and in most cases the error from the extrapolated value is worse than for the raw pc- n result.

An alternative extrapolation form can be derived by noting that the energy lowering in Figs. 1-3 depends approximately exponential on the number of functions. To the extent that the data can be represented by a straight line, the dependence is exactly exponential. The slope and intercept depend on the type of functions, and we may write an approximate energy function as

$$E(n_s, n_p, n_d, n_f, \dots) = E_\infty + A_s e^{-B_s n_s} + A_p e^{-B_p n_p} + A_d e^{-B_d n_d} + A_f e^{-B_f n_f} + \dots \quad (2)$$

We define an E_L function as including only functions with angular momentum up to L ,

$$\begin{aligned} E_{L=1}(n_s, n_p) &= E(n_s, n_p, 0, 0, \dots), \\ E_{L=2}(n_s, n_p, n_d) &= E(n_s, n_p, n_d, 0, 0, \dots), \\ E_{L=3}(n_s, n_p, n_d, n_f) &= E(n_s, n_p, n_d, n_f, 0, \dots). \end{aligned} \quad (3)$$

According to Eq. (2) this implies

$$\begin{aligned} E_{L=1}(n_s, n_p) &= E_\infty + A_s e^{-B_s n_s} + A_p e^{-B_p n_p} + A_d + A_f \\ &+ \dots, \\ E_{L=2}(n_s, n_p, n_d) &= E_\infty + A_s e^{-B_s n_s} + A_p e^{-B_p n_p} \\ &+ A_d e^{-B_d n_d} + A_f + \dots, \\ E_{L=3}(n_s, n_p, n_d, n_f) &= E_\infty + A_s e^{-B_s n_s} + A_p e^{-B_p n_p} \\ &+ A_d e^{-B_d n_d} + A_f e^{-B_f n_f} + \dots. \end{aligned} \quad (4)$$

For the pc- n basis sets, the principles of construction suggest that terms corresponding to $n_i=0$ are small compared to terms with $n_i \neq 0$, and furthermore, the $n_i \neq 0$ terms are of comparable magnitude, i.e., for $E_{L=1}$,

$$\begin{aligned} E_{L=1}(n_s, n_p) &= E_\infty + A_s e^{-B_s n_s} + A_p e^{-B_p n_p} + A_d + A_f \\ &+ \dots, \\ A_d, A_f &\gg A_s e^{-B_s n_s}, A_p e^{-B_p n_p}, \\ A_s e^{-B_s n_s} &\approx A_p e^{-B_p n_p}, \\ E_{L=1}(n_s) &\approx E_\infty + 2A_s e^{-B_s n_s}. \end{aligned} \quad (5)$$

Generalization of Eq. (5) suggests an extrapolation scheme of the form,

$$E_L(n_s) \approx E_\infty + A(L+1)e^{-B n_s}. \quad (6)$$

Equation (6) has an explicit dependence on the number of s -functions (and thereby indirectly also on the number of p -, d -, etc. functions), as well as the maximum angular momentum functions included in the basis set.

Although the data in Figs. 1-3 are reasonably well fitted by straight lines, there is a small but noticeable curvature. A somewhat better representation can be obtained by substituting n_i by the corresponding square root in Eq. (2),^{16,47} leading to Eq. (7),

$$E_L(n_s) \approx E_\infty + A(L+1)e^{-B\sqrt{n_s}}. \quad (7)$$

Extrapolation of total energies for the present systems (H_2 , C_2 , N_2 , O_2 , F_2 , NH , FH , CO , and NF) indicate that extrapolation in all cases improves the results. Furthermore, from the (limited) set of systems investigated it appears that the optimum B value is relatively constant, varying between 5.5 and 6.5. Using a standard value of 6 for B transforms Eq. (7) into a two-point extrapolation formula. Extrapolated results for the atomization energy with either a 2- or 3-point extrapolation by Eq. (7) are included in Fig. 5, with similar results for CO , FH , and NF . Extrapolation in all cases improves the agreement with the limiting value.

VII. CONTRACTION OF THE pc- n BASIS SETS

The size and composition of the basis set determines the limiting accuracy that can be obtained. In order to improve the computational efficiency, however, the primitive basis is normally contracted. Contraction will always involve some degradation in accuracy, and the question is how much loss in accuracy that is acceptable for a given improvement in performance. We note that Hartree-Fock and density func-

tional calculations are dominated by integral evaluations, and the computational saving by contraction is not as large as for correlated calculations.

For an isolated atom from the first row in the Periodic Table any basis set can be contracted to $2s1p$ without reduction in accuracy at the HF level by using the MO-coefficients for the occupied $1s$ -, $2s$ -, and $2p$ -orbitals. For molecules, however, this will introduce an error since the shape of the atomic orbitals change upon bond formation. The change is largest for the valence orbitals which are involved in bonding. The basis functions which primarily describe the core electrons can be contracted with little loss in accuracy.

We have investigated contraction of the pc- n basis sets by a general contraction scheme using the HF coefficients for the isolated atoms as contraction coefficients. The general procedure is to initially contract the sp -basis functions to $2s1p$, leaving the polarization functions uncontracted. The outer s - and p -functions are then sequentially removed from the contraction until the contraction error becomes acceptable. The acceptable error depends on the size of the underlying primitive basis set. The pc-0 basis set gives rather crude results, and a relatively large contraction error is thus acceptable. The pc-4 basis set, however, gives results of high accuracy (compared to the infinite limit), and only a small contraction error is therefore acceptable. A strong contraction of the pc-4 primitive basis set will result in a performance worse than the uncontracted pc-3 basis set, at a higher computational cost, and this is clearly undesirable.

The error relative to the limiting value for the atomization energy, equilibrium distance, and dipole moment for the NH molecules are given in Table XVI, and similar data for CO, FH, and NF are given as supplementary material.³⁵ For the pc-0 basis, which is $5s3p$ in terms of primitive functions, a contraction to a minimal $2s1p$ basis set increases the error by a factor of more than 3, which is unacceptable. A $3s2p$ contraction to a double zeta-type basis set degrades the results by an acceptable $\sim 10\%$.

A $3s2p1d$ contraction of the pc-1 $7s4p1d$ primitive basis set results in a quite large contraction error, $\sim 30\%$ for the atomization energy and up to a factor of two for the equilibrium distance. A $4s3p1d$ contraction gives much lower contraction errors, less than 2%, at the expense of increasing the number of independent basis functions. Given the inherent (in)accuracy of the uncontracted $7s4p1d$ basis set, the $3s2p1d$ contraction is recommended for computational efficiency.

The pc-2 basis set is $10s6p2d1f$ in its uncontracted version. A $3s2p$ contraction produces unacceptable errors, giving results worse than the uncontracted pc-1 basis set. A $4s3p$ contraction gives errors which are a reasonable compromise between accuracy and computational efficiency.

The uncontracted pc-3 basis set is $14s9p4d2f1g$ in composition. Since results with this basis set are approaching the limiting values, the acceptable contraction error is correspondingly smaller. Furthermore, the reduction in contraction error as more functions are left uncontracted is relatively slow. This partly reflects the fact that the basis function exponents get closer together as the size of the basis is increased. Taken in isolation, either a $5s4p$, $6s5p$ or $7s6p$

TABLE XVI. Contraction errors in the total atomization energy, equilibrium distance, and dipole moment relative to the HF limit for the NH molecule. The recommended contractions are underlined.

Basis set	M	ΔE_{atom} (a.u.)	ΔR_e (Å)	$\Delta\mu$ (D)	
HF-limit ^a		-0.077 651	1.017 313	1.6133	
pc-0	$5s3p$	17	0.016 581	0.027 843	0.2479
	$2s1p$	6	0.052 434	0.140 942	-0.2161
	<u>$3s2p$</u>	11	0.018 020	0.031 327	0.2470
pc-1	$7s4p1d$	31	0.004 503	0.005 210	0.0844
	<u>$3s2p1d$</u>	19	0.006 010	0.010 785	0.0560
	$4s3p1d$	24	0.004 560	0.005 235	0.0860
pc-2	$10s6p2d1f$	61	0.000 866	0.000 546	0.0636
	$3s2p2d1f$	39	0.004 602	0.014 579	0.0077
	<u>$4s3p2d1f$</u>	44	0.000 976	0.000 577	0.0651
	$5s4p2d1f$	49	0.000 895	0.000 501	0.0637
pc-3	$14s9p4d2f1g$	122	0.000 061	0.000 007	0.0082
	$4s3p4d2f1g$	88	0.000 458	0.001 500	0.0066
	$5s4p4d2f1g$	93	0.000 127	-0.000 031	0.0093
	<u>$6s5p4d2f1g$</u>	98	0.000 073	0.000 008	0.0082
	$7s6p4d2f1g$	103	0.000 061	0.000 008	0.0082
pc-4	$18s11p6d3f2g1h$	198	0.000 005	0.000 000	0.0006
	$5s4p6d3f2g1h$	157	0.000 079	-0.000 115	0.0014
	$6s5p6d3f2g1h$	162	0.000 034	-0.000 026	0.0008
	$7s6p6d3f2g1h$	167	0.000 008	0.000 002	0.0006
	<u>$8s7p6d3f2g1h$</u>	172	0.000 005	0.000 000	0.0006

^aReference 41. Error is defined as value-reference value. ΔE_{atom} and $\Delta\mu$ are for a fixed internuclear distance of 1.95 a.u.

contraction could be considered, however, we also want to be able to use the extrapolation procedure described above for improving the results. For extrapolation purposes it was found that the $6s5p$ contraction gave the best results, and this is therefore the recommended contraction.

The uncontracted pc-4 basis set is $18s11p6d3f2g1h$ in composition and produce results close to the limiting HF results. From the results in Table XVI it is clear that a $6s5p$ contraction of the $18s11p$ functions will produce results comparable in quality to the uncontracted pc-3 basis set, at a significantly higher computational cost. A $7s6p$ contraction gives fairly small contraction errors in an absolute sense, but quite large relative errors. Since the computational time will be dominated by the many polarization functions in the pc-4 basis set, and since such calculations primarily will be for calibration purposes, the recommended contraction is $8s7p6d4f2g1h$, which produces negligible contraction errors.

In terms of basis set exponents some of the polarization functions, especially the inner d -functions in the pc-3 and pc-4 basis sets, could also be considered as targets for contraction. Such a contraction would have to rely on information from molecular calculations, since the atomic energy is invariant to polarization functions. A sampling of some small molecules revealed that a contraction of polarization functions is not generally possible. A contraction which would give an acceptable error for one molecule produces large errors for other molecules, and vice versa.

The recommended contractions are underlined in Table XVI. Analogous to the uncontracted pc- n basis sets, the re-

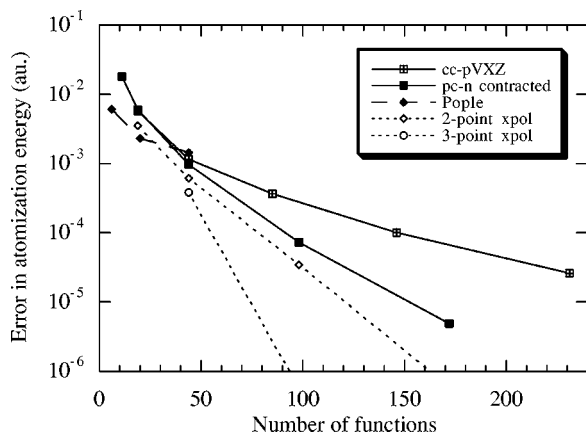


FIG. 8. Errors in atomization energy (a.u.) relative to the HF limit for NH for different contracted basis sets. Open symbols corresponds to either 2- or 3-point extrapolations based on the pc- n results and Eq. (7) in the text.

sults for the contracted basis sets can be extrapolated by Eq. (7), and the quality of the extrapolation is not changed significantly by the contraction. A comparison between the calculated atomization energies for NH for the contracted cc-pVXZ, pc- n and Pople style basis sets are shown in Fig. 8, which can be compared to the results in Fig. 5. For this system contraction actually improves the extrapolated results. Very similar results are obtained for the other three molecules, and the conclusions are identical to those for the uncontracted basis sets discussed above.

VIII. CONCLUSION

Based on analysis of the importance of polarization functions for Hartree–Fock energies of molecules, we propose a hierarchy of polarization consistent basis sets, suitable for systematically improving Hartree–Fock and density functional energies.³⁶ Compared to the correlation consistent basis sets, where the importance of the higher angular momentum functions decrease arithmetically, the polarization consistent basis set have the importance of the higher angular momentum functions decreasing roughly geometrically. Total energies, atomization energies, equilibrium distances, and dipole moments all converge monotonic towards the basis set limit. A general contraction scheme is proposed for improving the computational efficiency. It is possible to improve the energetic results by an exponential extrapolation. Based on these results polarization consistent basis sets appear to offer a faster convergence than other existing basis sets. Although the pc-3 and pc-4 type basis sets are quite large, continuing improvements in the computational efficiency of HF and DF methods, particularly linear scaling methods,⁴⁸ make calculations with large basis sets practical. We believe that it is important to have a systematic sequence of basis sets which are capable of approaching the limiting value for investigation of difficult systems and for development purposes.

ACKNOWLEDGMENTS

This work was supported by grants from the Danish Natural Science Research Council. The author is indebted to J. Kobus for making results from numerical Hartree–Fock calculations available.

- J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989).
- L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.* **94**, 7221 (1991).
- L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, *J. Chem. Phys.* **109**, 7764 (1998).
- J. A. Montgomery, Jr., J. W. Ochterski, and G. A. Petersson, *J. Chem. Phys.* **101**, 5900 (1994).
- J. W. Ochterski, G. A. Petersson, and J. A. Montgomery, Jr., *J. Chem. Phys.* **104**, 2598 (1996).
- J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, *J. Chem. Phys.* **110**, 2822 (1999).
- T. Helgaker, P. Jørgensen, and J. Olsen, *Modern Electronic-Structure Methods* (Wiley–Interscience, New York, 2000).
- C. Schwartz, *Phys. Rev.* **126**, 1015 (1962).
- R. N. Hill, *J. Chem. Phys.* **83**, 1173 (1985).
- W. Kutzelnigg, *Theor. Chim. Acta* **68**, 445 (1985).
- W. Kutzelnigg and J. D. Morgan III, *J. Chem. Phys.* **96**, 4484 (1992); **97**, 8821(E) (1992).
- T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- A. K. Wilson, T. van Mourik, and T. H. Dunning, Jr., *J. Mol. Struct.* **388**, 339 (1996).
- J. M. L. Martin and G. de Oliveira, *J. Chem. Phys.* **111**, 1843 (1999); S. Parthiban and J. M. L. Martin, *ibid.* **114**, 6014 (2001).
- F. Jensen, *J. Chem. Phys.* **110**, 6601 (1999).
- F. Jensen, *Theor. Chem. Acc.* **104**, 484 (2000).
- K. Aa. Christensen and F. Jensen, *Chem. Phys. Lett.* **317**, 400 (2000).
- R. Improta, V. Barone, K. N. Kudin, and G. E. Scuseria, *J. Chem. Phys.* **114**, 2541 (2001).
- J. Cioslowski, M. Schimeczek, G. Liu, and V. Stoyanov, *J. Chem. Phys.* **113**, 9377 (2000).
- G. Menconi, P. J. Wilson, and D. J. Tozer, *J. Chem. Phys.* **114**, 3958 (2001).
- A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, *Chem. Phys. Lett.* **302**, 437 (1999).
- A. Halkier, W. Klopper, T. Helgaker, and P. Jørgensen, *J. Chem. Phys.* **111**, 4424 (1999).
- A. Halkier, W. Klopper, T. Helgaker, P. Jørgensen, and P. R. Taylor, *J. Chem. Phys.* **111**, 9157 (1999).
- M. W. Schmidt and K. Ruedenberg, *J. Chem. Phys.* **71**, 3951 (1979).
- D. Moncrieff and S. Wilson, *J. Phys. B* **26**, 1605 (1993).
- D. Moncrieff and S. Wilson, *J. Phys. B* **27**, 1 (1994).
- D. Feller and E. R. Davidson, *Rev. Comput. Chem.* **1**, 1 (1990).
- T. Helgaker and P. R. Taylor, *Modern Electronic Structure Theory, Part II*, edited by D. Yarkony (World Scientific, Singapore, 1995), pp. 727–856.
- M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).
- J. Almlöf and P. R. Taylor, *Adv. Quantum Chem.* **22**, 301 (1991).
- R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **103**, 4572 (1995).
- M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 94, Gaussian, Inc., Pittsburgh, PA, 1995.
- T. Koga, S. Watanabe, K. Kanayama, R. Yasuda, and A. J. Thakkar, *J. Chem. Phys.* **103**, 3000 (1995).
- See EPAPS Document No. E-JCPSA6-115-305144 for supplementary tables and figures. This document may be retrieved via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>) or from <ftp.aip.org> in the directory /epaps. See the EPAPS homepage for more information.
- The present basis sets are not intended for general use. Basis sets specifically optimized for DFT applications along the present guidelines are currently being developed.
- J. Kobus, *Adv. Quantum Chem.* **28**, 1 (1997).
- W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.* **51**, 2657 (1969).

- ³⁹W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.* **56**, 2257 (1971); P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta* **28**, 213 (1973).
- ⁴⁰R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.* **72**, 650 (1980).
- ⁴¹J. Kobus (private communication).
- ⁴²A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, *Chem. Phys. Lett.* **286**, 243 (1998).
- ⁴³T. Helgaker, W. Klopper, H. Koch, and J. Noga, *J. Chem. Phys.* **106**, 9639 (1997).
- ⁴⁴J. S. Lee and S. Y. Park, *J. Chem. Phys.* **112**, 10746 (2000).
- ⁴⁵A. J. C. Varandas, *J. Chem. Phys.* **113**, 8880 (2000).
- ⁴⁶P. L. Fast, M. L. Sanches, and D. G. Truhlar, *J. Chem. Phys.* **111**, 2921 (1999).
- ⁴⁷W. Klopper and W. Kutzelnigg, *J. Mol. Struct.* **135**, 339 (1986).
- ⁴⁸S. Goedecker, *Rev. Mod. Phys.* **71**, 1085 (1999).